

1

The Practice of Medicine

The Editors

ENDURING VALUES OF THE MEDICAL PROFESSION

No greater opportunity, responsibility, or obligation can fall to the lot of a human being than to become a physician. In the care of the suffering, [the physician] needs technical skill, scientific knowledge, and human understanding. Tact, sympathy, and understanding are expected of the physician, for the patient is no mere collection of symptoms, signs, disordered functions, damaged organs, and disturbed emotions. [The patient] is human, fearful, and hopeful, seeking relief, help, and reassurance.

—Harrison's Principles of Internal Medicine, 1950

The practice of medicine has changed in significant ways since the first edition of this book was published in 1950. The advent of molecular genetics, sophisticated new imaging techniques, robotics, and advances in bioinformatics and information technology have contributed to an explosion of scientific information that has changed fundamentally the way physicians define, diagnose, treat, and attempt to prevent disease. This growth of scientific knowledge continues to evolve at an accelerated pace.

The widespread use of electronic medical records and the Internet have altered the way physicians and other health care providers access and exchange information as a routine part of medical education and practice (Fig. 1-1). As today's physicians strive to integrate an ever-expanding body of scientific knowledge into everyday practice, it is critically important to remember two key principles: first, the ultimate goal of medicine is to prevent disease and, when it occurs, to diagnose it early and provide effective treatment; and second, despite 70 years of scientific advances since the first edition of this text, a trusting relationship between physician and patient still lies at the heart of effective patient care.

THE SCIENCE AND ART OF MEDICINE

Deductive reasoning and applied technology form the foundation for the approach and solution to many clinical problems. Extraordinary advances in biochemistry, cell biology, immunology, and genomics,



FIGURE 1-1 *The Doctor* by Luke Fildes depicts the caring relationship between this Victorian physician and a very ill child. Painted in 1891, the painting reflects the death of the painter's young son from typhoid fever and was intended to reflect the compassionate care provided by the physician even when his tools were not able to influence the course of disease. (Source: History and Art Collection/Alamy Stock Photo.)

coupled with newly developed imaging techniques, provide a window into the most remote recesses of the body and allow access to the innermost parts of the cell. Revelations about the nature of genes and single cells have opened a portal for formulating a new molecular basis for the physiology of systems. Researchers are deciphering the complex mechanisms by which genes are regulated, and increasingly, physicians are learning how subtle changes in many different genes, acting in an integrative contextual way, can affect the function of cells and organisms. Clinicians have developed a new appreciation of the role of stem cells in normal tissue function, in the development of cancer and other disorders, and in the treatment of certain diseases. Entirely new areas of research, including studies of the human microbiome, epigenetics, and noncoding RNAs as regulatory features of the genome, have become important for understanding both health and disease. Information technology enables the interrogation of medical records from millions of individuals, yielding new insights into the etiology, characteristics, prognosis, and stratification of many diseases. With the increasing availability of very large data sets ("big data") from omic analyses and the electronic medical record, there is now a growing need for machine learning and artificial intelligence for unbiased analyses that enhance clinical predictive accuracy. The knowledge gleaned from the *science of medicine* continues to enhance the understanding by physicians of complex pathologic processes and to provide new approaches to disease prevention, diagnosis, and treatment. With continued refinement of unique omic signatures coupled with nuanced clinical pathophenotypes, the profession moves ever closer to practical precision medicine. Yet, skill in the most sophisticated applications of laboratory technology and in the use of the latest therapeutic modality alone does not make a good physician. Extraordinary advances in vaccine platform technology and the use of cryo-electron microscopy for the structure-based design of vaccine immunogens have transformed the field of vaccinology, resulting in the unprecedented speed and success with which COVID-19 vaccines were developed.

When a patient poses challenging clinical problems, an effective physician must be able to identify the crucial elements in a complex history and physical examination; order the appropriate laboratory, imaging, and diagnostic tests; and extract the key results from densely populated computer screens to determine whether to treat or to "watch." As the number of tests increases, so does the likelihood that some incidental finding, completely unrelated to the clinical problem at hand, will be uncovered. Deciding whether a clinical clue is worth pursuing or should be dismissed as a "red herring" and weighing whether a proposed test, preventive measure, or treatment entails a greater risk than the disease itself are essential judgments that a skilled clinician must make many times each day. This combination of medical knowledge, intuition, experience, and judgment defines the *art of medicine*, which is as necessary to the practice of medicine and the precision medicine of the future as is a sound scientific base, and as important for contemporary medical practice as it has been in earlier eras.

CLINICAL SKILLS

History-Taking The recorded history of an illness should include all the facts of medical significance in the life of the patient. Recent events should be given the most attention. Patients should, at some early point, have the opportunity to tell their own story of the illness without frequent interruption and, when appropriate, should receive expressions of interest, encouragement, and empathy from the physician. Any event related by a patient, however trivial or seemingly irrelevant, may provide the key to solving the medical problem. A methodical review of systems is important to elicit features of an underlying disease that might not be mentioned in the patient's narrative. In general, patients who feel comfortable with the physician will offer more complete information; thus, putting the patient at ease contributes substantially to obtaining an adequate history.

An informative history is more than eliciting an orderly listing of symptoms. By listening to patients and noting the ways in which they describe their symptoms, physicians can gain valuable insight. Inflections of voice, facial expression, gestures, and attitude (i.e., “body language”) may offer important clues to patients’ perception of and reaction to their symptoms. Because patients vary considerably in their medical sophistication and ability to recall facts, the reported medical history should be corroborated whenever possible. The social history also can provide important insights into the types of diseases that should be considered and can identify practical considerations for subsequent management. The family history not only identifies rare genetic disorders or common exposures, but often reveals risk factors for common disorders, such as coronary heart disease, hypertension, autoimmunity, and asthma. A thorough family history may require input from multiple relatives to ensure completeness and accuracy. An experienced clinician can usually formulate a relevant differential diagnosis from the history alone, using the physical examination and diagnostic tests to narrow the list or reveal unexpected findings that lead to more focused inquiry.

The very act of eliciting the history provides the physician with an opportunity to establish or enhance a unique bond that can form the basis for a good patient–physician relationship. This process helps the physician develop an appreciation of the patient’s view of the illness, the patient’s expectations of the physician and the health care system, and the financial and social implications of the illness for the patient. Although current health care settings may impose time constraints on patient visits, it is important not to rush the encounter. A hurried approach may lead patients to believe that what they are relating is not of importance to the physician, and, as a result, they may withhold relevant information. The confidentiality of the patient–physician relationship cannot be overemphasized.

Physical Examination The purpose of the physical examination is to identify physical signs of disease. The significance of these objective indications of disease is enhanced when they confirm a functional or structural change already suggested by the patient’s history. At times, however, physical signs may be the only evidence of disease and may not have been suggested by the history.

The physical examination should be methodical and thorough, with consideration given to the patient’s comfort and modesty. Although attention is often directed by the history to the diseased organ or part of the body, the examination of a new patient must extend from head to toe in an objective search for abnormalities. The results of the examination, like the details of the history, should be recorded at the time they are elicited—not hours later, when they are subject to the distortions of memory. Physical examination skills should be learned under direct observation of experienced clinicians. Even highly experienced clinicians can benefit from ongoing coaching and feedback. Simulation laboratories and standardized patients play an increasingly important role in the development of clinical skills. Although the skills of physical diagnosis are acquired with experience, it is not merely technique that determines success in identifying signs of disease. The detection of a few scattered petechiae, a faint diastolic murmur, or a small mass in the abdomen is not a question of keener eyes and ears or more sensitive fingers, but of a mind alert to those findings. Because physical findings can change with time, the physical examination should be repeated as frequently as the clinical situation warrants.

Given the many highly sensitive diagnostic tests now available (particularly imaging techniques), it may be tempting to place less emphasis on the physical examination. Some are critical of physical diagnosis based on perceived low levels of specificity and sensitivity. Indeed, many patients are seen by consultants only after a series of diagnostic tests have been performed and the results are known. This fact should not deter the physician from performing a thorough physical examination since important clinical findings may have escaped detection by diagnostic tests. Especially important, a thorough and thoughtful physical examination may render a laboratory finding unimportant (i.e., certain echocardiographic regurgitant lesions). The act of a hands-on examination of the patient also offers an opportunity

for communication and may have reassuring effects that foster the patient–physician relationship.

Diagnostic Studies Physicians rely increasingly on a wide array of laboratory and imaging tests to make diagnoses and ultimately to solve clinical problems; however, such information does not relieve the physician from the responsibility of carefully observing and examining the patient. It is also essential to appreciate the limitations of diagnostic tests. By virtue of their apparent precision, these tests often gain an aura of certainty regardless of the fallibility of the tests themselves, the instruments used in the tests, and the individuals performing or interpreting the tests. Physicians must weigh the expense involved in laboratory procedures against the value of the information these procedures are likely to provide.

Single laboratory tests are rarely ordered. Instead, physicians generally request “batteries” of multiple tests, which often prove useful and can be performed with a single specimen at relatively low cost. For example, abnormalities of hepatic function may provide the clue to nonspecific symptoms such as generalized weakness and increased fatigability, suggesting a diagnosis of chronic liver disease. Sometimes a single abnormality, such as an elevated serum calcium level, points to a particular disease, such as hyperparathyroidism.

The thoughtful use of screening tests (e.g., measurement of low-density lipoprotein cholesterol) may allow early intervention to prevent disease ([Chap. 6](#)). Screening tests are most informative when they are directed toward common diseases and when their results indicate whether other potentially useful—but often costly—tests or interventions are needed. On the one hand, biochemical measurements, together with simple laboratory determinations such as routine serum chemistries, blood counts, and urinalysis, often provide a major clue to the presence of a pathologic process. On the other hand, the physician must learn to evaluate occasional screening-test abnormalities that do not necessarily connote significant disease. An in-depth workup after the report of an isolated laboratory abnormality in a person who is otherwise well is often wasteful and unproductive. Because so many tests are performed routinely for screening purposes, it is not unusual for one or two values to be slightly abnormal. Nevertheless, even if there is no reason to suspect an underlying illness, tests yielding abnormal results ordinarily are repeated to rule out laboratory error. If an abnormality is confirmed, it is important to consider its potential significance in the context of the patient’s condition and other test results.

There is almost continual development of technically improved imaging studies with greater sensitivity and specificity. These tests provide remarkably detailed anatomic information that can be pivotal in informing medical decision-making. MRI, CT, ultrasonography, a variety of isotopic scans, and positron emission tomography (PET) have supplanted older, more invasive approaches and opened new diagnostic vistas. In light of their capabilities and the rapidity with which they can lead to a diagnosis, it is tempting to order a battery of imaging studies. All physicians have had experiences in which imaging studies revealed findings that led to an unexpected diagnosis. Nonetheless, patients must endure each of these tests, and the added cost of unnecessary testing is substantial. Furthermore, investigation of an unexpected abnormal finding may lead to an iatrogenic complication or to the diagnosis of an irrelevant or incidental problem. A skilled physician must learn to use these powerful diagnostic tools judiciously, always considering whether the results will alter management and benefit the patient.

■ MANAGEMENT OF PATIENT CARE

Team-Based Care Medical practice has long involved teams, particularly physicians working with nurses and, more recently, with physician assistants and nurse practitioners. Advances in medicine have increased our ability to manage very complex clinical situations (e.g., intensive care units [ICUs], bone marrow transplantation) and have shifted the burden of disease toward chronic illnesses. Because an individual patient may have multiple chronic diseases, he or she may be cared for by several specialists as well as a primary care physician. In the inpatient setting, care may involve multiple consultants along with

the primary admitting physician. Communication through the medical record is necessary but not sufficient, particularly when patients have complex medical problems or when difficult decisions need to be made about the optimal management plan. Physicians should optimally meet face-to-face or by phone to ensure clear communication and thoughtful planning. It is important to note that patients often receive or perceive different messages from various care providers; thus, attempts should be made to provide consistency among these messages to the patient. Management plans and treatment options should be outlined succinctly and clearly for the patient.

Another dimension of team-based care involves allied health professionals. It is not unusual for a hospitalized patient to encounter physical therapists, pharmacists, respiratory therapists, radiology technicians, social workers, dieticians, and transport personnel (among others) in addition to physicians and nurses. Each of these individuals contributes to clinical care as well as to the patient's experience with the health care system. In the outpatient setting, disease screening and chronic disease management are often carried out by nurses, physician assistants, or other allied health professionals.

The growth of team-based care has important implications for medical culture, student and resident training, and the organization of health care systems. Despite diversity in training, skills, and responsibilities among health care professionals, common values need to be espoused and reinforced. Many medical schools have incorporated interprofessional teamwork into their curricula. Effective communication is inevitably the most challenging aspect of implementing team-based care. While communication can be aided by electronic devices, including medical records, apps, or text messages, it is vitally important to balance efficiency with taking the necessary time to speak directly with colleagues.

The Dichotomy of Inpatient and Outpatient Internal Medicine The hospital environment has undergone sweeping changes over the past few decades. Emergency departments and critical care units have evolved to manage critically ill patients, allowing them to survive formerly fatal conditions. In parallel, there is increasing pressure to reduce the length of stay in the hospital and to manage complex disorders in the outpatient setting. This transition has been driven not only by efforts to reduce costs but also by the availability of new outpatient technologies, such as imaging and percutaneous infusion catheters for long-term antibiotics or nutrition, minimally invasive surgical procedures, and evidence that outcomes often are improved by reducing inpatient hospitalization.

In addition to traditional medical beds, hospitals now encompass multiple distinct levels of care, such as the emergency department, procedure rooms, overnight observation units, critical care units, and palliative care units. A consequence of this differentiation has been the emergence of new specialties (e.g., emergency medicine and end-of-life care) and the provision of in-hospital care by hospitalists and intensivists. Most *hospitalists* are board-certified internists who bear primary responsibility for the care of hospitalized patients and whose work is limited entirely to the hospital setting. The shortened length of hospital stay means that most patients receive only acute care while hospitalized; the increased complexities of inpatient medicine make the presence of an internist with specific training, skills, and experience in the hospital environment extremely beneficial. *Intensivists* are board-certified physicians who are further certified in critical care medicine and who direct and provide care for very ill patients in critical care units. Clearly, an important challenge in internal medicine today is to ensure the continuity of communication and information flow between a patient's primary care physician and those who are in charge of the patient's hospital care. Maintaining these channels of communication is frequently complicated by patient "handoffs"—i.e., transitions from the outpatient to the inpatient environment, from the critical care unit to a general medicine floor, from a medical to a surgical service and vice versa, from the hospital environment to the recently developed "home hospital" setting (for select patients with adequate home support), and from the hospital or home hospital to the outpatient environment.

The involvement of many care providers in conjunction with these transitions can threaten the traditional one-to-one relationship between patient and primary care physician. Of course, patients can benefit greatly from effective collaboration among a number of health care professionals; however, *it is the duty of the patient's principal or primary physician to provide cohesive guidance through an illness*. To meet this challenge, primary care physicians must be familiar with the techniques, skills, and objectives of specialist physicians and allied health professionals who care for their patients in the hospital. In addition, primary care physicians must ensure that their patients benefit from scientific advances and the expertise of specialists, both in and out of the hospital. Primary care physicians should explain the role of these specialists to reassure patients that they are in the hands of physicians best trained to manage their current illness. However, the primary care physician should assure patients and their families that decisions are being made in consultation with these specialists. The evolving concept of the "medical home" incorporates team-based primary care with subspecialty care in a cohesive environment that ensures smooth transitions of care.

Mitigating the Stress of Acute Illness Few people are prepared for a new diagnosis of cancer or anticipate the occurrence of a myocardial infarction, stroke, or major accident. The care of a frightened or distraught patient is confounded by these understandable responses to life-threatening events. The physician and other health providers can reduce the shock of life-changing events by providing information in a clear, calm, consistent, and reassuring manner. Often, information and reassurance need to be repeated. Caregivers should also recognize that, for the typical patient, hospital emergency rooms, operating rooms, ICUs, and general medical floors represent an intimidating environment. Hospitalized patients find themselves surrounded by air jets, buttons, and glaring lights; invaded by tubes and wires; and beset by the numerous members of the health care team—hospitalists, specialists, nurses, nurses' aides, physician assistants, social workers, technologists, physical therapists, medical students, house officers, attending and consulting physicians, and many others. They may be transported to special laboratories and imaging facilities replete with blinking lights, strange sounds, and unfamiliar personnel; they may be left unattended at times; and they may be obligated to share a room with other patients who have their own health problems. It is little wonder that patients may find this environment bewildering and stressful. The additive effects of an acute illness, unfamiliar environment, multiple medications, and sleep deprivation can lead to confusion or delirium, especially in older hospitalized patients. Physicians who appreciate the hospital experience from the patient's perspective and who make an effort to guide the patient through this experience may make a stressful situation more tolerable and enhance the patient's chances for an optimal recovery.

Medical Decision-Making Medical decision-making is a fundamental responsibility of the physician and occurs at each stage of the diagnostic and therapeutic process. The decision-making process involves the ordering of additional tests, requests for consultations, decisions about treatment, and predictions concerning prognosis. This process requires an in-depth understanding of the pathophysiology and natural history of disease. Formulating a differential diagnosis requires not only a broad knowledge base but also the ability to assess the relative probabilities of various diseases for a given patient. Application of the scientific method, including hypothesis formulation and data collection, is essential to the process of accepting or rejecting a particular diagnosis. Analysis of the differential diagnosis is an iterative process. As new information or test results are acquired, the group of disease processes being considered can be contracted or expanded appropriately. Whenever possible, decisions should be evidence-based, taking advantage of rigorously designed clinical trials or objective comparisons of different diagnostic tests. *Evidence-based medicine* stands in sharp contrast to anecdotal experience, which is often biased. Unless attuned to the importance of using larger, objective studies for making decisions, even the most experienced physicians can be influenced

to an undue extent by recent encounters with selected patients. Evidence-based medicine has become an increasingly important part of routine medical practice and has led to the publication of many useful practice guidelines. It is important to remember, however, that only a small fraction of the many decisions made in clinical practice are based on rigorous clinical trial evidence; other guideline recommendations are, therefore, predicated on expert consensus and weaker evidentiary support.

Thus, the importance of evidence-based medicine notwithstanding, much medical decision-making still relies on good clinical judgment, an attribute that is difficult to quantify or even to assess qualitatively. Physicians must use their knowledge and experience as a basis for weighing known factors, along with the inevitable uncertainties, and then making a sound judgment; this synthesis of information is particularly important when a relevant evidence base is not available. Several quantitative tools may be invaluable in synthesizing the available information, including diagnostic tests, Bayes' theorem (the probability of an event predicated on prior knowledge of conditions possibly related to the event), and multivariate statistical models (*Chap. 4*). Diagnostic tests serve to reduce uncertainty about an individual's diagnosis or prognosis and help the physician decide how best to manage that individual's condition. The battery of diagnostic tests complements the history and physical examination. The accuracy of a particular test is ascertained by determining its sensitivity (true-positive rate) and specificity (true-negative rate), as well as the predictive value of a positive and a negative result. *See Chap. 4 for a more thorough discussion of decision-making in clinical medicine.*

Practice Guidelines Many professional organizations and government agencies have developed formal clinical-practice guidelines to aid physicians and other caregivers in making diagnostic and therapeutic decisions that are evidence-based, cost-effective, and most appropriate to a particular patient and clinical situation. As the evidence base of medicine increases, guidelines can provide a useful framework for managing patients with particular diagnoses or symptoms. Clinical guidelines can protect patients—particularly those with inadequate health care benefits—from receiving substandard care. These guidelines also can protect conscientious caregivers from inappropriate charges of malpractice and society from the excessive costs associated with the overuse of medical resources. There are, however, caveats associated with clinical-practice guidelines since they tend to oversimplify the complexities of medicine. Furthermore, groups with different perspectives may develop divergent recommendations regarding issues as basic as the need for screening of women by mammography or of men with serum prostate-specific antigen (PSA) measurements. Finally, guidelines, as the term implies, do not—and cannot be expected to—account for the uniqueness of each individual and his or her illness. The physician's challenge is to integrate into clinical practice the useful recommendations offered by experts without accepting them blindly or being inappropriately constrained by them.

Precision Medicine The concept of *precision* or *personalized medicine* reflects the growing recognition that diseases once lumped together can be further stratified on the basis of genetic, biomarker, phenotypic, and/or psychosocial characteristics that distinguish a given patient from other patients with similar clinical presentations. Inherent in this concept is the goal of targeting therapies in a more specific way to improve clinical outcomes for the individual patient and minimize unnecessary side effects for those less likely to respond to a particular treatment. In some respects, precision medicine represents the evolution of clinical practice guidelines, which are usually developed for populations of patients or a particular diagnosis (e.g., hypertension, thyroid nodule). As the pathobiology, prognosis, and treatment responses of subgroups within these diagnoses become better understood (i.e., through refined genomic analysis or enhanced deep phenotyping), the relevant clinical guidelines incorporate progressively more refined recommendations for individuals within these subgroups. The role of precision medicine is best illustrated for cancers in which genetic testing is able to predict responses (or the lack thereof) to

targeted therapies (*Chap. 73*). One can anticipate similar applications of precision medicine in pharmacogenomics, immunologic disorders, and diseases in which biomarkers can predict treatment responses. *See Chap. 5 for a more thorough discussion of precision medicine.*

Evaluation of Outcomes Clinicians generally use *objective* and readily measurable parameters to judge the outcome of a therapeutic intervention. These measures may oversimplify the complexity of a clinical condition as patients often present with a major clinical problem in the context of multiple complicating background illnesses. For example, a patient may present with chest pain and cardiac ischemia, but with a background of chronic obstructive pulmonary disease and renal insufficiency. For this reason, outcome measures, such as mortality, length of hospital stay, or readmission rates, are typically risk-adjusted. An important point to remember is that patients usually seek medical attention for *subjective* reasons; they wish to obtain relief from pain, to preserve or regain function, and to enjoy life. The components of a patient's health status or quality of life can include bodily comfort, capacity for physical activity, personal and professional function, sexual function, cognitive function, and overall perception of health. Each of these important domains can be assessed through structured interviews or specially designed questionnaires. Such assessments provide useful parameters by which a physician can judge patients' subjective views of their disabilities and responses to treatment, particularly in chronic illness. The practice of medicine requires consideration and integration of both objective and subjective outcomes.

Many health systems use survey and patient feedback data to assess qualitative features such as patient satisfaction, access to care, and communication with nurses and physicians. In the United States, HCAHPS (Hospital Consumer Assessment of Healthcare Providers and Systems) surveys are used by many systems and are publicly reported. Social media is also being used to assess feedback in real time as well as to share patient experiences with health care systems, potentially enriching the information available for use in medical decisions.

Errors in the Delivery of Health Care A series of reports from the Institute of Medicine (now the National Academy of Medicine [NAM]) called for an ambitious agenda to reduce medical error rates and improve patient safety by designing and implementing fundamental changes in health care systems (*Chap. 8*). It is the responsibility of hospitals and health care organizations to develop systems to reduce risk and ensure patient safety. Medication errors can be reduced through the use of ordering systems that rely on electronic processes or, when electronic options are not available, that eliminate misreading of handwriting. Whatever the clinical situation, it is the physician's responsibility to use powerful therapeutic measures wisely, with due regard for their beneficial actions, potential dangers, and cost. Implementation of infection control systems, enforcement of hand-washing protocols, and careful oversight of antibiotic use can minimize the complications of nosocomial infections. Central-line infection rates and catheter-associated urinary tract infections have been dramatically reduced at many centers by careful adherence of trained personnel to standardized protocols for introducing and maintaining central lines and urinary catheters, respectively. Rates of surgical infection and wrong-site surgery can likewise be reduced by the use of standardized protocols and checklists. Falls by patients can be minimized by judicious use of sedatives and appropriate assistance with bed-to-chair and bed-to-bathroom transitions. Taken together, these and other measures are saving thousands of lives each year.

Electronic Medical Records Both the growing reliance on computers and the strength of information technology now play central roles in medicine, including efforts to reduce medical errors. Laboratory data are accessed almost universally through computers. Many medical centers now have electronic medical records (EMRs), computerized order entry, and bar-coded tracking of medications. Some of these systems are interactive, sending reminders or warning of potential medical errors.

EMRs offer rapid access to information that is invaluable in enhancing health care quality and patient safety, including relevant data,

historical and clinical information, imaging studies, laboratory results, and medication records. These data can be used to monitor and reduce unnecessary variations in care and to provide real-time information about processes of care and clinical outcomes. Ideally, patient records are easily transferred across the health care system; however, technological limitations and concerns about privacy and cost continue to limit broad-based use of EMRs in many clinical settings.

For all of the advantages of EMRs, they can create distance between the physician and patient if care is not taken to preserve face-to-face contact. EMRs also require training and time for data entry. Many providers spend significant time entering information to generate structured data and to meet billing requirements. They may feel pressured to take short cuts, such as “cutting and pasting” parts of earlier notes into the daily record, thereby increasing the risk of errors. EMRs also structure information in a manner that disrupts the traditional narrative flow across time and among providers. These features, which may be frustrating for some providers, must be weighed against the advantages of ready access to past medical history, imaging, laboratory data, and consultant notes. Furthermore, the effort, time, and attention needed to maintain and utilize the EMR have led to a growing sense of dissatisfaction among physicians, lessening professional and personal well-being as a result. Clearly, this is an area of daily practice that requires improvement both for the delivery of safe and optimal care and physician wellness.

It is important to emphasize that information technology is merely a tool and can never replace the clinical decisions that are best made by the physician. Clinical knowledge and an understanding of a patient’s needs, supplemented by quantitative tools, still represent the best approach to decision-making in the practice of medicine.

THE PATIENT-PHYSICIAN RELATIONSHIP

The significance of the intimate personal relationship between physician and patient cannot be too strongly emphasized, for in an extraordinarily large number of cases both the diagnosis and treatment are directly dependent on it. One of the essential qualities of the clinician is interest in humanity, for the secret of the care of the patient is in caring for the patient.

—Francis W. Peabody, October 21, 1925,
Lecture at Harvard Medical School

Physicians must never forget that patients are individuals with problems that all too often transcend their physical complaints. They are not “cases” or “admissions” or “diseases.” Patients do not fail treatments; treatments fail to benefit patients. This point is particularly important in this era of high technology in clinical medicine. Most patients are anxious and fearful. Physicians should instill confidence and offer reassurance, but they must never come across as arrogant, patronizing, impatient, or hurried. A professional attitude, coupled with warmth and openness, can do much to alleviate anxiety and to encourage patients to share all aspects of their medical history. Empathy and compassion are the essential features of a caring physician. The physician needs to consider the setting in which an illness occurs—in terms not only of patients themselves but also of their familial, social, and cultural backgrounds. The ideal patient–physician relationship is based on thorough knowledge of the patient, mutual trust, and the ability to communicate.

Informed Consent The fundamental principles of medical ethics require physicians to act in the patient’s best interest and to respect the patient’s autonomy. Both principles are reflected in the process of informed consent. Patients are required to sign consent forms for most diagnostic or therapeutic procedures. Many patients possess limited medical knowledge and must rely on their physicians for advice. Communicating in a clear and understandable manner, physicians must fully discuss the alternatives for care and explain the risks, benefits, and likely consequences of each alternative. The physician is responsible for ensuring that the patient thoroughly understands these risks and benefits; encouraging questions is an important part of this process. It may be necessary to go over certain issues with the patient more than once. This is the very definition of *informed consent*. Complete, clear

explanation and discussion of the proposed procedures and treatment can greatly mitigate the fear of the unknown that commonly accompanies hospitalization. Often the patient’s understanding is enhanced by repeatedly discussing the issues in an unthreatening and supportive way, answering new questions that occur to the patient as they arise. Continuing efforts to educate the patient are essential. Patients are frequently inhibited from understanding by the fear of an uncertain future and potential impact of the illness on themselves and their families. Clear communication can also help alleviate misunderstandings in situations where complications of intervention occur. Special care should also be taken to ensure that a physician seeking a patient’s informed consent has no real or apparent conflict of interest.

Approach to Grave Prognoses and Death No circumstance is more distressing than the diagnosis of an incurable disease, particularly when premature death is inevitable. What should the patient and family be told? What measures should be taken to maintain life? What can be done to optimize quality of life?

Transparency of information, delivered in an appropriate manner, is essential in the face of a terminal illness. Even patients who seem unaware of their medical circumstances, or whose family members have protected them from diagnoses or prognoses, often have keen insights into their condition. They may also have misunderstandings that can lead to additional anxiety. The patient must be given an opportunity to speak with the physician and ask questions. A wise and insightful physician uses such open communication as the basis for assessing what the patient wants to know and when he or she wants to know it. On the basis of the patient’s responses, the physician can assess the most appropriate time and pace for sharing information. Ultimately, the patient must understand the expected course of the disease so that appropriate plans and preparations can be made. The patient should participate in decision-making with an understanding of the goal of treatment (palliation) and its likely effects. The patient’s religious beliefs should be taken into consideration. Some patients may find it easier to share their feelings about death with their physician, nurses, or members of the clergy than with family members or friends.

The physician should provide or arrange for emotional, physical, and spiritual support, and must be compassionate, unhurried, and open. In many instances, there is much to be gained by the laying on of hands. Pain should be controlled adequately, human dignity maintained, and isolation from family and close friends avoided. These aspects of care tend to be overlooked in hospitals, where the intrusion of life-sustaining equipment can detract from attention to the individual person and encourage concentration instead on the life-threatening disease, against which the battle ultimately will be lost in any case. In the face of terminal illness, the goal of medicine must shift from *cure* to *care* in the broadest sense of the term. *Primum succurrere*, first to help, is a guiding principle. In offering care to a dying patient, a physician should be prepared to provide information to family members and deal with their grief and sometimes their feelings of guilt or even anger. It is important for the physician to assure the family that everything reasonable is being done. A substantial challenge in these discussions is that the physician often does not know exactly how to gauge the prognosis. In addition, various members of the health care team may offer different opinions. Good communication among providers is essential so that consistent information is provided to patients. This is especially important when the best path forward is uncertain. Advice from experts in palliative and terminal care should be sought whenever appropriate to ensure that clinicians are not providing patients with unrealistic expectations. **For a more complete discussion of end-of-life care, see Chap. 12.**

Maintaining Humanism and Professionalism Many trends in the delivery of health care tend to make medical care impersonal. These trends, some of which have been mentioned already, include (1) vigorous efforts to reduce the escalating costs of health care; (2) the growing number of managed-care programs, which are intended to reduce costs but where the patient may have little choice in selecting a physician; (3) increasing reliance on technological advances and

computerization; and (4) the need for numerous physicians and other health professionals to be involved in the care of most patients who are seriously ill.

In light of these changes in the medical care system, it is a major challenge for physicians to maintain the *humane* aspects of medical care. The American Board of Internal Medicine, working together with the American College of Physicians–American Society of Internal Medicine and the European Federation of Internal Medicine, has published a *Charter on Medical Professionalism* that underscores three main principles in physicians' contract with society: (1) the primacy of patient welfare, (2) patient autonomy, and (3) social justice. While medical schools appropriately place substantial emphasis on professionalism, a physician's personal attributes, including integrity, respect, and compassion, also are extremely important. In the United States, the Gold Humanism Society recognizes individuals who are exemplars of humanistic patient care and serve as role models for medical education and training.

Availability to the patient, expression of sincere concern, willingness to take the time to explain all aspects of the illness, and a nonjudgmental attitude when dealing with patients whose cultures, lifestyles, attitudes, and values differ from those of the physician are just a few of the characteristics of a humane physician. Every physician will, at times, be challenged by patients who evoke strongly negative or positive emotional responses. Physicians should be alert to their own reactions to such situations and should consciously monitor and control their behavior so that the patient's best interest remains the principal motivation for their actions at all times.

Another important aspect of patient care involves an appreciation of the patient's "quality of life," a subjective assessment of what each patient values most. This assessment requires detailed, sometimes intimate knowledge of the patient, which usually can be obtained only through deliberate, unhurried, and often repeated conversations. Time pressures will always threaten these interactions, but they should not diminish the importance of understanding and seeking to fulfill the priorities of the patient.

■ EXPANDING FRONTIERS IN MEDICAL PRACTICE

The Era of "Omics" In the spring of 2003, announcement of the complete sequencing of the human genome officially ushered in the genomic era. However, even before that landmark accomplishment, the practice of medicine had been evolving as a result of insights into both the human genome and the genomes of a wide variety of microbes. The clinical implications of these insights are illustrated by the complete genome sequencing of H1N1 influenza virus in 2009 and even faster sequencing of COVID-19 in early 2020, leading to the swift development and dissemination of effective vaccines. Today, gene expression profiles are being used to guide therapy and inform prognosis for a number of diseases, and genotyping is providing a new means to assess the risk of certain diseases as well as variations in response to a number of drugs. Despite these advances, the use of complex genomics in the diagnosis, prevention, and treatment of disease is still in its early stages. The task of physicians is complicated by the fact that phenotypes generally are determined not by genes alone but by the complex interactions among genes and gene products, and by the interplay of genetic and environmental factors.

Rapid progress is also being made in other areas of molecular medicine. *Epigenetics* is the study of alterations in chromatin and histone proteins and methylation of DNA sequences that influence gene expression (Chap. 483). Every cell of the body has identical DNA sequences; the diverse phenotypes a person's cells manifest are, in part, the result of epigenetic regulation of gene expression. Epigenetic alterations are associated with a number of cancers and other diseases. *Proteomics*, the study of the entire library of proteins made in a cell or organ and the complex relationship of these proteins to disease, is enhancing the repertoire of the 23,000 genes in the human genome through alternate splicing, posttranslational processing, and posttranslational modifications that often have unique functional consequences. The presence or absence of particular proteins in the circulation or in cells is being explored for many diagnostic and disease-screening

applications. *Microbiomics* is the study of the resident microbes in humans and other mammals, which together compose the microbiome. The human haploid genome has ~23,000 genes, whereas the microbes residing on and in the human body encompass more than 3–4 million genes; these resident microbes are likely to be of great significance with regard to health status. Ongoing research is demonstrating that the microbes inhabiting human mucosal and skin surfaces play a critical role in maturation of the immune system, in metabolic balance, in brain function, and in disease susceptibility. A variety of environmental factors, including the use and overuse of antibiotics, have been tied experimentally to substantial increases in disorders such as obesity, metabolic syndrome, atherosclerosis, and immune-mediated diseases in both adults and children. *Metagenomics*, of which microbiomics is a part, is the genomic study of environmental species that have the potential to influence human biology directly or indirectly. An example is the study of exposures to microorganisms in farm environments that may be responsible for the lower incidence of asthma among children raised on farms. *Metabolomics* is the study of the range of metabolites in cells or organs and the ways they are altered in disease states. The aging process itself may leave telltale metabolic footprints that allow the prediction (and possibly the prevention) of organ dysfunction and disease. It seems likely that disease-associated patterns will be found in lipids, carbohydrates, membranes, mitochondria and mitochondrial function, and other vital components of cells and tissues. *Exposomics* is the study of the exposome—i.e., the environmental exposures such as smoking, sunlight, diet, exercise, education, and violence that together have an enormous impact on health. All of this new information represents a challenge to the traditional reductionist approach to medical thinking. The variability of results in different patients, together with the large number of variables that can be assessed, creates challenges in identifying preclinical disease and defining disease states unequivocally. Accordingly, the tools of *systems biology* and *network medicine* are being applied to the enormous body of information ("big data") now obtainable for every patient and may eventually provide new approaches to classifying disease. **For a more complete discussion of a complex systems and network science approach to human disease, see Chap. 486.**

The rapidity of these advances may seem overwhelming to practicing physicians; however, physicians have an important role to play in ensuring that these powerful technologies and sources of new information are applied judiciously to patient care. Since omics are evolving so rapidly, physicians and other health care professionals must engage in continuous learning so that they can apply this new knowledge to the benefit of their patients' health and well-being. Genetic testing requires wise counsel based on an understanding of the value and limitations of the tests as well as the implications of their results for specific individuals. **For a more complete discussion of genetic testing, see Chap. 467.**

The Globalization of Medicine Physicians should be cognizant of diseases and health care services beyond local boundaries. Global travel has critical implications for disease spread, and it is not uncommon for diseases endemic to certain regions to be seen in other regions after a patient has traveled to and returned from those regions. The outbreak of Zika virus infections in the Americas is a cogent example of this phenomenon. In addition, factors such as wars, the migration of refugees, and increasing climate extremes are contributing to changing disease profiles worldwide. Patients have broader access to unique expertise or clinical trials at distant medical centers, even those in other countries, and the cost of travel may be offset by the quality of care at those distant locations. As much as any other factor influencing global aspects of medicine, the Internet has transformed the transfer of medical information throughout the world. This change has been accompanied by the transfer of technological skills through telemedicine and international consultation—for example, interpretation of radiologic images and pathologic specimens. **For a complete discussion of global issues, see Chap. 472.**

Medicine on the Internet On the whole, the Internet has had a positive effect on the practice of medicine; through personal computers, a wide range of information is available to physicians and patients

almost instantaneously at any time and from anywhere in the world. This medium holds enormous potential for the delivery of current information, practice guidelines, state-of-the-art conferences, journal content, textbooks (including this text), and direct communications with other physicians and specialists, expanding the depth and breadth of information available to the physician regarding the diagnosis and care of patients. Medical journals are now accessible online, providing rapid sources of new information. By bringing them into direct and timely contact with the latest developments in medical care, this medium also serves to lessen the information gap that has hampered physicians and health care providers in remote areas.

Patients, too, are turning to the Internet in increasing numbers to acquire information about their illnesses and therapies and to join Internet-based support groups. Patients often arrive at a clinic visit with sophisticated information about their illnesses. In this regard, physicians are challenged in a positive way to keep abreast of the latest relevant information while serving as an “editor” as patients navigate this seemingly endless source of information, the accuracy and validity of which are not uniform.

A critically important caveat is that virtually anything can be published on the Internet, with easy circumvention of the peer-review process that is an essential feature of academic publications. Both physicians and patients who search the Internet for medical information must be aware of this danger. Notwithstanding this limitation, appropriate use of the Internet is revolutionizing information access for physicians and patients, and in this regard represents a remarkable resource that was not available to practitioners a generation ago.

Public Expectations and Accountability The general public’s level of knowledge and sophistication regarding health issues has grown rapidly over the past few decades. As a result, expectations of the health care system in general and of physicians in particular have risen. Physicians are expected to master rapidly advancing fields (the *science* of medicine) while considering their patients’ unique needs (the *art* of medicine). Thus, physicians are held accountable not only for the technical aspects of the care they provide but also for their patients’ satisfaction with the delivery and costs of care.

In many parts of the world, physicians increasingly are expected to account for the way in which they practice medicine by meeting certain standards prescribed by federal and local governments. The hospitalization of patients whose health care costs are reimbursed by the government and other third parties is subjected to utilization review. Thus, a physician must defend the cause for and duration of a patient’s hospitalization if it falls outside certain “average” standards. Authorization for reimbursement increasingly is based on documentation of the nature and complexity of an illness, as reflected by recorded elements of the history and physical examination. A growing “pay-for-performance” movement seeks to link reimbursement to quality of care. The goal of this movement is to improve standards of health care and contain spiraling health care costs. In many parts of the United States, managed (capitated) care contracts with insurers have replaced traditional fee-for-service care, placing the onus of managing the cost of all care directly on the providers and increasing the emphasis on preventive strategies. In addition, physicians are expected to give evidence of their current competence through mandatory continuing education, patient record audits, maintenance of certification, and relicensing.

Medical Ethics and New Technologies The rapid pace of technological advances has profound implications for medical applications that go far beyond the traditional goals of disease prevention, treatment, and cure. Cloning, genetic engineering, gene therapy, human-computer interfaces, nanotechnology, and use of targeted therapies have the potential to modify inherited predispositions to disease, select desired characteristics in embryos, augment “normal” human performance, replace failing tissues, and substantially prolong life span. Given their unique training, physicians have a responsibility to help shape the debate on the appropriate uses of and limits placed on these new technologies and to consider carefully the ethical issues associated with the implementation of such interventions. As medicine becomes more complex, shared decision-making is increasingly important, not

only in areas such as genetic counseling and end-of-life care, but also in diagnostic and treatment options.

Learning Medicine More than a century has passed since the publication of the Flexner Report, a seminal study that transformed medical education and emphasized the scientific foundations of medicine as well as the acquisition of clinical skills. In an era of burgeoning information and access to medical simulation and informatics, many schools are implementing new curricula that emphasize lifelong learning and the acquisition of competencies in teamwork, communication skills, system-based practice, and professionalism. The tools of medicine also change continuously, necessitating formal training in the use of EMRs, large datasets, ultrasound, robotics, and new imaging techniques. These and other features of the medical school curriculum provide the foundation for many of the themes highlighted in this chapter and are expected to allow physicians to progress, with experience and learning over time, from competency to proficiency to mastery.

At a time when the amount of information that must be mastered to practice medicine continues to expand, increasing pressures both within and outside of medicine have led to the implementation of restrictions on the amount of time a physician-in-training can spend in the hospital and in clinics. Because the benefits associated with continuity of medical care and observation of a patient’s progress over time were thought to be outstripped by the stresses imposed on trainees by long hours and by fatigue-related errors, strict limits were set on the number of patients that trainees could be responsible for at one time, the number of new patients they could evaluate in a day on call, and the number of hours they could spend in the hospital. In 1980, residents in medicine worked in the hospital more than 90 hours per week on average. In 1989, their hours were restricted to no more than 80 per week. Resident physicians’ hours further decreased by ~10% between 1996 and 2008, and in 2010, the Accreditation Council for Graduate Medical Education further restricted (i.e., to 16 hours per shift) consecutive in-hospital duty hours for first-year residents. The impact of these changes is still being assessed, but the evidence that medical errors have decreased as a consequence is sparse. An unavoidable by-product of fewer hours at the bedside is an increase in the number of “handoffs” of patient responsibility from one physician to another. These transfers often involve a transition from a physician who knows the patient well, having evaluated that individual on admission, to a physician who knows the patient less well. It is imperative that these transitions of responsibility be handled with care and thoroughness, with all relevant information exchanged and acknowledged. These issues highlight the challenge our profession has in establishing a reliable measure of physician effectiveness.

The Physician as Perpetual Student From the time physicians graduate from medical school, it becomes all too apparent that this milestone is symbolic and that they must embrace the role of a “perpetual student.” This realization is at the same time exhilarating and anxiety-provoking. It is exhilarating because physicians can apply constantly expanding knowledge to the treatment of their patients; it is anxiety-provoking because physicians realize that they will never know as much as they want or need to know. Ideally, physicians will translate the latter feeling into energy through which they can continue to improve and reach their potential. It is the physician’s responsibility to pursue new knowledge continually by reading, attending conferences and courses, and consulting colleagues and the Internet. This is often a difficult task for a busy practitioner; however, a commitment to continued learning is an integral part of being a physician and must be given the highest priority.

The Physician as Citizen Being a physician is a privilege. The capacity to apply one’s skills for the benefit of fellow human beings is a noble calling. The physician-patient relationship is inherently unbalanced in the distribution of power. In light of their influence, physicians must always be aware of the potential impact of what they do and say, and must always strive to strip away individual biases and preferences to find what is best for their patients. To the extent possible, physicians should also act within their communities to promote health

and alleviate suffering. Meeting these goals begins by setting a healthy example and continues in taking action to deliver needed care even when personal financial compensation may not be available.

Research, Teaching, and the Practice of Medicine The word *doctor* is derived from the Latin *docere*, “to teach.” As teachers, physicians should share information and medical knowledge with colleagues, students of medicine and related professions, and their patients. The practice of medicine is dependent on the sum total of medical knowledge, which in turn is based on an unending chain of scientific discovery, clinical observation, analysis, and interpretation. Advances in medicine depend on the acquisition of new information through research, and improved medical care requires the transmission of that information. As part of their broader societal responsibilities, physicians should encourage patients to participate in ethical and properly approved clinical investigations if these studies do not impose undue hazard, discomfort, or inconvenience. Physicians engaged in clinical research must be alert to potential conflicts of interest between their research goals and their obligations to individual patients. The best interests of the patient must always take priority.

To wrest from nature the secrets which have perplexed philosophers in all ages, to track to their sources the causes of disease, to correlate the vast stores of knowledge, that they may be quickly available for the prevention and cure of disease—these are our ambitions.

—William Osler, 1849–1919

FURTHER READING

- CHESTO CC et al: Social media use in medical education: A systematic review. *Acad Med* 88:893, 2013.
- COOKE M et al: American medical education 100 years after the Flexner report. *N Engl J Med* 355:1339, 2006.
- EXCEL JL et al: Vaccine development for emerging infectious diseases. *Nat Med* 27:591, 2021.
- INSTITUTE OF MEDICINE: *Dying in America: Improving Quality and Honoring Individual Preferences Near the End of Life*. Washington, DC, National Academies Press, 2015.
- INSTITUTE OF MEDICINE: *Improving Diagnosis in Health Care*. Washington, DC, National Academies of Sciences, Engineering, and Medicine, 2015.
- LEVINE DM et al: Hospital-level care at home for acutely ill adults: A qualitative evaluation of a randomized controlled trial. *J Gen Intern Med* 36:1965, 2021.
- STERN DT, PAPADAKIS M: The developing physician—becoming a professional. *N Engl J Med* 355:1794, 2006.
- VICKREY BG et al: How neurologists think: A cognitive psychology perspective on missed diagnoses. *Ann Neurol* 67:425, 2010.
- WEST P et al: Intervention to promote physician well-being, job satisfaction, and professionalism. A randomized clinical trial. *JAMA Intern Med* 174:527, 2014.

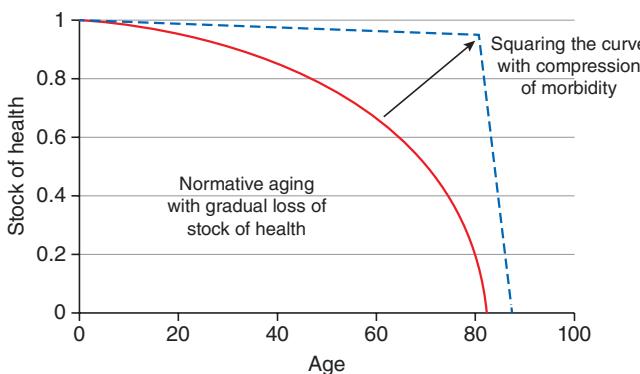


FIGURE 2-1 Loss of health with aging. Representation of normative aging with loss of the full stock of health with which individuals are born (indicating gain of morbidity), contrasted with a squared curve with greater longevity and fuller stock of health (less morbidity) until shortly before death. The “squared curve” represents the likely ideal situation for most patients.

worldwide over the last century (largely as a result of public health practices), increasing emphasis is placed on prevention for the purpose of preserving quality of life and extending the health span, not just the life span. Given that all patients will eventually die, the goal of prevention ultimately becomes compression of morbidity toward the end of the life span; that is, reduction of the amount of burden and time spent with disease prior to dying. As shown in Fig. 2-1, normative aging tends to involve a steady decline in the stock of health, with accelerating decline over time. Successful prevention offers the opportunity both to extend life and to extend healthy life, thus “squaring the curve” of health loss during aging.

Prevention strategies have been characterized as tertiary, secondary, primary, and primordial. *Tertiary prevention* requires rapid action to prevent imminent death in the setting of acute illness, such as through percutaneous coronary intervention in the setting of ST-segment elevation myocardial infarction. *Secondary prevention* strategies focus on avoiding the recurrence of disease and death in an individual who is already affected. For example, tamoxifen is recommended for women with surgically treated early-stage, estrogen receptor-positive breast cancer, because it reduces the risk of recurrent breast cancer (including in the contralateral breast) and death. *Primary prevention* attempts to reduce the risk of incident disease among individuals with one or more risk factors. Treatment of elevated blood pressure in individuals who have not yet experienced cardiovascular disease represents one example of primary prevention that has proven effective in reducing the incidence of stroke, heart failure, and coronary heart disease.

Primordial prevention is a more recent concept (first introduced in 1979) that focuses on prevention of the development of *risk factors* for disease, not just prevention of disease. Primordial prevention strategies emphasize upstream determinants of risk for chronic diseases, such as eating patterns, physical activity, and environmental and social determinants of health. It therefore encompasses medical treatment strategies for some individuals as well as a strong reliance on public health and social policy. It is increasingly clear that primordial prevention represents the ultimate means for reducing the burden of chronic diseases of aging. Once risk factors develop, it is difficult to restore risk to the low level of someone who never developed the risk factor. The time spent with adverse levels of the risk factor often causes irreversible damage that precludes complete restoration of low risk. For example, individuals with hypertension who are treated back to optimal levels ($<120/<80$ mmHg) do have a lower risk compared with untreated patients with hypertension, but they still have twice the risk of cardiovascular events as those who maintained optimal blood pressure without medications. Patients with elevated blood pressure that is subsequently treated have greater left ventricular mass index, worse renal function, and more evidence of atherosclerosis and other target organ damage as a result of the time spent with elevated blood pressure; such damage cannot be fully reversed despite efficacious therapy with antihypertensive medications. Conversely, as described below in greater detail, individuals who maintain optimal levels of all major

2

Promoting Good Health

Donald M. Lloyd-Jones,
Kathleen M. McKibbin

GOALS AND APPROACHES TO PREVENTION

Prevention of acute and chronic diseases before their onset has been recognized as one of the hallmarks of excellent medical practice for centuries and is now used as a metric for highly functioning health care systems. The ultimate goal of preventive strategies is to avoid premature death. However, as longevity has increased dramatically

cardiovascular risk factors into middle age through primordial prevention essentially abolish their lifetime risk of developing cardiovascular disease while also living substantially longer and having a lower burden and later onset of other comorbid illnesses (compression of morbidity).

Prevention strategies should be distinguished from disease screening strategies. Screening attempts to detect evidence of disease at its earliest stages, when treatment is likely to be more efficacious than for advanced disease (*Chap. 6*). Screening can be performed in service of prevention, especially if it aids in identifying preclinical markers, such as dyslipidemia or hyperglycemia, associated with elevated disease risk.

■ HEALTH PROMOTION

In recent decades, medical practice has increasingly focused on clinical and public health approaches to promote health, and not just prevent disease. Prevention of disease is a worthy individual and societal goal in and of itself, but it does not necessarily guarantee health. Health is a broader construct encompassing more than just absence of disease. It includes biologic, physiologic, and psychological domains (among others) in a continuum, rather than occurring as a dichotomous trait. Health is therefore somewhat subjective, but attempts have been made to use more objective criteria to define health in order to raise awareness, prevent disease, and promote healthy longevity.

For example, in 2010, the American Heart Association (AHA) defined a new construct of “cardiovascular health” based on evidence of associations with longevity, disease avoidance, healthy longevity, and quality of life. The definition of cardiovascular health is based on seven health behaviors and health factors (eating pattern, physical activity, smoking status, body mass index [BMI], and levels of blood pressure, blood cholesterol, and blood glucose) and includes a spectrum from poor to ideal. Individuals with optimal levels of all seven metrics simultaneously are considered to have ideal cardiovascular health. The state of cardiovascular health for an individual or a population can be assessed with simple scoring by counting the number of ideal metrics (out of seven) or applying 0 points for each poor metric, 1 point for each intermediate metric, and 2 points for each ideal metric, thus creating a composite cardiovascular health score ranging from 0 to 14 points. Higher cardiovascular health scores in younger and middle ages have been associated with greater longevity, lower incidence of cardiovascular disease, lower incidence of other chronic diseases of aging (including dementia, cancer, and more), compression of morbidity, greater quality of life, and lower health care costs, achieving both individual and societal goals for healthy aging and further establishing the critical importance of primordial prevention and cardiovascular health promotion.

Focusing on health promotion, rather than just disease prevention, may also provide greater motivation for patients to pursue lifestyle changes or adhere to clinician recommendations. Extensive literature suggests that providing patients solely with information regarding disease risk, or risk reduction with treatment, is unlikely to motivate desired behavior change. Empowering patients with strategies to achieve positive health goals after discussing risks can provide more effective adherence and better long-term outcomes. In the case of smoking cessation, enumerating only the risks of smoking can lead to patient inertia and therapeutic nihilism and has proven to be an ineffective approach, whereas strategies that incorporate positive health messaging, support, and feedback, with appropriate use of evidence-based therapies, have proven far more effective.

■ PRIORITIZING PREVENTION STRATEGIES

In secondary prevention, the patient already has manifest clinical disease and is therefore at high risk for progression. The approach should be to work with the patient to implement all evidence-based strategies that will help to prevent recurrence or progression. This will typically include drug therapy as well as therapeutic lifestyle changes to control ongoing risk factors that may have caused disease in the first place. Juggling priorities can be difficult, and barriers to implementation are many, including costs, time, patient health literacy, and patient and caregiver capacity to organize the regimen. Addressing these potential barriers with the patient can help to forge a therapeutic bond and

may improve adherence; ignoring them will likely lead to therapeutic failure. Numerous studies demonstrate that, even in high-functioning health systems, only ~50% of patients are taking recommended, evidence-based secondary prevention medications, such as statins, by 1 year after a myocardial infarction.

In patients who are eligible for primary prevention strategies, it is important to frame the discussion around the overall evidence base as well as an individual patient’s likelihood of benefit from a given preventive intervention. A first step is to understand the patient’s estimated absolute risk for disease in the foreseeable future or during their remaining life span. However, absolute risk estimation and presentation of those risks are generally insufficient to motivate behavior change. It is critical to assess the patient’s understanding and tolerance of the risk, their readiness to implement lifestyle changes or adhere to drug therapy, and their overall preferences regarding use of drug therapy to prevent an event (e.g., cancer, myocardial infarction, stroke). The clinician can help the patient by informing them of the risks for disease and potential for absolute benefits (and harms) from the available evidence-based choices. This may take more than one conversation, but given that diseases, such as cancer and cardiovascular disease, are the leading causes of premature death and disability, the time is well spent.

Partnering with the patient through motivational interviewing may assist in the process of selecting initial approaches to prevention. Selecting an area that the patient feels they are ready to change can lead to better adherence and greater achievement of success in the short and longer term. If the patient is uncertain what course to choose, prudence would dictate focusing on control of risk factors that may lead to the most rapid reduction in risk for acute events. For example, blood pressure is both a chronic risk factor and an acute trigger for cardiovascular events. Thus, if a patient has both significant elevations in blood pressure and dyslipidemia, it would be appropriate to focus initial efforts on blood pressure control. Likewise, a focus on smoking cessation can lead to more rapid reductions in risk for acute events than some other lifestyle interventions.

■ PREVENTION AND HEALTH PROMOTION ACROSS THE LIFE COURSE

Periodic Health Evaluations The “routine annual physical” has in many ways become an expected part of the patient-physician relationship in primary care practice. However, evidence for the efficacy of the periodic health evaluation in asymptomatic adults unselected for risk factors or disease is mixed and depends on the outcome. Systematic reviews and meta-analyses of published trials have consistently observed lack of benefit (and also lack of harm) in terms of total mortality in association with periodic health evaluations. Data are more heterogeneous but overall suggest no benefit for cancer- or cardiovascular-specific mortality, with the potential for either benefit or harm depending on number of evaluations and patient-level factors. Well-designed studies on nonfatal clinical events and morbidity have been sparsely reported, but there appear to be no large effects.

Periodic health evaluations do appear to lead to greater diagnosis of certain conditions such as hypertension and dyslipidemia, as expected. Likewise, periodic health examinations also improve the delivery of recommended preventive services, such as gynecologic examinations and Papanicolaou smears, fecal occult blood testing, and cholesterol screening. The benefits and risks associated with screening tests are discussed in detail in *Chap. 6*. Risks of routine evaluations include inappropriate testing or overtesting or false-positive findings that require follow-up and induce patients to worry. Periodic health examinations appear to be associated with less patient worry. On balance, given the lack of convincing evidence of harm and the potential for better delivery of appropriate screening, counseling, and preventive services, periodic health evaluations appear reasonable for general populations at average risk for chronic conditions.

It is important to note that routine annual comprehensive physical examinations of asymptomatic adult patients have very low yield and may take an inordinate amount of time in a wellness visit. Such time

TABLE 2-1 Guidelines and Key Recommendations from the Dietary Guidelines for Americans, 2020–2025

GUIDELINES	KEY RECOMMENDATIONS
<p>1. Follow a healthy dietary pattern at every life stage. For the first 6 months of life, infants should exclusively be fed human milk, or iron-fortified formula if human milk is unavailable. From 6 to 12 months, infants should be introduced to a variety of complementary nutrient-dense foods. From 12 months to older adulthood, the dietary pattern should meet nutrient needs, help achieve a healthy body weight, and reduce the risk of chronic disease.</p> <p>2. Customize and enjoy nutrient-dense food and beverage choices to reflect personal preferences, cultural traditions, and budgetary considerations. The Dietary Guidelines provide a framework of several dietary patterns intended to be customized to individual needs and preferences, as well as the foodways of the diverse cultures in the United States.</p> <p>3. Focus on meeting food group needs with nutrient-dense foods and beverages, and stay within calorie limits. Nutrient-dense foods provide vitamins, minerals, and other health-promoting components and have no or little added sugars, saturated fat, and sodium. A healthy dietary pattern consists of nutrient-dense forms of foods and beverages across all food groups, in recommended amounts, and within calorie limits.</p> <p>4. Limit foods and beverages higher in added sugars, saturated fat, and sodium, and limit alcoholic beverages. At every life stage, meeting food group recommendations, even with nutrient-dense choices, fulfills most of a person's daily calorie needs and sodium limits, with little room for extra added sugars, saturated fat, or sodium, or for alcoholic beverages.</p>	<p>The Dietary Guidelines' Key Recommendations for healthy eating patterns should be applied in their entirety, given the interconnected relationship that each dietary component can have with others. They are also intended as a framework to accommodate personal preferences, cultural traditions, and budgetary considerations.</p> <p>Focus on meeting food group needs with nutrient-dense foods and beverages, and stay within calorie limits to achieve a healthy weight and reduce the risk of chronic disease.</p> <p>The core elements that make up a healthy dietary pattern include:</p> <ul style="list-style-type: none"> Vegetables of all types—dark green; red and orange; beans, peas, and lentils; starchy; and other vegetables Fruits, especially whole fruit Grains, at least half of which are whole grain Dairy, including fat-free or low-fat milk, yogurt, and cheese, and/or lactose-free versions and fortified soy beverages and yogurt as alternatives Protein foods, including lean meats, poultry, and eggs; seafood; beans, peas, and lentils; and nuts, seeds, and soy products Oils, including vegetable oils and oils in food, such as seafood and nuts <p>A healthy eating pattern limits:</p> <ul style="list-style-type: none"> Added sugars—Less than 10% of calories per day starting at age 2. Avoid foods and beverages with added sugars for those younger than age 2. Saturated fat—Less than 10% of calories per day starting at age 2. Sodium—Less than 2300 mg per day—and even less for children younger than age 14. Alcoholic beverages—Adults of legal drinking age can choose not to drink or to drink in moderation by limiting intake to 2 drinks or less in a day for men and 1 drink or less in a day for women, when alcohol is consumed. Drinking less is better for health than drinking more. There are some adults who should not drink alcohol, such as women who are pregnant. <p>Meet the U.S. Department of Health and Human Services' <i>Physical Activity Guidelines for Americans</i></p> <p>In tandem with the recommendations above, Americans of all ages—children, adolescents, adults, and older adults—should meet the <i>Physical Activity Guidelines for Americans</i> to help promote health and reduce the risk of chronic disease. Americans should aim to achieve and maintain a healthy body weight. The relationship between diet and physical activity contributes to calorie balance and managing body weight.</p>

Source: Adapted from the *Dietary Guidelines for Americans, 2020–2025*. Washington, DC: U.S. Department of Agriculture and U.S. Department of Health and Human Services; 2020. Available at https://www.dietaryguidelines.gov/sites/default/files/2020-12/Dietary_Guidelines_for_Americans_2020-2025.pdf.

may be better spent on assessing and counseling the patient on other aspects of their health, as discussed below. Evidence-based components that should be included in periodic evaluations focused on health and prevention include a number of age-appropriate screening tests for chronic disease and risk factors, preventive interventions including immunizations and chemoprevention for at-risk individuals, and preventive counseling. The U.S. Preventive Services Task Force publishes its *Guide to Clinical Preventive Services*, which contains evidence-based recommendations from the Task Force on preventive services for which there is a high degree of certainty that the service provides at least moderate net clinical benefit (i.e., benefits outweigh harms significantly and to a reasonable magnitude).

Healthy Behaviors and Lifestyles Owing to the paucity of evidence, the heterogeneity of study designs, and the diverse nature of interventions studied, many clinicians are uncertain as to how to deliver advice regarding healthy behaviors and lifestyles. Nevertheless, adverse behaviors and lifestyles contribute to >75% of premature, preventable deaths and disability. Estimates from the U.S. National Health and Nutrition Examination Survey indicate that fewer than 1% of Americans achieve an optimal heart-healthy eating pattern. Thus, whereas there are many demands on time during a typical patient-clinician encounter, few things may have more impact on longevity, health, and quality of life for asymptomatic patients than an efficient approach to assessing, documenting, and improving patients' health behaviors. Indeed, the mere act of assessing health behaviors has been shown to affect patients' health behaviors. Facility with tools for assessment of lifestyle and with strategies for counseling are therefore of paramount importance.

Healthy Eating Patterns (see Chap. 332) Despite the existence of numerous "fad" diets and seemingly inconsistent recommendations

on dietary composition, there is remarkable agreement about what should constitute a healthy eating pattern for the broad population to avoid nutritional deficits (i.e., vitamin deficiency) and excesses (i.e., excessive caloric intake) and to maximize potential health (**Table 2-1**). Optimal eating patterns consist of whole fruits and vegetables, whole grains, lean proteins, and healthy oils, and allow for nonfat or low-fat dairy intake. They tend to exclude frequent ingestion of foods high in refined sugars and starches, saturated fat, and sodium. Since sodium and refined sugars and starches are the hallmark of much of the processed/packaged food supply, a simple rule of thumb is to provide or cook the majority of one's own meals starting from whole foods and emphasizing fruits and vegetables. Likewise, foods prepared outside of the home tend to have higher fat and sodium content, so special attention to menu choices focused on fruits, vegetables, lean proteins, and whole grains, while minimizing sauces and dressings, can help most individuals follow healthier eating patterns when eating food prepared outside the home. In all cases, sugar-sweetened beverages and nonnutritious snack foods should be minimized. If snacks are included, small amounts of healthy nuts and seeds or more fruits and vegetables should be encouraged.

Specific conditions and diseases, such as diabetes, other metabolic disorders, allergies, and gastrointestinal disorders, may require tailored approaches to diet. In counseling most patients, the general approach should focus on whole foods, eating patterns, and appropriate calorie balance, rather than on specific micronutrients such as electrolytes or selected vitamins. It should be remembered that most patients have difficulty understanding nutritional labels on packaged foods, with the attendant demands on numeracy and health literacy.

Dietary guidelines are published by the U.S. Department of Agriculture (USDA) and U.S. Department of Health and Human Services every 5 years, and these guidelines have undergone substantial

evolution over time. The current U.S. Dietary Guidelines and Key Recommendations for 2020–2025 are summarized in Table 2-1 and emphasize the importance of healthy eating patterns for every stage of life, to avoid chronic diseases including obesity, diabetes, cancer, and cardiovascular disease. The core elements include eating patterns with nutrient-dense (rather than calorie-dense) whole foods and appropriate caloric intake to achieve and maintain healthy weight. The USDA guidelines focus on the concept of a healthy plate (rather than the prior food pyramid) for ease of counseling and adoption. Fifty percent of the plate should consist of vegetables and whole fruits, with remaining portions for whole grains and lean protein foods. When using fat for cooking, it should be done by sauteing in healthier oils (e.g., canola oil), and addition of judicious amounts of healthy raw oils (e.g., olive oil, nuts) to dishes is appropriate. Recommendations also focus on limitation of foods and beverages higher in added sugars, saturated fat, and sodium, and moderation or avoidance of alcohol intake.

The USDA guidelines focus on specific healthy eating patterns that adhere to these broad recommendations and are appropriate for ~97% of the general population. They identify a “Healthy U.S.-Style Dietary Pattern” that adheres closely to the evidence-based Dietary Approaches to Stop Hypertension (DASH) eating pattern but is customizable for different cultural or personal preferences. Alternative patterns, which vary more in emphasis than in content, include a “Healthy Mediterranean-Style Dietary Pattern” and a “Healthy Vegetarian Dietary Pattern.”

AGE- AND SEX-SPECIFIC RECOMMENDATIONS Current dietary framework recommendations are generally similar for all life stages from ages ≥12 months, but recommended levels of caloric intake (and hence amounts of foods) differ by age, sex, and physical activity level. For example, recommended caloric intake ranges from 1000 calories/d for sedentary 2-year-old children to as high as 3200 calories/d for active 16- to 18-year-old young men. Recommended caloric intakes peak in late adolescence or early adulthood for men and women and gradually decrease over ensuing decades.

As with all lifestyle counseling aimed at behavior change, dietary approaches that partner with the patient and utilize motivational interviewing strategies and shared goals and commitments tend to work best, as described below (see “Approach to the Patient”).

Physical Activity Similar to the approach to counseling regarding healthy eating patterns, recommendations on participation in physical activity emphasize the point that any physical activity is better than none. A simple rule of thumb for patients is: “If you are doing nothing, do something; and if you are doing something, do more, every day.” The evidence base for physical activity indicates that the marginal benefits from physical activity are greatest in advancing from no activity to low levels of moderate activity. With increasing duration and intensity of activity, there is a continued curvilinear increase in health benefits, but the marginal gains for each additional minute of moderate-to-vigorous activity slowly diminish. Thus, for adults, the recommended amount of physical activity is 150 min of moderate-intensity or 75 min of vigorous-intensity aerobic activity per week, performed in episodes of at least 5 min, and preferably spread throughout the week, plus participation in muscle-strengthening activity at least 2 days per week. Additional health benefits can be realized by engaging in physical activity beyond this amount.

In counseling patients regarding physical activity, it is important to note that sedentary time (e.g., seated at work or at home in front of electronic screens) has adverse health consequences independent of the lack of physical activity during these episodes. Therefore, even modest efforts like standing at the desk and doing gentle stretching for periods during the day may be beneficial. It is also important to emphasize that participating in a variety of aerobic activities (biking, swimming, walking, jogging, rowing, elliptical training, stair-climbing, etc.) can be beneficial and may help to avoid overuse injuries and boredom with the exercise regimen. If patients choose to participate in muscle-strengthening activities for health improvement, emphasis should be placed on weights that allow more repetitions (e.g., 3 sets of 15–20 repetitions that can be performed comfortably, with a rest period in between) and on avoiding breath-holding and straining against a closed glottis.

SUDDEN CARDIAC DEATH RISK Patients may express concerns regarding the risk of sudden cardiac death during exercise. Whereas the risk of sudden death during exercise does increase directly with the amount of time spent exercising, this association is substantially mitigated by training effects. Thus, patients embarking on an exercise program should be encouraged to increase the duration of aerobic exercise gradually as tolerated, aiming for episodes of at least 30 min 5 times a week as an ideal. Once a comfortable duration is reached, incorporating interval training periods of more intensive activity interspersed during the exercise can provide greater fitness gains.

EXTREME ENDURANCE ACTIVITIES As with other forms of exercise, extreme endurance activities such as triathlons and marathons should be undertaken only with appropriate and graded training. Such activities tend to take a greater toll on the musculoskeletal system over time than less extreme activities, and they are also associated with measurable damage to the myocardium and greater risks for other organ damage. Athletes participating in endurance activities routinely have elevations in cardiac troponin (a specific circulating marker of myocardial cell damage and death) at the end of the race, although elevations are lower in those who are well trained. Patients and clinicians should consider the patient’s overall health, specific limitations, potential for injury, and ability to train in decision-making regarding participation in endurance events.

AGE-SPECIFIC RECOMMENDATIONS The U.S. Department of Health and Human Services’ *Physical Activity Guidelines for Americans*, second edition (2018) ([Table 2-2](#)), recommend that preschool-aged children (aged 3–5 years) should be physically active throughout the day in a variety of activity types to enhance growth and development. Children and adolescents aged 6–17 years should participate in ≥60 min of physical activity daily, most of which should be moderate- or vigorous-intensity aerobic activity, including vigorous, muscle-strengthening, and bone-strengthening activities at least 3 days a week each. As noted above, adults aged 18–64 years are recommended to pursue at least 150 min of moderate-intensity or 75 min of vigorous-intensity aerobic activity per week (or equivalent combinations), with at least 2 days of muscle-strengthening activities. Adults aged ≥65 years should follow the adult guidelines or be as active as possible as abilities and conditions allow. For older adults, special emphasis is also placed on multicomponent physical activity that includes balance training as well as aerobic and muscle-strengthening activities.

Sleep Hygiene Sleeping between 7 and 9 h per night appears to be optimal for health in adults aged ≥18 years. Sleeping <7 h is associated with adverse outcomes, including obesity, diabetes, elevated blood pressure, cardiovascular disease, depression, and all-cause mortality, as well as physiologic disturbances such as impaired immune function, increased pain sensitivity, and impaired cognitive performance. Conversely, achieving appropriate levels of sleep is associated with more success in weight loss, better blood pressure control among patients with hypertension, and improved mental health and performance. Regular sleep more than 9 h per night is appropriate for children and adolescents or individuals recovering from sleep deprivation or illness, but for most individuals, the effects on health are uncertain.

Patients often express concerns about the quantity and quality of their sleep. With aging, both aspects of sleep tend to decline, even without overt sleep disorders. Documentation of sleep using a sleep log may assist in understanding different types of insomnia and sleep disorders. Encouraging daily activity to promote fatigue, avoidance of eating and drinking alcohol too close to bedtime, and regular daily sleep habits may help patients achieve better sleep. Regular use of sedative medications should generally be discouraged given the high potential for dependence, addiction, and altered sleep quality.

DISORDERS OF SLEEP The prevalence of sleep-related breathing disorders, including obstructive sleep apnea (OSA), is poorly documented. A recent systematic review suggested that the prevalence of clinically important OSA in the general adult population may be between 9% and 38%, with higher rates in men versus women, older versus younger adults, and those with higher versus lower BMI.

TABLE 2-2 Recommendations from Physical Activity Guidelines for Americans

AGE	RECOMMENDATIONS
3–5 years	<ul style="list-style-type: none"> Preschool-aged children (ages 3 through 5 years) should be physically active throughout the day to enhance growth and development. Adult caregivers of preschool-aged children should encourage active play that includes a variety of activity types.
6–17 years	<ul style="list-style-type: none"> It is important to provide young people opportunities and encouragement to participate in physical activities that are appropriate for their age, that are enjoyable, and that offer variety. Children and adolescents ages 6 through 17 years should do 60 min (1 h) or more of moderate-to-vigorous physical activity daily: <ul style="list-style-type: none"> Aerobic: Most of the 60 min or more per day should be either moderate- or vigorous-intensity aerobic physical activity and should include vigorous-intensity physical activity on at least 3 days a week. Muscle-strengthening: As part of their 60 min or more of daily physical activity, children and adolescents should include muscle-strengthening physical activity on at least 3 days a week. Bone-strengthening: As part of their 60 min or more of daily physical activity, children and adolescents should include bone-strengthening physical activity on at least 3 days a week.
18–64 years	<ul style="list-style-type: none"> Adults should move more and sit less throughout the day. Some physical activity is better than none. Adults who sit less and do any amount of moderate-to-vigorous physical activity gain some health benefits. For substantial health benefits, adults should do at least 150 min (2 h and 30 min) to 300 min (5 hours) a week of moderate-intensity or 75 min (1 h and 15 min) to 150 min (2 h and 30 min) a week of vigorous-intensity aerobic physical activity, or an equivalent combination of moderate- and vigorous-intensity aerobic activity. Preferably, aerobic activity should be spread throughout the week. Additional health benefits are gained by engaging in physical activity beyond the equivalent of 300 min (5 h) of moderate-intensity physical activity a week. Adults should also do muscle-strengthening activities of moderate or greater intensity and that involve all major muscle groups on 2 or more days a week, as these activities provide additional health benefits.
≥65 years	<ul style="list-style-type: none"> The key guidelines for adults also apply to older adults. In addition, the following key guidelines are just for older adults: <ul style="list-style-type: none"> As part of their weekly physical activity, older adults should do multicomponent physical activity that includes balance training as well as aerobic and muscle-strengthening activities. Older adults should determine their level of effort for physical activity relative to their level of fitness. Older adults with chronic conditions should understand whether and how their conditions affect their ability to do regular physical activity safely. When older adults cannot do 150 min of moderate-intensity aerobic activity a week because of chronic conditions, they should be as physically active as their abilities and conditions allow.

Moderate-intensity physical activity: Aerobic activity that increases a person's heart rate and breathing to some extent. On a scale relative to a person's capacity, moderate-intensity activity is usually a 5 or 6 on a 0 to 10 scale. Brisk walking, dancing, swimming, or bicycling on a level terrain are examples. **Vigorous-intensity physical activity:** Aerobic activity that greatly increases a person's heart rate and breathing. On a scale relative to a person's capacity, vigorous-intensity activity is usually a 7 or 8 on a 0 to 10 scale. Jogging, singles tennis, swimming continuous laps, or bicycling uphill are examples. **Muscle-strengthening activity:** Physical activity, including exercise that increases skeletal muscle strength, power, endurance, and mass. It includes strength training, resistance training, and muscular strength and endurance exercises. **Bone-strengthening activity:** Physical activity that produces an impact or tension force on bones, which promotes bone growth and strength. Running, jumping rope, and lifting weights are examples.

Source: Adapted from U.S. Department of Health and Human Services. *Physical Activity Guidelines for Americans, 2nd edition*. Washington, DC: U.S. Department of Health and Human Services; 2018. Available at https://health.gov/sites/default/files/2019-09/Physical_Activity_Guidelines_2nd_edition.pdf.

Patients with persistent complaints of poor sleep quality or excessive daytime somnolence or with witnessed apneic spells may benefit from screening for sleep disorders, prior to consideration of a formal sleep study. A number of clinical tools have been developed to screen for sleep apnea, including the Epworth Sleepiness Scale, the STOP (snoring, tiredness, observed apnea, high blood pressure) Questionnaire, and the STOP-Bang Questionnaire (STOP plus assessment of BMI, age, neck circumference, and gender), among others. The U.S. Preventive Services Task Force found that current evidence is insufficient to assess the balance of benefits and harms of screening for OSA in asymptomatic adults owing to a lack of validation data in primary care settings. Nonetheless, the high prevalence and significant health consequences of sleep apnea suggest that clinicians should be alert for its potential presence, particularly in patients who are obese with symptoms of excessive daytime somnolence or witnessed apnea episodes. Other sleep disorders, such as restless leg syndrome, may be identified with simple history.

Weight Management Overweight and obesity are prevalent in epidemic proportions in the United States and other industrialized nations (**Chaps. 401 and 402**). Since 1985, the prevalence of obesity in the United States has increased from ~10 to ~35%, and the prevalence of overweight is now ~40%. Overweight and obesity disproportionately affect individuals in lower socioeconomic strata and in many underserved minority populations, including black Americans, Latino Americans, and American Indians. In all race/ethnic groups, both overweight and obesity are associated with adverse health consequences, including diabetes, certain cancers, cardiovascular diseases, and degenerative joint disease. Eating disorders such as anorexia and bulimia are much less common but pose major health consequences

for affected patients and should be suspected particularly in younger women with history of rapid weight shifts or underweight status.

Weight loss is one of the most difficult preventive interventions to achieve and sustain over time. However, several key factors can assist the patient and clinician, and early referral to a dietitian can be very helpful. The first therapeutic goal is to aim for weight stabilization. Many of the risks of overweight and obesity are driven more strongly by continued weight gain, rather than overweight/obese status per se. Working with the patient to find initial strategies for weight maintenance can be a successful initial step with success for many patients. For those who can progress to considering weight loss, it is critical to help the patient understand that there is no standard solution. Experimentation and documentation are key. Tools to assist patients can include food and weight logs, activity logs, and smart phone apps. Some patients respond best to structured approaches such as intermittent fasting regimens or commercial dietary programs where meals are provided. Any of these approaches can be tried with or without social group supports.

The key construct for weight loss is, of course, negative calorie balance. This is achieved through a combination of reduced caloric intake and increased physical activity. Patients may already understand, from prior weight loss attempts, what combination works best for them to achieve this. Some patients find that they cannot lose weight without increasing their exercise. For many, reduction of caloric intake is most efficient. Encouraging the patient to find what works for them is most important. The same principle holds for dietary content. Well-done feeding studies indicate that weight loss is dependent far more on the reduction of caloric intake than on the relative composition of fat, protein, and carbohydrate in the diet. There may be other medical reasons

to choose one approach over another, but if not, encouraging the patient to pick one approach and document the results is an important start. Once weight loss is achieved, increase in activity is often required for its successful maintenance.

Tobacco Cessation (see Chap. 454) Escaping nicotine dependence is another major, but critical, challenge to prevention and wellness efforts. The addictive effects of nicotine have been well documented, with effects that can last for years after successful cessation. Assessing a patient's past history of cessation attempts and current readiness for change are key first steps in forging a successful approach. Frequent follow-up and reinforcement, as well as use of nicotine replacement therapy and other cessation-promoting medications, are additional critical elements. Recidivism is the rule, and patients should expect to resume smoking and attempt again as they journey to tobacco cessation. Electronic cigarettes have some evidence for benefit in adult smoking cessation, but their potential for use by adolescents and young adults who are not smokers represents a major public health threat for a new generation of nicotine addiction, with unknown health consequences as a result of the high doses of nicotine delivered to developing organs, including the brain. Vaping of other substances, often in association with flavoring compounds, has also been associated with pulmonary and cardiovascular damage and should be actively discouraged.

VACCINATION (CHAP. 123)

One of the major advances in public health that has contributed to increases in health and longevity worldwide is the development of safe and effective vaccinations against endemic and epidemic infectious diseases. Patients should be counseled regarding age-appropriate vaccinations for their children and for themselves. Some individuals may be reluctant to receive a vaccination; in these cases, listening to the patient's concerns is important, followed by explanation of the benefits to the individual, their family, and their community and review of the low risk for potential harms. It is true to say that no current vaccines are ever worse than the disease they prevent, although side effects may occur rarely. Thorough knowledge of the data on side effect rates and of efficacy will aid the clinician in helping the patient make a fully informed decision.

MENTAL HEALTH AND ADDICTION

Assessment for depression and cognitive impairment is important to address when patients exhibit symptoms or they or their family members express concerns. Both of these common conditions play a major role in reducing quality of life and are high on patients' lists of concerns, even if not clearly expressed. Screening tools for depression are reviewed in *Chap. 452*. Cognitive function decline with aging or comorbid illness, including depression, should be anticipated. Assessment tools such as the General Practitioner Assessment of Cognition or the Mini-Cog™ test are widely available and effective rapid assessment tools.

Alcohol and Opioids (see Chaps. 453 and 456) Alcohol dependence and abuse are common and underdiagnosed. Rapid screening tools have proven efficacy for identifying patients with alcohol problems. In a systematic review, the CAGE (cut down, annoyed, guilty, eye opener) questionnaire was most effective at identifying alcohol abuse and dependence, with reasonable sensitivity and high specificity. The present opioid epidemic in the United States presents a new and substantial public health challenge given the high potential for dependency and abuse of these drugs. Rapid screening tools are available to assist clinicians in screening for opioid dependence.

ACCIDENTS AND SUICIDE

Regular assessment of patient safety through simple questions about seat belt use, domestic violence, and gun safety in the home continues to be an important part of health promotion and wellness. Long-standing recommendations for assessment of suicidal ideation among patients with depression or a history of suicide attempts also continue to be relevant.

APPROACH TO THE PATIENT

In the context of a clinical visit focused on health assessment, health promotion, and prevention, the basic skills of history-taking are of paramount importance. Much of the evaluation, counseling, and management that focus on health promotion and prevention also require engagement and buy-in from the patient in order to assist with recognition of contributing behaviors and to promote adherence to therapeutic plans. Therefore, in addition to standard history-taking, additional skills such as motivational interviewing and eliciting patient commitments and contracting may prove of significant value. The availability of additional tools to assist with screening, monitoring, and chronic management, both online and through wearable devices and mobile health technologies, is rapidly expanding, with uncertain implications for the future. Major research gaps exist in our understanding of how best to employ these newer technologies to improve health outcomes. Concepts of behavioral economics are being explored to better understand the psychology of decision-making and incentives as a means to improve lifestyle choices and adherence to treatment plans (*Chap. 481*).

The limited time available to clinicians and patients during a wellness visit or periodic health examination (not driven by specific patient issues) makes it important to prioritize assessment and counseling for factors that affect longevity, health span, and quality of life over approaches that may have low yield, such as the annual comprehensive physical examination in an asymptomatic patient. Setting clear expectations for the content of a wellness visit may be a first step, and scheduling follow-up visits for findings or to continue indicated counseling are important steps to achieving better health outcomes.

FURTHER READING

- BOULWA LE et al: Systematic review: The value of the periodic health evaluation. *Ann Intern Med* 146:289, 2007.
- DIETARY GUIDELINES FOR AMERICANS, 2020–2025. Washington, DC: U.S. Department of Agriculture and U.S. Department of Health and Human Services; 2020. Available at https://www.dietaryguidelines.gov/sites/default/files/2020-12/Dietary_Guidelines_for_Americans_2020-2025.pdf.
- IRISH LA et al: The role of sleep hygiene in promoting public health: A review of empirical evidence. *Sleep Med Rev* 22:23, 2015.
- KROGSBOLL LT et al: General health checks in adults for reducing morbidity and mortality from disease: Cochrane systematic review and meta-analysis. *BMJ* 345:e7191, 2012.
- U.S. DEPARTMENT OF HEALTH AND HUMAN SERVICES: *Physical Activity Guidelines for Americans*, 2nd ed. Washington, DC: U.S. Department of Health and Human Services; 2018. Available at https://health.gov/sites/default/files/2019-09/Physical_Activity_Guidelines_2nd_edition.pdf.
- U.S. PREVENTIVE SERVICES TASK FORCE webpage. Available at <https://www.uspreventiveservicestaskforce.org/uspstf/>.

3

Vaccine Opposition and Hesitancy

Julie A. Bettinger, Hana Mitchell

Vaccines have been recognized as one of the top public health achievements of the twentieth century. Dramatic declines in the morbidity and mortality of vaccine-preventable diseases have been observed, and the contribution of vaccines to the elimination, control, and prevention of infectious disease cannot be overstated. However, opposition and hesitancy to vaccines exist and are not new. Vaccine hesitancy has existed

since Edward Jenner introduced the first vaccine against smallpox in the eighteenth century. So why did the World Health Organization rank these attitudes as one of the ten greatest threats to public health in 2019? Are current opposition and hesitancy any different from what has been seen before? Many sociologists, public health experts, and health care providers (HCPs) argue yes. Recent social and cultural trends, combined with new communication formats, have converged to create a particularly potent form of hesitancy and what some have labeled a crisis of confidence. This crisis manifests as a lack of trust in specific vaccines, vaccine programs, researchers, HCPs, the health care system, pharmaceutical companies, academics, policymakers, governments, and authority in general. ([See “Focus: COVID-19 Vaccine Hesitancy,” below.](#))

The roots of modern vaccine hesitancy and opposition—defined as delay or rejection of vaccines in spite of availability—vary depending on the place and the population. For some individuals and communities, pseudoscience and false claims about the safety of existing vaccines (e.g., an unsupported link between measles vaccine and autism) have driven fears, increased hesitancy, and decreased acceptance. For others, real safety events, such as the association of narcolepsy with a specific pandemic influenza vaccine (Pandemrix), have justified concerns. In a few locations (e.g., Ukraine, Pakistan), vaccine hesitancy is the result of failed health systems or even state failures. Finally, for some groups, including some fundamentalist religious groups and alternative-culture communities, vaccine hesitancy and opposition reflect exclusion from and rejection of mainstream society and allopathic health care and manifest as a deep distrust of these institutions and their HCPs. Although the genesis of modern vaccine hesitancy is multifactorial, its outcomes are uniform: a decrease in vaccine demand and uptake, a decrease in coverage by childhood and adult vaccines, and an increase in vaccine-preventable diseases, outbreaks, and epidemics of disease. Addressing this crisis and moving people from vaccine hesitancy and refusal to acceptance and active demand require intervention at multiple levels: the individual, the health system (including public health), and the state.

This chapter will define vaccine hesitancy and briefly describe its determinants and effects in North America (the United States and Canada). Physicians and other HCPs are well positioned to address the

crisis of confidence many patients feel toward HCPs and the health care system. Studies demonstrate that an unambiguous, strong recommendation by trusted HCPs is most often the reason that patients, including those who are vaccine hesitant, choose to vaccinate. Strategies for counseling vaccine-hesitant and vaccine-resistant patients will be presented and examples of strong vaccine recommendations provided. Presenting strategies to increase vaccine demand at a system and policy level is beyond the scope of this chapter. While some physicians may have roles that allow them to act at this level, all physicians can act and influence their individual patients. Strategies to create active vaccine demand at the individual level alone will not solve vaccine hesitancy, but vaccine hesitancy cannot be addressed without these efforts. [For further discussion of immunization principles and vaccine use, see Chap. 123.](#)

VACCINE COVERAGE AND OUTBREAKS

The epidemiologic data from measles outbreaks over the past 10 years provide an interesting illustration of the effects of vaccine opposition and hesitancy. [For further discussion of measles, see Chap. 205.](#)

North America *Herd immunity* occurs when enough individuals in a population become immune to an infectious disease, usually through vaccination, that transmission of the infection stops. The level of immunity (or level of vaccine coverage) required to confer herd immunity varies with the specific infectious disease. Because measles is a highly contagious virus, a coverage rate of 93–95% must be achieved for vaccination to confer herd immunity and interrupt measles transmission. National coverage estimates place one-dose measles vaccine coverage rates in 2-year-old children at 92% in the United States and 88% in Canada. In spite of these relatively high levels of coverage in young children, numerous measles outbreaks have occurred in both countries since 2010 ([Table 3-1](#)).

The vast majority (>80%) of measles cases described in Table 3-1 occurred in under- or completely unvaccinated individuals. Of note, many of these outbreaks highlight pockets of significantly under- or unvaccinated individuals that are not apparent in national vaccine coverage statistics. Moreover, many of the outbreaks listed in Table 3-1 were ignited by unvaccinated returned travelers from areas with existing

TABLE 3-1 Measles Outbreaks in North America

YEAR/PLACE	NO. OF CASES	REASON
2010/Canada	70	An infected traveler to the 2010 Winter Olympics transmitted infection to an under- and unvaccinated local population in British Columbia.
2011/Canada	776	Disease was imported from France by an unvaccinated returned traveler to Quebec. The outbreak spread in a nonvaccinating religious community and outside that community. A majority of cases occurred in under- and unvaccinated persons.
2011/United States	118	Of 118 cases, 46 were in returned travelers from Europe and Asia/Pacific regions; 105 cases (89%) occurred in unvaccinated persons.
2013/United States	58	Disease was imported by a returned unvaccinated traveler from Europe. The outbreak spread in a nonvaccinating religious community in New York.
2014/Canada	433	Disease was imported from the Netherlands. The outbreak spread in a nonvaccinating religious community in British Columbia.
2014/United States	383	The outbreak occurred in nonvaccinating religious communities in Ohio.
2015/United States	147	A multistate/multicountry outbreak was linked to Disneyland amusement park. More than 80% of cases occurred in unvaccinated persons.
2015/Canada	159	Disease was imported from the United States (part of the Disneyland outbreak) by an unvaccinated traveler. The outbreak spread in a nonvaccinating religious community in Quebec.
2017/United States	75	The outbreak occurred in an under-vaccinated community in Minnesota; 95% of patients were unvaccinated.
2018/United States	375	Disease was imported by returned unvaccinated travelers from Israel. The outbreak spread in nonvaccinating religious communities in New York and New Jersey.
2019/Canada	31	Disease was imported from Vietnam by a returned traveler to British Columbia. The outbreak spread throughout local area schools in under- and unvaccinated persons and resulted in a province-wide measles mass immunization campaign for schoolchildren.
2019/United States	1282	Outbreaks occurred in 10 states; 73% of cases (~935) were linked to outbreaks in nonvaccinating religious communities in New York.

Source: Centers for Disease Control and Prevention and Public Health Agency of Canada.

outbreaks or epidemics, who spread disease into an unvaccinated or under-vaccinated community. Many of the outbreaks were contained within the nonvaccinating community, but several spread to other under-vaccinated communities geographically contiguous with the outbreak community. More concerning still are the cases and outbreaks originating in communities that had not previously been identified as nonvaccinating. These cases likely highlight pockets of unvaccinated individuals who object for cultural rather than religious reasons. In the past, these nonvaccinating individuals did not exist in large enough clusters to sustain the spread of measles. Of further concern is the number of individuals included in outbreak statistics who have had one or sometimes even two doses of vaccine and who were thought to be protected but who still end up with the disease. The assumption is that one or two doses provide full disease immunity, but this is not always true. Often, individual level characteristics (age, immune compromise, etc.) affect the individual's response to the vaccine and their level of protection. In other instances, vaccine protection can wane over time, thus leaving fully immunized individuals susceptible to infection. In fact, when herd immunity breaks (i.e., the level of immunity in a community becomes too low to prevent transmission of disease), the occurrence of cases even in fully immunized persons is seen, as reflected in outbreak statistics. As a result of decreased vaccination rates and the resulting disruption of herd immunity, these individuals may become more identifiable as non-immune.

Outside North America Although overall coverage rates may still be high in North America, they are lower in other parts of the world. In Samoa, for example, measles–mumps–rubella (MMR) vaccine coverage before a recent outbreak was 31%; in the Philippines, it was 67%. Twenty years ago, vaccine coverage was sufficiently high in some parts of the world, including Europe, that an unvaccinated traveler from a nonvaccinating community to most regions would have been protected by herd immunity at their destinations. Today that is not the case: such travelers are likely to become infected in a country with active measles transmission and return home to spread the infection into their communities and possibly beyond. Thus active measles transmission, whether at home or abroad, places individuals who rely on herd immunity (e.g., immunocompromised persons and young infants) at increased risk.

FACTORS IN VACCINE HESITANCY

Vaccination coverage rates provide an estimate of the proportion of children or adults in the population who have been vaccinated, but they do not indicate the proportion of individuals who are vaccine hesitant. An individual may be fully vaccinated but still be hesitant about the safety and effectiveness of vaccines, or an individual may be unvaccinated as a result of access issues but may not be hesitant. Therefore, in attempts to understand a patient's lack of vaccination, it is important to distinguish persons who are hesitant and refuse vaccines from those who need assistance to access the health care system and successfully complete vaccination. To this end, an understanding of vaccine hesitancy and its determinants is needed.

Vaccine hesitancy and opposition are defined by the World Health Organization's SAGE Working Group on Vaccine Hesitancy as a "delay in acceptance or refusal of vaccines despite availability of vaccination services." The SAGE group describes vaccine hesitancy as "complex and context specific, varying across time, place, and vaccines."

It is useful to frame vaccine acceptance as a continuum pyramid, with active demand for all vaccines representing the largest group at the bottom of the pyramid and outright refusal of all vaccines depicted in the smallest group at the top. In the middle lies vaccine hesitancy, in which the degree of vaccine demand and acceptance varies. Fortunately, for disease control efforts, most individuals fall within the active-demand category or, if they are hesitant, still accept all vaccines. Hesitancy can be influenced by complacency, convenience, and confidence (Fig. 3-1).

Complacency is self-satisfaction when accompanied by a lack of awareness for real dangers or deficiencies. Complacency exists in communities and individuals when the perceived risks of vaccine-preventable diseases are low and vaccination is not deemed a necessary preventive action. This attitude can apply to vaccination in general or to specific vaccines, such as influenza vaccines. Actual or perceived vaccine efficacy and effectiveness contribute to complacency. Patients who are complacent about vaccine-preventable diseases prioritize other lifestyle or health factors over vaccination. These individuals can be influenced toward vaccination by a strong recommendation from a trusted HCP or a local influenza outbreak. They can be influenced away from vaccination by a vaccine scare or misinformation on social

Characteristics

- Strong distrust of health system/pharmaceutical industry/government
 - Strong-willed and committed against vaccines
 - Negative or traumatic experiences with HCPs and health system
 - May use natural approach to health/alternative HCPs
 - May have strong religious/moral considerations for refusal
 - May cluster in communities (geographic and online)
 - Vaccination is very unlikely; alternative strategies to protect individual and community must be discussed.
-
- Questions safety and necessity of vaccines
 - Actively seeks information from many sources
 - Has conflicting feelings on whom to trust
 - Social norm is not vaccinating.
 - May have had negative or traumatic experience with health system
 - Vaccination may not occur; a strong trust relationship with HCP and many visits and conversations are required.
-
- Focused on vaccine risks
 - Conversation with trusted HCP strongly influential
 - Trusts HCPs
 - Actively seeking information and wants to verify it
 - Wants advice specific for their child
 - Confused by conflicting information
 - Social norm is vaccinating, but individual may feel conflicted by this norm.
 - Vaccination requires longer conversation and may require multiple visits.
-
- Focused toward vaccine risk
 - **Complacency:** low perceived benefits of vaccination
 - Can move up or down continuum as a result of various influences (HCP recommendation, vaccine scare, outbreak)
 - Trusts HCPs and health system
 - **Convenience:** need few barriers to vaccination
 - Vaccination requires longer conversation but likely can be performed at same visit; potential exists to move to active demand.
-
- **Confidence**
 - Considers vaccines important
 - Considers vaccines safe
 - Trusts HCP/vaccines/health system
 - Social norm is vaccinating
 - Very short conversation with HCP about vaccination, in which HCP should address any questions to maintain active-demand status

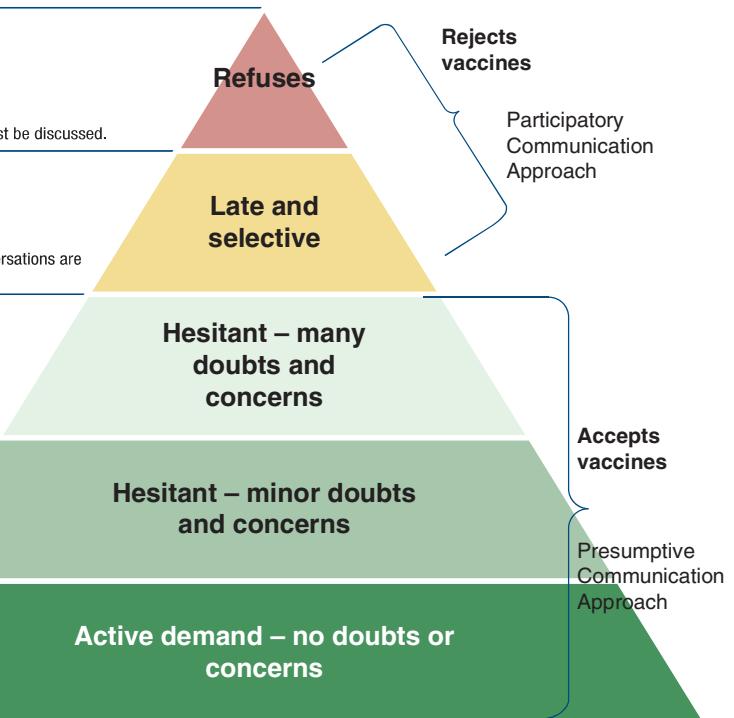


FIGURE 3-1 Vaccine acceptance continuum. HCPs, health care providers. (Adapted from J Leask et al: BMC Pediatrics 12:154, 2012; AL Benin et al: Pediatrics 117:1532, 2006; and E Dubé, NE MacDonald: The Vaccine Book, 2016, pp. 507-528.)

media. Finally, the real or perceived ability of patients to take the action required for vaccination (i.e., self-efficacy) influences the role complacency plays in hesitancy and willingness to seek vaccination.

Convenience is determined by the degree to which conversations about vaccination and other services can be provided in culturally safe contexts that are convenient and comfortable for the individual. Clearly, convenience varies by community, health clinic, and even patient. Persons who are criticized or scolded for not vaccinating themselves or their children may not feel comfortable or safe accessing health services. Factors such as affordability, geographic accessibility, language, and health literacy are important considerations when evaluating the convenience of existing clinical care. Any of these factors can affect vaccine acceptance and can push a patient who has some hesitancy toward vaccinating or not vaccinating.

Confidence is based on trust in the safety and efficacy of vaccines, in the health care system that delivers vaccines (including HCPs), and in the policymakers or governments who decide which vaccines are needed and used. A continual erosion of confidence around vaccination, health systems, and governments drives today's hesitancy and has been amplified by larger social and cultural trends in medicine, parenting, and information availability.

SOCIAL AND CULTURAL TRENDS

Individualized Health Care Over the past 30 years, the focus of medicine and health care has shifted to patient-oriented, individualized care, with an increasing emphasis on treatment and prevention options tailored to the individual patient. In vaccination programs, this shift has manifested as requests for individualized vaccine recommendations and customized immunization schedules. The increasing personalization of medicine, while positive overall, has forced public health away from a focus on the community and its common good and has created tension between individual rights and community health.

Parenting Trends The desire for an individualized approach to medicine and vaccination reflects broader cultural trends concerning individual risk management: accordingly, the individual is to blame for bad outcomes, and public institutions cannot be trusted to manage technological (i.e., vaccine-related) risks. This viewpoint is directly linked with cultural shifts in parenting and social norms defining what it means to be a "good parent." The image of a good parent has been reframed to refer to someone whom several investigators have described as "a critical consumer of health services and products, accounting for their own individual situation as they see it with little regard for the implications of their decision on other children." The archetypical good parent no longer unquestioningly trusts HCPs and other authorities and experts. According to this social norm, "good parents" should seek individual medical advice that is tailored for their child and specific to that child's needs. While in essence not a bad thing, this norm can conflict directly with public health vaccine recommendations and schedules that are organized to maximize community health and to facilitate efficient provision of care at a community level.

Traditional Media Newspapers, radio, and television have been criticized for their coverage of vaccines and in particular their coverage of the alleged link between MMR vaccine and autism. By offering equal coverage throughout the early to mid-2000s for both the scientific evidence and unproven claims of MMR vaccine harms, traditional media outlets provided a forum and a megaphone for the spread of pseudoscience. Equal coverage leads to false equivalencies. Celebrity advocates further amplified the message via this channel. The boost that traditional media provided to active vaccine resistance and, less directly, to vaccine hesitancy has not been adequately measured but must be considered in any discussion of vaccine hesitancy. After headlines about multiple outbreaks of measles and other vaccine-preventable diseases and continued direct criticism of the equal-coverage approach, some traditional media now reject and attempt to discredit pseudoscience. The effect this stance will have on increasing vaccine confidence is unknown.

The Internet and Social Media Approximately 90% of Americans and 91% of Canadians use the Internet, and 80% of Americans and 60% of Canadians have a social network profile. Widespread access to social media can be empowering, but it is also problematic. The Internet and social media require users to select their information sources, creating an environment described as an "echo chamber" in which individuals choose information sources harboring beliefs or opinions similar to their own and thereby reinforcing their existing views. This situation has created a new platform for further spread of vaccine *misinformation* (inaccuracies due to error) and *disinformation* (deliberate lies) and has provided a forum for vaccine-resistant individuals, including celebrities, to organize and raise funds to support their efforts. The harmful effects of Internet and social media use on vaccine hesitancy have been well documented. Vaccine hesitancy increases for parents who seek their information from the Internet. Unfortunately, public health and health care institutions have been slow to adapt to this new communication medium and to recognize its influence and impact. In this medium, personal stories and anecdotes are now viewed as data and disproportionately influence vaccine decision-making, while traditional, more authoritative, fact-based information sources are deemphasized. Centralized monitoring by jurisdiction of vaccine misinformation and disinformation, with summaries of the relevant discourses and rebuttals provided to HCPs, has been proposed as a potential way to counter the influence of social media on vaccine hesitancy. While such strategies have been applied in single jurisdictions and appear to have had some success, their applicability to a broader context is unknown. Moreover, the resources for such a coordinated response have not been made available, and individual HCPs have been left to counter popular, shifting, viral communications on their own, patient by patient.

As with traditional media, the social media landscape appears to be shifting. In 2019, the proliferation of anti-vaccination information combined with measles outbreaks in North America and increasing pressure from health leaders led large social media companies (Facebook, Instagram, Pinterest) to deemphasize anti-vaccination information by removing relevant advertisements and recommendations and decreasing their prominence in search results. While it is too soon to determine the effects of these measures, critics are skeptical that they will have the intended result of reducing vaccine misinformation and disinformation. Early evidence shows that misleading content is still widely available, with anti-vaccine advertisements now using the term "vaccine choice" to avoid censorship. More disturbingly, public health advertisements in support of vaccination have been included in social media and removed from social media sites.

In a more grassroots effort, providers and vaccine supporters have united on social media to provide online support and evidence-based facts to providers and others who support vaccines when they are attacked digitally by anti-vaccine supporters. For example, Shots Heard Round the World (www.shotsheard.com) is an effort led by two U.S. pediatricians to provide advice and support for HCPs who speak out about the importance of vaccines. Such efforts harness the power of social media in ways similar to those used by vaccine opponents and may prove successful in combating vaccine hesitancy.

Given these social and cultural trends, no one should be surprised when individuals now question vaccination, express confusion about conflicting information and information sources, and feel unsure whom to trust. Their broader social context is telling them they should question everything and trust no one. This message is reinforced via misinformation and disinformation on social media. Recent vaccine-preventable disease outbreaks illustrate that effective engagement with individuals cannot be accomplished through one-way, top-down information provision (which still is often the de facto choice for health system communication), but rather requires a dialogue that takes into account the social processes surrounding individual vaccination decisions. It is at the interface between the individual and the health system in which conversations between HCPs and their patients can have the greatest impact. It is critical for all HCPs to discuss vaccines and provide strong vaccine recommendations—including HCPs who do not administer vaccines but who have established trust with their patients.

APPROACH TO THE PATIENT

An ideal vaccine-hesitancy intervention would result in full compliance with vaccination, the patient's satisfaction with the health care encounter, and sustained trust in the HCP's recommendations. On a programmatic level, vaccine-hesitancy interventions should be multicomponent, dialogue based, and tailored to specific under-vaccinated populations.

Communicating with vaccine-hesitant individuals can be challenging and time-consuming. HCPs may feel that vaccine-hesitant patients cast doubt on their personal and professional integrity, their authority as medical experts, and their competence as communicators. Some HCPs may be reluctant to initiate conversations about vaccination because of concerns that discussing a sensitive topic may compromise their clinical rapport with their patients. Other HCPs may believe that they have not received sufficient training to confidently recommend vaccines and answer questions. Discussing vaccines with hesitant patients, while not always easy, provides an opportunity to honor the principles of patient-centered care by demonstrating an interest in patients' opinions, engaging in dialogue, and ideally increasing patients' confidence in vaccine recommendations.

FACTORS IN EFFECTIVE VACCINE RECOMMENDATIONS

Vaccine recommendations ideally should be made within an established, trusting patient-provider relationship in which patients are comfortable asking questions and voicing concerns, even if their views on vaccines contradict the HCP's recommendations. Recommending vaccines requires both provision of information and effective communication. There is no single "best practice" for how providers should approach recommending vaccines to vaccine-hesitant individuals. In general, all vaccine recommendations should be (1) strong, making it clear that the provider supports and recommends vaccination; (2) tailored, acknowledging the vaccine attitudes and potential concerns of individual patients; (3) transparent and accurate, highlighting the benefits of vaccines while also communicating the risks; (4) supported by trustworthy information resources that patients can access and review after the clinical encounter; and (5) revisited, with repetition and reinforcement during follow-up health care encounters.

Strength of the Recommendation HCPs should make it explicit (in the absence of medical contraindications) that vaccination based on the recommended schedule is the best option. While HCPs should take time to elicit patients' questions and address concerns, the recommendation for vaccination should be made in clear and unambiguous terms.

Tailored Communication Vaccine hesitancy occurs on a continuum (Fig. 3-1). Therefore, it is helpful for HCPs to have some understanding of their patients' attitudes toward vaccination at the start of the health care appointment. Unfortunately, vaccine-hesitancy surveys for use as part of vaccine consultation visits have not been validated on a large scale. However, the following are some examples of questions that can be asked, depending on the setting. (1) Did you have a chance to review the vaccine leaflet we provided? Did you have any questions about it? (2) Have you ever been reluctant or hesitant about getting a vaccination for yourself or your child? If so, what were the reasons? (3) Are there other pressures in your life that prevent you from getting yourself or your child immunized on time? (4) Whom/what resources do you trust the most for information about vaccines? Whom/what resources do you trust the least?

Communication style and content for patients in the active-demand category for vaccination will be different from those for individuals who are hesitant, late and selective, or strongly inclined to refuse vaccines. Two communication styles have been proposed for vaccine recommendations. Evidence shows that a *presumptive/directive* approach ("Your child is due for MMR vaccination.") results in higher rates of vaccine uptake than a *participatory/guiding* approach ("What are your thoughts about the MMR vaccine?").

However, adopting a strictly presumptive/directive approach may alienate some patients, especially those who are higher up on the hesitancy pyramid and who may feel that they are being pressured into vaccination before their concerns have been heard and addressed. Adopting a participatory/guiding approach and clarifying receptivity to vaccines may be more suitable for hesitant individuals with many doubts and concerns, persons with a late or selective attitude, and those who are strongly inclined to refuse vaccines. In addition, a participatory/guiding approach provides an opportunity for ongoing clinical rapport and dialogue between unvaccinated or under-vaccinated patients and their HCPs, even when it does not result in immediate vaccine uptake. Regardless of which approach is used, a strong vaccine recommendation should be made at each encounter.

Transparency and Accuracy Vaccine recommendations should be transparent, should include accurate information about both the benefits and the risks of the vaccine, and should emphasize why the benefits outweigh the risks. For example, when evidence supports an association between a vaccine and an adverse event, the occurrence of the adverse event is often very rare and the event quickly resolves (Chap. 123). U.S. Federal law (under the National Childhood Vaccine Injury Act) requires HCPs to provide a copy of the current Vaccine Information Statement from the Centers for Disease Control and Prevention (CDC), which describes both benefits and risks of vaccines to an adult patient or to a child's parent/legal representative before vaccination.

CDC Vaccine Information Statements should not replace a discussion with the HCP. Depending on the provider and the patient, a description of benefits and risks may include words and numbers, graphics, and personal anecdotes (e.g., why the provider vaccinates his or her own children). Personal anecdotes are powerful, and many hesitant patients seek and are influenced by them.

A discussion of benefits and risks provides an opportunity to address specific misconceptions about a particular vaccine or about vaccines overall. For example, patients may be concerned about adverse events following vaccination that are not supported by evidence, such as autism following MMR vaccination or myocardial infarction following influenza vaccination in the elderly.

Most adults—even those whose children are fully immunized—still have questions, misconceptions, or concerns about vaccines that should be addressed. A risk/benefit discussion allows HCPs to describe the vaccine safety monitoring systems in place. Providers should emphasize that vaccines are developed and approved through a highly regulated process that includes prelicensure clinical trials, review and approval by designated regulatory authorities (e.g., the U.S. Food and Drug Administration, Health Canada), strict manufacturing regulations, and ongoing postmarketing safety surveillance.

Support from Accessible Information Sources All vaccine recommendations should be supported by additional information sources patients can assess after the health care encounter. HCPs play an important role as information intermediaries for their patients. They can navigate information (and misinformation) about vaccines and direct patients toward reliable, appropriate resources. HCPs should consider what resources will be suitable for a patient or patient population. Vaccine information resources are available in different media formats and use a combination of images and text to communicate the information to various audiences. See "Further Reading," below, for suggestions or refer to resources provided by local health authorities.

Revisiting and Reinforcement of Vaccine Recommendations All health care encounters offer an opportunity to revisit and reinforce vaccine recommendations. Vaccine-hesitant individuals who do not accept vaccines but are willing to review information should be offered a follow-up appointment to reinforce previously made recommendations and address further questions. Vaccine-hesitant

patients who accept vaccines should be seen at a follow-up appointment to confirm and document vaccine receipt (if vaccine is not given at the point of care), ascertain whether the vaccine was well tolerated, and reinforce the message about vaccine safety and effectiveness. Patients who actively demand vaccines usually do not require much follow-up other than to confirm and document the receipt of vaccine (if it is not given at the point of care) and to address additional questions or concerns arising subsequent to vaccination. Often this follow-up can be covered without an office visit.

WHAT TO SAY TO VACCINE-HESITANT PATIENTS

Engaging vaccine-hesitant individuals requires confidence, knowledge, skills, time, and creativity to tailor the approach to each individual patient. Examples for each part of the vaccine recommendation are listed in **Table 3-2**.

■ OTHER CONSIDERATIONS DURING CLINICAL ENCOUNTERS

Missed Opportunities The World Health Organization defines a missed opportunity for vaccination as “any contact with health services by an individual (child or person of any age) who is eligible

for vaccination (e.g., unvaccinated or partially vaccinated and free of contraindications to vaccination), which does not result in the person receiving one or more of the vaccine doses for which he or she is eligible.” HCPs who do not offer point-of-care vaccination frequently miss the opportunity to recommend vaccines to their patients. Missed opportunities for recommending and providing vaccines during routine health care encounters contribute to under-vaccination. Studies show that up to 45% of under-vaccinated children could be up to date with all age-appropriate vaccines and up to 90% of female adolescents could be up to date with human papillomavirus (HPV) vaccination if all opportunities to vaccinate were taken.

Vaccine counseling and vaccination should be incorporated into clinical care for individuals of all ages, not just young children. Because many adolescents and adults do not have regular health care follow-up, providers need to take advantage of every health care encounter to recommend and provide vaccines. For example, a visit to an emergency department, a routine follow-up visit at a diabetes clinic, or a visit planning for elective orthopedic surgery offer opportunities to inquire about the patient’s vaccination status and to recommend vaccines.

HCPs should make preemptive vaccine recommendations (e.g., initiating discussions about infant vaccines during pregnancy, informing parents about HPV vaccine before their child becomes eligible). Such advance discussions may be especially helpful in identifying

TABLE 3-2 Sample Vaccine Conversations

STRONG VACCINE RECOMMENDATION

“We are headed into the flu season. Getting flu vaccine not only protects you, but it helps protect other people around you who can get very sick from flu. I strongly recommend you get your flu shot. Do you know where to get it?”

“You will be turning 50 next year. This means you will be eligible for a vaccine that prevents shingles, and I strongly recommend you receive it. Have you heard about this vaccine before? Can I answer your questions about it?”

“I know you are not comfortable getting vaccinated today. I do want to make it clear that I recommend vaccines because I am convinced they are the best way to protect you from some serious diseases. Is there something that would lead you to think about getting vaccinated in the future?”

TAILORED COMMUNICATION

“I recommend that children and adults stay up to date on recommended vaccines. I see from your vaccine record that you’ve had your childhood vaccines, but you haven’t gotten any adult vaccines. I wanted to clarify whether this is because you decided not to get vaccines or something else prevented you from getting vaccinated.”

“I understand that you are here for your pneumococcal vaccine. This is the best way to protect yourself and those around you from pneumonia. Do you have any questions before I give you the vaccine?”

“I understand you have some concerns about vaccines. What are you most concerned about? Would you like me to explain why I recommend giving your child these vaccines?”

TRANSPARENCY AND ACCURACY

“Serious side effects can develop after MMR vaccination but are very rare. On average, 3 out of 10,000 children who get MMR vaccine will have a febrile seizure/convulsion in the days after vaccination. Febrile seizures can be frightening, but nearly all children who have a febrile seizure recover very quickly and without any long-term consequences. On the other hand, 1 out of 1000 children who get measles will develop encephalitis (brain inflammation) that not only causes seizures but can also lead to permanent damage.”

“About 10 out of every 10,000 Americans who do not get vaccinated against flu die because of influenza every year, and many more are hospitalized. While flu vaccine does not prevent all cases of influenza, it is the most effective vaccine we have. By getting the vaccine, you also help protect people around you from getting sick.”

“You are correct, aluminum is used in some vaccines to help the body’s immune system respond. However, aluminum is also present in food and drinking water. In fact, the amount of aluminum present in vaccines is similar to or less than what is present in breast milk or infant formulas.”

SUPPORT FROM ACCESSIBLE INFORMATION SOURCES

“Your child and other boys and girls his age will be eligible for the human papillomavirus vaccine this coming school year. Have you heard about this vaccine before? What questions do you have about it? Here’s a list of websites for parents and teenagers that explain what it is about.”

“There’s a lot of information about vaccines on the internet, and a lot of that information is not based on facts. Here is a list of websites that have been reviewed by health care professionals and accurately describe benefits and risks of each vaccine. The information is written in lay language and includes helpful illustrations.”

REVISITING AND REINFORCEMENT OF THE RECOMMENDATION

“During our last visit, we talked about MMR vaccine for your son and some of the concerns you had about potential side effects. Have you had a chance to look at the take-home information I gave you? Was there anything else you would like to ask about? I recommend that we vaccinate your child today.”

“During our last visit, we talked about receiving a pertussis booster during pregnancy and where you can get vaccinated. Have you had a chance to get your pertussis vaccine?”

“I see that you got your vaccines at the public health clinic last week. How did it go? Did you have any questions?”

“It’s possible that the symptoms you experienced after receiving the vaccine were an adverse reaction to the vaccine. I will report this to the health authority. Let’s discuss what we can do next time to prevent symptoms from occurring again.”

Note: Specific vaccine recommendations, vaccine eligibility guidelines, and statistics used to communicate benefits and risks will vary with the health jurisdiction and the country. Several sample statements here are adapted from the Australian National Centre for Immunisation Research and Surveillance website (www.talkingaboutimmunisation.org.au). For patient vaccine information resources, see also the Immunization Action Coalition website for the public developed in partnership with the CDC (vaccineinformation.org).

vaccine-hesitant patients and ensuring that they have enough time to ask questions and make decisions before vaccines are due.

HCPs should ensure that a vaccine recommendation is followed by vaccination. Providers who recommend vaccines but do not vaccinate at the point of care should inform patients where they can be vaccinated. This discussion may include information about public health clinics, travel clinics, and pharmacies or a referral to another provider. HCPs should follow up with their patients at subsequent appointments to confirm that they were vaccinated.

Adverse Events Following Vaccination Although rare, adverse events ([Chap. 123](#)) may influence vaccine acceptance and willingness to be vaccinated in the future. It is important for providers to identify and follow up with all patients who experience an adverse event, regardless of the patients' vaccine attitudes prior to the event. Adverse events following vaccination should be reported to the relevant vaccine monitoring system: the U.S. Vaccine Adverse Event Reporting System or the Canadian Adverse Event Following Immunization Surveillance System.

Addressing Inequities In Vaccine Access Discrepancies in access to health care services create inequitable access to vaccines for children and adults and contribute to under-vaccination. A U.S. study found that socially disadvantaged individuals were more likely than other persons to be under-vaccinated, in part because of a lack of access to health care services. HCPs must recognize that socially disadvantaged individuals and populations are often at greater risk of vaccine-preventable diseases (e.g., as a result of crowded living conditions, limited access to sanitation, poor nutrition, or substance abuse) and also at greater risk of being under-vaccinated because they have limited access to health care services. In addition, specific vaccines may be recommended for some socially disadvantaged populations or communities. For example, in the wake of several outbreaks of hepatitis A among the U.S. homeless population, the CDC now recommends that everyone >1 year of age experiencing homelessness receive hepatitis A vaccine.

Depending on the setting and the patient, some recommended vaccines may not be covered through public funding or private insurance coverage. HCPs should be aware of alternative funding models, such as the Vaccines for Children Program, which provides free vaccines for U.S. children (<19 years of age) with financial barriers to vaccine access. When vaccines are not publicly funded or covered by private insurance and patients perceive that they cannot afford a vaccine, HCPs should not withhold a vaccine recommendation. The risks and benefits of vaccination still need to be communicated, with a strong recommendation, and the patient should be provided the opportunity to decide whether they can afford the vaccine.

Further Communication With Patients Who Refuse Vaccines Fortunately, the proportion of people who completely refuse all vaccines and are not willing to talk to their HCP is small. Nevertheless, in some cases, attempts to initiate discussion and address vaccine refusal may be futile. When possible, HCPs should focus on the common goals of care and preserve the therapeutic relationship. Vaccine refusal should be well documented in the patient's chart. The HCP should continue with tailored communication and be open to future discussions. Vaccine demand and vaccine refusal are rarely static over time. ([See “Focus: COVID-19 Vaccine Hesitancy,” below.](#))

■ CONCLUSION

In summary, vaccine hesitancy is complex and context specific. It varies with time, place, patient, and vaccine. HCPs are well positioned to address vaccine hesitancy and should develop the skills, knowledge, and confidence to make strong vaccine recommendations to their patients.

■ FOCUS: COVID-19 VACCINE HESITANCY

As COVID-19 vaccines are used to control SARS-CoV-2, some individuals will have concerns about these vaccines and a proportion of

the population will reject them. While worrisome, hesitancy about COVID-19 vaccines is not unexpected; it mirrors public concerns expressed about past pandemic influenza vaccines and other newly introduced vaccines. It has been established that the newness of any vaccine, be it a pandemic influenza vaccine or a COVID-19 vaccine, raises concern in a large percentage of the population. Politicization of COVID-19 vaccines raises additional issues for some patients.

Past Experience with New Vaccines Past experience with new vaccines, including the H1N1 pandemic influenza vaccine in 2009 and the human papillomavirus vaccine in the early 2000s, provides a guide to topics that need to be addressed with regard to COVID-19 vaccines. While resistance is often framed as uncertainty about a vaccine's "newness," further discussion translates this uncertainty into concern about the new vaccine's safety. This concern encompasses both short- and long-term side effects. Frequent, acute adverse effects can be captured in clinical trial data, whereas worries about rare and long-term side effects can be addressed only by direct evidence after the initiation of a new vaccination program. In addition to queries about the overall safety of the vaccine, HCPs can expect specific questions regarding the safety of individual ingredients included in the vaccine, whether or not these ingredients are new and whether or not relevant safety data are available. Information on the incidence of common or expected health events in an unvaccinated population (i.e., background rates) over a 4-week period is helpful in distinguishing what is normal and expected from a point of concern. Studies that have examined this issue with regard to other vaccines can be used as a basis for presenting background rates of expected events in the context of COVID-19 vaccines for some groups; however, it is important to ensure that more specific background-rate information is available to HCPs with regard to the individual groups being vaccinated. HCPs, public health programs, and vaccine manufacturers can anticipate these questions and should develop answers and information to respond to them.

Specific Concerns about COVID-19 Vaccines While some concerns can be anticipated on the basis of past experience with new vaccines, several characteristics of COVID-19 vaccines require new approaches to adequately address individual concerns, and HCPs need to educate themselves in several specific areas. First, an overwhelming amount of attention has been paid to the speed of development of COVID-19 vaccines, with some jurisdictions even skipping the usual clinical-trial steps in an effort to provide vaccine more rapidly to their populations. This situation directly increases concerns about the "newness" of the vaccine and its safety and, unfortunately, raises questions about the entire vaccine development process. Education is required to explain how a process that normally requires 5–10 years was condensed to this degree. (See Lurie et al [2020] for an excellent explanation of the COVID-19 vaccine development process.) In addition, transparency with regard to clinical trial data is required to enable scientists, HCPs, and consumers to read and understand the development and evaluation processes. The usually shrouded, proprietary development process is unsuitable if the final vaccine product is to garner public trust. Education on existing vaccine-safety monitoring systems also needs to be provided. HCPs must familiarize themselves with the vaccine development process and safety monitoring systems if they are to present this information to their patients.

Second, several newer vaccine platforms that are being used for COVID-19 vaccines (e.g., nucleic acid-based vaccines, viral vector) have not been used in the past. This novelty exacerbates public concern about the unfamiliarity of new vaccines and further heightens misgivings about vaccine safety and the potential for long-term adverse effects. Again, HCPs need to familiarize themselves with the new technology and develop effective messaging for their patients. Public health officials have developed resources to address this issue ([see \[www.cdc.gov/vaccines/covid-19/vaccinate-with-confidence.html\]\(http://www.cdc.gov/vaccines/covid-19/vaccinate-with-confidence.html\)](http://www.cdc.gov/vaccines/covid-19/vaccinate-with-confidence.html)), but, even in the absence of such resources, HCPs can anticipate questions about the new technology involved and become comfortable explaining it.

Third, clinical trial safety and efficacy data were lacking for all groups initially prioritized to receive the vaccine. For example, long-term-care residents were prioritized for vaccine receipt, but clinical trial data were not available for the range of chronic health conditions that exist in older adults. While observational studies have filled some of these gaps, HCPs need to extrapolate on the basis of available evidence in considering individual patients and must make a recommendation without knowing all the answers.

Fourth, some minority and marginalized communities who have been disproportionately affected by COVID-19 express hesitancy or reject COVID-19 vaccines. For some Black, Indigenous, Latinx, and other communities, COVID-19 hesitancy stems directly from systematic discrimination, racism, and mistreatment in the health care system. Black and Indigenous communities also share a horrific legacy of unethical medical experimentation,¹ which, when combined with current discrimination and overt racism, creates a powerful climate of mistrust in HCPs, the medical system, and science.

Social and Cultural Trends The social and cultural trends already discussed in this chapter—in particular, traditional media, the Internet, and social media—are exerting influence and pressure that did not affect the introduction of older vaccines, even the H1N1 pandemic vaccines. The media attention given to the development of transverse myelitis in one clinical-trial participant following receipt of COVID-19 vaccine is but one example of the intense media scrutiny of the vaccine development process. Unfortunately, in the United States, efforts to control COVID-19, including vaccine development, have become highly politicized. This degree of politicization has not occurred with past vaccines, so HCPs are in uncharted territory in terms of how to address it or even to understand its potential influence on vaccine acceptance. Again, individual HCPs need to navigate complex conversations with their patients and possibly their communities. Below are some suggestions that may prove helpful in formulating these conversations.

Tips for Discussion of COVID-19 Vaccines • ADDRESS CONCERNs ABOUT “NEWNESS” HCPs need to understand and be able to explain the newer vaccine platforms (mRNA, DNA, and viral vector vaccines) and to provide examples of other, older vaccines that have been developed by similar techniques. This information makes COVID-19 vaccines more familiar.

ADDRESS CONCERNs ABOUT VACCINE SAFETY HCPs need to understand and explain how vaccines are evaluated before being approved for use and how vaccine safety is monitored after vaccines are used in the population. It is important to be honest and state that potential rare and long-term effects are not yet known, but then to speak to what is from the animal and clinical trial data and to comment on background rates for rare events. Placing potential vaccine risks in the context of known COVID-19 disease risks is helpful for some patients.

Depending on the context, explain why specific high-risk groups may have been prioritized to receive the vaccine. Patients who have been prioritized may still need a strong recommendation from an HCP to accept the vaccine. An HCP recommendation is as important here as it is for acceptance of routine vaccines. As with other vaccines, many patients' decision to accept a COVID-19 vaccine rests upon whether their HCP recommends it.

¹The Tuskegee Syphilis Study is the most infamous example of medical experimentation in Black communities in the United States. (See Brandt [1978] for details.) Numerous examples of medical experimentation on Indigenous peoples are available. For example, a 12-year trial of an experimental bacille Calmette-Guérin vaccine for tuberculosis was conducted on Cree and Nakoda Oyadebi infants in Saskatchewan during the 1930s. (See Lux [2016] for details.)

Address implicit or overt racism and systemic discrimination in the medical system and create culturally safe health care spaces. HCPs need to be aware of the legacy of discrimination, racism, and medical experimentation and the distrust it fosters in some communities. While SARS-CoV-2 has critically highlighted fractures in our health care system for minority and marginalized communities, addressing these underlying issues goes beyond addressing vaccine hesitancy and is clearly needed for all types of medical care in these communities.

EMPHASIZE THE IMPORTANCE OF KEEPING UP TO DATE WITH OTHER ROUTINE VACCINES DURING THE COVID-19 PANDEMIC These vaccines include but are not limited to seasonal influenza vaccine and the childhood primary vaccination series.

FURTHER READING

Vaccine Hesitancy

- AMERICAN ACADEMY OF PEDIATRICS: Vaccine hesitant parents. Available at www.aap.org/en-us/advocacy-and-policy/aap-health-initiatives/immunizations/Pages/vaccine-hesitant-parents.aspx. Accessed October 23, 2020.
- DESTEFANO F et al: Principal controversies in vaccine safety in the United States. *Clin Infect Dis* 69:726, 2019.
- DUDLEY MZ et al: The state of vaccine safety science: Systematic reviews of the evidence. *Lancet Infect Dis* 20:e80, 2020.
- IMMUNIZATION ACTION COALITION: For healthcare professionals. Available at www.immunize.org. Accessed October 23, 2020.
- IMMUNIZATION ACTION COALITION: For the public: Vaccine information you need. Available at vaccineinformation.org. Accessed October 23, 2020.
- JAMISON AM et al: Vaccine-related advertising in the Facebook Ad Archive. *Vaccine* 38:512, 2020.
- LEASK J et al: Communicating with parents about vaccination: A framework for health professionals. *BMC Pediatr* 12:154, 2012.
- MACDONALD N et al: Vaccine hesitancy: Definition, scope and determinants. *Vaccine* 33:4161, 2015.
- WORLD HEALTH ORGANIZATION: Vaccine hesitancy survey questions related to SAGE vaccine hesitancy. Available at www.who.int/immunization/programmes_systems/Survey_Questions_Hesitancy.pdf. Accessed October 23, 2020.
- WORLD HEALTH ORGANIZATION: Improving vaccination demand and addressing hesitancy. Available at www.who.int/immunization/programmes_systems/vaccine_hesitancy/en/. Accessed October 23, 2020.
- WORLD HEALTH ORGANIZATION: Missed opportunities for vaccination (MOV) strategy. Available at www.who.int/immunization/programmes_systems/policies_strategies/MOV/en/. Accessed October 23, 2020.

COVID-19 Vaccine Hesitancy

- BRANDT AM: Racism and research: The case of the Tuskegee Syphilis Study. *Hastings Cent Rep* 8:21, 1978.
- CENTERS FOR DISEASE CONTROL AND PREVENTION: Vaccinate with confidence: Strategy to reinforce confidence in Covid-19 vaccines. Available at www.cdc.gov/vaccines/covid-19/vaccinate-with-confidence.html. Accessed April 5, 2021.
- LURIE N et al: Developing Covid-19 vaccines at pandemic speed. *N Engl J Med* 382:21, 2020.
- LUX MK: Separate beds: A history of Indian hospitals in Canada, 1920s–1980s. Toronto, University of Toronto Press, 2016.
- MOSBY I et al: Medical experimentation and the roots of COVID-19 vaccine hesitancy among Indigenous Peoples in Canada. *CMAJ* 193:E381, 2021.

4

Decision-Making in Clinical Medicine

Daniel B. Mark, John B. Wong

Practicing medicine at its core requires making decisions. What makes medical practice so difficult is not only the specialized technical knowledge required but also the intrinsic uncertainty that surrounds each decision. Mastering the technical aspects of medicine alone, unfortunately, does not ensure a mastery of the practice of medicine. Sir William Osler's familiar quote "Medicine is a science of uncertainty and an art of probability" captures well this complex duality. Although the science of medicine is often taught as if the mechanisms of the human body operate with Newtonian predictability, every aspect of medical practice is infused with an element of irreducible uncertainty that the clinician ignores at her peril. Although deeply rooted in science, more than 100 years after the practice of medicine took its modern form, it remains at its core a craft, to which individual doctors bring varying levels of skill and understanding. With the exponential growth in medical literature and other technical information and an ever-increasing number of testing and treatment options, twenty-first century physicians who seek excellence in their craft must master a more diverse and complex set of skills than any of the generations that preceded them. This chapter provides an introduction to three of the pillars upon which the craft of modern medicine rests: (1) expertise in clinical reasoning (what it is and how it can be developed); (2) rational diagnostic test use and interpretation; and (3) integration of the best available research evidence with clinical judgment in the care of individual patients (evidence-based medicine [EBM]).

BRIEF INTRODUCTION TO CLINICAL REASONING

Clinical Expertise Defining "clinical expertise" remains surprisingly difficult. Chess has an objective ranking system based on skill and performance criteria. Athletics, similarly, have ranking systems to distinguish novices from Olympians. But in medicine, after physicians complete training and pass the boards (or get recertified), no tests or benchmarks are used to identify those who have attained the highest levels of clinical performance. At each institution, there are often a few "elite" clinicians who are known for their "special problem-solving prowess" when particularly difficult or obscure cases have baffled everyone else. Yet despite their skill, even such master clinicians typically cannot explain their exact processes and methods, thereby limiting the acquisition and dissemination of the expertise used to achieve their impressive results. Furthermore, clinical virtuosity appears not to be generalizable, e.g., an expert on hypertrophic cardiomyopathy may be no better (and possibly worse) than a first-year medical resident at diagnosing and managing a patient with neutropenia, fever, and hypotension.

Broadly construed, clinical expertise encompasses not only cognitive dimensions involving the integration of disease knowledge with verbal and visual cues and test interpretation but also potentially the complex fine-motor skills necessary for invasive procedures and tests. In addition, "the complete package" of expertise in medicine requires effective communication and care coordination with patients and members of the medical team. Research on medical expertise remains sparse overall and mostly centered on diagnostic reasoning, so in this chapter, we focus primarily on the cognitive elements of clinical reasoning.

Because clinical reasoning occurs in the heads of clinicians, objective study of the process is difficult. One research method used for this area asks clinicians to "think out loud" as they receive increments of clinical information in a manner meant to simulate a clinical encounter. Another research approach focuses on how doctors should reason diagnostically, to identify remediable "errors," rather than on how they actually do reason. Much of what is known about clinical

reasoning comes from empirical studies of nonmedical problem-solving behavior. Because of the diverse perspectives contributing to this area, with important contributions from cognitive psychology, medical education, behavioral economics, sociology, informatics, and decision sciences, no single integrated model of clinical reasoning exists, and not infrequently, different terms and reasoning models describe similar phenomena.

Intuitive Versus Analytic Reasoning A useful contemporary model of reasoning, the dual-process theory distinguishes two general conceptual modes of thinking as fast or slow. *Intuition* (System 1) provides rapid effortless judgments from memorized associations using pattern recognition and other simplifying "rules of thumb" (i.e., heuristics). For example, a very simple pattern that could be useful in certain situations is "black woman plus hilar adenopathy equals sarcoid." Because no effort is involved in recalling the pattern, the clinician is often unable to say how those judgments were formulated. In contrast, *Analysis* (System 2), the other form of reasoning in the dual-process model, is slow, methodical, deliberative, and effortful. A student might read about causes of hilar adenopathy and from that list (e.g., Chap. 66), identify diseases more common in black women or examine the patient for skin or eye findings that occur with sarcoid. These dual processes, of course, represent two exemplars taken from the cognitive continuum. They provide helpful descriptive insights but very little guidance in how to develop expertise in clinical reasoning. How these idealized systems interact in different decision problems, how experts use them differently from novices, and when their use can lead to errors in judgment remain the subject of study and considerable debate.

Pattern recognition, an important part of System 1 reasoning, is a complex cognitive process that appears largely effortless. One can recognize people's faces, the breed of a dog, an automobile model, or a piece of music from just a few notes within milliseconds without necessarily being able to articulate the specific features that prompted the recognition. Analogously, experienced clinicians often recognize familiar diagnostic patterns very quickly. The key here is having a large library of stored patterns that can be rapidly accessed. In the absence of an extensive stored repertoire of diagnostic patterns, students (as well as experienced clinicians operating outside their area of expertise and familiarity) often must use the more laborious System 2 analytic approach along with more intensive and comprehensive data collection to reach the diagnosis.

The following brief patient scenarios illustrate three distinct patterns associated with hemoptysis that experienced clinicians recognize without effort:

- A 46-year-old man presents to his internist with a chief complaint of hemoptysis. An otherwise healthy, nonsmoker, he is recovering from an apparent viral bronchitis. This presentation pattern suggests that the small amount of blood-streaked sputum is due to acute bronchitis, so that a chest x-ray provides sufficient reassurance that a more serious disorder is absent.
- In the second scenario, a 46-year-old patient who has the same chief complaint but with a 100-pack-year smoking history, a productive morning cough with blood-streaked sputum, and weight loss fits the pattern of carcinoma of the lung. Consequently, along with the chest x-ray, the clinician obtains a sputum cytology examination and refers this patient for a chest CT scan.
- In the third scenario, the clinician hears a soft diastolic rumbling murmur at the apex on cardiac auscultation in a 46-year-old patient with hemoptysis who immigrated from a developing country and orders an echocardiogram as well, because of possible pulmonary hypertension from suspected rheumatic mitral stenosis.

Pattern recognition by itself is not, however, sufficient for secure diagnosis. Without deliberative systematic reflection, undisciplined pattern recognition can result in premature closure: mistakenly jumping to the conclusion that one has the correct diagnosis before all the relevant data are in. A critical second step, therefore, even when the diagnosis seems obvious, is *diagnostic verification*: considering whether

the diagnosis adequately accounts for the presenting symptoms and signs and can explain all the ancillary findings. The following case based on a real clinical encounter provides an example of premature closure. A 45-year-old man presents with a 3-week history of a “flulike” upper respiratory infection (URI) including dyspnea and a productive cough. The emergency department (ED) clinician pulled out a “URI assessment form,” which defines and standardizes the information gathered. After quickly acquiring the requisite structured examination components and noting in particular the absence of fever and a clear chest examination, the physician prescribed a cough suppressant for acute bronchitis and reassured the patient that his illness was not serious. Following a sleepless night at home with significant dyspnea, the patient developed nausea and vomiting and collapsed. He was brought back to the ED in cardiac arrest and was unable to be resuscitated. His autopsy showed a posterior wall myocardial infarction (MI) and a fresh thrombus in an atherosclerotic right coronary artery. What went wrong? Presumably, the ED clinician felt that the patient was basically healthy (one can be misled by the way the patient appears on examination—a patient that does not appear “sick” may be incorrectly assumed to have an innocuous illness). So, in this case, the physician, upon hearing the overview of the patient from the triage nurse, elected to use the URI assessment protocol even before starting the history, closing consideration of the broader range of possibilities and associated tests required to confirm or refute these possibilities. In particular, by concentrating on the abbreviated and focused URI protocol, the clinician failed to elicit the full dyspnea history, which was precipitated by exertion and accompanied by chest heaviness and relieved by rest, suggesting a far more serious disorder.

Heuristics or rules of thumb are a part of the intuitive system. These cognitive shortcuts provide a quick and easy path to reaching conclusions and making choices, but when used improperly, they can lead to errors. Two major research programs have studied heuristics in a mostly nonmedical context and have reached very different conclusions about the value of these cognitive tools. The “heuristics and biases” program focuses on how these mental shortcuts can lead to incorrect judgments. So far, however, little evidence exists that educating physicians and other decision makers to watch for the >100 cognitive biases identified to date has had any effect on the rate of diagnostic errors. In contrast, the “fast and frugal heuristics” research program explores how and when relying on simple heuristics can produce good decisions. Although many heuristics have relevance to clinical reasoning, only four will be mentioned here.

When diagnosing patients, clinicians usually develop diagnostic hypotheses based on the similarity of that patient’s symptoms, signs, and other data to their mental representations (memorized patterns) of the disease possibilities. In other words, clinicians pattern match to identify the diagnoses that share the most similar findings to the patient at hand. This cognitive shortcut is called the representativeness heuristic. Consider a patient with hypertension who has headache, palpitations, and diaphoresis. Based on the representativeness heuristic, clinicians might judge pheochromocytoma to be quite likely given this classic presenting symptom triad suggesting pheochromocytoma. Doing so, however, would be incorrect given that other causes of hypertension are much more common than pheochromocytoma and this triad of symptoms can occur in patients who do not have it. Thus, clinicians using the representativeness heuristic may overestimate the likelihood of a particular disease based on the presence of representative symptoms and signs, failing to account for its low underlying prevalence (i.e., the prior, or pretest, probabilities). Conversely, atypical presentations of common diseases may lead to underestimating the likelihood of a particular disease. Thus, inexperience with a specific disease and with the breadth of its presentations may also lead to diagnostic delays or errors, e.g., diseases that affect multiple organ systems, such as sarcoid or tuberculosis, may be particularly challenging to diagnose because of the many different patterns they may manifest.

A second commonly used cognitive shortcut, the availability heuristic, involves judgments based on how easily prior similar cases or outcomes can be brought to mind. For example, a clinician may recall

a case from a morbidity and mortality conference in which an elderly patient presented with painless dyspnea of acute onset and was evaluated for a pulmonary cause but was eventually found to have acute MI, with the diagnostic delay likely contributing to the development of ischemic cardiomyopathy. If the case was associated with a malpractice accusation, such examples may be even more memorable. Errors with the availability heuristic arise from several sources of recall bias. Rare catastrophic outcomes become memorable cases with a clarity and force disproportionate to their likelihood for future diagnosis—for example, a patient with a sore throat eventually found to have leukemia or a young athlete with leg pain subsequently found to have an osteosarcoma—and those publicized in the media or recently experienced are, of course, easier to recall and therefore more influential on clinical judgments.

The third commonly used cognitive shortcut, the anchoring heuristic (also called conservatism or stickiness), involves insufficiently adjusting the initial probability of disease up (or down) following a positive (or negative test) when compared with Bayes’ theorem, i.e., sticking to the initial diagnosis. For example, a clinician may still judge the probability of coronary artery disease (CAD) to be high despite a negative exercise perfusion test and go on to cardiac catheterization (see “Measures of Disease Probability and Bayes’ Rule,” below).

The fourth heuristic states that clinicians should use the simplest explanation possible that will adequately account for the patient’s symptoms and findings (Occam’s razor or, alternatively, the simplicity heuristic). Although this is an attractive and often used principle, it is important to remember that no biologic basis for it exists. Errors from the simplicity heuristic include premature closure leading to the neglect of unexplained significant symptoms or findings.

For complex or unfamiliar diagnostic problems, clinicians typically resort to analytic reasoning processes (System 2) and proceed methodically using the *hypothetico-deductive model of reasoning*. Based on the patient’s stated reasons for seeking medical attention, clinicians develop an initial list of diagnostic possibilities in *hypothesis generation*. During the history of the present illness, the initial hypotheses evolve in *diagnostic refinement* as emerging information is tested against the mental models of the diseases being considered with diagnoses increasing and decreasing in likelihood or even being dropped from consideration as the working hypotheses of the moment. These mental models often generate additional questions that distinguish the diagnostic possibilities from one another. The focused physical examination contributes to further distinguishing the working hypotheses. Is the spleen enlarged? How big is the liver? Is it tender? Are there any palpable masses or nodules? *Diagnostic verification* involves testing the adequacy (whether the diagnosis accounts for all symptoms and signs) and coherency (whether the signs and symptoms are consistent with the underlying pathophysiologic causal mechanism) of the working diagnosis. For example, if the enlarged and quite tender liver felt on physical examination is due to acute hepatitis (the hypothesis), then certain specific liver function tests will be markedly elevated (the prediction). Should the tests come back normal, the hypothesis may have to be discarded and others reconsidered.

Although often neglected, negative findings are as important as positive ones because they reduce the likelihood of the diagnostic hypotheses under consideration. Chest discomfort that is not provoked or worsened by exertion and not relieved by rest in an active patient lowers the likelihood that chronic ischemic heart disease is the underlying cause. The absence of a resting tachycardia and thyroid gland enlargement reduces the likelihood of hyperthyroidism in a patient with paroxysmal atrial fibrillation.

The acuity of a patient’s illness may override considerations of prevalence and the other issues described above. “Diagnostic imperatives” recognize the significance of relatively rare but potentially catastrophic conditions if undiagnosed and untreated. For example, clinicians should consider aortic dissection routinely as a possible cause of acute severe chest discomfort. Although the typical presenting symptoms of dissection differ from those of MI, dissection may mimic MI, and because it is far less prevalent and potentially fatal if mistreated, diagnosing dissection remains a challenging diagnostic imperative (**Chap. 280**).

Clinicians taking care of acute, severe chest pain patients should explicitly and routinely inquire about symptoms suggestive of dissection, measure blood pressures in both arms for discrepancies, and examine for pulse deficits. When these are all negative, clinicians may feel sufficiently reassured to discard the aortic dissection hypothesis. If, however, the chest x-ray shows a possible widened mediastinum, the hypothesis should be reinstated and an appropriate imaging test ordered (e.g., thoracic computed tomography [CT] scan or transesophageal echocardiogram). In nonacute situations, the prevalence of potential alternative diagnoses should play a much more prominent role in diagnostic hypothesis generation.

Cognitive scientists studying the thought processes of expert clinicians have observed that clinicians group data into packets, or “chunks,” that are stored in short-term or “working memory” and manipulated to generate diagnostic hypotheses. Because short-term memory is limited (classically humans can accurately repeat a list of 7 ± 2 numbers read to them), the number of diagnoses that can be actively considered in hypothesis-generating activities is similarly limited. For this reason, the cognitive shortcuts discussed above play a key role in the generation of diagnostic hypotheses, many of which are discarded as rapidly as they are formed, thereby demonstrating that the distinction between analytic and intuitive reasoning is an arbitrary and simplistic, but nonetheless useful, representation of cognition.

Research into the hypothetico-deductive model of reasoning has had difficulty identifying the elements of the reasoning process that distinguish experts from novices. This has led to a shift from examining the problem-solving process of experts to analyzing the organization of their knowledge for pattern matching as exemplars, prototypes, and illness scripts. For example, diagnosis may be based on the resemblance of a new case to patients seen previously (exemplars). As abstract mental models of disease, prototypes incorporate the likelihood of various disease features. Illness scripts include risk factors, pathophysiology, and symptoms and signs. Experts have a much larger store of exemplar and prototype cases, an example of which is the visual long-term memory of experienced radiologists. However, clinicians do not simply rely on literal recall of specific cases but have constructed elaborate conceptual networks of memorized information or models of disease to aid in arriving at their conclusions (illness scripts). That is, expertise involves an enhanced ability to connect symptoms, signs, and risk factors to one another in meaningful ways; relate those findings to possible diagnoses; and identify the additional information necessary to confirm the diagnosis.

No single theory accounts for all the key features of expertise in medical diagnosis. Experts have more knowledge about presenting symptoms of diseases and a larger repertoire of cognitive tools to employ in problem solving than nonexperts. One definition of expertise highlights the ability to make powerful distinctions. In this sense, expertise involves a working knowledge of the diagnostic possibilities and those features that distinguish one disease from another. Memorization alone is insufficient, e.g., photographic memory of a medical textbook would not make one an expert. But having access to detailed case-specific relevant information is critically important. In the past, clinicians primarily acquired clinical knowledge through their patient experiences, but now clinicians have access to a plethora of information sources. Clinicians of the future will be able to leverage the experiences of large numbers of other clinicians using electronic tools, but, as with the memorized textbook, the data alone will be insufficient for becoming an expert. Nonetheless, availability of these data removes one barrier for acquiring experience with connecting symptoms, signs, and risk factors to the possible diagnoses and identifying the additional distinguishing information necessary to confirm the diagnosis, thereby potentially facilitating the development of the working knowledge necessary for becoming an expert.

Despite all of the research seeking to understand expertise in medicine and other disciplines, it remains uncertain whether any didactic program can actually accelerate the progression from novice to expert or from experienced clinician to master clinician. Deliberate effortful practice (over an extended period of time, sometimes said to be 10 years or 10,000 practice hours) and personal coaching are two strategies often used outside medicine (e.g., music, athletics, chess) to cultivate

expertise. Their use in developing medical expertise and maintaining or enhancing it has not yet been adequately explored. Some studies in medicine suggest that the most beneficial approach to education exposes students to both the signs and symptoms of specific diseases (disease pattern recognition) and, in addition, the lists of diseases that can present with specific symptoms and signs (differential diagnosis). Active learning opportunities useful for those in training include developing a personal learning system, e.g., systematically reflecting on diagnostic processes used (metacognition) and following-up to identify diagnoses and treatments for patients in their care.

■ DIAGNOSTIC VERSUS THERAPEUTIC DECISION-MAKING

The modern ideal of medical therapeutic decision-making is to “personalize” treatment recommendations. In the abstract, personalizing treatment involves combining the best available evidence about what works with an individual patient’s unique features (e.g., risk factors, genomics, and comorbidities) and his or her preferences and health goals to craft an optimal treatment recommendation with the patient. Operationally, two different and complementary levels of personalization are possible: individualizing the risk of harm and benefit for the options being considered based on the specific patient characteristics (precision medicine), and personalizing the therapeutic decision process by incorporating the patient’s preferences and values for the possible health outcomes. This latter process is sometimes referred to as shared decision-making and typically involves clinicians sharing their knowledge about the options and the associated consequences and trade-offs and patients sharing their health goals (e.g., avoiding a short-term risk of dying from coronary artery bypass grafting to see their grandchild get married in a few months).

Individualizing the evidence about therapy **does not** mean relying on physician impressions of benefit and harm from their personal experience. Because of small sample sizes and rare events, the chance of drawing erroneous causal inferences from one’s own clinical experience is very high. For most chronic diseases, therapeutic effectiveness is only demonstrable statistically in large patient populations. It would be incorrect to infer with any certainty, for example, that treating a hypertensive patient with angiotensin-converting enzyme (ACE) inhibitors necessarily prevented a stroke from occurring during treatment, or that an untreated patient would definitely have avoided their stroke had they been treated. For many chronic diseases, a majority of patients will remain event free regardless of treatment choices; some will have events regardless of which treatment is selected; and those who avoided having an event through treatment cannot be individually identified. Blood pressure lowering, a readily observable surrogate endpoint, does not have a tightly coupled relationship with strokes prevented. Consequently, in most situations, demonstrating therapeutic effectiveness cannot rely simply on observing the outcome of an individual patient but should instead be based on large groups of patients carefully studied and properly analyzed.

Therapeutic decision-making, therefore, should be based on the best available evidence from clinical trials and well-done outcome studies. Trustworthy clinical practice guidelines that synthesize such evidence offer normative guidance for many testing and treatment decisions. However, all guidelines recognize that “one size fits all” recommendations may not apply to individual patients. Increased research into the heterogeneity of treatment effects seeks to understand how best to adjust group-level clinical evidence of treatment harms and benefits to account for the absolute level of risks faced by subgroups and even by individual patients, using, for example, validated clinical risk scores.

■ NONCLINICAL INFLUENCES ON CLINICAL DECISION-MAKING

More than three decades of research on variations in clinician practice patterns has identified important nonclinical forces that shape clinical decisions. These factors can be grouped conceptually into three overlapping categories: (1) factors related to an individual physician’s practice, (2) factors related to practice setting, and (3) factors related to payment systems.

Factors Related to Practice Style To ensure that necessary care is provided at a high level of quality, physicians fulfill a key role in medical care by serving as the patient's advocate. Factors that influence performance in this role include the physician's knowledge, training, and experience. Clearly, physicians cannot practice EBM if they are unfamiliar with the evidence. As would be expected, specialists generally know the evidence in their field better than do generalists. Beyond published evidence and practice guidelines, a major set of influences on physician practice can be subsumed under the general concept of "practice style." The practice style serves to define norms of clinical behavior. Differing practice styles may be based on training, personal experience, and medical evidence. Beliefs about effectiveness of different therapies and preferred patterns of diagnostic test use are examples of different facets of a practice style. For example, cardiologists evaluating patients with lower risk chest pain symptoms often conceptualize their primary diagnostic objective as maximizing the detection of ischemia. For this reason, they may strongly favor stress imaging. Internists caring for the same patients may be more comfortable with initial use of exercise ECG testing without imaging. This latter practice style focuses less on ischemia detection and more on following guideline recommendations that indicate no outcome advantage for stress imaging in this context. Cardiologist may also favor a more liberal use of coronary angiography and revascularization in patients with stable ischemic symptoms relative to general internists.

Beyond the patient's welfare, physician perceptions about the risk of a malpractice suit resulting from either an erroneous decision or a bad outcome may drive clinical decisions and create a practice referred to as defensive medicine. This practice involves ordering tests and therapies with very small marginal benefits, ostensibly to preclude future criticism should an adverse outcome occur. With conscious or unconscious awareness of a connection to the risk of litigation or to payment, however, over time, such patterns of care may become accepted as part of the practice norm, thereby perpetuating their overuse, e.g., annual cardiac exercise testing in asymptomatic patients.

Practice Setting Factors Factors in this category relate to work systems including tasks and workflow (interruptions, inefficiencies, workload), technology (poor design or implementation, errors in use, failure, misuse), organizational characteristics (e.g., culture, leadership, staffing, scheduling), and the physical environment (e.g., noise, lighting, layout). *Physician-induced demand* is a term that refers to the repeated observation that once medical facilities and technologies become available to physicians, they will find ways to use them. Other environmental factors that can influence decision-making include the local availability of specialists for consultations and procedures; "high-tech" advanced imaging or procedure facilities such as MRI machines and proton beam therapy centers; and fragmentation of care.

Payment Systems Economic incentives are closely related to the other two categories of practice-modifying factors. Financial issues can exert both stimulatory and inhibitory influences on clinical practice. Historically, physicians are paid on a fee-for-service, capitation, or salary basis. In fee-for-service, physicians who do more get paid more, thereby encouraging overuse, consciously or unconsciously. When fees are reduced (discounted reimbursement), clinicians tend to increase the number of services provided to maintain revenue. Capitation, in contrast, provides a fixed payment per patient per year to encourage physicians to consider a global population budget in managing individual patients and ideally reducing the use of interventions with small marginal benefit. To discourage volume-based excessive utilization, fixed salary compensation plans pay physicians the same regardless of the clinical effort expended but may provide an (unintended) incentive to see fewer patients. In recognition of the nonsustainability of continued growth in medical expenditures and the opportunity costs associated with that (funds that might be more beneficially applied to education, energy, social welfare, or defense), current efforts seek to transition to a value-based payment system to reduce overuse and to reflect benefit. Work to define how to actually tie payment to value has mostly focused so far on "pay for performance" models. High-quality

clinical trial evidence for the effectiveness of these models is still mostly lacking.

■ INTERPRETATION OF DIAGNOSTIC TESTS

Despite impressive technological advances in medicine over the past century, uncertainty still abounds and challenges all aspects of medical decision-making. Compounding this challenge, massive information overload characterizes modern medicine. Clinicians on average subscribe to seven journals, presenting them with >2500 new articles each year, and need access to 2 million pieces of information to practice medicine. Of course, to be useful, this information must be sifted for quality and examined for applicability for integration into patient-specific care. Although computers appear to offer an obvious solution both for information management and for quantification of medical care uncertainties, many practical problems remain to be solved before computerized decision support can be routinely incorporated into the clinical reasoning process in a way that demonstrably improves the quality of care. For the present, understanding the nature of diagnostic test information can help clinicians become more efficient users of such data. The next section reviews select concepts related to diagnostic testing.

■ DIAGNOSTIC TESTING: MEASURES OF TEST ACCURACY

The purpose of performing a test on a patient is to reduce uncertainty about the patient's diagnosis or prognosis in order to facilitate appropriate management. Although diagnostic tests commonly refer to laboratory (e.g., blood count) or imaging tests or procedures (e.g., colonoscopy or bronchoscopy), any information that changes a provider's understanding of the patient's problem qualifies as a diagnostic test. Thus, even the history and physical examination can be considered as diagnostic tests. In clinical medicine, it is common to reduce the results of a test to a dichotomous outcome, such as positive or negative, normal or abnormal. Although this simplification often suppresses useful information (such as the degree of abnormality), it facilitates illustrating some important principles of test interpretation that are described below.

The accuracy of any diagnostic test is assessed relative to a "gold standard," where a positive gold standard test defines the patients who have disease and a negative test securely rules out disease (**Table 4-1**). Characterizing the diagnostic performance of a new test requires identifying an appropriate population (ideally, patients representative of those in whom the new test would be used) and applying both the new and the gold standard tests to all subjects. Biased estimates of test performance occur when diagnostic accuracy is defined using an inappropriate population or one in which gold standard determination of disease status is incomplete. The accuracy of the new test in distinguishing disease from health is determined relative to the gold standard results and summarized in four estimates. The sensitivity or true-positive rate reflects how well the new test identifies patients with disease. It is the proportion of patients with disease (defined by the gold standard) who have a positive test. The proportion of patients with disease who have a negative test is the false-negative rate, calculated as $1 - \text{sensitivity}$. The specificity, or true-negative rate, reflects how well

TABLE 4-1 Measures of Diagnostic Test Accuracy

TEST RESULT	DISEASE STATUS	
	PRESENT	ABSENT
Positive	True positives (TP)	False positives (FP)
Negative	False negatives (FN)	True negatives (TN)
Test Characteristics in Patients with Disease		
True-positive rate (sensitivity) = $TP/(TP + FN)$		
False-negative rate = $FN/(TP + FN) = 1 - \text{true-positive rate}$		
Test Characteristics in Patients without Disease		
True-negative rate (specificity) = $TN/(TN + FP)$		
False-positive rate = $FP/(TN + FP) = 1 - \text{true-negative rate}$		

the new test correctly identifies patients without disease. It is the proportion of patients without disease (defined by the gold standard) who have a negative test. The proportion of patients without disease who have positive test is the false-positive rate, calculated as $1 - \text{specificity}$. In theory, a perfect test would be one with a sensitivity of 100% and a specificity of 100% and would completely distinguish patients with disease from those without it. A useful mnemonic to help remember the somewhat paradoxical relationship between what the test is best at technically versus what it is most useful for clinically is: a test with a very high sensitivity (Sn) when *negative* (N) helps *rule out* (out) disease (Sn_{Nout}), and a test with a very high specificity (Sp) when *positive* (P) helps *rule in* (in) disease (Sp_{Pin}).

Calculating sensitivity and specificity requires selection of a threshold value or cut point above which the test is considered “positive.” Making the cut point “stricter” (e.g., raising it) lowers sensitivity but improves specificity, while making it “laxer” (e.g., lowering it) raises sensitivity but lowers specificity. This dynamic trade-off between more accurate identification of subjects with disease versus those without disease is often displayed graphically as a receiver operating characteristic (ROC) curve (Fig. 4-1) by plotting sensitivity (y axis) versus $1 - \text{specificity}$ (x axis). Each point on the curve represents a potential cut point with an associated sensitivity and specificity value. The area under the ROC curve often is used as a quantitative measure of the information content of a test. Values range from 0.5 (no diagnostic information from testing at all; the test is equivalent to flipping a coin) to 1.0 (perfect test). The choice of cut point should in theory reflect the relative harms and benefits of treatment for those without versus those with disease. For example, if treatment was safe with substantial benefit, then choosing a high-sensitivity cut point (upper right of the ROC curve) for a low-risk test may be appropriate (e.g., phenylketonuria in newborns), but if treatment had substantial risk for harm, then choosing a high-specificity cut point (lower left of the ROC curve) may be appropriate (e.g., chemotherapy for cancer). The choice of cut point may also depend on the prevalence of disease, with low prevalence placing a greater emphasis on the harms of false-positive tests (e.g., HIV testing in marriage applicants) or the harms of false-negative tests (e.g., HIV testing in blood donors).

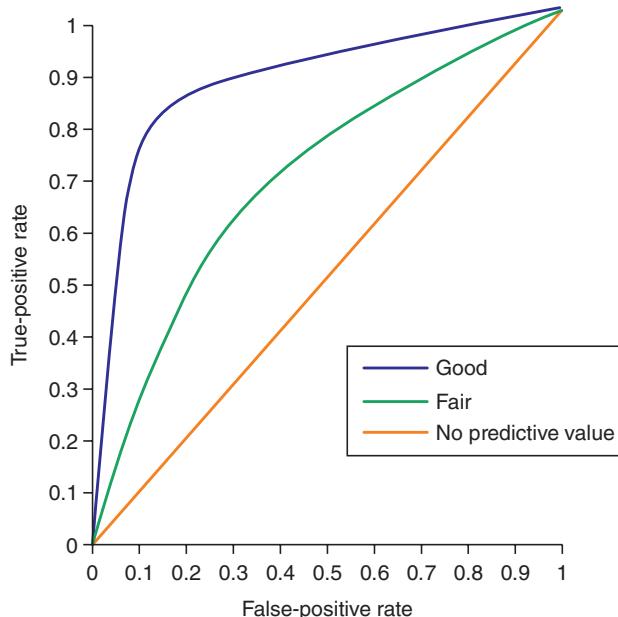


FIGURE 4-1 Each receiver operating characteristic (ROC curve) illustrates a trade-off that occurs between improved test sensitivity (accurate detection of patients with disease) and improved test specificity (accurate detection of patients without disease), as the test value defining when the test turns from “negative” to “positive” is varied. A 45° line would indicate a test with no predictive value (sensitivity = specificity at every test value). The area under each ROC curve is a measure of the information content of the test. Thus, a larger ROC area signifies increased diagnostic accuracy.

MEASURES OF DISEASE PROBABILITY AND BAYES’ RULE

In the absence of perfect tests, the true disease state of the patient remains uncertain after every test. Bayes’ rule provides a way to quantify the revised uncertainty using simple probability mathematics (and thereby avoid anchoring bias). It calculates the *posttest probability*, or likelihood of disease after a test result, from three parameters: the pretest probability of disease, the test sensitivity, and the test specificity. The *pretest probability* is a quantitative estimate of the likelihood of the diagnosis before the test is performed and is usually estimated from the prevalence of the disease in the underlying population (if known) or clinical context (e.g., age, sex, and type of chest pain). For some common conditions, such as CAD, existing nomograms and statistical models generate estimates of pretest probability that account for history, physical examination, and test findings. The posttest probability (also called the predictive value of the test, see below) is a recalibrated statement of the probability of the diagnosis, accounting for both pretest probability and test results. For the probability of disease following a positive test (i.e., positive predictive value), Bayes’ rule is calculated as:

$$\text{Posttest probability} = \frac{\text{Pretest probability} \times \text{test sensitivity}}{\text{Pretest probability} \times \text{test sensitivity} + (1 - \text{Pretest probability}) \times (\text{false-positive test rate})}$$

For example, consider a 64-year-old woman with atypical chest pain who has a pretest probability of 0.50 and a “positive” diagnostic test result (assuming test sensitivity = 0.90 and specificity = 0.90).

$$\text{Posttest probability} = \frac{(0.50)(0.90)}{(0.50)(0.90) + (0.50)(0.10)} = 0.90$$

The term *predictive value* has often been used as a synonym for the posttest probability. Unfortunately, clinicians commonly misinterpret reported predictive values as intrinsic measures of test accuracy rather than calculated probabilities. Studies of diagnostic test performance compound the confusion by calculating predictive values from the same sample used to measure sensitivity and specificity. Such calculations are misleading unless the test is applied subsequently to populations with exactly the same disease prevalence. For these reasons, the term *predictive value* is best avoided in favor of the more descriptive posttest probability following a positive or a negative test result.

The nomogram version of Bayes’ rule (Fig. 4-2) helps us to understand at a conceptual level how it estimates the posttest probability of disease. In this nomogram, the impact of the diagnostic test result is summarized by the likelihood ratio, which is defined as the ratio of the probability of a given test result (e.g., “positive” or “negative”) in a patient with disease to the probability of that result in a patient without disease, thereby providing a measure of how well the test distinguishes those with from those without disease.

The *likelihood ratio for a positive test* is calculated as the ratio of the true-positive rate to the false-positive rate (or sensitivity/[$1 - \text{specificity}$]). For example, a test with a sensitivity of 0.90 and a specificity of 0.90 has a likelihood ratio of $0.90/(1 - 0.90)$, or 9. Thus, for this hypothetical test, a “positive” result is 9 times more likely in a patient with the disease than in a patient without it. Most tests in medicine have likelihood ratios for a positive result between 1.5 and 20. Higher values are associated with tests that more substantially increase the posttest likelihood of disease. A very high likelihood ratio positive (>10) usually implies high specificity, so a positive high specificity test helps “rule in” disease (the “ Sp_{Pin} ” mnemonic introduced earlier). If sensitivity is excellent but specificity is less so, the likelihood ratio positive will be reduced substantially (e.g., with a 90% sensitivity but a 55% specificity, the likelihood ratio positive is 2.0).

The corresponding *likelihood ratio for a negative test* is the ratio of the false-negative rate to the true-negative rate (or [$1 - \text{sensitivity}$]/specificity).

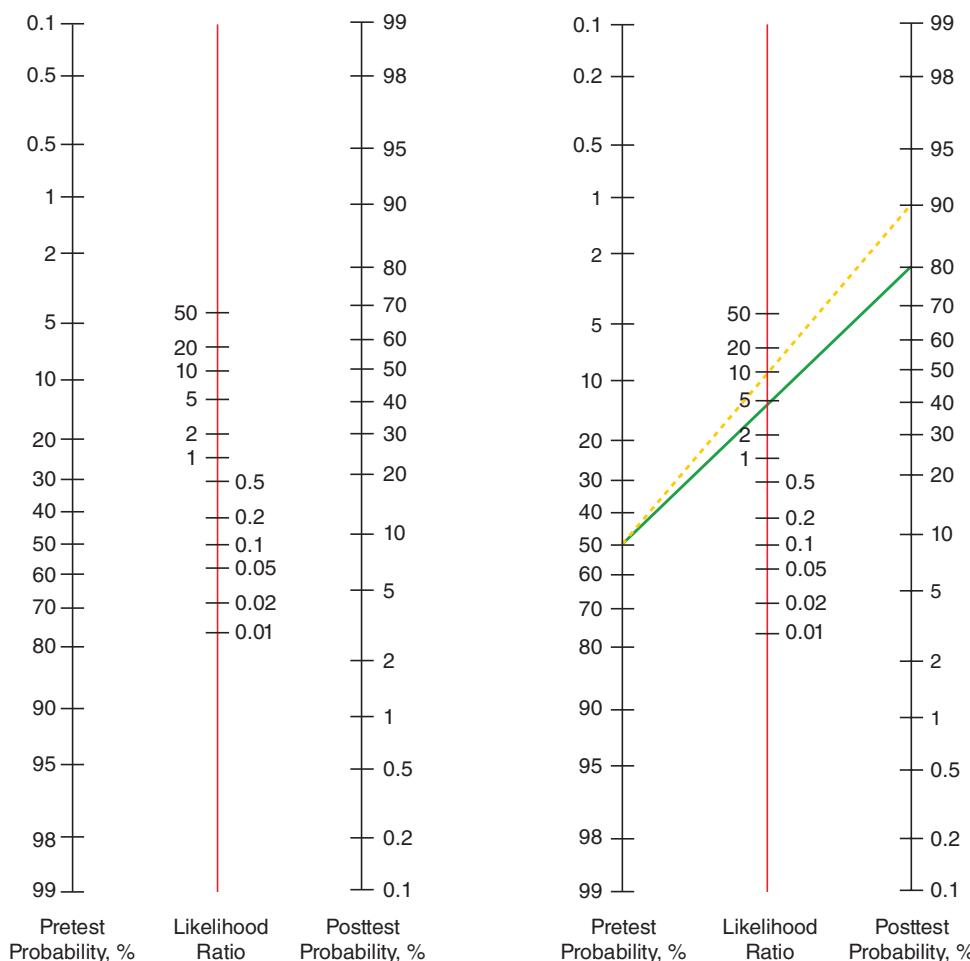


FIGURE 4-2 Nomogram version of Bayes' theorem used to predict the posttest probability of disease (right-hand scale) using the pretest probability of disease (left-hand scale) and the likelihood ratio for a positive or a negative test (middle scale). See text for information on calculation of likelihood ratios. To use, place a straightedge connecting the pretest probability and the likelihood ratio and read off the posttest probability. The right-hand part of the figure illustrates the value of a positive exercise treadmill test (likelihood ratio 4, green line) and a positive exercise thallium single-photon emission CT perfusion study (likelihood ratio 9, broken yellow line) in a patient with a pretest probability of coronary artery disease of 50%. (Adapted from Centre for Evidence-Based Medicine: Likelihood ratios. Available at <http://www.cebm.net/likelihood-ratios/>.)

Lower likelihood ratio negative values more substantially lower the posttest likelihood of disease. A very low likelihood ratio negative (falling below 0.10) usually implies high sensitivity, so a negative high sensitivity test helps “rule out” disease (the SnNout mnemonic). The hypothetical test considered above with a sensitivity of 0.9 and a specificity of 0.9 would have a likelihood ratio for a negative test result of $(1 - 0.9)/0.9$, or 0.11, meaning that a negative result is about one-tenth as likely in patients with disease than in those without disease (or about 10 times more likely in those without disease than in those with disease).

APPLICATIONS TO DIAGNOSTIC TESTING IN CAD

Consider two tests commonly used in the diagnosis of CAD: an exercise treadmill and an exercise single-photon emission CT (SPECT) myocardial perfusion imaging test (Chap. 241). A positive treadmill ST-segment response has an average sensitivity of ~60% and an average specificity of ~75%, yielding a likelihood ratio positive of 2.4 ($0.60/[1 - 0.75]$) (consistent with modest discriminatory ability because it falls between 2 and 5). For a 41-year-old man with nonanginal pain and a 10% pretest probability of CAD, the posttest probability of disease after a positive result rises to only ~30%. For a 60-year-old woman with typical angina and a pretest probability of CAD of 80%, a positive test result raises the posttest probability of disease to ~95%.

In contrast, exercise SPECT myocardial perfusion test is more accurate for diagnosis of CAD. For simplicity, assume that the finding of a reversible exercise-induced perfusion defect has both a sensitivity and

a specificity of 90% (a bit higher than reported), yielding a likelihood ratio for a positive test of 9.0 ($0.90/[1 - 0.90]$) (consistent with intermediate discriminatory ability because it falls between 5 and 10). For the same 10% pretest probability patient, a positive test raises the probability of CAD to 50% (Fig. 4-2). However, despite the differences in posttest probabilities between these two tests (30 vs 50%), the more accurate test may not improve diagnostic likelihood enough to change patient management (e.g., decision to refer to cardiac catheterization) because the more accurate test has only moved the physician from being fairly certain that the patient did not have CAD to a 50:50 chance of disease. In a patient with a pretest probability of 80%, exercise SPECT test raises the posttest probability to 97% (compared with 95% for the exercise treadmill). Again, the more accurate test does not provide enough improvement in posttest confidence to alter management, and neither test has improved much on what was known from clinical data alone.

In general, positive results with an accurate test (e.g., likelihood ratio for a positive test of 10) when the pretest probability is low (e.g., 20%) do not move the posttest probability to a range high enough to rule in disease (e.g., 80%). In screening situations, pretest probabilities are often particularly low because patients are asymptomatic. In such cases, specificity becomes especially important. For example, in screening first-time female blood donors without risk factors for HIV, a positive test raised the likelihood of HIV to only 67% despite a specificity of 99.995%

because the prevalence was 0.01%. Conversely, with a high pretest probability, a negative test may not rule out disease adequately if it is not sufficiently sensitive. Thus, the largest change in diagnostic likelihood following a test result occurs when the clinician is most uncertain (i.e., pretest probability between 30 and 70%). For example, in patients with a pretest probability for CAD of 50%, a positive exercise treadmill test moves the posttest probability to 80% and a positive exercise SPECT perfusion test moves it to 90% (Fig. 4-2).

As presented above, Bayes' rule employs a number of important simplifications that should be considered. First, few tests provide only “positive” or “negative” results. Many tests have multidimensional outcomes (e.g., extent of ST-segment depression, exercise duration, and exercise-induced symptoms with exercise testing). Although Bayes' theorem can be adapted to this more detailed test result format, it is computationally more complex to do so. Similarly, when multiple sequential tests are performed, the posttest probability may be used as the pretest probability to interpret the second test. However, this simplification assumes conditional independence—that is, that the results of the first test do not affect the likelihood of the second test result—and this is often not true.

Finally, many texts assert that sensitivity and specificity are prevalence-independent parameters of test accuracy. This statistically useful assumption, however, is often incorrect. A treadmill exercise test, for example, has a sensitivity of ~30% in a population of patients with one-vessel CAD, whereas its sensitivity in patients with severe three-vessel CAD approaches 80%. Thus, the best estimate of sensitivity

to use in a particular decision may vary, depending on the severity of disease in the local population. A hospitalized, symptomatic, or referral population typically has a higher prevalence of disease and, in particular, a higher prevalence of more advanced disease than does an outpatient population. Consequently, test sensitivity will likely be higher in hospitalized patients and test specificity higher in outpatients.

■ STATISTICAL PREDICTION MODELS

Bayes' rule, when used as presented above, is useful in studying diagnostic testing concepts, but predictions based on multivariable statistical models can more accurately address these more complex problems by simultaneously accounting for additional relevant patient characteristics. In particular, these models explicitly account for multiple, even possibly overlapping, pieces of patient-specific information and assign a relative weight to each on the basis of its unique independent contribution to the prediction in question. For example, a logistic regression model to predict the probability of CAD ideally considers all the relevant independent factors from the clinical examination and diagnostic testing and their relative importance instead of the limited data that clinicians can manage in their heads or with Bayes' rule. However, despite this strength, prediction models are usually too complex computationally to use without a calculator or computer. Guideline-driven treatment recommendations based on statistical prediction models available online, e.g., the American College of Cardiology/American Heart Association risk calculator for primary prevention with statins and the CHA₂DS₂-VASC calculator for anticoagulation for atrial fibrillation, have generated more widespread usage. When electronic health records (EHRs) will provide sufficient platform support to allow for routine use of predictive models in clinical practice and increase their impact on clinical encounters and outcomes remains uncertain.

One reason for limited clinical use is that, to date, only a handful of prediction models have been validated sufficiently (for example, Wells criteria for pulmonary embolism; **Table 4-2**). The importance of independent validation in a population separate from the one used to develop the model cannot be overstated. An unvalidated prediction model should be viewed with the skepticism appropriate for any new drug or medical device that has not had rigorous clinical trial testing.

When statistical survival models in cancer and heart disease have been compared directly with clinicians' predictions, the survival models have been found to be more consistent, as would be expected, but not always more accurate. On the other hand, comparison of clinicians with websites and apps that generate lists of possible diagnoses to help patients with self-diagnosis found that physicians outperformed the currently available programs. For students and less-experienced clinicians, the biggest value of diagnostic decision support may be in extending diagnostic possibilities and triggering "rational override," but their impact on knowledge, information-seeking, and problem-solving needs additional research.

TABLE 4-2 Wells Clinical Prediction Rule for Pulmonary Embolism (PE)

CLINICAL FEATURE	POINTS
Clinical signs of deep-vein thrombosis	3
Alternative diagnosis is less likely than PE	3
Heart rate >100 beats/min	1.5
Immobilization ≥3 days or surgery in previous 4 weeks	1.5
History of deep-vein thrombosis or pulmonary embolism	1.5
Hemoptysis	1
Malignancy (with treatment within 6 months) or palliative	1
INTERPRETATION	
Score >6.0	High
Score 2.0–6.0	Intermediate
Score <2.0	Low

FORMAL DECISION SUPPORT TOOLS

■ DECISION SUPPORT SYSTEMS

Over the past 50 years, many attempts have been made to develop computer systems to aid clinical decision-making and patient management. Conceptually, computers offer several levels of potentially useful support for clinicians. At the most basic level, they provide ready access to vast reservoirs of information, which may, however, be quite difficult to sort through to find what is needed. At higher levels, computers can support care management decisions by making accurate predictions of outcome, or can simulate the whole decision process, and provide algorithmic guidance. Computer-based predictions using Bayesian or statistical regression models inform a clinical decision but do not actually reach a "conclusion" or "recommendation." Machine learning methods are being applied to pattern recognition tasks such as the examination of skin lesions and the interpretation of x-rays. Artificial intelligence (AI) systems attempt to simulate or replace human reasoning with a computer-based analogue. Natural language processing allows the system to access and process large amounts of data, both from the EHR and from the medical literature. To date, such approaches have achieved only limited success. The most prominent example, IBM's Watson program, introduced publicly in 2011, has yet to produce persuasive evidence of clinical decision support utility. Reminder or protocol-directed systems do not make predictions but use existing algorithms, such as guidelines or appropriate utilization criteria, to direct clinical practice. In general, however, decision support systems have so far had little impact on practice. Reminder systems built into EHRs have shown the most promise, particularly in correcting drug dosing and promoting adherence to guidelines. Checklists may also help avoid or reduce errors.

■ DECISION ANALYSIS

Compared with the decision support methods discussed earlier, decision analysis represents a normative prescriptive approach to decision-making in the face of uncertainty. Its principal application is in complex decisions. For example, public health policy decisions often involve *trade-offs* in length versus quality of life, benefits versus resource use, population versus individual health, and *uncertainty* regarding efficacy, effectiveness, and adverse events as well as *values* or preferences regarding mortality and morbidity outcomes.

One recent analysis using this approach involved the optimal screening strategy for breast cancer, which has remained controversial, in part because a randomized controlled trial to determine when to begin screening and how often to repeat screening mammography is impractical. In 2016, the National Cancer Institute–sponsored Cancer Intervention and Surveillance Network (CISNET) examined eight strategies differing by whether to initiate mammography screening at age 40, 45, or 50 years and whether to screen annually, biennially, or annually for women in their forties and biennially thereafter (hybrid). The six simulation models found biennial strategies to be the most efficient for average-risk women. Biennial screening for 1000 women from age 50–74 years versus no screening avoided seven breast cancer deaths. Screening annually from age 40–74 years avoided three additional deaths but required 20,000 additional mammograms and yielded 1988 more false-positive results. Factors that influenced the results included patients with a 2–4-fold higher risk for developing breast cancer in whom annual screening from age 40–74 years yielded similar benefits as biennial screening from age 50–74. For average-risk patients with moderate or severe comorbidities, screening could be stopped earlier, at age 66–68 years.

This analysis involved six models that reproduced epidemiologic trends and a screening trial result, accounted for digital technology and treatments advances, and considered quality of life, risk factors, breast density, and comorbidity. It provided novel insights into a public health problem in the absence of a randomized clinical trial and helped weigh the pros and cons of such a health policy recommendation. Although such models have been developed for selected clinical problems, their benefit and application to individual real-time clinical management has yet to be demonstrated.

DIAGNOSIS AS AN ELEMENT OF QUALITY OF CARE

High-quality medical care begins with accurate diagnosis. The incidence of diagnostic errors has been estimated by a variety of methods including postmortem examinations, medical record reviews, and medical malpractice claims, with each yielding complementary but different estimates of this quality of care patient-safety problem. In the past, diagnostic errors tended to be viewed as a failure of individual clinicians. The modern view is that they are mostly a system of care deficiencies. Current estimates suggest that nearly everyone will experience at least one diagnostic error in their lifetime, leading to mortality, morbidity, unnecessary tests and procedures, costs, and anxiety.

Solutions to the “diagnostic errors as a system of care” problem have focused on system-level approaches, such as decision support and other tools integrated into EHRs. The use of checklists has been proposed as a means of reducing some of the cognitive errors discussed earlier in the chapter, such as premature closure. While checklists have been shown to be useful in certain medical contexts, such as operating rooms and intensive care units, their value in preventing diagnostic errors that lead to patient adverse events remains to be shown.

EVIDENCE-BASED MEDICINE

Clinical medicine is defined traditionally as a practice combining medical knowledge (including scientific evidence), intuition, and judgment in the care of patients (Chap. 1). Evidence-based medicine (EBM) updates this construct by placing much greater emphasis on the processes by which clinicians gain knowledge of the most up-to-date and relevant clinical research to determine for themselves whether medical interventions alter the disease course and improve the length or quality of life. The phrase “evidence-based medicine” is now used so often and in so many different contexts that many practitioners are unaware of its original meaning. The intention of the EBM program, as described in the early 1990s by its founding proponents at McMaster University, becomes clearer through an examination of its four key steps:

1. Formulating the management question to be answered
2. Searching the literature and online databases for applicable research data
3. Appraising the evidence gathered with regard to its validity and relevance
4. Integrating this appraisal with knowledge about the unique aspects of the patient (including the patient’s preferences about the possible outcomes)

The process of searching the world’s research literature and appraising the quality and relevance of studies can be time-consuming and requires skills and training that most clinicians do not possess. In a busy clinical practice, the work required is also logistically not feasible. This has led to a focus on finding recent systematic overviews of the problem in question as a useful shortcut in the EBM process. Systematic reviews are regarded by some as the highest level of evidence in the EBM hierarchy because they are intended to comprehensively summarize the available evidence on a particular topic. To avoid the potential biases found in narrative review articles, predefined reproducible explicit search strategies and inclusion and exclusion criteria seek to find all of the relevant scientific research and grade its quality. The prototype for this kind of resource is the Cochrane Database of Systematic Reviews. When appropriate, a meta-analysis is used to quantitatively summarize the systematic review findings (discussed further below).

Unfortunately, systematic reviews are not uniformly the acme of the EBM process they were initially envisioned to be. In select circumstances, they can provide a much clearer picture of the state of the evidence than is available from any individual clinical report, but their value is less clear when only a few trials are available, when trials and observational studies are mixed, or when the evidence base is only observational. They cannot compensate for deficiencies in the underlying research available, and many are created without the requisite clinical insights. The medical literature is now flooded with systematic

reviews of varying quality and clinical utility. The peer review system has, unfortunately, not proved to be an effective arbiter of quality of these papers. Therefore, systematic reviews should be used with circumspection in conjunction with selective reading of some of the best empirical studies.

SOURCES OF EVIDENCE: CLINICAL TRIALS AND REGISTRIES

The notion of learning from observation of patients is as old as medicine itself. Over the past 50 years, physicians’ understanding of how best to turn raw observation into useful evidence has evolved considerably. Medicine has received a hard refresher lesson in this process from COVID-19 pandemic. Starting in the spring of 2020, case reports, personal and institutional anecdotal experience, and small single-center case series started appearing in the peer-reviewed literature and within months turned into a flood of confusing and often contradictory evidence. Observational reports of treatments for COVID-19 fueled the confusion. Despite >40,000 publications appearing in the first 7 months of the pandemic, an enormous amount of uncertainty around prevention, diagnosis, treatment, and prognosis of the disease remained. Many of the early 2020 publications were either small observational series or reviews of published series, neither of which can resolve the key uncertainties clinicians need to address in caring for these patients. These small observational studies often have substantial limitations in validity and generalizability, and although they may generate important hypotheses or be the first reports of adverse events or therapeutic benefit, they have no role in formulating modern standards of practice. The major tools used to develop reliable evidence consist of randomized clinical trials supplemented strategically by large (high-quality) observational registries. A registry or database typically is focused on a disease or syndrome (e.g., different types of cancer, acute or chronic CAD, pacemaker capture, or chronic heart failure), a clinical procedure (e.g., bone marrow transplantation, coronary revascularization), or an administrative process (e.g., claims data used for billing and reimbursement).

By definition, in observational data, the investigator does not control patient care. Carefully collected prospective observational data, however, can at times achieve a level of evidence quality approaching that of major clinical trial data. At the other end of the spectrum, data collected retrospectively (e.g., chart review) are limited in form and content to what previous observers recorded and may not include the specific research data being sought (e.g., claims data). Advantages of observational data include the inclusion of a broader population as encountered in practice than is typically represented in clinical trials because of their restrictive inclusion and exclusion criteria. In addition, observational data provide primary evidence for research questions when a randomized trial cannot be performed. For example, it would be difficult to randomize patients to test diagnostic or therapeutic strategies that are unproven but widely accepted in practice, and it would be unethical to randomize based on sex, racial/ethnic group, socioeconomic status, or country of residence or to randomize patients to a potentially harmful intervention, such as smoking or deliberately overeating to develop obesity.

A well-done prospective observational study of a particular management strategy differs from a well-done randomized clinical trial most importantly by its lack of protection from treatment selection bias. The use of observational data to compare diagnostic or therapeutic strategies assumes that sufficient uncertainty and heterogeneity exists in clinical practice to ensure that similar patients will be managed differently by diverse physicians. In short, the analysis assumes that a sufficient element of randomness (in the sense of disorder rather than in the formal statistical sense) exists in clinical management. In such cases, statistical models attempt to adjust for important imbalances to “level the playing field” so that a fair comparison among treatment options can be made. When management is clearly not random (e.g., all eligible left main CAD patients are referred for coronary bypass surgery), the problem may be too confounded (biased) for statistical correction, and observational data may not provide reliable evidence.

In general, the use of concurrent controls is vastly preferable to that of historical controls. For example, comparison of current surgical management of left main CAD with medically treated patients with left main CAD during the 1970s (the last time these patients were routinely treated with medicine alone) would be extremely misleading because “medical therapy” has substantially improved in the interim.

Randomized controlled clinical trials include the careful prospective design features of the best observational data studies but also include the use of random allocation of treatment. This design provides the best protection against measured and unmeasured confounding due to treatment selection bias (a major aspect of internal validity). However, the randomized trial may not have good external validity (generalizability) if the process of recruitment into the trial resulted in the exclusion of many potentially eligible subjects or if the nominal eligibility for the trial describes a very heterogeneous population.

Consumers of medical evidence need to be aware that randomized trials vary widely in their quality and applicability to practice. The process of designing such a trial often involves many compromises. For example, trials designed to gain U.S. Food and Drug Administration (FDA) approval for an investigational drug or device must fulfill regulatory requirements (such as the use of a placebo control) that may result in a trial population and design that differ substantially from what practicing clinicians would find most useful.

META-ANALYSIS

The Greek prefix *meta* signifies something at a later or higher stage of development. Meta-analysis is research that combines and summarizes the available evidence quantitatively. Although it is used to examine nonrandomized studies, meta-analysis is most useful for summarizing all available randomized trials examining a particular therapy used in a specific clinical context. Ideally, unpublished trials should be identified and included to avoid publication bias (i.e., missing “negative” trials that may not be published). Furthermore, the best meta-analyses obtain and analyze individual patient-level data from all trials rather than using only the summary data from published reports. Nonetheless, not all published meta-analyses yield reliable evidence for a particular problem, so their methodology should be scrutinized carefully to ensure proper study design and analysis. The results of a well-done meta-analysis are likely to be most persuasive if they include at least several large-scale, properly performed randomized trials. Meta-analysis can especially help detect benefits when individual trials are inadequately powered (e.g., the benefits of streptokinase thrombolytic therapy in acute MI demonstrated by ISIS-2 in 1988 were evident by the early 1970s through meta-analysis). However, in cases in which the available trials are small or poorly done, meta-analysis should not be viewed as a remedy for deficiencies in primary trial data or trial design.

Meta-analyses typically focus on summary measures of relative treatment benefit, such as odds ratios or relative risks. Clinicians should also examine what absolute risk reduction (ARR) can be expected from the therapy. A metric of absolute treatment benefit that is frequently reported is the number needed to treat (NNT) to prevent one adverse outcome event (e.g., death, stroke). NNT should not be interpreted literally as a causal statement. NNT is simply 1/ARR. For example, if a hypothetical therapy reduced mortality rates over a 5-year follow-up by 33% (the relative treatment benefit) from 12% (control arm) to 8% (treatment arm), the ARR would be 12% – 8% = 4% and the NNT would be 1/0.04, or 25. This does not mean literally that 1 patient benefits and 24 do not. However, it can be conceptualized as an informal measure of treatment efficiency. If the hypothetical treatment was applied to a lower-risk population, say, with a 6% 5-year mortality, the 33% relative treatment benefit would reduce absolute mortality by 2% (from 6% to 4%), and the NNT for the same therapy in this lower-risk group of patients would be 50. Although not always made explicit, comparisons of NNT estimates from different studies should account for the duration of follow-up used to create each estimate. In addition, the NNT concept assumes a homogeneity in response to treatment that may not be accurate. The NNT is simply another way of summarizing the absolute treatment difference and does not provide any unique information.

CLINICAL PRACTICE GUIDELINES

Per the 1990 Institute of Medicine definition, clinical practice guidelines are “systematically developed statements to assist practitioner and patient decisions about appropriate health care for specific clinical circumstances.” This definition emphasizes several crucial features of modern guideline development. First, guidelines are created by using the tools of EBM. In particular, the core of the development process is a systematic literature search followed by a review of the relevant peer-reviewed literature. Second, guidelines usually are focused on a clinical disorder (e.g., diabetes mellitus, stable angina pectoris) or a health care intervention (e.g., cancer screening). Third, the primary objective of guidelines is to improve the quality of medical care by identifying care practices that should be routinely implemented, based on high-quality evidence and high benefit-to-harm ratios for the interventions. Guidelines are intended to “assist” decision-making, not to define explicitly what decisions should be made in a particular situation, in part because guideline-level evidence alone is never sufficient for clinical decision-making (e.g., deciding whether to intubate and administer antibiotics for pneumonia in a terminally ill individual, in an individual with dementia, or in an otherwise healthy 30-year-old mother).

Guidelines are narrative documents constructed by expert panels whose composition often is determined by interested professional organizations. These panels vary in expertise and in the degree to which they represent all relevant stakeholders. The guideline documents consist of a series of specific management recommendations, a summary indication of the quantity and quality of evidence supporting each recommendation, an assessment of the benefit-to-harm ratio for the recommendation, and a narrative discussion of the recommendations. Many recommendations simply reflect the expert consensus of the guideline panel because literature-based evidence is insufficient or absent. A recent examination of this issue in cardiovascular guidelines showed that <15% of guideline recommendations were based on the highest level of clinical trial evidence, and this proportion had not improved in 10 years despite a substantial number of trials being conducted and published. The final step in guideline construction is peer review, followed by a final revision in response to the critiques provided.

Guidelines are closely tied to the process of quality improvement in medicine through their identification of evidence-based best practices. Such practices can be used as quality indicators. Examples include the proportion of acute MI patients who receive aspirin upon admission to a hospital and the proportion of heart failure patients with a depressed ejection fraction treated with an ACE inhibitor.

CONCLUSIONS

Thirty years after the introduction of the EBM movement, it is tempting to think that all the difficult decisions practitioners face have been or soon will be solved and digested into practice guidelines and computerized reminders. However, EBM provides practitioners with an ideal rather than a finished set of tools with which to manage patients. Moreover, even with such evidence, it is always worth remembering that the response to therapy of the “average” patient represented by the summary clinical trial outcomes may not be what can be expected for the specific patient sitting in front of a provider in the clinic or hospital. In addition, meta-analyses cannot generate evidence when there are no adequate randomized trials, and most of what clinicians confront in practice will never be thoroughly tested in a randomized trial. For the foreseeable future, excellent clinical reasoning skills and experience supplemented by well-designed quantitative tools and a keen appreciation for the role of individual patient preferences in their health care will continue to be of paramount importance in the practice of clinical medicine.

FURTHER READING

- CROSKERRY P: A universal model of diagnostic reasoning. *Acad Med* 84:1022, 2009.
- DHALIWAL G, DETSKY AS: The evolution of the master diagnostician. *JAMA* 310:579, 2013.

- FANAROFF AC et al: Levels of evidence supporting American College of Cardiology/American Heart Association and European Society of Cardiology Guidelines, 2008-2018. *JAMA* 321:1069, 2019.
- HUNINK MGM et al: *Decision Making in Health and Medicine: Integrating Evidence and Values*, 2nd ed. Cambridge, Cambridge University Press, 2014.
- KAHNEMAN D: *Thinking Fast and Slow*. New York, Farrar, Straus and Giroux, 2013.
- KASSIRER JP et al: *Learning Clinical Reasoning*, 2nd ed. Baltimore, Lippincott Williams & Wilkins, 2009.
- MANDELBLATT JS et al: Collaborative modeling of the benefits and harms of associated with different U.S. breast cancer screening strategies. *Ann Intern Med* 164:215, 2016.
- MONTEIRO S et al: The 3 faces of clinical reasoning: Epistemological explorations of disparate error reduction strategies. *J Eval Clin Pract* 24:666, 2018.
- MURTHY VK et al: An inquiry into the early careers of master clinicians. *J Grad Med Educ* 10:500, 2018.
- RICHARDS JB et al: Teaching clinical reasoning and critical thinking: From cognitive theory to practical application. *Chest* 158:1617, 2020.
- ROYCE CS et al: Teaching critical thinking: A case for instruction in cognitive biases to reduce diagnostic errors and improve patient safety. *Acad Med* 94:187, 2019.
- SAPOSNIK G et al: Cognitive biases associated with medical decisions: A systematic review. *BMC Med Inform Decis Mak* 16:138, 2016.
- SCHUWIRTH LWT et al: Assessment of clinical reasoning: three evolutions of thought. *Diagnosis (Berl)* 7:191, 2020.

5

Precision Medicine and Clinical Care

The Editors



DISEASE NOSOLOGY AND PRECISION MEDICINE

Modern disease nosology arose in the late nineteenth century and represented a clear departure from the holistic, limited descriptions of disease dating to Galen. In this rubric, the definition of any disease is largely based on clinicopathologic observation. As the correlation between clinical signs and symptoms with pathoanatomy required autopsy material, diseases tended to be characterized by the end organ in which the primary syndrome was manifest and by late-stage presentations. Morgagni institutionalized this framework with the publication of *De Sedibus et Causis Morborum per Anatomam Indagatis* in 1761, in which he correlated the clinical features of patients with more than 600 autopsies at the University of Padua, demonstrating an anatomic basis for disease pathophysiology. Clinicopathologic observation served as the basis for inductive generalization coupled with the application of Occam's razor in which disease complexity was reduced to its simplest possible form. While this approach to defining human disease has held sway for over a century and facilitated the conquest of many diseases previously considered incurable, overly inclusive and simplified Oslerian diagnostics suffer from significant shortcomings. These include, but are not limited to, failure to distinguish the underlying etiology of different diseases with common pathophenotypes. For example, many different diseases can cause end-stage kidney disease or heart failure. Over time, the classification of neurodegenerative disorders or lymphomas, as well as many other diseases, is becoming more refined and precise as the underlying etiologies are identified. These distinctions are important for providing predictable prognostic information for individual patients with even highly prevalent diseases. Additionally, therapies may be ineffective owing to a lack of understanding of the often subtle molecular complexities of specific disease drivers.

Beginning in the mid-twentieth century, the era of molecular medicine offered the idealized possibility of identifying the underlying molecular basis of every disease. Using a conventional reductionist paradigm, physician-scientists explored disease mechanism at ever-increasing molecular depth, seeking the single (or limited number of) molecular cause(s) of many human diseases. Yet, as effective as this now conventional scientific approach was at uncovering many disease mechanisms, the clinical manifestations of very few diseases could be explained on the basis of a single molecular mechanism. Even knowledge of the globin β chain mutation that causes sickle cell disease does not predict the many different manifestations of the disease (stroke syndrome, painful crises, and hemolytic crisis, among others). Clearly, the profession had expected too much from oversimplified reductionism and failed to take into consideration the extraordinary biologic variety and its accompanying molecular and genetic complexity that underpin both normal and pathologic diversity. The promise of the Human Genome Project provided new tools and approaches and unleashed efforts to identify a monogenic, oligogenic, or polygenic cause for every disease (allowing for environmental modulation). Yet, once again, disappointment reigned as the pool of genomes expanded without the expected revelations (aside from rare variants). The arc of progressive reductionism (as illustrated for tuberculosis in Fig. 5-1) in refining and explaining disease reached a humbling plateau, revealing the need for new approaches to understand better the etiology, manifestations, and progression of most diseases. The stage was set for a return to holism. However, in contrast to the holism of ancient physicians, we adopted one that is integrative, taking genomic context into account in all dimensions. In the course of elaborating this complex pathobiologic landscape, disease definition must become more precise and progressively more individualized, setting the stage for what we term *precision medicine*.

Oversimplification of phenotype is a natural outgrowth of the observational scientific method. Categorizing individuals as falling into groups or clusters that are reasonably similar simplifies the task of the diagnostician and also facilitates the application of "specific" therapies more broadly. Biomedicine has been viewed as less quantitative and precise than other scientific disciplines, with biologic and pathobiologic diversity (biologic "noise") viewed as the norm. Thus, distilling such observational complexity to a fundamental group of symptoms or signs that are reasonably invariant across a group of sick individuals has served as the basis for the approach to disease and its treatment since the earliest days of medicine. This approach to diagnosis and therapy has remained in place into the twenty-first century, serving as the basis for the development of standard diagnostic tests and of broadly applied drug therapies. Targeting larger groups of patients is efficient when applied to large populations. As successful as this approach has been in advancing medical care, it is important to point out its limitations, which include significant predictive inaccuracies and sizeable segments of the disease population who do not respond to the most "effective" drugs (upward of 60% by some estimates). Clearly, a more nuanced approach to diagnosis and therapy is required to achieve better prognostic and therapeutic outcomes.

Turning first to phenotype, astute clinicians know full well the subtle and vivid differences in presentation that are often manifest among individuals with the same disease. In some cases, these differences in pathophenotype lead to new subclassifications of the disease, such as heart failure with preserved ejection fraction versus heart failure with reduced ejection fraction. Often, these relatively crude efforts at making diagnoses more precise are driven by new technologies or new ways of applying established technologies. In other cases, differences in pathophenotype are more subtle, not necessarily clinically apparent, and often driven by measures of endophenotype, such as distinctions among vasculitides facilitated by refinements in serologies or immunophenotyping. The impetus to create these subclasses of disease is largely determined by the need to improve prognosis and apply more precise and effective therapies. Based on these guiding principles, many experienced clinicians will argue—and rightly so—that they have been practicing personalized, precision medicine throughout their careers: they characterize each patient's illness in great detail, and choose therapies that respect and are guided by those individualized clinical and laboratory features, limited though they may be.

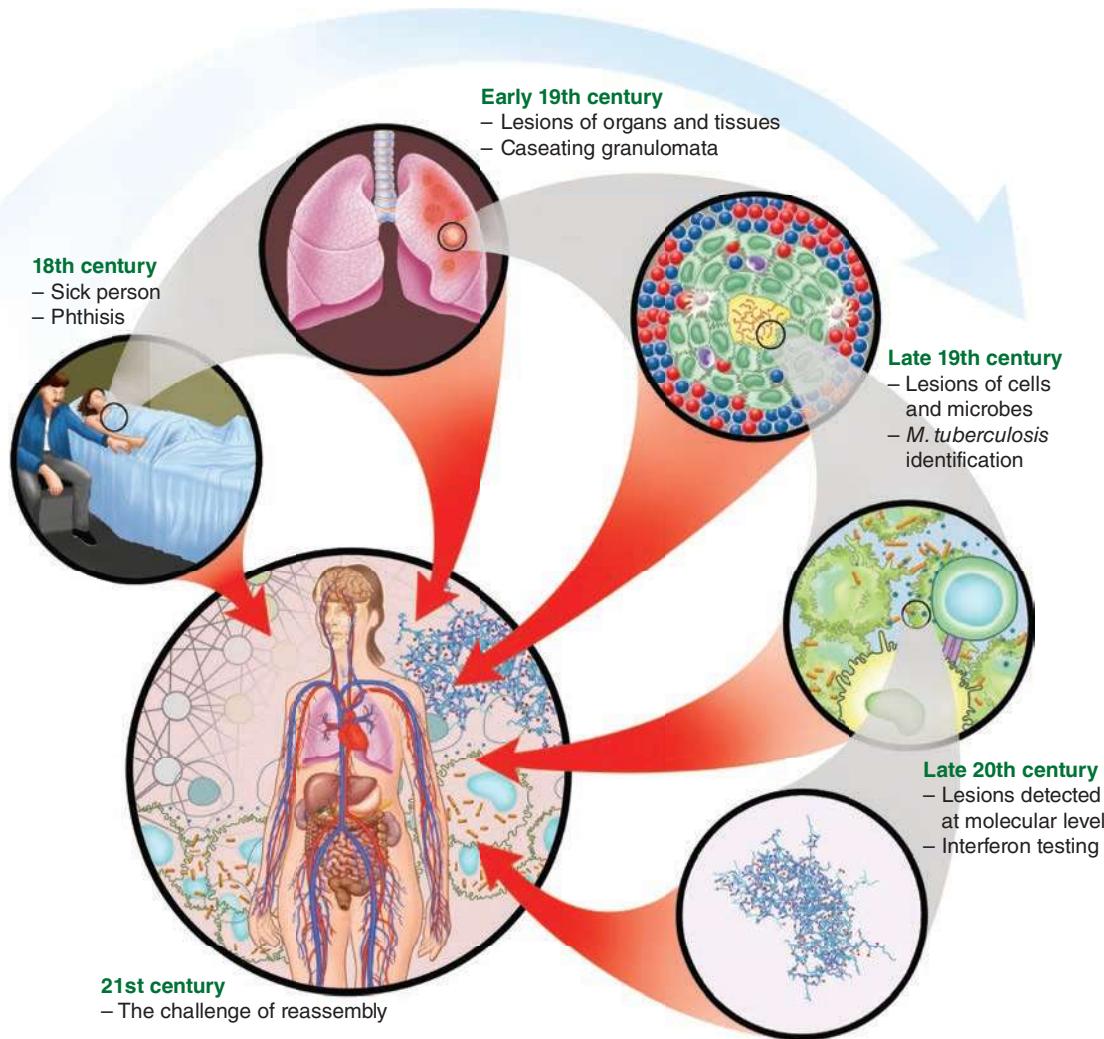


FIGURE 5-1 Arc of reductionism in medicine. (From JA Greene, J Loscalzo. Putting the patient back together—social medicine, network medicine, and the limits of reductionism. *N Engl J Med* 377:2493, 2017. Copyright © 2017 Massachusetts Medical Society. Reprinted with permission from Massachusetts Medical Society.)

For many diseases, genomic variation, whether inherited or acquired, provides opportunities to refine diagnostic precision with even greater fidelity and predictive accuracy. For this reason, the field of precision medicine has now entered a new era that couples the molecular reductionism of the last century with an integrative, systems-level understanding of the basis for pathophenotype. Equally important, modern genomics has established that genomic context, sometimes referred to as modifier genes, is distinctive for each individual person; hence, understanding that context provides the insight necessary to predict how a primary disease driver or drivers may manifest a clinical pathophenotype—e.g., why some individuals with sickle cell anemia will develop stroke, while others will develop acute chest syndrome. *This concept that primary genetic and/or environmental drivers of a disease differentially affect disease expression based on an individual's unique genomic context serves as the ultimate basis for much of what we denote as precision medicine.*

To develop a precision medicine strategy for any disease, the clinician needs to be aware of two important, confounding principles. First, patients with different diseases can manifest similar pathophenotypes, i.e., *convergent phenotypes*. Examples of this principle include the hypertrophied myocardium found in hypertrophic cardiomyopathy, infiltrative cardiomyopathies, critical aortic stenosis, and untreated, long-standing hypertension; and the thrombotic microangiopathy found in malignant hypertension, scleroderma renal crisis, thrombotic thrombocytopenic purpura, eclampsia, and antiphospholipid syndrome. Second, patients with the same basic disease can manifest very different pathophenotypes, i.e., *divergent phenotypes* (Chap. 466). Examples of this principle include the different clinical manifestations of cystic fibrosis or sickle cell disease and the incomplete penetrance of

many common genetic diseases. These common presentations of different diseases and different presentations of the same disease are both a consequence of genomic context coupled with unique exposures over an individual's lifetime (Fig. 5-2). Understanding the interplay among these many complex molecular determinants of disease expression is essential for the success of precision medicine.

Given the complexity of the genomic and environmental context of an individual, one must ask the question: How precise do we need to be in order to practice effective precision medicine? Complete knowledge of a person's comprehensive genome (DNA, gene expression, mitochondrial function, proteome, metabolome, posttranslational modification of the proteome, and metagenome, among others) and quantitative assessments of environmental and social history are not possible to acquire; yet, this shortcoming does not render the general problem intractable. Owing to the fact that the molecular networks that govern phenotype are overdetermined (i.e., redundant) and that there are primary drivers of disease expression that are modified in a weighted way by other genomic features of an individual, the practice of precision medicine can be realized without complete knowledge of all dimensions of the genome. Examples of how best to realize this strategy are discussed later in this chapter.

■ REQUIREMENTS FOR PRECISION MEDICINE

The essential elements of any precision medicine effort include phenotyping, endophenotyping (defining the characteristics of a disorder that are not readily observable), and genomic profiling (Fig. 5-3). While subtle distinctions among individuals with the same disease are well known to clinicians, formalizing these nuanced differences is critical for achieving more precise phenotypes. Deep phenotyping requires a

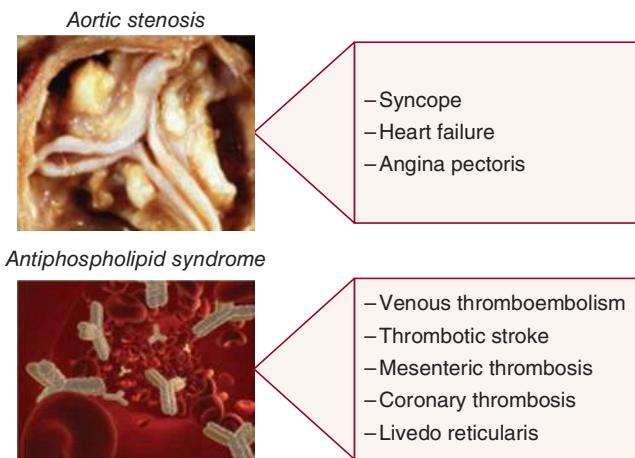
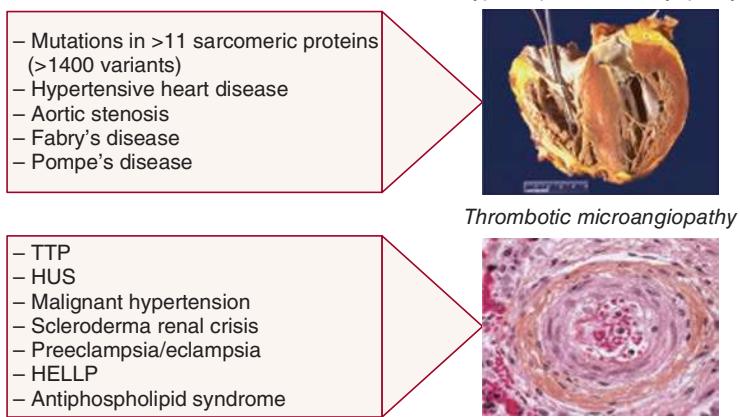


FIGURE 5-2 Convergent and divergent phenotypes. Examples of the former (**A**) include hypertrophic cardiomyopathy and thrombotic microangiopathy, and examples of the latter, and (**B**) include aortic stenosis and antiphospholipid syndrome, each of which can have several distinct clinical presentations. HELLP, hemolysis, elevated liver enzymes, and a low platelet count; HUS, hemolytic-uremic syndrome; TTP, thrombotic thrombocytopenic purpura.

detailed history, including family history and environmental exposures, as well as relevant (physiologic) functional studies and imaging, including molecular imaging where appropriate. Biochemical, immunologic, and molecular tests of body fluids provide additional detail to the overall phenotype. Importantly, these objective laboratory tests together with functional studies compose an assessment of the endophenotype (or endotype) of an individual, refining the overall discriminant power of the evaluation. One additional concept that has gained traction in recent years is the notion of orthogonal phenotyping, i.e., assessing clinical, molecular, imaging, or functional (endo)phenotypes seemingly unrelated to the clinical presentation. These features further enhance the ability to distinguish (sub) phenotypes and derive from the fact that diseases can be subtly (subclinically) manifest in organ systems different from that in which the primary symptoms or signs are expressed. While some diseases are well known to affect multiple organ systems (e.g., systemic lupus erythematosus) and in many cases involvement of those many systems is assessed at initial diagnosis, such is not the case for most other diseases. As we begin to understand the differences in the organ-specific expression of genomic variants that drive or modify disease, it is becoming increasingly apparent that orthogonal—or more appropriately, unbiased comprehensive—phenotyping should become the norm.

Genomic profiling must next be coupled to detailed phenotyping. The complex levels of genomic assessment continue to mature and include DNA sequencing (exomic, whole genome), gene expression (mRNA and protein expression), and metabolomics. In addition, the epigenome, the posttranslationally modified proteome, and the metagenome (the personal microbiome of an individual) are gaining traction as additional elements of comprehensive genomics (Chap. 483). Not all of these genomic features are yet available for clinical laboratory testing, and those that are available are largely confined to blood testing. While DNA sequencing using whole blood would generally apply to any organ-based disease,

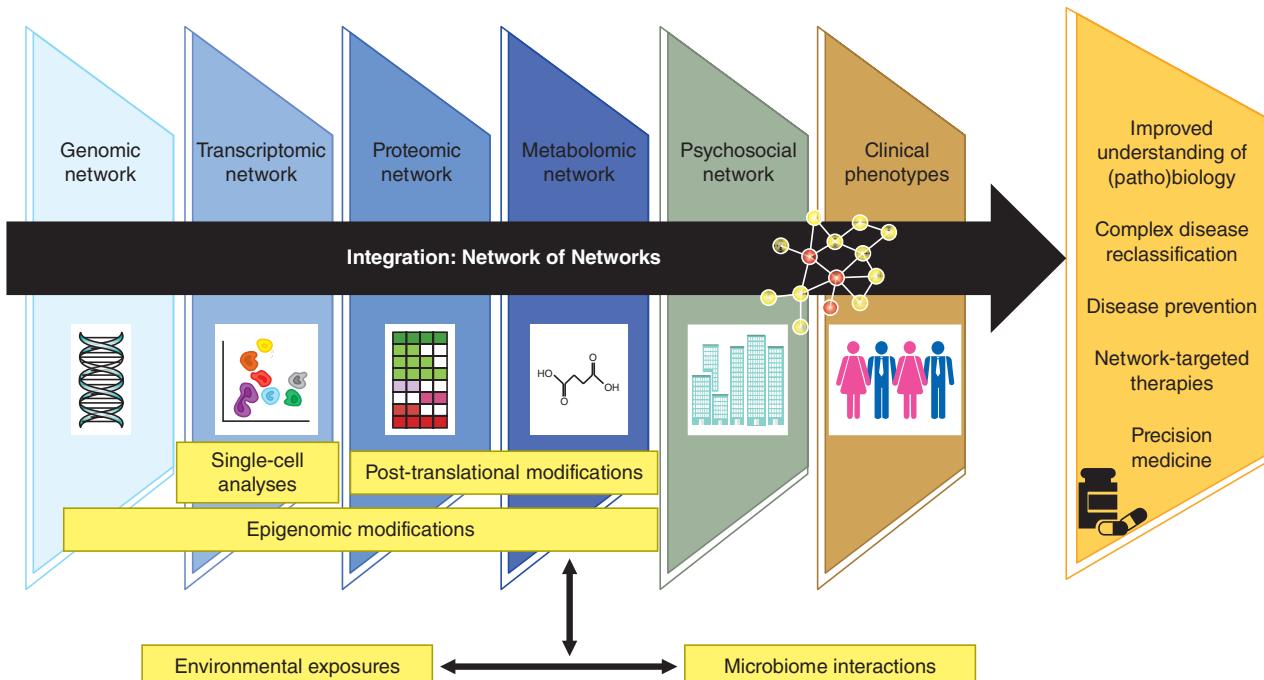


FIGURE 5-3 Universe of precision medicine. The totality of precision medicine incorporates multidimensional biologic networks, the integration of which leads to a network of networks whose components interact with each other and with environmental exposures to yield a distinctive phenotype or pathophenotype. (Reproduced with permission from LY-H Lee, J Loscalzo: Network medicine in pathobiology. *Am J Pathol* 189:1311, 2019.)

gene expression, metabolomics, and epigenetics are often tissue specific. As tissue specimens cannot always or easily be obtained from the organ of interest, attempts at correlating whole-blood mRNA, protein, or metabolite profiles with those of the involved organ are critical for precise prognostics and therapeutic choices. In many cases, systemic consequences to an organ-specific disease (e.g., systemic inflammatory responses in individuals with atherosclerosis) can be ascertained and may provide useful prognostic information or therapeutic strategies. These biomarker signatures are the subject of ongoing discovery and have provided useful guidance toward improved diagnostic precision in many diseases. However, in many diseases, the correlations between these plasma or blood markers and organ-based diseases are weak, indicating a need to analyze each condition and each resulting signature before applying it to clinical decision-making. It is important to note that one of the key determinants of the functional consequences of a genetic variant believed to drive a disease phenotype is not simply its expression in a tissue of interest but, more importantly, the coexpression of protein binding partners in that same tissue comprising specific (dys)functional pathways that govern phenotype (Fig. 5-4). An alternative strategy currently under investigation is the conversion of induced pluripotent stem cells from a patient into a cell type of interest for gene expression or metabolomics study. As rational as this approach seems from first principles, it is important to note that gene expression patterns in these induced, differentiated cell types are not completely consonant with their native counterparts, offering often limited additional information at potentially great additional expense.

While phenotype features of many chronic diseases are assessed over time, genomic features tend to be limited to single time point sampling. Time trajectories are extremely informative in precision genotyping and phenotyping, with gene expression patterns and phenotypes changing over time in different ways among different patients with the same overarching phenotype. Cost, feasible sampling frequency, predictive power, and therapeutic choices will all drive the optimal strategy for the acquisition of timed samples in any given patient; however, with continued cost reduction in genomics technologies, this limitation may be progressively mitigated and clinical application may become a reality.

One important class of diseases that does not have most of these limitations in genomic profiling is cancer. Cancers can be (and are) sampled (biopsied) frequently to monitor temporal changes in the somatically mutating oncogenome and its consequences for the limited number of well-defined oncogenic driver pathways (Chap. 68). A unique limitation of cancer in this regard, however, is that the frequency of somatic mutations over time (and, especially, with treatment) is great and the functional consequences of many of these mutations unknown. Equally important, assessment of single-cell mRNA sequencing patterns demonstrates great variability between apparently similar cells, challenging functional interpretation. Lastly, in solid tumors, stromal cells interact in a variety of ways (e.g., metabolically) with the associated malignant cells, and their gene expression signatures are also modified by the changing somatic mutational landscape of the primary malignancy. Thus, while much more information can be obtained over time in most cancer patients, the interpretation of these rich data sets continues to remain largely semi-empirical.

The possibility of identifying specific therapeutic targets remains a major goal of precision medicine. Doing so requires more than simple DNA sequencing and must include analysis of some level of gene expression, ideally in the involved organ(s). In addition to demonstrating the expression of a variant protein in the organ, one must ideally also demonstrate its functional consequences, which requires ascertaining the expression of binding partner proteins and the functional pathways they comprise. To achieve this goal, a variety of approaches have been tried, one of the most successful of which is the construction of the protein-protein interactome (the interactome), which is a comprehensive network map of the protein-protein interactions in a cell or organ of interest (Chap. 486). This template provides information on the subnetworks that govern a disease phenotype (disease modules), which can be further individualized by incorporating individual variants and differentially expressed proteins that are patient specific. This type of analysis leads to the creation of an individual “reticulome” or

reticulotype, which links the genotype to the phenotype of an individual (Fig. 5-5). Using this approach, one can identify potential drug targets in a rational way or can even repurpose existing drugs by demonstrating the proximity of a known drug target to a disease module of interest (Fig. 5-6). For example, in multicentric Castleman's disease, a disorder of unclear etiology, recognition that the PI3K/Akt/mTOR pathway is highly activated led to trials with an existing, approved drug, sirolimus. Precision medicine offers additional opportunities for optimizing the utilization of a drug by assessing the individualized pharmacogenomics of its disposition and metabolism, as demonstrated for the adverse consequences of variants in *TPMT* on azathioprine metabolism and in *CYP2C19* on clopidogrel metabolism (Chap. 68).

■ EXAMPLES OF PRECISION MEDICINE APPLICATIONS

The field of precision medicine did not appear abruptly in medical history but, rather, evolved gradually as clinicians became more aware of differences among patients with the same disease. With the advent of modern genomics, in the ideal situation, these phenotype differences can now be mapped to genotype differences. Thus, we can consider precision medicine from the perspective of the pregenomic era and the postgenomic era. Pregenomic precision medicine was applied to many diseases as therapeutic classes expanded for those disorders. A prime example of this approach is in the field of heart failure, where diuretics, digoxin, beta blockers, afterload-reducing agents, venodilators, renin-angiotensin-aldosterone inhibitors, and brain natriuretic peptide (nesiritide) are commonly used in some combination for most patients. The choice of agents is governed by the evidence basis for their use, but tailored to the primary pathophysiologic phenotypes manifest in a patient, such as congestion, hypertension, and impaired contractility. These treatments were developed in the latter half of the last century based on empiric observation, reductionist experiments of specific pathways believed to be involved in the pathophysiology, and clinical response in prospective trials. As phenotyping became more refined (e.g., echocardiographic assessments of ventricular function and tissue Doppler characterization of ventricular relaxation), the syndrome was subclassified into heart failure with reduced ejection fraction and heart failure with preserved ejection fraction, the latter of which does not respond well to any of the classes of therapeutic agents currently available. In the postgenomic era, ever more refined and detailed methods are under investigation to characterize pathophenotypes as well as genotypes, which may then be matched to the idealized combination of therapeutic classes of agents.

Pulmonary arterial hypertension is another disease for which definitive therapies straddle the pre- and postgenomic eras of precision medicine. Prior to the 1990s, there were no effective therapies for this highly morbid and lethal condition. With the advent of molecular and biochemical characterization of vascular abnormalities in individuals with established disease, however, therapies with agents that restored normal vascular function improved morbidity and mortality. These included calcium channel blockers, prostacyclin congeners, and endothelin receptor antagonists. As genomic characterization of the disease has progressed over the past two decades, there is increasing recognition of distinct genotypes that yield unique phenotypes (Chap. 283), such as the demonstration of a primarily fibrotic endophenotype governed by the (oxidized) scaffold protein NEDD-9 and its aldosterone-dependent, TGF- β -independent enhancement of collagen III expression. This approach will continue to evolve as therapies become more effective (e.g., for perivascular fibrosis) and therapeutic choices better targeted to individual patients.

Precision genomics has also led to a new classification of the dementias, conditions previously thought to have a single cause with varied clinical expression. These disorders can now be categorized based on the genes and pathways involved and the site where aggregated proteins first form and then spread in the nervous system. For example, the varied clinical presentations of frontotemporal dementia, including progressive aphasia, behavioral disturbances, and dementia with amyotrophic lateral sclerosis, can now be linked to specific genotypes and susceptible cells (Chap. 432). In prion diseases, the clinical phenotype

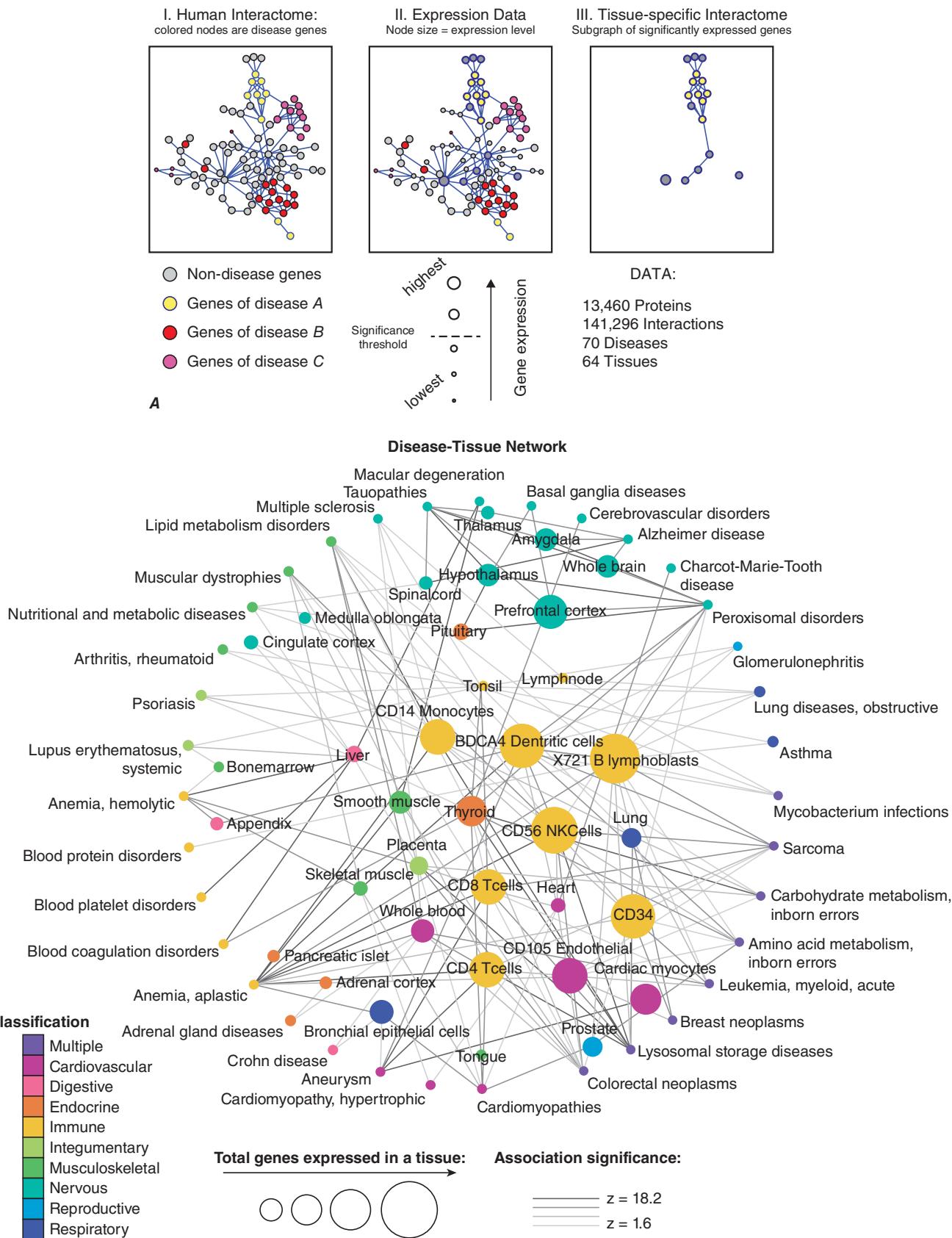


FIGURE 5-4 Gene expression and phenotype. **A.** The human protein-protein interactome is constructed, and a specific disease module is identified (I); gene expression within this module is ascertained (II); and the tissue specificity of gene expression is determined (III). This analysis leads to a reduction of the total number of disease module genes that govern phenotype in a specific organ, which is a reflection of the specific pathway (or pathways) that is (or are) expressed in their functional entirety in that tissue. **B.** A disease-tissue bipartite network is constructed wherein specific tissues are placed within the circle and linked to diseases shown on the circumference. Nodes are colored according to tissue classification, the sizes of nodes are proportional to the total number of genes expressed in them, and the widths (shades) of the lines or edges correspond to the significance of the associations with specific diseases. (From M Kitsak et al: Tissue Specificity of Human Disease Module. *Sci Rep* 6: 35241, 2016, Figure 4.)

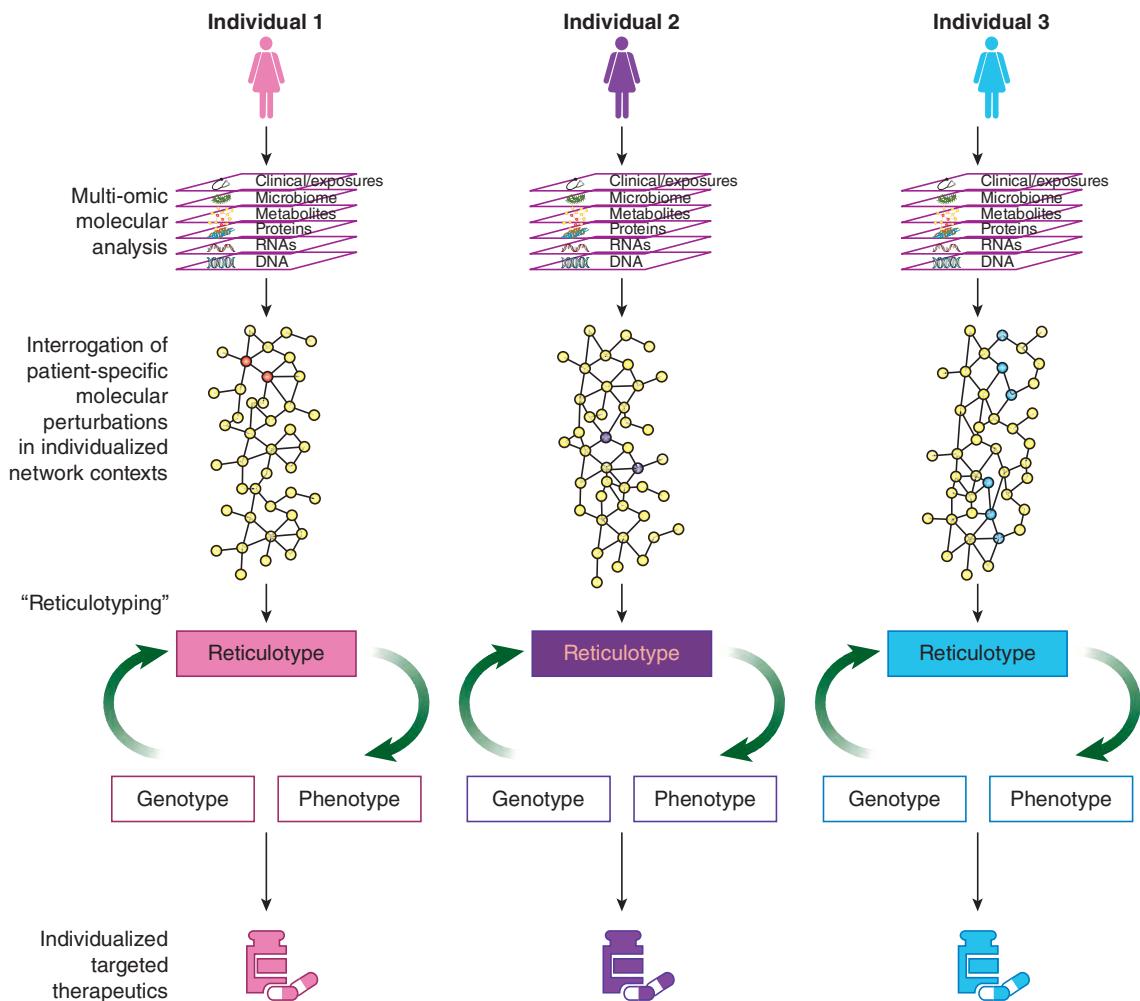


FIGURE 5-5 Reticulotyping. Patient-specific genotype-phenotype relationships by multiomic network structures are depicted for three individuals. Each individual's unique molecular perturbations (genetic variants, differentially expressed genes) are examined within the context of the subject's unique integrative biologic network or reticulome derived from these multiomic analyses. These unique reticulotypes then serve as the basis for patient-specific, precision therapies. (Reproduced with permission from LYH Lee, J Loscalzo: *Network Medicine in Pathobiology*. Am J Pathol 189:1311, 2019.)

is determined by specific germline mutations present in the prion protein (**Chap. 438**). Discovery of autoantibodies against aquaporin-4 (AQP-4) and myelin oligodendrocyte glycoprotein (MOG) has allowed neuromyelitis optica, previously considered a multiple sclerosis-like disorder, to be classified as a separate entity requiring different treatment (**Chap. 445**). Similarly, in myasthenia gravis, the identification of novel autoantibodies now permits stratification and a more finely tuned precision approach to therapy (**Chap. 448**).

Precision medicine approaches to cancers have, of course, become the prime example of the opportunity that this strategy offers. In the pregenomic era, chemotherapy was widely used with variable success despite continued efforts to characterize the molecular features of the specific tumors and their semi-empiric responses to specific chemotherapeutic agents. As cancer genome sequencing evolved, however, it became apparent that there are a limited number of oncogenic pathways (<20) that are represented in the great majority of malignancies, without regard for the organ in which the disease was primarily manifest. These genomic signatures served as a template for precisely targeted therapies that have led to dramatic changes in response to treatment, including, for example, imatinib (and congeners) for Bcr-Abl tyrosine kinase activity in chronic myelogenous leukemia, erlotinib for EGFR-mutant non-small cell lung cancers, and ibrutinib for Bruton tyrosine kinase in chronic lymphocytic leukemia, among many others.

As exciting as these approaches have been, there are at least three primary challenges associated with precision therapeutics that are unique to cancer: (1) the mutational landscape continues to evolve as the disease progresses, and therapy often (if not invariably) leads to selection for resistant clones; (2) the likelihood that any cancer can be definitively cured by any single agent, no matter its exquisite precision, is quite limited, necessitating

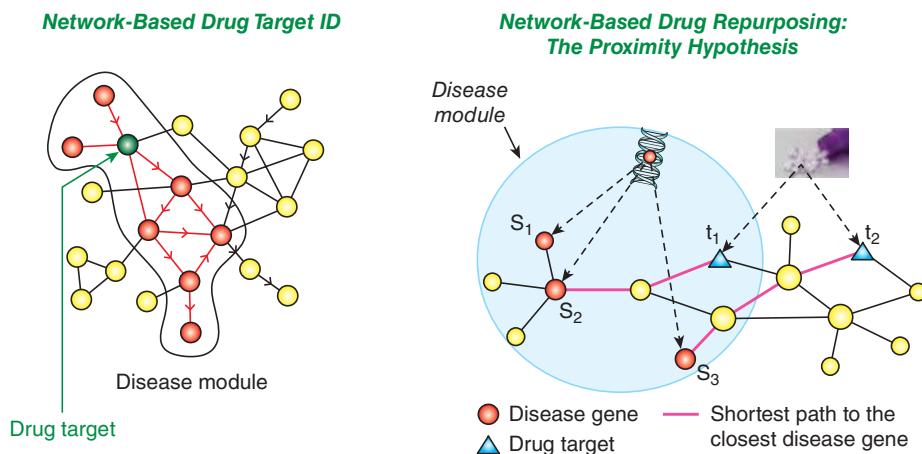


FIGURE 5-6 Network-based precision drug repurposing. (Adapted from F Cheng et al: A genome-wide positioning systems network algorithm for *in silico* drug repurposing. Nat Commun 10:3476, 2019.)

the development of rational polypharmaceutical approaches that take into account alternative pathways that achieve the same oncogenic goals as the primary targeted pathway, complicating drug development; and (3) there is marked genomic heterogeneity in many malignancies arguing that targeting a specific pathway—even with multiple drugs—may not ultimately succeed over the long term owing to the continued and heterogeneous evolution of the genomic landscape within a tumor within a patient. Despite these serious shortcomings, the application of progressively more refined and precisely targeted therapies used alone and in combination, such as with immune modulators, continues to offer great promise for the treatment of these diseases. In some ways, these approaches in cancer mirror earlier strategies in the treatment of infectious diseases in which the identification of the causative organism and its sensitivity to potential antimicrobials allows precision approaches to treatment. Combinatorial antimicrobial treatments represent an effective strategy to address acquired resistance. These diagnostic and therapeutic strategies can be applied without detailed knowledge of personalized responses to the infection or treatment (aside from serious adverse effects) with good outcomes in most cases. Yet, individuals do respond differently to specific infections and their treatments, possibly driven by different endophenotypes (e.g., different inflammatory responses), suggesting that more precise knowledge of these precise mechanistic differences may yield improved prognosis and therapeutic approaches. As with cancer, immune modulation, particularly for immune exhaustion in chronic infections, represents a new frontier, again amenable to the personalized, precise analyses described above.

THE FUTURE OF PRECISION MEDICINE

Precision medicine clearly holds great promise for the future of the practice of medicine. For precision medicine to continue to evolve successfully, however, several requirements will need to be met. First, both deeply refined personal phenotypic data and genomic data are essential as the information with which precision analysis is performed. These data sets are quite large and require sufficient storage for analysis, especially for individuals in whom time trajectories are acquired (as should be the case for every person). Equally important, the analytical methods required to extract useful information from these data sets are evolving and themselves quite complex. While great progress has been made in genomics and biochemical testing, our ability to capture meaningful immunologic endophenotypes and environmental exposures is limited by comparison. Machine learning and artificial (auxiliary) intelligence methods will be essential for extracting optimal information from these data sets, which include not only pathways that can be uniquely targeted therapeutically but also individualized genomic or phenotypic signatures that are highly predictive of outcome, with or without therapy. Gathering sufficient information on the “normal” segments of the population is also required to ensure appropriate comparison data sets for optimal prediction.

Second, phenotyping must continue to expand and become dimensionally richer. The phenotypic features included in this data gathering must incorporate not only data relevant to the clinical presentation but also orthogonal phenotypic data that may yield useful information on disease trajectory or preclinical disease markers. Personal device data, environmental exposure history, social network interactions, and health system data will all be incorporated increasingly in defining phenotype and will require great efforts on the part of the medical informatics community to harmonize data sets, standardize data collection, and optimize/standardize data analysis (Fig. 5-7).

Third, perhaps the greatest challenge to making precision medicine the standard approach to illness will be to determine the minimal data set required to predict outcome and response to therapy. Gathering data is comparatively simple; however, analyzing it to eliminate redundant information in these overdetermined biologic systems, weighting the determinants of an outcome, and using the data as phenomic/genomic signatures that are easier to collect than comprehensive, unbiased data sets are the ideal goals—a major challenge, but not insurmountable. Rapidly evolving machine learning and artificial intelligence strategies will also be essential for maximal success.

To return to the question of how precise precision medicine needs to be in order to be useful, please refer to Fig. 5-8 where the approaches

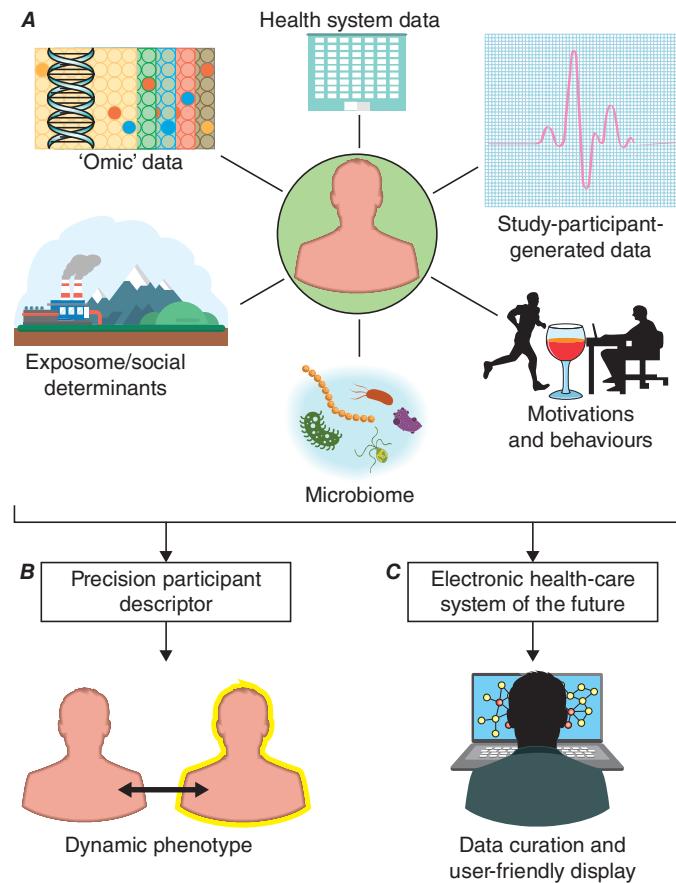


FIGURE 5-7 Big data in precision medicine. **A**, Six dimensions by which individuals may be characterized in the precision medicine era are described. **B**, The precision participant descriptor integrates the data from these six dimensions and varies over time. **C**, The electronic medical record increasingly must evolve to provide curated precision data in a user-friendly way. (Reproduced with permission from EM Antman, J Loscalzo: Precision medicine in cardiology. *Nat Rev Cardiol* 13:591, 2016.)

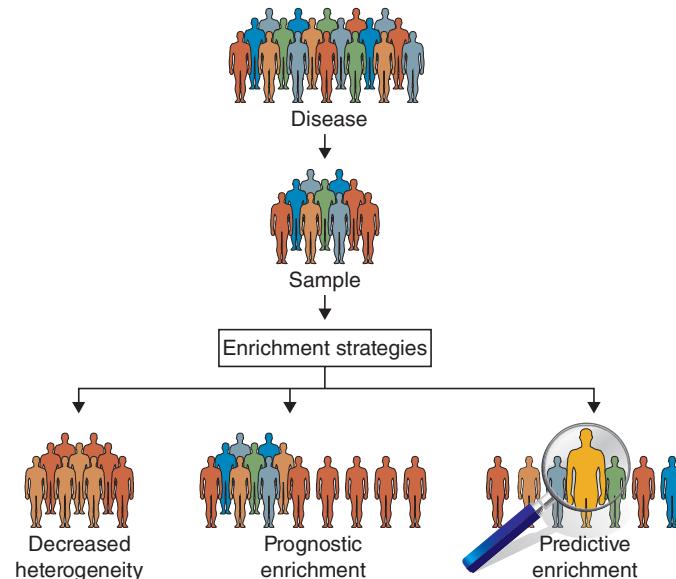


FIGURE 5-8 The basis for precision medicine. The notion of precision medicine evolved, in part, from clinical trial design. From the entire population of patients with the disease of interest, a sample cohort of individuals is enrolled in the trial that ideally is representative of the entire distribution. Enrichment strategies developed to decrease heterogeneity or increase the representation of individuals with a high risk of observed outcomes (prognostic enrichment) facilitate trial conduct but do not necessarily improve precision in defining treatment response. The predictive enrichment strategy utilizes both trial participant characteristics and data from experiments conducted before or during (adaptive design) the trial to improve the prediction of who is likely to have a more pronounced response to the treatment under study. (Reproduced with permission from EM Antman, J Loscalzo: Precision medicine in cardiology. *Nat Rev Cardiol* 13:591, 2016.)

to clinical trial design meant to improve therapeutic signal are illustrated. Decreasing heterogeneity and enriching the study population will enhance the effect size, but these strategies are based on analyses of prior data sets that define those individuals who are more likely than not to respond to a therapy. By contrast, the notion of predictive enrichment follows from the information provided by a detailed, big data–driven analysis of individuals that explores phenotypic and genomic features used to predict response. These features need not be precisely met by each patient; however, they can be collated or clustered to define a reasonably sized cohort predicted to respond in a particular way within certain confidence bounds. In this way, the boundaries to the practice of precision medicine are imprecise strictly speaking, but sufficiently predictive to be practical from the perspectives of clinical care and cost-effectiveness.

FURTHER READING

- ANTMAN EM, LOSCALZO J: Precision medicine in cardiology. *Nat Rev Cardiol* 13:591, 2016.
- CHENG F et al: Comprehensive characterization of protein-protein interactions perturbed by disease mutations. *Nat Genet* 53:342, 2021.
- CHENG F et al: A genome-wide positioning systems network algorithm for in silico drug repurposing. *Nat Commun* 10:3476, 2019.
- GREENE JA, LOSCALZO J: Putting the patient back together—Social medicine, network medicine, and the limits of reductionism. *N Engl J Med* 377:2493, 2017.
- KITSAK M et al: Tissue specificity of human disease module. *Sci Rep* 6:35241, 2016.
- LEE LY, LOSCALZO J: Network medicine in pathobiology. *Am J Pathol* 189:1311, 2019.
- LEOPOLD JA et al: The application of big data to cardiovascular disease: Paths to precision medicine. *J Clin Invest* 130:29, 2020.
- LOSCALZO J et al: Human disease classification in the postgenomic era: A complex systems approach to human pathobiology. *Mol Syst Biol* 3:124, 2007.
- MARON BA et al: Individualized interactomes for network-based precision medicine in hypertrophic cardiomyopathy with implications for other clinical pathophenotypes. *Nat Commun* 12:873, 2021.
- MENCHE J et al: Disease networks: Uncovering disease-disease relationships through the incomplete interactome. *Science* 347:1257601, 2015.
- SAMOKHIN AO et al: NEDD9 targets COL3A1 to promote endothelial fibrosis and pulmonary arterial hypertension. *Sci Transl Med* 10:eaap7294, 2018.

intervention requires a particularly high bar of evidence that testing and intervention are both practical and effective.

 Because population-based screening and prevention strategies must be extremely low risk to have an acceptable benefit-to-harm ratio, the ability to target individuals who are more likely to develop disease could enable the application of a wider set of potential approaches and increase efficiency. Currently, there are many types of data that can predict disease incidence in an asymptomatic individual. Germline genomic data have received the most attention to date, at least in part because mutations in high-penetrance genes have clear implications for preventive care ([Chap. 467](#)). Women with mutations in either *BRCA1* or *BRCA2*, the two major breast cancer susceptibility genes identified to date, have a markedly increased risk (five- to twentyfold) of breast and ovarian cancer. Screening and prevention recommendations include prophylactic oophorectomy and breast magnetic resonance imaging (MRI), both of which are considered to incur too much harm for women at average cancer risk. Some women with *BRCA* mutations opt for prophylactic mastectomy to dramatically reduce their breast cancer risk. Although the proportion of common disease explained by high-penetrance genes appears to be relatively small (5–10% of most diseases), mutations in rare, moderate-penetrance genes, and variants in low-penetrance genes, also contribute to the prediction of disease risk. Most recently, polygenic risk scores combining information about variants across hundreds of genes are being evaluated for identifying individuals at high risk of coronary heart disease and other conditions. The advent of affordable whole exome/whole genome sequencing is likely to speed the dissemination of these tests into clinical practice and may transform the delivery of preventive care.

Other forms of “omic” data also have the potential to provide important predictive information. Proteomics and metabolomics can provide insight into gene function, but it has proven challenging to develop reliable, predictive measures using these platforms. More recently, it has become possible to measure the presence of mutations in DNA circulating in the bloodstream and in stool, with early promising evidence that these assays can be used to detect cancer before existing screening tests.

In addition to “omic” data, imaging data are increasingly being integrated into risk-stratified prevention approaches as evidence grows about the predictive ability of these data. For example, coronary computed tomography (CT) scans are used in many preventive cardiology programs to inform decisions about beginning statin therapy when there is conflicting or uncertain information from other risk assessment approaches. Of course, these data may also be helpful in predicting the risk of harms from screening or prevention, such as the risk of a false-positive mammogram.

In addition to advances in risk prediction, there are several other reasons that screening and prevention are likely to gain importance in medical care in the near term. New imaging modalities are being developed that promise to detect changes at the cellular and subcellular levels, greatly increasing the probability that early detection improves outcomes. The rapidly growing understanding of the biologic pathways underlying initiation and progression of many common diseases has the potential to transform the development of preventive interventions, including chemoprevention. Furthermore, screening and prevention offer the promise of both improving health and sparing the costs of disease treatment, an issue that will continue to gain importance as long as health care costs in the United States remain a concern to patients, government agencies, and insurers.

This chapter will review the basic principles of screening and prevention in the primary care setting. Recommendations for specific disorders such as cardiovascular disease, diabetes, and cancer are provided in the chapters dedicated to those topics.

BASIC PRINCIPLES OF SCREENING

The basic principles of screening populations for disease were published by the World Health Organization in 1968 ([Table 6-1](#)).

In general, screening is most effective when applied to relatively common disorders that carry a large disease burden ([Table 6-2](#)). The five leading causes of mortality in the United States are heart diseases, malignant neoplasms, chronic obstructive pulmonary disease,

6

Screening and Prevention of Disease

Katrina A. Armstrong, Gary J. Martin

A primary goal of health care is to prevent disease or detect it early enough that intervention will be more effective. Tremendous progress has been made toward this goal over the past 50 years. Screening tests are available for many common diseases and encompass biochemical (e.g., cholesterol, glucose), physiologic (e.g., blood pressure, growth curves), radiologic (e.g., mammogram, bone densitometry), and cytologic (e.g., Pap smear) approaches. Effective preventive interventions have resulted in dramatic declines in mortality from many diseases, particularly infections. Preventive interventions include counseling about risk behaviors, vaccinations, medications, and, in some relatively uncommon settings, surgery. Preventive services (including screening tests, preventive interventions, and counseling) are different than other medical interventions because they are proactively administered to healthy individuals instead of in response to a symptom, sign, or diagnosis. Thus, the decision to recommend a screening test or preventive

TABLE 6-1 Principles of Screening

The condition should be an important health problem.
There should be a treatment for the condition.
Facilities for diagnosis and treatment should be available.
There should be a latent stage of the disease.
There should be a test or examination for the condition.
The test should be acceptable to the population.
The natural history of the disease should be adequately understood.
There should be an agreed policy on whom to treat.
The cost of finding a case should be balanced in relation to overall medical expenditure.

accidents, and cerebrovascular diseases. Thus, many screening strategies are targeted at these conditions. From a global health perspective, these conditions are priorities, but malaria, malnutrition, AIDS, tuberculosis, and violence also carry a heavy disease burden (*Chap. 472*).

Having an effective treatment for early disease has proven challenging for some common diseases. For example, although Alzheimer's disease is the sixth leading cause of death in the United States, there are no curative treatments and no evidence that early treatment improves outcomes. Lack of facilities for diagnosis and treatment is a particular challenge for developing countries and may change screening strategies, including the development of "see and treat" approaches such as those currently used for cervical cancer screening in some countries. A long latent or preclinical phase where early treatment increases the chance of cure is a hallmark of many cancers; for example, polypectomy prevents progression to colon cancer. Similarly, early identification of hypertension or hyperlipidemia allows therapeutic interventions that reduce the long-term risk of cardiovascular or cerebrovascular events. In contrast, lung cancer screening has historically proven more challenging because most tumors are not curable by the time they can be detected on a chest x-ray. However, the length of the preclinical phase also depends on the level of resolution of the screening test, and this situation changed with the development of chest CT. Low-dose chest CT scanning can detect tumors earlier and has been demonstrated to reduce lung cancer mortality by 20% in individuals who had at least a 30-pack-year history of smoking. The short interval between the ability to detect disease on a screening test and the development of incurable disease also contributes to the limited effectiveness of mammography screening in reducing deaths from some forms of breast cancer. At the other end of the spectrum, the early detection of prostate cancer may not lead to a difference in the mortality rate because the disease is often indolent and competing morbidities, such as coronary artery disease, may ultimately cause mortality (*Chap. 70*). This uncertainty about the natural history is also reflected in the controversy about treatment of prostate cancer, further contributing to the challenge of screening in this disease. Finally, screening programs can incur significant economic costs that must be considered in the context of the available resources and alternative strategies for improving health outcomes.

METHODS OF MEASURING HEALTH BENEFITS

Because screening and preventive interventions are recommended to asymptomatic individuals, they are held to a high standard for demonstrating a favorable risk-benefit ratio before implementation. In general, the principles of evidence-based medicine apply to demonstrating the efficacy of screening tests and preventive interventions, where randomized controlled trials (RCTs) with mortality outcomes are the gold standard. However, because RCTs are often not feasible,

observational studies, such as case-control designs, have been used to assess the effectiveness of some interventions such as colonoscopy for colorectal cancer screening. For some strategies, such as Pap smear screening for cervical cancer, the only data available are ecologic data demonstrating dramatic declines in mortality.

Irrespective of the study design used to assess the effectiveness of screening, it is critical that disease incidence or mortality is the primary endpoint rather than length of disease survival. This is important because lead time bias and length time bias can create the appearance of an improvement in disease survival from a screening test when there is no actual effect. Lead time bias occurs because screening identifies a case before it would have presented clinically, thereby creating the perception that a patient lived longer after diagnosis simply by moving the date of diagnosis earlier rather than the date of death later. Length time bias occurs because screening is more likely to identify slowly progressive disease than rapidly progressive disease. Thus, within a fixed period of time, a screened population will have a greater proportion of these slowly progressive cases and will appear to have better disease survival than an unscreened population.

A variety of endpoints are used to assess the potential gain from screening and preventive interventions.

1. *The absolute and relative impact of screening on disease incidence or mortality.* The absolute difference in disease incidence or mortality between a screened and nonscreened group allows the comparison of size of the benefit across preventive services. A meta-analysis of Swedish mammography trials (ages 40–70) found that ~1.2 fewer women per 1000 would die from breast cancer if they were screened over a 12-year period. By comparison, at least ~3 lives per 1000 would be saved from colon cancer in a population (aged 50–75) screened with annual fecal occult blood testing (FOBT) over a 13-year period, and an estimated 20–24 lives per 1000 would be saved over the entire 25-year period. Based on this analysis, colon cancer screening may actually save more women's lives than does mammography. However, the relative impact of FOBT (30% reduction in colon cancer death) is similar to the relative impact of mammography (14–32% reduction in breast cancer death), emphasizing the importance of both relative and absolute comparisons.
2. *The number of subjects screened to prevent disease or death in one individual.* The inverse of the absolute difference in mortality is the number of subjects who would need to be screened or receive a preventive intervention to prevent one death. For example, 731 women aged 65–69 would need to be screened by dual-energy x-ray absorptiometry (DEXA) (and treated appropriately) to prevent one hip fracture from osteoporosis.
3. *Increase in average life expectancy for a population.* Predicted increases in life expectancy for various screening and preventive interventions are listed in *Table 6-3*. It should be noted, however, that the increase in life expectancy is an average that applies to a population, not to an individual. In reality, the vast majority of the population does not derive any benefit from a screening test. A small subset of patients, however, will benefit greatly. For example, Pap smears do not benefit the 98% of women who never develop cancer of the cervix. However, for the 2% who would have developed cervical cancer, Pap smears may add as much as 25 years to their lives. Some studies suggest that a 1-month gain of life expectancy is a reasonable goal for a population-based screening or prevention strategy.

TABLE 6-3 Estimated Average Increase in Life Expectancy for a Population

SCREENING OR PREVENTIVE INTERVENTION	AVERAGE INCREASE
Mammography:	
Women, 40–50 years	0–5 days
Women, 50–70 years	1 month
Pap smears, age 18–65	2–3 months
Getting a 35-year-old smoker to quit	3–5 years
Beginning regular exercise for a 40-year-old man (30 min, 3 times a week)	9 months–2 years

^aAssuming an unscreened population.

TABLE 6-2 Lifetime Cumulative Risk

Breast cancer for women	10%
Colon cancer	6%
Cervical cancer for women ^a	2%
Domestic violence for women	Up to 15%
Hip fracture for white women	16%

ASSESSING THE HARMS OF SCREENING AND PREVENTION

Just as with most aspects of medical care, screening and preventive interventions also incur the possibility of adverse outcomes. These adverse outcomes include side effects from preventive medications and vaccinations, false-positive screening tests, overdiagnosis of disease from screening tests, anxiety, radiation exposure from some screening tests, and discomfort from some interventions and screening tests. The risk of side effects from preventive medications is analogous to the use of medications in therapeutic settings and is considered in the U.S. Food and Drug Administration (FDA) approval process. Side effects from currently recommended vaccinations are primarily limited to discomfort and minor immune reactions. However, the concern about associations between vaccinations and serious adverse outcomes continues to limit the acceptance of many vaccinations despite the lack of data supporting the causal nature of these associations.

The possibility of a false-positive test occurs with nearly all screening tests, although the definition of what constitutes a false-positive result often varies across settings. For some tests such as screening mammography and screening chest CT, a false-positive result occurs when an abnormality is identified that is not malignant, requiring either a biopsy diagnosis or short-term follow-up. For other tests such as Pap smears, a false-positive result occurs because the test identifies a wide range of potentially premalignant states, only a small percentage of which would ever progress to an invasive cancer. This risk is closely tied to the risk of overdiagnosis in which the screening test identifies disease that would not have presented clinically in the patient's lifetime. Assessing the degree of overdiagnosis from a screening test is very difficult given the need for long-term follow-up of an unscreened population to determine the true incidence of disease over time. Recent estimates suggest that as much as 15–40% of breast cancers identified

by mammography screening and 15–37% of prostate cancers identified by prostate-specific antigen testing may never have presented clinically. Screening tests also have the potential to create unwarranted anxiety, particularly in conjunction with false-positive findings. Although multiple studies have documented increased anxiety through the screening process, there are few data suggesting this anxiety has long-term adverse consequences, including subsequent screening behavior. Screening tests that involve radiation (e.g., mammography, chest CT) add to the cumulative radiation exposure for the screened individual. The absolute amount of radiation is very small from any of these tests, but the overall impact of repeated exposure from multiple sources is still being determined. Some preventive interventions (e.g., vaccinations) and screening tests (e.g., mammography) may lead to discomfort at the time of administration, but again, there is little evidence of long-term adverse consequences.

WEIGHING THE BENEFITS AND HARMS

The decision to implement a population-based screening and prevention strategy requires weighing the benefits and harms, including the economic impact of the strategy. The costs include not only the expense of the intervention but also time away from work, downstream costs from false-positive results, “incidentalomas” or adverse events, and other potential harms. Cost-effectiveness is typically assessed by calculating the cost per year of life saved, with adjustment for the quality of life impact of different interventions and disease states (i.e., quality-adjusted life-year). Typically, strategies that cost \$50,000–100,000 per quality-adjusted year of life saved are considered “cost-effective” ([Chap. 4](#)).

The U.S. Preventive Services Task Force (USPSTF) is an independent panel of experts in preventive care that provides evidence-based recommendations for screening and preventive strategies based on an assessment of the benefit-to-harm ratio ([Tables 6-4 and 6-5](#)). Because

TABLE 6-4 Screening Tests Recommended by the U.S. Preventive Services Task Force for Average-Risk Adults

DISEASE	TEST	POPULATION	FREQUENCY	CHAPTER
Abdominal aortic aneurysm	Ultrasound	Men 65–75 who have ever smoked	Once	
Alcohol misuse	Alcohol Use Disorders Identification Test	All adults	Unknown	453
Breast cancer	Mammography with or without clinical breast examination	Women 50–75	Every 2 years	
Cervical cancer	Pap smear	Women 21–65	Every 3 years	70
	Pap smear and/or HPV testing	Women 30–65	Every 5 years if HPV negative	
Chlamydia/gonorrhea	Nucleic acid amplification test on urine or cervical swab	Sexually active women <25	Unknown	189
Colorectal cancer	Fecal occult blood testing	45–75	Every year	70, 81
	Fecal immunochemical-DNA	45–75	Every 1–3 years	
	Sigmoidoscopy	45–75	Every 5 years	
	Colonoscopy (or occult blood testing combined with sigmoidoscopy)	45–75	Every 10 years	
Depression	Screening questions	All adults	Periodically	
Diabetes	Fasting blood glucose or HgbA1c	Adults overweight, obese, or with hypertension	Every 3 years	403
Hepatitis C	Anti-HCV antibody followed by confirmatory PCR	18–79	Once	
HIV	Reactive immunoassay or rapid HIV followed by confirmatory test	15–65	At least once	
Hyperlipidemia	Cholesterol	40–75	Unknown	407
Hypertension	Blood pressure	All adults	Periodically	277
Intimate partner violence	Screening questions	Women of childbearing age	Unknown	
Lung cancer	Low-dose computed tomography	Adults 50 to 80 years who have a 20 pack-year smoking history and currently smoke or have quit within the past 15 years	Yearly	
Obesity	Body mass index	All adults	Unknown	
Osteoporosis	DEXA	Women >65 or >60 with risk factors	Unknown	411

Abbreviations: DEXA, dual-energy x-ray absorptiometry; HCV, hepatitis C virus; HPV, human papillomavirus; PCR, polymerase chain reaction.

Source: Adapted from the U.S. Preventive Services Task Force 2017. www.uspreventiveservicestaskforce.org/Page/Name/uspstf-a-and-b-recommendations/.

TABLE 6-5 Preventive Interventions Recommended for Average-Risk Adults

INTERVENTION	DISEASE	POPULATION	FREQUENCY	CHAPTER
Adult immunization				
Tetanus-diphtheria		>18	Every 10 years	
Varicella		Susceptibles only, >18	Two doses	
Measles-mumps-rubella		Women, childbearing age	One dose	
Pneumococcal		>64	13 followed by 23 valent	
Influenza		>18	Yearly	
Human papillomavirus		Up to age 27	If not done prior	
Zoster		>60	Once	
Chemoprevention				
Aspirin	Cardiovascular disease	Aged 50–59 years with a ≥10% 10-year cardiovascular disease risk (bleeding risk may = benefit for some groups)		
Folic acid	Neural tube defects in baby	Women planning or capable of pregnancy		
Tamoxifen/raloxifene	Breast cancer	Women at high risk for breast cancer		
Vitamin D	Fracture/falls	>64 at increased risk for falls		

there are multiple advisory organizations providing recommendations for preventive services, the agreement among the organizations varies across the different services. For example, all advisory groups support screening for hyperlipidemia and colorectal cancer, whereas consensus is lower for breast cancer screening among women in their forties and for prostate cancer screening. Because the guidelines are only updated periodically, differences across advisory organizations may also reflect the data that were available when the guideline was issued.

For many screening tests and preventive interventions, the balance of benefits and harms may be uncertain for the average-risk population but more favorable for individuals at higher risk for disease. Although age is the most commonly used risk factor for determining screening and prevention recommendations, the USPSTF also recommends some screening tests in populations based upon the presence of other risk factors for the disease. In addition, being at increased risk for the disease often supports initiating screening at an earlier age than that recommended for the average-risk population. For example, when there is a significant family history of colon cancer, it is prudent to initiate screening 10 years before the age at which the youngest family member was diagnosed with cancer.

Although informed consent is important for all aspects of medical care, shared decision-making may be a particularly important approach to decisions about preventive services when the benefit-to-harm ratio is uncertain for a specific population. For example, many expert groups, including the American Cancer Society, recommend an individualized discussion about prostate cancer screening, because the decision-making process is complex and relies heavily on personal issues. Some men may decline screening, whereas others may be more willing to accept the risks of an early detection strategy. Recent analysis suggests that many men may be better off not screening for prostate cancer because watchful waiting was the preferred strategy when quality-adjusted life-years were considered. Another example of shared decision-making involves the choice of techniques for colon cancer screening (**Chap. 70**). In controlled studies, the use of annual FOBT reduces colon cancer deaths by 15–30%. Flexible sigmoidoscopy reduces colon cancer deaths by ~40–60%. Colonoscopy appears to offer a greater benefit than flexible sigmoidoscopy with a reduction in risk of ~70%, but its use incurs additional costs and risks. These screening procedures have not been compared directly in the same population, but models suggest that appropriate frequencies of each technique may be associated with similar numbers of lives saved and cost to society per life saved (\$10,000–25,000). Thus, although one patient may prefer the ease of preparation, less time disruption, and the lower risk of flexible sigmoidoscopy, others may prefer the sedation, thoroughness, and time interval of colonoscopy.

COUNSELING ON HEALTHY BEHAVIORS

In considering the impact of preventive services, it is important to recognize that tobacco and alcohol use, diet, and exercise constitute the

vast majority of factors that influence preventable deaths in developed countries. Perhaps the single greatest preventive health care measure is to help patients quit smoking (**Chap. 454**). However, efforts in these areas frequently require behavior changes (e.g., weight loss, exercise) or the management of addictive conditions (e.g., tobacco and alcohol use) that are often recalcitrant to intervention. Although these are challenging problems, evidence strongly supports the role of counseling by health care providers (**Table 6-6**) in effecting health behavior change. Educational campaigns, public policy changes, and community-based interventions have also proven to be important parts of a strategy for addressing these factors in some settings. Although the USPSTF found that the evidence was conclusive to recommend a relatively small set of counseling activities, counseling in areas such as physical activity and injury prevention (including seat belts and bicycle and motorcycle helmets) has become a routine part of primary care practice.

IMPLEMENTING DISEASE PREVENTION AND SCREENING

The implementation of disease prevention and screening strategies in practice is challenging. A number of techniques can assist physicians with the delivery of these services. An appropriately configured electronic health record can provide reminder systems that make it easier for physicians to track and meet guidelines. Some systems give patients secure access to their medical records, providing an additional means to enhance adherence to routine screening. Systems that provide nurses and other staff with standing orders are effective for immunizations. The USPSTF has developed flow sheets and electronic tools to assist clinicians (<https://www.uspreventiveservicestaskforce.org/uspstf/information-health-professionals>). Many of these tools use age categories to help guide implementation. Age-specific recommendations for screening and counseling are summarized in **Table 6-7**.

Many patients see a physician for ongoing care of chronic illnesses, and this visit provides an opportunity to include a “measure of prevention” for other health problems. For example, a patient seen for management of hypertension or diabetes can have breast cancer

TABLE 6-6 Preventive Counseling Recommended by the U.S. Preventive Services Task Force (USPSTF)

TOPIC	CHAPTER REFERENCE
Alcohol and drug use	453, 456, 457
Genetic counseling for <i>BRCA1/2</i> testing among women at increased risk for deleterious mutations	79, 467
Nutrition and diet	332, 333
Sexually transmitted infections	136, 202
Sun exposure	61
Tobacco use	454

TABLE 6-7 Age-Specific Causes of Mortality and Corresponding Preventive Options

AGE GROUP	LEADING CAUSES OF AGE-SPECIFIC MORTALITY	SCREENING PREVENTION INTERVENTIONS TO CONSIDER FOR EACH SPECIFIC POPULATION
15–24	1. Accident 2. Homicide 3. Suicide 4. Malignancy 5. Heart disease	<ul style="list-style-type: none"> Counseling on routine seat belt use, bicycle/motorcycle/ATV helmets (1) Counseling on diet and exercise (5) Discuss dangers of alcohol use while driving, swimming, boating (1) Assess and update vaccination status (tetanus, diphtheria, hepatitis B, MMR, rubella, varicella, meningitis, HPV) Ask about gun use and/or gun possession (2,3) Assess for substance abuse history including alcohol (2,3) Screen for domestic violence (2,3) Screen for depression and/or suicidal/homicidal ideation (2,3) Pap smear for cervical cancer screening after age 21 (4) Discuss skin, breast awareness, and testicular self-examinations (4) Recommend UV light avoidance and regular sunscreen use (4) Measurement of blood pressure, height, weight, and body mass index (5) Discuss health risks of tobacco use, consider emphasis on cosmetic and economic issues to improve quit rates for younger smokers (4,5) Chlamydia and gonorrhea screening and contraceptive counseling for sexually active females, discuss STD prevention Hepatitis B, and syphilis testing if there is high-risk sexual behavior(s) or any prior history of sexually transmitted disease Hepatitis C screening starting at age 18 to 79 HIV testing Continue annual influenza vaccination
25–44	1. Accident 2. Malignancy 3. Heart disease 4. Suicide 5. Homicide 6. HIV	<p><i>As above plus consider the following:</i></p> <ul style="list-style-type: none"> Readdress smoking status, encourage cessation at every visit (2,3) Obtain detailed family history of malignancies and begin early screening/prevention program if patient is at significant increased risk (2) Assess all cardiac risk factors (including screening for diabetes and hyperlipidemia) and consider primary prevention with aspirin for patients at >3% 5-year risk of a vascular event (3) and statin therapy for higher risk patients Assess for chronic alcohol abuse, risk factors for viral hepatitis, or other risks for development of chronic liver disease Consider individualized breast cancer screening with mammography at age 40 (2)
45–64	1. Malignancy 2. Heart disease 3. Accident 4. Diabetes mellitus 5. Cerebrovascular disease 6. Chronic lower respiratory disease 7. Chronic liver disease and cirrhosis 8. Suicide	<ul style="list-style-type: none"> Consider prostate cancer screen with annual PSA and digital rectal examination at age 50 (or possibly earlier in African Americans or patients with family history) (1) Begin colorectal cancer screening at age 45 or 50 with fecal occult blood testing, flexible sigmoidoscopy, or colonoscopy (1) Reassess and update vaccination status at age 50 and vaccinate all smokers against <i>Streptococcus pneumoniae</i> at age 50 (6) Consider screening for coronary disease in higher-risk patients (2,5) Zoster vaccination at age 60 Begin mammography screening by age 50 Lung cancer screening at age 50 to 80 years if a 20 pack-year smoking history and currently smoke or have quit within the past 15 years, yearly.
≥65	1. Heart disease 2. Malignancy 3. Cerebrovascular disease 4. Chronic lower respiratory disease 5. Alzheimer's disease 6. Influenza and pneumonia 7. Diabetes mellitus 8. Kidney disease 9. Accidents 10. Septicemia	<p><i>As above plus consider the following:</i></p> <ul style="list-style-type: none"> Readdress smoking status, encourage cessation at every visit (1,2,3,4) One-time ultrasound for AAA in men 65–75 who have ever smoked Consider pulmonary function testing for all long-term smokers to assess for development of chronic obstructive pulmonary disease (4,6) Screen all postmenopausal women (and all men with risk factors) for osteoporosis Continue annual influenza vaccination and vaccinate against <i>S. pneumoniae</i> at age 65 (4,6) Screen for visual and hearing problems, home safety issues, and elder abuse (9) Consider fall prevention exercise intervention if at higher risk (9)

Note: The numbers in parentheses refer to areas of risk in the mortality column affected by the specified intervention.

Abbreviations: AAA, abdominal aortic aneurysm; ATV, all-terrain vehicle; HPV, human papillomavirus; MMR, measles-mumps-rubella; PSA, prostate-specific antigen; STD, sexually transmitted disease; UV, ultraviolet.

screening incorporated into one visit and a discussion about colon cancer screening at the next visit. Other patients may respond more favorably to a clearly defined visit that addresses all relevant screening and prevention interventions. Because of age or comorbidities, it may be appropriate with some patients to abandon certain screening and

prevention activities, although there are fewer data about when to “sunset” these services. For many screening tests, the benefit of screening does not accrue until 5–10 years of follow-up, and there are generally few data to support continuing screening for most diseases past age 75. In addition, for patients with advanced diseases and limited life

expectancy, there is considerable benefit from shifting the focus from screening procedures to the conditions and interventions more likely to affect quality and length of life.

FURTHER READING

- BRETTHAUER M et al: America, we are confused: The updated U.S. Preventive Services Task Force recommendation on colorectal cancer screening. *Ann Intern Med* 166:139, 2017.
- HAYES JH et al: Observation versus initial treatment for men with localized, low-risk prostate cancer: A cost-effectiveness analysis. *Ann Intern Med* 158:853, 2013.
- HUGOSSON J et al: Mortality results from the Goteborg randomized population-based prostate-cancer screening trial. *Lancet Oncol* 11:725, 2010.
- OEFFINGER KC et al: Breast cancer screening for women at average risk 2015. Guideline update from the American Cancer Society. *JAMA* 314:1599, 2015.
- US PREVENTIVE SERVICES TASK FORCE: Screening for colorectal cancer. US Preventive Services Task Force recommendation statement. *JAMA* 315:2564, 2016.

insurance with individual medical savings accounts (patterned after Singapore), promote private insurance, and expand private hospitals.

In lower-income developing countries, health services are typically provided by tax-funded public institutions, often with considerable inadequacies and sometimes with substantial copayments. It is important to note that governmentally organized systems in nearly all developing countries, as well as in former Soviet bloc countries and, to a lesser degree, in tax-funded developed countries, are supplemented to varying extents by a mix of private and/or employer-paid insurers and providers.

This chapter focuses on the individual patient care system: on the financing and delivery of individual clinical and preventive services. The individual patient care system is composed of the financing and delivery of necessary services to prevent death or serious harm ("rule of rescue"); to maintain quality of life; and to manage, reduce, and/or prevent the burden of illness on individual patients. While the technical dimensions of most clinical services are similar across countries, their organizational, social, and economic characteristics range widely. Health systems in both developed and developing countries exhibit substantial differences, for example, in access to care; in the design and reliance on quality assurance and provider payment mechanisms; in the relationship of primary care to hospital services; in the coordination of health care with home care and nursing home services; in the design and use of provider management strategies; in the way physicians work and are paid; in the decision-making roles of politically elected officials and of national, regional, and municipal governments; and in participation of both citizens and patients. These wide-ranging institutional and organizational characteristics reflect differing country contexts (geographical, social, economic, and political), differences in national culture (consisting of prioritized norms and values), and substantial variation in how health sector institutions are structured.

7

Global Diversity of Health System Financing and Delivery

Richard B. Saltman

Health care systems are highly complex organizations, with many interdependent components. In developed countries, health systems have traditionally been classified by their type of financing—i.e., either predominantly tax-funded (such as the National Health Service in England and publicly operated regional care systems in the four European Nordic countries) or predominantly statutory social health insurance (SHI)-funded (such as in Germany, the Netherlands, and France). Over the past several decades, however, there has been structural convergence in the technical characteristics of both funding arrangements and in the associated delivery systems, making analytic observations about differences across national systems more difficult.

A second confounding factor has been that former Soviet Bloc countries in Central and Eastern Europe, including the Russian Federation, have, since 1991, replaced their former Soviet-style Semashko models (a top-down, national government-controlled funding and delivery structure with a parallel Communist Party administrative apparatus) with various hybrid arrangements built on national government-run SHI financing. Distinctions across developed country health systems, especially in Europe, have been further compressed by inadequate resources in many publicly funded systems in an era of rapid clinical and technological change, triggering increased private sector funding and provision.

In middle-income developing countries, institutional structures in the health sector typically reflect the country's preindependence administrative framework. Mexico, for example, has a Spanish-derived configuration with health insurance as part of social insurance for formally employed workers (via Instituto Mexicano del Seguro Social), supplemented by tax-funded health services (Seguro Popular) provided for those with informal employment and all other citizens, as well as a separate program (Instituto de Seguridad y Servicios Sociales de los Trabajadores del Estado) for public employees. Countries such as India and Egypt, reflecting British influence, have predominantly tax-funded and publicly operated health systems. China is an exception, with an internally generated system that is publicly funded and operated, although recent Communist Party policy has been to introduce SHI-based

FINANCING INDIVIDUAL PATIENT CARE SERVICES IN DEVELOPED COUNTRIES

Funding for individual care services in developed countries comes from the particular national mix among four possible sources of revenue: national, regional, and/or municipal taxes; mandatory SHI; private health insurance (including employer-paid insurances); and out-of-pocket payments. Most countries have one preponderant payer, which then defines its funding arrangements and serves to frame the structure of its delivery system as well.

Total Health Expenditures The Organization for Economic Co-operation and Development (OECD) data from 2017 (adjusted for purchasing power parities) show that total health care expenditures in developed countries vary across a considerable range, tied to health system structure as well as national history and culture (**Table 7-1**).

Per capita health expenditure figures provide a different, specific measurement of available funds in a country's health sector (**Table 7-2**).

Tax-Funded Systems In the United Kingdom, 79% of all health care funding was furnished through general tax revenues allocated by the national government in its annual budget process (all figures from OECD for 2017). In Sweden, all public taxes combined raised 83.7% of total health care spending. Sweden's 21 regional-level elected governments provide approximately 70% of that 83.7%, with the remaining 13.7% of total health spending raised by national and municipal taxes. In Canada, 71% of total health spending was raised by tax revenues, with 66% of that 71% coming from provincial or territorial taxes, while 5% came from national and local government taxes.

In most tax-funded countries, a segment of the population also has individual-, company-, or union-purchased private complementary and/or supplemental insurance coverage. In Sweden, 2019 estimates are that about 600,000 individuals have private complementary policies in a total population of 9 million. In Denmark, 50% of the population purchase supplemental insurance, while 30% have complementary insurance (often purchased by employers) that pays for private sector services enabling them to bypass public sector queues. In Finland, many middle-class families purchase separate private health insurance for their children to enable them to bypass long waiting times

TABLE 7-1 Developed Country Total Health Expenditure (% GDP)

TAX FUNDED IN WESTERN EUROPE	SHI FUNDED IN WESTERN EUROPE	CENTRAL EUROPEAN	DEVELOPED ASIAN	DEVELOPED NORTH AMERICAN					
Ireland	7.2%	Belgium	10.3%	Latvia	6.0%	Singapore	4.5%	Canada	10.7%
Spain	8.9%	Netherlands	10.1%	Poland	6.5%	South Korea	7.6%	United States	17.1%
UK	9.6%	Germany	11.2%	Czech Republic	7.2%	Japan	10.9%		
Finland	9.6%	Switzerland	12.3%	Slovenia	8.2%				
Denmark	10.1%								
Sweden	11.0%								

Abbreviations: GDP, gross domestic product; SHI, social health insurance; UK, United Kingdom.

Source: The Organization for Economic Co-operation and Development (OECD) data.

for primary and secondary pediatric health care services. More than 400,000 Finnish children (in a total population of 5 million) have privately purchased policies. In England in 2015, individual-, employer-, and union-purchased private complementary insurance covered an estimated 10.5% of the population, or about 6 million people. In Canada, individuals are not allowed by law to purchase private complementary insurance (except for Supreme Court-ordered insurance for three backlogged surgical procedures in Quebec Province—2005 Chaoulli decision); however, approximately 65% of the population have employer-, union-, or private group-purchased supplemental insurance for non-publicly covered services such as outpatient pharmaceutical prescriptions and home care.

Social Insurance-Funded Systems In Western Europe, SHI funds have traditionally been organized on a private not-for-profit basis, but with statutory responsibilities under national law. When former Soviet Bloc countries in Eastern Europe regained their independence in 1991, they returned to pre-World War II SHI models, but because there was no remaining organizational infrastructure, these post-1991 arrangements typically became a single SHI fund, run as an arm of the national government. In the United States, the Medicare social insurance system for citizens over age 65, enacted in 1965, is organized as a single fund tied to the national Social Security (public pension) Administration, an independent agency within the national government, with reimbursement arrangements supervised by the Centers for Medicare and Medicaid Services (CMS) inside the Department of Health and Human Services. Medicare covers inpatient hospital care plus limited post-hospital nursing home services (Medicare Part A). Supplemental private insurance policies are bought by covered individuals to help pay for outpatient physician visits (Medicare Part B) and for outpatient pharmaceuticals (Medicare Part D).

In Germany, 85% of the population is enrolled in one of 120 not-for-profit, monthly premium-based private SHI funds. This figure includes all individuals with annual incomes below 54,500 euros, who are required by law to join an SHI fund, as well as those with higher incomes who choose to enroll or remain. Eleven percent of the population—all having annual incomes above the mandatory SHI enrollment ceiling of 54,500 euros—have opted out of the SHI system to voluntarily enroll in claims-based private health insurance, whereas 4% of the citizenry is enrolled in sector-specific public programs such as the military. Since 2009, all SHI members pay a flat tax on gross monthly income as a contribution (8.2% in 2018, up to an upper income limit of 49,500 euros), which is transferred by their SHI fund to a national pool,

and then redistributed back to their chosen fund on an individual risk-adjusted basis. Employers send 7.3% of each employee's salary to the same national pool. Special arrangements exist for payments from self-employed, retired, and unemployed workers. Since 1995, there has been a separate mandatory social insurance fund for long-term care (LTC), with an annual premium of 1.95% of each adult's gross monthly income, split 50%–50% with their employer. Pensioners since 2004 are required to pay the full 1.95% from their pensions. Childless SHI enrollees pay a surcharge of 0.25% of monthly gross income. Overall, 78% of all health care expenditures in Germany were paid from public and/or mandatory private SHI sources.

In the Netherlands since 2006, all adult citizens pay a fixed premium (about 1453 euros in 2019) to their choice among 35 private health insurers (not-for-profit and for-profit), with four large insurance groups having over 1 million members each. In addition, employers pay 6.95% of salary below 51,400 euros for each employee into a national health insurance fund. Self-employed individuals pay 4.85% into the national fund for taxable income up to the same limit. Retired and unemployed individuals also make payments. In addition to the individual premiums paid to their choice of private insurance fund, payments from the national health insurance fund, adjusted by individual age, sex, and health characteristics, also are made to the individual's chosen insurer. The Netherlands has a separate mandatory social insurance fund for LTC (the ABWZ, since 2015 the WLZ, and now only for residential nursing home care) to which each employee pays 9.5% of taxable income beneath 33,600 euros every year. Self-employed, unemployed, and retired individuals also are required to pay premiums to the WLZ. Overall, including SHI revenues, public spending provided 87% of total health expenditures in 2014.

In Estonia, a former Soviet Republic that re-established an SHI system in 1991 upon regaining its independence, there is one national SHI fund that is an arm of the national government. This fund collects mandatory payments of 13% from salaried workers and 20% from self-employed individuals, covering both health care and retirement pensions. Overall, including SHI revenues, public spending accounted for 74.5% of total health expenditures in 2017.

Singapore, Japan, South Korea, and Taiwan have predominantly SHI systems of funding for individual care services. In these Asian countries (except Japan), there is one SHI fund that typically is operated as an arm of the national government.

In Singapore, starting in 1983, all employees up to age 50 have been required to place 20% of their income (employers add 16% more) into a personal health savings account to pay for direct health care costs,

TABLE 7-2 Developed Country Per Capita Health Expenditures

TAX FUNDED IN WESTERN EUROPE	SHI FUNDED IN WESTERN EUROPE	CENTRAL EUROPEAN	DEVELOPED ASIAN	DEVELOPED NORTH AMERICAN					
Spain	\$2738	Belgium	\$4149	Latvia	\$874	South Korea	\$2043	Canada	\$4458
Italy	\$2738	Germany	\$4714	Poland	\$809	Singapore	\$4083	United States	\$9869
UK	\$3958	Netherlands	\$4742	Czech Republic	\$1321	Japan	\$4233		
Denmark	\$5565	Switzerland	\$9835	Slovenia	\$1834				
Sweden	\$5710								

Abbreviations: SHI, social health insurance; UK, United Kingdom.

managed in their name by the Singapore government, called a Medisave account. Medisave accounts have a maximum amount, are tax-exempt, and receive interest payments (currently set at 4%). Consistent with a Confucian emphasis on family, the funds that accumulate in the Medisave account can be spent on health care for family members as well. If the accumulated funds are not spent on health care during the insured's life, they become part of the individual's personal estate and are distributed as a tax-free inheritance to his or her designated heirs. In addition, Singaporean citizens are also automatically enrolled into a second government-run health insurance plan called MediShield that pays for supplemental catastrophic, chronic, and long-term care. While citizens can opt out, 90% of citizens remain in the program. The Singapore government also operates a third, wholly tax-funded payer called Medifund that, with approval of a local neighborhood committee, will pay hospital costs for 3–4% of the population who are recognized as indigent. In part reflecting the high level of mandatory individual saving, public funding provided only 54.5% of total health expenditures in 2016.

In South Korea, a state-run SHI system was established in 1977, which in 1990 covered 30.9% of total health care costs. This percentage paid by the SHI system rose to 40.5% of total costs in 2017, with national tax revenue covering 16.9%, leaving out-of-pocket expenses at a relatively high 34.4% of total costs. Although there are legal ceilings on total out-of-pocket copayments for each 6-month period, over 70% of Korean adults purchase an additional private Voluntary Health Insurance policy to cover these additional direct expenditures. In 2000, three types of public SHI funds were merged into a single national state-run fund. As of 2018, 6% of an employee's salary must be paid as a social insurance contribution into this fund, with employees and employers each paying 50% of that amount. In 2008, an additional SHI fund was introduced to pay for LTC, operated by the main state-run SHI fund to reduce administrative costs. Contributions to the LTC fund are set at 6.55% of the individual's regular SHI contribution, coupled with 20% copayments for institutional care and 15% copayments for home care services.

The United States There is no single preponderant source of health care spending in the United States. The federal government's CMS reported that, for 2017, private health insurance covered 34% of total health expenditures, Medicare (mandatory SHI program for all citizens over 65) covered 20%, Medicaid (a joint federal-state welfare program for low-income citizens) covered 17%, and out-of-pocket paid 10%. Sources of funds for these programs were 28% from the federal government, 17% from state and local governments, 28% from private households, and 20% from private business (e.g., employers). The World Bank set public funding in the United States at 50.2% of total health expenditures in 2017.

In 2010, the passage of the Affordable Care Act (ACA) extended privately provided but heavily regulated and federally subsidized health insurance to many low- and middle-income uninsured individuals and families. Since the same act reduced the availability of existing individually purchased private health insurance, the total increase in the number of newly covered individuals was less than expected. Insurance premium increases for 2017 rose from 20% to over 100%, depending on the particular state, with additional increases in up-front deductible requirements, raising questions about the long-term sustainability of the ACA initiative. The recent Republican administration sought to repeal major financial and tax elements of the ACA and to replace existing subsidy arrangements with a system of refundable tax credits toward the establishment of individual health savings accounts and/or purchase of private health insurance on open cross-state markets (currently, private health insurance in the United States remains controlled at the separate 50-state level of government).

■ DELIVERING INDIVIDUAL PATIENT CARE SERVICES IN DEVELOPED COUNTRIES

Hospital Services In Europe, hospitals in both tax-funded and SHI-funded health systems are mostly publicly owned and operated by regional or municipal governments. In tax-funded health systems,

most hospital-based physicians are civil servants, employed on a negotiated salary basis (often by a physician labor union), and subject to most of the usual advantages and disadvantages of being a public sector employee. There are somewhat more private hospitals in SHI-funded health systems. However, most larger hospitals are public institutions operated by local governments, and most hospital physicians (with the notable exception of the Netherlands, where they are private contractors organized in private group practices) are, like those in tax-funded systems, public sector employees. In most tax-funded European countries (but not continental SHI-funded countries), few specialist physicians have office-based practices, and in both tax- and SHI-funded systems, office-based specialists do not have admitting privileges to publicly operated hospitals.

Most public hospitals in both tax-funded and SHI-funded health systems are single free-standing institutions that can be classified into three broad categories by complexity of patients admitted and number of specialties available: (1) district hospitals (four specialties: internal medicine, general surgery, obstetrics, and psychiatry); (2) regional hospitals (20 specialties); and (3) university hospitals (>40 specialties). In addition, many countries have a number of small, 15- to 20-bed, freestanding, private (typically for-profit) clinics. Recently, some tax-funded countries have begun to merge district and regional hospitals in an effort to improve the quality of care and create financial efficiencies (for example, Norway; planned for Denmark, also for Ireland; however, failed Parliamentary passage and brought down the coalition government, in Finland in 2019). Institutional mergers can be difficult to negotiate among publicly operated hospitals, due to the role that these large institutions play as important care providers and as large employers in smaller cities and towns, especially given political and union concerns about maintaining current employment levels. In the United States, financial and reimbursement pressures triggered by the implementation of the 2010 ACA have generated a number of private sector hospital mergers into larger hospital groups.

In tax-funded health systems, publicly funded patients who are admitted for an elective procedure cannot choose their specialist physician (except private-pay patients in "pay beds" in National Health Service [NHS] hospitals in England). Specialists are assigned by the clinic to a patient based on availability, with both junior and senior doctors placed in rotation.

Capital costs (buildings, large medical equipment) are publicly funded in all tax-funded systems and in most traditional SHI systems. For example, in Germany, capital costs for public hospitals are paid for by the regional governments. As a result, new capital investment is often allocated politically, according to location and political priorities. In Finland, local politicians in the 1980s would say that it "takes 10 years to build a hospital," meaning that it took that long to become a political priority for the regional government that controlled capital expenditures. Local politicians would therefore regularly overbuild when they got their one opportunity to obtain new capital.

Recently, efforts have been made to make public hospitals more responsible for their use of capital. In the Netherlands, public hospitals were shifted into private not-for-profit entities that are expected either to fund new capital from operating surplus or to borrow the funds from a bank based on a viable business plan. In England, more than 100 hospitals have been built using the Public Finance Initiative (PFI) program, in which private developers build turn-key facilities (thus taking capital costs off the public borrowing limit), and then rent these facilities back to the NHS and/or the relevant NHS Foundation Trust. In Sweden and Finland, while capital equipment is now a cost on hospital operating budgets, large new capital equipment and major building renovations remain politically driven processes often with extensive delays. In Stockholm County, the New Karolinska University Hospital opened in 2018 was built and is managed by a separate nonprofit public-private company.

In Singapore and South Korea, both of which are SHI funded, larger hospitals are publicly operated. However, there are a substantial number of smaller private clinics typically owned by specialist physicians. In the United States, the passage of the 2010 ACA has triggered the selling of many private specialist group practices to hospital groups,

transforming previously independent practicing physicians into hospital employees.

Primary Care Services Most primary health care in SHI-funded health systems, and also in an increasing number of tax-funded health systems (except in low-income areas of some large cities), is delivered by independent private general practitioners (GPs), working either individually or in small privately owned group practices. Recent changes in tax-funded health systems include Norway, where most primary care moved from municipally employed physicians to private-practice GPs in 2003, and Sweden, where, following a 2010 change in national reimbursement requirements, new privately owned not-for-profit and for-profit GP practices were established and now deliver 50% of all primary care visits.

In England, most primary care physicians are private GPs who are contractors to the NHS, working either independently or in small group practices. These private GPs own their own practices, which they can sell when they retire. However, as part of the original agreement to convince physicians to support the establishment of the NHS in 1948 (which most physicians strongly opposed), private GPs also receive a national government pension upon retirement. In the inner cities in England, there are some larger primary health clinics.

In 2001, England's private primary care doctors were organized into geographically based Primary Care Trusts (PCTs). These PCTs were allocated 80% of the total NHS budget to contract for elective hospital services required by their patients with both NHS hospital trusts as well as private hospitals. In 2013, PCTs were restructured into Clinical Commissioning Groups with similar contracting responsibilities.

In 2004, the Quality Outcomes Framework (QOF) was introduced as a quality of care-tied approach to providing additional income for NHS GPs. This regulatory mechanism in 2010 set 134 different standards for best practice primary care in four main domains: 86 clinical, 36 organizational, 4 preventive service, and 3 patient experience. GP income grew on average by 25% through the introduction of the QOF with general practices averaging 96% of possible QOF points. Total spending on QOF in 2014 in England consumed 15% of all primary care expenditures.

In April 2019, a slightly revised QOF contract was implemented, which retired 28 low-value indicators, introduced 15 new more clinically appropriate indicators, added two Quality Improvement modules, and added a new personalized care adjustment option. Funding was only changed marginally.

Access for individuals to primary care services is considered good in SHI-funded systems such as those in Germany and the Netherlands. One often-cited reason is that private office-based physicians (both GPs and specialists) in these countries are paid on a modified fee-for-service basis. In Germany, office-based physicians are paid on a quarterly basis by the Sickness funds, acting jointly at the Länder (regional) level through a point-based system. A national agreement between the physician association and the association of sick funds establishes points for each clinical act. Similarly, the association of sick funds (led in each of Germany's 16 Länder by the fund with the most subscribers in that region) establishes a fixed budget for all office-based physician services for all sick fund patients each 3-month period. Retrospectively at the end of each period, the total number of points is divided into the sick funds' fixed allocation for office-based physicians for that Länder for that quarter, establishing the value of a point for that quarter. Subsequently, each office-based physician's point total is multiplied by that quarterly point value, resulting in that physician's total payment from the statutory sick funds.

In contrast to SHI systems, seeing a primary care doctor in a number of tax-funded health systems has become increasingly difficult over the past decade. In Sweden, in 2005, a "care guarantee" was introduced that required its predominantly publicly operated health centers to see a patient within 7 days after calling for an appointment. In Finland, where public primary health care centers used to provide most primary care visits, delays in getting public health center appointments have pushed up to 40% of all visits into a parallel occupational health system, as well as to publicly employed primary care physicians working privately in the afternoons.

In England in 2019, access to GP services has been labeled a "crisis," aggravated by a 6% fall in the number of practicing GPs, leading to delays of up to 30 days for an appointment in urban areas like London. A 2019 report by the King's Fund found that only 1 in 20 trainee GPs planned to work full time. Also in 2019, the Nuffield Trust published a report suggesting that future planning for primary care services in England should assume a permanent shortage of GPs, requiring large numbers of new nurse practitioners and other auxiliary personnel. In Central European countries that were formerly within the Soviet Bloc, primary care provision had to be newly established after independence was regained in 1991, since first-line care in the former Semashko model was provided in specialist polyclinics. Primary care doctors rapidly emerged as almost entirely private for-profit GPs, working on contract from the national SHI fund (Estonia, Hungary, North Macedonia), from state-regulated private insurance companies (Czech Republic), or from regional/municipal public payers (Poland). Private GPs in most Central European countries now are paid on a per-visit-tied basis. This arrangement was heavily influenced by the structure of primary care in Germany, where private office-based GPs are paid according to a point-system-tied framework.

In Asian countries such as Singapore, South Korea, and Japan, most primary care is provided by private for-profit GPs working independently or in small group practices. Private GPs are reimbursed at a set per-service fee by the national SHI fund(s). Access to primary care physicians is considered good.

Developed countries have varying policies regarding access to individual preventive services. Health systems in most countries provide vaccinations and mammography as part of funded health care services. In the United States, most insured individuals—and in Canada, most covered residents—automatically receive an annual physical exam including full blood profiles. In Norway and Denmark, adult physical exams are provided only upon special request by the individual, and in Sweden, adult physical exams are provided only to pregnant women. In Sweden, adults who wish to know their cholesterol or prostate-specific antigen (PSA) levels have begun to purchase blood tests out-of-pocket from private laboratories. In England in 2019, the NHS announced it would stop providing PSA screening tests for prostate cancer, even to men who requested one, similarly forcing concerned patients to purchase private laboratory testing.

Patients must make copayments to see a primary care doctor in some tax-funded health systems and in most SHI countries. In tax-funded systems, for example, Swedish patients are required to make a county-council-set copayment for each primary care visit up to a national-government-set annual ceiling, after which ambulatory visits (both primary and outpatient specialist) are not charged. Finland has a fixed copay for public health center visits, while Denmark's private GP visits do not have a copayment. In England, there is no copayment for GP visits.

In SHI health care systems in Europe and in Asia, patients usually are responsible for a copayment for both primary and office-based specialist care. To defray these charges (and to pay for other nonfunded services), a high percentage of citizens typically purchase additional supplemental health insurance. In France, where 95% of patients in 2015 purchased private supplemental insurance, patients paid directly the full fee for 65% of outpatient primary and specialist services, reimbursed subsequently by both their SHI fund and their supplemental insurance carrier for all payments (after deductibles), while for 35% of services (for low-income individuals and certain high-cost procedures), full agreed prices were paid directly to providers by SHI.

Access to Elective Specialist Care Approximately half of all European health care systems have a gatekeeping system that requires referrals from primary care physicians in order to book hospital specialist visits (for publicly paid visits). In most tax-funded health systems (although not in most SHI systems), there are substantial waiting times, typically several months or more, for elective specialist appointments as well as for high-tech diagnostic and treatment procedures. Waiting times can be particularly long for cancer and other elective surgical or high-demand services. In Sweden, government figures from the

summer of 2017 showed that, nationally, only 5–10% of prostate cancer operations were performed within 60 days after diagnosis.

In the English NHS, waiting lists for elective surgery in 2019 were often 6 months or longer. In August 2017, there were over 4,000,000 patients on NHS waiting lists. In January 2018, what administrators termed “a severe flu season,” during which hospital emergency rooms were overwhelmed with elderly patients requiring admission, led to a national-level NHS decision to cancel all elective operating room procedures in all hospitals in England (>50,000 procedures in 1200 hospitals) for the entire month of January, further lengthening waiting lists. Regarding quality of care, again in England, a March 2018 report from the national Office of Health Economics found that, in 2016 and 2017, up to three-quarters of patients who could have undergone key-hole procedures were forced to undergo open surgery, resulting in an estimated 1 million procedures each year that were more invasive than clinically necessary.

Delays in some tax-funded systems also are procedural. In England, for example, a patient who requires a further consultation with a second specialist typically has to return to their primary care physician for a second referral and then has to wait in the regular patient queue for that second appointment.

There is also substantial waiting time for radiologic imaging services in most tax-funded systems. In Malta, the tax-funded health system’s recent efforts to prioritize elective MRI investigations have succeeded in reducing waiting times from 18 months to 4 months. In both the Alberta and British Columbia Provinces in Canada, waiting times for a publicly funded nonemergency MRI can extend up to several months, whereas privately paid MRIs were available in both provinces within 1 week.

This issue of waiting times for specialist services in tax-funded health systems reflects a combination of growing demand (increasing/aging populations and changing clinical indications), financial constraints, and insufficient capacity, including inadequate physician working hours. For example, in the 1980s, when several surgical procedures for the elderly became more routine practice (e.g., hip replacement, coronary artery bypass graft, corneal lens implantation), the waiting list problem worsened. It had been mitigated somewhat through increased service capacity by the early 2000s, only to return as a growing policy challenge once public sector financial resources became constrained again after the 2008 global financial crisis. Timely cancer diagnosis and care continue to be a particularly sensitive issue, with tax-funded systems often taking several months for a patient to see an oncologist and then months more to begin treatment. In 2013 in Sweden, a newspaper journalist set off a political storm when he described women patients in one large county council (Malmö) who had to wait more than 40 days to receive the results from their breast cancer biopsy. In September 2019 in England, only 76.9% of patients with suspected cancer began treatment within 2 months of an urgent referral from a GP.

In response to pressure from national patient associations, a number of tax-funded health care systems introduced maximum waiting times for elective hospital procedures in the early 2000s. (Most Western European SHI systems do not have long waiting times or treatment guarantees for hospital care.) These maximum waiting times typically include initial primary care visits as well as specialist evaluations and treatment. In Denmark, a patient has the right to go to a different Danish public hospital for care after waiting 30 days without treatment. In Sweden, under the 2005 “waiting time guarantee,” an untreated patient’s local county council is required to pay for care in another county’s hospital after 180 days. In a parallel process at the European Union (EU) level, beginning in 1997, the EU Court of Justice steadily expanded the right of all EU citizens to travel to another EU country in order to receive “timely” care, with their home country health system required to pay for that care.

In private not-for-profit SHI-funded health systems such as in Germany and Switzerland, waiting times for specialist visits and hospital procedures are typically a few weeks to 1 month. In the SHI system in France, which is more centrally organized and funded (part of the Napoleonic tradition of public administration), ongoing disputes

about insufficient central government funding for public hospitals and staff salaries led in March 2019 to 9 months of hospital staff strikes, particularly in accident and emergency departments. In November 2019, the national government announced that it would take over 10 billion euros in public hospital debt as part of an effort to reverse staff cutbacks, bed and operating theater closures, and personnel flight to the private sector.

Long-Term Care Services LTC (consisting of residential and home-based services) consumes a relatively small but increasing proportion of gross domestic product (GDP) in developed countries. In 2016, Norway (2.95% GDP), Sweden (2.87% GDP), and the Netherlands (2.64% GDP) all spent more than one-fourth of their total health expenditures on LTC (Eurostat and OECD figures). More than one-fifth of all health care expenditures went to LTC in Belgium (2.16% GDP), Ireland (1.55% GDP), and Denmark (2.5% GDP). Lower-spending countries included the United Kingdom (18% of health expenditures; 1.75% GDP), Germany (12% of expenditures; 1.33% GDP), and Spain (9% of expenditures; 0.81% GDP). In the United States, official figures put total LTC expenditures in 2016 at 4.9% of total health expenditures, or 0.9% of total GDP. (Note that these figures do not include emergency, inpatient, or outpatient hospital costs generated by elderly patients.)

Since nursing home care is more expensive than home care (nursing home care requires the provision of housing, food, and around-the-clock care providers), government policymakers seek to keep the elderly and the chronically ill out of nursing homes for as long as feasible. Moreover, in developed countries like Sweden, Norway, and the United States, some 70% of all home care services come from informal caregivers: spouses, children (typically daughters), neighbors, and nonprofit community groups. While some SHI systems (e.g., Germany) have separate public LTC insurance (funded by mandatory premiums paid by all adults) that make available cash payments for LTC that can be used to compensate informal caregivers, most policymakers work hard to not monetize what is a large amount of essentially free care. Indeed, policymakers actively seek to encourage those providing these services to continue to do so as long as possible, trying to postpone caregiver burnout by providing support services such as free respite care, special call-in lines for caregiving advice, pension points toward retirement for the informal caregiver (Nordic countries), and free day-care center services.

In most tax-funded and SHI-funded European countries, home care services are organized at the municipal government level. In tax-funded systems, these services are also delivered mostly by municipal employees, working according to union-negotiated protocols. In some European SHI systems, and recently in tax-funded Sweden and Finland, private companies also provide home care services on contract to municipal governments. In combination with national legislation, these municipal systems also provide important support for informal caregivers, since the financial costs of caring for adults in their own home are substantially less than providing housing, food, and caregiver support in publicly funded homes for the aged or in nursing homes.

A high proportion of nursing homes in European tax-funded and SHI-funded health systems are publicly owned facilities operated by municipal governments; in some instances, in SHI-funded systems (Israel, the Netherlands), they are operated by private not-for-profit organizations. Recently, in some tax-funded systems (e.g., Sweden), private for-profit chains have begun to open nursing homes that are funded on a contract basis with local municipal governments. Costs for nursing home care can be expensive: in Norway, the cost per patient is often over \$100,000 per year in a publicly funded home, with the patient responsible for paying up to 80% depending on the family’s economic status. In Sweden, patients living in publicly funded nursing homes in Stockholm County pay a relatively small official fee, but they also pay room rent and up to 2706 Swedish krona (SEK) per month (about \$270 U.S. dollars [USD]) for food out of their monthly public pension payments.

In 2012, in an effort to reduce demand for expensive hospital and nursing home services, Norway and Denmark began elderly care reforms that shifted service delivery as well as funding responsibilities

to municipal governments. Among innovations in Norway, municipalities are required to establish a municipal acute bed unit (MAU) to treat stable elderly patients and provide observation beds for evaluation. Partial funding for these units is provided by the four public regional health care administrations. Some municipalities have also embedded primary care units inside their regional hospital to arrange discharge and to coordinate care for the chronically ill elderly. Norwegian municipalities are also responsible through their contracted (mostly private) primary care physicians to implement the National Pathways Program, which established treatment protocols for cross-sector conditions such as diabetes and cardiovascular conditions.

A differently configured structural innovation to better integrate LTC for the chronically ill elderly with clinical individual health services has been to consolidate both social and health care services within the same public administrative organization. In 2019, as part of health reforms in Ireland and Denmark and a proposed (unenacted) reform in Finland, as well as a pilot decentralization program in England for 2.8 million people in Greater Manchester, social and health care programs are to be administered by a single responsible agency.

In the SHI-funded system in the Netherlands, almost 7% of the population live in a residential home. National government legislation revised the structure of nursing home funding and care in 2015. Three acts restructured the separate public LTC SHI fund, which requires mandatory payments by 100% of Dutch adults, and introduced delivery-related reforms that reduced the number and overall cost of nursing home patients paid for by the fund. Determination of eligibility for public payment for nursing home care is now made by an independent national assessment body (the Centre for Needs Assessment). Moreover, municipal governments now play a stronger role in funding and delivering home care services. The reforms created social care teams that hold “kitchen table talks” to steer the elderly first toward seeking care from family, neighbors, churches, and other local community organizations before they qualify for publicly paid in-home care. In 2012, some 1.5 million people (12% of total population) provided informal care to ill or disabled persons, averaging 22 hours per week of care per person.

Home care recipients in the Netherlands can choose to set up a “personal budget,” using their public funding allocation to select their preferred individual care personnel (either publicly employed or publicly approved private providers). This arrangement also enables these home care recipients to determine the particular mix of services they want, as well as to augment the allocated public funds with personal funds. A number of innovative not-for-profit nursing homes have been created to provide additional services to elderly living in their neighborhood (primary care home visits), as well as terminal hospice care (e.g., the Saffier De Residentie Groep residences in The Haag).

In the United States, nursing home and home care are funded and delivered in a variety of different ways. For individuals who have minimal financial assets, nursing home costs are paid by a joint federal-regional (state) welfare program called Medicaid. Most state government Medicaid programs pay out more than 40% of their total budget for nursing home care. In the past, Medicaid did not pay for home care services. However, some states have programs with private for-profit and not-for-profit providers that provide home care as a way to forestall the need for the more expensive nursing home care.

Many private individuals take out private LTC insurance, typically from commercial insurance companies. These policies require individuals to make premium payments for years in advance (often 20 or more) before the individual learns whether they will, in fact, require home or nursing home care. Some private insurers have also raised premiums after individuals have paid in for many years and canceled policies if the new higher rate is not affordable. The 2010 ACA contained a new public LTC insurance program. However, the program was designed to be voluntary, and U.S. Department of Health and Human Services administrators decided in 2013 not to implement that portion of the law.

In addition to the tax-funded Medicaid program and privately purchased LTC insurance, many middle-class families pay for care from savings, by selling the elderly person’s home, or by direct contribution

from children and other family members. Expenses can reach between \$60,000 and \$100,000 per year depending on the location of a facility and who operates it.

Nursing home care in the United States is provided by a wide mix of private not-for-profit and for-profit providers, ranging from church-owned single-site homes to large stock market-listed companies. Many of these homes are purpose-built as assisted-living or memory-care facilities. Home care services are delivered by a mix of private not-for-profit and for-profit providers.

In Japan, a national LTC insurance fund was introduced in 2000. Although the new fund applies uniformly across the country, the program is administered by municipal governments and the premium level differs across municipalities, with an average monthly premium of 3000 yen (about \$30 USD). In South Korea, an SHI fund for LTC is funded by mandatory contributions of 4.78% of a person’s regular national health insurance contribution, with an additional 20% of total LTC expenditures provided by national government funds. The client copayment for home care is set at 15% of expenses and at 20% for residential care.

■ PHARMACEUTICALS

Pharmaceutical expenditures in developed countries (inpatient and outpatient combined) vary widely across different health system types, as well as between different countries within each institutional type. OECD figures for 2018 show drug expenditures in tax-funded countries in Western Europe ranging from 6.3% of total health expenditures (THE) in Denmark to 11.9% of THE in the United Kingdom and 18.6% of THE in Spain. In SHI-funded Western European systems, pharmaceuticals absorbed 7.5% of THE in the Netherlands, while in Germany, that figure was 14.1%. In the hybrid tax-funded SHI systems of Central Europe, the pharmaceutical percentage of THE is higher: 18.2% of THE in Estonia to 27.9% of THE in Hungary. Similarly, in Asian SHI systems, pharmaceuticals consumed 20.7% of THE in South Korea and 18.6% of THE in Japan. The OECD’s 2018 figures for pharmaceutical spending in North America are 12.0% of THE in the United States and 16.7% in Canada.

Contributing factors to this wide-ranging variation are (1) differences in national practice and prescription patterns reflecting differing cultural expectations; (2) the ratio problem (relatively fixed level of pharmaceutical costs due to international prices—the numerator—divided by a greatly varying per capita health expenditure cost in different developed country health systems); (3) the range and type of pharmaceutical price controls in each country; and (4) the degree of limitation placed on pharmaceutical supply, tied to formularies and/or explicit forms of drug rationing.

Most European health systems have tight national controls on the cost and, in some tax-based countries, on the availability of pharmaceuticals. Most European countries also use a number of different regulatory measures to limit prices and/or availability of both inpatient and outpatient drugs, including mandatory generic prescribing, reference pricing, patient copays (sometimes with an annual ceiling, after which copayments are no longer required), and (particularly in tax-funded systems) national formularies tied to clinical effectiveness. Norway, for example, allows only about 2300 different preparations—including dosage, delivery method, and box size—to be stocked by pharmacies. Prices for drugs can vary considerably across different European countries, tied to economic development and domestic pricing patterns. One consequence of these differential national pricing controls has been the development of a parallel import market, in which drug wholesalers and pharmacists in the more expensive countries purchase supplies from a cheaper market elsewhere in Europe.

Access to expensive drugs has also been intentionally limited in some tax-funded health systems in Europe. One basis for rationing has been rationing tied to quality-adjusted life-years (QALYs). Rationing also reflects a clash between strained public drug budgets and public pressure. For example, in the case of cancer drugs in England, the recommendation of the National Institute for Health and Care Excellence (NICE) against funding the breast cancer drug trastuzumab (Herceptin) was subsequently overturned by the Minister of Health.

Expensive cancer drugs continue to be rationed in England where the NHS Cancer Drug Fund, established in 2011 to provide access to non-NHS-provided drugs on a case-by-case basis, ran out of funds in 2015, forcing it to drop 25 of 83 covered drugs and close down for 3 months to restructure its operations.

As part of earlier medical patterns in Asian countries, office-based physicians traditionally filled prescriptions as well as prescribing drugs to patients. These sales also served to supplement their income in a setting of relatively low per-visit payments from state-run SHI funds. Concerned about cost and overuse, both Taiwan (in 1997, except for emergency cases or rural regions) and South Korea (for the whole country in 2003) implemented “separation reforms,” which ended these physician sales. In Japan, a series of fee and reimbursement reforms have trimmed the percentage of all prescriptions dispensed in 2016 by physicians to 26% of prescriptions filled.

■ GOVERNANCE AND REGULATION

Health care services in developed countries are steered, constrained, monitored, and (to varying degrees) assessed by governments and governmentally established and/or empowered bodies. Although these measures apply particularly to the financial efficiency of government-funded services, they also seek to promote patient and community safety, equity of access, and high-quality clinical outcomes. This oversight is often strongly focused on privately operated and contracted providers and insurers, although in principle, it applies to publicly operated organizations as well.

Governance consists of macro national-level policy, meso institutional-level management, and micro clinic-level care decisions. This complex mix of governance decisions is often shared among different national, regional, and local governments, depending on the degree of centralization, decentralization, or, recently, recentralization (e.g., Norway and Denmark). While most systems officially prioritize “good governance,” governance activities frequently comingle with political objectives as core policy concepts are developed and transformed into concrete organizational targets.

In Sweden, health system governance is shared among national, regional (county), and local municipal governments. The national government has responsibility to pass “frame” legislation, which establishes the basic structure of the system. To cite one example, until recently, the national government had limited an adult patient’s total copayments for outpatient physician care (specialist and primary care) and pharmaceuticals to 2800 SEK (about \$280 USD) for a 12-month period. The 20 regional governments, in turn, made policy decisions within that legislation, deciding how to apportion the specific copayments for each primary care and specialist outpatient visit. Since Swedes can self-refer to specialists, some counties double the copayment to hospital-based doctors to discourage unnecessary appointments. Similarly, fiscal policy normally is shared between the regional government, which raises about 70% of total health expenditures through its own county-set flat income tax, and the national government, which provides additional purpose-tied funds for national objectives such as consolidating open-heart surgery across county lines as well as supplementing lower tax receipts in rural counties with smaller working populations. However, this normal funding relationship across governments can change. In the early 1990s, the national government placed a “stop” on raising county taxes prior to Sweden’s admission in 1995 to the EU. In 2016, each of the 20 counties could set their own ceilings, which were almost all at 3300 SEK (about \$330 USD).

In Spain’s tax-funded health system (71.1% publicly funded in 2015), 17 regional “autonomous communities” were given full managerial responsibility for the provision of health services in a 1990s decentralization process, along with ownership of all publicly operated hospitals. The national government generates a substantial proportion of health care resources, which are included in the broad block grants it allocates to the regional governments, which then add regional tax revenue to make up the full public sector budget. In a mechanism to steer regional government operating policies in this decentralized environment, the national Spanish government established a joint federal-regional council to review quality and performance data (through the 2003 Health

System Cohesion and Quality Act). Italy’s tax-funded health system (75.8% publicly funded in 2014) similarly shares governance responsibilities between national and regional governments. Health services are provided by local health authorities (Azienda Sanitaria Locale) supervised by 20 regional governments within a nationally established governance framework, financed through a complicated mix of national and nationally stipulated but regionally collected taxes. Again, like Spain, the national government established a federal-regional government council, seeking to better coordinate care standards and information among the regions and with national government agencies. In 2006, the national government imposed strict financial plans on 10 regions that were systematically in deficit.

In Germany, where funding for its SHI-based health system is predominantly the responsibility of 120 private not-for-profit sickness funds, governance decisions are shared among these private sector sickness funds and public sector national, regional, and municipal governments. The sickness funds receive a risk-adjusted premium payment for each enrolled individual, according to a national government-determined formula, and from a national government-run health insurance pool. Most hospitals are owned and operated by municipal governments, while investment capital for structural renovations and new building comes from the 16 regional Länder taken from their tax revenues. Payment frameworks and amounts for public hospitals are negotiated between associations of these municipally owned hospitals and associations of the private sickness funds, without formal government participation.

Regulation is an essential element of an effective health care system and a key component of overall health system governance. Regulation incorporates both broad standard requirements that affect all organizations that operate in a country (e.g., hiring, firing, and wage decisions) as well as specific health sector-related regulations (e.g., proper handling, use, and disposal of low-grade nuclear waste from radiation treatments). Recent examples of health sector regulation in England, for example, include the following:

1. Requiring all cancer drugs adopted for use in the NHS to cost no more than \$41,268/QALY;
2. Requiring in their employment contract that junior doctors in hospitals work a specific number of Sundays; and
3. Requiring that all emergency department patients receive care within 4 h of their arrival.

A powerful tool that has the force of law, regulation can have substantial negative as well as positive effects. A well-known political science corollary of regulatory power is that “the right to regulate is also the right to destroy.” For example, in the United States, the federal Environmental Protection Agency, as part of its pursuit of cleaner air, issued wide-ranging regulatory orders setting performance standards that resulted in the closing of many West Virginia coal mines, with the loss of tens of millions of dollars of productive capacity and thousands of high-paying jobs, and likely contributing to social conditions that helped spawn that state’s high rates of opioid abuse among unemployed males. Similarly, in some tax-funded European systems, such as those in Sweden and England, there is growing pressure from public health advocates for national regulations to prohibit the making of a profit from publicly paid funds. In Sweden, the national government’s Reepalu report in 2016 honored a pledge made by the Social Democratic government to its Left (socialist) Party ally by calling for a legislated ban on profit-making in the provision of publicly funded health care services. The report’s publication triggered substantial divestment of existing investor-owned primary care, nursing home, and home care companies.

■ FUTURE CHALLENGES

Health systems in developed countries face continued challenges in the coming years. These include financial, organizational, and policy dilemmas for which institutionally viable, financially sustainable, and politically supportable solutions will be complicated to develop and difficult to implement. On the delivery side, a key question is whether

privately structured GP-based primary care is more efficient and effective than various clinic-based forms of primary care services. Recent movement in Northern and Central Europe toward more private GPs, along with continued private office-based primary care in much of Canada, the United States, and economically developed countries in Asia, raises complex policy issues for international organizations like the World Health Organization (WHO), as well as national policymakers. In the hospital sector, existing levels of clinical quality and patient responsiveness in publicly operated command-and-control institutions will increasingly have to compete with those of semi-autonomous public hospitals, as well as various types of private, sometimes very innovative providers. In the financing arena, continued pressure on publicly raised health system revenues is likely to erode longtime commitments in some tax-funded health systems to minimal patient copayments and low out-of-pocket funding.

An additional set of challenges will arise from recent commitments by international organizations like WHO to restructure health systems in developed countries to better address the social determinants of health. This new, incomplete strategy calls for a dramatic expansion of health sector responsibility to include a wide range of existing institutional arrangements in housing, education, work-life, and social and political decision-making. The influential 2010 Strategic Review of Health Inequalities in England entitled “Fair Society, Healthy Lives,” led by Sir Michael Marmot, a British epidemiologist, called for the elimination of all “inequities in power, money, and resources.” Separate from the political dimensions of this proposed new paradigm, how such fundamental societal change will be funded and implemented has yet to be addressed.

Looking forward, among the most essential challenges to national decision-makers in the coming period will be four specific health system imperatives:

- 1. Finding a more sustainable balance between ethics and funding.** Policymakers in publicly funded health systems face a growing gap between patient expectations of high-quality clinical care, staff expectations of better compensation, and the economic imperative of no new taxes. Recent research has suggested that SHI-funded health systems, faced with increasing aging and thus proportionally fewer employed, face a similar gap. While the present solidaristic foundation for raising collective revenues is insufficient, available nonsolidaristic tools (copayments, supplemental insurance, private pay) inevitably contribute to overall inequality. But what then are the realistic policy alternatives? The minimalist new policy goal necessarily will have to become one of raising new revenues while doing the least economic and social harm.

- 2. Developing better strategies to steer provider diversity.**

Health systems in developed countries are becoming more diverse with more and different types of public owners: hospital trusts, state enterprises, and mixed public-private hospital owners/managers. There also are more and different types of private providers: not-for-profit community groups, foundations, and cooperatives, as well as for-profit small local entrepreneurs, large international companies, and risk capital funds (venture capital). Furthermore, new innovative delivery models are reorganizing traditional service boundaries: not-for-profit private nursing homes in the Netherlands also provide outpatient primary care to neighborhood elderly patients, as well as hospice care; Israeli technology companies combine high-tech home-based patient monitoring with standard medical and custodial home care services. Public pressure from citizens for more choice and better outcomes will pressure policymakers toward new, more accommodative health system arrangements. A 2019 national government report in Sweden on the hospital sector recommended a new emphasis on better access to out-of-office hours and out-of-hospital acute care by private as well as public providers.

- 3. Ensuring better coordination between social and health services.** Tax-funded and SHI-funded systems alike are under intense policy pressure to develop better strategies to integrate services for the chronically ill elderly, as a way to improve the quality of services that these patients receive and to keep them at home healthier and

longer, reducing expensive acute visits to hospitals and emergency departments. The clear delivery system goal will increasingly be to keep the elderly out of nursing homes and acute care facilities for as long as possible.

4. Building labor unions into provider innovation.

In many developed countries, health sector staff, including hospital physicians, are members of labor unions. Effective policymaking will require finding mechanisms to build these personnel unions into accelerated health system restructuring processes. This process will necessarily involve integrating unions into more innovative, flexible, fiscally sustainable organizational arrangements with contracts that reward active participation in organizational change, contracts that pay incentives to more productive employees, quicker reassignment and redundancy procedures (firing health sector workers can take a year or longer in some European health systems), and establishing profit-sharing payments to teams/unions, also in public sector organizations.

While the structure and complexity of resolving these specific organizational challenges will vary depending on a country’s cultural and institutional context, the commonality of these problems suggests that health systems in the developed world require a new, broader range of targeted policy strategies and solutions.

■ FINANCING AND PROVIDING HEALTH SERVICES IN DEVELOPING COUNTRIES (See also Chap. 474)

Health systems in developing countries reflect a complex combination of the same core elements found in developed country systems (hospitals, primary care facilities, medical staff, pharmaceuticals) adapted to different, widely varying organizational, social, political, and economic contexts and conditions. System structure and provider institutions typically vary by differing national characteristics including historical relationships (Anglophone/Francophone/Hispanic/Soviet Semashko/American institutional and educational links); GDP and per capita annual national income (low- or middle-income developing countries); political norms and values; and ethnic and/or cultural mix. Predominantly public sector funding, particularly in lower-income countries, typically generates substantially lower levels of resources per capita than in developed countries and tends to be less reliable, particularly in countries where the economy is dependent on commodity exports.

Service delivery arrangements in developing countries, in turn, typically have higher provider-to-population ratios as well as, in public sector institutions, more mixed quality of care. In a number of middle-income developing countries, migration of trained medical staff to practice in higher-paying developed country health systems (often going to countries with historical relationships and/or where they received advanced training) further depletes available medical resources. In nearly all developing countries, private sector providers play an important supplemental role, with some middle-income developing countries like China currently encouraging their further development.

Most middle- and lower-income developing countries struggle to fund high-quality individual health services. Recent emphasis on universal health coverage has intensified that struggle. In middle-income developing countries (Table 7-3), World Bank data from 2016 show

TABLE 7-3 Middle-Income Developing Countries: Total Health Expenditure (% of gross domestic product)

Middle-Income Developing Countries	
Kazakhstan	3.53%
Thailand	3.71%
Malaysia	3.80%
Turkey	4.31%
China	4.98%
Botswana	5.46%
Mexico	5.47%
Colombia	5.91%

TABLE 7-4 Low-Income Developing Countries: Total Health Expenditure (% of gross domestic product)

Low-Income Developing Countries	
Nigeria	3.65%
India	3.66%
Ethiopia	3.97%
Nepal	6.29%
Honduras	8.40%

a range of health expenditure rates as a percentage of GDP, including Kazakhstan at 3.53% of GDP, Thailand at 3.71%, Malaysia at 3.80%, Turkey at 4.31%, China at 4.98%, Botswana at 5.46%, Mexico at 5.47%, and Colombia at 5.91%. Total health spending in low-income developing countries (**Table 7-4**) ranges from 3.65% of GDP for Nigeria, 3.66% for India, 3.97% for Ethiopia, 6.29% for Nepal, to 8.40% for Honduras.

Given lower aggregate GDP levels, per capita annual expenditures are considerably less than those found in developed countries. In middle-income developing countries (**Table 7-5**), Thailand spent (2016 data in adjusted USD) \$221 annually per person, Kazakhstan spent \$262, Colombia spent \$340, Malaysia spent \$361, Botswana spent \$379, China spent \$398, Mexico spent \$461, and Turkey spent \$468. Among low-income developing countries (**Table 7-6**), Ethiopia spent \$27 per person annually, Nepal spent \$45, India spent \$62, and Nigeria spent \$79, whereas Honduras spent \$199.

China provides an interesting example of financing and service delivery development in middle-income developing countries. Financing reforms replaced fully publicly funded services with three new arrangements tied to work status and residence: (1) Urban Employee Basic Medical Insurance in 1998 (incorporating privately funded medical savings accounts—a concept pioneered in Singapore); (2) Urban Resident Basic Medical Insurance in 2007; and (3) New Rural Cooperative Medical Scheme in 2007. The urban employee program is an SHI model reflecting the rapid rate of economic growth and increasing incomes for urban workers. Starting in 2013, the Chinese government increasingly emphasized the development of new private hospitals and promotion of private insurance in urban areas. These and other health sector reforms became possible as continued strong economic growth over 30 years raised an estimated 300 million Chinese into the middle class, generating the requisite private as well as public revenues to underpin major structural health sector change.

Service delivery in developing countries varies widely in access, quality, and outcomes across and also within many developing countries. Medical services and tertiary institutions in urban areas of China, for example, operate at a substantially higher standard of service than those typically available in poorer rural regions. Similar disparities exist in wealthier parts of India such as Rajasthan, whereas in poorer states such as Bihar, primary care is mostly delivered by community “volunteers” with basic medical training, supervised by a GP.

Two critical challenges for all developing country health systems are contingent on generating adequate future funding flows. First, the current push from United Nations agencies to achieve universal health coverage will require additional public and private sector funding to

TABLE 7-5 Middle-Income Developing Countries: Per Capita Health Expenditures

Middle-Income Developing Countries	
Thailand	\$221
Kazakhstan	\$262
Colombia	\$340
Malaysia	\$361
Botswana	\$379
China	\$398
Mexico	\$461
Turkey	\$468

TABLE 7-6 Low-Income Developing Countries: Per Capital Health Expenditures

Low-Income Developing Countries	
Ethiopia	\$27
Nepal	\$45
India	\$62
Nigeria	\$79
Honduras	\$199

pay for the necessary new providers and services. Second, available funding will need to be more effectively targeted on needed and appropriate services, with minimized managerial inefficiencies and substantially less political corruption.

Both forms of expanded funding will be dependent on strong national and global economic growth, which in turn will require continued country-level economic and political reforms. Achieving both funding-related objectives will require considerable international as well as national effort.

FURTHER READING

BARBER SL et al: *Price Setting and Price Regulation in Health Care: Lessons for Advancing Universal Health Coverage*. Geneva, World Health Organization, Organization for Economic Co-operation and Development, 2019. <https://apps.who.int/iris/bitstream/handle/10665/325547/9789241515924-eng.pdf>.

FIGUERAS J, MCKEE M (eds): *Health Systems, Health, Wealth, and Societal Well-Being: Assessing the Case for Investing in Health Systems*. Maidenhead, Open University Press/McGraw-Hill Education, 2011. www.euro.who.int/_data/assets/pdf_file/0007/164383/e96159.pdf.

HASELTINE W: *Affordable Excellence: The Singapore Health Story*. Washington, Brookings Institution Press, 2013. www.brookings.edu/wp-content/uploads/2016/07/AffordableExcellencePDF.pdf.

KUHLMANN E et al (eds): *The Palgrave International Handbook on Healthcare Policy and Governance*. London, Palgrave MacMillan, 2015.

RICE T et al: *United States of America: Health System Review*. Health in Transition (HiT) Series 15 (3). Brussels, European Observatory on Health Systems and Policies, 2013. www.euro.who.int/_data/assets/pdf_file/0019/215155/HiT-United-States-of-America.pdf.

8

The Safety and Quality of Health Care

David W. Bates



Safety and quality are two of the central dimensions of health care. In recent years, it has become easier to measure safety and quality, and it is increasingly clear that performance in both dimensions could be much better. The public is—with good justification—demanding measurement and accountability, and payment for services will increasingly be based on performance in these areas. Thus, physicians must learn about these two domains, how they can be improved, and the relative strengths and limitations of the current ability to measure them.

Safety and quality are closely related but do not completely overlap. The Institute of Medicine has suggested in a seminal series of reports that safety is the first part of quality and that the health care system must first and foremost guarantee that it will deliver safe care, although quality is also pivotal. In the end, it is likely that more net clinical

benefit will be derived from improving quality than from improving safety, though both are important and safety is in many ways more tangible to the public. The first section of this chapter will address issues relating to the safety of care and the second will cover quality of care.

■ SAFETY IN HEALTH CARE

Safety Theory and Systems Theory *Safety theory* clearly points out that individuals make errors all the time. Think of driving home from the hospital: you intend to stop and pick up a quart of milk on the way home but find yourself entering your driveway without realizing how you got there. Everybody uses low-level, semiautomatic behavior for many activities in daily life; this kind of error is called a *slip*. Slips occur often during care delivery—e.g., when people intend to write an order but forget because they must complete another action first. *Mistakes*, by contrast, are errors of a higher level; they occur in new or nonstereotypic situations in which conscious decisions are being made. An example would be dosing of a medication with which a physician is not familiar. The strategies used to prevent slips and mistakes are often different.

Systems theory suggests that most accidents occur as the result of a series of small failures that happen to line up in an individual instance so that an accident can occur (Fig. 8-1). It also suggests that most individuals in an industry such as health care are trying to do the right thing (e.g., deliver safe care) and that most accidents thus result from defects in systems. Systems should be designed both to make errors less likely and to identify those that do inevitably occur.

Factors That Increase the Likelihood of Errors Many factors ubiquitous in health care systems can increase the likelihood of errors, including fatigue, stress, interruptions, complexity, and transitions. The effects of fatigue in other industries are clear, but its effects in health care have been more controversial until recently. For example, the accident rate among truck drivers increases dramatically if they work over a certain number of hours in a week, especially with prolonged shifts. A recent study of house officers in the intensive care unit demonstrated that they were about one-third more likely to make errors when they were on a 24-h shift than when they were on a schedule that allowed them to sleep 8 h the previous night. The American College of Graduate Medical Education has moved to address this issue by putting in place the 80-h workweek. Although this stipulation is a step forward, it does not address the most important cause of fatigue-related errors: extended-duty shifts. High levels of stress and heavy workloads also can increase error rates. Thus, in extremely high-pressure situations, such as cardiac arrests, errors are more likely to occur. Strategies such as using protocols in these settings can be helpful, as can simple recognition that the situation is stressful.

Interruptions also increase the likelihood of error and occur frequently in health care delivery. It is common to forget to complete an

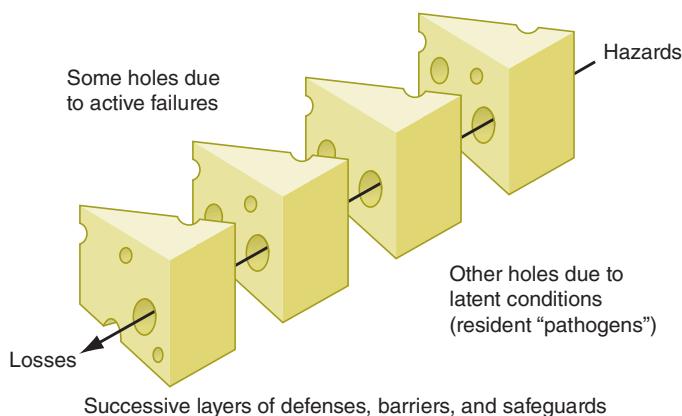


FIGURE 8-1 “Swiss cheese” diagram. Reason argues that most accidents occur when a series of “latent failures” are present in a system and happen to line up in a given instance, resulting in an accident. Examples of latent failures in the case of a fall might be that the unit is unusually busy and the floor happens to be wet. (Adapted from J Reason: BMJ 320:768, 2000.)

action when one is interrupted partway through it by a page, for example. Approaches that may be helpful in this area include minimizing interruptions and setting up tools that help define the urgency of an interruption.

Complexity represents a key issue that contributes to errors. Providers are confronted by streams of data (e.g., laboratory tests and vital signs), many of which provide little useful information but some of which are important and require action or suggest a specific diagnosis. Tools that emphasize specific abnormalities or combinations of abnormalities may be helpful in this area.

Transitions between providers and settings are also common in health care, especially with the advent of the 80-h workweek, and generally represent points of vulnerability. Tools that provide structure in exchanging information—for example, when transferring care between providers—may be helpful.

The Frequency of Adverse Events in Health Care Most large studies focusing on the frequency and consequences of adverse events have been performed in the inpatient setting; some data are available for nursing homes, but much less information is available about the outpatient setting. The Harvard Medical Practice Study, one of the largest studies to address this issue, was performed with hospitalized patients in New York. The primary outcome was the adverse event: an injury caused by medical management rather than by the patient's underlying disease. In this study, an event either resulted in death or disability at discharge or prolonged the length of hospital stay by at least 2 days. Key findings were that the adverse event rate was 3.7% and that 58% of the adverse events were considered preventable. Although New York is not representative of the United States as a whole, the study was replicated later in Colorado and Utah, where the rates were essentially similar. Since then, other studies using analogous methodologies have been performed in various developed nations, and the rates of adverse events in these countries appear to be ~10%. Rates of safety issues appear to be even higher in developing and transitional countries; thus, this is clearly an issue of global proportions.

In the Harvard Medical Practice Study, adverse drug events (ADEs) were most common, accounting for 19% of all adverse events, and were followed in frequency by wound infections (14%) and technical complications (13%). Almost half of adverse events were associated with a surgical procedure. Among nonoperative events, 37% were ADEs, 15% were diagnostic mishaps, 14% were therapeutic mishaps, 13% were procedure-related mishaps, and 5% were falls.

ADEs have been studied more than any other error category. Studies focusing specifically on ADEs have found that they appear to be much more common than was suggested by the Harvard Medical Practice Study, although most other studies use more inclusive criteria. Detection approaches in the research setting include chart review and the use of a computerized ADE monitor, a tool that explores the database and identifies signals that suggest an ADE may have occurred. Studies that use multiple approaches find more ADEs than does any individual approach, and this discrepancy suggests that the true underlying rate in the population is higher than would be identified by a single approach. About 6–10% of patients admitted to U.S. hospitals experience an ADE.

Injuries caused by drugs are also common in the outpatient setting. One study found a rate of 21 ADEs per every 100 patients per year when patients were called to assess whether they had had a problem with one of their medications. The severity level was lower than in the inpatient setting, but approximately one-third of these ADEs were preventable.

The period immediately after a patient is discharged from the hospital appears to be very risky. A recent study of patients hospitalized on a medical service found an adverse event rate of 19%; about one-third of those events were preventable, and another one-third were ameliorable (i.e., they could have been made less severe). ADEs were the single leading error category.

Prevention Strategies Most work on strategies to prevent adverse events has targeted specific types of events in the inpatient setting, with nosocomial infections and ADEs having received the most attention. Nosocomial infection rates have been reduced greatly in intensive care

settings, especially by using checklists. For ADEs, several strategies have been found to reduce the medication error rate, although it has been harder to demonstrate that they reduce the ADE rate overall, and no studies with adequate power to show a clinically meaningful reduction have been published.

Implementation of checklists to ensure that specific actions are carried out has had a major impact on rates of catheter-associated blood-stream infection and ventilator-associated pneumonia, two of the most serious complications occurring in intensive care units. The checklist concept is based on the premise that several specific actions can reduce the frequency of these issues; when these actions are all taken for every patient, the result has been an extreme reduction in the frequency of the associated complication. These practices have been disseminated across wide areas in the state of Michigan.

Computerized physician order entry (CPOE) linked with clinical decision support reduces the rate of serious medication errors, defined as those that harm someone or have the potential to do so. In one study, CPOE, even with limited decision support, decreased the serious medication error rate by 55%. CPOE can prevent medication errors by suggesting a default dose, ensuring that all orders are complete (e.g., that they include dose, route, and frequency), and checking orders for allergies, drug-drug interactions, and drug-laboratory issues. In addition, clinical decision support can suggest the right dose for a patient, tailoring it to the level of renal function and age. In one study, patients with renal insufficiency received the appropriate dose only one-third of the time without decision support, whereas that fraction increased to approximately two-thirds with decision support; moreover, with such support, patients with renal insufficiency were discharged from the hospital half a day earlier. As of 2019, over 95% of U.S. hospitals had implemented CPOE, although the decision support often is still limited.

Another technology that can improve medication safety is bar coding linked with an electronic medication administration record. Bar coding can help ensure that the right patient gets the right medication at the right time. Electronic medication administration records can make it much easier to determine what medications a patient has received. Studies to assess the impact of bar coding on medication safety are under way, and the early results are promising. Another technology to improve medication safety is “smart pumps.” These pumps can be set according to which medication is being given and at what dose; the health care professional will receive a warning if too high a dose is about to be administered.

The National Safety Picture Several organizations, including the National Quality Forum and The Joint Commission, have made recommendations for improving safety. The National Quality Forum has released recommendations to U.S. hospitals about what practices will most improve the safety of care, and all hospitals are expected to implement these recommendations. Many of these practices arise frequently in routine care. One example is “readback,” the practice of recording all verbal orders and immediately reading them back to the physician to verify the accuracy of what was heard. Another is the consistent use of standard abbreviations and dose designations; some abbreviations and dose designations are particularly prone to error (e.g., 7U may be read as 70).

Measurement of Safety Measuring the safety of care is difficult and expensive, since adverse events are, fortunately, rare. Most hospitals rely on spontaneous reporting to identify errors and adverse events, but the sensitivity of this approach is very low, with only ~1 in 20 ADEs reported. Promising research techniques involve searching the electronic record for signals suggesting that an adverse event has occurred. These methods are not yet in wide use but will probably be used routinely in the future. Claims data have been used to identify the frequency of adverse events; this approach works much better for surgical care than for medical care and requires additional validation. The net result is that, except for a few specific types of events (e.g., falls and nosocomial infections), hospitals have little idea about the true frequency of safety issues.

Nonetheless, all providers have the responsibility to report problems with safety as they are identified. All hospitals have spontaneous reporting systems, and if providers report events as they occur, those events can serve as lessons for subsequent improvement.

Conclusions about Safety It is abundantly clear that the safety of health care can be improved substantially. As more areas are studied closely, more problems are identified. Much more is known about the epidemiology of safety in the inpatient setting than in outpatient settings. A number of effective strategies for improving inpatient safety have been identified and are increasingly being applied. Some effective strategies are also available for the outpatient setting. Transitions appear to be especially risky. The solutions to improving care often entail the consistent use of systematic techniques such as checklists and often involve leveraging of information technology. Nevertheless, solutions will also include many other domains, such as human factors techniques, team training, and a culture of safety.

■ QUALITY IN HEALTH CARE

Assessment of quality of care has remained somewhat elusive, although the tools for this purpose have increasingly improved. Selection of health care and measurement of its quality are components of a complex process.

Quality Theory Donabedian has suggested that quality of care can be categorized by type of measurement into structure, process, and outcome. *Structure* refers to whether a particular characteristic is applicable in a particular setting—e.g., whether a hospital has a catheterization laboratory or whether a clinic uses an electronic health record. *Process* refers to the way care is delivered; examples of process measures are whether a Pap smear was performed at the recommended interval or whether an aspirin was given to a patient with a suspected myocardial infarction. *Outcome* refers to what happens—e.g., the mortality rate in myocardial infarction. It is important to note that good structure and process do not always result in a good outcome. For instance, a patient may present with a suspected myocardial infarction to an institution with a catheterization laboratory and receive recommended care, including aspirin, but still die because of the infarction.

Quality theory also suggests that overall quality will be improved more in the aggregate if the performance level of all providers is raised rather than if a few poor performers are identified and punished. This view suggests that systems changes are especially likely to be helpful in improving quality, since large numbers of providers may be affected simultaneously.

The theory of *continuous quality improvement* suggests that organizations should be evaluating the care they deliver on an ongoing basis and continually making small changes to improve their individual processes. This approach can be very powerful if embraced over time.

Several specific tools have been developed to help improve process performance. One of the most important is the Plan-Do-Check-Act cycle (Fig. 8-2). This approach can be used for “rapid cycle” improvement of a process—e.g., the time that elapses between a diagnosis of pneumonia and administration of antibiotics to the patient. Some statistical tools, such as control charts, are often used in conjunction to determine whether progress is being made. Because most medical care includes one or many processes, this tool is especially important for improvement.

Factors Relating to Quality Many factors can decrease the level of quality, including stress to providers, high or low levels of production pressure, and poor systems. Stress can have an adverse effect on quality because it can lead providers to omit important steps, as can a high level of production pressure. Low levels of production pressure sometimes can result in worse quality, as providers may be bored or have little experience with a specific problem. Poor systems can have a tremendous impact on quality, and even extremely dedicated providers typically cannot achieve high levels of performance if they are operating within a poor system.

Data about the Current State of Quality A study published by the RAND Corporation in 2006 provided the most complete picture of

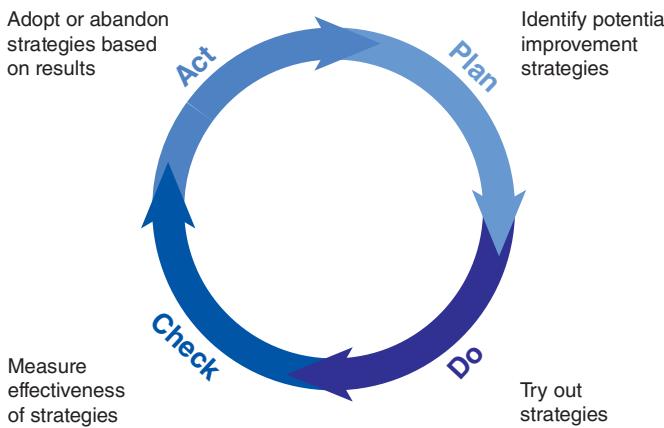


FIGURE 8-2 Plan-Do-Check-Act cycle. This approach can be used to improve a specific process rapidly. First, planning is undertaken, and several potential improvement strategies are identified. Next, these strategies are evaluated in small “tests of change.” “Checking” entails measuring whether the strategies have appeared to make a difference, and “acting” refers to acting on the results.

quality of care delivered in the United States to date. The results were sobering. The authors found that, across a wide range of quality parameters, patients in the United States received only 55% of recommended care overall; there was little variation by subtype, with scores of 54% for preventive care, 54% for acute care, and 56% for care of chronic conditions. The authors concluded that, in broad terms, the chances of getting high-quality care in the United States were little better than those of winning a coin flip.

Work from the Dartmouth Atlas of Health Care evaluating geographic variation in use and quality of care demonstrates that, despite large variations in utilization, there is no positive correlation between the two variables at the regional level. An array of data demonstrate, however, that providers with larger volumes for specific conditions, especially for surgical conditions, do have better outcomes.

Strategies for Improving Quality and Performance Many specific strategies can be used to improve quality at the individual level, including rationing, education, feedback, incentives, and penalties. *Rationing* has been effective in some specific areas, such as persuading physicians to prescribe within a formulary, but it generally has been resisted. *Education* is effective in the short run and is necessary for changing opinions, but its effect decays fairly rapidly with time. *Feedback* on performance can be given at either the group or the individual level. Feedback is most effective if it is individualized and is given in close temporal proximity to the original events. *Incentives* can be effective, and many believe that they will prove to be a key to improving quality, especially if pay-for-performance with sufficient incentives is broadly implemented (see below). *Penalties* produce provider resentment and are rarely used in health care.

Another set of strategies for improving quality involves changing the systems of care. An example would be introducing reminders about which specific actions need to be taken at a visit for a specific patient—a strategy that has been demonstrated to improve performance in certain situations, such as the delivery of preventive services. Another approach that has been effective is the development of “bundles” or groups of quality measures that can be implemented together with a high degree of fidelity. Many hospitals have implemented a bundle for ventilator-associated pneumonia in the intensive care unit that includes five measures (e.g., ensuring that the head of the bed is elevated). These hospitals have been able to improve performance substantially. Another technique is SCAMPs, or Standardized Clinical Assessment and Management Plans. These are care guidelines developed by clinicians who identify key steps in workflow and decisions to help improve the process outcomes.

Perhaps the most pressing need is to improve the quality of care for chronic diseases. The Chronic Care Model has been developed by Wagner and colleagues (Fig. 8-3); it suggests that a combination of

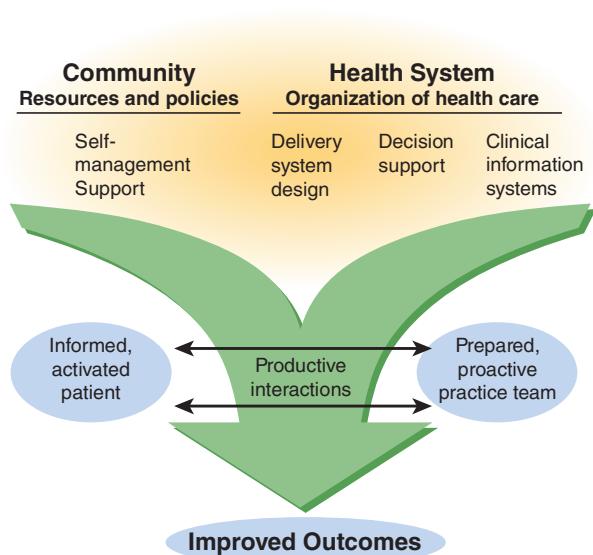


FIGURE 8-3 The Chronic Care Model, which focuses on improving care for chronic diseases, suggests that (1) delivery of high-quality care requires a range of strategies that must closely involve and engage the patient and (2) team care is essential. (From EH Wagner et al: *Eff Clin Pract* 1:2, 1998.)

strategies is necessary (including self-management support, changes in delivery system design, decision support, and information systems) and that these strategies must be delivered by a practice team composed of several providers, not just a physician.

Available evidence about the relative efficacy of strategies in reducing hemoglobin A_{1c} (HbA_{1c}) in outpatient diabetes care supports this general premise. It is especially notable that the outcome was the HbA_{1c} level, as it has generally been much more difficult to improve outcome measures than process measures (such as whether HbA_{1c} was measured). In this meta-analysis, a variety of strategies were effective, but the most effective ones were the use of team changes and the use of a case manager. When cost-effectiveness is considered in addition, it appears likely that an amalgam of strategies will be needed. However, the more expensive strategies, such as the use of case managers, probably will be implemented widely only if pay-for-performance takes hold.

The evidence linking better performance on quality metrics assessing process and outcomes varies greatly by condition. For example, there is strong evidence that performing Pap smears results in better outcomes in patients who develop cervical cancer, but the evidence for many other conditions is far more tenuous.

National State of Quality Measurement In the inpatient setting, quality measurement is now being performed by a very large proportion of hospitals for several conditions, including myocardial infarction, congestive heart failure, pneumonia, and surgical infection prevention; 20 measures are included in all. This is the result of the Hospital Quality Initiative, which represents a collaboration among many entities, including the Hospital Quality Alliance, The Joint Commission, the National Quality Forum, and the Agency for Healthcare Research and Quality. The data are housed at the Centers for Medicare and Medicaid Services, which publicly releases performance data on the measures on a website called *Hospital Compare* (www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/Hospital-QualityInits/HospitalCompare.html). These data are reported voluntarily and are available for a very high proportion of the nation's hospitals. Analyses demonstrate substantial regional variation in quality and important differences among hospitals. Analyses by The Joint Commission for similar indicators reveal that performance on measures by hospitals has improved over time and that, as might be hoped, lower performers have improved more than higher performers.

Public Reporting Overall, public reporting of quality data is becoming increasingly common. There are now commercial websites that have quality-related data for most regions of the United States, and

these data can be accessed for a fee. Similarly, national data for hospitals are available. The evidence to date indicates that patients have not made much use of such data, but that the data have had an important effect on provider and organization behavior. Instead, patients have relied on provider reputation to make choices, partly because little information was available until very recently and the information that was available was not necessarily presented in ways that were easy for patients to access. Problems still exist with quality metrics; many can be “gamed,” and even though providers are now nearly universally using electronic health records (EHRs), most metrics come from claims that include many inaccuracies. More metrics that leverage EHRs are sorely needed. However, many authorities think that, as more information about quality becomes available, it will become increasingly central to patients’ choices about where to access care.

Pay-for-Performance Currently, providers in the United States get paid the same amount for a specific service, regardless of the quality of care delivered. The pay-for-performance theory suggests that, if providers are paid more for higher-quality care, they will invest in strategies that enable them to deliver that care. The current key issues in the pay-for-performance debate relate to (1) how effective it is, (2) what levels of incentives are needed, and (3) what perverse consequences are produced. The evidence on effectiveness is limited, although a number of studies are ongoing. With respect to incentive levels, most quality-based performance incentives have accounted for merely 1–2% of total payment in the United States to date. In the United Kingdom, however, 40% of general practitioners’ salaries have been placed at risk according to performance across a wide array of parameters; this approach has been associated with substantial improvements in reported quality performance, although it is still unclear to what extent this change represents better performance versus better reporting. The potential for perverse consequences exists with any incentive scheme. One problem is that, if incentives are tied to outcomes, there may be a tendency to transfer the sickest patients to other providers and systems. Another concern is that providers will pay too much attention to quality measures with incentives and ignore the rest of the quality picture. The validity of these concerns remains to be determined. Nonetheless, it appears likely that, under health care reform, the use of various pay-for-performance schemes is likely to increase.

CONCLUSIONS

The safety and quality of care in the United States could be improved substantially. A number of available interventions have been shown to improve the safety of care and should be used more widely; others are undergoing evaluation or soon will be. Quality also could be dramatically better, and the science of quality improvement continues to mature. Implementation of value-based approaches such as accountable care that include pay-for-performance related to safety and quality should make it much easier for organizations to justify investments in improving safety and quality parameters, including health information technology. However, many improvements will also require changing the structure of care—e.g., moving to a more team-oriented approach and ensuring that patients are more involved in their own care. Payment reform focusing on value seems very likely to progress and will likely include both positive incentives and penalties related to safety and quality performance. Measures of safety are still relatively immature and could be made much more robust; it would be particularly useful if organizations had measures they could use in routine operations to assess safety at a reasonable cost, and substantial research is addressing this. Although the quality measures available are more robust than those for safety, they still cover a relatively small proportion of the entire domain of quality, and more measures need to be developed. The public and payers are demanding better information about safety and quality as well as better performance in these areas. The clear implication is that these domains will have to be addressed directly by providers.

FURTHER READING

BATES DW et al: Effect of computerized physician order entry and a team intervention on prevention of serious medication errors. *JAMA* 280:1311, 1998.

BATES DW et al: Two decades since to err is human: An assessment of progress and emerging priorities in patient safety. *Health Aff* (Millwood) 37:1736, 2018.

BERWICK DM: Era 3 for medicine and health care. *JAMA* 315:1329, 2016.

BRENNAN TA et al: Incidence of adverse events and negligence in hospitalized patients. Results of the Harvard Medical Practice Study I. *N Engl J Med* 324:370, 1991.

CHERTOW GM et al: Guided medication dosing for inpatients with renal insufficiency. *JAMA* 286:2839, 2001.

INSTITUTE OF MEDICINE: Report: To err is human: Building a safer health system. 1999. <https://www.nap.edu/resource/9728/To-Err-is-Human-1999--report-brief.pdf>.

INSTITUTE OF MEDICINE: Crossing the quality chasm: A new health system for the 21st century. 2001. <https://www.nap.edu/catalog/10027/crossing-the-quality-chasm-a-new-health-system-for-the>.

LANDRIGAN C et al: Effect of reducing interns’ work hours on serious medical errors in intensive care units. *N Engl J Med* 351:1838, 2004.

MCGLYNN EA et al: The quality of health care delivered to adults in the United States. *N Engl J Med* 348:2635, 2003.

PRONOVOST P et al: An intervention to decrease catheter-related blood-stream infections in the ICU. *N Engl J Med* 355:2725, 2006. Erratum in: *N Engl J Med* 356:2660, 2007.

STARMER AJ et al: Rates of medical errors and preventable adverse events among hospitalized children following implementation of a resident handoff bundle. *JAMA* 310:2262, 2013.

9

Diagnosis: Reducing Errors and Improving Quality

Gordon Schiff



Diagnosing patients’ illnesses is the essence of medicine. Patients present to doctors seeking an answer to the question, “What is wrong with me?” Ideally, no clinician would want to treat a patient without knowing the diagnosis or, worse yet, erroneously treat a misdiagnosed illness. From the earliest moments of medical school, the defining quest toward becoming a knowledgeable and proficient physician is learning how to put a diagnostic label on patients’ symptoms and physical findings, and clinicians pride themselves on being “good diagnosticians.” Yet the centuries-old paradigm of mastering a long list of diseases, understanding their pathophysiology, and knowing the cardinal ways they manifest themselves in signs and symptoms, while still of fundamental importance, is being challenged by new insights illuminated by the glaring spotlight of diagnostic errors. Basic internal medicine diseases, such as asthma, pulmonary embolism, congestive heart failure, seizures, strokes, ruptured aneurysms, depression, and cancer, are misdiagnosed at shockingly high rates, often with 20–50% of patients either being mislabeled as having these conditions (false-positive diagnoses) or having their diagnosis missed or delayed (false negatives). How and why do physicians so often get it wrong, and what can we do to both diagnose and treat the problem of delayed diagnosis or misdiagnosis?

Diagnosis is both an ancient art and a modern science. The current science of diagnosis, however, goes far beyond what typically comes to clinicians’ and patients’ minds when they conjure up images of state-of-the-art molecular, genetic, or imaging technologies. Improvements in diagnosis are just as likely to come from other areas, many with origins outside of medicine, as they are from advanced diagnostic testing modalities. These diverse sciences that the field of diagnostic safety has, and must, draw from include systems and human factors

engineering, reliability science, cognitive psychology, decision sciences, forensic science, clinical epidemiology, health services research, decision analysis, network medicine, learning health systems theory, medical sociology, team dynamics and communication, risk assessment and communication, information and knowledge management, and health information technology, especially artificial intelligence and clinical decision support. A clinician reading this chapter is likely to find this list of overlapping and intersecting domains quite daunting. However, rather than feeling overwhelmed, we urge readers to view them as the basic science supports that will ultimately make their lives easier and diagnosis more accurate and timely. Rather than feeling intimidated, clinicians should feel a sense of relief and assurance in understanding that good diagnosis does not rest entirely on their shoulders. Instead, it is a systems property, where an infrastructure and a team, one that especially includes the patient, can in a coordinated way work together to achieve more reliable and optimal diagnosis.

■ EMERGENCE OF DIAGNOSIS ERROR AS AN IMPORTANT PATIENT SAFETY ISSUE

Over the past decade, a series of studies culminating in a landmark report from the U.S. National Academy of Medicine (NAM), *Improving Diagnosis in Health Care*, have shone a spotlight on diagnostic errors. Reports from patient surveys, malpractice claims, and safety organizations, such as the ECRI and the National Patient Safety Foundation (now part of Institute for Healthcare Improvement), have found that diagnostic errors are the leading type of medical error. Although errors in diagnosis are defined in various ways, the NAM Committee defined diagnostic error as “the failure to (a) establish an accurate and timely explanation of the patient’s health problem(s) or (b) communicate that explanation to the patient.” One way to visualize diagnostic errors is through a Venn diagram (Fig. 9-1), which illustrates the fact that many things can go wrong in the diagnostic process (e.g., failure to ask an important history question, physical examination sign overlooked, laboratory specimen erroneously switched between two patients, x-ray not followed up), but this usually does not result in a wrong diagnosis or patient harm. Similarly, a patient can be misdiagnosed but unharmed, without any identifiable error in the care received. Our greatest concern is where these three circles intersect, with conservative estimates suggesting that 40,000–80,000 patients die each year in U.S. hospitals alone from diagnostic errors. The NAM report outlined eight recommendations that are the foundation for this chapter (Table 9-1).

■ NEW WAYS TO THINK ABOUT DIAGNOSIS AND DIAGNOSTIC ERRORS

Medical textbooks have historically given attention to “clinician reasoning” and associated cognitive heuristics and biases. Errors in clinical reasoning can be summarized in three broad groups: (1) hasty judgments, (2) biased judgments, and (3) inaccurate probability estimates. Research from cognitive psychology has identified scores of common mental shortcuts or “heuristics” humans are prone to use in everyday life, many of which are useful for efficient diagnosis but

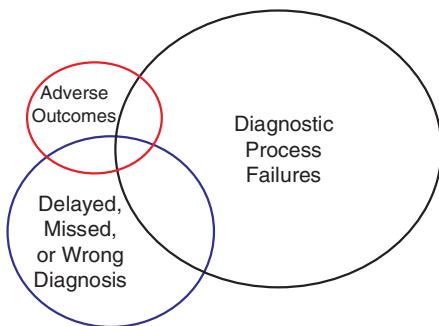


FIGURE 9-1 What is a diagnosis error? (Adapted from GD Schiff et al: Diagnosing diagnosis errors: Lessons from a multi-institutional collaborative project, in *Advances in Patient Safety: from Research to Implementation*. Vol. 2 Concepts and Methodology, Rockville, MD, 2005, pp. 255–278, and GD Schiff, L Leape: Acad Med 87:135, 2012.)

TABLE 9-1 National Academy of Medicine Recommendations for Improving Diagnosis in Health Care

1. Facilitate more effective teamwork in the diagnostic process among health care professionals, patients, and their families.
2. Enhance professional education and training in the diagnostic process in areas such as clinical reasoning; teamwork; communication with patients, families, and other health care professionals; and appropriate use of diagnostic tests.
3. Ensure that health information technologies support patients and health care professionals in the diagnostic process.
4. Develop and deploy approaches to identify, learn from, and reduce diagnostic errors and near misses in clinical practice including providing systematic feedback on diagnostic performance.
5. Establish a work system and culture that supports the diagnostic process and improvements in diagnostic performance.
6. Develop a reporting environment and medical liability system that facilitates improved diagnosis by learning from diagnostic errors and near misses.
7. Design a payment and care delivery environment that supports the diagnostic process.
8. Provide dedicated funding for research on the diagnostic process and diagnostic errors.

can also lead to biases and errors. Table 9-2 lists some of the common cognitive biases that can lead diagnosis astray (this topic is discussed further in Chap. 4).

However, clinicians will also benefit from having a better understanding of diagnosis as a “system” rather than just what takes place in clinicians’ minds. Classic teaching exhorting trainees and practicing physicians to have a broad differential and “high index of suspicion” for various diseases is challenged not only by these unconscious biases but also by limitations of human memory, information shortfalls, constrained encounter time, system process failures, and the myriad nonspecific symptoms that patients bring to clinicians. Many symptoms are self-limited, defy a precise diagnosis or etiology, and do not portend harmful outcomes. Insights from safety and cognitive sciences call for rethinking traditional approaches to diagnosis and suggest new approaches to overcome current limitations (Table 9-3).

■ UNCERTAINTY IN DIAGNOSIS

Given variations and overlap in ways patients present, illnesses evolve, and tests perform, it is often not possible or practical to “make” a definitive diagnosis, particularly in the primary care setting early in the course of a patient’s illness. Clinicians need to harness these uncertainties to both engineer situational awareness of where things

TABLE 9-2 Selected Cognitive Biases Contributing to Diagnostic Errors

1. Premature closure: accepting a diagnosis before it has been fully verified
2. Anchoring: tendency to fixate on a specific symptom or piece of information early in the diagnostic process with subsequent failure to appropriately adjust
3. Confirmation bias: tendency to look for confirming evidence to support one’s diagnostic hypothesis, rather than disconfirming evidence to refute it
4. Search satisficing: tendency to call off a search, satisfied once a piece of data or presumed explanation is found, and not considering/searching for additional findings or diagnoses
5. Availability bias: tendency to give too much weight to diagnoses that come more readily to mind (e.g., recent dramatic case)
6. Base-rate neglect: failing to adequately take into account prevalence of a particular disease (e.g., erroneously interpreting a positive test as indicating disease in a low-prevalence population using a test with 5% false-positive rate)
7. Knowledge deficit (on part of provider, with accompanying lack of awareness)
8. Framing bias: judgement overly influenced by the way the problem was presented (how it was framed in words, settings, situations)
9. Social/demographic/stereotype bias: biases from personal or cultural beliefs about women, minorities, or other patient groups for whom prejudices may distort diagnostic assessment

TABLE 9-3 New Models for Conceptualizing Diagnosis and Diagnosis Improvement

TRADITIONAL WAYS OF THINKING ABOUT DIAGNOSIS AND DIAGNOSTIC ERROR	NEW PARADIGMS/BETTER WAYS TO THINK ABOUT DIAGNOSIS AND IMPROVING DIAGNOSIS
General	
A good diagnostician gets it right the first time, almost all of the time	Diagnosis is an inexact science with inherent uncertainties Goal is to minimize errors and delays via more reliable systems and follow-up
Lore of masterful/skillful academic expert diagnostician who knows/recalls everything; need to look to them if seeking diagnostic excellence	Less reliance on (fallible) human memory Quality diagnosis is based on well-coordinated distributed network/team of people and reliable processes All patients entitled to receive quality diagnosis, regardless of where and from whom they receive care
Diagnosis is the doctor's job	Co-production of diagnosis among clinicians (including lab, radiology, specialists, nurses, social workers) and, especially, the patient and family
Patients often viewed as overly anxious, exaggerating, time-consuming, questioning, with sometimes unreasonable demands and expectations	Patients are key allies in diagnosis; hold key information Need to address understandable/legitimate fears, desires for explanations Leveraging patient questions and questioning of diagnosis to stimulate rethinking the diagnosis where needed
Diagnosis and treatment as separate stages in patient care (i.e., make a diagnosis, then treat)	Prioritizing diagnostic efforts to target treatable conditions More integrated strategies and timing for testing and treatment depending on urgency for treatment
Clinical practices	
Order lots of tests to avoid missing diagnoses	Judicious ordering: targeted, well-organized data and testing Appreciation of test limitations (false positive or negative, incidental findings, overdiagnosis, test risks) and resulting harms
More referrals to avoid missing rarer/specialized diagnoses; concomitant utilization barriers (copays, prior authorization) to minimize overuse	"Pull systems" to lower barriers and make it easier to pose questions, obtain real-time virtual consults Co-management approaches to enable collaborative watch-and-wait conservative strategies where appropriate
Frequent empirical drug trials when uncertain of diagnosis	Conservative use of drugs to avoid confusing clinical picture or labeling patients with diseases they may not have
Physician attention/efforts to ensure disease screening	Automating, delegating clerical functions; teamwork to free up physician cognitive time
Diagnosis errors and challenges	
Diagnostic error viewed as a personal failing Errors classified as either "system" or "cognitive"	Many errors/delays rooted in processes and system design failures Errors multifactorial with interwoven, interacting, and inseparable cognitive and system factors
Errors are infrequent; hit-and-miss ways to learn about errors	Errors are common; systematic proactive follow-up is needed to recognize potential for errors Surveilling of high-risk situations and one's own diagnostic performance and outcomes
Clinicians' reactions: denial, defensive, others to blame, pointing to others also making similar errors	Culture of actively and nondefensively seeking to uncover, dig deep to learn from, and share errors and lessons
Dreading complex, frustrating diagnostic dilemmas	Welcoming/enjoying intellectual/professional challenges Adequate support (time, help, consultations) for more complex patients
Diagnoses as distinct labels, events	Diagnoses can be indistinct, interacting comorbidities, socially constructed, multifactorial, evolving over time, or have overlapping genotype-phenotype expressions
Documentation/communication	
Viewed as time-consuming, mindless, primarily to document for billing code and/or bulwark against malpractice claims	Documentation as useful tool for reflecting, crafting, sharing assessments, differential diagnosis, reflecting about unanswered questions Opportunities for decision support interacting with computer Notes open for patients to read to help understand and critique diagnosis
Say and write as little as possible about uncertainties, lest it be used against you in malpractice allegation	Share uncertainties to maximize communication and engagement with other caregivers, patients
Don't let patient know about errors so they don't become angry, mistrustful, or sue	Patients have right to honest disclosure; often find out about errors anyway (e.g., cancer evolves); anticipate, engage their concerns
Patients advised to call if not better; no news is good news (test results: "We'll call if anything is abnormal.")	Systematic proactive follow-up to close loop on tests and symptoms, to check how patient is doing, monitor outcomes
Global remedies	
Knowing/memorizing more medical knowledge	Knowing more about the patient (including psychosocial, past history, environmental contexts)
Attention to the "objective" data (physical exam, tests) to reliably make diagnoses	Renewed emphasis on history, history-taking, listening Acknowledgement of ubiquitous subjective cognitive biases; efforts to anticipate, recognize, counteract
Exhortations to have "high index of suspicion" of various diagnoses	Less reliance on memory recall of lectures/reading; more just-in-time info look-up Affordances, alerts to red flags engineered into workflow Delineation of "don't miss" diagnoses with design of context-relevant decision support reminders
Ensuring physician is copied on everything, thorough/voluminous notes, widespread reminders/alerts	Biggest problem is no longer lack of access to information, but rather information overload; strategies to organize, minimize
Continuing medical education (CME) courses to expand medical knowledge	Real-time, context-aware reminders of pitfalls, critical differential diagnoses, and key differentiating features. Ready access to medical references, second opinions

(Continued)

TABLE 9-3 New Models for Conceptualizing Diagnosis and Diagnosis Improvement (Continued)

TRADITIONAL WAYS OF THINKING ABOUT DIAGNOSIS AND DIAGNOSTIC ERROR	NEW PARADIGMS/BETTER WAYS TO THINK ABOUT DIAGNOSIS AND IMPROVING DIAGNOSIS
Redundancies, double-checks	Recognition that single, highly reliable systems are often better than multiple halfway solutions. Clear delineation of responsibilities for follow-up tasks
Fear of malpractice suits to motivate physicians to be more careful and practice defensive medicine	Drive out fear, making it safe to learn from and share errors Shared situational awareness of where pitfalls lurk
More accountability, financial incentives, and penalties tied to performance metrics	Clinician engagement in improvement based on trust, collaboration, professionalism, financial neutrality Metric modesty, recognizing many best practices yet to be defined/proven
More rules, requirements; target outliers for better compliance	Standardization with flexibility; learning from deviations
More time with patients	Better time spent with patients: offloading distractions, more efficient history collection/organization, longitudinal continuity, and, where needed, additional time to talk/think/explain during, before, or after visits Easier access for patients to reach or be seen by clinicians when experiencing symptoms
Reflex changes in response to errors	Avoiding “tampering,” which entails understanding/diagnosing difference between “special cause” versus “common cause” (random) variation

Source: Modified from GD Schiff: Quality and Safety in Health Care 2013.

can go wrong and create safety nets to protect patients against harms from delayed diagnosis and misdiagnosis. Terms such as *preliminary diagnosis*, *working diagnosis*, *differential diagnosis*, *deferred diagnosis*, *undiagnosed illness*, *diagnoses with uncertain or multifactorial etiologies*, *intermittent diagnoses*, *multiple/dual diagnoses*, *self-diagnosis*, or at times *contested diagnosis* need to be part of our vocabulary, thinking, and communications with patients to convey that diagnosis is often imprecise. Anxious patients worried about a condition, for example, cancer, COVID-19 infection, or a diagnosis to which a relative or a friend has recently succumbed, come seeking reassurance and may not welcome an uncertain answer. Thus, we have to work with patients, listen to and respect their concerns, and take their symptoms seriously yet modestly acknowledge our limitations. We need to tailor this approach to patients’ differing levels of health literacy, trust in our clinical advice, and experiences with the health system.

DON’T MISS DIAGNOSES AND RED FLAGS

Uncertainty should not be a license for complacency. Particularly for diseases that (1) progress rapidly, (2) require specific treatments that depend on making the correct diagnosis, or (3) have public health or contagion implications, clinicians need to be poised, and systems designed, to consider and, where appropriate, pursue critical “don’t miss” diagnoses. While clinicians are generally aware of more common “don’t miss” diagnoses (e.g., acute myocardial infarction, sepsis), **Table 9-4** illustrates examples of less common diagnoses that warrant similar consideration. Throughout this textbook, readers should orient themselves to recognize such critical diagnoses and think about presentations and syndromes where they may be lurking.

An important related concept is so-called “red flags” or “alarm symptoms.” This construct has its origins in guidelines for back pain but has increasingly been applied to many other problems, such as headache, red eye, swollen joint, or even abdominal pain and chest pain. Examples of widely cited red flags for back pain that should trigger consideration of more serious etiologies include fever, weight loss, history of malignancy or intravenous drug use, or neurologic signs and symptoms. In theory, many presenting syndromes could benefit from identification of such clues to more serious diagnoses. Evidence-based medicine calls for better data on the sensitivity, specificity, yield, and discriminatory ability of various clinical “red flag” clues; yet, few have been rigorously evaluated. Nonetheless, clinicians find them useful as simple ways to reassure themselves and their patients that a common symptom such as back pain or headache is, or is not, likely an indicator of more urgent or serious pathology.

Interwoven with the challenges of not missing critical diagnoses is the problem of overtreatment and overdiagnosis—performing unnecessary and even potentially harmful tests whose benefit does not justify the risks or costs or that may lead to diagnoses that would have never

caused any symptoms or problems. Thoughtful diagnosticians need to weigh carefully this “other side of the coin” of missed diagnosis to avoid such harms and expenses.

DIAGNOSTIC PITFALLS

One of the important ways of learning in medicine is learning from the missteps of those who have walked the path ahead of us. By learning about commonly missed diagnoses and the ways accurate, timely diagnosis went astray, we can avoid making similar mistakes. Anticipating the potential for similar types of errors can both create situational awareness of traps to avoid and contribute to learning from our own personal and collective patterns of mistakes. Several studies have examined common or recurring pitfalls in diagnosis. An example of a common disease-specific diagnostic pitfall in breast cancer diagnosis is ordering a mammogram for a woman with a palpable breast lump and, when the mammogram returns as normal, reassuring her that cancer has been “ruled out” by the negative test. Any mass or lesion palpable

TABLE 9-4 Examples of “Don’t Miss” Diagnoses

INFECTIONS/ INFLAMMATION	CARDIAC/ISCHEMIC/ BLEEDING	METABOLIC/ HEMATOLOGIC/ ENVIRONMENTAL
Spinal epidural abscess	Aortic dissection Leaking/ruptured abdominal aortic aneurysm	Diabetes ketoacidosis Hyperosmolar hyperglycemia
Necrotizing fasciitis	Pericardial tamponade	Myxedema/thyrototoxicosis
Meningitis	Wolff-Parkinson-White Prolonged QT	Addison’s disease
Endocarditis	Pulmonary embolism	B ₁₂ deficiency anemia
Peritonsillar abscess	Tension pneumothorax	von Willebrand’s disease
Tuberculosis-active pulmonary, other	Acute mesenteric ischemia Sigmoid volvulus	Hemochromatosis
COVID-19 infection	Esophageal, bowel perforation	Celiac sprue
Guillain-Barré syndrome	Cerebellar hemorrhage	Carbon monoxide poisoning
Ebola infection	Spinal cord compression	Food poisoning
Temporal arteritis	Testicular, ovarian torsion	Malignant hyperthermia
Rhabdomyolysis	Ectopic pregnancy	Alcohol, benzodiazepine, barbiturate withdrawal
Angioedema	Retroperitoneal hemorrhage	Tumor lysis syndrome Hypo-/hypercalcemia

TABLE 9-5 Generic Types of Diagnostic Pitfalls

PITFALL	EXAMPLES
Disease A mistaken for disease B Diseases often mistaken/misdiagnosed with each other	<ul style="list-style-type: none"> Aortic dissection misdiagnosed as acute myocardial infarction Bipolar disorder misdiagnosed as depression
Misinterpretation of test result(s) False-positive or false-negative results with failure to recognize test limitations	<ul style="list-style-type: none"> Breast lump dismissed after negative mammogram Negative COVID-19 test early or late in course
Failure to recognize atypical presentation, signs, and symptoms	<ul style="list-style-type: none"> Apathetic hyperthyroidism Sepsis in elderly patient who is afibrile or hypothermic
Failure to assess appropriately the urgency of diagnosis Urgency of the clinical situation was not appreciated and/or delays critical diagnoses	<ul style="list-style-type: none"> Compartment syndrome Pericardial tamponade Tension pneumothorax
Perils of intermittent symptoms or misleading evolution Intermittent symptoms dismissed due to normal findings (exam, lab, electrocardiogram) when initially seen	<ul style="list-style-type: none"> "Lucid interval" in traumatic epidural hematoma Paroxysmal arrhythmias Intermittent hydrocephalus (Bruns' syndrome)
Confusion arising from response/masking by empiric treatment	<ul style="list-style-type: none"> Empiric treatment with steroids, proton pump inhibitors, antibiotics, pain medication erroneously masking serious diagnosis
Chronic disease or comorbidity presumed to account for new symptoms Especially in medically complex patients	<ul style="list-style-type: none"> Septic joint signs misattributed to chronic rheumatoid arthritis Mental status change due to infection or medication misattributed to underlying dementia
Rare diagnosis: failure to consider or know	<ul style="list-style-type: none"> Many; fortunately, by definition, rare, but still warrant consideration especially if urgent or treatable
Drug or environmental factor not considered/overlooked Underlying etiology causing/contributing to symptoms, or disease progression not sought, uncovered	<ul style="list-style-type: none"> Ventricular arrhythmia related to QT-prolonging drug Achilles tendon rupture related to quinolone drugs
Failure to appreciate risk factors for particular disease	<ul style="list-style-type: none"> Family history of breast, colorectal cancer not solicited and/or weighed in diagnostic evaluation or screening
Failure to appreciate limitations of physical exam Now with ↑ telemedicine, missing physical exam entirely	<ul style="list-style-type: none"> Overweighing absence of tenderness, swelling in deep vein thrombosis Missing pill-rolling tremor during telemedicine visit

on physical examination probably needs more careful assessment proceeding all the way to invasive biopsy, if necessary. Diagnostic pitfalls can be classified into a number of generic scenarios (**Table 9-5**). We now have large databases that have the potential to track "diagnoses outcomes"—i.e., whether a new diagnosis emerges that suggests an initial diagnosis was incorrect or a diagnosis of a patient's symptoms was suboptimally delayed. This should, in the future, allow us to more rigorously focus on these cases, to identify contributing factors and recurring patterns, and to help point the way for systemwide improvement strategies.

■ DIAGNOSIS SAFETY CULTURE

Just as diagnosing bacterial infections relies on a proper culture medium to grow and identify etiologic organisms, good diagnosis also requires a healthy safety culture that will allow it to grow and flourish. While clinicians may be inclined to view "safety culture" as something too subjective to be important in their quest to make a definitive diagnosis, this view is misguided. Multiple studies have demonstrated adverse consequences resulting from organizational cultures that inhibit openness, learning, and sharing and create a climate where staff

and patients are afraid to speak up when they observe problems or have questions. Most importantly, patients need to be encouraged to question diagnoses and be heard, particularly when they are not responding to treatment as expected or developing symptoms that are either not consistent with the diagnosis or represent possible red flags for other diagnoses or complications.

Studies examining "high-reliability organizations" outside of medicine and "learning health care organizations" have distilled a series of fundamental properties that are correlated with more reliable and safer outcomes. Just as a thermometer or recording of a pulse can suggest how ill a patient is, we now have instruments that can measure safety culture. These safety measurement tools typically are validated staff surveys that assess (1) communication about errors with staff willingness to report mistakes because they do not feel these mistakes are held against them; (2) openness and encouragement to talk about hospital/office problems; (3) existence of a learning culture that seeks to learn from errors and improve based on lessons learned; (4) leadership commitment to safety, prioritizing safety over production speed and the "bottom line" by providing adequate staffing and resources to operate safely; and (5) accountability and transparency for following up safety events and concerns. Each of these generic culture attributes translates into specific implications for diagnostic safety. These include the following:

- Making it "safe" for clinicians to admit and share diagnostic errors
- Proactive identification, ownership, and accountability regarding error-prone diagnostic workflow processes (particularly around test results, referrals, and patient follow-up)
- Leadership making diagnosis improvement a top priority based on recognition that patients and malpractice insurers report that diagnostic errors are the leading patient safety problem
- Mutual trust and respect for challenges that clinicians often face in making diagnoses and caution in applying the lens of hindsight bias in judging what in retrospect might seem like an "obvious" diagnosis that a clinician initially missed

■ HEALTH INFORMATION TECHNOLOGY AND THE FUTURE OF DIAGNOSIS

Clinicians now spend more time interacting with computers than they do interacting with patients. This is especially true for diagnosis and will likely be even more so in the future. Interactions with patients, consultants, and other staff are increasingly mediated through the computer. Key activities, such as collecting patients' history (past and current), interpreting data to make a diagnosis, conveying diagnostic assessments (to others on the team and, increasingly, to the patient via open notes), and tracking diagnostic trajectories as they evolve over time, are now computer based. With the rise of telemedicine, even elements of the physical examination have been rerouted to electronic encounters.

While many complain the computer has "gotten in the way" of good diagnosis, distracting clinicians from quality time listening to patients and miring doctors in reading and writing notes filled with copied/pasted/templated information of questionable currency and accuracy, medicine needs to harness the computer's capabilities to improve diagnosis (**Table 9-6**). Although these basic diagnosis-supporting capabilities should be the foundation of the design of health information technology and everyday workflow, electronic medical records have historically been largely designed around other needs, such as ordering medications and billing and malpractice documentation. They need to be radically redesigned to better support diagnostic processes, as well as save, rather than squander, clinicians' time.

■ DIAGNOSIS OF DIAGNOSIS ERRORS AND SAFETY: PRACTICAL CONCLUSIONS

In practice, there are frequent and meaningful opportunities for improving diagnosis in each of the three NAM-defined areas to make it a) more reliable, b) timely, and c) to improve diagnosis-related communication with patients. Clinicians in training, practicing physicians, nurses, and others should develop the habit of regularly asking

TABLE 9-6 Areas Where Health Information Technology Has Potential to Help Improve Diagnosis and Reduce Errors

FUNCTION	EXAMPLES
Facilitate collection/gathering of information	<ul style="list-style-type: none"> Quickly access past history from prior care at same and outside institutions Electronic collection of history of present illness, review of systems, and social determinant risks in advance of visits
Enhanced information entry, organization, and display	<ul style="list-style-type: none"> Visually enhanced flowsheets showing trends, relationships to treatment Reorganized notes to facilitate summarization and simplification and prevent items from getting lost
Generating differential diagnosis	<ul style="list-style-type: none"> Automated creation of lists of diagnoses to consider based on patient's symptoms, demographics, risks
Weighing diagnoses likelihoods	<ul style="list-style-type: none"> Tools to assist in calculation of posttest (Bayesian) probabilities
Aids for formulating diagnostic plan, intelligent test ordering	<ul style="list-style-type: none"> Entering a diagnostic consideration (e.g., celiac disease, pheochromocytoma) and computer suggests most appropriate diagnostic test(s) and how to order
Access to diagnostic reference information	<ul style="list-style-type: none"> Info-buttons instantly linking symptom or diagnosis relevant questions to Harrison's, Up-to-Date chapters, references
Ensuring more reliable follow-up	<ul style="list-style-type: none"> Hardwiring "closed loops" to ensure abnormal labs, missed referrals, worrisome symptoms are tracked and followed up
Support screening for early detection	<ul style="list-style-type: none"> Collaborative tools that patients, clinicians, and offices can use to know when due, order and track screening based on individualized demographics, risk factors, prior tests
Collaborative diagnosis; access to specialist	<ul style="list-style-type: none"> Real-time posing/answering of questions Electronic consults; virtual co-management
Facilitating feedback on diagnoses	<ul style="list-style-type: none"> Feeding back new diagnoses (from downstream providers, patients) that emerge suggesting potential misdiagnosis/errors to clinicians, ERs who saw patient previously

Abbreviation: ERs, emergency rooms.

Source: Modified from G Schiff, DW Bates: N Engl J Med 362:1066, 2010, and R El-Karab et al: BMJ Qual Saf Suppl 2:i40, 2013.

themselves three questions about individual patients in their care, and another three questions regarding the systems in which they work. For each patient being assessed, clinicians should ask:

1. What else might this be? (forcing a differential diagnosis to be made)
2. What doesn't fit? (making sure unexplained abnormal findings are not dismissed)
3. What critical diagnoses are important not to miss? (injecting consideration of "don't miss" diagnoses, red flags, and known pitfalls)

and to diagnose safely, each practitioner must recognize that he or she is working within a larger system. Questions to be asking continually, ensuring we are maximizing reliability and timeliness and minimizing potential for errors, include:

1. Do we have reliable "closed loop" systems to provide reliable, ideally automated tracking and following up of patients' symptoms, abnormal laboratory or imaging findings, and critical referrals that we order?
2. What is the culture-of-safety climate in our organization, office, or clinic?
3. How does the electronic (or even paper) medical record as currently implemented help versus impair efficient, timely, accurate, and fail-safe diagnosis, and how can it be improved?

To take these questions to the next stage, an international movement dedicated to studying and improving diagnosis has emerged. These efforts include annual conferences of clinicians, researchers, and patients; the formation of the Society for Improving Diagnosis in Medicine (SIDM); and convening of a broad coalition of organizations,

including the American Board of Internal Medicine (ABIM), the American College of Physicians (ACP), and the Society of Hospital Medicine (SHM), committed to increasing awareness and action. Ultimately, collectively tackling the challenges of improving the quality of diagnosis will transform the way clinicians and patients work together to co-produce better diagnoses.

FURTHER READING

GANDHI TK, SINGH H: Reducing the risk of diagnostic error in the COVID-19 era. J Hosp Med 15:363, 2020.

GRABER ML et al: The impact of electronic health records on diagnosis. Diagnosis (Berl) 4:211, 2017.

NATIONAL ACADEMIES OF SCIENCES, ENGINEERING, AND MEDICINE. 2015.

Improving Diagnosis in Health Care. <https://doi.org/10.17226/21794>.

Adapted and reproduced with permission from the National Academy of Sciences, Courtesy of the National Academies Press, Washington, DC.

NEWMAN-TOKER DE et al: Serious misdiagnosis-related harms in malpractice claims: The "big three"—Vascular events, infections, and cancers. Diagnosis (Berl) 6:227, 2019.

SCHIFF GD et al: Diagnosing diagnosis errors: Lessons from a multi-institutional collaborative project, in *Advances in Patient Safety: From Research to Implementation*. Vol 2: Concepts and Methodology. Rockville, MD, Agency for Healthcare Research and Quality, 2005.

SCHIFF GD et al: Ten principles for more conservative, care-full diagnosis. Ann Intern Med 169:643, 2018.

10

Racial and Ethnic Disparities in Health Care

Lenny López, Joseph R. Betancourt



Over the course of its history, the United States has experienced dramatic improvements in overall health and life expectancy, largely as a result of initiatives in public health, health promotion, disease prevention, and chronic care management. Our ability to prevent, detect, and treat diseases in their early stages has allowed us to target and reduce rates of morbidity and mortality. Despite interventions that have improved the overall health of the majority of Americans, racial and ethnic minorities (blacks, Hispanics/Latinos, Native Americans/Alaskan Natives, Asian/Pacific Islanders) have benefited less from these advances than whites and have suffered poorer health outcomes from many major diseases, including cardiovascular disease, cancer, and diabetes. These disparities highlight the importance of recognizing and addressing the multiple factors that impact health outcomes, including structural racism, *social determinants of health* (SDOH), access to care, and health care quality. On this last point, research has revealed that minorities may receive less care and lower-quality care than whites, even when confounders such as stage of presentation, comorbidities, and health insurance are controlled. These differences in quality are called *racial and ethnic disparities in health care*. These health care disparities have taken on greater importance with the significant transformation of the U.S. health care system and value-based purchasing. The shift toward creating financial incentives and disincentives to achieve quality goals makes focusing on those who receive lower-quality care more important than ever before. This chapter will provide an overview of racial and ethnic disparities in health and health care, identify root causes, and provide key recommendations to address these disparities at both the clinical and health system levels.

NATURE AND EXTENT OF DISPARITIES

Life expectancy at birth is an important measure of the health of a nation's population. Although the overall life expectancy in the United States has been increasing since 1900, differences due to

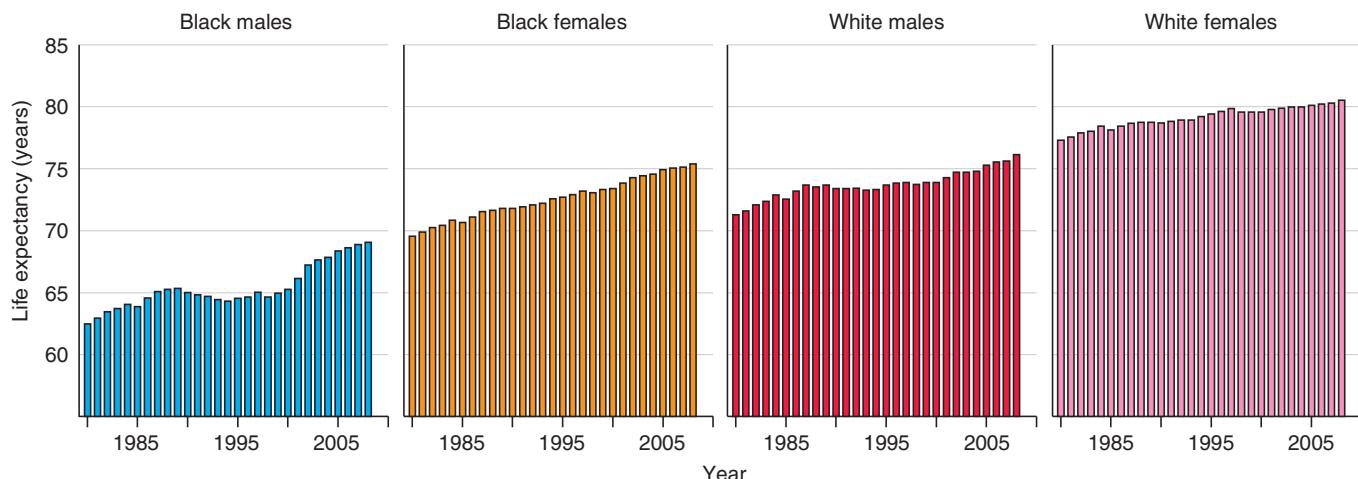


FIGURE 10-1 Life expectancy at birth among black and white males and females in the United States, 1975–2003. (Adapted from S Harper, J Lynch, S Burris, GD Smith: Trends in the black-white life expectancy gap in the United States, 1983–2003. *JAMA* 297:1224, 2007.)

race/ethnicity, education, and socioeconomic status have persisted. For example, at every level of education and income, African Americans have lower life expectancy at age 25 than whites and Hispanics/Latinos. Blacks with a college degree or more education have lower life expectancy than whites and Hispanics who graduated from high school. Blacks have had lower life expectancy compared to whites for as long as data have been collected. From 1975 to 2003, the largest difference in life expectancy between blacks and whites was substantial (6.3 years for males and 4.5 years for females) (Fig. 10-1). The gap in life expectancy between the black and white populations decreased by 2.3 years between 1999 and 2013 from 5.9 to 3.6 years (4.4 years for males and 3.0 years for women) (Fig. 10-2).

The life expectancy gap is augmented by worse health and higher disease burden. Cardiovascular-related diseases remain the leading cause of black-white differences in life expectancy. If all cardiovascular causes and diabetes are considered together, they account for 35% and 52% of the gap for males and females, respectively. Finally, place matters for health. Analysis of data from 2010 to 2015 demonstrate large geographic life expectancy gap variation at the census tract level (Fig. 10-3). Socioeconomic and race/ethnicity factors, behavioral and metabolic risk factors (prevalence of obesity, leisure-time physical inactivity, cigarette smoking, hypertension, diabetes), and health care factors (percentage of the population younger than 65 years who are insured, primary care access and quality, number of physicians per capita) explained 60%, 74%, and 27% of county-level variation in life expectancy, respectively. Combined, these factors explained 74% of this variation. Most of the association between socioeconomic and race/ethnicity factors and life expectancy was mediated through behavioral and metabolic risk factors.

In addition to racial and ethnic disparities in *health*, there are racial and ethnic disparities in the *quality of care* for persons with access to

the health care system. Seminal studies over several decades have consistently documented disparities in health care. For instance, studies have documented disparities in the treatment of pneumonia and congestive heart failure, with blacks receiving less optimal care than whites when hospitalized for these conditions. Moreover, blacks with end-stage renal disease are referred less often to the transplant list than are their white counterparts (Fig. 10-4). Disparities have been found, for example, in the use of cardiac diagnostic and therapeutic procedures (with blacks being referred less often than whites for cardiac catheterization and bypass grafting), prescription of analgesia for pain control (with blacks and Hispanics/Latinos receiving less pain medication than whites for long-bone fractures and cancer), and surgical treatment of lung cancer (with blacks receiving less curative surgery than whites for non-small-cell lung cancer). Again, many of these disparities have occurred even when variations in factors such as insurance status, income, age, comorbid conditions, and symptom expression are taken into account. Finally, disparities in the quality of care provided at the sites where minorities tend to receive care have been shown to be an important additional contributor to overall disparities.

The 2019 National Healthcare Quality and Disparities Report, released by the Agency for Healthcare Research and Quality, tracks about 250 health care process, outcome, and access measures, across many diseases and settings. This annual report is particularly important because most studies of disparities have not been longitudinally repeated with the same methodology to document trends and changes in disparities over time. This report found that some disparities were getting smaller from 2000 through 2016–2018, but disparities persisted and some even worsened, especially for poor and uninsured populations. For about 40% of quality measures, blacks (82 of 202 measures) and American Indians and Alaska Natives (47 of 116 measures) received worse care than whites. For more than one-third of quality measures, Hispanics (61 of 177 measures) received worse care than whites. Asians and Native Hawaiians/Pacific Islanders received worse care than whites for about 30% of quality measures, but Asians also received better care for about 30% of quality measures (Fig. 10-5). Of note, for those quality measures that demonstrated disparities at baseline, >90% of these measures showed no improvement since 2000 (Fig. 10-6).

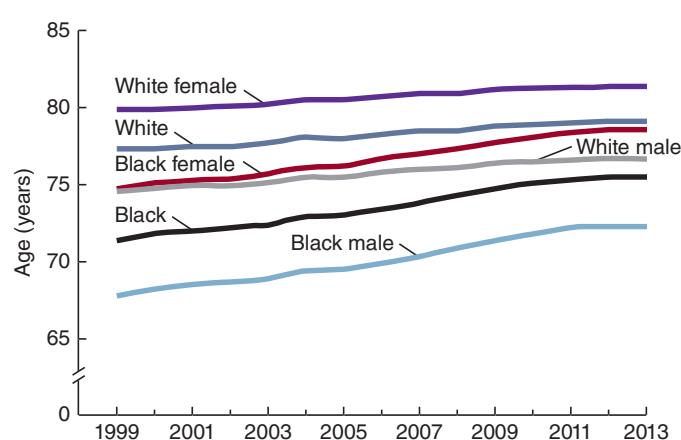


FIGURE 10-2 Life expectancy, by race and sex: United States, 1999–2013. (From KD Kochanek et al: NCHS Data Brief 218:1, 2015.)

■ ROOT CAUSES OF DISPARITIES

Race, Racism, and Health Race and racism are core elements of any explanatory model on racial and ethnic disparities in health and health care. Our nation's history of slavery, segregation, separate but "equal" health care, and medical experimentation, among a myriad of other ways in which racism has manifested in the United States, has played a key role in the existence and persistence of these disparities. It is now well accepted that race is a social category without biologic foundation and a product of historical racism. Nevertheless, it is clear that racism has a biologic impact as a form of psychosocial stress. It is now well established

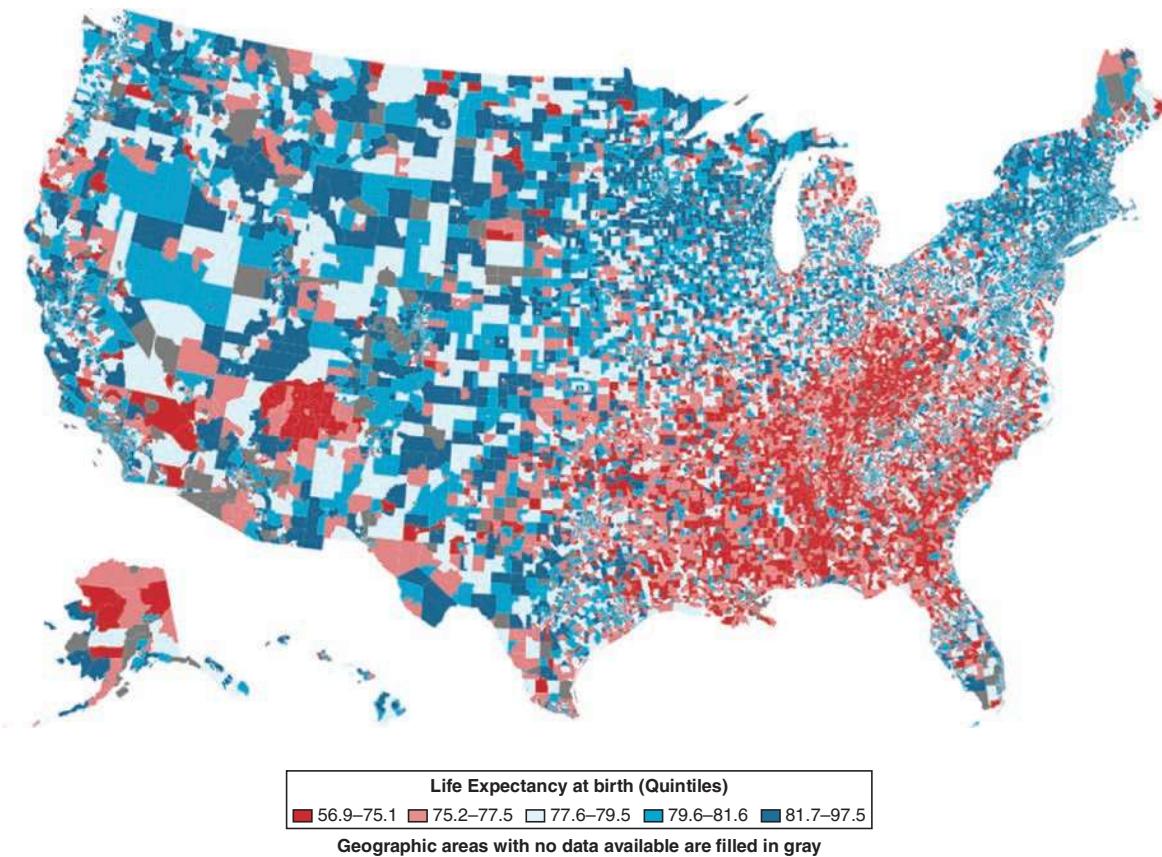


FIGURE 10-3 Life expectancy at birth for U.S. census tracts, 2010-2015. (A New View of Life Expectancy, Surveillance and Data - Blogs and Stories, Centers for Disease Control and Prevention. Retrieved from <https://www.cdc.gov/surveillance/blogs-stories/life-expectancy.html>.)

that psychosocial stress negatively impacts health through psychophysiological reactivity causing hyperstimulation of the sympathetic-adrenal-medullary system and the hypothalamic-pituitary-adrenal axis, leading to vascular inflammation, endothelial dysfunction, and neurohormonal dysregulation causing an acceleration of cardiovascular disease. Behavioral changes occurring as adaptations or coping responses to stressors such as increased smoking, decreased exercise and sleep, and

poorer adherence to medical regimens provide an additional important pathway through which stressors influence disease risk. This accelerated disease risk, aging, and premature death has been termed the *weathering effect*.

While most empiric research focuses on interpersonal racial/ethnic discrimination, structural racism (sometimes called institutional racism) provides a more holistic framework. Structural racism refers to the totality of ways that a society fosters, sustains, and reinforces discrimination through sociopolitical, legal, economic, and health structures that determine differential access to risks, opportunities, and resources that drive health and health care disparities. Structural racism explains how racism's structure and ideology can persist in governmental and institutional policies in the absence of individual actors who are explicitly racially prejudiced. For example, the history of residential segregation has had lasting negative effects generationally on equal access for racial/ethnic minorities to employment, banking, earnings, high-quality education, and health care. Policies that do not address root structural causes will not address health and health care inequities.

With the promise of individualizing clinical decisions, the use of race in clinical and risk assessment algorithms has long been a part of modern medicine. The evidence is now clear that race is not a reliable proxy for genetic difference and that race adjustment has the potential to create inadvertent disparities in health care. One clinical example is from nephrology. Blacks have higher rates of end-stage kidney disease and death due to kidney failure than the overall population. The most widely used cohort-derived equation to estimate glomerular filtration rate (GFR), the Chronic Kidney Disease Epidemiology Collaboration (CKD-EPI) equation, has the limitation that it produces 80–90% estimated GFR (eGFR) values that are within $\pm 30\%$ of a patient's measured GFR. In addition, this equation uses a black race-related factor, which increases eGFR for any given serum creatinine by 15.9% compared to a nonblack patient with the same age, sex, and serum creatinine. The increase in eGFR is likely to disadvantage blacks for early referral to a nephrologist, early treatment of advanced chronic kidney disease, and

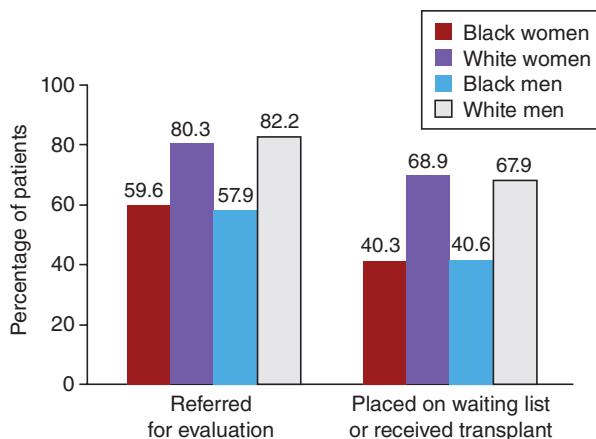


FIGURE 10-4 Referral for evaluation at a transplantation center or placement on a waiting list/receipt of a renal transplant within 18 months after the start of dialysis among patients who wanted a transplant, according to race and sex. The reference population consisted of 239 black women, 280 white women, 271 black men, and 271 white men. Racial differences were statistically significant among both the women and the men ($p < .0001$ for each comparison). (From JZ Ayanian, PD Cleary, JS Weissman, AM Epstein: The effect of patients' preferences on racial differences in access to renal transplantation. *N Engl J Med* 341:1661,1999. Copyright © 1999 Massachusetts Medical Society. Reprinted with permission from Massachusetts Medical Society.)

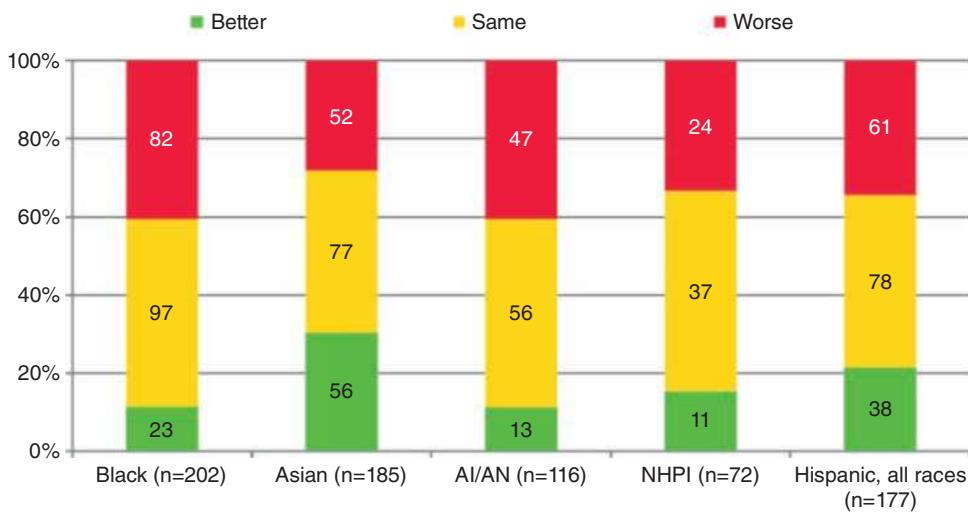


FIGURE 10-5 Number and percentage of quality measures for which members of selected groups experienced better, same, or worse quality of care compared with reference group (white) for the most recent data year, 2014, 2016, 2017, or 2018. AI/AN, American Indian or Alaska Native; NHPI, Native Hawaiian/Pacific Islander (From 2019 National Healthcare Quality and Disparities Report. Rockville, MD: Agency for Healthcare Research and Quality; December 2020. AHRQ Pub. No. 20(21)-0045-EF.)

kidney transplantation. It is also not clear how to apply the race factor when the patient's race is unknown and/or ambiguous, as in those who are multiracial. This disparity-inducing scenario could be avoided through the use of cystatin C-based eGFR estimation, which has been demonstrated to be more accurate than the CKD-EPI equation and for which race is not required in estimation.

The application of artificial intelligence (AI) analytics to large amounts of clinical electronic data—big data—holds the promise to better understand health care costs, utilization, resource allocation, and population health monitoring. Machine learning models can identify the statistical patterns in large amounts of historically collected data. These data naturally contain the patterning of preexisting health care disparities created by socially and historically structured inequities. This biased patterning can lead to incorrect predictions, withholding of resources, and worse outcomes for vulnerable populations. Recently, analysis of a commercial, national, proprietary prediction algorithm, affecting millions of patients, exhibited racial bias. Historical cost data were used to predict clinical risk and allocate additional clinical services for high-cost patients. Algorithmic bias arose because black patients historically have less access to health care and thus less money is spent on their care compared to white patients. Thus, blacks, who tended to be sicker than white patients, received lower clinical risk scores and thus were less likely to receive additional clinical services. The observed allocation bias was remedied using direct measures of illness and illness severity. Thus, machine learning algorithms are not inherently free of bias and should be assessed for accuracy and fairness.

In summary, there are many ways in which racism has contributed and does and will continue to contribute to racial and ethnic disparities in health and health care.

SOCIAL DETERMINANTS OF HEALTH

Minority Americans have poorer health outcomes than whites from preventable and treatable conditions such as cardiovascular disease, diabetes, asthma, cancer, and HIV/AIDS. Multiple factors contribute to these racial and ethnic disparities in health. The landmark

National Academy of Medicine (formerly, the Institute of Medicine [IOM]) report, *Unequal Treatment: Confronting Racial and Ethnic Disparities in Health Care*, published in 2002, summarized the scientific evidence on health disparities and provided an important framework for conceptualizing and defining racial/ethnic disparities. Since the *Unequal Treatment* report, there has been a growing empiric evidence base on how racism and the SDOH, often working in synergy, create and sustain disparities. Mechanistically, the biopsychosocial model brings together the social and physical characteristics of the environment with individual physical and psychological attributes. These environmental and individual characteristics, in turn, influence health behaviors and stress-related physiologic pathways that directly impact health. The National Institute on Minority Health and Health Disparities SDOH model builds on prior models and adds

the time element across the life course of the individual in recognition of the long-lasting health effects of socioeconomic exposures (Fig. 10-7). The resulting matrix has the domains of influence of health (biological, behavioral, physical and built environment, sociocultural environment, health care system) along the y-axis and the levels of influence on health (individual, interpersonal, community, societal) along the x-axis. Cells are not mutually exclusive, and examples of factors within each cell are illustrative and not comprehensive. This framework emphasizes the complex multidomain etiologies of disparities across the factors in the conceptual matrix thus highlighting the limitation of individual-level focused research and policy.

In addition to race and racism, *Unequal Treatment* identified a set of root causes that included health system, provider-level, and patient-level factors.

Health System Factors • HEALTH SYSTEM COMPLEXITY Even among persons who are insured and educated and who have a high degree of health literacy, navigating the U.S. health care system can be complicated and confusing. Some individuals may be at higher risk for receiving substandard care because of their difficulty navigating the system's complexities. These individuals may include those from cultures

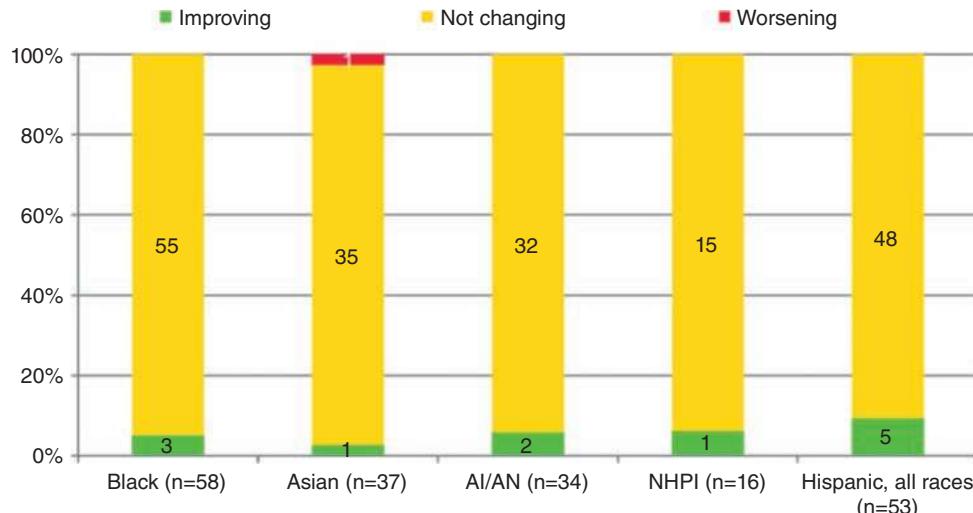


FIGURE 10-6 Number and percentage of quality measures with disparity at baseline for which disparities related to race and ethnicity were improving, not changing, or worsening over time, 2000 through 2014, 2015, 2016, 2017, or 2018. AI/AN, American Indian or Alaska Native; NHPI, Native Hawaiian/Pacific Islander. (From 2019 National Healthcare Quality and Disparities Report. Rockville, MD: Agency for Healthcare Research and Quality; December 2020. AHRQ Pub. No. 20(21)-0045-EF.)

		Levels of Influence*			
		Individual	Interpersonal	Community	Societal
Domains of Influence (Over the Lifecourse)	Biological	Biological Vulnerability and Mechanisms	Caregiver–Child Interaction Family Microbiome	Community Illness Exposure Herd Immunity	Sanitation Immunization Pathogen Exposure
	Behavioral	Health Behaviors Coping Strategies	Family Functioning School/Work Functioning	Community Functioning	Policies and Laws
	Physical/Built Environment	Personal Environment	Household Environment School/Work Environment	Community Environment Community Resources	Societal Structure
	Sociocultural Environment	Sociodemographics Limited English Cultural Identity Response to Discrimination	Social Networks Family/Peer Norms Interpersonal Discrimination	Community Norms Local Structural Discrimination	Social Norms Societal Structural Discrimination
	Health Care System	Insurance Coverage Health Literacy Treatment Preferences	Patient–Clinician Relationship Medical Decision-Making	Availability of Services Safety Net Services	Quality of Care Health Care Policies
	Health Outcomes	 Individual Health	 Family/ Organizational Health	 Community Health	 Population Health

FIGURE 10-7 National Institute on Minority Health and Health Disparities social determinants research framework. *Health disparity populations: race/ethnicity, low socioeconomic status, rural, sexual and gender minority. Other fundamental characteristics: sex and gender, disability, geographic region. (From National Institute on Minority Health and Health Disparities. NIMHD Research Framework. 2017. Retrieved from <https://www.nimhd.nih.gov/about/overview/research-framework.html>.)

unfamiliar with the Western model of health care delivery, those with limited English proficiency, those with low health literacy, and those who are mistrustful of the health care system. These individuals may have difficulty knowing how and where to go for a referral to a specialist; how to prepare for a procedure such as a colonoscopy; or how to follow up on an abnormal test result such as a mammogram. Since people of color in the United States tend to be overrepresented among the groups listed above, the inherent complexity of navigating the health care system has been seen as a root cause for racial/ethnic disparities in health care.

OTHER HEALTH SYSTEM FACTORS Racial/ethnic disparities are due not only to differences in care provided within hospitals but also to where and from whom minorities receive their care; i.e., certain specific providers, geographic regions, or hospitals are lower-performing on certain aspects of quality. For example, one study showed that 25% of hospitals cared for 90% of black Medicare patients in the United States and that these hospitals tended to have lower performance scores on certain quality measures than other hospitals. That said, health systems generally are not well prepared to measure, report, and intervene to reduce disparities in care. Few hospitals or health plans stratify their quality data by race/ethnicity or language to measure disparities, and even fewer use data of this type to develop disparity-targeted interventions. Similarly, despite regulations concerning the need for professional interpreters, research demonstrates that many health care organizations and providers fail to routinely provide this service for patients with limited English proficiency. Despite the link between limited English proficiency and health care quality and safety, few providers or institutions monitor performance for patients in these areas.

Provider-Level Factors • PROVIDER-PATIENT COMMUNICATION Significant evidence highlights the impact of sociocultural factors, race, ethnicity, and limited English proficiency on health and clinical care. Health care professionals frequently care for diverse populations with varied perspectives, values, beliefs, and behaviors regarding health and well-being. The differences include variations in the recognition of symptoms, thresholds for seeking care, comprehension of management strategies, expectations of care (including preferences for or against diagnostic and therapeutic procedures), and adherence to preventive

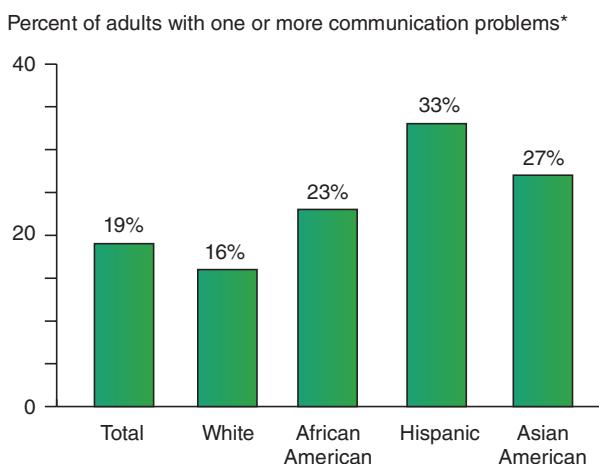
measures and medications. In addition, sociocultural differences between patient and provider influence communication and clinical decision-making and are especially pertinent: evidence clearly links provider–patient communication to improved patient satisfaction, regimen adherence, and better health outcomes (Fig. 10-8). Thus, when sociocultural differences between patient and provider are not appreciated, explored, understood, or communicated effectively during the medical encounter, patient dissatisfaction, poor adherence, poorer health outcomes, and racial/ethnic disparities in care may result.

A survey of 6722 Americans ≥18 years of age is particularly relevant to this important link between provider–patient communication and health outcomes. Whites, African Americans, Hispanics/Latinos, and Asian Americans who had made a medical visit in the past 2 years were asked whether they had trouble understanding their doctors; whether they felt the doctors did not listen; and whether they had medical questions they were afraid to ask. The survey found that 19% of all patients experienced one or more of these problems, yet whites experienced them 16% of the time as opposed to 23% of the time for African Americans, 33% for Hispanics/Latinos, and 27% for Asian Americans (Fig. 10-9).

How do we link communication to outcomes?



FIGURE 10-8 The link between effective communication and patient satisfaction, adherence, and health outcomes. (Institute of Medicine. 2003. *Unequal Treatment: Confronting Racial and Ethnic Disparities in Health Care*. <https://doi.org/10.17226/12875>. Adapted and reproduced with permission from the National Academy of Sciences, Courtesy of the National Academies Press, Washington, D.C.)



Base: Adults with health care visit in past two years

*Problems include understanding doctor, feeling doctor listened, had questions but did not ask.

FIGURE 10-9 Communication difficulties with physicians, by race/ethnicity. The reference population consisted of 6722 Americans ≥ 18 years of age who had made a medical visit in the previous 2 years and were asked whether they had had trouble understanding their doctors, whether they felt that the doctors had not listened, and whether they had had medical questions they were afraid to ask. (Reproduced with permission from the Commonwealth Fund Health Care Quality Survey, 2001.)

In addition, in the setting of even a minimal language barrier, provider–patient communication without an interpreter is recognized as a major challenge to effective health care delivery. These communication barriers for patients with limited English proficiency lead to frequent misunderstanding of diagnosis, treatment, and follow-up plans; inappropriate use of medications; lack of informed consent for surgical procedures; high rates of adverse events with more serious clinical consequences; and a lower-quality health care experience than is provided to patients who speak fluent English. Physicians who have access to trained interpreters report a significantly higher quality of patient–physician communication than physicians who use other methods. Communication issues related to discordant language disproportionately affect minorities and likely contribute to racial/ethnic disparities in health care.

CLINICAL DECISION-MAKING Theory and research suggest that variations in clinical decision-making may contribute to racial and ethnic disparities in health care. Two factors are central to this process: clinical uncertainty and stereotyping.

First, a doctor's decision-making process is nested in *clinical uncertainty*. Doctors depend on inferences about severity based on what they understand about illness and the information obtained from the patient. A doctor caring for a patient whose symptoms he or she has difficulty understanding and whose “signals”—the set of clues and indications that physicians rely on to make clinical decisions—are hard to read may make a decision different from the one that would be made for another patient who presents with exactly the same clinical condition. Given that the expression of symptoms may differ among cultural and racial groups, doctors—the overwhelming majority of whom are white—may understand symptoms best when expressed by patients of their own racial/ethnic groups. The consequence is that white patients may be treated differently from minority patients. Differences in clinical decisions can arise from this mechanism even when the doctor has the same regard for each patient (i.e., is not prejudiced).

Second, the literature on social cognitive theory highlights how natural tendencies to stereotype may influence clinical decision-making. *Stereotyping* can be defined as the way in which people use social categories (e.g., race, gender, age) in acquiring, processing, and recalling information about others. Faced with enormous information loads and the need to make many decisions, people often subconsciously simplify the decision-making process and lessen cognitive effort by using “categories” or “stereotypes” that bundle information into groups or

types that can be processed more quickly. Although functional, stereotyping can be systematically biased, as people are automatically classified into social categories based on dimensions such as *race*, *gender*, and *age*. Many people may not be aware of their attitudes, may not consciously endorse specific stereotypes, and paradoxically may consider themselves egalitarian and not prejudiced.

Stereotypes may be strongly influenced by the messages presented consciously and unconsciously in society. For instance, if the media and our social/professional contacts tend to present images of minorities as being less educated, more violent, and nonadherent to health care recommendations, these impressions may generate stereotypes that unnaturally and unjustly impact clinical decision-making. As signs of racism, classism, gender bias, and ageism are experienced (consciously or unconsciously) in our society, stereotypes may be created that impact the way doctors manage patients from these groups. On the basis of training or practice location, doctors may develop certain perceptions about race/ethnicity, culture, and class that may evolve into stereotypes. For example, many medical students and residents are trained—and minorities cared for—in academic health centers or public hospitals located in socioeconomically disadvantaged areas. As a result, doctors may begin to equate certain races and ethnicities with specific health beliefs and behaviors (e.g., “these patients” engage in risky behaviors, “those patients” tend to be noncompliant) that are more associated with the social environment (e.g., poverty) than with a patient's racial/ethnic background or cultural traditions. This “conditioning” phenomenon may also be operative if doctors are faced with certain racial/ethnic patient groups who frequently do not choose aggressive forms of diagnostic or therapeutic intervention. The result over time may be that doctors begin to believe that “these patients” do not like invasive procedures; thus, they may not offer these procedures as options. A wide range of studies have documented the potential for provider biases to contribute to racial/ethnic disparities in health care. For example, one study measured physicians' unconscious (or implicit) biases and showed that these were related to differences in decisions to provide thrombolysis for a hypothetical black or white patient with a myocardial infarction.

It is important to differentiate stereotyping from prejudice and discrimination. *Prejudice* is a conscious prejudgment of individuals that may lead to disparate treatment, and *discrimination* is conscious and intentional disparate treatment. All individuals *stereotype* subconsciously, yet, if left unquestioned, these subconscious assumptions may lead to lower-quality care for certain groups because of differences in clinical decision-making or differences in communication and patient-centeredness. For example, one study tested physicians' unconscious racial/ethnic biases and showed that patients perceived more biased physicians as being less patient-centered in their communication. What is particularly salient is that stereotypes tend to be activated most in environments where the individual is stressed, multitasking, and under time pressure—the hallmarks of the clinical encounter. In fact, in a survey of close to 16,000 physicians, 42% admitted that bias—including by race and ethnicity—impacted their clinical decision-making. Interestingly, emergency medicine physicians, who work in environments of stress, time pressure, risk, and where they are multitasking, topped the list by discipline at 62%.

Patient-Level Factors Lack of trust has become a major concern for many health care institutions today. For example, an IOM report, *To Err Is Human: Building a Safer Health System*, documented alarming rates of medical errors that made patients feel vulnerable and less trusting of the U.S. health care system. The increased media and academic attention to problems related to quality of care (and of disparities themselves) has clearly diminished trust in doctors and nurses.

Trust is a crucial element in the therapeutic alliance between patient and health care provider. It facilitates open communication and is directly correlated with adherence to the physician's recommendations and the patient's satisfaction. In other words, patients who mistrust their health care providers are less satisfied with the care they receive, and mistrust of the health care system greatly affects patients' use of services. Mistrust can also result in inconsistent care, “doctor-shopping.”

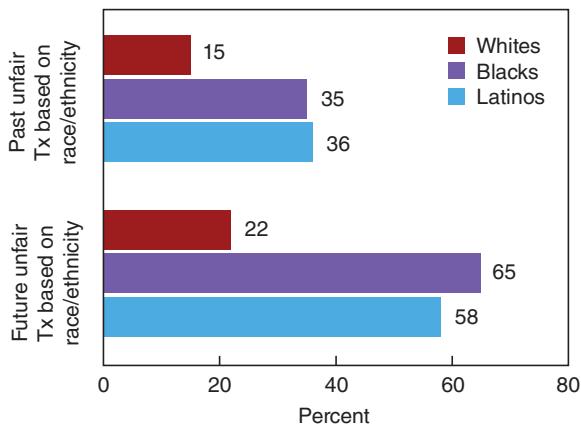


FIGURE 10-10 Patient perspectives regarding unfair treatment (Tx) based on race/ethnicity. The reference population consisted of 3884 individuals surveyed about how fairly they had been treated in the health care system in the past and how fairly they felt they would be treated in the future on the basis of their race/ethnicity. (From *Race, Ethnicity & Medical Care: A Survey of Public Perceptions and Experiences*. Kaiser Family Foundation, 2005.)

self-medication, and an increased demand by patients for referrals and diagnostic tests.

On the basis of historic factors such as discrimination, segregation, and medical experimentation, blacks may be especially mistrustful of providers. The exploitation of blacks by the U.S. Public Health Service during the Tuskegee syphilis study from 1932 to 1972 left a legacy of mistrust that persists even today among this population. Other populations, including Native Americans/Alaskan Natives, Hispanics/Latinos, and Asian Americans, also harbor significant mistrust of the health care system. A national survey conducted by the Kaiser Family Foundation found that there is significant mistrust for the health care system among minority populations. Of the 3884 individuals surveyed, 36% of Hispanics and 35% of blacks (compared to 15% of whites) felt they were treated unfairly in the health care system in the past based on their race and ethnicity. Perhaps even more alarming—65% of blacks and 58% of Hispanics (compared to 22% of whites) were afraid of being treated unfairly in the future based on their race/ethnicity (Fig. 10-10).

This mistrust may contribute to wariness in accepting or following recommendations, undergoing invasive procedures, or participating in clinical research, and these choices, in turn, may lead to misunderstanding and the perpetuation of stereotypes among health professionals.

KEY RECOMMENDATIONS TO ADDRESS RACIAL/ETHNIC DISPARITIES IN HEALTH CARE

Unequal Treatment provides recommendations to address the root causes of racial/ethnic disparities organized as *health system interventions*, *provider interventions*, *patient interventions*, and *general recommendations*.

Health System Interventions • COLLECTING, REPORTING, AND TRACKING OF DATA ON HEALTH CARE ACCESS AND USE, BY PATIENTS' RACE/ETHNICITY *Unequal Treatment* found that the appropriate systems to track and monitor racial and ethnic disparities in health care are lacking and that less is known about the disparities affecting minority groups other than African Americans (Hispanics, Asian Americans, Pacific Islanders, Native Americans, and Alaskan Natives). For instance, only in the mid-1980s did the Medicare database begin to collect data on patient groups outside the standard categories of “white,” “black,” and “other.” Federal, private, and state-supported data-collection efforts are scattered and unsystematic, and many health care systems and hospitals still do not collect data on the race, ethnicity, or primary language of enrollees or patients. A survey by the Institute for Diversity in Health Management and the Health Research and Educational Trust in 2015 found that 98% of 1083 U.S. hospitals collected information on race, 95% collected data on ethnicity, and 94% collected data on primary language. However, only 45% collected

data on race, 40% collected data on ethnicity, and 38% collected data on primary language to benchmark gaps in care. A survey by America's Health Insurance Plans Foundation in 2008 and 2010 showed that the proportion of enrollees in plans that collected race/ethnicity data of some type increased from 75 to 79%; however, the total percentage of plan enrollees whose race/ethnicity and language are recorded is still much lower than these figures.

COLLECTING, REPORTING, AND TRACKING OF SDOH DATA In 2014, the IOM Committee on Recommended Social and Behavioral Domains and Measures for Electronic Health Records recommended the routine collection, in the electronic health record, of a parsimonious panel of clinically significant SDOH measures that may be obtained by self-report in advance of or during the health care encounter and, when used together, provide a psychosocial vital sign. The IOM-recommended questionnaire includes 25 items addressing the following domains: race and ethnicity, education, financial resource strain, stress, depression, physical activity, tobacco use, alcohol use, social connection or isolation, intimate partner violence, residential address, and geocoded census tract median income. Implementation studies have demonstrated that collection of these data takes about 5 minutes, and both patients and providers saw this data collection as appropriate and important. Given that data access and monitoring is an essential component to disparities elimination, we highlight several important sources of up-to-date racial/ethnic disparities monitoring initiatives that are available to the general public and are updated regularly. We highlight only three examples of national data sources.

- Since 2003, the Agency for Healthcare Research and Quality has led the yearly compilation of *The National Healthcare Quality and Disparities Report*, which reports trends for measures related to access to health care, affordable care, care coordination, healthy living, patient safety, and the quality of care across acute and chronic disease management by race/ethnicity, income, and other SDOH (<https://www.ahrq.gov/research/findings/nhqrdr/index.html>).
- Since 2011, the Geospatial Research, Analysis, and Services Program (GRASP) created and maintains the Centers for Disease Control and Prevention Social Vulnerability Index. This database maps, for all U.S. Census tracts, 15 social factors (grouped in four SDOH categories: socioeconomic status, housing composition and disability, minority status and language, and housing and transportation) and is updated every 2 years (<https://www.atsdr.cdc.gov/placeandhealth/svi/index.html>).
- Launched in 2018, the Health Opportunity and Equity (HOPE) Initiative benchmarks and tracks 27 indicators by race, ethnicity, and socioeconomic status. The indicators measure social and economic factors, community and safety, physical environment, access to health care, and health outcomes for the United States (<https://www.nationalcollaborative.org/our-programs/hope-initiative-project/>).

INCREASE INSURANCE COVERAGE AND ACCESS Lack of access to high-quality health care is an important driver of racial/ethnic disparities. Signed into law in 2010, the Affordable Care Act (ACA) fundamentally transformed health insurance by decreasing the uninsured population from 16.3% in 2010 (~49.9 million) to 8.8% in 2016 (~28.1 million). This represents the largest expansion of health insurance since the creation of Medicare and Medicaid in 1965. Prior to the ACA, non-Hispanic blacks were 70% and Hispanics nearly three times more likely to be uninsured than non-Hispanic whites. Of note, Medicaid expansion accounted for an estimated 60% of the ACA's effect through a combination of expanded eligibility and increased enrollment of previously eligible but unenrolled people. This is important given the higher number of racial/ethnic minorities who obtain insurance through Medicaid. Many studies have demonstrated that increased insurance coverage has also translated to greater improvement for blacks and Hispanics in access to care, more access to a usual source of care, and improved health outcomes.

ENCOURAGEMENT OF THE USE OF EVIDENCE-BASED GUIDELINES AND QUALITY IMPROVEMENT *Unequal Treatment* highlights the subjectivity of clinical decision-making as a potential cause of racial and

ethnic disparities in health care by describing how clinicians—despite the existence of well-delineated practice guidelines—may offer (consciously or unconsciously) different diagnostic and therapeutic options to different patients on the basis of their race or ethnicity. Therefore, the widespread adoption and implementation of evidence-based guidelines is a key recommendation in eliminating disparities. For instance, evidence-based guidelines are now available for the management of diabetes, HIV/AIDS, cardiovascular diseases, cancer screening and management, and asthma—all areas where significant disparities exist. As part of ongoing quality-improvement efforts, particular attention should be paid to the implementation of evidence-based guidelines for all patients, regardless of their race and ethnicity.

SUPPORT FOR THE USE OF LANGUAGE INTERPRETATION SERVICES IN THE CLINICAL SETTING As described previously, a lack of efficient and effective interpreter services in a health care system can lead to patient dissatisfaction, to poor comprehension and adherence, and thus to ineffective/lower-quality care for patients with limited English proficiency. *Unequal Treatment's* recommendation to support the use of interpretation services has clear implications for delivery of quality health care by improving doctors' ability to communicate effectively with these patients.

INCREASES IN THE PROPORTION OF UNDERREPRESENTED MINORITIES IN THE HEALTH CARE WORKFORCE Data for 2018 from the Association of American Medical Colleges indicate that of active physicians, 56.2% identified as white, 5.8% identified as Hispanic, 5.0% identified as black or African American, and 0.3% identified as Native American or Alaskan Natives. Furthermore, U.S. national data show that only 3.6% of full-time faculty are black or African American, and 5.5% are Hispanic, Latino, or of Spanish origin (alone or in combination with another race/ethnicity), compared to 63.9% who identified as white. Longitudinal data demonstrate that minority faculty are more likely to be at or below the rank of assistant professor, while whites composed the highest proportion of full professors. Similarly, several studies have found that both Hispanic and black faculty were promoted at lower rates than their white counterparts. Despite representing ~30% of the U.S. population (a number projected to almost double by 2050), minority students are still underrepresented in medical schools. In 2018, matriculates to U.S. medical schools were 6.2% Latino, 7.1% African American, 0.1% Native Hawaiian or Other Pacific Islander, and 0.2% Native American or Alaskan Native. These percentages have decreased or remained nearly the same since 2007. It will be difficult to develop a diverse physician workforce that can meet the needs of an increasingly diverse population without dramatic changes in the racial and ethnic composition of medical student bodies. Long-term investment in pipeline programs and the nearly universal adoption of holistic admissions (a process by which schools consider each applicant individually to determine how they might contribute to the learning environment and the workforce instead of relying just on test scores and grades) have produced modest results. Institutional change in medical schools, focused on creating nurturing, inclusive, and equity-focused environments that dismantle the structural racism that has created the opportunity gap faced by many minority students, is needed to address this important workforce challenge.

Provider Interventions • INTEGRATION OF CROSS-CULTURAL EDUCATION INTO THE TRAINING OF ALL HEALTH CARE PROFESSIONALS The goal of cross-cultural education is to improve providers' ability to understand, communicate with, and care for patients from diverse backgrounds. Such education focuses on enhancing awareness of sociocultural influences on health beliefs and behaviors and on building skills to facilitate understanding and management of these factors in the medical encounter. Cross-cultural education includes curricula on health care disparities, use of interpreters, and effective communication and negotiation across cultures. These curricula can be incorporated into health professions training in medical schools, residency programs, nursing schools, and other health professions programs, and can be offered as a component of continuing education. Despite the importance of this area of education and the attention it has attracted from medical education accreditation bodies, a national survey of

senior resident physicians by Weissman and colleagues found that up to 28% felt unprepared to deal with cross-cultural issues, including caring for patients who have religious beliefs that may affect treatment, patients who use complementary medicine, patients who have health beliefs at odds with Western medicine, patients who mistrust the health care system, and new immigrants. In a study at one medical school, 70% of fourth-year students felt inadequately prepared to care for patients with limited English proficiency. Efforts to incorporate cross-cultural education into medical education will contribute to improving communication and to providing a better quality of care for all patients.

INCORPORATION OF TEACHING ON THE IMPACT OF RACE, ETHNICITY, AND CULTURE ON CLINICAL DECISION-MAKING *Unequal Treatment* and more recent studies found that stereotyping by health care providers can lead to disparate treatment based on a patient's race or ethnicity. The Liaison Committee on Medical Education, which accredits medical schools, issued a directive that medical education should include instruction on how a patient's race, ethnicity, and culture might unconsciously impact communication and clinical decision-making.

Patient Interventions Difficulty navigating the health care system and obtaining access to care can be a hindrance to all populations, particularly to minorities. Similarly, lack of empowerment or involvement in the medical encounter by minorities can be a barrier to care. Patients need to be educated on how to navigate the health care system and how best to access care. Interventions should be used to increase patients' participation in treatment decisions.

General Recommendations • INCREASE AWARENESS OF RACIAL/ETHNIC DISPARITIES IN HEALTH CARE Efforts to raise awareness of racial/ethnic health care disparities have done little for the general public but have been fairly successful among physicians, according to a Kaiser Family Foundation report. In 2006, nearly 6 in 10 people surveyed believed that blacks received the same quality of care as whites, and 5 in 10 believed that Latinos received the same quality of care as whites. These estimates are similar to findings in a 1999 survey. Despite this lack of awareness, most people believed that all Americans deserve quality care, regardless of their background. In contrast, the level of awareness among physicians has risen sharply. In 2002, the majority (69%) of physicians said that the health care system "rarely or never" treated people unfairly on the basis of their racial/ethnic background. In 2005, less than one-quarter (24%) of physicians disagreed with the statement that "minority patients generally receive lower-quality care than white patients." More recently, a survey by WebMD showed that 42% of 16,000 physicians admitted that their own personal biases impact their clinical decision-making, including on characteristics such as race and ethnicity. Increasing awareness of racial and ethnic health disparities, and their root causes, among health care professionals and the public is an important first step in addressing these disparities. The ultimate goals are to generate discourse and to mobilize action to address disparities at multiple levels, including health policymakers, health systems, and the community.

CONDUCT FURTHER RESEARCH TO IDENTIFY SOURCES OF DISPARITIES AND PROMISING INTERVENTIONS While the literature that formed the basis for the findings reported and recommendations made in *Unequal Treatment* provided significant evidence for racial and ethnic disparities, additional research is needed in several areas. First, most of the literature on disparities focuses on black-versus-white differences; much less is known about the experiences of other minority groups. Improving the ability to collect racial and ethnic patient data should facilitate this process. However, in instances where the necessary systems are not yet in place, racial and ethnic patient data may be collected prospectively in the setting of clinical or health services research to more fully elucidate disparities for other populations. Second, much of the literature on disparities to date has focused on defining areas in which these disparities exist, but less has been done to identify the multiple factors that contribute to the disparities or to test interventions to address these factors. There is clearly a need for research that identifies promising practices and solutions to disparities.

■ IMPLICATIONS FOR CLINICAL PRACTICE

Individual health care providers can do several things in the clinical encounter to address racial and ethnic disparities in health care.

Be Aware That Disparities Exist Increasing awareness of racial and ethnic disparities among health care professionals is an important first step in addressing disparities in health care. Only with greater awareness can care providers be attuned to their behavior in clinical practice and thus monitor that behavior and ensure that all patients receive the highest quality of care, regardless of race, ethnicity, or culture.

Practice Culturally Competent Care Previous efforts have been made to teach clinicians about the attitudes, values, beliefs, and behaviors of certain cultural groups—the key practice “dos and don’ts” in caring for “the Hispanic patient” or the “Asian patient,” for example. In certain situations, learning about a particular local community or cultural group, with a goal of following the principles of community-oriented primary care, can be helpful; when broadly and uncritically applied, however, this approach can actually lead to stereotyping and oversimplification of culture, without respect for its complexity.

Cultural competence has thus evolved from merely learning information and making assumptions about patients on the basis of their backgrounds to focusing on the development of skills that follow the principles of patient-centered care. *Patient-centeredness* encompasses the qualities of compassion, empathy, and responsiveness to the needs, values, and expressed preferences of the individual patient. *Cultural competence* aims to take things a step further by expanding the repertoire of knowledge and skills classically defined as “patient-centered” to include those that are especially useful in cross-cultural interactions (and that, in fact, are vital in all clinical encounters). This repertoire includes effectively using interpreter services, eliciting the patient’s understanding of his or her condition, assessing decision-making preferences and the role of family, determining the patient’s views about biomedicine versus complementary and alternative medicine, recognizing sexual and gender issues, and building trust. For example, while it is important to understand all patients’ beliefs about health, it may be particularly crucial to understand the health beliefs of patients who come from a different culture or have a different health care experience. With the individual patient as teacher, the physician can adjust his or her practice style to meet the patient’s specific needs.

Avoid Stereotyping Several strategies can allow health care providers to counteract, both systemically and individually, the normal tendency to stereotype. For example, when racially/ethnically/culturally/socially diverse teams in which each member is given equal power are assembled and are tasked to achieve a common goal, a sense of camaraderie develops and prevents the development of stereotypes based on race/ethnicity, gender, culture, or class. Thus, health care providers should aim to gain experiences working with and learning from a diverse set of colleagues. In addition, simply being aware of the operation of social cognitive factors allows providers to actively check up on or monitor their behavior. Physicians can constantly reevaluate to ensure that they are offering the same things, in the same ways, to all patients. Understanding one’s own susceptibility to stereotyping—and how disparities may result—is essential in providing equitable, high-quality care to all patients.

Work to Build Trust Patients’ mistrust of the health care system and of health care providers impacts multiple facets of the medical encounter, with effects ranging from decreased patient satisfaction to delayed care. Although the historic legacy of discrimination can never be erased, several steps can be taken to build trust with patients and to address disparities. First, providers must be aware that mistrust exists and is more prevalent among minority populations, given the history of discrimination in the United States and other countries. Second, providers must reassure patients that they come first, that everything possible will be done to ensure that they always get the best care available, and that their caregivers will serve as their advocates. Third, interpersonal skills and communication techniques that demonstrate honesty, openness, compassion, and respect on the part of the health

care provider are essential tools in dismantling mistrust. Finally, patients indicate that trust is built when there is shared, participatory decision-making and the provider makes a concerted effort to understand the patient’s background. When the doctor–patient relationship is reframed as one of solidarity, the patient’s sense of vulnerability can be transformed into one of trust. The successful elimination of disparities requires trust-building interventions and strengthening of this relationship.

■ CONCLUSION

The issue of racial and ethnic disparities in health care has gained national prominence, both with the release of the IOM report *Unequal Treatment* and with more recent articles that have confirmed their persistence and explored their root causes. Furthermore, another influential IOM report, *Crossing the Quality Chasm*, has highlighted the importance of equity—i.e., no variations in quality of care due to personal characteristics, including race and ethnicity—as a central principle of quality. Current efforts in health care reform and transformation, including a greater focus on value (high-quality care and cost-control), will sharpen the nation’s focus on the care of populations who experience low-quality, costly care. Addressing disparities will become a major focus, and there will be many obvious opportunities for interventions to eliminate them. Greater attention to addressing the root causes of disparities will improve the care provided to all patients, not just those who belong to racial and ethnic minorities.

■ FURTHER READING

- BUCHMUELLER TC et al: The ACA’s impact on racial and ethnic disparities in health insurance coverage and access to care. *Health Aff* (Millwood) 39:395, 2020.
- CARNETHON MR et al: Cardiovascular health in African Americans: A scientific statement from the American Heart Association. *Circulation* 136:e393, 2017.
- DWYER-LINDGREN L et al: Inequalities in life expectancy among us counties, 1980 to 2014: Temporal trends and key drivers. *JAMA Intern Med* 177:1003, 2017.
- KREUTER MW et al: Addressing social needs in health care settings: Evidence, challenges and opportunities for public health. *Annu Rev Public Health* 42:11, 2021.
- KRIEGER N: Measures of racism, sexism, heterosexism, and gender binarism for health equity research: from structural injustice to embodied harm: An ecosocial analysis. *Annu Rev Public Health* 41:37, 2020.
- MEDSCAPE: Medscape Lifestyle Report 2016: Bias and burnout. <http://www.medscape.com/features/slideshow/lifestyle/2016/public/overview>.
- VIAS DA et al: Hidden in plain sight: Reconsidering the use of race correction in clinical algorithms. *N Engl J Med* 383:874, 2020.
- WILLIAMS DR et al: Racism and health: Evidence and needed research. *Annu Rev Public Health* 40:105, 2019.

11

Ethical Issues in Clinical Medicine

Christine Grady, Bernard Lo

Physicians face novel ethical dilemmas that can be perplexing and emotionally draining. For example, telemedicine, artificial intelligence, handheld personal devices, and learning health care systems all hold the promise of more coordinated and comprehensive care, but also raise concerns about confidentiality, the doctor–patient relationship, and responsibility. This chapter presents approaches and principles that physicians can use to address important vexing ethical issues they

encounter in their work. Physicians make ethical judgments about clinical situations every day. They should prepare for lifelong learning about ethical issues so they can respond appropriately. Traditional professional codes and ethical principles provide instructive guidance for physicians but need to be interpreted and applied to each situation. When facing or struggling with a challenging ethical issue, physicians may need to reevaluate their basic convictions, tolerate uncertainty, and maintain their integrity while respecting the opinions of others. Physicians should articulate their concerns and reasoning, discuss and listen to the views of others involved in the case, and utilize available resources, including other health care team members, palliative care, social work, and spiritual care. Moreover, ethics consultation services or a hospital ethics committee can help to clarify issues and identify strategies for resolution, including improving communication and dealing with strong or conflicting emotions. Through these efforts, physicians can gain deeper insight into the ethical issues they face and usually reach mutually acceptable resolutions to complex problems.

APPROACHES TO ETHICAL PROBLEMS

Several approaches are useful for resolving ethical issues, including approaches based on ethical principles, virtue ethics, professional oaths, and personal values. These various sources of guidance may seem to conflict in a particular case, leaving the physician in a quandary. In a diverse society, different individuals may turn to different sources of moral guidance. In addition, general moral precepts often need to be interpreted and applied to a particular clinical situation.

ETHICAL PRINCIPLES

Ethical principles can serve as general guidelines to help physicians determine the right thing to do.

Respecting Patients Physicians should always treat patients with respect, which entails understanding patients' goals, providing information, communicating effectively, obtaining informed and voluntary consent, respecting informed refusals, and protecting confidentiality. Different clinical goals and approaches are often feasible, and interventions can result in both benefit and harm. Individuals differ in how they value health and medical care and how they weigh the benefits and risks of medical interventions. Generally, physicians should respect patients' values and informed choices. Treating patients with respect is especially important when patients are responding to experiences of, or fears about, disrespect and discrimination.

GOALS AND TREATMENT DECISIONS Physicians should provide relevant and accurate information for patients about diagnoses, current clinical circumstances, expected future course, prognosis, treatment options, and uncertainties, and discuss patients' goals of care. Physicians may be tempted to withhold a serious diagnosis, misrepresent it by using ambiguous terms, or limit discussions of prognosis or risks for fear that patients will become anxious or depressed. Providing honest information about clinical situations promotes patients' autonomy and trust as well as sound communication with patients and colleagues. When physicians have to share bad news with patients, they should adjust the pace of disclosure, offer empathy and hope, provide emotional support, and call on other resources such as spiritual care or social work to help patients cope. Some patients may choose not to receive such information or may ask surrogates to make decisions on their behalf, as is common with serious diagnoses in some traditional cultures.

SHARED DECISION-MAKING AND OBTAINING INFORMED CONSENT Physicians should engage their patients in shared decision-making about their health and their care, whenever appropriate. Physicians should discuss with patients the nature, risks, and benefits of proposed care; any alternative; and the likely consequences of each option. Physicians promote shared decision-making by informing and educating patients, answering their questions, checking that they understand key issues, making recommendations, and helping them to deliberate. Medical jargon, needlessly complicated explanations, or the provision of too much information at once may overwhelm patients. Increasingly, decision aids can assist patients in playing a more active role in

decision-making, improving the accuracy of their perception of risk and benefit, and helping them feel better informed and clearer about their values. Informed consent is more than obtaining signatures on consent forms and involves disclosure of honest and understandable information to promote understanding and choice. Competent, informed patients may refuse recommended interventions and choose among reasonable alternatives. In an emergency, treatment can be given without informed consent if patients cannot give their own consent and delaying treatment while surrogates are contacted would jeopardize patients' lives or health. People are presumed to want such emergency care unless they have previously indicated otherwise.

Respect for patients does not entitle patients to insist on any care or treatment that they want. Physicians are not obligated to provide interventions that have no physiologic rationale, that have already failed, or that are contrary to evidence-based practice recommendations or good clinical judgment. Public policies and laws also dictate certain decisions—e.g., allocation of scarce medical resources during a public health crisis such as the COVID-19 pandemic, use of cadaveric organs for transplantation, and requests for physician aid in dying.

CARING FOR PATIENTS WHO LACK DECISION-MAKING CAPACITY Some patients are unable to make informed decisions because of unconsciousness, advanced dementia, delirium, or other medical conditions. Courts have the legal authority to determine that a patient is legally incompetent, but in practice, physicians usually determine when patients lack the capacity to make particular health care decisions and arrange for authorized surrogates to make decisions, without involving the courts. Patients with decision-making capacity can express a choice and appreciate their medical situation; the nature, risks, and benefits of proposed care; and the consequences of each alternative. Patient choices should be consistent with their values and not the result of delusions, hallucinations, or misinformation. Physicians should use available and validated assessment tools, resources such as psychiatry or ethics consultation, and clinical judgment to ascertain whether individuals have the capacity to make decisions for themselves. Patients should not be assumed to lack capacity if they disagree with recommendations or refuse treatment. Such decisions should be probed, however, to ensure the patient is not deciding based on misunderstandings and has the capacity to make an informed decision. When impairments are fluctuating or reversible, decisions should be postponed if possible until the patient recovers decision-making capacity.

When a patient lacks decision-making capacity, physicians seek an appropriate surrogate. Patients may designate a health care proxy through an advance directive or on a Physician Orders for Life-Sustaining Treatment form; such choices should be respected (see Chap. 12). For patients who lack decision-making capacity and have not previously designated a health care proxy, family members usually serve as surrogates. Statutes in most U.S. states delineate a prioritized list of relatives to make medical decisions. Patients' values, goals, and previously expressed preferences guide surrogate decisions. However, the patient's current best interests may sometimes justify overriding earlier preferences if an intervention is likely to provide significant benefit, previous statements do not fit the situation well, or the patient gave the surrogate leeway in decisions.

MAINTAINING CONFIDENTIALITY Maintaining confidentiality is essential to respecting patients' autonomy and privacy; it encourages patients to seek treatment and to discuss problems candidly. However, confidentiality may be overridden to prevent serious harm to third parties or the patient. Exceptions to confidentiality are justified when the risk to others is serious and probable, no less restrictive measures can avert risk, and the adverse effects of overriding confidentiality are minimized and deemed acceptable by society. For example, laws require physicians to report cases of tuberculosis, sexually transmitted infection, elder or child abuse, and domestic violence.

Beneficence or Acting in Patients' Best Interests The principle of *beneficence* requires physicians to act for the patient's benefit. Patients typically lack medical expertise, and illness may make them vulnerable. Patients rely on and trust physicians to treat them with

compassion and provide sound recommendations and treatments aimed to promote their well-being. Physicians encourage such trust and have a fiduciary duty to act in the best interests of patients, which should prevail over physicians' self-interest or the interests of third parties such as hospitals or insurers. A principle related to beneficence, "first do no harm," obliges physicians to prevent unnecessary harm by recommending interventions that maximize benefit and minimize harm and forbids physicians from providing known ineffective interventions or acting without due care. Although often cited, this precept alone provides limited guidance because many beneficial interventions also pose serious risks.

Physicians increasingly provide care within interdisciplinary teams and rely on consultation with or referral to specialists. Team members and consultants contribute different types of expertise to the provision of comprehensive, high-quality care for patients. Physicians should collaborate with and respect the contributions of the various interdisciplinary team members and should initiate and participate in regular communication and planning to avoid diffusion of responsibility and ensure accountability for quality patient care.

INFLUENCES ON PATIENTS' BEST INTERESTS Conflicts arise when patients' refusal or request of interventions thwarts their own goals for care, causes serious harm, or conflicts with their best medical interests. For example, simply accepting a young asthmatic adult's refusal of mechanical ventilation for reversible respiratory failure, in the name of respecting autonomy, is morally constricted. Physicians should elicit patients' expectations and concerns, correct their misunderstandings, and try to persuade them to accept beneficial therapies. If disagreements persist after such efforts, physicians should call on institutional resources for assistance, but patients' informed choices and views of their own best interests should prevail.

Drug prices and out-of-pocket expenses for patients have been escalating in many parts of the world and may compromise care that is in the patients' best interests. Physicians should recognize that patients, especially those with high copayments or inadequate insurance, may not be able to afford prescribed tests and interventions. Physicians should strive to prescribe medications that are affordable and acceptable to the patient. Knowing what kind of insurance, if any, the patient has and whether certain medications are likely to be covered may help in determining appropriate prescriptions. Available alternatives should be considered and discussed. Physicians should follow up with patients who don't fill prescriptions, don't take their medications, or skip doses to explore whether cost and affordability are obstacles. It may be reasonable for physicians to advocate for coverage of nonformulary products for sound reasons, such as when the formulary drugs are less effective or not tolerated or are too costly for the patient to pay for out of pocket. These should be shared decisions with the patient to the extent possible.

Organizational policies and workplace conditions may sometimes conflict with patients' best interests. Physicians' focus and dedication to the well-being and interests of patients may be negatively influenced by perceived or actual staffing inadequacies, unfair wages, infrastructural deficiencies or lack of equipment, work-hour limitations, corporate culture, and threats to personal security in the workplace. Physicians should work with institutional leaders to ensure that policies and practices support their ability to provide quality care focused on patients' best interests.

Patients' interests are served by improvements in overall quality of care and the increasing use of evidence-based practice guidelines and performance benchmarking. However, practice guideline recommendations may not serve the interests of each individual patient, especially when another plan of care may provide substantially greater benefits. In prioritizing their duty to act in the patient's best interests, physicians should be familiar with relevant practice guidelines, be able to recognize situations that might justify exceptions, and advocate for reasonable exceptions.

Acting Justly The principle of *justice* provides guidance to physicians about how to ethically treat patients and make decisions about allocating important resources, including their own time. *Justice* in a

general sense means fairness: people should receive what they deserve. In addition, it is important to act consistently in cases that are similar in ethically relevant ways, in order to avoid arbitrary, biased, and unfair decisions. Justice forbids discrimination in health care based on race, religion, gender, sexual orientation, disability, age, or other personal characteristics (**Chap. 10**).

ALLOCATION OF RESOURCES Justice also requires fair allocation of limited health care resources. Universal access to medically needed health care remains an unrealized moral aspiration in the United States and many countries around the world. Patients with no or inadequate health insurance often cannot afford health care and lack access to safety-net services. Even among insured patients, insurers may deny coverage for interventions recommended by their physician. In this situation, physicians should advocate for patients' affordable access to indicated care, try to help patients obtain needed care, and work with institutions and policies to promote wider access. Doctors might consider—or patients might request—the use of lies or deception to obtain such benefits, for example, signing a disability form for a patient who does not meet disability criteria. Although motivated by a desire to help the patient, such deception breaches basic ethical guidelines and undermines physicians' credibility and trustworthiness.

Allocation of health care resources is unavoidable when resources are limited. Allocation policies should be fair, transparent, accountable, responsive to the concerns of those affected, and proportionate to the situation, including the supply relative to the need. In the 2019–2020 SARS-CoV-2/COVID-19 pandemic, some epicenters anticipated or faced shortages of staff, protective equipment, hospital and critical care beds, and ventilators, even after increasing supplies and modifying usual clinical procedures. Many jurisdictions developed guidelines for implementing crisis standards of care to allocate limited interventions and services. Under crisis standards of care, some aspects of conventional care are not possible and interventions may not be provided to all who might benefit or wish to receive them. Crisis standards of care aim to promote the good of the community by saving the most lives in the short term, using evidence-based criteria.

When demand for medications or other interventions exceeds the supply, allocation should be fair, strive to avoid discrimination, and mitigate health disparities. First-come, first-served allocation is not fair, because it disadvantages patients who experience barriers to accessing care. To avoid discrimination, allocation decisions should not consider personal social characteristics such as race, gender, or disability, nor consider insurance status or wealth. Allocation policies also should aspire to reduce health care disparities. U.S. African-American, Latino-American, and Native-American patients suffered a disproportionate number of COVID-19 cases and deaths, likely due in part to being employed in jobs that cannot be done remotely or with physical distancing, crowded housing, lack of health benefits, and poor access to health care.

Fair and well-considered guidelines help mitigate any emotional and moral distress that clinicians may experience making difficult allocation decisions. Authorizing triage officers or committees to make allocation decisions according to policies determined with public input allows treating physicians and nurses to dedicate their efforts to their patients. Ad hoc resource allocation by physicians at the bedside may be inconsistent, unfair, and ineffective. At the bedside, physicians should act as patient advocates within constraints set by society, reasonable insurance policies, and evidence-based practice. Many allocation decisions are made at the level of public policy, with physician and public input. For example, the United Network for Organ Sharing (www.unos.org) provides criteria for allocating scarce organs.

VIRTUE ETHICS

Virtue ethics focuses on physicians' character and qualities, with the expectation that doctors will cultivate virtues such as compassion, trustworthiness, intellectual honesty, humility, and integrity. Proponents argue that, if such characteristics become ingrained, they help guide physicians in unforeseen situations. Moreover, following ethical precepts or principles without any of these virtues could lead to uncaring doctor–patient relationships.

■ PROFESSIONAL OATHS AND CODES

Professional oaths and codes are useful guides for physicians. Most physicians take oaths during their medical training, and many are members of professional societies that have professional codes. Physicians pledge to the public and to their patients that they will be guided by the principles and values in these oaths or codes and commit to the spirit of the ethical ideals and precepts represented in oaths and professional codes of ethics.

■ PERSONAL VALUES

Personal values, cultural traditions, and religious beliefs are important sources of personal morality that help physicians address ethical issues and cope with any moral distress they may experience in practice. While essential, personal morality alone is a limited ethical guide in clinical practice. Physicians have role-specific ethical obligations that go beyond their obligations as good people, including the duties to obtain informed consent and maintain confidentiality discussed earlier. Furthermore, in a culturally and religiously diverse world, physicians should expect that some patients and colleagues will have personal moral beliefs that differ from their own.

ETHICALLY COMPLEX PROFESSIONAL ISSUES FOR PHYSICIANS

■ CLAIMS OF CONSCIENCE

Some physicians, based on their personal values, have conscientious objections to providing, or referring patients for, certain treatments such as contraception or physician aid in dying. Although physicians should not be asked to violate deeply held moral beliefs or religious convictions, patients need medically appropriate, timely care and should always be treated with respect. Institutions such as clinics and hospitals have a collective ethical duty to provide care that patients need while making reasonable attempts to accommodate health care workers' conscientious objections—for example, when possible by arranging for another professional to provide the service in question. Patients seeking a relationship with a doctor or health care institution should be notified in advance of any conscientious objections to the provision of specific interventions. Since insurance often constrains patients' selection of physicians or health care facilities, switching providers can be burdensome. There are also important limits on claims of conscience. Health care workers may not insist that patients receive unwanted medical interventions. They also may not refuse to treat or discriminate against patients because of their race, ethnicity, disability, genetic information, or diagnosis. Such discrimination is illegal and violates physicians' duties to respect patients. Refusal to treat patients for other reasons such as sexual orientation, gender identity, or other personal characteristics is legally more controversial, yet ethically inappropriate because it falls short of helping patients in need and respecting them as persons.

■ PHYSICIAN AS GATEKEEPER

In some cases, patients may ask their physicians to facilitate access to services that the physician has ethical qualms about providing. For example, a patient might request a prescription for a cognitively enhancing medication to temporarily augment his cognitive abilities in order to take an exam or apply for employment. Patients may request more pain medication than the physician believes is warranted for the given situation or marijuana to facilitate sleep. Patients may ask their physician to sign a waiver to avoid vaccines for reasons that are not included in state exceptions (see [Chap. 3](#)). A physician may feel uncomfortable prescribing attention-deficit/hyperactivity disorder medications to a young child because she is not convinced that the possible benefit justifies the risks to the child despite the parent's request. In these circumstances, the physician should work with the patient or parent to understand the reasons for their requests, some of which might be legitimate. In addition to considering possible risks and benefits to the patient, the physician should consider how meeting the request might affect other patients, societal values, and public trust in the medical profession. If the physician determines that fulfilling the request requires deception, is unfair, jeopardizes her professional

responsibilities, or is inconsistent with the patient's best medical interests, the physician should decline and explain the reasons to the patient.

■ MORAL DISTRESS

Health care providers, including residents, medical students, and experienced physicians, may experience moral distress when they feel that ethically appropriate action is hindered by institutional policies or culture, decision-making hierarchies, limited resources, or other reasons. Moral distress can lead to anger, anxiety, depression, frustration, fatigue, work dissatisfaction, and burnout. A physician's health and well-being can affect how he or she cares for patients. Discussing complex or unfamiliar clinical situations with colleagues and seeking assistance with difficult decisions can help alleviate moral distress, as can a healthy work environment characterized by open communication, mutual respect, and emphasis on the common goal of good patient care. In addition, physicians should take good care of their own well-being and be aware of the personal and system factors associated with stress, burnout, and depression. Health care organizations should provide a supportive work environment, counseling, and other support services when needed.

■ OCCUPATIONAL RISKS AND BURDENS

Physicians accept some physical risk in fulfilling their professional responsibilities, including exposure to infectious agents or toxic substances, violence in the workplace, and musculoskeletal injury. Nonetheless, most physicians, nurses, and other hospital staff willingly care for patients, despite personal risk and fear, grueling hours, and sometimes inadequate personal protective equipment or information. During the COVID-19 pandemic, many communities honored clinicians' dedication to professional ideals, and some medical students who were relieved from in-person patient care responsibilities volunteered to support front-line workers in other ways. The burdens of navigating professional and personal responsibilities fall more heavily on women health care providers. Health care institutions are responsible for reducing occupational risk and burden by providing proper information, training and supervision, protective equipment, infrastructure and workflow modifications, and emotional and psychological support to physicians. Clinical leaders need to acknowledge fears about personal safety and take steps to mitigate the impact of work on family responsibilities, moral distress, and burnout.

■ USE OF SOCIAL MEDIA AND PATIENT PORTALS

Increasingly, physicians use social and electronic media to share information and advice with patients and other providers. Social networking may be especially useful in reaching young or otherwise hard-to-access patients. Patients increasingly access their physicians' notes through patient portals, which aim to transparently share information, promote patient engagement, and increase adherence. Physicians should be professional and respectful and consider patient confidentiality, professional boundaries, and therapeutic relationships when posting to social media or writing notes for the portal. Overall, appropriate use of these platforms can enhance communication and transparency while avoiding misunderstandings or harmful consequences for patients, physicians, or their colleagues. Unprofessional or careless posts that express frustration or anger over work incidents, disparage patients or colleagues, use offensive or discriminatory language, or reveal inappropriate personal information about the physician can have negative consequences. Physicians should separate professional from personal websites and accounts and follow institutional and professional society guidelines when communicating with patients.

CONFLICTS OF INTEREST

Acting in patients' best interests may sometimes conflict with a physician's self-interest or the interests of third parties such as insurers or hospitals. From an ethical viewpoint, patients' interests are paramount. Transparency, appropriate disclosure, and management of conflicts of interest are essential to maintain the trust of colleagues and the public. Disclosure requirements vary for different purposes, and software has

been developed to assist physicians in complying with specific requirements. Importantly, not all conflicts are financial. Physicians sometimes face conflicts of commitment between their patient's interests and their own personal interests, professional goals, responsibilities, and aspirations. As mentioned earlier, physicians should prioritize patients' interests while recognizing possible conflicts and using disclosure, discussion with the chief of service, and management of the conflict or recusal when appropriate.

In addition to individual physicians, medical institutions may have conflicts of interest arising from patent rights, industry-funded research programs, and donations from individuals and companies. Institutions need to be transparent about the presence and amount of such relationships and make clear the steps taken to prevent such relationships from having an impact on clinical or financial decisions. If there is good evidence that a donor acted in ways that breached ethical or legal standards, the institution should take steps not to benefit from the donation or honor the donor.

■ FINANCIAL INCENTIVES

Physicians have financial incentives to improve the quality or efficiency of care that might lead some to avoid patients who are older, are chronically ill, or have more complicated problems, or to focus on benchmarked outcomes even when not in the best interests of individual patients. In contrast, fee-for-service payments might encourage physicians to order more interventions than necessary or to refer patients to laboratory, imaging, or surgical facilities in which they have a financial stake. Regardless of financial incentives, physicians should recommend available care that is in the patient's best interests—no more and no less.

■ RELATIONSHIPS WITH PHARMACEUTICAL COMPANIES

Financial relationships between physicians and industry are increasingly scrutinized. Many academic medical centers have banned drug-company gifts, including branded pens and notepads and meals to physicians, to reduce inappropriate risk of undue influence or subconscious feelings of reciprocity and to decrease possible influences on public trust or the costs of health care.

The federal Open Payments website provides public information on the payments and amounts that drug and device companies give to individual physicians by name. The challenge is to distinguish payments for scientific consulting and research contracts—which should be encouraged as consistent with professional and academic missions—from those for promotional speaking and consulting whose goal is to increase sales of company products.

■ LEARNING CLINICAL SKILLS

Medical students', residents', and physicians' interests in learning, which fosters the long-term goal of benefiting future patients, may sometimes conflict with the short-term goal of providing optimal care to current patients. When trainees are learning procedures on patients, they lack the proficiency of experienced physicians, and patients may experience inconvenience, discomfort, longer procedures, or increased risk. Increasingly, institutions are developing clinical skills laboratories for simulation-based medical education and requiring students to demonstrate proficiency before carrying out procedures such as venipuncture and intravenous lines in patients. Furthermore, teaching hospitals are establishing proceduralist services in which procedure-specialist faculty members directly supervise interns for procedures such as lumbar puncture and thoracentesis and certify their proficiency. Medical students may need to defer learning such invasive procedures until internship. Seeking patients' consent for trainee participation in their care is always important and is particularly important for intimate examinations, such as pelvic, rectal, breast, and testicular examinations, and for invasive procedures. Patients should be told who is providing care and how trainees are supervised. Failing to introduce students or not telling patients that trainees will be performing procedures undermines trust, may lead to more elaborate deception, and makes it difficult for patients to make informed choices about their care. Most patients, when informed, allow trainees to play an active role in their care.

■ RESPONSE TO MEDICAL ERRORS

Errors are inevitable in clinical medicine, and some errors cause harm to patients. Most errors are caused by lapses of attention or flaws in the system of delivering health care; only a small number result from blameworthy individual behavior. Many health care institutions have adopted a just culture system, which encourages open and honest reporting of errors as essential to quality learning and shifts the focus from individual blame to system design for improvement in quality and safety (*Chap. 8*). This approach is more likely than a punitive approach to improve patient safety. However, professional discipline is appropriate for cases of gross incompetence, reckless behavior, physician impairment, and boundary violations. Physicians and students may fear that disclosing errors will damage their careers. Physicians and health care institutions show respect for patients by disclosing and explaining errors, offering an apology, offering appropriate compensation for harm done, and using errors as opportunities to improve the quality of care.

■ PHYSICIAN IMPAIRMENT

Physicians may hesitate to intervene when colleagues impaired by alcohol, drugs, or psychiatric or medical illness place patients at risk. However, society relies on physicians to regulate themselves. Colleagues of an impaired physician should take steps to protect patients and help their impaired colleague, starting with reporting their concerns to their clinical supervisor or director.

ETHICAL ISSUES IN CLINICAL RESEARCH

Clinical research is essential to translate scientific discoveries into beneficial interventions for patients. However, clinical research raises ethical concerns because participants face inconvenience and risks in research designed to advance scientific knowledge and not specifically to benefit them. Ethical guidelines require researchers to rigorously design and conduct research, minimize risk to participants, and obtain informed and voluntary consent from participants and approval from an institutional review board (IRB). IRBs determine that risks to participants are acceptable and have been minimized and recommend appropriate additional protections when research includes vulnerable participants.

Physicians may be clinical research investigators themselves or may be in a position to refer or recommend clinical trial participation to their patients. Physician-investigators are likely to feel some inherent tension between conducting research and providing health care. Awareness of this tension, familiarity with research ethics, collaboration with research and clinical team members, and utilizing research ethics consultation can help to mitigate tensions. Before starting clinical research, investigators should complete training in the ethics of clinical research, which is widely available.

Physicians also should be critical consumers of clinical research results and keep up with research advances that change standards of practice. Precision medicine initiatives aim to individualize clinical care by combining clinical information from electronic health records, genomic sequencing, and data from personal mobile devices. Furthermore, physicians and health care institutions are analyzing data routinely collected and available in electronic health records, leftover clinical specimens, and administrative data. Such studies encompass traditional discovery research as well as quality improvement, comparative effectiveness research, and learning health care systems. Efforts to improve the quality of care in real-world clinical settings are important but also raise new issues about informed consent, privacy, and risk.

EMERGING TECHNOLOGIES

Scientific advances in genome sequencing, gene editing (e.g., with CRISPR-Cas9), machine learning, artificial intelligence, computer-brain interfaces, and other technologies offer great promise for research and clinical care with the ultimate goal of improving the prediction, prevention, and treatment of disease. Groundbreaking innovations that have strong scientific plausibility need to be evaluated in rigorous clinical studies for efficacy and safety.

Physicians should keep up to date on the status of novel and often complex technologies as research evolves, data emerge, and technologies are incorporated into clinical practice. They can help their patients understand research findings and the evidence for clinical use, correct any misunderstandings, facilitate shared decision-making, and advocate for fair access to such therapies. Further, physicians should engage in professional and public discussion related to allocation of resources and fair access to expensive new therapies and emerging technologies and their impact on overall health care affordability.

Certain cell-based therapies, such as peripheral blood stem cell transplantation (**Chap. 114**) and chimeric antigen receptor (CAR)-T cell therapy (**Chap. 69**), are approved for use in several serious hematologic cancers, and gene therapies have been approved as safe and effective for clinical use in certain serious inherited diseases and cancers. Patients may request these and other complex, highly technical, and expensive therapies for unproven indications. Yet, claims of cures through unproven stem cell or gene-based “therapies” pose significant health and financial risks to patients without evidence of benefit. Physicians should help patients distinguish approved therapies from unproven claims and refer interested patients to well-designed clinical trials.

Medical applications of CRISPR-Cas9 are promising, and their safety and efficacy for particular clinical conditions are being carefully evaluated in clinical trials. Applications of CRISPR genome editing in somatic cells to modify or correct problematic genes could lay the foundation for treating a variety of serious diseases, including blood disorders, HIV, cancer, and hereditary blindness. Germline gene editing in blastocysts or embryos raises many ethical questions and is currently not permitted in the United States in clinical trials or clinical practice.

In artificial intelligence (AI), computers carry out tasks typically done by humans. Machine learning (ML) is a type of AI that automatically learns and improves its performance without explicit programming. Clinical algorithms using AI and ML can make diagnoses from radiology images, retinal scans, or skin photographs and identify patients at increased risk for surgical complications, critical care, or hospital readmission. However, such algorithms can also pose risks. Bias may occur if an algorithm was derived or validated from a data set in which groups who suffer from health disparities or poor health outcomes are underrepresented or if the algorithm predicts outcomes that are not clinically meaningful. To address these ethical concerns, researchers should assess AI algorithms in well-designed randomized clinical trials with clinical endpoints. Institutions should integrate validated and unbiased algorithms into clinical workflow without unduly burdening physicians and nurses and should check effectiveness and safety in their particular settings and patient populations.

Physicians should stay informed of emerging evidence about such technologies and the ethical challenges that accompany their use and always keep their patients’ best interests and preferences at the forefront.

GLOBAL CONSIDERATIONS

■ INTERNATIONAL RESEARCH

Clinical research is often conducted across multiple sites and across national borders. Societal, legal, and cultural norms and perspectives about research may vary, and there are many ethical challenges. Physician-investigators involved in international research should be familiar with international guidelines, such as the Declaration of Helsinki, the Council for International Organizations of Medical Sciences (CIOMS) guidelines, and the International Council on Harmonisation Good Clinical Practice guidelines, as well as national and local laws where research is taking place. Partnering with local researchers and communities is essential not only to demonstrate respect but also to facilitate successful clinical research.

■ INTERNATIONAL CLINICAL EXPERIENCES

Many physicians and trainees gain valuable experience providing patient care in international settings through international training opportunities or volunteering for humanitarian or other international

clinical work. Such arrangements, however, raise ethical challenges—for example, as a result of differences in beliefs about health and illness, expectations regarding health care and physicians’ roles, standards of clinical practice, resource limitations, and norms for disclosure of serious diagnoses. Additional dilemmas arise if visiting physicians and trainees take on responsibilities beyond their expertise or if donated drugs and equipment are not appropriate to local needs. Visiting physicians and trainees should prepare well for these experiences, receive training and mentoring, learn about cultural and clinical practices in the host community, respect local customs and values, collaborate closely with local professionals and staff, and be explicit and humble about their own skills, knowledge, and limits. Leaders of global health field experiences should ensure that participating physicians receive training on ethical and cultural issues, as well as mentoring, backup, and debriefing upon return home.

■ CONCLUSION

Ethical issues are common in clinical medicine and occur in circumstances that may be foreseeable, novel, or unexpected. Physicians address these ethical issues by being prepared, informed, and thoughtful and using appropriate available resources.

■ FURTHER READING

- BEAUCHAMP T, CHILDRESS J: *Principles of Biomedical Ethics*, 8th ed. New York, Oxford University Press, 2019.
- DEJONG C et al: An ethical framework for allocating scarce medications for COVID-19 in the US. *JAMA* 323:2367, 2020.
- MATHENY M et al (eds): *Artificial Intelligence in Health Care: The Hope, the Hype, the Promise, the Peril*. NAM Special Publication. Washington, DC, National Academy of Medicine, 2019.
- ULRICH C, GRADY C: *Moral Distress in the Health Professions*. Cham, Switzerland, Springer-Nature International, 2018.
- WASSERMAN J et al: Responding to unprofessional behavior by trainees: a “just culture” framework. *N Engl J Med* 382:773, 2020.
- WICCLAIR MR: Conscientious objection, moral integrity, and professional obligations. *Perspect Biol Med* 62:543, 2019.

12

Palliative and End-of-Life Care

Ezekiel J. Emanuel



EPIDEMIOLOGY

■ CAUSES OF DEATH

In 2019, 2,854,838 individuals died in the United States (**Table 12-1**). Approximately 74% of these deaths occurred in those aged ≥65 years. The epidemiology of death has changed significantly since 1900 and even since 1980. In 1900, heart disease caused ~8% of all deaths, and cancer accounted for <4% of all deaths. In 1980, heart disease accounted for 38.2% of all deaths, cancer 20.9%, and cerebrovascular disease 8.6% of all deaths. By 2019, there had been a dramatic drop in deaths from cardiovascular and cerebrovascular diseases. In 2019, 23.1% of all deaths were from cardiovascular disease and just 5.3% from cerebrovascular disease. Deaths attributable to cancer, however, had increased slightly to 21.0%. The proportions of deaths due to chronic lower respiratory disease, diabetes, Alzheimer’s, and suicide have increased. Interestingly, in 2019, HIV/AIDS accounted for <0.18% of all U.S. deaths. While unlikely to continue being a leading cause of death in the future, COVID-19 was also the cause for >600,000 deaths in 2020–2021, and the official figure is almost certainly an undercount of the actual death toll.

TABLE 12-1 Ten Leading Causes of Death in the United States and Britain

CAUSE OF DEATH	UNITED STATES (2019)		ENGLAND AND WALES (2019)	
	NUMBER OF DEATHS, ALL AGES (%)	NUMBER OF DEATHS, PEOPLE ≥65 YEARS OF AGE	NUMBER OF DEATHS, ALL AGES (%)	NUMBER OF DEATHS, PEOPLE ≥65 YEARS OF AGE
All deaths	2,854,838	2,117,332	530,841	449,047
Heart disease ^a	659,041 (23.1)	531,583 (25.1)	87,095 (16.4)	74,967 (16.7)
Malignant neoplasms	599,601 (21.0)	435,462 (20.6)	147,419 (27.8)	118,982 (26.5)
Chronic lower respiratory diseases	156,979 (5.5)	133,246 (6.3)	31,221 (5.9)	28,235 (6.3)
Accidents	173,040 (6.1)	60,527 (2.9)	15,141 (2.9)	8999 (2.0)
Cerebrovascular diseases	150,005 (5.3)	129,193 (6.1)	29,816 (5.6)	27,210 (6.0)
Alzheimer's disease	121,499 (4.3)	120,090 (5.7)	20,400 (3.8)	20,279 (4.5)
Diabetes mellitus	87,647 (3.1)	62,397 (2.9)	6528 (1.2)	5552 (1.2)
Influenza and pneumonia	49,783 (1.7)	40,399 (1.9)	26,398 (5.0)	24,269 (5.4)
Nephritis, nephritic syndrome, nephrosis	51,565 (1.8)	42,230 (2.0)	3575 (0.7)	3323 (0.7)
Intentional self-harm	47,511 (1.7)	—	4832 (0.9)	751 (0.2)

^aCalculated using International Classification of Diseases codes I00–I09, I11, I13, I20–I51.

Source: National Center for Health Statistics (United States, 2019), <http://www.cdc.gov/nchs>; National Statistics (Great Britain, 2019), <http://www.statistics.gov.uk>.

This change in the epidemiology of death is also reflected in the costs of illness. In the United States, ~84% of all health care spending goes to patients with chronic illnesses, and 12% of total personal health care spending—slightly less than \$400 billion in 2015—goes to the 0.83% of the population in the last year of their lives.

In upper-middle- and upper-income countries, an estimated 70% of all deaths are preceded by a disease or condition, making it reasonable to plan for dying in the foreseeable future. Cancer has served as the paradigm for terminal care, but it is not the only type of illness with a recognizable and predictable terminal phase. Since heart failure, chronic obstructive pulmonary disease (COPD), chronic liver failure, dementia, and many other conditions have recognizable terminal phases, a systematic approach to end-of-life care should be part of all medical specialties. Many patients with chronic illness-related symptoms and suffering also can benefit from palliative care regardless of prognosis. Ideally, palliative care should be considered part of comprehensive care for all chronically ill patients. Strong evidence demonstrates that palliative care can be improved by coordination between caregivers, doctors, and patients for advance care planning, as well as dedicated teams of physicians, nurses, and other providers.

SITE OF DEATH

Where patients die varies by country. In Belgium and Canada, for instance, over half of all cancer patients still die in the hospital. The past few decades have seen a steady shift, both in the United States and other countries like the Netherlands, out of the hospital, as patients and their families list their own homes as the preferred site of death. In the early 1980s, ~70% of American cancer patients died in the hospital. Today, that percentage is ~25% (Fig. 12-1). A recent report shows that since 2000, there has been a shift in the United States from inpatient to home deaths, especially for patients with cancer, COPD, and dementia. For instance, among Medicare beneficiaries, 30.1% of deaths due to cancer in 2000 occurred in acute care hospitals; by 2009, this figure had dropped to 22.1%; by 2015, it was 19.8%.

Paradoxically, while deaths in acute care hospitals have declined in the United States since 2000, both hospitalizations in the last 90 days of life and—even more troublingly—admission to the intensive care unit (ICU) in the last 30 days have actually increased. Over 40% of cancer patients in the United States are admitted to the ICU in their last 6 months of life, and >25% of cancer patients are admitted to the hospital in the last 30 days.

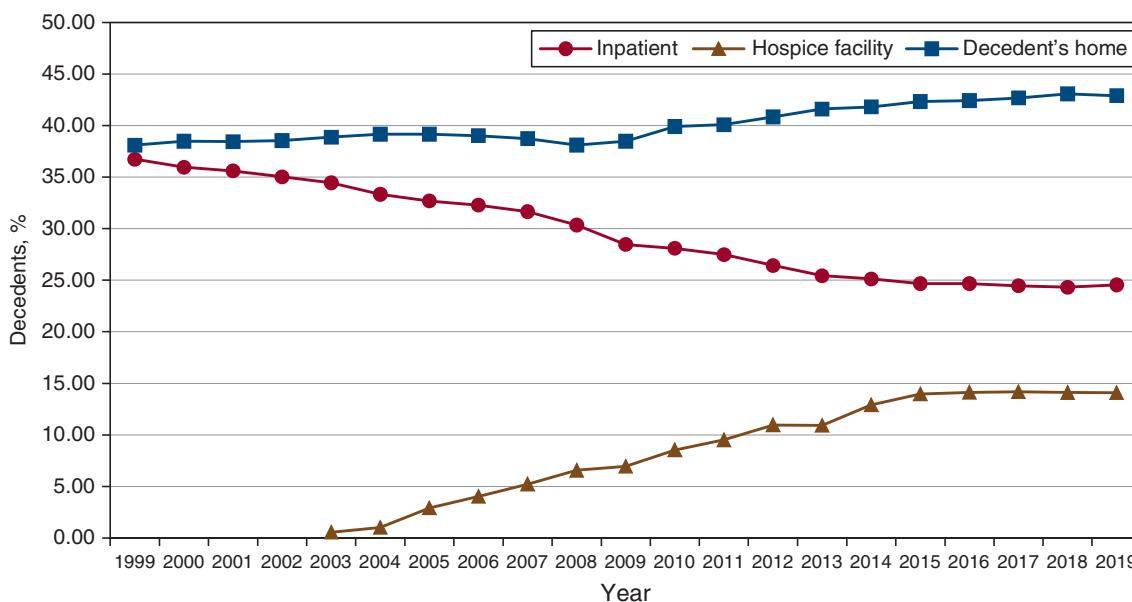


FIGURE 12-1 Graph showing trends in cancer decedents' site of death 1999–2019. (Source: Centers for Disease Control and Prevention, National Center for Health Statistics. Underlying Cause of Death 1999–2019 on CDC WONDER Online Database. <http://wonder.cdc.gov>.)

The shift in deaths out of the hospital has been accompanied by an increase in the use of hospice in the United States. In 2000, 21.6% of Medicare decedents used hospice at the time of death; by 2009, 42.2% were using hospice; and by 2018, 50.7% of Medicare decedents were enrolled in hospice at the time of death. Among cancer patients, ~60% were using hospice at the time of death. Hospice is also increasingly being used by noncancer patients. Today, cancer patients constitute ~20% of hospice users. But since 2014, the proportion of patients with other diagnoses using hospice has grown substantially, including those with circulatory/heart disease (17.4% in 2018 vs 13.8% in 2014), stroke (9.5% vs 6.2%), and respiratory disease (11.0% vs 9.4%). Of 2018 Medicare hospice decedents, 51.5% died at home, 17.4% in a nursing facility, 12.8% in a hospice inpatient facility, and 12.3% in assisted living.

Unfortunately, significant racial disparities exist in end-of-life care and the use of hospice, especially for noncancer deaths. Racial and ethnic minorities are less likely to receive hospice services than white decedents and are more likely to receive invasive or aggressive care in end-of-life treatment. Of people who died of head and neck cancers between 1999 and 2017, African Americans and Asians/Pacific Islanders were less likely to die at home or in hospice. Among Medicare beneficiaries who had a pancreatectomy for pancreatic cancer and lived at least 30 days, racial and ethnic minority patients remained 22% less likely than white patients to initiate hospice before death.

In 2008, for the first time, the American Board of Medical Specialties (ABMS) offered certification in hospice and palliative medicine. With the shortening of hospital stays, many serious conditions are now being treated at home or on an outpatient basis. Consequently, providing optimal palliative and end-of-life care requires ensuring that appropriate services are available in a variety of settings, including noninstitutional settings.

HOSPICE AND THE PALLIATIVE CARE FRAMEWORK

Central to this type of care is an interdisciplinary team approach that typically encompasses pain and symptom management, spiritual and psychological care for the patient, and support for family caregivers during the patient's illness and the bereavement period.

One of the more important changes in this field is beginning palliative care many months before death in order to focus on symptom relief and then switching to hospice in the patient's last few months. This approach avoids leaving hospice until the very end by introducing palliative care earlier, thereby allowing patients and families time to accommodate and transition. Phasing palliative care into end-of-life care means that patients will often receive palliative interventions long before they are formally diagnosed as terminally ill, or likely to die within 6 months.

Fundamental to ensuring quality palliative and end-of-life care is a focus on four broad domains: (1) physical symptoms; (2) psychological symptoms; (3) social needs that include interpersonal relationships, caregiving, and economic concerns; and (4) existential or spiritual needs.

ASSESSMENT AND CARE PLANNING

Comprehensive Assessment Standardized methods for conducting a comprehensive assessment focus on evaluating the patient's condition in all four domains affected by the illness: physical, psychological, social, and spiritual.

A comprehensive assessment should follow a modified version of the traditional medical history and physical examination and should emphasize both physical and mental symptoms. Questions should aim to elucidate symptoms, discern sources of suffering, and gauge how much those symptoms interfere with the patient's quality of life. Standardized and repeated assessments to evaluate the effectiveness of interventions are critical. Thus, clinicians should use shorter, validated instruments, such as (1) the revised Edmonton Symptom Assessment Scale; (2) Condensed Memorial Symptom Assessment Scale (MSAS); (3) MD Anderson Brief Symptom Inventory; (4) Rotterdam Symptom Checklist; (5) Symptom Distress Scale; (6) Patient-Reported Outcomes Measurement Information System; and (7) Interactive Symptom Assessment and Collection (ISAAC) tool.

MENTAL HEALTH With respect to mental health, many practices use the Patient Health Questionnaire-9 (PHQ-9) to screen for depression and the Generalized Anxiety Disorder-7 (GAD-7) to screen for anxiety. Using such tools ensures that the assessment is comprehensive and does not focus excessively on only pain.

INVASIVE TESTS Invasive tests are best avoided in end-of-life care, and even minimally invasive tests should be evaluated carefully for their benefit-to-burden ratio for the patient. Aspects of the physical examination that are uncomfortable and unlikely to yield useful information that change patient management should be omitted.

SOCIAL NEEDS Health care providers should also assess the status of important relationships, financial burdens, caregiving needs, and access to medical care. Relevant questions include the following: *How often is there someone to feel close to? How has this illness been for your family? How has it affected your relationships? How much help do you need with things like getting meals and getting around? How much trouble do you have getting the medical care you need?*

EXISTENTIAL NEEDS To determine a patient's existential needs, providers should assess distress, the patient's sense of emotional and existential well-being, and whether the patient believes he or she has found purpose or meaning. Helpful assessment questions can include the following: *How much are you able to find meaning since your illness began? What things are most important to you at this stage?*

PERCEPTION OF CARE In addition, it can be helpful to ask how the patient perceives his or her care: *How much do you feel your doctors and nurses respect you? How clear is the information from us about what to expect regarding your illness? How much do you feel that the medical care you are getting fits with your goals?* If concern is detected in any of these areas, deeper evaluative questions are warranted.

Communication Particularly when an illness is life-threatening, there exists the potential for many emotionally charged and potentially conflict-creating moments—collectively called “bad news” situations—in which empathic and effective communication skills are essential. Those moments include the sharing of a terminal diagnosis with the patient and/or family, the discussion of the patient's prognosis and any treatment failures, the consideration of deemphasizing efforts to cure and prolong life while focusing more on symptom management and palliation, advance care planning, and the patient's actual death. Although these conversations can be difficult, research indicates that end-of-life discussions can lead to earlier hospice referrals, rather than overly aggressive treatment, ultimately benefiting quality of life for patients and improving the bereavement process for families.

Just as surgeons prepare for major operations and investigators rehearse a presentation of research results, physicians and health care providers caring for patients with significant or advanced illnesses should develop a standardized approach for sharing important information and planning interventions. In addition, physicians must be aware that families often care not only about how prepared the physician was to deliver bad news, but also the setting in which it was delivered. For instance, one study found that 27% of families making critical decisions for patients in an ICU desired better and more private physical space to communicate with physicians.

One structured seven-step procedure for communicating bad news goes by the acronym P-SPIKES: (1) prepare for the discussion, (2) set up a suitable environment, (3) begin the discussion by finding out what the patient and/or family understand, (4) determine how they will comprehend new information best and how much they want to know, (5) provide needed new knowledge accordingly, (6) allow for emotional responses, and (7) share plans for the next steps in care (**Table 12-2**).

Continuous Goal Assessment Major barriers to providing high-quality palliative and end-of-life care include the difficulty in determining an accurate prognosis and the emotional resistance of patients and their families to accepting the implications of a poor prognosis. A practical solution to these barriers is to integrate palliative care interventions or home visits from a palliative care visiting nurse months before the estimated final 6 months of life. Under this

TABLE 12-2 Elements of Communicating Bad News—The P-SPIKES Approach

ACRONYM	STEPS	AIM OF THE INTERACTION	PREPARATIONS, QUESTIONS, OR PHRASES
P	Preparation	Mentally prepare for the interaction with the patient and/or family.	Review what information needs to be communicated. Plan how you will provide emotional support. Rehearse key steps and phrases in the interaction.
S	Setting of the interaction	Ensure the appropriate setting for a serious and potentially emotionally charged discussion.	Ensure that patient, family, and appropriate social supports are present. Devote sufficient time. Ensure privacy and prevent interruptions by people or beeper. Bring a box of tissues.
P	Patient's perception and preparation	Begin the discussion by establishing the baseline and whether the patient and family can grasp the information. Ease tension by having the patient and family contribute.	Start with open-ended questions to encourage participation. Possible questions to use: <i>What do you understand about your illness? When you first had symptom X, what did you think it might be? What did Dr. X tell you when he or she sent you here? What do you think is going to happen?</i>
I	Invitation and information needs	Discover what information needs the patient and/or family have and what limits they want regarding the bad information.	Possible questions to use: <i>If this condition turns out to be something serious, do you want to know? Would you like me to tell you all the details of your condition? If not, who would you like me to talk to?</i>
K	Knowledge of the condition	Provide the bad news or other information to the patient and/or family sensitively.	Do not just dump the information on the patient and family. Check for patient and family understanding. Possible phrases to use: <i>I feel badly to have to tell you this, but... Unfortunately, the tests showed... I'm afraid the news is not good...</i>
E	Empathy and exploration	Identify the cause of the emotions—e.g., poor prognosis. Empathize with the patient's and/or family's feelings. Explore by asking open-ended questions.	Strong feelings in reaction to bad news are normal. Acknowledge what the patient and family are feeling. Remind them such feelings are normal, even if frightening. Give them time to respond. Remind the patient and family you won't abandon them. Possible phrases to use: <i>I imagine this is very hard for you to hear. You look very upset. Tell me how you are feeling. I wish the news were different. We'll do whatever we can to help you.</i>
S	Summary and planning	Delineate for the patient and the family the next steps, including additional tests or interventions.	It is the unknown and uncertain that can increase anxiety. Recommend a schedule with goals and landmarks. Provide your rationale for the patient and/or family to accept (or reject). If the patient and/or family are not ready to discuss the next steps, schedule a follow-up visit.

Source: Adapted from R Buckman: *How to Break Bad News: A Guide for Health Care Professionals*. Baltimore, Johns Hopkins University Press, 1992.

approach, palliative care no longer conveys the message of failure, having no more treatments, or “giving up hope.” The transition from palliative to end-of-life care or hospice also feels less hasty and unexpected to the family. Fundamental to integrating palliative care with curative therapy is the inclusion of a continuous goal assessment as part of the routine patient reassessments that occur at most patient-physician encounters.

Goals for care are numerous, ranging from curing a specific disease, to prolonging life, to relieving a particular symptom, to adapting to a progressive disability without disrupting the family, to finding peace of mind or personal meaning, to dying in a manner that leaves loved ones with positive memories. Discerning a patient’s goals for care can be approached through a seven-step protocol: (1) ensure that medical and other information is as complete as reasonably possible and is understood by all relevant parties (see above); (2) explore what the patient and/or family is hoping for, while also identifying relevant and realistic goals; (3) share all the options with the patient and family; (4) respond with empathy as they adjust to changing expectations; (5) make a plan that emphasizes what can be done to achieve the realistic goals; (6) follow through with the plan; and (7) periodically review the plan and consider at every encounter whether the goals of care should be revised with the patient and/or family. Each of these steps need not be followed in rote order, but together they provide a helpful framework for interactions with patients and their families regarding

their goals for care. Such interactions can be especially challenging if a patient or family member has difficulty letting go of an unrealistic goal. In such cases, the provider should help them refocus on more realistic goals and should also suggest that while it is fine to hope for the best, it is still prudent to plan for other outcomes as well.

Advance Care Planning • **PRACTICES** Advance care planning is the process of planning for future medical care in case the patient becomes incapable of making medical decisions. A 2010 study of adults aged ≥60 who died between 2000 and 2006 found that while 42% of adults were required to make treatment decisions in their final days of life, 70% lacked decision-making capacity. Among those lacking decision-making capacity, approximately one-third did not have advance planning directives. Ideally, such planning would occur before a health care crisis or the terminal phase of an illness. Unfortunately, diverse barriers prevent this. Approximately 80% of Americans endorse advance care planning and living wills. However, according to a 2013 Pew survey, only 35% of adults have written down their end-of-life wishes. Other studies report that even fewer Americans—with some estimates as low as 26% of adults—have filled out advance care directives. A review of studies suggests that the percentage of Americans who had written advance directives did not change between 2011 and 2016 and remains slightly over one-third of Americans. Larger numbers of adults, between 50 and 70%, claim to have talked with someone

about their treatment wishes. Americans aged 65 and older are more likely to complete an advance directive compared to younger adults (46% vs 32%).

Effective advance care planning should follow six key steps: (1) introducing the topic, (2) structuring a discussion, (3) reviewing plans that have been discussed by the patient and family, (4) documenting the plans, (5) updating them periodically, and (6) implementing the advance care directives (**Table 12-3**). Two of the main barriers to advance care planning are problems in raising the topic and difficulty in structuring a succinct discussion. Raising the topic can be done efficiently as a routine matter, noting that it is recommended for all patients, analogous to purchasing insurance or estate planning. Many of the most difficult cases have involved unexpected, acute episodes of brain damage in young individuals.

Structuring a focused discussion is an important communication skill. To do so, a provider must first identify the health care proxy and recommend his or her involvement in the advance care planning process. Next, a worksheet must be selected that has been demonstrated to produce reliable and valid expressions of patient preferences, and the patient and proxy must be oriented to it. Such worksheets exist for both general and disease-specific situations. The provider should then discuss with the patient and proxy one example scenario to demonstrate how to think about the issues. It is often helpful to begin with a scenario in which the patient is likely to have settled preferences for care, such as being in a persistent vegetative state. Once the patient's preferences

for interventions in this scenario are determined, the provider should suggest that the patient and proxy discuss and complete the worksheet for each other. If appropriate, the patient and proxy should consider involving other family members in the discussion. During a subsequent return visit, the provider should go over the patient's preferences, checking and resolving any inconsistencies. After having the patient and proxy sign the document, the provider should place the document in the patient's medical chart and make sure that copies are provided to relevant family members and care sites. Since patients' preferences can change, these documents must be reviewed periodically.

TYPES OF DOCUMENTS Advance care planning documents are of two broad types. The first includes living wills, also known as instructional directives; these are advisory documents that describe the types of decisions that should direct a patient's care. Some are more specific, delineating different scenarios and interventions for the patient to choose from. Among these, some are for general use and others are designed for use by patients with a specific type of disease, such as cancer, renal failure, or HIV. Less specific directives can be general statements, such as not wanting life-sustaining interventions, or forms that describe the values that should guide specific discussions about terminal care. The second type of advance directive allows the designation of a health care proxy (sometimes also referred to as a durable attorney for health care), an individual selected by the patient to make decisions. The choice is not either/or; a combined directive that includes a living

TABLE 12-3 Steps in Advance Care Planning

STEP	GOALS TO BE ACHIEVED AND MEASURES TO COVER	USEFUL PHRASES OR POINTS TO MAKE
Introduce advance care planning	Ask the patient what he or she knows about advance care planning and if he or she has already completed an advance care directive.	<i>I'd like to talk with you about something I try to discuss with all my patients. It's called advance care planning. In fact, I feel that this is such an important topic that I have done this myself. Are you familiar with advance care planning or living wills?</i>
	Indicate that you as a physician have completed advance care planning.	<i>Have you thought about the type of care you would want if you ever became too sick to speak for yourself? That is the purpose of advance care planning.</i>
	Indicate that you try to perform advance care planning with all patients regardless of prognosis.	<i>There is no change in health that we have not discussed. I am bringing this up now because it is sensible for everyone, no matter how well or ill, old or young.</i>
	Explain the goals of the process as empowering the patient and ensuring that you and the proxy understand the patient's preferences.	<i>Have many copies of advance care directives available, including in the waiting room, for patients and families.</i>
	Provide the patient relevant literature, including the advance care directive that you prefer to use.	<i>Know resources for state-specific forms (available at www.nhpco.org).</i>
	Recommend the patient identify a proxy decision-maker who should attend the next meeting.	
Have a structured discussion of scenarios with the patient	Affirm that the goal of the process is to follow the patient's wishes if the patient loses decision-making capacity.	<i>Use a structured worksheet with typical scenarios.</i>
	Elicit the patient's overall goals related to health care.	<i>Begin the discussion with persistent vegetative state and consider other scenarios, such as recovery from an acute event with serious disability; then ask the patient about his or her preferences regarding specific interventions, such as ventilators, artificial nutrition, and CPR; finally, proceeding to less invasive interventions, such as blood transfusions and antibiotics.</i>
	Elicit the patient's preferences for specific interventions in a few salient and common scenarios.	
	Help the patient define the threshold for withdrawing and withholding interventions.	
	Define the patient's preference for the role of the proxy.	
Review the patient's preferences	After the patient has made choices of interventions, review them to ensure they are consistent and the proxy is aware of them.	
Document the patient's preferences	Formally complete the advance care directive and have a witness sign it.	
	Provide a copy for the patient and the proxy.	
	Insert a copy into the patient's medical record and summarize it in a progress note.	
Update the directive	Periodically, and with major changes in health status, review the directive with the patient and make any modifications.	
Apply the directive	The directive goes into effect only when the patient becomes unable to make medical decisions for himself or herself.	
	Reread the directive to be sure about its content.	
	Discuss your proposed actions based on the directive with the proxy.	

Abbreviation: CPR, cardiopulmonary resuscitation.

will and designates a proxy is often used, and the directive should indicate clearly whether the specified patient preferences or the proxy's choice takes precedence if they conflict. Some states have begun to put into practice a "Physician Orders for Life-Sustaining Treatment (POLST)" directive, which builds on communication between providers and patients by including guidance for end-of-life care in a color-coordinated form that follows the patient across treatment settings. The procedures for completing advance care planning documents vary according to state law.

A potentially misleading distinction relates to statutory, as opposed to advisory, documents. Statutory documents are drafted to fulfill relevant state laws. Advisory documents are drafted to reflect the patient's wishes. Both are legal, the former under state law and the latter under common or constitutional law.

LEGAL ASPECTS As of 2021, 48 states and the District of Columbia had enacted living will legislation. Massachusetts and Michigan are the two states without living will legislation. Indiana has a life-prolonging procedures declaration. States differ in the requirements for advanced directives, including whether they need to be witnessed and, if so, by how many witnesses and whether they need to be notarized. Importantly, in 25 states, the laws state that the living will is not valid if a woman is pregnant. All states except Alaska have enacted durable power of attorney for health care laws that permit patients to designate a proxy decision-maker with authority to terminate life-sustaining treatments. Only in Alaska does the law prohibit proxies from terminating life-sustaining treatments for pregnant women.

The U.S. Supreme Court has ruled that patients have a constitutional right to decide any issues related to refusing or terminating medical interventions, including life-sustaining interventions, and that mentally incompetent patients can exercise this right by providing "clear and convincing evidence" of their preferences. Since advance care directives permit patients to provide such evidence, commentators agree that they are constitutionally protected. Most commentators believe that a state is required to honor any clear advance care directive, regardless of whether it is written on an "official" form. Many states have enacted laws for the explicit purpose of honoring out-of-state directives. If a patient is not using a statutory form, it may be advisable to attach a statutory form to the advance care directive being used. State-specific forms are readily available free of charge for health care providers, patients, and families through the website of the National Hospice and Palliative Care Organization (<http://www.nhpco.org>).

REIMBURSEMENT As of January 1, 2016, the Centers for Medicare and Medicaid Services amended the physician fee schedule to reimburse discussions of advance care planning under Current Procedural Terminology codes 99497 and 99498. The session must be voluntary and include an explanation of advance care planning but need not include a completed advance care document. There can be multiple bills for the discussion if it extends over several encounters. A study found that patients who engaged in a billed advance care planning encounter were more likely to be enrolled in hospice and less likely to receive intensive therapies, despite being more likely to be hospitalized in the ICU. However, a billing incentive in and of itself may not increase advance care planning discussions by clinicians. In 2016, just 1.6% of Medicare Advantage patients had a discussion of advance care planning that was billed. Factors beyond reimbursement, such as clinicians' lack of comfort and skill in carrying out advance care planning discussions and lack of time, appear to impede discussions of advance care planning.

INTERVENTIONS

■ PHYSICAL SYMPTOMS AND THEIR MANAGEMENT

Great emphasis has been placed on addressing dying patients' pain. In order to emphasize its importance, pain assessment has frequently been included as the fifth vital sign. Heightened consideration of pain has been advocated by large health care systems such as the Veterans' Administration and accrediting bodies such as The Joint Commission. Although this embrace of pain has been symbolically important,

TABLE 12-4 Common Physical and Psychological Symptoms of Terminally Ill Patients

PHYSICAL SYMPTOMS	PSYCHOLOGICAL SYMPTOMS
Pain	Anxiety
Fatigue and weakness	Depression
Dyspnea	Hopelessness
Insomnia	Meaninglessness
Dry mouth	Irritability
Anorexia	Impaired concentration
Nausea and vomiting	Confusion
Constipation	Delirium
Cough	Loss of libido
Swelling of arms or legs	
Itching	
Diarrhea	
Dysphagia	
Dizziness	
Fecal and urinary incontinence	
Numbness/tingling in hands/feet	

available data suggest that making pain the fifth vital sign does not lead to improved pain management practices. In light of the opioid crisis in the United States, the emphasis on pain management has begun to be reexamined. For instance, in 2017 draft standards, The Joint Commission recommends nonpharmacologic pain treatment as well as identification of psychosocial risk factors for addiction. Importantly, good palliative care requires much more than good pain management. The frequency of symptoms varies by disease and other factors. The most common physical and psychological symptoms among all terminally ill patients include pain, fatigue, insomnia, anorexia, dyspnea, depression, anxiety, nausea, and vomiting. In the last days of life, terminal delirium is also common. Assessments of patients with advanced cancer have shown that patients experienced an average of 11.5 different physical and psychological symptoms (**Table 12-4**).

In the vast majority of cases, evaluations to determine the etiology of these symptoms should be limited to the history and physical examination. In some cases, radiologic or other diagnostic examinations will provide sufficient benefit in directing optimal palliative care to warrant the risks, potential discomfort, and inconvenience, especially to a seriously ill patient. Only a few of the common symptoms that present difficult management issues will be addressed in this chapter. **Additional information on the management of other symptoms, such as nausea and vomiting, insomnia, and diarrhea, can be found in Chaps. 45, 31, and 46, respectively. Information on the management of patients with cancer is provided in Chap. 69.**

Pain • FREQUENCY The frequency of pain among terminally ill patients varies significantly. Cancer (~85%), congestive heart failure (CHF; ~75%), and AIDS have been associated with a higher prevalence of pain compared to other advanced illnesses, such as COPD (~45%), chronic kidney disease (~40%), and dementia (~40%). One meta-analysis of adults with advanced or terminal illness found pain prevalence of 30–94% in patients with cancer, compared to 21–77% for COPD, 14–78% for CHF, 11–83% for end-stage renal disease, 14–63% for dementia, and 30–98% for AIDS.

ETIOLOGY There are two types of pain: nociceptive and neuropathic. Nociceptive pain is further divided into somatic or visceral pain. *Somatic pain* is the result of direct mechanical or chemical stimulation of nociceptors and normal neural signaling to the brain. It tends to be localized, aching, throbbing, and cramping. The classic example is bone metastases. *Visceral pain* is caused by nociceptors in gastrointestinal (GI), respiratory, and other organ systems. It is a deep or colicky type of pain classically associated with pancreatitis, myocardial infarction, or tumor invasion of viscera. *Neuropathic pain* arises from

disordered nerve signals. It is described by patients as burning, electrical, or shock-like pain. Classic examples are post-stroke pain, tumor invasion of the brachial plexus, and herpetic neuralgia.

ASSESSMENT Pain is a subjective experience. Depending on the patient's circumstances, perspective, and physiologic condition, the same physical lesion or disease state can produce different levels of reported pain and need for pain relief. Systematic assessment includes eliciting the following: (1) type: throbbing, cramping, burning, etc.; (2) periodicity: continuous, with or without exacerbations, or incident; (3) location; (4) intensity; (5) modifying factors; (6) effects of treatments; (7) functional impact; and (8) impact on patient. Several validated pain assessment measures may be used, including the Visual Analogue Scale (VAS), the Brief Pain Inventory (BPI), or the Numerical Pain Rating Scale (NRS-11). Other scales have been developed for neuropathic pain, such as the Neuropathic Pain Scale and the DN4 Questionnaire. Frequent reassessments on a consistent scale are essential to assess the impact of and need to readjust interventions.

INTERVENTIONS Interventions for pain must be tailored to each individual, with the goal of preempting chronic pain and relieving breakthrough pain. At the end of life, there is rarely reason to doubt a patient's report of pain. With the opioid crisis in the United States, there is more emphasis on making opioids one component of multimodal analgesia. Nevertheless, at the end of life, pain medications, especially opioids, remain the cornerstone of management (Fig. 12-2). If they are failing and nonpharmacologic interventions—including radiotherapy and anesthetic or neurosurgical procedures such as peripheral nerve blocks or epidural medications—are required, a pain consultation is appropriate.

Pharmacologic interventions still largely follow the World Health Organization three-step, "analgesic ladder" approach, which involves nonopioid analgesics, "mild" opioids, and "strong" opioids, with or without adjuvants (Chap. 13). Nonopioid analgesics, especially nonsteroidal anti-inflammatory drugs (NSAIDs), are the initial treatments for mild pain. They work primarily by inhibiting peripheral prostaglandins

and reducing inflammation but may also have central nervous system (CNS) effects. Additionally, NSAIDs have a ceiling effect. Ibuprofen, up to 2400 mg/d qid, has a minimal risk of causing bleeding and renal impairment and is a good initial choice. In patients with a history of severe GI or other bleeding, however, ibuprofen should be avoided. In patients with a history of mild gastritis or gastroesophageal reflux disease (GERD), acid-lowering therapy, such as a proton pump inhibitor, should be used. Acetaminophen is an alternative in patients with a history of GI bleeding and can be used safely at up to 4 g/d qid. In patients with liver dysfunction due to metastases or other causes and in patients with heavy alcohol use, doses should be reduced.

If nonopioid analgesics are insufficient, opioids should be introduced. Opioids primarily work by interacting with μ opioid receptors to activate pain-inhibitory neurons in the CNS, although they also interact variably with δ and κ receptors. Receptor agonists, such as morphine, codeine, and fentanyl, produce analgesia by activating pain-inhibitory neurons in the CNS. Partial agonists, such as buprenorphine, have a ceiling effect for analgesia and a lower potential for abuse. They are useful for postacute pain but should not be used for chronic pain in end-of-life care. Pure antagonists, such as naloxone and methylnaltrexone, are used for reversal of opioid effects.

Traditionally, "weak" opioids such as codeine were used first. If they failed to relieve pain after dose escalation, "strong" opioids like morphine were used in doses of 5–10 mg every 4 h. However, this breakdown between "weak" and "strong" opioids is no longer commonly accepted, with smaller doses of "stronger" opioids frequently being preferred over similar or larger doses of "weaker" opioids, and different pain syndromes having different preferred therapies. Regardless, non-opioid analgesics should be combined with opioids, as they potentiate the effect of opioids.

Importantly, the goal is to prevent patients from experiencing pain. Consequently, for continuous pain, opioids should be administered on a regular, around-the-clock basis consistent with their duration of analgesia, and the next dose should occur before the effect of the previous dose wears off. They should not be provided only when the patient

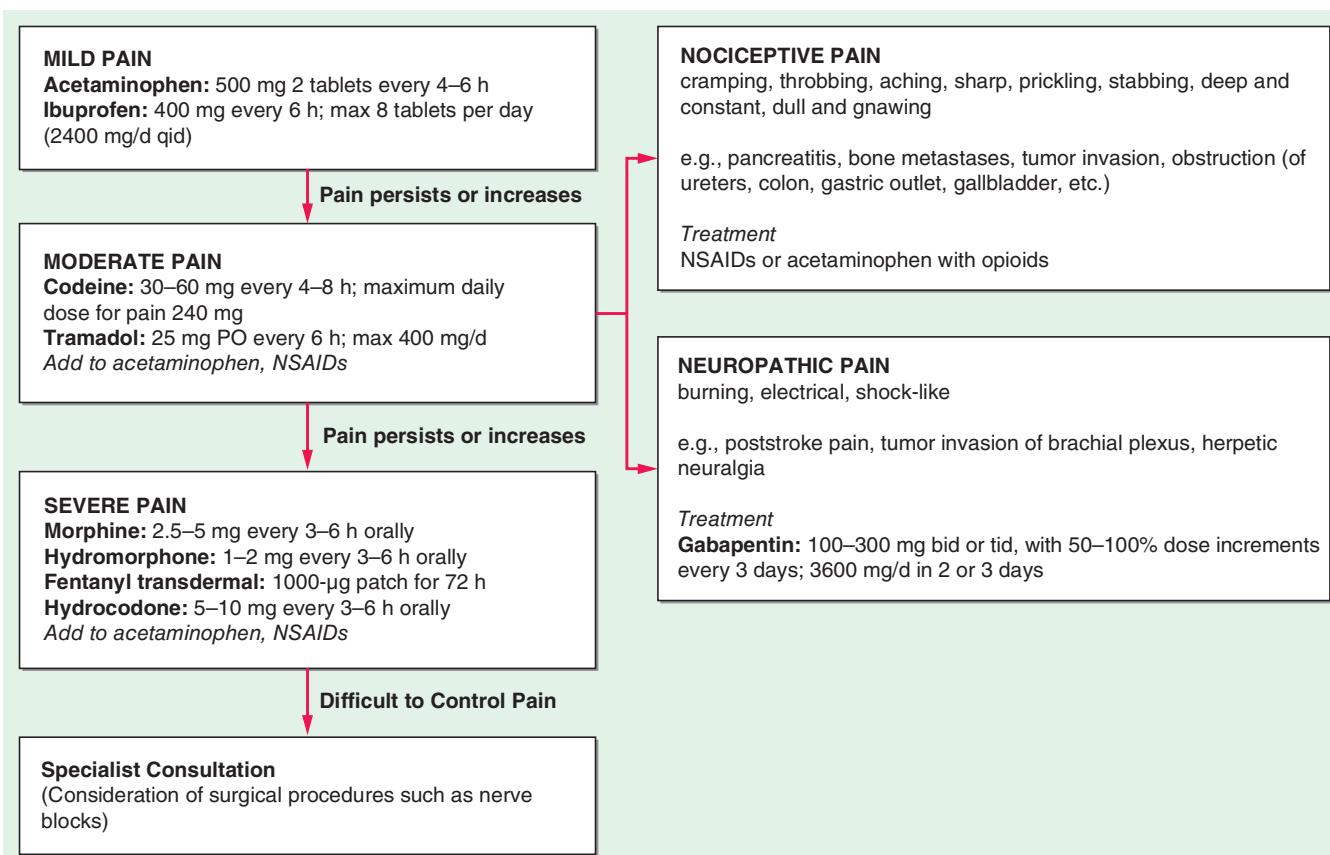


FIGURE 12-2 Terminal pain management flow chart. NSAIDs, nonsteroidal anti-inflammatory drugs.

experiences pain. Patients should also be provided rescue medication, such as liquid morphine, for breakthrough pain, generally at 20% of the baseline dose. Patients should be informed that using the rescue medication does not obviate the need to take the next standard dose of pain medication. If the patient's pain remains uncontrolled after 24 h and recurs before the next dose, requiring the patient to utilize the rescue medication, the daily opioid dose can be increased by the total dose of rescue medications used by the patient, or by 50% of the standing opioid daily dose for moderate pain and 100% for severe pain.

It is inappropriate to start with extended-release preparations. Instead, an initial focus on using short-acting preparations to determine how much is required in the first 24–48 h will allow clinicians to determine opioid needs. Once pain relief is obtained using short-acting preparations, the switch should be made to extended-release preparations. Even with a stable extended-release preparation regimen, the patient may experience incident pain, such as during movement or dressing changes. Short-acting preparations should be taken before such predictable episodes. Although less common, patients may have "end-of-dose failure" with long-acting opioids, meaning that they develop pain after 8 h in the case of an every-12-h medication. In these cases, a trial of giving an every-12-h medication every 8 h is appropriate.

Due to differences in opioid receptors, cross-tolerance among opioids is incomplete, and patients may experience different side effects with different opioids. Therefore, if a patient is not experiencing pain relief or is experiencing too many side effects, a change to another opioid preparation is appropriate. When switching, one should begin with 50–75% of the published equianalgesic dose of the new opioid.

Unlike NSAIDs, opioids have no ceiling effect; therefore, there is no maximum dose, no matter how many milligrams the patient is receiving. The appropriate dose is the dose needed to achieve pain relief. This is an important point for clinicians to explain to patients and families. Addiction or excessive respiratory depression is extremely unlikely in the terminally ill; fear of these side effects should neither prevent escalating opioid medications when the patient is experiencing insufficient pain relief nor justify using opioid antagonists.

Opioid side effects should be anticipated and treated preemptively. Nearly all patients experience constipation that can be debilitating (see below). Failure to prevent constipation often results in noncompliance with opioid therapy. The preferred treatment is prevention. Cathartics (senna 2 tbsp qHS), stool softeners (docusate 100 mg PO qd), and/or laxatives (laxitolose 30 mL qd) are considered first-line treatment. For refractory cases, opioid antagonists or other therapies, such as lubiprostone, should be considered.

Methylnaltrexone is the best-studied opioid antagonist for use in refractory opioid-induced constipation. It reverses opioid-induced constipation by blocking peripheral opioid receptors, but not central receptors, for analgesia. In placebo-controlled trials, it has been shown to cause laxation within 24 h of administration. As with the use of opioids, about a third of patients using methylnaltrexone experience nausea and vomiting, but unlike with opioid usage, tolerance usually develops within a week. Therefore, when one is beginning opioids, an antiemetic such as metoclopramide or a serotonin antagonist is often prescribed prophylactically and stopped after 1 week. Olanzapine has also been shown to have antinausea properties and can be effective in countering delirium or anxiety, with the advantage of some weight gain.

Drowsiness, a common side effect of opioids, also usually abates within a week. For refractory or severe cases, pharmacologic therapy should be considered. The best-studied agents are the psychostimulants dextroamphetamine, methylphenidate, and modafinil, although evidence regarding their efficacy is weak. Modafinil has the advantage of once-a-day dosing compared to methylphenidate's twice daily dosing.

Seriously ill patients who require chronic pain relief rarely become addicted. Suspicion of addiction should not be a reason to withhold pain medications from terminally ill patients. Nonetheless, patients and families may withhold prescribed opioids for fear of addiction or dependence. Physicians and health care providers should reassure patients and families that the patient will not become addicted to

opioids if they are used as prescribed for pain relief; this fear should not prevent the patient from taking the medications around the clock. However, diversion of drugs for use by other family members or illicit sale may occur. It may be necessary to advise the patient and caregiver about secure storage of opioids. Contract writing with the patient and family can help. If that fails, transfer to a safe facility may be necessary.

Tolerance describes the need to increase medication dosage for the same pain relief without a concurrent change in disease. In the case of patients with advanced disease, the need for increasing opioid dosage for pain relief usually is caused by disease progression rather than tolerance. Physical dependence is indicated by symptoms resulting from the abrupt withdrawal of opioids and should not be confused with addiction.

In recent years, the potential dangers of opioid drugs have become increasingly apparent. To help mitigate the risk of these powerful drugs, several strategies should be used to reduce the risk of aberrant drug use. To start, all patients should be assessed for their individual levels of risk. While there are multiple surveys available, including the Opioid Risk Tool, none have gained widespread use or validation. In general, however, it is important to screen for prior substance abuse and major psychiatric disorders.

For patients deemed to be high risk, a multidisciplinary effort should be pursued to reduce the risk of adverse consequences, such as addiction and diversion. Prescribing strategies include selecting opioids with longer durations of action and lower street values, such as methadone, and prescribing smaller quantities with more frequent follow-up. Monitoring options include periodic urine screening and referral to pain specialists. In some cases, it may also be reasonable to consider not offering short-acting opioids for breakthrough pain. In no situation, however, should adequate pain relief be withheld due to risk.

Adjuvant analgesic medications are nonopiods that potentiate the analgesic effects of opioids. They are especially important in the management of neuropathic pain. Gabapentin, an anticonvulsant initially studied in the setting of herpetic neuralgia, is now the first-line treatment for neuropathic pain resulting from a variety of causes. It is begun at 100–300 mg bid or tid, with 50–100% dose increments every 3 days. Usually 900–3600 mg/d in two or three doses is effective. The combination of gabapentin and nortriptyline may be more effective than gabapentin alone. Two potential side effects of gabapentin to be aware of are confusion and drowsiness, especially in the elderly. Other effective adjuvant medications include pregabalin, which has the same mechanism of action as gabapentin but is absorbed more efficiently from the GI tract. Lamotrigine is a novel agent whose mechanism of action is unknown but has been shown to be effective. It is recommended to begin at 25–50 mg/d, increasing to 100 mg/d. Carbamazepine, a first-generation agent, has been proven effective in randomized trials for neuropathic pain. Other potentially effective anticonvulsant adjuvants include topiramate (25–50 mg qd or bid, rising to 100–300 mg/d) and oxcarbazepine (75–300 mg bid, rising to 1200 mg bid).

Glucocorticoids, preferably dexamethasone given once a day, can be useful in reducing inflammation that causes pain, while also elevating mood, energy, and appetite. Its main side effects include confusion, sleep difficulties, and fluid retention. Glucocorticoids are especially effective for bone pain and abdominal pain from distention of the GI tract or liver. Other drugs, including clonidine and baclofen, can be effective in providing pain relief. These drugs are adjuvants and generally should be used in conjunction with—not instead of—opioids. Methadone, carefully dosed because of its unpredictable half-life in many patients, has activity at the N-methyl-D-aspartate (NMDA) receptor and is useful for complex pain syndromes and neuropathic pain. It is generally reserved for cases in which first-line opioids (morphine, oxycodone, hydromorphone) are either ineffective or unavailable.

Radiation therapy can treat bone pain from single metastatic lesions. Bone pain from multiple metastases can be amenable to radiopharmaceuticals such as strontium-89 and samarium-153. Bisphosphonates, such as pamidronate (90 mg every 4 weeks) and calcitonin (200 IU intranasally once or twice a day), also provide relief from bone pain but have multiday onsets of action.

Constipation • FREQUENCY Constipation is reported in up to 70–100% of patients requiring palliative care.

ETIOLOGY Although hypercalcemia and other factors can cause constipation, it is most frequently a predictable consequence of the use of opioids for pain and dyspnea relief and of the anticholinergic effects of tricyclic antidepressants, as well as due to the inactivity and poor diets common among seriously ill patients. If left untreated, constipation can cause substantial pain and vomiting and also is associated with confusion and delirium. Whenever opioids and other medications known to cause constipation are used, preemptive treatment for constipation should be instituted.

ASSESSMENT Assessing constipation can be difficult because people describe it differently. Four commonly used assessment scales are the Bristol Stool Form Scale, the Constipation Assessment Scale, the Constipation Visual Analogue Scale, and the Eton Scale Risk Assessment for Constipation. The Bowel Function Index can be used to quantify opioid-induced constipation. The physician should establish the patient's previous bowel habits, as well as any changes in subjective and objective qualities such as bloating or decreased frequency. Abdominal and rectal examinations should be performed to exclude impaction or an acute abdomen. Radiographic assessments beyond a simple flat plate of the abdomen in cases in which obstruction is suspected are rarely necessary.

INTERVENTION Any measure to address constipation during end-of-life care should include interventions to reestablish comfortable bowel habits and to relieve pain or discomfort. Although physical activity, adequate hydration, and dietary treatments with fiber can be helpful, each is limited in its effectiveness for most seriously ill patients, and fiber may exacerbate problems in the setting of dehydration or if impaired motility is the etiology. Fiber is contraindicated in the presence of opioid use. Stimulant and osmotic laxatives, stool softeners, fluids, and enemas are the mainstays of therapy (Table 12-5). To prevent constipation from opioids and other medications, a combination of a laxative and a stool softener (such as senna and docusate) should be used. If after several days of treatment a bowel movement has not occurred, a rectal examination to remove impacted stool and place a suppository is necessary. For patients with impending bowel

obstruction or gastric stasis, octreotide to reduce secretions can be helpful. For patients in whom the suspected mechanism is dysmotility, metoclopramide can be helpful.

Nausea • FREQUENCY Up to 70% of patients with advanced cancer have nausea, defined as the subjective sensation of wanting to vomit.

ETIOLOGY Nausea and vomiting are both caused by stimulation at one of four sites: the GI tract, the vestibular system, the chemoreceptor trigger zone (CTZ), and the cerebral cortex. Medical treatments for nausea are aimed at receptors at each of these sites: the GI tract contains mechanoreceptors, chemoreceptors, and 5-hydroxytryptamine type 3 (5-HT₃) receptors; the vestibular system probably contains histamine and acetylcholine receptors; and the CTZ contains chemoreceptors, dopamine type 2 receptors, and 5-HT3 receptors. An example of nausea that most likely is mediated by the cortex is anticipatory nausea before a dose of chemotherapy or other noxious stimuli.

Specific causes of nausea include metabolic changes (liver failure, uremia from renal failure, hypercalcemia), bowel obstruction, constipation, infection, GERD, vestibular disease, brain metastases, medications (including antibiotics, NSAIDs, proton pump inhibitors, opioids, and chemotherapy), and radiation therapy. Anxiety can also contribute to nausea.

INTERVENTION Medical treatment of nausea is directed at the anatomic and receptor-mediated cause revealed by a careful history and physical examination. When no specific cause of nausea is identified, many advocate beginning treatment with metoclopramide; a serotonin type 3 (5-HT₃) receptor antagonist such as ondansetron, granisetron, palonosetron, dolasetron, tropisetron, or ramosetron; or a dopamine antagonist such as chlorpromazine, haloperidol, or prochlorperazine. When decreased motility is suspected, metoclopramide can be an effective treatment. When inflammation of the GI tract is suspected, glucocorticoids, such as dexamethasone, are an appropriate treatment. For nausea that follows chemotherapy and radiation therapy, one of the 5-HT₃ receptor antagonists or neurokinin-1 antagonists, such as aprepitant or fosaprepitant, is recommended. Clinicians should attempt prevention of postchemotherapy nausea, rather than simply providing treatment after the fact. Current clinical guidelines recommend tailoring the strength of treatments to the specific emetic risk posed by a specific chemotherapy drug. When a vestibular cause (such as "motion sickness" or labyrinthitis) is suspected, antihistamines, such as meclizine (whose primary side effect is drowsiness), or anticholinergics, such as scopolamine, can be effective. In anticipatory nausea, patients can benefit from nonpharmacologic interventions, such as biofeedback and hypnosis. The most common pharmacologic intervention for anticipatory nausea is a benzodiazepine, such as lorazepam. As with antihistamines, drowsiness and confusion are the main side effects.

The use of medical marijuana or oral cannabinoids for palliative treatment of nausea is controversial, as there are no controlled trials showing its effectiveness for patients at the end of life. A 2015 meta-analysis showed "low-quality evidence suggesting that cannabinoids were associated with improvements in nausea and vomiting due to chemotherapy," and such treatments are not as good as 5-HT₃ receptor antagonists and can sometimes even cause cannabis hyperemesis syndrome. Older patients, who compose the vast majority of dying patients, seem to tolerate cannabinoids poorly.

Dyspnea • FREQUENCY Dyspnea is the subjective experience of being short of breath. Over 50%, and as many as 75%, of dying patients, especially those with lung cancer, metastases to the lung, CHF, and COPD, experience dyspnea at some point near the end of life. Dyspnea is among the most distressing of physical symptoms and can be even more distressing than pain.

ASSESSMENT As with pain, dyspnea is a subjective experience that may not correlate with objective measures of PO₂, PCO₂, or respiratory rate. Consequently, measurements of oxygen saturation through pulse oximetry or blood gases are rarely helpful in guiding therapy. Despite the limitations of existing assessment methods, physicians should regularly assess and document patients' experience of dyspnea

TABLE 12-5 Medications for the Management of Constipation

INTERVENTION	DOSE	COMMENT
Stimulant laxatives		These agents directly stimulate peristalsis and may reduce colonic absorption of water.
Prune juice	120–240 mL/d	Work in 6–12 h.
Senna (Senokot)	2–8 tablets PO bid	
Bisacodyl	5–15 mg/d PO, PR	
Osmotic laxatives		These agents are not absorbed. They attract and retain water in the gastrointestinal tract.
Lactulose	15–30 mL PO q4–8h	Lactulose may cause flatulence and bloating.
Magnesium hydroxide (Milk of Magnesia)	15–30 mL/d PO	Lactulose works in 1 day, magnesium products in 6 h.
Magnesium citrate	125–250 mL/d PO	
Stool softeners		These agents work by increasing water secretion and as detergents, increasing water penetration into the stool.
Sodium docusate (Colace)	300–600 mg/d PO	Work in 1–3 days.
Calcium docusate	300–600 mg/d PO	
Suppositories and enemas		
Bisacodyl	10–15 PR qd	
Sodium phosphate enema	PR qd	Fixed dose, 4.5 oz, Fleet's.

and its intensity. Guidelines recommend visual analogue dyspnea scales to assess the severity of symptoms and the effects of treatment. Potentially reversible or treatable causes of dyspnea include infection, pleural effusions, pulmonary emboli, pulmonary edema, asthma, and tumor encroachment on the airway. However, the risk-versus-benefit ratio of the diagnostic and therapeutic interventions for patients with little time left to live must be considered carefully before undertaking diagnostic steps. Frequently, the specific etiology cannot be identified, and dyspnea is the consequence of progression of the underlying disease that cannot be treated. The anxiety caused by dyspnea and the choking sensation can significantly exacerbate the underlying dyspnea in a negatively reinforcing cycle.

INTERVENTIONS When reversible or treatable etiologies are diagnosed, they should be treated as long as the side effects of treatment, such as repeated drainage of effusions or anticoagulants, are less burdensome than the dyspnea itself. More aggressive treatments such as stenting a bronchial lesion may be warranted if it is clear that the dyspnea is due to tumor invasion at that site and if the patient and family understand the risks of such a procedure.

Usually, treatment will be symptomatic (**Table 12-6**). Supplemental oxygen does not appear to be effective. “A systematic review of the literature failed to demonstrate a consistent beneficial effect of oxygen inhalation over air inhalation for study participants with dyspnea due to end-stage cancer or cardiac failure.” Therefore, oxygen may be no more than an expensive placebo. Low-dose opioids reduce the sensitivity of the central respiratory center and relieve the sensation of dyspnea. If patients are not receiving opioids, weak opioids can be initiated; if patients are already receiving opioids, morphine or other stronger opioids should be used. Controlled trials do not support the use of nebulized opioids for dyspnea at the end of life. Phenothiazines and chlorpromazine may be helpful when combined with opioids. Benzodiazepines can be helpful in treating dyspnea, but only if anxiety is present. Benzodiazepines should not be used as first-line therapy or if there is no anxiety. If the patient has a history of COPD or asthma, inhaled bronchodilators and glucocorticoids may be helpful. If the patient has pulmonary edema due to heart failure, diuresis with a medication such as furosemide is indicated. Excess secretions can be transdermally or intravenously dried with scopolamine. More general interventions that medical staff can perform include sitting the patient upright, removing smoke or other irritants like perfume, ensuring a supply of fresh air with sufficient humidity, and minimizing other factors that can increase anxiety.

TABLE 12-6 Medications for the Management of Dyspnea

INTERVENTION	DOSE	COMMENTS
Weak opioids		For patients with mild dyspnea
Codeine (or codeine with 325 mg acetaminophen)	30 mg PO q4h	For opioid-naïve patients
Hydrocodone	5 mg PO q4h	
Strong opioids		For opioid-naïve patients with moderate to severe dyspnea
Morphine	5–10 mg PO q4h	For patients already taking opioids for pain or other symptoms
	30–50% of baseline opioid dose q4h	
Oxycodone	5–10 mg PO q4h	
Hydromorphone	1–2 mg PO q4h	
Anxiolytics		Give a dose every hour until the patient is relaxed; then provide a dose for maintenance
Lorazepam	0.5–2.0 mg PO/SL/IV qh then q4–6h	
Clonazepam	0.25–2.0 mg PO q12h	
Midazolam	0.5 mg IV q15min	

Fatigue • FREQUENCY Fatigue is one of the most commonly reported symptoms not only of cancer treatment but also of the palliative care of multiple sclerosis, COPD, heart failure, and HIV. More than 90% of terminally ill patients experience fatigue and/or weakness. Fatigue is frequently cited as one of the most distressing symptoms in these patients.

ETIOLOGY The multiple causes of fatigue in the terminally ill can be categorized as resulting from the underlying disease; from disease-induced factors such as tumor necrosis factor and other cytokines; and from secondary factors such as dehydration, anemia, infection, hypothyroidism, and drug side effects. In addition to low caloric intake, loss of muscle mass and changes in muscle enzymes may play an important role in fatigue during terminal illness. The importance of changes in the CNS, especially the reticular activating system, have been hypothesized based on reports of fatigue in patients receiving cranial radiation, experiencing depression, or having chronic pain in the absence of cachexia or other physiologic changes. Finally, depression and other causes of psychological distress can contribute to fatigue.

ASSESSMENT Like pain and dyspnea, fatigue is subjective, as it represents a patient’s sense of tiredness and decreased capacity for physical work. Objective changes, even in body mass, may be absent. Consequently, assessment must rely on patient self-reporting. Scales used to measure fatigue, such as the Edmonton Functional Assessment Tool, the Fatigue Self-Report Scales, and the Rhoten Fatigue Scale, are usually appropriate for research but not clinical purposes. In clinical practice, a simple performance assessment such as the Karnofsky performance status or the Eastern Cooperative Oncology Group (ECOG)’s question “How much of the day does the patient spend in bed?” may be the best measure. In the ECOG 0–4 performance status assessment, 0 = normal activity; 1 = symptomatic without being bedridden; 2 = requiring some, but <50%, bed time; 3 = bedbound more than half the day; and 4 = bedbound all the time. Such a scale allows for assessment over time and correlates with overall disease severity and prognosis. A 2008 review by the European Association of Palliative Care also described several longer assessment tools that contained 9–20 items, including the Piper Fatigue Inventory, the Multidimensional Fatigue Inventory, and the Brief Fatigue Inventory (BFI).

INTERVENTIONS Reversible causes of fatigue, such as anemia and infection, should be treated. However, at the end of life, it must be realistically acknowledged that fatigue will not be “cured.” The goal is to ameliorate fatigue and help patients and families adjust expectations. Behavioral interventions should be utilized to avoid blaming the patient for inactivity and to educate both the family and the patient that the underlying disease causes physiologic changes that produce low energy levels. Understanding that the problem is physiologic and not psychological can help alter expectations regarding the patient’s level of physical activity. Practically, this may mean reducing routine activities such as housework, cooking, and social events outside the house and making it acceptable to receive guests while lying on a couch. At the same time, the implementation of exercise regimens and physical therapy can raise endorphins, reduce muscle wasting, and decrease the risk of depression. In addition, ensuring good hydration without worsening edema may help reduce fatigue. Discontinuing medications that worsen fatigue may help, including cardiac medications, benzodiazepines, certain antidepressants, or opioids if the patient’s pain is well-controlled. As end-of-life care proceeds into its final stages, fatigue may protect patients from further suffering, and continued treatment could be detrimental.

Only a few pharmacologic interventions target fatigue and weakness. Randomized controlled trials suggest glucocorticoids can increase energy and enhance mood. Dexamethasone (8 mg/d) is preferred for its once-a-day dosing and minimal mineralocorticoid activity. Benefit, if any, is usually seen within the first month. For fatigue related to anorexia, megestrol (480–800 mg) can be helpful. Psychostimulants such as dextroamphetamine (5–10 mg PO) and methylphenidate (2.5–5 mg PO) may enhance energy levels, although controlled trials have not shown these drugs to be effective for fatigue induced by mild

to moderate cancer. Doses should be given in the morning and at noon to minimize the risk of counterproductive insomnia. Modafinil and armodafinil, developed for narcolepsy, have shown promise in the treatment of fatigue and have the advantage of once-daily dosing. Their precise role in fatigue at the end of life has not been documented but may be worth trying if other interventions are not beneficial. Anecdotal evidence suggests that L-carnitine may improve fatigue, depression, and sleep disruption.

PALLIATIVE SEDATION

Palliative sedation is used in distressing situations that cannot be addressed in other ways. When patients experience severe symptoms, such as pain or dyspnea, that cannot be relieved by conventional interventions or experience acute catastrophic symptoms, such as uncontrolled seizures, then palliative sedation should be considered as an intervention of last resort. It can be abused if done to hasten death (which it usually does not), when done at the request of the family rather than according to the patient's wishes, or when there are other interventions that could still be tried. The use of palliative sedation in cases of extreme existential or spiritual distress remains controversial. Typically, palliative sedation should be introduced only after the patient and family have been assured that all other interventions have been tried and after the patient and their loved ones have been able to "say goodbye."

Palliative sedation can be achieved by significantly increasing opioid doses until patients become unconscious and then putting them on a continuous infusion. Another commonly used medication for palliative sedation is midazolam at 1–5 mg IV every 5–15 min to calm the patient, followed by a continuous IV or subcutaneous infusion of 1 mg/h. In hospital settings, a continuous propofol infusion of 5 µg/kg per min can be used. There are also other, less commonly used medications for palliative sedation that include levomepromazine, chlorpromazine, and phenobarbital.

PSYCHOLOGICAL SYMPTOMS AND THEIR MANAGEMENT

Depression • FREQUENCY AND IMPACT Depression at the end of life presents an apparently paradoxical situation. Many people believe that depression is normal among seriously ill patients because they are dying. People frequently say, "Wouldn't you be depressed?" Although sadness, anxiety, anger, and irritability are normal responses to a serious condition, they are typically of modest intensity and transient. Persistent sadness and anxiety and the physically disabling symptoms that they can lead to are abnormal and suggestive of major depression. The precise number of terminally ill patients who are depressed is uncertain, primarily due to a lack of consistent diagnostic criteria and screening. Careful follow-up of patients suggests that while as many as 75% of terminally ill patients experience depressive symptoms, ~25% of terminally ill patients have major depression. Depression at the end of life is concerning because it can decrease the quality of life, interfere with closure in relationships and other separation work, obstruct adherence to medical interventions, and amplify the suffering associated with pain and other symptoms.

ETIOLOGY Previous history of depression, family history of depression or bipolar disorder, and prior suicide attempts are associated with increased risk for depression among terminally ill patients. Other symptoms, such as pain and fatigue, are associated with higher rates of depression; uncontrolled pain can exacerbate depression, and depression can cause patients to be more distressed by pain. Many medications used in the terminal stages, including glucocorticoids, and some anticancer agents, such as tamoxifen, interleukin 2, interferon α, and vincristine, also are associated with depression. Some terminal conditions, such as pancreatic cancer, certain strokes, and heart failure, have been reported to be associated with higher rates of depression, although this is controversial. Finally, depression may be attributable to grief over the loss of a role or function, social isolation, or loneliness.

ASSESSMENT Unfortunately, many studies suggest that most depressed patients at the end of life are not diagnosed, or if they are diagnosed,

they are not properly treated. Diagnosing depression among seriously ill patients is complicated, as many of the vegetative symptoms in the *DSM-V (Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition)* criteria for clinical depression—insomnia, anorexia and weight loss, fatigue, decreased libido, and difficulty concentrating—are associated with the process of dying itself. The assessment of depression in seriously ill patients therefore should focus on the dysphoric mood, helplessness, hopelessness, and lack of interest, enjoyment, and concentration in normal activities. It is now recommended that patients near the end of life should be screened with either the PHQ-9 or the PHQ-2, which asks "Over the past 2 weeks, how often have you been bothered by any of the following problems? (1) Little interest or pleasure in doing things and (2) feeling down, depressed or hopeless." The answer categories are as follows: not at all, several days, more than half the days, nearly every day. Other possible diagnostic tools include the short form of the Beck Depression Index or a visual analogue scale.

Certain conditions may be confused with depression. Endocrinopathies, such as hypothyroidism and Cushing's syndrome, electrolyte abnormalities, such as hypercalcemia, and akathisia, especially from dopamine-blocking antiemetics such as metoclopramide and prochlorperazine, can mimic depression and should be excluded.

INTERVENTIONS Undertreatment of depressed, terminally ill patients is common. Physicians must treat any physical symptom, such as pain, that may be causing or exacerbating depression. Fostering adaptation to the many losses that the patient is experiencing can also be helpful. Unfortunately, there are few randomized trials to guide such interventions. Thus, treatment typically follows the treatment used for non-terminally ill depressed patients.

In the absence of randomized controlled trials, nonpharmacologic interventions, including group or individual psychological counseling, and behavioral therapies such as relaxation and imagery can be helpful, especially in combination with drug therapy.

Pharmacologic interventions remain at the core of therapy. The same medications are used to treat depression in terminally ill as in non-terminally ill patients. Psychostimulants may be preferred for patients with a poor prognosis or for those with fatigue or opioid-induced somnolence. Psychostimulants are comparatively fast-acting, working within a few days instead of the weeks required for selective serotonin reuptake inhibitors (SSRIs). Dextroamphetamine or methylphenidate should be started at 2.5–5.0 mg in the morning and at noon, the same starting doses used for treating fatigue. The doses can eventually be escalated up to 15 mg bid. Modafinil is started at 100 mg qd and can be increased to 200 mg if there is no effect at the lower dose. Pemoline is a nonamphetamine psychostimulant with minimal abuse potential. It is also effective as an antidepressant beginning at 18.75 mg in the morning and at noon. Because it can be absorbed through the buccal mucosa, it is preferred for patients with intestinal obstruction or dysphagia. If it is used for prolonged periods, liver function must be monitored. The psychostimulants can also be combined with more traditional antidepressants while waiting for the antidepressants to become effective, then tapered down after a few weeks if necessary. Psychostimulants have side effects, particularly initial anxiety, insomnia, and very rarely paranoia, which may necessitate lowering the dose or discontinuing treatment.

Mirtazapine, an antagonist at the postsynaptic serotonin receptors, is a promising psychostimulant. It should be started at 7.5 mg before bed and titrated up no more than once every 1–2 weeks to a maximal dose of 45 mg/d. It has sedating, antiemetic, and anxiolytic properties, with few drug interactions. Its side effect of weight gain may be beneficial for seriously ill patients; it is available in orally disintegrating tablets.

For patients with a prognosis of several months or longer, SSRIs, including fluoxetine, sertraline, paroxetine, escitalopram, and citalopram, and serotonin-noradrenaline reuptake inhibitors, such as venlafaxine and duloxetine, are the preferred treatments, due to their efficacy and comparatively few side effects. Because low doses of these medications may be effective for seriously ill patients, one should use half the usual starting dose as for healthy adults. The starting dose for

fluoxetine is 10 mg once a day. In most cases, once-a-day dosing is possible. The choice of which SSRI to use should be driven by (1) the patient's past success or failure with the specific medication and (2) the most favorable side effect profile for that specific agent. For instance, for a patient in whom fatigue is a major symptom, a more activating SSRI (fluoxetine) would be appropriate. For a patient in whom anxiety and sleeplessness are major symptoms, a more sedating SSRI (paroxetine) would be appropriate. Importantly, it can take up to 4 weeks for these drugs to have an effect.

Atypical antidepressants are recommended only in select circumstances, usually with the assistance of a specialty consultation. Trazodone can be an effective antidepressant but is sedating and can cause orthostatic hypotension and, occasionally, priapism. Therefore, it should be used before bed and only when a sedating effect is desired and is often used for patients with insomnia at a dose starting at 25 mg. Bupropion can also be used. In addition to its antidepressant effects, bupropion is energizing, making it useful for depressed patients who experience fatigue. However, it can cause seizures, preventing its use for patients with a risk of CNS neoplasms or terminal delirium. Finally, alprazolam, a benzodiazepine, starting at 0.25–1.0 mg tid, can be effective in seriously ill patients who have a combination of anxiety and depression. Although it is potent and works quickly, it has many drug interactions and may cause delirium, especially among very ill patients, because of its strong binding to the benzodiazepine- γ -aminobutyric acid (GABA) receptor complex.

Unless used as adjuvants for the treatment of pain, tricyclic antidepressants are not recommended. While they can be effective, their therapeutic window and serious side effects typically limit their utility. Similarly, monoamine oxidase (MAO) inhibitors are not recommended because of their side effects and dangerous drug interactions.

Delirium (See Chap. 27) • **FREQUENCY** In the weeks or months before death, delirium is uncommon, although it may be significantly underdiagnosed. However, delirium becomes relatively common in the days and hours immediately before death. Up to 85% of patients dying from cancer may experience terminal delirium.

ETIOLOGY Delirium is a global cerebral dysfunction characterized by alterations in cognition and consciousness. It is frequently preceded by anxiety, changes in sleep patterns (especially reversal of day and night), and decreased attention. In contrast to dementia, delirium has an acute onset, is characterized by fluctuating consciousness and inattention, and is reversible, although reversibility may be more theoretical than real for patients near death. Delirium may occur in a patient with dementia; indeed, patients with dementia are more vulnerable to delirium.

Causes of delirium include metabolic encephalopathy arising from liver or renal failure, hypoxemia, or infection; electrolyte imbalances such as hypercalcemia; paraneoplastic syndromes; dehydration; and primary brain tumors, brain metastases, or leptomeningeal spread of tumor. Among dying patients, delirium is commonly caused by side effects of treatments, including radiation for brain metastases and medications, such as opioids, glucocorticoids, anticholinergic drugs, antihistamines, antiemetics, benzodiazepines, and chemotherapeutic agents. The etiology may be multifactorial; e.g., dehydration may exacerbate opioid-induced delirium.

ASSESSMENT Delirium should be recognized in any terminally ill patient exhibiting new onset of disorientation, impaired cognition, somnolence, fluctuating levels of consciousness, or delusions with or without agitation. Delirium must be distinguished from acute anxiety, depression, and dementia. The central distinguishing feature is altered consciousness, which usually is not noted in anxiety, depression, or dementia. Although "hyperactive" delirium, characterized by overt confusion and agitation, is probably more common, patients should also be assessed for "hypoactive" delirium, which is characterized by sleep-wake reversal and decreased alertness.

In some cases, use of formal assessment tools such as the Mini-Mental Status Examination (which does not distinguish delirium from dementia) and the Delirium Rating Scale (which does distinguish

delirium from dementia) may be helpful in distinguishing delirium from other processes. The patient's list of medications must be evaluated carefully. Nonetheless, a reversible etiologic factor for delirium is found in fewer than half of all terminally ill patients. Given that most terminally ill patients experiencing delirium are very close to death and often at home, extensive diagnostic evaluations such as lumbar punctures and neuroradiologic examinations are inappropriate.

INTERVENTIONS One of the most important objectives of terminal care is to provide terminally ill patients the lucidity to say goodbye to the people they love. Delirium, especially when in combination with agitation during the final days, is distressing to family and caregivers. A strong determinant of bereavement difficulties is witnessing a difficult death. Thus, terminal delirium should be treated aggressively.

At the first sign of delirium, such as day-night reversal with slight changes in mentation, the physician should let the family members know that it is time to be sure that everything they want to say has been said. The family should be informed that delirium is common just before death.

If medications are suspected of being a cause of the delirium, unnecessary agents should be discontinued. Other potentially reversible causes, such as constipation, urinary retention, and metabolic abnormalities, should be treated. Supportive measures aimed at providing a familiar environment should be instituted, including restricting visits only to individuals with whom the patient is familiar and eliminating new experiences; orienting the patient, if possible, by providing a clock and calendar; and gently correcting the patient's hallucinations or cognitive mistakes.

Pharmacologic management focuses on the use of neuroleptics and, in extreme cases, anesthetics (Table 12-7). Haloperidol remains the first-line therapy. Usually, patients can be controlled with a low dose (1–3 mg/d), given every 6 h, although some may require as much as 20 mg/d. Haloperidol can be administered PO, SC, or IV. IM injections should not be used, except when this is the only way to address a patient's delirium. Olanzapine, an atypical neuroleptic, has shown significant effectiveness in completely resolving delirium in cancer patients. It also has other beneficial effects for terminally ill patients, including antinausea, antianxiety, and weight gain. Olanzapine is useful for patients with longer anticipated life expectancies because it is less likely to cause dysphoria and has a lower risk of dystonic reactions. Additionally, because olanzapine is metabolized through multiple pathways, it can be used in patients with hepatic and renal dysfunction. Olanzapine has the disadvantage that it is only available orally and takes a week to reach steady state. The usual dose is 2.5–5 mg PO bid. Chlorpromazine (10–25 mg every 4–6 h) can be useful if sedation is desired and can be administered IV or PR in addition to PO. Dystonic reactions resulting from dopamine blockade are a side effect of neuroleptics, although they are reported to be rare when these drugs are used to treat terminal delirium. If patients develop dystonic reactions, benztropine should be administered. Neuroleptics may be

TABLE 12-7 Medications for the Management of Delirium

INTERVENTIONS	DOSE
Neuroleptics	
Haloperidol	0.5–5 mg q2–12h, PO/IV/SC/IM
Thioridazine	10–75 mg q4–8h, PO
Chlorpromazine	12.5–50 mg q4–12h, PO/IV/IM
Atypical neuroleptics	
Olanzapine	2.5–5 mg qd or bid, PO
Risperidone	1–3 mg q12h, PO
Anxiolytics	
Lorazepam	0.5–2 mg q1–4h, PO/IV/IM
Midazolam	1–5 mg/h continuous infusion, IV/SC
Anesthetics	
Propofol	0.3–2.0 mg/h continuous infusion, IV

combined with lorazepam to reduce agitation when the delirium is the result of alcohol or sedative withdrawal.

If no response to first-line therapy is observed, a specialty consultation should be obtained with a goal to change to a different medication. If the patient fails to improve after a second neuroleptic, sedation with either an anesthetic such as propofol or continuous-infusion midazolam may be necessary. By some estimates, as many as 25% of patients at the very end of life who experience delirium, especially restless delirium with myoclonus or convulsions, may require sedation.

Physical restraints should be used with great reluctance and only when patients' violence is threatening to themselves or others. If restraints are used, their appropriateness should be frequently reevaluated.

Insomnia • FREQUENCY Sleep disorders, defined as difficulty initiating sleep or maintaining sleep, sleep difficulty at least 3 nights a week, or sleep difficulty that causes impairment of daytime functioning, occurs in 19–63% of patients with advanced cancer. Some 30–74% of patients with other end-stage conditions, including AIDS, heart disease, COPD, and renal disease, experience insomnia.

ETIOLOGY Patients with cancer may experience changes in sleep efficiency, such as an increase in stage I sleep. Insomnia may also coexist with both physical illnesses, like thyroid disease, and psychological illnesses, like depression and anxiety. Medications, including antidepressants, psychostimulants, glucocorticoids, and β agonists, are significant contributors to sleep disorders, as are caffeine and alcohol. Multiple over-the-counter medications contain caffeine and antihistamines, which can contribute to sleep disorders.

ASSESSMENT Assessments should include specific questions concerning sleep onset, sleep maintenance, and early-morning wakening, as these will provide clues to both the causative agents and management of insomnia. Patients should be asked about previous sleep problems, screened for depression and anxiety, and asked about symptoms of thyroid disease. Caffeine and alcohol are prominent causes of sleep problems, and a careful history of the use of these substances should be obtained. Both excessive use and withdrawal from alcohol can be causes of sleep problems.

INTERVENTIONS The mainstays of any intervention include improvement of sleep hygiene (encouragement of regular time for sleep, decreased nighttime distractions, elimination of caffeine and other stimulants and alcohol), interventions to treat anxiety and depression, and treatment for the insomnia itself. For patients with depression who have insomnia and anxiety, a sedating antidepressant such as mirtazapine can be helpful. In the elderly, trazodone, beginning at 25 mg at nighttime, is an effective sleep aid at doses lower than those that cause its antidepressant effect. Zolpidem may have a decreased incidence of delirium in patients compared with traditional benzodiazepines, but this has not been clearly established. When benzodiazepines are prescribed, short-acting ones (such as lorazepam) are favored over longer-acting ones (such as diazepam). Patients who receive these medications should be observed for signs of increased confusion and delirium.

SOCIAL NEEDS AND THEIR MANAGEMENT

Financial Burdens • FREQUENCY Dying can impose substantial economic strains on patients and families, potentially causing distress. This is known as financial toxicity. In the United States, which has the least comprehensive health insurance systems among wealthy countries, a quarter of families coping with end-stage cancer report that care was a major financial burden and a third used up most of their savings. Among Medicare beneficiaries, average out-of-pocket costs were >\$8000. Between 10% and 30% of families are forced to sell assets, use savings, or take out a mortgage to pay for the patient's health care costs.

The patient is likely to reduce hours worked and eventually stop working altogether. In 20% of cases, a family member of the terminally ill patient also must stop working to provide care. The major underlying causes of economic burden are related to poor physical functioning and care needs, such as the need for housekeeping, nursing, and personal care. More debilitated patients and poor patients experience greater economic burdens.

INTERVENTION The economic burden of end-of-life care should not be ignored as a private matter. It has been associated with a number of adverse health outcomes, including preferring comfort care over life-prolonging care, as well as consideration of euthanasia or physician-assisted suicide (PAS). Economic burdens increase the psychological distress of the families and caregivers of terminally ill patients, and poverty is associated with many adverse health outcomes. Importantly, studies have found that "patients with advanced cancer who reported having end-of-life conversations with physicians had significantly lower health care costs in their final week of life. Higher costs were associated with worse quality of death." Assistance from a social worker, early on if possible, to ensure access to all available benefits may be helpful. Many patients, families, and health care providers are unaware of options for long-term care insurance, respite care, the Family Medical Leave Act (FMLA), and other sources of assistance. Some of these options (such as respite care) may be part of a formal hospice program, but others (such as the FMLA) do not require enrollment in a hospice program.

Relationships • FREQUENCY Settling personal issues and closing the narrative of lived relationships are universal needs. When asked if sudden death or death after an illness is preferable, respondents often initially select the former, but soon change to the latter as they reflect on the importance of saying goodbye. Bereaved family members who have not had the chance to say goodbye often have a more difficult grief process.

INTERVENTIONS Care of seriously ill patients requires efforts to facilitate the types of encounters and time spent with family and friends that are necessary to meet those needs. Family and close friends may need to be accommodated in hospitals and other facilities with unrestricted visiting hours, which may include sleeping near the patient, even in otherwise regimented institutional settings. Physicians and other health care providers may be able to facilitate and resolve strained interactions between the patient and other family members. Assistance for patients and family members who are unsure about how to create or help preserve memories, whether by providing materials such as a scrapbook or memory box or by offering them suggestions and informational resources, can be deeply appreciated. Taking photographs and creating videos can be especially helpful to terminally ill patients who have younger children or grandchildren.

Family Caregivers • FREQUENCY Caring for seriously ill patients places a heavy burden on families. Families are frequently required to provide transportation and homemaking, as well as other services. Typically, paid professionals, such as home health nurses and hospice workers, supplement family care; only about a quarter of all caregiving consists of exclusively paid professional assistance. Over the past 40 years, there has been a significant decline in the United States of deaths occurring in hospitals, with a simultaneous increase in deaths in other facilities and at home. Over a third of deaths occur in patients' homes. This increase in out-of-hospital deaths increases reliance on families for end-of-life care. Increasingly, family members are being called upon to provide physical care (such as moving and bathing patients) and medical care (such as assessing symptoms and giving medications) in addition to emotional care and support.

Three-quarters of family caregivers of terminally ill patients are women—wives, daughters, sisters, and even daughters-in-law. Since many are widowed, women tend to be able to rely less on family for caregiving assistance and may need more paid assistance. About 20% of terminally ill patients report substantial unmet needs for nursing and personal care. The impact of caregiving on family caregivers is substantial: both bereaved and current caregivers have a higher mortality rate than that of non-caregiving controls.

INTERVENTIONS It is imperative to inquire about unmet needs and to try to ensure that those needs are met either through the family or by paid professional services when possible. Community assistance through houses of worship or other community groups often can be mobilized by telephone calls from the medical team to someone the patient or family identifies. Sources of support specifically for family

caregivers should be identified through local sources or nationally through groups such as the National Family Caregivers Association (www.nfcacares.org), the American Cancer Society (www.cancer.org), and the Alzheimer's Association (www.alz.org).

■ EXISTENTIAL NEEDS AND THEIR MANAGEMENT

Frequency Religion and spirituality are often important to dying patients. Nearly 70% of patients report becoming more religious or spiritual when they became terminally ill, and many find comfort in religious or spiritual practices such as prayer. However, ~20% of terminally ill patients become less religious, frequently feeling cheated or betrayed by becoming terminally ill. For other patients, the need is for existential meaning and purpose that is distinct from, and may even be antithetical to, religion or spirituality. When asked, patients and family caregivers frequently report wanting their professional caregivers to be more attentive to religion and spirituality.

Assessment Health care providers are often hesitant about involving themselves in the religious, spiritual, and existential experiences of their patients because it may seem private or not relevant to the current illness. But physicians and other members of the care team should be able at least to detect spiritual and existential needs. Screening questions have been developed for a physician's spiritual history taking. Spiritual distress can amplify other types of suffering and even masquerade as intractable physical pain, anxiety, or depression. The screening questions in the comprehensive assessment are usually sufficient. Deeper evaluation and intervention are rarely appropriate for the physician unless no other member of a care team is available or suitable. Pastoral care providers may be helpful, whether from the medical institution or from the patient's own community.

Interventions Precisely how religious practices, spirituality, and existential explorations can be facilitated and improve end-of-life care is not well established. What is clear is that for physicians, one main intervention is to inquire about the role and importance of spirituality and religion in a patient's life. This will help a patient feel heard and help physicians identify specific needs. In one study, only 36% of respondents indicated that a clergy member would be comforting. Nevertheless, the increase in religious and spiritual interest among a substantial fraction of dying patients suggests inquiring of individual patients how this need can be addressed. Some evidence supports specific methods of addressing existential needs in patients, ranging from establishing a supportive group environment for terminal patients to individual treatments emphasizing a patient's dignity and sources of meaning.

MANAGING THE LAST STAGES

■ PALLIATIVE CARE SERVICES: HOW AND WHERE

Determining the best approach to providing palliative care to patients will depend on patient preferences, the availability of caregivers and specialized services in close proximity, institutional resources, and reimbursement. Hospice is a leading, but not the only, model of palliative care services. In the United States, slightly more than a third—35.7%—of hospice care is provided in private residential homes with 14.5% of hospice care in nursing homes. In the United States, Medicare pays for hospice services under Part A, the hospital insurance part of reimbursement. Two physicians must certify that the patient has a prognosis of ≤6 months if the disease runs its usual course. Prognoses are probabilistic by their nature; patients are not required to die within 6 months but rather to have a condition from which half the individuals with it would not be alive within 6 months. Patients sign a hospice enrollment form that states their intent to forgo curative services related to their terminal illness but can still receive medical services for other comorbid conditions. Patients also can withdraw enrollment and reenroll later; the hospice Medicare benefit can be revoked later to secure traditional Medicare benefits. Payments to the hospice are per diem (or capitated), not fee-for-service. Payments are intended to cover physician services for the medical direction of the care team;

regular home care visits by registered nurses and licensed practical nurses; home health aide and homemaker services; chaplain services; social work services; bereavement counseling; and medical equipment, supplies, and medications. No specific therapy is excluded, and the goal is for each therapy to be considered for its symptomatic (as opposed to disease-modifying) effect. Additional clinical care, including services of the primary physician, is covered by Medicare Part B even while the hospice Medicare benefit is in place.

The Affordable Care Act directs the secretary of Health and Human Services to gather data on Medicare hospice reimbursement with the goal of reforming payment rates to account for resource use over an entire episode of care. The legislation also requires additional evaluations and reviews of eligibility for hospice care by hospice physicians or nurses. The Center for Medicare and Medicaid Innovation (CMMI) sponsors and carries out demonstration projects to test models and evaluate the potential of new methods. In 2016, CMMI started a 5-year test of concurrent hospice and palliative care services with curative treatment for terminally ill patients who have a life expectancy of ≤6 months. A 4-year test initiated in 2021 will examine the inclusion of hospice in Medicare Advantage covering 8% of the market and include important health plans.

By 2018, the average length of enrollment in a hospice for Medicare beneficiaries was 90 days. However, the median length of stay was just 18 days, suggesting most patients are in hospice for a short time. Such short stays create barriers to establishing high-quality palliative services in patients' homes and also place financial strains on hospice providers since the initial assessments are resource intensive. Physicians should initiate early referrals to the hospice to allow more time for patients to receive palliative care.

In the United States, hospice care has been the main method for securing palliative services for terminally ill patients. However, leading physicians have increasingly emphasized the need to introduce palliative care much earlier in patients' illness, and efforts are being made to develop palliative care services that can be provided before the last 6 months of life and across a variety of settings. Studies of terminally ill patients indicate that those who received in-home palliative care delivered by an interdisciplinary team compared to usual care were more satisfied, more likely to die at home, and had fewer visits to the emergency room and lower per-day costs. More companies and home health agencies are now offering nonhospice palliative care services in patients' homes in an effort to increase quality of life and forestall emergency room visits and hospitalizations. Similarly, palliative care services are increasingly available via consultation, rather than being available only in hospital, day care, outpatient, and nursing home settings. Palliative care consultations for nonhospice patients can be billed as for other consultations under Medicare Part B. It is argued that using palliative care earlier in patients' illness allows patients and family members to become more acculturated to avoiding life-sustaining treatments, facilitating a smoother transition to hospice care closer to death.

■ WITHDRAWING AND WITHHOLDING LIFE-SUSTAINING TREATMENT

Legal Aspects For centuries, it has been deemed ethical to withhold or withdraw life-sustaining interventions. The current legal consensus in the United States and most wealthy countries is that patients have a moral as well as legal right to refuse medical interventions. American courts also have held that incompetent patients have a right to refuse medical interventions. For patients who are incompetent and terminally ill and who have not completed an advance care directive, next of kin can exercise that right, although this may be restricted in some states, depending on how clear and convincing the evidence is of the patient's preferences. Courts have limited families' ability to terminate life-sustaining treatments in patients who are conscious and incompetent but not terminally ill. In theory, patients' right to refuse medical therapy can be limited by four countervailing interests: (1) preservation of life, (2) prevention of suicide, (3) protection of third parties such as children, and (4) preservation of the integrity of the

medical profession. In practice, these interests almost never override the right of competent patients and incompetent patients who have left explicit wishes or advance care directives.

For incompetent patients who either appointed a proxy without specific indications of their wishes or never completed an advance care directive, three criteria have been suggested to guide the decision to terminate medical interventions. First, some commentators suggest that ordinary care should be administered but extraordinary care could be terminated. Because the ordinary/extraordinary distinction is too vague, courts and commentators widely agree that it should not be used to justify decisions about stopping treatment. Second, many courts have advocated the use of the substituted-judgment criterion, which holds that the proxy decision-makers should try to imagine what the incompetent patient would do if he or she were competent. However, multiple studies indicate that many proxies, even close family members, cannot accurately predict what the patient would have wanted. Therefore, substituted judgment becomes more of a guessing game than a way of fulfilling the patient's wishes. Finally, the best-interests criterion holds that proxies should evaluate treatments by balancing their benefits and risks and select those treatments where the benefits maximally outweigh the burdens of treatment. Clinicians have a clear and crucial role in this by carefully and dispassionately explaining the known benefits and burdens of specific treatments. Yet even when that information is as clear as possible, different individuals can have very different views of what is in the patient's best interests, and families may have disagreements or even overt conflicts. This criterion has been criticized because there is no single way to determine the balance between benefits and burdens; it depends on a patient's personal values. For instance, for some people, being alive even if mentally incapacitated is a benefit, whereas for others, it may be the worst possible existence. As a matter of practice, physicians rely on family members to make decisions that they feel are best and object only if those decisions seem to demand treatments that the physicians consider not beneficial.

Practices Withholding and withdrawing acutely life-sustaining medical interventions from terminally ill patients are now standard practice. More than 90% of American patients die without cardiopulmonary resuscitation (CPR), and just as many forgo other potentially life-sustaining interventions. For instance, in ICUs in the period of 1987–1988, CPR was performed 49% of the time, but it was performed only 10% of the time in 1992–1993 and on just 1.8% of admissions from 2001 to 2008. On average, 3.8 interventions, such as vasopressors and transfusions, were stopped for each dying ICU patient. However, up to 19% of decedents in hospitals received interventions such as extubation, ventilation, and surgery in the 48 h preceding death. There is wide variation in practices among hospitals and ICUs, suggesting an important element of physician preferences rather than consistent adherence to professional society recommendations.

Mechanical ventilation may be the most challenging intervention to withdraw. The two approaches are *terminal extubation*, which is the removal of the endotracheal tube, and *terminal weaning*, which is the gradual reduction of the fraction of inspired oxygen (FIO_2) or ventilator rate. One-third of ICU physicians prefer to use the terminal weaning technique, and 13% extubate; the majority of physicians utilize both techniques. The American Thoracic Society's 2008 clinical policy guidelines note that there is no single correct process of ventilator withdrawal and that physicians use and should be proficient in both methods but that the chosen approach should carefully balance benefits and burdens as well as patient and caregiver preferences. Some recommend terminal weaning because patients do not develop upper airway obstruction and the distress caused by secretions or stridor; however, terminal weaning can prolong the dying process and not allow a patient's family to be with the patient unencumbered by an endotracheal tube. To ensure comfort for conscious or semiconscious patients before withdrawal of the ventilator, neuromuscular blocking agents should be terminated and sedatives and analgesics administered. Removing the neuromuscular blocking agents permits patients to show discomfort, facilitating the titration of sedatives and analgesics; it also permits interactions between patients and their families. A common

practice is to inject a bolus of midazolam (2–4 mg) or lorazepam (2–4 mg) before withdrawal, followed by a bolus of 5–10 mg of morphine and continuous infusion of morphine (50% of the bolus dose per hour) during weaning. In patients who have significant upper airway secretions, IV scopolamine at a rate of 100 $\mu\text{g}/\text{h}$ can be administered. Additional boluses of morphine or increases in the infusion rate should be administered for respiratory distress or signs of pain. Higher doses will be needed for patients already receiving sedatives and opioids.

The median time to death after stopping of the ventilator is 1 h. However, up to 10% of patients unexpectedly survive for 1 day or more after mechanical ventilation is stopped. Women and older patients tend to survive longer after extubation. Families need to be reassured about both the continuations of treatments for common symptoms, such as dyspnea and agitation, after withdrawal of ventilatory support and the uncertainty of length of survival after withdrawal of ventilatory support.

FUTILE CARE

Beginning in the late 1980s, some commentators argued that physicians could terminate futile treatments demanded by the families of terminally ill patients. Although no objective definition or standard of futility exists, several categories have been proposed. Physiologic futility means that an intervention will have no physiologic effect. Some have defined qualitative futility as applying to procedures that "fail to end a patient's total dependence on intensive medical care." Quantitative futility occurs "when physicians conclude (through personal experience, experiences shared with colleagues, or consideration of reported empiric data) that in the last 100 cases, a medical treatment has been useless." The term conceals subjective value judgments about when a treatment is "not beneficial." Deciding whether a treatment that obtains an additional 6 weeks of life or a 1% survival advantage confers benefit depends on patients' preferences and goals. Furthermore, physicians' predictions of when treatments are futile deviate markedly from the quantitative definition. When residents thought CPR was quantitatively futile, more than one in five patients had a >10% chance of survival to hospital discharge. Most studies that purport to guide determinations of futility are based on insufficient data and therefore cannot provide statistical confidence for clinical decision-making. Quantitative futility rarely applies in ICU settings.

Many commentators reject using futility as a criterion for withdrawing care, preferring instead to consider futility situations as ones that represent conflict that calls for careful negotiation between families and health care providers. The American Medical Association and other professional societies have developed process-based approaches to resolving cases clinicians feel are futile. These process-based measures mainly suggest involving consultants and/or ethics committees when there are seemingly irresolvable differences. Some hospitals have enacted "unilateral do-not-resuscitate" policies to allow clinicians to provide a do-not-resuscitate order in cases in which consensus cannot be reached with families and medical opinion is that resuscitation would be futile if attempted. This type of a policy is not a replacement for careful and patient communication and negotiation but recognizes that agreement cannot always be reached.

In 1999, Texas enacted the so-called Futile Care Act. Other states, such as Virginia, Maryland, and California, have also enacted such laws that provide physicians a "safe harbor" from liability if they refuse a patient's or family's request for life-sustaining interventions. For instance, in Texas, when a disagreement about terminating interventions between the medical team and the family has not been resolved by an ethics consultation, the physician is tasked with trying to facilitate transfer of the patient to an institution willing to provide treatment. If this fails after 10 days, the hospital and physician may unilaterally withdraw treatments determined to be futile. The family may appeal to a state court. Early data suggest that the law increases futility consultations for the ethics committee and that, although most families concur with withdrawal, ~10–15% of families refuse to withdraw treatment. As of 2007, there had been 974 ethics committee consultations on medical futility cases and 65 in which committees ruled against families and gave notice that treatment would be terminated. In 2007,

a survey of Texas hospitals showed that 30% of hospitals had used the futility law in 213 adult cases and 42 pediatric cases. Treatment was withdrawn for 27 of those patients, and the remainder were transferred to other facilities or died while awaiting transfer.

EUTHANASIA AND PHYSICIAN-ASSISTED SUICIDE

Euthanasia and PAS are defined in **Table 12-8**. Terminating life-sustaining care and providing opioid medications to manage symptoms such as pain or dyspnea have long been considered ethical by the medical profession and legal by courts and should not be conflated with euthanasia or PAS.

Legal Aspects Euthanasia and PAS are legal in the Netherlands, Belgium, Luxembourg, Colombia, Canada, Spain, Western Australia, and New Zealand. Euthanasia was legalized in the Northern Territory of Australia in 1996, but that legislation was repealed 9 months later in 1997. Under certain conditions, a layperson in Switzerland or Germany can legally elect assisted suicide. In the United States, PAS is legal in Washington, D.C., and 10 states: Oregon, Washington State, Montana, Vermont, California, Colorado, Hawaii, Maine, New Jersey, and New Mexico. No state in the United States has legalized euthanasia. In the United States, multiple criteria must be met for PAS: the patient must have a terminal condition of <6 months and must be determined eligible through a process that includes a 15-day waiting period. In 2009, the state supreme court of Montana ruled that state law permits PAS for terminally ill patients. Many other countries, such as Portugal, are actively debating the legalization of euthanasia and/or PAS.

Practices Fewer than 10–20% of terminally ill patients actually consider euthanasia and/or PAS for themselves. Use of euthanasia and PAS is increasing but remains relatively rare. In all countries, even the Netherlands and Belgium where these practices have been tolerated and legal for many years, <5% of death occur by euthanasia or PAS. As of the most recent data, 4.7% of all deaths were by euthanasia or PAS in the Netherlands (2015) and 4.6% in Belgium (2013). Just 0.50% of all deaths in Oregon in 2019 (188 of 37,397 deaths) and 0.36% of all deaths in Washington State in 2018 (203 of 56,913 deaths) were reported to be by PAS, although these may be underestimates since the cause of some deaths of patients who received medications could not be verified.

In Belgium, the Netherlands, Oregon, and Washington, >70% of patients utilizing these interventions are dying of cancer; <10% of deaths by euthanasia or PAS involve patients with AIDS or amyotrophic

lateral sclerosis. While the numbers are small, in the Netherlands, the numbers of euthanasia or PAS cases in patients with psychiatric disorders, dementia, and the accumulation of health issues are increasing.

Pain is not the primary motivator for patients' requests for or interest in euthanasia and/or PAS. Among the first patients to receive PAS in Oregon, only 1 of the 15 patients had inadequate pain control, compared with 15 of the 43 patients in a control group who experienced inadequate pain relief. About 33% of patients in Oregon seeking PAS currently cite pain or fear of pain as their main reason for doing so. Conversely, depression and hopelessness are strongly associated with patient interest in euthanasia and PAS. Concerns about loss of dignity or autonomy or being a burden on family members appear to be more important factors motivating a desire for euthanasia or PAS. Losing autonomy (87% Oregon [OR], 85% Washington [WA]), not being able to enjoy activities (90% OR, 84% WA), and fear of losing dignity (72% OR, 69% WA) are the most-cited end-of-life concerns in both states. A high percentage of patients seeking PAS note being a burden on family (59% OR, 51% WA). A study from the Netherlands showed that depressed terminally ill cancer patients were four times more likely to request euthanasia and confirmed that uncontrolled pain was not associated with greater interest in euthanasia.

Euthanasia and PAS are no guarantee of a painless, quick death. Data from the Netherlands indicate that in as many as 20% of euthanasia and PAS cases technical and other problems arose, including patients waking from coma, not becoming comatose, regurgitating medications, and experiencing a prolonged time to death. Data from Oregon between 1998 and 2017 and Washington between 2009 and 2017 indicate that of patients who received PAS prescriptions, 81% died at home and prescribers were present in 9.7% of cases. The time between drug intake and coma ranged from 1 min to 11 h, and the time from drug intake to death ranged from 1 min to 104 h. The median time from ingestion to coma was 5 min and from ingestion to death was 25 min. In Oregon between 1998 and 2015, 53% of patients had no complications, 44% of patients had no data on complications, and 2.4% of patients had regurgitation after taking the prescribed medicine as the only complication. In addition, six patients awakened. In Washington State between 2014 and 2015, 1.4% of patients had regurgitation, one patient had a seizure, and the reported range of time to death extended to 30 h. In the Netherlands, problems were significantly more common in PAS, sometimes requiring the physician to intervene and provide euthanasia.

Regardless of whether they practice in a setting where euthanasia is legal or not, many physicians over the course of their careers will receive a patient request for euthanasia or PAS. In the United States, 18% of physicians have received a request for PAS and 11% have received a request for euthanasia. Three percent complied with a request for PAS, while 5% complied with a request for euthanasia. In the Netherlands, where the practices are legal, 77% of physicians have received a request for PAS or euthanasia and 60% have performed these interventions.

Competency in dealing with such a request is crucial. Although challenging, the request can also provide a chance to address intense suffering. After receiving a request for euthanasia and/or PAS, health care providers should carefully clarify the request with empathetic, open-ended questions to help elucidate the underlying cause for the request, such as, "What makes you want to consider this option?" Endorsing either moral opposition or moral support for the act tends to be counterproductive, giving an impression of being judgmental or of endorsing the idea that the patient's life is worthless. Health care providers must reassure the patient of continued care and commitment. The patient should be educated about alternative, less laden options, such as symptom management and withdrawing any unwanted treatments, and the reality of euthanasia and/or PAS, since the patient may have misconceptions about their effectiveness as well as the legal implications of the choice. Depression, hopelessness, and other symptoms of psychological distress, as well as physical suffering and economic burdens, are likely factors motivating the request, and such factors should be assessed and treated aggressively. After these interventions and clarification of options, most patients proceed with another approach,

TABLE 12-8 Definitions of Physician-Assisted Suicide and Euthanasia

TERM	DEFINITION	LEGAL STATUS
Voluntary active euthanasia	Intentionally administering medications or other interventions to cause the patient's death with the patient's informed consent	Netherlands, Belgium, Luxembourg, Canada, Colombia, Spain, Western Australia, New Zealand
Involuntary active euthanasia	Intentionally administering medications or other interventions to cause the patient's death when the patient was competent to consent but did not—e.g., the patient may not have been asked	Nowhere
Passive euthanasia	Withholding or withdrawing life-sustaining medical treatments from a patient to let him or her die (terminating life-sustaining treatments)	Everywhere
Physician-assisted suicide	A physician provides medications or other interventions to a patient with the understanding that the patient can use them to commit suicide	Netherlands, Belgium, Luxembourg, Canada, Colombia, Germany, Switzerland, Oregon, Washington, Montana, Vermont, California, Colorado, District of Columbia, Hawaii, Maine, New Jersey, New Mexico

declining life-sustaining interventions, possibly including refusal of nutrition and hydration.

CARE DURING THE LAST HOURS

Most laypersons have limited experiences with the actual dying process and death. They frequently do not know what to expect of the final hours and afterward. The family and other caregivers must be prepared, especially if the plan is for the patient to die at home.

Patients in the last days of life typically experience extreme weakness and fatigue and become bedbound; this can lead to pressure sores. The issue of turning patients who are near the end of life, however, must be balanced against the potential discomfort that movement may cause. Patients stop eating and drinking with drying of mucosal membranes and dysphagia. Careful attention to oral swabbing, lubricants for lips, and use of artificial tears can provide a form of care to substitute for attempts at feeding the patient. With loss of the gag reflex and dysphagia, patients may also experience accumulation of oral secretions, producing noises during respiration sometimes called “the death rattle.” Scopolamine can reduce the secretions. Patients also experience changes in respiration with periods of apnea or Cheyne-Stokes

breathing. Decreased intravascular volume and cardiac output cause tachycardia, hypotension, peripheral coolness, and livedo reticularis (skin mottling). Patients can have urinary and, less frequently, fecal incontinence. Changes in consciousness and neurologic function generally lead to two different paths to death.

Each of these terminal changes can cause patients and families distress, requiring reassurance and targeted interventions (Table 12-9). Informing families that these changes might occur and providing them with an information sheet can help preempt problems and minimize distress. Understanding that patients stop eating because they are dying, not dying because they have stopped eating, can reduce family and caregiver anxiety. Similarly, informing the family and caregivers that the “death rattle” may occur and that it is not indicative of suffocation, choking, or pain can reduce their worry from the breathing sounds.

Families and caregivers may also feel guilty about stopping treatments, fearing that they are “killing” the patient. This may lead to demands for interventions, such as feeding tubes, that may be ineffective. In such cases, the physician should remind the family and caregivers about the inevitability of events and the palliative goals.

TABLE 12-9 Managing Changes in the Patient’s Condition during the Final Days and Hours

CHANGES IN THE PATIENT’S CONDITION	POTENTIAL COMPLICATION	FAMILY’S POSSIBLE REACTION AND CONCERN	ADVICE AND INTERVENTION
Profound fatigue	Bedbound with development of pressure ulcers that are prone to infection, malodor, and pain, and joint pain	Patient is lazy and giving up.	Reassure family and caregivers that terminal fatigue will not respond to interventions and should not be resisted. Use an air mattress if necessary.
Anorexia	None	Patient is giving up; patient will suffer from hunger and will starve to death.	Reassure family and caregivers that the patient is not eating because he or she is dying; not eating at the end of life does not cause suffering or death. Forced feeding, whether oral, parenteral, or enteral, does not reduce symptoms or prolong life.
Dehydration	Dry mucosal membranes (see below)	Patient will suffer from thirst and die of dehydration.	Reassure family and caregivers that dehydration at the end of life does not cause suffering because patients lose consciousness before any symptom distress. Intravenous hydration can worsen symptoms of dyspnea by pulmonary edema and peripheral edema as well as prolong the dying process.
Dysphagia	Inability to swallow oral medications needed for palliative care		Do not force oral intake. Discontinue unnecessary medications that may have been continued, including antibiotics, diuretics, antidepressants, and laxatives. If swallowing pills is difficult, convert essential medications (analgesics, antiemetics, anxiolytics, and psychotropics) to oral solutions, buccal, sublingual, or rectal administration.
“Death rattle”—noisy breathing		Patient is choking and suffocating.	Reassure the family and caregivers that this is caused by secretions in the oropharynx and the patient is not choking. Reduce secretions with scopolamine (0.2–0.4 mg SC q4h or 1–3 patches q3d). Reposition patient to permit drainage of secretions. Do not suction. Suction can cause patient and family discomfort and is usually ineffective.
Apnea, Cheyne-Stokes respirations, dyspnea		Patient is suffocating.	Reassure family and caregivers that unconscious patients do not experience suffocation or air hunger. Apneic episodes are frequently a premorbid change. Opioids or anxiolytics may be used for dyspnea. Oxygen is unlikely to relieve dyspneic symptoms and may prolong the dying process.
Urinary or fecal incontinence	Skin breakdown if days until death Potential transmission of infectious agents to caregivers	Patient is dirty, malodorous, and physically repellent.	Remind family and caregivers to use universal precautions. Frequent changes of bedclothes and bedding. Use diapers, urinary catheter, or rectal tube if diarrhea or high urine output.
Agitation or delirium	Day/night reversal Hurt self or caregivers	Patient is in horrible pain and going to have a horrible death.	Reassure family and caregivers that agitation and delirium do not necessarily connote physical pain. Depending on the prognosis and goals of treatment, consider evaluating for causes of delirium and modifying medications. Manage symptoms with haloperidol, chlorpromazine, diazepam, or midazolam.
Dry mucosal membranes	Cracked lips, mouth sores, and candidiasis can also cause pain. Odor	Patient may be malodorous, physically repellent.	Use baking soda mouthwash or saliva preparation q15–30 min. Use topical nystatin for candidiasis. Coat lips and nasal mucosa with petroleum jelly q60–90 min. Use ophthalmic lubricants q4h or artificial tears q30 min.

Interventions may prolong the dying process and cause discomfort. Physicians also should emphasize that withholding treatments is both legal and ethical and that the family members are not the cause of the patient's death. This reassurance may have to be provided multiple times.

Hearing and touch are said to be the last senses to stop functioning. Whether this is the case or not, families and caregivers can be encouraged to communicate with the dying patient. Encouraging them to talk directly to the patient, even if he or she is unconscious, and hold the patient's hand or demonstrate affection in other ways can be an effective way to channel their urge "to do something" for the patient.

When the plan is for the patient to die at home, the physician must inform the family and caregivers how to determine that the patient has died. The cardinal signs are cessation of cardiac function and respiration; the pupils become fixed; the body becomes cool; muscles relax; and incontinence may occur. Remind the family and caregivers that the eyes may remain open even after the patient has died.

The physician should establish a plan for who the family or caregivers will contact when the patient is dying or has died. Without a plan, family members may panic and call 911, unleashing a cascade of unwanted events, from arrival of emergency personnel and resuscitation to hospital admission. The family and caregivers should be instructed to contact the hospice (if one is involved), the covering physician, or the on-call member of the palliative care team. They should also be told that the medical examiner need not be called unless the state requires it for all deaths. Unless foul play is suspected, the health care team need not contact the medical examiner either.

Just after the patient dies, even the best-prepared family may experience shock and loss and be emotionally distraught. They need time to assimilate the event and be comforted. Health care providers are likely to find it meaningful to write a bereavement card or letter to the family. The purpose is to communicate about the patient, perhaps emphasizing the patient's virtues and the honor it was to care for the patient, and to express concern for the family's hardship. Some physicians attend the funerals of their patients. Although this is beyond any medical obligation, the presence of the physician can be a source of support to the grieving family and provides an opportunity for closure for the physician.

Death of a spouse is a strong predictor of poor health, and even mortality, for the surviving spouse. It may be important to alert the spouse's physician about the death so that he or she is aware of symptoms that might require professional attention.

FURTHER READING

EMANUEL E et al: Attitudes and practices of euthanasia and physician-assisted suicide in the United States, Canada, and Europe. *JAMA* 316:79, 2016.

KELLEY AS, MEIER DE: Palliative care—A shifting paradigm. *N Engl J Med* 363:781, 2010.

KELLEY AS et al: Hospice enrollment saves money for Medicare and improves care quality across a number of different lengths-of-stay. *Health Aff* 32:552, 2012.

KELLEY AS et al: Palliative care for the seriously ill. *N Engl J Med* 373:747, 2015.

MACK JW et al: Associations between end-of-life discussion characteristics and care received near death: A prospective cohort study. *J Clin Oncol* 30:4387, 2012.

MURRAY SA et al: Illness trajectories and palliative care. *BMJ* 330:1007, 2005.

NEUMAN P et al: Medicare per capita spending by age and service: New data highlights oldest beneficiaries. *Health Aff (Millwood)* 34:335, 2015.

NICHOLAS LH et al: Regional variation in the association between advance directives and end-of-life Medicare expenditures. *JAMA* 306:1447, 2011.

ORNSTEIN KA et al: Evaluation of racial disparities in hospice use and end-of-life treatment intensity in the REGARDS cohort. *JAMA Netw Open* 3(8):e2014639, 2020.

QUINN KL et al: Association of receipt of palliative care interventions with health care use, quality of life, and symptom burden among adults with chronic noncancer illness: A systematic review and meta-analysis. *JAMA* 324:1439, 2020.

TENO JM et al: Change in end-of-life care for medicare beneficiaries: Site of death, place of care, and health transitions in 2000, 2005, and 2009. *JAMA* 309:470, 2013.

TENO JM et al: Site of death, place of care, and health care transitions among US Medicare beneficiaries, 2000–2015. *JAMA* 320:264, 2018.

VAN DEN BEUKEN-VANEVERDINGEN MH et al: Update on prevalence of pain in patients with cancer: Systematic review and meta-analysis. *J Pain Symptom Manage* 51:1070, 2016.

WEBSITES

AMERICAN ACADEMY OF HOSPICE AND PALLIATIVE MEDICINE: www.aahpm.org

CENTER TO ADVANCE PALLIATIVE CARE: <http://www.capc.org>

EDUCATION IN PALLIATIVE AND END OF LIFE CARE (EPEC): <http://www.epec.net>

FAMILY CAREGIVER ALLIANCE: <http://www.caregiver.org>

NATIONAL HOSPICE AND PALLIATIVE CARE ORGANIZATION (including state-specific advance directives): <http://www.nhpco.org>

NCCN: The National Comprehensive Cancer Network palliative care guidelines: <http://www.nccn.org>

OUR CARE WISHES ADVANCE CARE PLANNING TOOL: <https://www.ourcarewishes.org>

This page intentionally left blank