

Reference Manual on Scientific Evidence

Third Edition

Committee on the Development of the Third Edition of the
Reference Manual on Scientific Evidence

Committee on Science, Technology, and Law
Policy and Global Affairs

FEDERAL JUDICIAL CENTER

NATIONAL RESEARCH COUNCIL
OF THE NATIONAL ACADEMIES

THE NATIONAL ACADEMIES PRESS
Washington, D.C.
www.nap.edu

THE NATIONAL ACADEMIES PRESS 500 Fifth Street, N.W. Washington, DC 20001

The Federal Judicial Center contributed to this publication in furtherance of the Center's statutory mission to develop and conduct educational programs for judicial branch employees. The views expressed are those of the authors and not necessarily those of the Federal Judicial Center.

NOTICE: The project that is the subject of this report was approved by the Governing Board of the National Research Council, whose members are drawn from the councils of the National Academy of Sciences, the National Academy of Engineering, and the Institute of Medicine. The members of the committee responsible for the report were chosen for their special competences and with regard for appropriate balance.

The development of the third edition of the *Reference Manual on Scientific Evidence* was supported by Contract No. B5727.R02 between the National Academy of Sciences and the Carnegie Corporation of New York and a grant from the Starr Foundation. The views expressed in this publication are those of the authors and do not necessarily reflect those of the National Academies or the organizations that provided support for the project.

International Standard Book Number-13: 978-0-309-21421-6

International Standard Book Number-10: 0-309-21421-1

Library of Congress Cataloging-in-Publication Data

Reference manual on scientific evidence. — 3rd ed.

p. cm.

Includes bibliographical references and index.

ISBN-13: 978-0-309-21421-6 (pbk.)

ISBN-10: 0-309-21421-1 (pbk.)

1. Evidence, Expert—United States. I. Federal Judicial Center.

KF8961.R44 2011

347.73'67—dc23

2011031458

Additional copies of this report are available from the National Academies Press, 500 Fifth Street, N.W., Lockbox 285, Washington, DC 20055; (800) 624-6242 or (202) 334-3313 (in the Washington metropolitan area); Internet, <http://www.nap.edu>.

Copyright 2011 by the National Academy of Sciences. All rights reserved.

Printed in the United States of America

Summary Table of Contents

A detailed Table of Contents appears at the front of each chapter.

Introduction, 1	Stephen Breyer
The Admissibility of Expert Testimony, 11	Margaret A. Berger
How Science Works, 37	David Goodstein
Reference Guide on Forensic Identification Expertise, 55	Paul C. Giannelli, Edward J. Imwinkelried, & Joseph L. Peterson
Reference Guide on DNA Identification Evidence, 129	David H. Kaye & George Sensabaugh
Reference Guide on Statistics, 211	David H. Kaye & David A. Freedman
Reference Guide on Multiple Regression, 303	Daniel L. Rubinfeld
Reference Guide on Survey Research, 359	Shari Seidman Diamond
Reference Guide on Estimation of Economic Damages, 425	Mark A. Allen, Robert E. Hall, & Victoria A. Lazear
Reference Guide on Exposure Science, 503	Joseph V. Rodricks
Reference Guide on Epidemiology, 549	Michael D. Green, D. Michal Freedman, & Leon Gordis
Reference Guide on Toxicology, 633	Bernard D. Goldstein & Mary Sue Henifin
Reference Guide on Medical Testimony, 687	John B. Wong, Lawrence O. Gostin, & Oscar A. Cabrera
Reference Guide on Neuroscience, 747	Henry T. Greely & Anthony D. Wagner
Reference Guide on Mental Health Evidence, 813	Paul S. Appelbaum
Reference Guide on Engineering, 897	Channing R. Robertson, John E. Moalli, & David L. Black
Appendix A. Biographical Information of Committee and Staff, 961	

Reference Guide on Medical Testimony

JOHN B. WONG, LAWRENCE O. GOSTIN, AND
OSCAR A. CABRERA

John B. Wong, M.D., is Chief of the Division of Clinical Decision Making, Informatics, and Telemedicine at the Institute for Clinical Research and Health Policy Studies, Tufts Medical Center, and Professor of Medicine at Tufts University School of Medicine.

Lawrence O. Gostin, J.D., is Linda D. and Timothy J. O'Neill Professor of Global Health Law and Faculty Director of O'Neill Institute for National and Global Health Law, Georgetown University Law Center.

Oscar A. Cabrera, Abogado, LL.M., is Deputy Director of the O'Neill Institute for National and Global Health Law and Adjunct Professor of Law, Georgetown University Law Center.

CONTENTS

- I. Introduction, 689
- II. Medical Testimony Introduction, 689
 - A. Medical Versus Legal Terminology, 689
 - B. Applicability of *Daubert v. Merrell Dow Pharmaceuticals, Inc.*, 692
 - C. Relationship of Medical Reasoning to Legal Reasoning, 693
- III. Medical Care, 695
 - A. Medical Education and Training, 695
 - 1. Medical school, 695
 - 2. Postgraduate training, 697
 - 3. Licensure and credentialing, 698
 - 4. Continuing medical education, 700
 - B. Organization of Medical Care, 700
 - C. Patient Care, 702
 - 1. Goals, 702
 - 2. Patient-physician encounters, 703
- IV. Medical Decisionmaking, 704
 - A. Diagnostic Reasoning, 704
 - 1. Clinical reasoning process, 705
 - 2. Probabilistic reasoning and Bayes' rule, 707
 - 3. Causal reasoning, 714

- B. Testing, 717
 - 1. Screening, 717
 - 2. Diagnostic testing, 719
 - 3. Prognostic testing, 721
 - C. Judgment and Uncertainty in Medicine, 721
 - 1. Variation in medical care, 721
 - 2. Evidence-based medicine, 722
 - 3. Hierarchy of medical evidence, 723
 - 4. Guidelines, 726
 - 5. Vicissitudes of therapeutic decisionmaking, 728
 - D. Informed Consent, 734
 - 1. Principles and standards, 734
 - 2. Risk communication, 737
 - 3. Shared decisionmaking, 739
 - V. Summary and Future Directions, 740
- Glossary of Terms, 742
- References on Medical Testimony, 745

I. Introduction

Physicians are a common sight in today's courtroom. A survey of federal judges published in 2002 indicated that medical and mental health experts constituted more than 40% of the total number of testifying experts.¹ Medical evidence is a common element in product liability suits,² workers' compensation disputes,³ medical malpractice suits,⁴ and personal injury cases.⁵ Medical testimony may also be critical in certain kinds of criminal cases.⁶ The goal of this reference guide is to introduce the basic concepts of diagnostic reasoning and clinical decisionmaking, as well as the types of evidence that physicians use to make judgments as treating physicians or as experts retained by one of the parties in a case. Following this introduction (Section I), Section II identifies a few overarching theoretical issues that courts face in translating the methods and techniques customary in the medical profession in a manner that will serve the court's inquiry. Sections III and IV describe medical education and training, the organization of medical care, the elements of patient care, and the processes of diagnostic reasoning and medical judgment. When relevant, each subsection includes examples from case law illustrating how the topic relates to legal issues.

II. Medical Testimony Introduction

A. Medical Versus Legal Terminology

Because medical testimony is common in the courtroom generally and indispensable to certain kinds of cases, courts have employed some medical terms in ways

1. Joe S. Cecil, *Ten Years of Judicial Gatekeeping Under Daubert*, 95 Am. J. Pub. Health S74–S80 (2005).

2. See, e.g., *In re Bextra & Celebrex Mktg. Sales Practices and Prod. Liab.*, 524 F. Supp. 2d 1166 (N.D. Cal. 2007) (thoroughly reviewing the proffered testimony of plaintiff's expert cardiologist and neurologist in a products liability suit alleging that defendant's arthritis pain medication caused serious cardiovascular injury).

3. See, e.g., *AT&T Alascom v. Orchitt*, 161 P.3d 1232 (Alaska 2007) (affirming the decision of the state workers' compensation board and rejecting appellant's challenges to worker's experts).

4. *Schneider ex rel. Estate of Schneider v. Fried*, 320 F.3d 396 (3d Cir. 2003) (allowing a physician to testify in a malpractice case regarding whether administering a particular drug during angioplasty was within the standard of care).

5. See, e.g., *Epp v. Lauby*, 715 N.W.2d 501 (Neb. 2006) (detailing the opinions of two physicians regarding whether plaintiff's fibromyalgia resulted from an automobile accident with two defendants).

6. Medical evidence will be at issue in numerous kinds of criminal cases. See *State v. Price*, 171 P.3d 293 (Mont. 2007) (an assault case in which a physician testified regarding the potential for a stun gun to cause serious bodily harm); *People v. Unger*, 749 N.W.2d 272 (Mich. Ct. App. 2008) (a second-degree murder case involving testimony of a forensic pathologist and neuropathologist); *State v. Greene*, 951 So. 2d 1226 (La. Ct. App. 2007) (a child sexual battery and child rape case involving the testimony of a board-certified pediatrician).

that differ from their use by the medical profession. Differential diagnosis, for example, is an accepted method that a medical expert may employ to offer expert testimony that satisfies *Daubert*.⁷ In the legal context, differential diagnosis refers to a technique “in which physician first rules in all scientifically plausible causes of plaintiff’s injury, then rules out least plausible causes of injury until the most likely cause remains, thereby reaching conclusion as to whether defendant’s product caused injury. . . .”⁸ In the medical context, by contrast, differential diagnosis

7. See, e.g., Feliciano-Hill v. Principi, 439 F.3d 18, 25 (1st Cir. 2006) (“[W]hen an examining physician calls upon training and experience to offer a differential diagnosis . . . most courts have found no *Daubert* problem.”); Clausen v. M/V New Carissa, 339 F.3d 1049, 1058–59 (9th Cir. 2003) (recognizing differential diagnosis as a valid methodology); Mattis v. Carlon Elec. Prods., 295 F.3d 856, 861 (8th Cir. 2002) (“A medical opinion based upon a proper differential diagnosis is sufficiently reliable to satisfy [*Daubert*.]”); Westberry v. Gislaved Gummi AB, 178 F.3d 257, 262 (4th Cir. 1999) (recognizing differential diagnosis as a reliable technique).

8. Wilson v. Taser Int’l, Inc. 2008 WL 5215991, at *5 (11th Cir. Dec. 16, 2008) (“[N]onetheless, Dr. Meier did not perform a differential diagnosis or any tests on Wilson to rule out osteoporosis and these corresponding alternative mechanisms of injury. Although a medical expert need not rule out every possible alternative in order to form an opinion on causation, expert opinion testimony is properly excluded as unreliable if the doctor ‘engaged in very few standard diagnostic techniques by which doctors normally rule out alternative causes and the doctor offered no good explanation as to why his or her conclusion remained reliable’ or if ‘the defendants pointed to some likely cause of the plaintiff’s illness other than the defendants’ action and [the doctor] offered no reasonable explanation as to why he or she still believed that the defendants’ actions were a substantial factor in bringing about that illness.’”); Williams v. Allen, 542 F.3d 1326, 1333 (11th Cir. 2008) (“Williams also offered testimony from Dr. Eliot Gelwan, a psychiatrist specializing in psychopathology and differential diagnosis. Dr. Gelwan conducted a thorough investigation into Williams’ background, relying on a wide range of data sources. He conducted extensive interviews with Williams and with fourteen other individuals who knew Williams at various points in his life.”) (involving a capital murder defendant petitioning for habeas corpus offering supporting expert witness); Bland v. Verizon Wireless, L.L.C., 538 F.3d 893, 897 (8th Cir. 2008) (“Bland asserts Dr. Sprince conducted a differential diagnosis which supports Dr. Sprince’s causation opinion. We have held, ‘a medical opinion about causation, based upon a proper differential diagnosis is sufficiently reliable to satisfy *Daubert*.’ A ‘differential diagnosis [is] a technique that identifies the cause of a medical condition by eliminating the likely causes until the most probable cause is isolated.’”) (stating expert’s incomplete execution of differential diagnosis procedure rendered expert testimony unsatisfactory for *Daubert* standard) (citations omitted); Lash v. Hollis 525 F.3d 636, 640 (8th Cir. 2008) (“Further, even if the treating physician had specifically opined that the Taser discharges caused rhabdomyolysis in Lash Sr., the physician offered no explanation of a differential diagnosis or other scientific methodology tending to show that the Taser shocks were a more likely cause than the myriad other possible causes suggested by the evidence.”) (finding lack of expert testimony with differential diagnosis enough to render evidence insufficient for jury to find causation in personal injury suit); Feit v. Great West Life & Annuity Ins. Co., 271 Fed. App’x. 246, 254 (3d Cir. 2008) (“However, although this Court generally recognizes differential diagnosis as a reliable methodology the differential diagnosis must be properly performed in order to be reliable. To properly perform a differential diagnosis, an expert must perform two steps: (1) ‘Rule in’ all possible causes of Dr. Feit’s death and (2) ‘Rule out’ causes through a process of elimination whereby the last remaining potential cause is deemed the most likely cause of death.”) (ruling that district court not in error for excluding expert medical testimony that relied on an improperly performed differential diagnosis) (citations omitted); Glastetter v. Novartis Pharms. Corp., 252 F.3d 986 (8th Cir. 2001).

refers to a set of diseases that physicians consider as possible causes for symptoms the patient is suffering or signs that the patient exhibits.⁹ By identifying the likely potential causes of the patient's disease or condition and weighing the risks and benefits of additional testing or treatment, physicians then try to determine the most appropriate approach—testing, medication, or surgery, for example.¹⁰

Less commonly, courts often have used the term “differential etiology” interchangeably with differential diagnosis.¹¹ In medicine, etiology refers to the study of causation in disease,¹² but differential etiology is a legal invention not used by physicians. In general, both differential etiology and differential diagnosis are concerned with establishing or refuting causation between an external cause and a plaintiff's condition. Depending on the type of case and the legal standard, a medical expert may testify in regard to specific causation, general causation, or both. General causation refers to whether the plaintiff's injury could have been caused by the defendant, or a product produced by the defendant, while specific causation is established only when the defendant's action or product actually caused the harm.¹³ An opinion by a testifying physician may be offered in support of both kinds of causation.¹⁴

Courts also refer to medical certainty or probability in ways that differ from their use in medicine. The standards “reasonable medical certainty” and “reasonable medical probability” are also terms of art in the law that have no analog for a practicing physician.¹⁵ As is detailed in Section IV, diagnostic reasoning and medi-

9. Steadman's Medical Dictionary 531 (28th ed. 2006) (defining differential diagnosis as “the determination of which of two or more diseases with similar symptoms is the one from which the patient is suffering, by a systematic comparison and contrasting of the clinical findings.”).

10. The Concise Dictionary of Medical-Legal Terms 36 (1998) (definition of differential diagnosis).

11. See *Proctor v. Fluor Enters., Inc.* 494 F.3d 1337 (11th Cir. 2007) (testifying medical expert employed differential etiology to reach a conclusion regarding the cause of plaintiff's stroke). *But see McClain v. Metabolife Int'l, Inc.*, 401 F.3d 1233, 1252 (11th Cir. 2005) (distinguishing differential diagnosis from differential etiology, with the former closer to the medical definition and the latter employed as a technique to determine external causation).

12. Steadman's Medical Dictionary 675 (28th ed. 2006) (defining etiology as “the science and study of the causes of disease and their mode of operation. . .”). For a discussion of the term “etiology” in epidemiology studies, see Michael D. Green et al., *Reference Guide on Epidemiology*, Section I, in this manual.

13. See *Amorgianos v. Nat'l R.R. Passenger Corp.*, 303 F.3d 256, 268 (2d Cir. 2002).

14. See, e.g., *Ruggiero v. Warner-Lambert Co.* 424 F.3d 249 (2d Cir. 2005) (excluding testifying expert's differential diagnosis in support of a theory of general causation because it was not supported by sufficient evidence).

15. See, e.g., *Dallas v. Burlington N., Inc.*, 689 P.2d 273, 277 (Mont. 1984) (“[R]easonable medical certainty’ standard; the term is not well understood by the medical profession. Little, if anything, is ‘certain’ in science. The term was adopted in law to assure that testimony received by the fact finder was not merely conjectural but rather was sufficiently probative to be reliable”). This reference guide will not probe substantive legal standards in any detail, but there are substantive differences in admissibility standards for medical evidence between federal and state courts. See Robin Dundis Craig, *When Daubert Gets Erie: Medical Certainty and Medical Expert Testimony in Federal Court*, 77 Denv. U. L. Rev. 69 (1999).

cal evidence are aimed at recommending the best therapeutic option for a patient. Although most courts have interpreted “reasonable medical certainty” to mean a preponderance of the evidence,¹⁶ physicians often work with multiple hypotheses while diagnosing and treating a patient without any “standard of proof” to satisfy.

Statutes and administrative regulations may also contain terms that are borrowed, often imperfectly, from the medical profession. In these cases, the court may need to examine the intent of the legislature and the term’s usage in the medical profession.¹⁷ If no intent is apparent, the court may need to determine whether the medical definition is the most appropriate one to apply to the statutory language. Whether the language is a term of art or a question of law will often dictate the admissibility and weight of evidence.¹⁸

B. *Applicability of Daubert v. Merrell Dow Pharmaceuticals, Inc.*

The Supreme Court’s decision in *Daubert v. Merrell Dow Pharmaceuticals, Inc.*,¹⁹ changed the way that judges screen expert testimony. A 2002 study by the RAND Corporation indicated that after *Daubert*, judges began scrutinizing expert testimony much more closely and began more aggressively excluding evidence that does not meet its standards.²⁰ Despite the Court’s subsequent decisions in *General Electric Co. v. Joiner*²¹ and *Kumho Tire Co. v. Carmichael*²² further defining the

16. See, e.g., *Sharpe v. United States*, 230 F.R.D. 452, 460 (E.D. Va. 2005) (“It is not enough for the plaintiff’s expert to testify that the defendant’s negligence might or may have caused the injury on which the plaintiff bases her claim. The expert must establish that the defendant’s negligence was ‘more likely’ or ‘more probably’ the cause of the plaintiff’s injury . . .”).

17. See, e.g., *Feltner v. Lamar Adver., Inc.*, 83 F. App’x 101 (6th Cir. 2003) (holding that the statutory definition of “permanent total disability” under the Tennessee Workers Compensation Act was not the same as the medical definition); *Endorf v. Bohlender*, 995 P.2d 896 (Kan. Ct. App. 2000) (a medical malpractice case reversing a lower court’s interpretation of the statutory phrase “clinical practice” because it did not comport with the legislature’s intent that the statutory meaning reflect the medical definition).

18. See, e.g., *Coleman v. Workers’ Comp. Appeal Bd. (Ind. Hosp.)*, 842 A.2d 349 (Pa. 2004) (holding that since the legislature did not define the medical term “physical examination,” the common usage of the term is more appropriate than the strict medical definition).

19. 509 U.S. 579 (1993).

20. Lloyd Dixon & Brian Gill, *Changes in the Standards for Admitting Expert Evidence in Federal Civil Cases Since the Daubert Decision* (2002).

21. 522 U.S. 136 (1997) (holding that the trial court had properly excluded expert testimony extrapolated from animal studies and epidemiological studies).

22. 526 U.S. 137 (1999). In *Kumho*, the Court made clear that *Daubert* applies to all expert testimony and not just “scientific” testimony. Although the case involved a defect in tires, courts before *Kumho* were divided on whether expert medical opinion based on experience or clinical medical testimony were subject to *Daubert*. See also Joe S. Cecil, *Ten Years of Judicial Gatekeeping Under Daubert*, 95 Am. J. Pub. Health S74–S80 (2005). See also Lawrence O. Gostin, *Public Health Law: Power, Duty, Restraint* (2d ed. 2008).

Daubert standard, federal and state courts have sometimes employed conflicting interpretations of what *Daubert* requires from testifying physicians.

The standard of review is an important factor in understanding how *Daubert* has engendered seemingly inconsistent results. The Supreme Court adopted an abuse of discretion standard in *Joiner*²³ and affirmed it in *Kumho*.²⁴ Although in most product liability cases the courts reached the same conclusion, inconsistent determinations regarding the admissibility of similar evidence may not constitute an abuse of discretion under the federal standard of review or in states with a similar standard.²⁵

C. Relationship of Medical Reasoning to Legal Reasoning

As Section II.A suggested, the goal that guides the physician—recommending the best therapeutic options for the patient—means that diagnostic reasoning and the process of ongoing patient care and treatment involve probabilistic judgments concerning several working hypotheses, often simultaneously. When a court requires a testifying physician to offer evidence “to a reasonable medical certainty” or “reasonable medical probability,” it is supplying the expert with a legal rule to which his or her testimony must conform.²⁶ In other words, a lawyer often will

23. 522 U.S. at 143.

24. 526 U.S. at 142.

25. *Hollander v. Sandoz Pharm. Corp.*, 289 F.3d 1193, 1207 (10th Cir. 2002); *see also* *Brasher v. Sandoz Pharm. Corp.*, 160 F. Supp. 2d 1291, 1298 n.17 (N.D. Ala. 2001); *Reichert v. Phipps*, 84 P.3d 353, 358 (Wyo. 2004).

26. Courts have occasionally noted the tension between the medical reasoning and legal reasoning when applying the reasonable medical certainty or reasonable medical probability standards. *See Clark v. Arizona*, 548 U.S. 735, 777 (2006) (“When . . . ‘ultimate issue’ questions are formulated by the law and put to the expert witness who must then say ‘yea’ or ‘nay,’ then the expert witness is required to make a leap in logic. He no longer addresses himself to medical concepts but instead must infer or intuit what is in fact unspeakable, namely, the *probable relationship* between medical concepts and legal or moral constructs such as free will. These impermissible leaps in logic made by expert witnesses confuse the jury. . . .”); *Rios v. City of San Jose*, 2008 U.S. Dist. LEXIS 84923, at *4 (N.D. Cal. Oct. 9, 2008) (“In their fifth motion, plaintiffs seek to exclude the testimony of Dr. Brian Peterson who defendants designated to testify, among other subjects, about the ‘proximate cause’ of Rios’ death. As the use of terms that also carry legal significance could confuse the jury, the motion is granted in part, and defendants are instructed to distinguish between medical and legal terms such as proximate cause to the extent possible. Where such terms must be used by the witness consistent with the language employed in his field of expertise, the parties shall craft a limiting instruction to advise the jury of the distinction between those terms and the issues they will be called upon to determine.”); *Norland v. Wash. Gen. Hosp.*, 461 F.2d 694, 697 (8th Cir. 1972) (“The use of the terms ‘probable’ and ‘possible’ as a basis for test of qualification or lack of qualification in respect to a medical opinion has frequently converted this aspect of a trial into a mere semantic ritual or hassle. The courts have come to recognize that the competency of a physician’s testimony cannot soundly be permitted to turn on a mechanical rule of law as to which of the two terms he has employed. Regardless of which term he may have used, if his testimony is such in nature and basis of hypothesis as to judicially impress that the opinion expressed represents his professional judgment as to the most likely one among the

need to explain the legal standard to the physician, who will then shape the form and content of his or her testimony in a manner that serves the legal inquiry.²⁷

Legal standards will shape how physicians testify in a number of other ways. Although treating physicians generally are concerned less about discovering the actual causes of the disease than treating the patient, the testifying medical expert will need to tailor his or her opinions in a way that conforms to the legal standard of causation. As Section IV will demonstrate, when analyzing the patient's symptoms and making a judgment based on the available medical evidence, a physician will not expressly identify a "proximate cause" or "substantial factor." For example, in order to recommend treatment, a physician does not necessarily need to determine whether a patient's lung ailment was more likely the result of a long history of tobacco use or prolonged exposure to asbestos if the optimal treatment is the same. In contrast, when testifying as an expert in a case in which an employee with a long history of tobacco use is suing his employer for possible injuries as a result of asbestos exposure in the workplace, physicians may need to make judgments regarding the likelihood that either tobacco or asbestos—or both—could have contributed to the injury.²⁸

Physicians often will be asked to testify about patients from whom they have never taken a medical history or examined and make estimates about proximate cause, increased risk of injury, or likely future injuries.²⁹ The doctor may even need to make medical judgments about a deceased litigant.³⁰ Testifying in all such cases requires making judgments that physicians do not ordinarily make in their profession, making these judgments outside of physicians' customary patient encounters, and adapting the opinion in a way that fits the legal standard. The purpose of this guide is not to describe or recommend competing legal standards, whether it be the standard of proof, causation, admissibility, or the applicable standard of care in medical malpractice cases. Instead, it aims to introduce the practice of medicine to federal and state judges, emphasizing the tools and methods that

possible causes of the physical condition involved, the court is entitled to admit the opinion and leave its weight to the jury.").

27. There are several cases that demonstrate the difficulty that physicians sometimes have in adapting their testimony to the legal standard. See *Schranz v. Luancing*, 527 A.2d 967 (N.J. Super. Ct. Law Div. 1986) (malpractice case in which the medical expert's opinion was inadequate because of her understanding of "reasonable medical certainty").

28. Physicians will testify as experts in cases in which the plaintiff's condition may be the result of multiple causes. In these cases, the divergence between medical reasoning and legal reasoning are very apparent. See, e.g., *Tompkin v. Philip Morris USA, Inc.*, 362 F.3d 882 (6th Cir. 2004) (affirming district court's conclusion that testimony offered by the defendant's expert regarding the decedent's work-related asbestos exposure was not prejudicial in a suit against a tobacco company on behalf of plaintiff's deceased husband); *Mobil Oil Corp. v. Bailey*, 187 S.W.3d 265 (Tex. Ct. App. 2006) (involving claims from a worker who had a long history of tobacco use that exposure to asbestos increased his risk of cancer).

29. See, e.g., *Tompkin*, 362 F.3d 882.

30. See, e.g., *id.*

doctors use to make decisions and highlighting the challenges in adapting them when testifying as medical experts.

Sections III and IV of this guide explain in great detail the practice of medicine, including medical education, the structure of health care, and, most importantly, the methods that physicians use to diagnose and treat their patients. Special attention is given to the physician–patient relationship and to the types of evidence that physicians use to make medical judgments. In an effort to make each issue more salient, examples from case law are offered when they are illustrative.

III. Medical Care

A. Medical Education and Training

1. Medical school

The Association of American Medical Colleges (AAMC) consists of 133 accredited U.S. medical schools and 17 Canadian medical schools.³¹ The Liaison Committee on Medical Education performs the accreditation for AAMC and assesses the quality of postsecondary education by determining whether each institution or program meets established standards for function, structure, and performance. The goal of medical school is to prepare students in the art and science of medicine for graduate medical education.³² Of the 4 years of medical school, the first 2 years are typically spent studying preclinical basic sciences involving the study of the normal structure and function of human systems (e.g., through anatomy, biochemistry, physiology, behavioral science, and neuroscience), followed by the study of abnormalities and therapeutic principles (e.g., through microbiology, immunology, pharmacology, and pathology). The final 2 years involve clinical experience, including rotations in patient care settings such as clinics or hospitals with required “core” clerkships in internal medicine, pediatrics, psychiatry, surgery, obstetrics/gynecology, and family medicine. All physicians who wish to be licensed must pass the United States Medical Licensing Examination Steps 1, 2, and 3.³³

31. Association of American Medical Colleges, Membership, available at <https://www.aamc.org/about/membership/> (last visited Feb. 12, 2011).

32. See *Davis v. Houston Cnty., Ala. Bd. of Educ.*, 2008 WL 410619 (M.D. Ala. Feb. 13, 2008) (finding that an individual with no medical training was not qualified to give expert testimony).

33. *Planned Parenthood Cincinnati Region v. Taft* 444 F.3d 502, 515 (6th Cir. 2006), (“The State has not appealed the district court’s order refusing to recognize Dr. Crockett as an expert in the critical review of medical literature. Although that order has not been placed before us, the only reason the district court gave for her ruling was that Dr. Crockett did not have any specific training in the critical review of medical literature beyond the training incorporated in her general medical school and residency training. This ruling ignored Dr. Crockett’s testimony that her residency program at Georgetown University put particular emphasis on training residents in the critical review of medical literature, that she had taught classes on the subject, that she had done extensive reading and

In the United States, besides the more than 941,000 physicians, there are more than 61,000 doctors of osteopathy. The Commission on Osteopathic College Accreditation accredits 25 colleges of osteopathic medicine. Training is similar to that for medical physicians but with additional “special attention on the musculoskeletal system which reflects and influences the condition of all other body systems.”³⁴ About 25% of current U.S. physicians are foreign medical graduates that include both U.S. citizens and foreign nationals.³⁵ Because educational standards and curricula outside the United States and Canada vary, the Education Commission for Foreign Medical Graduates has developed a certification exam to assess whether these graduates may enter Accreditation Council for Graduate Medical Education (ACGME) accredited residency and fellowship programs.³⁶

self-education on the subject, and that she had critically reviewed medical literature for the FDA. If these qualifications are not sufficient to demonstrate expertise, this court is hard-pressed to imagine what qualifications would suffice.”); *Davis v. Houston Cnty., Ala. Bd. of Educ.*, 2008 WL 410619, at *4 (M.D. Ala. Feb. 13, 2008) (“The Board has moved to exclude all evidence of Freet’s opinions and conclusions related to the cause of Joshua Davis’s behavior at the football game contained in his deposition as well as Freet’s letter to Malcolm Newman. The Board argues that Freet is not qualified to give expert testimony, and that Plaintiff failed to comply with Fed. R. Civ. P. 26(a)(2)(B) by not providing a report of Freet’s testimony that includes all of the information required by Rule 26(a)(2)(B). . . . In order to consider Freet’s expert opinions, this Court must find that Freet meets the requirements of Fed. R. Evid. 702. Rule 702 requires an expert to be qualified by ‘knowledge, skill, experience, training, or education.’ Freet is not a medical doctor and never attended medical school. The only evidence of Freet’s qualifications are: approximately five years working for the Department of Veterans Affairs in the vocational rehabilitation program, followed by approximately seven years working in private practice as a ‘licensed professional counselor.’ There is no evidence in the record of Freet’s educational background, or any details of the exact nature of Freet’s work experience.”); *Therrien v. Town of Jay*, 489 F. Supp. 2d 116, 117 (D. Me. 2007) (“Citing *Daubert v. Merrell Dow Pharmaceuticals, Inc.*, 509 U.S. 579, 113 S. Ct. 2786, 125 L. Ed. 2d 469 (1993) and Rule 702 of the Federal Rules of Evidence, Officer Gould’s first objection is that Dr. Harding does not possess sufficient expertise to express expert opinions about ‘the mechanism and timing of Plaintiff’s injuries.’ This objection is not well taken. Dr. Harding was graduated from Dartmouth College and Georgetown Medical School; he completed a residency in internal medicine, is board certified in internal medicine, and has been licensed to practice medicine in the state of Maine since 1978.”). United States Medical Licensing Examination, Examinations, available at <http://www.usmle.org/Examinations/index.html> (last visited Aug. 9, 2011).

34. Association of American Medical Colleges, What is a DO? available at <http://www.osteopathic.org/osteopathic-health/about-dos/what-is-a-do/Pages/default.aspx> (last visited Feb. 12, 2011); Association of American Medical Colleges, About Osteopathic Medicine, available at <http://www.osteopathic.org/osteopathic-health/about-dos/about-osteopathic-medicine/Pages/default.aspx> (last visited Feb. 12, 2011).

35. American Medical Association, Physician Characteristics and Distribution in the U.S. (2009).

36. Commission for Foreign Medical Graduates, About ECFMG, available at <http://www.ecfmg.org/about.html> (last visited Feb. 12, 2011).

2. Postgraduate training

After graduating from medical school, most physicians undergo additional training in a residency program in a chosen specialty.³⁷ Residencies typically range from 3 to 7 years at teaching hospitals and academic medical centers where residents care for patients while being supervised by physician faculty and participating in educational and research activities.³⁸ After graduating from an accredited residency program, physicians become eligible to take their board certification examinations.³⁹ Physician licensure in many states requires the completion of a residency program accredited by the ACGME, the organization which is responsible for accrediting the more than 8700 residency programs in 26 specialties and 130 subspecialties.⁴⁰ Following residency, some physicians opt for additional subspecialty fellowship training. ACGME divides fellowship training⁴¹ into (1) Dependent Subspecialty Programs in which the program functions in conjunction with an accredited specialty/core program and (2) Independent Subspecialty Programs in which the program does not depend on the accreditation status of a specialty program.⁴² For osteopathic physicians, the American Osteopathic Association approves osteopathic postdoctoral

37. See *Brown v. Harmot Med. Ctr.*, 2008 WL 55999 (W.D. Pa. Jan. 3, 2008). American Medical Association, Requirements for Becoming a Physician, *available at* <http://www.ama-assn.org/ama/pub/education-careers/becoming-physician.page?> (last visited Aug. 9, 2011).

38. See *Planned Parenthood Cincinnati Region v. Taft*, 444 F.3d 502, 515 (6th Cir. 2006). American Medical Association, Requirements for Becoming a Physician, *available at* <http://www.ama-assn.org/ama/pub/education-careers/becoming-physician.page?> (last visited Aug. 9, 2011).

39. See *Therrien v. Town of Jay*, 489 F. Supp. 2d 116, 117 (D. Me. 2007) (finding that a physician who completed a residency in internal medicine was qualified to give his opinion on trauma related to a § 1983 claim against a police department). American Medical Association, Requirements for Becoming a Physician, *available at* <http://www.ama-assn.org/ama/pub/education-careers/becoming-physician.page?> (last visited Aug. 9, 2011).

40. Accreditation Council for Graduate Medical Education, The ACGME at a Glance, *available at* http://www.acgme.org/acWebsite/newsRoom/newsRm_acGlance.asp (last visited Feb. 12, 2011).

41. Accreditation Council for Graduate Medical Education, Specialty Programs with Dependent and Independent Subspecialties, *available at* http://www.acgme.org/acWebsite/RRC_sharedDocs/sh_progs_depIndSubs.asp (last visited Feb. 12, 2011).

42. *John Doe 21 v. Sec'y of Health and Human Servs.*, 84 Fed. Cl. 19, 35–36 (Fed. Cl. 2008) (“The Government’s expert, Dr. Wiznitzer, is a board-certified neurologist by the American Board of Psychiatry and Neurology, with a special qualification in Child Neurology. In addition, Dr. Wiznitzer is certified by the American Board of Pediatrics. Since 1986, Dr. Wiznitzer has been an Associate Pediatrician and an Associate Neurologist at University Hospital of Cleveland, Ohio. And, since 1992, Dr. Wiznitzer has been Director of the Autism Center at Rainbow Babies and Children’s Hospital in Cleveland, Ohio. During the past 24 years, Dr. Wiznitzer also has been an Associate Professor of Pediatrics and Associate Professor of Neurology at Case Western Reserve University. Dr. Wiznitzer completed his residency in Pediatrics from Children’s Hospital Medical Center in Cincinnati and served as a Fellow in Developmental Disorders, Pediatric Neurology, and Higher Cortical Functions. Dr. Wiznitzer also has received numerous awards and honors in the neurology field and his work has been widely published.”) (citations omitted); *Brown v. Hamot Med. Ctr.*, 2008 WL 55999, at *8–9

training programs.⁴³ The American Osteopathic Association established the Osteopathic Postdoctoral Training Institutions (OPTI), wherein each OPTI partners a community-based training consortium with one or more colleges of osteopathic medicine and one or more hospitals and possibly ambulatory care facilities.⁴⁴

3. *Licensure and credentialing*

Medical Practice Acts defining the practice of medicine and delegating enforcement to state medical boards exist for each of the 50 states, the District of Columbia, and the U.S. territories. Besides awarding medical licenses, state medical boards also investigate complaints, discipline physicians who violate the law, and evaluate and rehabilitate physicians. The Federation of State Medical Boards represents the 70 medical boards of the United States and its territories, and its mission is “promoting excellence in medical practice, licensure, and regulation as the national resource and voice on behalf of state medical boards in their protection of the public.”⁴⁵

Credentialing typically involves verifying medical education, postgraduate training, board certification, professional experience, state licensure, prior credentialing outcomes, medical board actions, malpractice, and adverse clinical events. Credentialing or recredentialing by hospitals involves an assessment of a physician’s professional or technical competence and performance by evaluating and monitoring the quality of patient care. This credentialing process defines physicians’ scope of practice and hospital privileges, that is, the clinical services they may provide.

The American Board of Medical Specialties (ABMS) provides certification in 24 medical specialties (e.g., emergency medicine, internal medicine, obstetrics and gynecology, family medicine, pediatrics, surgery, and others) to provide⁴⁶ “assurance of a physician’s expertise in a particular specialty and/or subspecialty

(W.D. Pa. Jan. 3, 2008) (“As the United States Court of Appeals for the Fifth Circuit has explained in another context, a medical residency is primarily an academic enterprise:

[a] residency program is distinct from other types of employment in that the resident’s “work” is what is academically supervised and evaluated. [T]he primary purpose of a residency program is not employment or a stipend, but the academic training and the academic certification for successful completion of the program. The certificate . . . tells the world that the resident has successfully completed a course of training and is qualified to pursue further specialized training or to practice in specified areas. . . . Successful completion of the residency program depends upon subjective evaluations by trained faculty members into areas of expertise that courts are poorly equipped to undertake in the first instance or to review. . . .”).

43. American Osteopathic Association, Postdoctoral Training, *available at* <http://www.osteopathic.org/inside-aoa/Education/postdoctoral-training/Pages/default.aspx> (last visited Feb. 12, 2011).

44. *Id.*

45. Federation of State Medical Boards, FSMB Mission and Goals, *available at* <http://www.fsmb.org/mission.html> (last visited Feb. 12, 2011).

46. American Board of Medical Specialties, Who We Are and What We Do, *available at* http://www.abms.org/About_ABMS/who_we_are.aspx (last visited Feb. 12, 2011).

of medical practice.”⁴⁷ Although the criteria vary depending on the field, board eligibility requires the completion of an appropriate residency, an institutional or valid license to practice medicine, and evaluation with written and—in some cases—oral examinations. Many boards also require an evaluation of practice performance for initial certification. Board certification documents the fulfillment of all criteria including passing the examinations. Originally, board certificates had no expiration, but a program of periodic recertification (every 6 to 10 years) was subsequently initiated to ensure that physicians remained current in their specialty. In 2006, the ABMS recertification process became the Maintenance of Certification to emphasize continuous professional development through a four-part process:

1. Licensure and professional standing;
2. Lifelong learning;
3. Cognitive expertise; and
4. Practice performance assessment in six core competencies
 - a. patient care,
 - b. medical knowledge,
 - c. practice-based learning,
 - d. interpersonal and communications skills,
 - e. professionalism, and
 - f. systems-based practice.⁴⁸

In some cases, specialty organizations have opted to develop their own certification process outside of the ABMS (e.g., the American Board of Bariatric Medicine).⁴⁹

The American Osteopathic Association (AOA) certifies osteopathic physicians in 18 osteopathic specialty boards (e.g., emergency medicine, internal medicine, obstetrics and gynecology, family medicine, pediatrics, surgery, and others).⁵⁰ The osteopathic continuous certification process involves (1) unrestricted licensure, (2) lifelong learning/continuing medical education, (3) cognitive assessment, (4) practice performance assessment and improvement, and (5) continuous AOA membership.⁵¹

47. Although specialization is a hallmark of modern medical practice, courts have not always required that medical testimony come from a specialist. See *Gaydar v. Sociedad Instituto Gineco-Quirurgico y Planificacion Familiar*, 245 F.3d 15, 24–25 (1st Cir. 2003) (“The proffered expert physician need not be a specialist in a particular medical discipline to render expert testimony relating to that discipline.”).

48. American Board of Medical Specialties, ABMS Maintenance of Certification, available at http://www.abms.org/Maintenance_of_Certification/ABMS_MOC.aspx (last visited Feb. 12, 2011).

49. American Board of Bariatric Medicine, Certification, available at <http://www.abmcertification.org/> (last visited Feb. 12, 2011).

50. American Osteopathic Association, AOA Specialty Certifying Boards, available at <http://www.osteopathic.org/inside-aoa/development/aoa-board-certification/Pages/aoa-specialty-boards.aspx> (last visited Feb. 12, 2011).

51. *Id.*

4. Continuing medical education

For relicensure, state medical boards require continuing medical education so that physicians can acquire new knowledge and maintain clinical competence. The Accreditation Council for Continuing Medical Education (ACCME) identifies, develops, and promotes quality standards for continuing medical education for physicians. ACCME requires certain elements of structure, method, and organization in the development of continuing medical education materials to ensure uniformity across states and to help assure physicians, state medical boards, medical societies, state legislatures, continuing medical education providers, and the public that the education meets certain quality standards. For osteopathic physicians, the AOA Board of Trustees also oversees accreditation for osteopathic CME sponsors through the Council on Continuing Medical Education (CCME).⁵² The AOA's Healthcare Facilities Accreditation Program (HFAP) reviews services delivered by medical facilities.⁵³

B. Organization of Medical Care

The delivery of health care in the United States is highly decentralized and fragmented,⁵⁴ and is provided through clinics, hospitals, managed care organizations, medical groups, multispecialty clinics, integrated delivery systems, specialty standalone hospitals, imaging facilities, skilled nursing facilities, rehabilitation hospitals, emergency departments, and pharmacy-based and other walk-in clinics. When surveyed in 1996, patients viewed the health care system as a “nightmare to navigate.”⁵⁵ Transitioning care from outpatient to inpatient hospitalization to recovery often involves multiple handoffs among different physicians and care providers with the need for accurate, timely, and complete transfer of information about the patient's acute and chronic medical conditions, medications, and treatments. Although hospitals increasingly belong to a network or system, most community physicians belong to practices involving 10 or fewer physicians.⁵⁶

Concerns about the safety of the organization of medical care first arose from the Harvard Medical Practice Study which found that adverse events occurred in

52. American Osteopathic Association, Continuing Medical Education, *available at* <http://www.osteopathic.org/inside-aoa/development/continuing-medical-education/Pages/default.aspx> (last visited Feb. 12, 2011).

53. Healthcare Facilities Accreditation Program, About HFAP, *available at* <http://www.hfap.org/about/overview.aspx> (last visited Feb. 12, 2011).

54. Committee on Quality of Health Care in America, Institute of Medicine, Crossing the Quality Chasm: A New Health System for the 21st Century (2001) (hereinafter “2001 CQHCA Report”).

55. *Id.* at 28.

56. *Id.* at 28.

3.7% of hospitalizations.⁵⁷ Following some highly publicized errors (fatal medication overdoses and amputation of the limb on the wrong side), the Institute of Medicine estimated that errors resulted in as many as 98,000 deaths in patients hospitalized during 1997.⁵⁸ The report highlights “The decentralized and fragmented nature of the health care delivery system (some would say ‘nonsystem’) also contributes to unsafe conditions for patients, and serves as an impediment to efforts to improve safety.” While recognizing that “not all errors result in harm,” the report defines safety as “freedom from accidental injury” and specifies two types of error: “the failure of a planned action to be completed as intended or the use of a wrong plan to achieve an aim.”⁵⁹

Subsequently, the Institute of Medicine recommended development of a learning health care delivery system “a system that both prevents errors and learns from them when they occur. The development of such a system requires, first, a commitment by all stakeholders to a culture of safety and, second, improved information systems.”⁶⁰ Government and nongovernment institutions such as the Agency for Healthcare Research and Quality (designated as the federal lead for patient safety by the Healthcare Research and Quality Act of 1999 to “(1) identify the causes of preventable health care errors and patient injury in health care delivery; (2) develop, demonstrate, and evaluate strategies for reducing errors and improving patient safety; and (3) disseminate such effective strategies throughout the health care industry.”),⁶¹ the National Quality Forum (a nonprofit organization with multiple stakeholders developing and measuring performance standards), the Joint Commission (independent not-for-profit organization accrediting and certifying care quality and safety), Institute of Healthcare Improvement (independent not-for-profit organization fostering innovation that improves care), and the Leapfrog Group (a coalition of large employers rewarding performance) all have adopted as parts of their mission the assessment and promotion of safety at the healthcare system level. To deliver safe, effective, and efficient care, medical delivery systems having increasingly incorporated allied health professions, including nurses, nurse practitioners, physicians’ assistants, pharmacists, and therapists into care delivery.

57. Troyen A. Brennan et al., *Incidence of Adverse Events and Negligence in Hospitalized Patients: Results of the Harvard Medical Practice Study I*, 324 *New Eng. J. Med.* 370–76 (1991); Lucian L. Leape et al., *The Nature of Adverse Events in Hospitalized Patients: Results of the Harvard Medical Practice Study II*, 324 *New Eng. J. Med.* 377–84 (1991).

58. Committee on Quality of Health Care in America, Institute of Medicine, *To Err Is Human: Building a Safer Health System* 26 (2000) (hereinafter “2000 CQHCA Report”).

59. *Id.* at 4, 54, 58.

60. Committee on Data Standards for Patient Safety, Institute of Medicine, *Patient Safety: Achieving a New Standard for Care* 1 (2005).

61. Agency for Healthcare Research and Quality, *Advancing Patient Safety: A Decade of Evidence, Design and Implementation at 1*, available at <http://www.ahrq.gov/qual/advptsafety.htm> (last visited Feb. 12, 2011.)

C. Patient Care

1. Goals

The Institute of Medicine (IOM) describes quality health care delivery as “[t]he degree to which health services for individuals and populations increase the likelihood of desired health outcomes and are consistent with current professional knowledge.” The six specific aims for improving health care include

1. “Safe: avoiding injuries to patients from the care that is intended to help them;”
2. “Effective: providing services based on scientific knowledge to all who could benefit, and refraining from providing services to those not likely to benefit;”
3. “Patient-centered: providing care that is respectful of and responsive to individual patient preferences, needs, and values, and ensuring that patient values guide all clinical decisions;”
4. “Timely: reducing waits and sometimes harmful delays for both those who receive and those who give care;”
5. “Efficient: avoiding waste, including waste of equipment, supplies, ideas, and energy;” and
6. “Equitable: providing care that does not vary in quality because of personal characteristics such as gender, ethnicity, geographic location, and socioeconomic status.”⁶²

Health outcome goals include (1) improving longevity or life expectancy, (2) relieving symptoms (improving quality of life or reducing morbidity), and (3) preventing disease. These goals, however, may conflict with one another. For example, some patients may be willing to accept the chance of a reduced length of life to try to obtain a higher quality of life (e.g., if normal volunteers had a vocal cord cancer, about 20% of them would prefer radiation therapy instead of surgery to preserve their voice despite a reduction in survival⁶³), whereas others may accept reduced quality of life to try to extend life (e.g., cancer chemotherapy). Some may accept a risk of dying from a procedure to prolong life or relieve symptoms (e.g., coronary revascularization), whereas others may prefer to avoid the near-term risk of the procedure or surgery despite future benefit (risk aversion). In *Crossing the Quality Chasm*, the IOM emphasized care delivery that should accommodate individual patient choices and preferences and be customized on the basis of patients needs and values.⁶⁴

62. 2001 CQHCA Report, *supra* note 54, at 44, 5-6.

63. Barbara J. McNeil et al., *Speech and Survival: Tradeoffs Between Quality and Quantity of Life in Laryngeal Cancer*, 305 *New Eng. J. Med.* 982-87 (1981) (hereinafter “McNeil”).

64. 2001 CQHCA Report, *supra* note 54, at 49.

The Charter on Medical Professionalism avers three fundamental principles: (1) patient welfare or serving the interest of the patient, (2) patient autonomy or empowering patients to make informed decisions, and (3) social justice or fair distribution of health care resources.⁶⁵ At times, the primacy of patient welfare places the physician in conflict with social justice—for example, a patient with an acute heart attack is in the emergency room with no coronary care unit (CCU) beds available, and the most stable patient in the CCU has a 2-day-old heart attack. Transferring the patient out of the CCU places him or her at a small risk for a complication, but the CCU bed is a limited societal resource that other patients should be able to access.⁶⁶ Similarly, patients may insist on an unneeded and costly test or treatment, and the first two principles would encourage physicians to acquiesce, yet these unnecessary tests or treatments expose patients to harm and expense and also diminish resources that would otherwise be available to others.⁶⁷

2. Patient-physician encounters

A patient-physician encounter typically consists of four components: (1) patient history, (2) physical examination, (3) medical decisionmaking, and (4) counseling.⁶⁸ In many cases, patients seek medical attention because of a change in health that led to symptoms. During the patient history, physicians identify the chief complaint as the particular symptom that led the patient to seek medical evaluation. The history of the present illness includes the onset and progression of symptoms over time and may include eliciting pertinent symptoms that the patient does not exhibit. These “pertinent negatives” reduce the likelihood of certain competing diagnoses. A comprehensive encounter includes past medical history of prior illnesses, hospitalizations, surgeries, current medications, drug allergies, and lifestyle habits including smoking, alcohol use, illicit drug use, dietary habits, and exercise habits. Family history considers illnesses that have been diagnosed in related family members to identify potential genetic predispositions for disease. Social history usually includes education, employment, and social relationships and provides a socioeconomic context for developing or coping with illness and an employment context for exposure to environmental or toxin risks. Finally, the review of systems is a comprehensive checklist of symptoms that might or might not arise from the various organ systems and is an ancillary means to capture symp-

65. Medical Professionalism Project: ABIM Foundation, *Medical Professionalism in the New Millennium: A Physician Charter*, 136 *Annals Internal Med.* 243, 244 (2002).

66. Harold C. Sox et al., *Medical Decision Making* (2007).

67. Harold C. Sox, *Medical Professionalism and the Parable of the Craft Guilds*, 147 *Annals Internal Med.* 809–10 (2007).

68. See generally *Davoll v. Webb*, 194 F.3d 1116, 1138 (10th Cir. 1999) (“A treating physician is not considered an expert witness if he or she testifies about observations based on personal knowledge, including treatment of the party.”).

toms that the patient may have unintentionally neglected to mention, but which may lead physicians to consider additional diagnostic possibilities.

Patients, particularly the elderly, also may seek care to monitor multiple chronic conditions. This places an emphasis on collaborative and continuous care that involves patients (and their families) and providers, long-term care goals and plans, and self-management training and support.⁶⁹ The organizational needs for condition management, however, differ substantially from those necessary to deliver health services for acute episodic complaints. Taking a patient history in this case involves determining the status of the multiple conditions and whether symptoms from those conditions have progressed, improved, or stabilized and of the ability of patients to manage their condition.

The physical examination may be directed or complete. Physical findings are referred to as signs (distinct from symptoms noted by the patient). Directed physical examination refers to the examination of the relevant organ systems that may cause the symptoms or that may have positive or negative findings related to suspected diseases. When the disease is a chronic condition, the examination may be used to monitor disease progression or resolution. The complete physical examination of all organ systems may be performed as part of any annual examination, for difficult diagnoses, or for diseases that affect multiple organ systems.

The medical decisionmaking step of the encounter involves performing an assessment and plan. After the history and physical examination—based on the diagnostic possibilities, their likelihood, and the risks and benefits of treatment for each—the physician decides whether to recommend diagnostic testing, empiric treatment or referral to specialty or subspecialty care for further diagnostic evaluation, or a therapeutic intervention. Particularly challenging diagnoses are those that present with atypical symptoms, occur rarely, mimic other diseases, or involve multiple organ systems. For example, symptoms may arise from different organ systems: Wheezing, which is consistent with asthma, could be caused by acid going up from the stomach into the esophagus and then into the lungs (gastroesophageal reflux), congestive heart failure, or vocal cord dysfunction, among other diagnostic possibilities. The final step in the encounter is counseling the patient regarding diagnoses, tests, and treatments including dietary and lifestyle changes, medications, medical devices, and procedural interventions.

IV. Medical Decisionmaking

A. Diagnostic Reasoning

Uncertainty in defining a disease makes diagnosis difficult: (1) the difference between normal and abnormal is not always well demarcated; (2) many diseases

69. 2001 CQHCA Report, *supra* note 54, at 27.

do not progress with certainty (e.g., progression of ductal carcinoma in situ of the breast to invasive breast cancer occurs less than 50% of the time) but rather increase the risk of a poor outcome (e.g., hypertension raises the risk of developing heart disease or stroke); and (3) symptoms, signs, and findings for one disease overlap with others.⁷⁰ Variation also exists in the ability of physicians to elicit particular symptoms (e.g., in a group of patients interviewed by many physicians, 23% to 40% of the physicians reported cough as being present), observe signs (e.g., only 53% of physicians detected cyanosis—a blue or purple discoloration of the skin resulting from lack of oxygen—when present), or interpret tests (e.g., only 51% of pathologists agreed with each other when examining PAP smear slides with cells taken from a woman’s cervix to look for signs of cervical cancer).⁷¹ Moreover, prognosis (response to disease or treatment) with alternative therapies is in many cases uncertain. In a report by the Royal College of Physicians:

The practice of medicine is distinguished by the need for judgement in the face of uncertainty. Doctors take responsibility for these judgements and their consequences. A doctor’s up-to-date knowledge and skill provide the explicit scientific and often tacit experiential basis for such judgements. But because so much of medicine’s unpredictability calls for wisdom as well as technical ability, doctors are vulnerable to the charge that their decisions are neither transparent nor accountable.⁷²

1. Clinical reasoning process

Studies of clinical problem solving suggest that physicians employ combinations of two diagnostic approaches ranging from hypothetico-deductive (deliberative and analytical) to pattern recognition (quick and intuitive).⁷³ In the hypothetico-deductive approach, based on partial information, such as patient age, gender, and chief complaint, physicians⁷⁴ begin to generate a limited list of potential diagnostic hypotheses (hypothesis generation). Over the past 50 years, cognitive scientists

70. David M. Eddy, *Variations in Physician Practice: The Role of Uncertainty*, 3 *Health Affairs* 74, 75–76 (1984).

71. *Id.* at 77–78.

72. Royal College of Physicians, RCP Bookshop. *Doctors in Society. Medical Professionalism in a Changing World technical supplement full text* at 11, available at <http://bookshop.rcplondon.ac.uk/contents/pub75-411c044b-3eee-462d-936d-1dad7313e4a0.pdf> (last visited Feb. 12, 2011).

73. Jerome P. Kassirer et al., *Learning Clinical Reasoning* (2d ed. 2009) (hereinafter “Kassirer et al.”); Arthur S. Elstein & Alan Schwartz, *Clinical Problem Solving and Diagnostic Decision Making: Selective Review of the Cognitive Literature*, 324 *BMJ* 729–32 (2002) (hereinafter “Elstein”); Jerome P. Kassirer & G. Anthony Gorry, *Clinical Problem Solving: A Behavioral Analysis*, 89 *Annals Internal Med.* 245 (1978); Geoffrey Norman, *Research in Clinical Reasoning: Past History and Current Trends*, 39 *Med Educ.* 418–27 (2005).

74. Steven N. Goodman, *Toward Evidence-Based Medical Statistics, 1: The p Value Fallacy*, 130 *Annals Internal Med.* 995–1004 (1999) (hereinafter “Goodman”).

have demonstrated that human short-term memory capacity is limited,⁷⁵ and so this initial list of possible diagnoses is a cognitive necessity and provides an initial context that physicians use to evaluate subsequent data. Based on their knowledge of the diagnoses on that list, physicians have expectations about what symptoms, risk factors, disease course, signs, or test results would be consistent with each diagnosis (deductive inference).

As physicians gather additional information, they evaluate those data for their consistency with the possibilities on their initial list and whether those data would increase or decrease the likelihood of each possibility (hypothesis refinement). If the data are inconsistent, additional diagnostic possibilities are considered (hypothesis modification). The information gathering continues as an iterative process at the same visit or over time during multiple visits with the same or other physicians. The final cognitive step (diagnostic verification) involves testing the validity of the diagnosis for its coherency (consistency with predisposing risk factors, physiological mechanisms, and resulting manifestations), its adequacy (the ability to account for all normal and abnormal findings and the disease time course), and its parsimony (the simplest single explanation as opposed to requiring the simultaneous occurrence of two or more diseases to explain the findings).⁷⁶

At the other end of clinical reasoning are heuristics, quick automatic “rules of thumb” or cognitive shortcuts. In such cases, pattern recognition leads to rapid recognition and a quick diagnosis, improving cognitive efficiency.⁷⁷ For example, a black woman with large shadows of lymph nodes in her chest x ray would trigger a diagnosis of a disease known as sarcoidosis for many physicians. The simplifying assumptions involved in heuristics, however, are subject to cognitive biases. For example, episodic headache, sweating, and a rapid heartbeat form the classic triad seen in patients with a rare adrenal tumor known as a pheochromocytoma that also can cause hypertension. Physicians finding those three symptoms in a patient with hypertension may overestimate the patient’s likelihood of having pheochromocytoma based on representativeness bias, overestimating the likelihood of a less common disease just because case findings resemble those found in that disease.⁷⁸ Other cognitive errors include availability (overestimating the

75. Elstein, *supra* note 73; George A. Miller, *The Magical Number Seven Plus or Minus Two: Some Limits on Our Capacity for Processing Information*, 63 *Psychol. Rev.* 81–97 (1956).

76. Kassirer et al., *supra* note 73, at 5–6.

77. Stephen G. Pauker & John B. Wong, *How (Should) Physicians Think? A Journey from Behavioral Economics to the Bedside*, 304 *JAMA* 1233–35 (2010).

78. For additional discussion and definition of terms, see Section IV.A.2. Applying Bayes’ rule, about 100 in 100,000 patients with hypertension have pheochromocytoma; this symptom triad occurs in 91% of patients with pheochromocytoma (sensitivity) and does not occur in 94% of those without pheochromocytoma (specificity), and so 6% of those without pheochromocytoma would have this symptom triad. On the basis of Bayes’ rule, 91 of the 100 individuals with pheochromocytoma (91% times 100) would have this triad, and 5994 without a pheochromocytoma (6% times 99,900) will have the triad. Thus, among the 100,000 hypertensive patients, 6085 will have the classic triad, suggesting the possibility of pheochromocytoma, but only 91 out of the 6085 or 1.5%, will indeed have pheochromocytoma.

likelihood of memorable diseases because of severity or media attention and underestimating common or routine diseases) and anchoring (insufficient adjustment of the initial likelihood of disease).⁷⁹

Clinical intuition refers to rapid, unconscious processes that select the pertinent findings out of the multitude of available data.⁸⁰ Such expertise results from practice, is context sensitive, and cannot always be reduced to cause and effect.⁸¹ Cognitive research into the development of expertise suggests two competing hypotheses. In instance- or exemplar-based memory, physicians store scripts or “stories” of prior recalled case examples, for example, visual information such as that in pathology, dermatology, or radiology, and match new cases to those stories. The alternative prototype memory hypothesis is based on a mental model of disease wherein experts store structured “facts” about the disease to create abstractions. These “prototypes” enable experts to link findings to one another, to connect findings to the possible diagnoses, and to predict additional findings necessary to confirm the diagnosis, even in the absence of prior experience with exactly such a case.⁸²

Physicians typically apply hypothetico-deductive approaches when seeing patients with problems outside of their expertise or difficult problems with atypical issues within their expertise and apply intuitive pattern recognition for cases within their expertise or less challenging cases. However, diagnostic accuracy appears to depend more on mastery of domain knowledge than on the particular problem-solving method.⁸³

2. Probabilistic reasoning and Bayes’ rule

There is no correlation between physicians’ ability to collect data thoroughly and their ability to interpret the data accurately.⁸⁴ Making quantitative predictions or interpretation of test results constitutes probabilistic reasoning and avoids the use of ambiguous qualitative terms such as “low” or “always” that may contribute to different management decisions.⁸⁵

Over 200 years ago, the Reverend Bayes first wrote a paper published posthumously which now forms a critical concept in modern medicine. Ignored for

79. Kassirer et al., *supra* note 73; Elstein, *supra* note 73.

80. Trisha Greenhalgh, *Intuition and Evidence—Uneasy Bedfellows?* 52 *Brit. J. Gen. Practice* 395–400 (2002).

81. *Id.* at 396.

82. Kassirer et al., *supra* note 73; Elstein, *supra* note 73.

83. Elstein, *supra* note 73.

84. Arthur S. Elstein & Alan Schwartz, *Clinical Reasoning in Medicine*, in *Clinical Reasoning in the Health Professions* 223–34 (Joy Higgs et al. eds., 3d ed. 2008).

85. When physicians were asked to quantify “low probability,” the estimates had a mean of ~37% with a range from 0% to ~80% and when asked to quantify “always,” physicians had a mean of ~88% with a range from 70% to 100%. Geoffrey D. Bryant & Geoffrey R. Norman, *Expressions of Probability: Words and Numbers*, 302 *New Eng. J. Med.* 411 (1980).

nearly two centuries, his paper showed how to estimate the likelihood of disease following a test result using the likelihood of disease prior to testing and the specific test result obtained. Thus, Bayesian analysis refers to a method of combining existing evidence or a prior belief with additional evidence, for example, from test results. The additional evidence may be the presence or absence of a symptom, sign, test, or research study results.

The pretest suspicion of disease or, equivalently, the likelihood or prior probability of disease may be objective, that is, related to incidence (new cases over a specified period of time) or prevalence (existing cases at a particular point in time); based on clinical prediction rules (e.g., mathematical predictive models to estimate the likelihood of developing heart disease over the next 10 years using data from the Framingham Study); or subjective, that is, based on a clinician's estimated likelihood of disease prior to any testing.⁸⁶ Bayes' rule then combines that pretest suspicion with the observed test result. Those who have disease and a positive test are said to have true-positive test results. Those without disease who have a negative test are said to have true-negative test results. Tests, however, are almost always not perfectly accurate. That is, not everyone with disease has a positive test; these are called false-negative test results. Similarly, some individuals who are healthy may mistakenly have positive tests; these are called false-positive test results.

For example, consider screening mammography which is positive in 90% of women with breast cancer, and so the true-positive rate (or "sensitivity") of 90% is the likelihood of a positive test among those with disease. Mammography is negative in 93% of women without breast cancer, and so the true-negative rate (or "specificity") of 93% is the likelihood of a negative test among those who do not have disease (see Table 1).⁸⁷ Note that if the test is not negative, it must be positive, or vice versa, so that the sum of the columns in Table 1 must equal 100%.

Because a positive mammogram can occur among individuals with or without breast cancer, the interpretation of the likelihood of breast cancer with a positive mammogram can be problematic. Given that the prevalence of breast cancer among asymptomatic 40- to 50-year-old women is 8 in 1000, or 0.8%, Bayes' rule calculates the likelihood of breast cancer following a test result, for example, a positive mammogram (see Figures 1 and 2, Table 2).⁸⁸ This analysis helps explain in part why mammogram screening is controversial in women under age 50.

86. See *Gonzalez v. Metro. Transp. Auth.*, 174 F.3d 1016, 1023 (9th Cir. 1999) (describing the implications of Bayes' rule for drug testing and noting that a test with the same false-positive rate will generate a higher proportion of false positives to true positives in a population with fewer drug users); see generally Michael O. Finkelstein & William B. Fairley, *A Bayesian Approach to Identification Evidence*, 83 Harv. L. Rev. 489 (1970). For a discussion of Bayesian statistics, see David H. Kaye & David A. Freedman, Reference Guide on Statistics, Section IV.D, in this manual.

87. Gerd Gigerenzer, *Calculated Risks: How to Know When Numbers Deceive You* (2002) at 41 (hereinafter "Gigerenzer").

88. *Id.* at 45-48.

Table 1. 2 × 2 Test Characteristics of Screening Mammogram for Use in Bayes' Rule

	Breast Cancer	No Breast Cancer
Positive mammogram	90 true positives	7 false positives
Negative mammogram	10 false negatives	93 true negatives

Figure 1. Screening 1000 women for breast cancer.

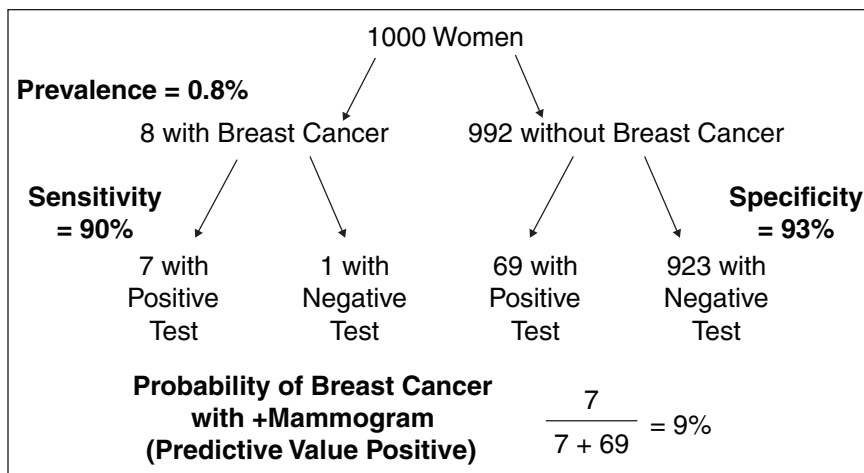


Figure 2. Likelihood of breast cancer after a positive or a negative mammogram.

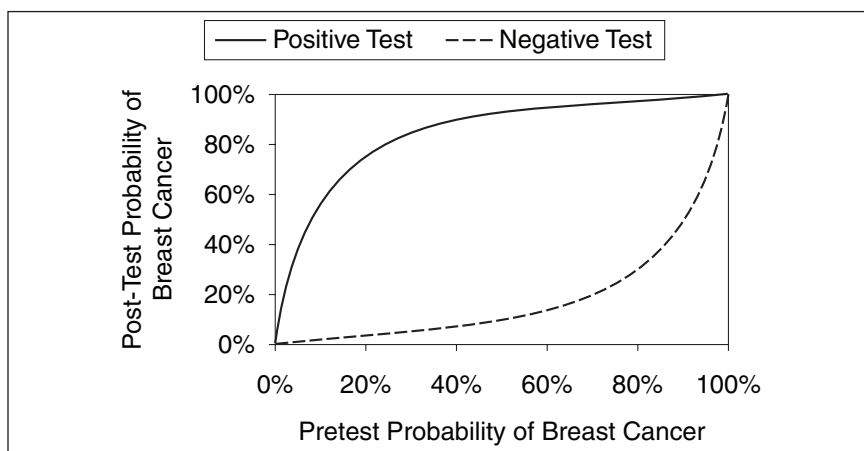


Table 2. Tabular and Formula Forms of Bayes’ Rule

Tabular Form of Bayes’ Rule

Condition	Pretest or Prior Probability (%)	Conditional Probability of Positive Test for the Condition (%)	Product of the Pretest and the Conditional Probabilities (%)	Posttest or Posterior Probability (%)
Breast cancer	0.8	90 sensitivity	0.72	9 = 0.72 ÷ 7.6
No breast cancer	99.2	7 1 – specificity	6.9	
			Sum = 7.6	

Formula Form of Bayes’ Rule

$$\frac{pD+ \star pT+ | D+}{(pD+ \star pT+ | D+) + ((1-pD+) \star (1-pT- | D-))}$$

pD+ = prior probability of disease = 0.8%
 pT+ | D+ = Sensitivity = True Positive Rate = 90%
 pT- | D- = Specificity = True Negative Rate = 93%

$$\frac{0.008 \star 0.90}{(0.008 \star 0.90) + ((1-0.008) \star (1-0.93))}$$

$$= 9\%$$

Despite a test that has a 90% or higher rate on both sensitivity and specificity, a calculation using Bayes’ theorem shows that having a low probability of breast cancer before testing means that even with a positive result on a screening mammogram, the likelihood that an average woman under age 50 has breast cancer is less than 10%.

The probability of breast cancer among those with a positive mammogram is termed the “predictive value positive.” Similarly, if the test were negative, the likelihood of breast cancer in those with a negative mammogram (“false reassurance rate”) would be 1 divided by 924 (1 woman with breast cancer and a negative test and 923 women without breast cancer who have negative tests in Figure 1), or about 0.1%. Interpreting a medical test result then depends on the pretest likelihood of disease and the test’s sensitivity and specificity. Figure 2

illustrates the likelihood of breast cancer for differing pretest or prior probabilities of breast cancer.

The discriminating ability of a test can be succinctly summarized as a likelihood ratio. The likelihood ratio positive expresses how much more likely disease is to be present following a positive test result. It is the ratio of the true-positive rate to the false-positive rate (sensitivity divided by 1 minus the specificity), e.g., 12.5 (0.90 divided by $1 - 0.93$) in the case of mammography. The likelihood ratio negative expresses how much less likely disease is to be present following a negative test result. It is 1 minus the ratio of the false-negative rate to the true-negative rate (1 minus the sensitivity divided by the specificity) or 0.11 ($1 - 0.90$ divided by 0.93) in the case of mammography. Likelihood ratios exceeding 10 or falling below 0.1 are believed to be strong discriminators causing “large” changes in the likelihood of disease; those between 5 and 10 or 0.1 and 0.2 cause “moderate” changes; and those between 2 and 5 or 0.2 and 0.5 cause “small” changes.⁸⁹ Note that even for a strongly discriminating test such as mammography, a positive or a negative test result does not change the likelihood of disease substantially for very low or very high probabilities of disease (see Figure 2), thereby highlighting the importance of the pretest likelihood of disease in interpreting test results.

Terms such as “sensitivity,” “specificity,” and “predictive value negative or positive” are called conditional probabilities because they express the likelihood of a particular result based on a particular condition (e.g., a positive test result among those with disease) or the likelihood of a particular condition among those with a particular result (e.g., disease among those with a positive test).⁹⁰ These kinds of expression, however, remove the base case probability (the pretest probability of disease, sometimes referred to as the prior probability of disease) as part of “normalization,” so that Bayes’ rule is required to interpret a test result. Moreover, confusion between sensitivity and predictive value positive may lead to errors in the interpretation of test results; for example, a 90% likelihood of having a positive mammogram in patients with breast cancer—the sensitivity—may be misinterpreted as the predictive value positive, implying that a woman with a positive mammogram has a 90% chance of having cancer. This misinterpretation ignores the role for pretest suspicion or likelihood of disease (or assumes that all

89. David A Grimes & Kenneth F Schulz, *Refining Clinical Diagnosis with Likelihood Ratios*, 365 *Lancet* 1500–05 (2005).

90. This terminology may be confusing. The predictive value negative (negative predictive value) is defined as the probability of no disease among those with a negative test. It also equals 1 minus the false reassurance rate. The false-alarm rate is defined as the probability of no disease among those with a positive test. It is also 1 minus the predictive value positive. The false reassurance rate may be confused with the false negative rate (among those with disease, the likelihood of a negative test) because both involve those with negative tests and those with disease but in one case the denominator is individuals with negative tests (false reassurance rate) and in the other case individuals with disease (false negative rate). Similarly, the false alarm rate may be confused with the false positive rate (among those with no disease, the likelihood of a positive test).

women undergoing the test have the disease). This confusion can be avoided by translating Bayes' rule into natural frequency expressions.⁹¹ The natural frequency expression incorporates both the pretest likelihood and the conditional probabilities of the test results to yield the following statements (see Figure 1): Of 1000 women between 40 and 50 years old, 8 have breast cancer, and 7 of these will test positive. Of the remaining 992 who do not have breast cancer, about 69 will also test positive. When presented as a natural frequency (including the likelihood of disease), the likelihood of breast cancer becomes more transparent; thus 76 women will test positive, and 7 of the 76 will have breast cancer. When 48 physicians with an average of 14 years of professional experience were presented with the natural frequency version or the conditional probability version, 16 of 24 estimated the likelihood of breast cancer to exceed 50% with the conditional probability (sensitivity, specificity) version but only 5 of 24 did so with the natural frequency information.⁹²

Just as mammography test results may be misinterpreted if Bayes' rule is not applied, the prosecutor's fallacy involves the misinterpretation of probabilistic information. For example, in *People v. Collins*, the prosecutor argued that 1 in 3 girls have blonde hair, 1 in 10 girls have a pony tail, 1 in 10 automobiles are partly yellow, 1 in 4 men have a mustache, 1 in 10 black men have a beard, and 1 in 1000 cars have an interracial couple in the car.⁹³ Multiplying these six probabilities together yields a 1 in 12 million joint probability of having all conditions present. Aside from being simply estimates and from assuming that the probabilities were independent of one another, the prosecutor made the statement that "The probability of the defendant matching on these six characteristics is 1 in 12 million," thereby assuming that someone other than the defendant being guilty is the same 1 in 12 million. However, if translated into natural frequency terms, 1 out of every 12 million couples would have these six characteristics, and so assuming that there are 24 million couples, there would be a 1 in 2 chance that the Collinses are innocent. The error results from confusing the probability of a positive test (having all six characteristics) among those with the disease (being guilty) and the probability of the disease (being guilty) among those with a positive test (having all six characteristics), that is, confusing the conditional probabilities—sensitivity and positive predictive value.

Bayes' rule becomes even more relevant in the genomic medicine era.⁹⁴ Suppose a genetic test has a sensitivity and specificity of 99.9%, and suppose the probability of disease is 1 in 1000 if a positive family history is present and 1 in 100,000 if no family history is present. Screening 1000 individuals with a positive family

91. Gigerenzer, *supra* note 87, at 42.

92. *Id.* at 43.

93. *Id.* at 152.

94. Isaac S. Kohane et al., *The Incidentalome. A Threat to Genomic Medicine*, 296 JAMA 212–15 (2006).

history for the gene results in 2 positive tests: 1 individual truly has disease, and in the other the test is a false positive. Screening 10 million individuals without a family history results in 10,100 positive tests in which 100 individuals have disease and 10,000 do not. Even with a specificity of 99.99%, if a test screens for 10,000 genes simultaneously, then 63% of individuals will have at least one false-positive test result. Based simply on the genetic test results alone, neither individuals nor physicians would be able to distinguish those with true-positive results from those with false-positive results, thereby potentially leading to inappropriate monitoring or treatment for all with positive test results.

Although a test is commonly thought of as a sample from a bodily fluid, tissue, or image, a test also could be the presence or absence of a symptom or physical sign. For example, both inhalation anthrax and influenza can cause symptoms of muscle aches, fever, and malaise. However, a critical symptom that helps distinguish one from the other is runny nose, which occurs in 14% of those with inhalation anthrax but in 78% to 89% of those with influenza or influenza-like illness. Thus, when faced with distinguishing between these diagnoses, patients with a runny nose given this symptom alone are about six times more likely to have influenza or a flu-like illness than to have anthrax.⁹⁵

Sensitivity and specificity rely on setting a positivity criterion, the threshold level for determining normal above which tests are positive and below which the test is negative. If the criterion is made stricter (e.g., what is considered to be abnormal requires a higher test result), then sensitivity falls and specificity increases, and if the criterion is made laxer, then sensitivity rises and specificity falls. Depending on the context of the testing, it may be more appropriate to choose a laxer criterion (e.g., screening donated blood for HIV infection where the benefit is reducing transfusion-associated HIV transmission, and the risk is discarding some uninfected units of donated blood) or a stricter one (e.g., screening a low-prevalence population for HIV infection where the benefit is reducing false-positive diagnoses and the risk is missing some truly HIV-infected individuals).⁹⁶ Thus the benefits of finding and treating a person with disease versus the risk of treating a person without disease should help establish what is considered normal or abnormal.

The terms “sensitivity” and “specificity” apply to the simple situation in which disease is present or absent and a test can be positive or negative, but terminology and interpretation become more complicated when multiple diseases are under consideration and when multiple test results may occur.⁹⁷ For example, consider blood in the urine (hematuria), which could be caused by a urinary tract infection, a kidney stone, or a bladder cancer, among many other diseases. The

95. Nathaniel Hupert et al., *Accuracy of Screening for Inhalational Anthrax After a Bioterrorist Attack*, 139 *Annals Internal Med.* 337–45 (2003).

96. Klemens M. Meyer & Stephen G. Pauker, *Screening for HIV: Can We Afford the False Positive Rate?* 317 *New Eng. J. Med.* 238–41 (1987).

97. Kassirer et al., *supra* note 73, at 21–22.

terms “sensitivity” and “specificity” are no longer appropriate because disease is not simply present or absent. Instead, they are replaced by the term conditional probabilities, that is, sensitivity is replaced by the likelihood of blood in the urine with a urinary tract infection, or with a kidney stone, or with a bladder cancer. Similarly, a very positive test has a different interpretation than a weakly positive test, and Bayes’ rule can quantify the difference. Results from multiple tests can be combined with Bayes’ rule by applying Bayes’ rule to the first test result and then reapplying Bayes’ rule to subsequent test results. This approach assumes that the result of the first test does not affect the test characteristics (sensitivity or specificity) of the second test (i.e., that there is conditional independence of each test). When two tests are available, screening will usually occur first with the high-sensitivity test to detect a high proportion of those with disease (true positives), or “ruling in” disease. Those with a positive first test will then undergo a high-specificity test to reduce the number of individuals who do not have disease but a positive first test (false positive), or “ruling out” disease.

3. Causal reasoning

To select the most appropriate therapy, physicians seek to identify the cause of a patient’s complaints and findings. While considering the presence or absence of risk factors (e.g., the presence of male gender, advanced age, high cholesterol, high blood pressure, diabetes mellitus, and smoking for the medical condition coronary heart disease), physicians will often use any type of evidence⁹⁸ that might support causation, for example, biological plausibility,⁹⁹ physiological drug effects, case reports, or temporal proximity¹⁰⁰ to an exposure.¹⁰¹ Although physicians use epidemiological studies in their decisionmaking, “they are accustomed to using *any* reliable data to assess causality, no matter what their source” because they must make care decisions even in the face of uncertainty.¹⁰² This is in contrast to the courts which require a higher standard than clinicians or regulators, and wherein causation cannot just be “possible” but where “a ‘preponderance of evidence’ establishes that an injury was caused by an alleged exposure.”¹⁰³ For physicians, causal reasoning typically involves

98. Jerome P. Kassirer & Joe S. Cecil, *Inconsistency in Evidentiary Standards for Medical Testimony: Disorder in the Courts*, 288 JAMA 1382–87 (2002) (hereinafter “Kassirer & Cecil”); see also Section IV.C.2, for levels of evidence.

99. See *Kennan v. Sec’y of Health & Human Servs.*, 2007 WL 1231592 (Ct. Fed. Cl. Apr. 5, 2007).

100. *But see Wilson v. Taser Int’l, Inc.*, 303 F. App’x 708, 714 (11th Cir. 2008) (“[A]lthough a doctor usually may primarily base his opinion as to the cause of a plaintiff’s injuries on this history where the patient ‘has sustained a common injury in a way that it commonly occurs,’ . . . Dr. Meier could not rely upon the temporal connection between the two events to support his causation opinion in this case.”).

101. Kassirer & Cecil, *supra* note 98, at 1384.

102. *Id.* at 1394.

103. *Id.* at 1384.

understanding how abnormalities in physiology, anatomy, genetics, or biochemistry lead to the clinical manifestations of disease. Through such reasoning, physicians develop a “causal cascade” or “chain or web of causation” linking a sequence of plausible cause-and-effect mechanisms to arrive at the pathogenesis or pathophysiology of a disease. For example, kidney failure leads to poor drug excretion, resulting in symptoms or signs of drug toxicity.¹⁰⁴ Although probabilistic reasoning typically dominates initial hypothesis generation by physicians based on prevalence or incidence, pattern recognition of concomitant symptoms and signs could trigger a diagnosis. For example, cough, lung lesions, and enlarged breasts (gynecomastia) in a 37-year-old man could trigger the diagnosis of metastatic germ cell cancer.¹⁰⁵ More typically, physicians use causal reasoning in diagnostic refinement and verification to examine a diagnosis for its coherency, namely, asking whether its physiological mechanism would be expected to lead to the observed manifestations and whether it is adequate to account for all normal and abnormal findings and the disease time course. Once treatment has been implemented, physicians must make causal judgments in determining whether an alteration in patient status is the result of progression of disease or an adverse consequence of treatment, or whether the absence of improvement results from therapeutic ineffectiveness that should prompt a change in therapy or even reconsideration of the diagnosis.

Pathophysiological reasoning, however, also can lead to incorrect conclusions. In patients with heart failure with a weakened heart, a class of medications called beta blockers had been thought to be contraindicated because beta blockers would decrease the strength of the heart muscle contraction. Subsequent studies found that beta blockers in patients with heart failure usually had no ill effect and actually increased survival. Similarly, physicians once thought that atherosclerotic blockages in heart arteries slowly progressed to cause a heart attack, so that revascularizing those plaques through heart bypass surgery would prevent heart attacks.¹⁰⁶ Over the past 15 years, however, scientific evidence has emerged that small vulnerable atherosclerotic plaques (not amenable to revascularization because of their small size) can suddenly rupture and cause heart attacks. Not surprisingly, revascularization trials involving either bypass surgery or percutaneous interventions such as stenting or angioplasty do not diminish the risk of having a heart attack or improve survival for most patients.¹⁰⁷

Although treating physicians¹⁰⁸ may testify with regard to both general and specific causation, as with use of evidence for causation, their standards for evi-

104. Kassirer et al., *supra* note 73, at 63–66.

105. *Id.* at 29.

106. David S. Jones, *Visions of a Cure: Visualization, Clinical Trials, and Controversies in Cardiac Therapeutics, 1968–1998*, 91 *Isis* 504–41 (2000).

107. Thomas A. Trikalinos et al., *Percutaneous Coronary Interventions for Non-acute Coronary Artery Disease: A Quantitative 20-Year Synopsis and a Network Meta-analysis*, 373 *Lancet* 911–18 (2009).

108. See generally *Bland v. Verizon Wireless, LLC*, 538 F.3d 893 (8th Cir. 2008) (upholding the district court’s decision to reject a treating physician’s evidence of causation under *Daubert*).

dence vary.¹⁰⁹ For example, some physicians may stop using a drug after the first reports of adverse effects, and others may continue to use a drug despite evidence of harm from randomized controlled trials. Determining whether an effect is a class effect or drug specific can be difficult. When considering beta blockers for patients with a weakened heart (heart failure), many studies have consistently demonstrated the benefit of beta blockers in reducing mortality in those with heart attacks often resulting in weakened heart function. However, in a randomized trial limited to patients with documented weakened heart, one particular beta blocker was found to not confer a survival benefit, and as a result the heart failure guidelines limited their beta blocker recommendation to just those three drugs with documented mortality benefit in trials.¹¹⁰

Although treating physicians may be aware of patient-specific risk factors such as smoking or family history, they may not routinely review specialized aspects of such data, for example, toxicology, industrial hygiene, environment, and some aspects of epidemiology. Additional experts may assist in distinguishing general from specific causation by using their specialized knowledge to weigh the relative contribution of each putative causative factor to determine “reasonable medical certainty” or “reasonable medical probability.” The determination of general causation involves medical and scientific literature review and the evaluation of epidemiological data, toxicological data, and dose–response relationships. Consider for example, hormone replacement therapy for postmenopausal women. Multiple observational studies using methods such as case–control, cross-sectional, and cohort designs¹¹¹ suggested an association between hormone therapy and reduction in heart attack, but such designs are subject to confounding and bias and are particularly weak for causation because in case–control and cross-sectional studies, the sequence of the exposure and outcome is unknown. To resolve the question, the Women’s Health Initiative (WHI) study randomized women to hormone replacement therapy or placebo and found a statistically significant increase in clot-related disorders—heart attack, stroke, and heart-related mortality over 5 years but most notable in the first year after initiation of hormone therapy.¹¹² Heart attacks are caused by blood clots and plaque rupture, and so the results were consistent with the known biological mechanism of estrogens in the clotting cascade. However, patients in the WHI were, on average, 63 years old and therefore not perimenopausal as analyzed in the observational studies. In a novel

109. Kassirer & Cecil, *supra* note 98, at 1384.

110. Mariell Jessup et al., *2009 Focused Update: ACCF/AHA Guidelines for the Diagnosis and Management of Heart Failure in Adults: A Report of the American College of Cardiology Foundation/American Heart Association Task Force on Practice Guidelines*, 119 *Circulation* 1977–2016 (2009).

111. See Michael D. Green et al., Reference Guide on Epidemiology, in this manual.

112. Jacques E. Rossouw et al., *Risks and Benefits of Estrogen Plus Progestin in Healthy Postmenopausal Women: Principal Results from the Women’s Health Initiative Randomized Controlled Trial*, 288 *JAMA* 321–33 (2002); JoAnn E. Manson et al., *Estrogen Plus Progestin and the Risk of Coronary Heart Disease*, 349 *New Eng. J. Med.* 523–34 (2003).

approach, the observational Nurses' Health Study attempted to emulate the design and intention-to-treat (ITT) analysis aspect of the WHI randomized trial, and saw that the hormone replacement treatment effects were similar to those from the randomized trial, suggesting that "the discrepancies between the WHI and the Nurses' Health Study ITT estimates could be largely explained by differences in the distribution of time since menopause and length of followup."¹¹³

B. Testing

1. Screening

Screening on a population basis requires that (1) the condition be present in the population and affect quality and length of life; (2) the incidence or prevalence be sufficiently high to justify any risks associated with the test; (3) preventive or early treatment should be available; (4) an asymptomatic period for early detection must exist; (5) the screening test should be accurate, acceptable, and affordable; and (6) screening benefits should exceed harms. Screening for disease in asymptomatic, otherwise healthy patients has become widely accepted and promulgated.¹¹⁴ Screening differs from diagnostic testing used to elucidate the cause of symptoms or loss of function because screening involves apparently healthy individuals.¹¹⁵ Although screening may prevent the development of disease-related morbidity and mortality, positive test results (both false positive and true positive) may lead to interventions that could be unnecessary or even risky because of overdiagnosis and overtreatment.¹¹⁶

Normal ranges for biochemical tests are often based on the 95% confidence intervals in a normal healthy population—that is, although everyone is healthy, by convention, values outside the 2.5% lower and upper extremes are considered to be abnormal. Consequently, ordering six blood tests in a normal healthy individual yields only a 74% chance that all six tests will be normal; that is, there is a 26% chance that one or more may be abnormal. Similarly, when ordering 12 tests in a normal person, there is a 54% chance that all 12 will be normal and a 46% chance that 1 or more will be abnormal. So simply ordering tests in healthy individuals or in the absence of clinical suspicion of a disease may result in many

113. Miguel A. Hernán et al., *Observational Studies Analyzed Like Randomized Experiments: An Application to Postmenopausal Hormone Therapy and Coronary Heart Disease*, 19 *Epidemiology* 766–79 (2008).

114. Lisa M. Schwartz et al., *Enthusiasm for Cancer Screening in the United States*, 291 *JAMA* 71–78 (2004).

115. David A. Grimes & Kenneth F. Schulz, *Uses and Abuses of Screening Tests*, 359 *Lancet* 881–84 (2002) (hereinafter Grimes and Schulz); William C. Black, *Overdiagnosis: An Under Recognized Cause of Confusion and Harm in Cancer Screening*, 92 *J. Nat'l Cancer Inst.* 1280–82 (2000) (hereinafter "Black").

116. Grimes & Schulz, *supra* note 115, at 884; Black, *supra* note 115, at 1280.

false-positive test results that can lead to false alarms, anxiety, additional testing, and possible morbidity or mortality from subsequent testing or interventions.¹¹⁷

Even a valueless screening test may appear to be beneficial because of “lead-time bias.” If screened or unscreened patients have the same prognosis from the time of onset of symptoms to death, then screened patients only appear to live longer because the time elapsed from diagnosis by screening to death exceeds that from diagnosis made at the time of symptom onset to death. A second bias, “length bias,” also leads to overestimation of the benefit from screening.¹¹⁸ Suppose that a randomized trial of screening or no screening is conducted over a limited length of time from study initiation to termination. The screening test detects patients with both aggressive and indolent forms of the disease. Among the unscreened patients, however, disease only becomes evident through the development of symptoms, which would be more likely in patients who have the aggressive form of the disease and a poorer prognosis. Thus screened patients with disease appear to have a better prognosis than unscreened patients with disease because a higher proportion of the screened patients have more indolent disease. Extending the concept of length bias further, screening can result in “pseudodisease” or “overdiagnosis,” such as the identification of slow-growing cancers that even if untreated would never cause symptoms or reduce survival.¹¹⁹ Although lung cancer is commonly thought to be one of the more aggressive cancers, an autopsy study found that one-third of lung cancers were unsuspected prior to autopsy, and nearly all of these patients with unsuspected lung cancer prior to autopsy died from other causes.¹²⁰ Lung cancer screening in these individuals would have resulted in pseudodisease or overdiagnosis because screening would have diagnosed their cancer but they would have died of something else (or from a severe adverse effect of the cancer treatment) before the cancer became evident.

To further illustrate bias in screening studies, the Mayo Lung Project was a randomized trial comparing screening for lung cancer with periodic chest X rays and sputum samples versus usual care. It found that screening did improve the likelihood of survival 5 years after diagnosis in those with lung cancer but surprisingly did not affect lung cancer deaths. Further analysis of the randomized trial found that the survival advantage of screening was attributable to the 46 extra

117. A radiologist described his own experience to illustrate the clinical aphorism that “the only ‘normal’ patient is one who has not yet undergone a complete work-up.” He had a negative CT scan of the colon examination, but the CT scan also provided images outside the liver with radiologists identifying lesions in the kidneys, liver, and lungs. This resulted in additional CT scans, a liver biopsy, PET scan, video-aided thoracoscopy (a flexible scope inserted into the chest), and three wedge resections of the lung leading to multiple tubes, medications, and “excruciating pain” that required 5 weeks for recovery. William J. Casarella, *A Patient’s Viewpoint on a Current Controversy*, 224 *Radiology* 927 (2002).

118. Grimes & Schulz, *supra* note 115, at 884.

119. Black, *supra* note 115, at 1280.

120. Charles K. Chan et al., *More Lung Cancer but Better Survival: Implications of Secular Trends in “Necropsy Surprise” Rates*, 96 *Chest* 291–96 (1989).

lung cancer cases detected by screening. These 46 cases had indolent (or, at worst, very slowly progressive) lung cancer; that is, these patients would have a normal life expectancy, and so, including their prognosis in those with screen-detected lung cancer inflates the apparent 5-year survival with screening because of length bias and overdiagnosis.¹²¹ More recently, CT scan screening found lung cancer to be present in the same proportion of nonsmokers as smokers,¹²² suggesting that many of the cancers detected in the nonsmokers were ones that would have never progressed. This overdiagnosis can lead to morbidity and mortality: CT scan screening for lung cancer results in a threefold increase in diagnosis and threefold increase in surgery with an average surgical mortality of 5% and serious complication rate exceeding 20%,¹²³ as well as potential risk from radiation exposure. A similar phenomenon occurs with breast cancer where screening increases surgeries by about one-third from overdiagnosis and with prostate cancer where the lifetime risk of dying from prostate cancer is about 3%, yet 60% of men in their sixties have prostate cancer, and so, screening and detecting all men with prostate cancer in their sixties would lead to treatment of many men who would not have died from prostate cancer.¹²⁴ In patients found to have cancer by screening, it is not possible to distinguish those whose cancers would have progressed from those in whom the cancer-appearing cells would not have progressed or spread.

2. Diagnostic testing

Based on the history and physical examination, physicians will establish diagnostic possibilities. They may then request additional tests to reduce uncertainty and to confirm the diagnosis, as part of diagnostic verification. Although, theoretically, all tests could be ordered, tests should be chosen on the basis of a clinical suspicion because of possible morbidity or even mortality from inappropriate testing. Normative prescriptive decision models for reasoning in the presence of uncertainty suggest that whether and which tests get ordered should depend on the sensitivity and specificity of the test as discussed in Section IV.A.2, *supra*, but also the risk of mortality or morbidity from the test, and the benefit and risk of treatment.¹²⁵ In general, for sufficiently low probabilities of disease, no tests should be ordered and no treatment given. For sufficiently high probabilities of disease,

121. Black, *supra* note 115.

122. William C. Black & John A. Baron, *CT Screening for Lung Cancer: Spiraling into Confusion?* 297 *JAMA* 995–97 (2007).

123. *Id.* at 996.

124. Karsten J. Jørgensen & Peter C. Gøtzsche, *Overdiagnosis in Publicly Organised Mammography Screening Programmes: Systematic Review of Incidence Trends*, 339 *BMJ* b2587 (2009); Michael J. Barry, *Prostate-Specific-Antigen Testing for Early Diagnosis of Prostate Cancer*, 344 *New Eng. J. Med.* 1373–77 (2001).

125. Stephen G. Pauker & Jerome P. Kassirer, *The Threshold Approach to Clinical Decision Making*, 302 *New Eng. J. Med.* 1109–17 (1980).

testing is unnecessary and treatment should be administered. For intermediate probabilities of disease, testing should be performed. When testing carries risks, the probabilities of disease for which testing should be done become narrower, and so physicians should be more likely to treat empirically or neither test nor treat. As sensitivity and specificity increase, the range of probabilities in which testing should be done expands.

Although an abnormal test result may be found, that abnormality may not be causing symptoms. For example, herniated lumbar discs are found in approximately 25% of healthy individuals without back pain; thus finding a herniated disc in patients with back pain may be an incidental finding. If signs such as a foot drop develop, additional muscle and nerve conduction studies might confirm evidence of nerve compromise from the herniated disc, but such tests are painful. Over time, sequential images show that the herniated disc has partial or complete resolution after 6 months without surgery. Therefore, a herniated disc may be seen with CT or MRI scanning in patients with or without symptoms, and so just having symptoms and evidence of a herniated disc would be an insufficient indication for back surgery.¹²⁶ In the absence of severe or progressive neurological deficits, elective disc surgery could be considered for patients with probable herniated discs who have persistent symptoms and findings consistent with sciatica (not just low back pain) for 4 to 6 weeks, but such “patients should be involved in decision making” (see Section IV.D.3, *infra*).¹²⁷

Just as some therapies may eventually be found to be harmful or not beneficial, tests initially felt to be useful may be found to be less valuable.¹²⁸ Among other potential biases,¹²⁹ this may occur because of the choice of study population used to determine the test’s sensitivity and specificity. For example, an FDA-approved rapid test for HIV infection has a reported specificity of 100%, implying that any positive tests must indicate truly infected individuals, yet one of the populations in which testing is recommended is women who have had prior children and are in labor but have not yet had an HIV test during the pregnancy.¹³⁰ In 15 multiparous women, this rapid HIV test resulted in one false-positive test result in the 15 women tested, yielding a specificity of 93%,¹³¹ and so not all pregnant women with positive tests can be assumed to be truly infected.

126. Richard A. Deyo & James N. Weinstein, *Low Back Pain*, 344 *New Eng. J. Med.* 363–70 (2001); Richard A. Deyo et al., *Trends, Major Medical Complications, and Charges Associated with Surgery for Lumbar Spinal Stenosis in Older Adults*, 303 *JAMA* 1259–65 (2010).

127. Deyo & Weinstein, *supra* note 126, at 368.

128. David F. Ransohoff & Alvan R. Feinstein, *Problems of Spectrum and Bias in Evaluating the Efficacy of Diagnostic Tests*, 299 *New Eng. J. Med.* 926–30 (1978).

129. Penny Whiting et al., *Sources of Variation and Bias in Studies of Diagnostic Accuracy: A Systematic Review*, 140 *Annals Internal Med.* 189–202 (2004).

130. Food and Drug Administration, OraQuick® Rapid HIV-1 Antibody Test, available at <http://www.fda.gov/downloads/BiologicsBloodVaccines/BloodBloodProducts/ApprovedProducts/PremarketApprovalsPMAs/ucm092001.pdf> (last visited Mar. 2, 2011).

131. *Id.*

3. Prognostic testing

Once a diagnosis has been established, additional prognostic testing may be performed to establish the extent of disease (e.g., staging of a cancer) or to monitor response to therapy. Molecular profiling of disease may not only characterize prognosis but also treatment response. In women with breast cancer, for example, finding a genetic marker called the human epidermal growth factor receptor type 2 (HER2, also called HER2/neu) gene identified patients who responded poorly to any of the standard chemotherapeutic agents and hence had a poor prognosis. Illustrative of the emerging era of pharmacogenomics, adjuvant chemotherapy combined with a monoclonal antibody in HER2-positive breast cancer patients has been found to delay progression and prolong survival.¹³²

C. Judgment and Uncertainty in Medicine

1. Variation in medical care

Studies over the past several decades show substantial geographic variation in the utilization rates for medical care within small areas or local regions (e.g., a three- to fourfold variation in the use of surgical procedures such as tonsillectomy when comparing children living in adjacent areas of similar demographics)¹³³ and between large areas or widespread regions (e.g., a 10-fold variation in the performance of other discretionary surgical procedures such as lower extremity revascularization, carotid endarterectomy, back surgery, and radical prostatectomy).¹³⁴ Even when limiting the analysis to 77 U.S. hospitals with reputations for high-quality care in managing chronic illness, the care that patients received in their last 6 months of life varied extensively, ranging from hospital stays of 9 to 27 days (threefold variation), intensive care unit stays of 2 to 10 days (fivefold variation); and physician visits of 18 to 76 (fourfold variation), depending on the hospital at which patients received their care.¹³⁵

Four categories of variation are recognized: (1) underuse of effective care, (2) issues of patient safety, (3) concern for preference-sensitive care, and (4) notions of supply-sensitive services.¹³⁶ Effective care refers to treatments that are known to be beneficial and that nearly all patients should receive with little influence

132. Dennis J. Slamon et al., *Use of Chemotherapy Plus a Monoclonal Antibody Against HER2 for Metastatic Breast Cancer That Overexpresses HER2*, 344 *New Eng. J. Med.* 783–92 (2001).

133. John Wennberg & Alan Gittelsohn, *Small Area Variations in Health Care Delivery*, 182 *Science* 1102–08 (1973) (hereinafter “Wennberg & Gittelsohn”).

134. John D. Birkmeyer et al., *Variation Profiles of Common Surgical Procedures*, 124 *Surgery* 917–23 (1998).

135. John E. Wennberg et al., *Use of Hospitals, Physician Visits, and Hospice Care During Last Six Months of Life Among Cohorts Loyal to Highly Respected Hospitals in the United States*, 328 *BMJ* 607 (2004).

136. John E. Wennberg, *Unwarranted Variations in Healthcare Delivery: Implications for Academic Medical Centres*, 325 *BMJ* 961–64 (2002) (hereinafter “Wennberg”).

of patient preferences, for example, use of beta blockers following myocardial infarction. The *underuse of effective care* was illustrated by one prominent study that identified 439 high-quality process measures for 30 conditions and preventive care. In assessing the use of measures that were clearly recommended (i.e., clearly beneficial), they found that only about 50% of patients received these highly recommended care processes.¹³⁷ Issues of *patient safety* refer to the execution of care and the occurrence of iatrogenic complications (i.e., complications resulting from health care interventions). The IOM estimates that hospitalized patients risk one medication error for every day they are hospitalized, resulting in an estimated 7000 deaths annually (more than from workplace injuries) at an annual cost of \$3.5 billion in 2006 dollars.¹³⁸ Concern for *preference-sensitive care* refers to treatment choices that should depend on patient health goals or preferences. Prostate surgery helps relieve symptoms of an enlarged prostate (such as frequent urination, waking up at night to urinate) but carries a risk of losing sexual function. Separate from the probability of losing sexual function, in preference-sensitive care, the decision to have prostate surgery depends on how much the enlarged prostate symptoms bother the patient and on how important sexual function is to them, that is, their preferences and values.¹³⁹ Finally, *supply-sensitive services* refer to care that depends not on evidence of effectiveness or patient preferences, but rather on the availability of services. Specifically, patients living in areas with more doctors or more hospitals experience more office visits, tests, and hospitalizations.¹⁴⁰

2. Evidence-based medicine

The exceptional variation in the delivery of medical care was a major factor that led to a careful reexamination of physician diagnostic strategies, therapeutic decision making, and the use of medical evidence, but it was not the only one. Other circumstances that set the stage for an intense focus on medical evidence included (1) the development of medical research, including randomized controlled trials and other observational study designs; (2) the growth of diagnostic and therapeutic interventions;¹⁴¹ (3) interest in understanding medical decisionmaking and how physicians reason;¹⁴² and (4) the acceptance of meta-analysis as a method to com-

137. Elizabeth A. McGlynn et al., *The Quality of Health Care Delivered to Adults in the United States*, 348 *New Eng. J. Med.* 2635–45 (2003).

138. Committee on Identifying and Preventing Medication Errors, Institute of Medicine, *Preventing Medication Errors* (2006); 2000 CQHCA Report, *supra* note 58.

139. Michael J. Barry et al., *Patient Reactions to a Program Designed to Facilitate Patient Participation in Treatment Decisions for Benign Prostatic Hyperplasia*, 1995 *Med. Care* 771–82 (1995).

140. Wennberg, *supra* note 136, at 142.

141. Cynthia D. Mulrow & K.N. Lohr, *Proof and Policy from Medical Research Evidence*, 26 *J. Health Pol., Pol'y & L.* 249–66 (2001) (hereinafter “Mulrow & Lohr”).

142. Robert S. Ledley & Lee B. Lusted, *Reasoning Foundations of Medical Diagnosis; Symbolic Logic, Probability, and Value Theory Aid Our Understanding of How Physicians Reason*, 130 *Science* 9–21 (1959).

bine data from multiple randomized trials.¹⁴³ In response to the above conditions, “evidence-based medicine” gained prominence in 1992.¹⁴⁴ It is aptly defined as “the conscientious, explicit and judicious use of current best evidence in making decisions about the care of the individual patient. It means integrating individual clinical expertise with the best available external clinical evidence from systematic research.”¹⁴⁵

Evidence-based medicine contrasts with the traditional informal method of practicing based on anecdotes, applying the most recently read articles, doing what a group of eminent experts recommend, or minimizing costs.¹⁴⁶ Rather, it is “the use of mathematical estimates of the risks of benefit and harm, derived from high-quality research on population samples, to inform clinical decision making in the diagnosis, investigation or management of individual patients.”¹⁴⁷ In a paper from a joint workshop held by IOM and the Agency for Healthcare Research and Quality¹⁴⁸ that addressed what physicians consider to be sufficient evidence to justify their clinical practice and treatment decisions, Mulrow and Lohr wrote “evidence-based medicine stresses a structured critical examination of medical research literature: relatively speaking, it deemphasizes average practice as an adequate standard and personal heuristics.”¹⁴⁹

3. Hierarchy of medical evidence

With the explosion of available medical evidence, increased emphasis has been placed on assembling, evaluating, and interpreting medical research evidence. A fundamental principle of evidence-based medicine (*see also* Section IV.C.5, *infra*) is that the strength of medical evidence supporting a therapy or strategy is hierarchical. When ordered from strongest to weakest, systematic review of randomized trials (meta-analysis) is at the top, followed by single randomized trials, systematic reviews of observational studies, single observational studies,

143. See Michael D. Green et al., Reference Guide on Epidemiology, Section VI, in this manual; *Video Software Dealers Ass'n v. Schwarzenegger*, 556 F.3d 950, 963 (9th Cir. 2009) (analyzing a meta-analysis of studies on video games and adolescent behavior); *Kennecott Greens Creek Min. Co. v. Mine Safety & Health Admin.*, 476 F.3d 946, 953 (D.C. Cir. 2007) (reviewing the Mine Safety and Health Administration's reliance on epidemiological studies and two meta-analyses).

144. Evidence-Based Medicine Working Group, *Evidence-Based Medicine. A New Approach to Teaching the Practice of Medicine*, 268 JAMA 2420–25 (1992).

145. David L. Sackett et al., *Evidence Based Medicine: What It Is and What It Isn't*, 312 BMJ 71–72, 71 (1996).

146. Trisha Greenhalgh, *How to Read a Paper: The Basics of Evidence-Based Medicine* (3d ed. 2006).

147. *Id.* at 1.

148. Clark C. Havighurst et al., *Evidence: Its Meanings in Health Care and in Law*, 26 J. Health Pol., Pol'y & L. 195–215 (2001).

149. Mulrow & Lohr, *supra* note 141, at 253.

physiological studies, and unsystematic clinical observations.¹⁵⁰ An analysis of the frequency with which various study designs are cited by others provides empirical evidence supporting the influence of meta-analysis followed by randomized controlled trials in the medical evidence hierarchy.¹⁵¹ Although they are at the bottom of the evidence hierarchy, unsystematic clinical observations or case reports may be the first signals of adverse events or associations that are later confirmed with larger or controlled epidemiological studies (e.g., aplastic anemia caused by chloramphenicol,¹⁵² or lung cancer caused by asbestos¹⁵³). Nonetheless, subsequent studies may not confirm initial reports (e.g., the putative association between coffee consumption and pancreatic cancer).¹⁵⁴

Just as in laboratory experiments, evidence about the benefits and risks of medical interventions arises through repetitive observations. A single randomized controlled trial relies on hypothesis testing, specifically assuming the null hypothesis that a new drug is equivalent to the comparator (e.g., placebo). As conceived nearly 100 years ago, interpreting the trial involved calculating the likelihood of the alpha error (p -value) wherein the study suggests that the drug or device is beneficial but the “truth” is that it is not, that is, a false-positive study result. Similarly, a beta error (1 minus power) is the likelihood of a study finding that the drug or device is not beneficial when the “truth” is that it is, that is, a false-negative study result (Table 3).

Table 3. Analogy Between Interpreting a Diagnostic Test and a Drug Study

	Truth	
	Drug +	Drug –
Study +	Power (true positive)	α Type I error (false positive)
Study –	β Type II error (false negative)	True negative

The choice of which specific error rates to use (e.g., false positive or p -value or alpha of 0.05) was suppose to depend on a judgment of the relative consequences of the two errors, missing an effective drug (Type II beta error) or

150. Gordon H. Guyatt et al., *Users' Guides to the Medical Literature: A Manual for Evidence-Based Clinical Practice* (2d ed. 2008) (hereinafter “Guyatt”); see also Michael D. Green et al., Reference Guide on Epidemiology, in this manual.

151. Nikolaos A. Patsopoulos et al., *Relative Citation Impact of Various Study Designs in the Health Sciences*, 293 JAMA 2362–66 (2005).

152. W.T.W. Clarke, *Fatal Aplastic Anemia and Chloramphenicol*, 97 Can. Med. Ass'n J. 815 (1967) (hereinafter “Clarke”).

153. Michael Gochfeld, *Asbestos Exposure in Buildings*, *Envtl. Med.* 438, 440 (1995).

154. Brian MacMahon et al., *Coffee and Cancer of the Pancreas*, 304 New Eng. J. Med. 630–33 (1981) (hereinafter “MacMahon”).

considering an ineffective drug to be effective (Type I alpha error).¹⁵⁵ The null hypothesis, however, assumes equivalence, and so it does not provide any measure of evidence outside of the particular study (e.g., prior studies or biological mechanism or plausibility). Thus, the null hypothesis assumption necessitates abandoning the ability to measure evidence or determine “truth” from a single experiment, so that hypothesis testing is thereby “equivalent to a system of justice that is not concerned with which individual defendant is found guilty or innocent (that is, ‘whether each separate hypothesis is true or false’) but tries instead to control the overall number of incorrect verdicts.”¹⁵⁶ From a Bayesian perspective, the interpretation of a new study depends on whether prior studies showed benefit or harm and on the existence of a biological mechanism or plausibility (e.g., the association between coffee consumption and pancreatic cancer was a “false-positive” result because in further testing the initial finding was not validated and there was no known plausible biological mechanism).¹⁵⁷

Cumulative meta-analysis of treatments enables the accumulation of randomized trial evidence to examine trends in efficacy or risks, overcoming issues of underpowered trials that have insufficient numbers of patients enrolled to reliably detect a benefit. For example, between 1959 and 1988, 33 randomized trials with streptokinase for acute myocardial infarction involving over 35,000 patients had been published. By combining the results of each trial as they occurred, a cumulative meta-analysis found “a consistent, statistically significant reduction in total mortality” with streptokinase use by 1973.¹⁵⁸ In contrast, for many years, physicians used a drug called lidocaine to prevent life-threatening heart rhythm disturbances, yet none of the randomized trials of lidocaine demonstrated any benefit, and finally cumulative meta-analysis found a trend toward harm. When the results of meta-analysis were compared with comments in textbooks and review articles,

discrepancies were detected between the meta-analytic patterns of effectiveness in the randomized trials and the recommendations of reviewers [the review article author]. Review articles often failed to mention important advances or exhibited delays in recommending effective preventive measures. In some cases, treatments that have no effect on mortality or are potentially harmful continued to be recommended by several clinical experts.¹⁵⁹

155. Goodman, *supra* note 74, at 998.

156. *Id.* at 998.

157. MacMahon, *supra* note 154, at 630.

158. Joseph Lau et al., *Cumulative Meta-Analysis of Therapeutic Trials for Myocardial Infarction*, 327 *New Eng. J. Med.* 248–54 (1992).

159. Elliott M. Antman et al., *A Comparison of Results of Meta-Analyses of Randomized Control Trials and Recommendations of Clinical Experts: Treatments for Myocardial Infarction*, 268 *JAMA* 240, 240 (1992).

4. Guidelines

Clinical practice guidelines are “systematically developed statements to assist practitioner and patient decisions about appropriate health care for specific clinical circumstances.”¹⁶⁰ Such guidelines have been widely developed and issued by medical specialty associations, professional societies, government agencies, or health care organizations.¹⁶¹ To avoid biases inherent in review articles (particularly single-authored ones) and to encourage transparency and acceptance, a standard method to develop clinical practice guidelines has emerged. It involves systematically searching for and reviewing the evidence (summarizing the evidence), grading the quality of evidence for each outcome (the certainty of the recommendation), and assessing the balance of benefits versus risks (the size of the treatment effect or the strength of the recommendation).¹⁶² Additional considerations include values and preferences (patient health goals) and costs (resource allocation) where increasing variability or uncertainty in preferences or the presence of higher costs reduces the likelihood of making a strong recommendation.¹⁶³ The number, length, and diversity of guidelines developed by various professional organizations challenge practicing physicians. An attempt to quantify guideline development found exponential growth, with 8 guidelines published in 1990, 138 in 1996, and 855 by mid-1997, including 160 that were more than 10 pages long.¹⁶⁴

With this proliferation, different professional organizations may issue guidelines on the same topic, but with competing recommendations. The composition of the panel and the processes for developing guideline recommendations may differ. For example, the U.S. Preventive Services Task Force (USPSTF) is “an independent panel of non-Federal experts in prevention and evidence-based medicine and is composed of primary care providers (such as internists, pediatricians, family physicians, gynecologists/obstetricians, nurses, and health behavior specialists).”¹⁶⁵ In their evaluation of mammography, the USPSTF “recommends against routine screening mammography in women aged 40 to 49 years” (see

160. Committee to Advise the Public Health Service on Clinical Practice Guidelines, Institute of Medicine, *Clinical Practice Guidelines: Directions for a New Program* 8 (Marilyn J. Field & Kathleen N. Lohr, eds. 1994).

161. See generally *Sofamor Danek Group v. Gaus*, 61 F.3d 929 (D.C. Cir. 1995) (reviewing guidelines issued by the Agency for Health Care Policy and Research in light of the Federal Advisory Committee Act); *Levine v. Rosen*, 616 A.2d 623 (Pa. 1992) (finding that differing guidance from two groups was evidence that reasonable physicians could follow either school of thought); Michelle M. Mello, *Of Swords and Shields: The Role of Clinical Practice Guidelines in Medical Malpractice Litigation*, 149 U. Pa. L. Rev. 645 (2001).

162. David Atkins et al., *Grading Quality of Evidence and Strength of Recommendations*, 328 *BMJ* 1490 (2004).

163. Gordon H. Guyatt et al., *Going from Evidence to Recommendations*, 336 *BMJ* 1049–51 (2008).

164. Arthur Hibble et al., *Guidelines in General Practice: The New Tower of Babel?* 317 *BMJ* 862–63 (1998).

165. U.S. Preventive Services Task Force (USPSTF), Agency for Healthcare Research and Quality, available at <http://www.ahrq.gov/clinic/uspstfix.htm> (last visited Mar. 2, 2011).

also Section IV.D.2).¹⁶⁶ In contrast, based on a writing group composed of its members who are “directly responsible for performing these screening tests,” the Society of Breast Imaging and the American College of Radiology recommend “annual screening from age 40” with mammography for “women at average risk for breast cancer.”¹⁶⁷ Similarly, for prostate cancer screening, the USPSTF update “concludes that the current evidence is insufficient to assess the balance of benefits and harms of prostate cancer screening in men younger than age 75 years.”¹⁶⁸ In the American Urological Association update, a statement panel composed of urologists, oncologists, and other physicians made two recommendations: “The decision to use PSA for the early detection of prostate cancer should be individualized. Patients should be informed of the known risks and the potential benefits” and “Early detection and risk assessment of prostate cancer should be offered to asymptomatic men 40 years of age or older who wish to be screened with an estimated life expectancy of more than 10 years.”¹⁶⁹

Practice guidelines provide recommendations on how to evaluate and treat patients, but because they apply to the general case, their recommendations may not apply to a particular individual patient, or some extrapolation may be required, particularly when multiple diseases exist, as they frequently do in the elderly,¹⁷⁰ or when treatment entails competing risks. For example, anticoagulation is generally recommended for patients with atrial fibrillation (an abnormal heart rhythm disturbance) to prevent blood clots that could cause a stroke, yet anticoagulation can also lead to life-threatening bleeding; therefore, for individual patients, physicians must weigh the risk of developing clots versus the risk of bleeding. Consequently, guidelines typically include statements such as “clinical or policy decisions involve more considerations than this body of evidence alone. Clinicians and policymakers should understand the evidence but individualize decision making to the specific patient or situation.”¹⁷¹ Some physicians who rely on personal style, review articles, and colleagues to influence their clinical practice have been concerned with how guidelines affect clinical autonomy and health care costs.¹⁷²

166. U.S. Preventive Services Task Force, *Screening for Breast Cancer: U.S. Preventive Services Task Force Recommendation Statement*, 151 *Annals Internal Med.* 716–26 (2009).

167. Carol H. Lee et al., *Breast Cancer Screening with Imaging: Recommendations from the Society of Breast Imaging and the ACR on the Use of Mammography, Breast MRI, Breast Ultrasound, and Other Technologies for the Detection of Clinically Occult Breast Cancer*, 7 *J. Am. C. Radiology* 18–27 (2010).

168. U.S. Preventive Services Task Force, *Screening for Prostate Cancer: U.S. Preventive Services Task Force Recommendation Statement*, 149 *Annals Internal Med.* 185–91 (2008).

169. American Urological Association, *Prostate-Specific Antigen Best Practice Statement* (rev. 2009), available at <http://www.auanet.org/content/media/psa09.pdf> (last visited Mar. 2, 2011).

170. Cynthia M. Boyd et al., *Clinical Practice Guidelines and Quality of Care for Older Patients with Multiple Comorbid Diseases: Implications for Pay for Performance*, 294 *JAMA* 716–24 (2005).

171. U.S. Preventive Services Task Force, *Screening for Carotid Artery Stenosis: U.S. Preventive Services Task Force Recommendation Statement*, 147 *Annals Internal Med.* 854–59 (2007).

172. Sean R. Tunis et al., *Internists' Attitudes About Clinical Practice Guidelines*, 120 *Annals Internal Med.* 956–63 (1994).

However, just as clinicians have been reluctant to apply guidelines in practice, courts have generally been slow to apply them in deciding cases.¹⁷³ There are political and legal issues that can arise with the development of guidelines.¹⁷⁴ Political sensitivities, conflicts of interest, and potential lawsuits often silence otherwise innovative and potentially useful guidelines. In 2006, the Connecticut Attorney General launched an antitrust suit against the Infectious Disease Society of America (IDSA) after IDSA promulgated guidelines recommending against the use of long-term antibiotics for the treatment of “chronic Lyme disease (CLD).”¹⁷⁵ Although the Centers for Disease Control and Prevention and the Food and Drug Administration (FDA) findings seemed to concur with IDSA’s guidelines, a strong lobby representing patients afflicted with CLD and the physicians who treated them colored the Attorney General’s decision to file suit.¹⁷⁶ Organizations can violate antitrust laws if their guideline-setting process is an unreasonable attempt to advance their members’ economic interests by suppressing competition. IDSA settled without admitting guilt, but it is clear that organizations must be careful to maintain transparency in the guideline development process.¹⁷⁷

Besides clinical practice guidelines, IOM defines other types of statements: (1) *medical review criteria* are systematically developed statements that can be used to assess the appropriateness of specific health care decisions, services, and outcomes; (2) *standards of quality* are authoritative statements of minimum levels of acceptable performance or results, excellent levels of performance or results, or the range of acceptable performance or results; and (3) *performance measures* are methods or instruments to estimate or monitor the extent to which the actions of a health care practitioner or provider conform to practice guidelines, medical review criteria, or standards of quality.

5. *Vicissitudes of therapeutic decisionmaking*

Medical decisionmaking often involves complexity, uncertainty, and tradeoffs¹⁷⁸ because of unique genetic factors, lifestyle habits, known conditions, medication histories, and ambiguity about possible diagnoses, test results, treatment benefits,

173. Arnold J. Rosoff, *Evidence-Based Medicine and the Law: The Courts Confront Clinical Practice Guidelines*, 26 J. Health Pol., Pol’y & L. 327–68 (2001).

174. One element in the near demise of the Agency for Health Care Policy and Research was a political audience receptive to complaints from an association of back surgeons who disagreed with the AHCPR practice guideline conclusions regarding low back pain. B.H. Gray et al., *AHCPR and the Changing Politics of Health Services Research*, Health Affairs, Suppl. Web Exclusives W3-283-307 (June 2003).

175. John D. Kraemer & Lawrence O. Gostin, *Science, Politics, and Values: The Politicization of Professional Practice Guidelines*, 301 JAMA 665–67 (2009).

176. *Id.* at 666.

177. *Id.* at 666.

178. John P.A. Ioannidis & Joseph Lau, *Systematic Review of Medical Evidence*, 12 J.L. & Pol’y 509–35 (2004).

and therapeutic harms. Given inherent diagnostic and therapeutic uncertainty, physicians often make treatment decisions in the face of uncertainty.

Donald Schön argued that regardless of the professional field, “An artful practice of the unique case appears anomalous when professional competence is modeled in terms of application of established techniques to recurrent events” and that specialization “fosters selective inattention to practical competence and professional artistry.”¹⁷⁹ In the case of a patient with peanut allergies and heart disease, allergy guidelines recommend avoiding beta blockers, but heart disease guidelines recommend beta blockers because they have been shown to prolong life in patients with heart disease. An allergist would recommend against taking a beta blocker, yet a cardiologist would recommend taking it.¹⁸⁰

Well-performed randomized trials provide the least biased estimates of treatment benefit and harm by creating groups with equivalent prognoses. Sticking strictly to the scientific evidence, some physicians may limit their use of medications to the specific drug at the specific doses found to be beneficial in such trials. Others may assume class effects until proven otherwise. Still others may consider additional factors such as out-of-pocket costs for patients or patient preferences. When physicians evaluate patients who might benefit from a treatment but who would have been excluded from the study in which the benefit was demonstrated, they must weigh the risks and benefits in the absence of definitive evidence of benefit or of harm. Indeed, because few medical recommendations are based on randomized trials (the least biased level of evidence) physicians frequently and necessarily face uncertainty in making testing and treatment decisions and trade-offs: Very few treatments come without some risk, and in many disciplines, clear evidence of efficacy and risks of treatment are lacking. In cardiology (one of the better studied areas of medical care), nearly one-half of guideline recommendations are based on expert opinion, case studies, or standards of care.¹⁸¹

Applying well-designed studies to populations of patients represents another problem. The Randomized Aldactone Evaluation Study demonstrated that spironolactone reduced mortality and hospitalizations for heart failure and improved quality of life with minimal risk of seriously high levels of potassium (hyperkalemia).¹⁸² Published in a prominent medical journal, prescriptions for spironolactone rose quickly because of familiarity with the medication and the poor prognosis of patients with heart failure. As opposed to the study population, however, community individuals were older, more frequently women, often

179. Donald A. Schön, *The Reflective Practitioner: How Professionals Think in Action*, at vii (1983).

180. John A. TenBrook et al., *Should Beta-Blockers Be Given to Patients with Heart Disease and Peanut-Induced Anaphylaxis? A Decision Analysis*, 113 *J. Allergy & Clin. Immunol.* 977–82 (2004).

181. Pierluigi Tricoci et al., *Scientific Evidence Underlying the ACC/AHA Clinical Practice Guidelines*, 301 *JAMA* 831–41 (2009).

182. Bertram Pitt et al., *The Effect of Spironolactone on Morbidity and Mortality in Patients with Severe Heart Failure. Randomized Aldactone Evaluation Study Investigators*, 341 *New Eng. J. Med.* 709–17 (1999).

had absolute or relative contraindications to treatment, and had not had tests of their heart function to establish the indication to treat or of their potassium level and kidney function to determine their risk for high potassium levels from treatment.¹⁸³ These factors increased the risk that spironolactone therapy in these patients might lead to high potassium levels that could be life-threatening. Indeed, hospitalizations per 1000 patients for high potassium rose from 2.4 in 1994 to 11.0 in 2001, resulting in an estimated 560 additional hospitalizations for high potassium and 73 additional hospital deaths in older patients with heart failure in Ontario.¹⁸⁴ Criteria for entry into randomized trials of drugs typically exclude individuals with concomitant medication use, medical comorbidities, and female gender, and they may limit participation by socioeconomic status or race and ethnicity, thereby limiting the ability to generalize the results of a trial to the clinical population being treated.¹⁸⁵ Physicians refer to randomized controlled studies as assessments of drug “efficacy” in restricted patient populations, whereas treatment in general clinical populations are often referred to as “effectiveness” studies.

To be sufficiently powered to demonstrate statistical significance,¹⁸⁶ randomized controlled trials usually require high event rates, prolonged followup, or large numbers of patients. Because of impracticality, expense, and the time period needed to obtain long-term outcomes, these trials may often choose a surrogate marker that is associated with a clinically important event or with survival. For example, statins were approved on the basis of their safety and efficacy in lowering cholesterol but were only demonstrated to improve survival in patients with known coronary heart disease years later.¹⁸⁷ Fast-track approval of new drugs for HIV infection was based on safety and efficacy in reducing viral levels (as a surrogate or substitute outcome measure felt to be related to survival) as opposed to demonstration of improved survival.

On the other hand, in the late 1970s, patients with frequent extra heartbeats (ventricular premature contractions) following a heart attack had an increased risk for sudden death. On that basis, those in the then-emerging field of cardiac electrophysiology believed that reducing ventricular premature beats (as a surrogate outcome measure) would decrease subsequent sudden cardiac death. In early randomized controlled trials, oral antiarrhythmic drugs such as encainide and flecainide were approved by FDA on the basis of their ability to suppress these extra heartbeats in patients who had had a myocardial infarction. Years after

183. Dennis T. Ko et al., *Appropriateness of Spironolactone Prescribing in Heart Failure Patients: A Population-Based Study*, 12 J. Cardiac Failure 205–10 (2006).

184. David N. Juurlink et al., *Rates of Hyperkalemia After Publication of the Randomized Aldactone Evaluation Study*, 351 New Eng. J. Med. 543–51 (2004).

185. Harriette G.C. Van Spall et al., *Eligibility Criteria of Randomized Controlled Trials Published in High-Impact General Medical Journals: A Systematic Sampling Review*, 297 JAMA 1233–40 (2007).

186. See Michael D. Green et al., Reference Guide on Epidemiology, in this manual.

187. *Randomised Trial of Cholesterol Lowering in 4444 Patients with Coronary Heart Disease: The Scandinavian Simvastatin Survival Study (4S)*, 344 Lancet 1383–89 (1994).

approval of these drugs, however, a randomized controlled trial designed to demonstrate a survival benefit of these drugs was discontinued after only 10 months because of a statistically significant higher rate of mortality in patients receiving the drugs. Although these drugs effectively suppressed the extra heartbeats, the study found that they also increased the likelihood of fatal heart rhythm disturbances.¹⁸⁸

Prior to approval by FDA, drugs and devices must undergo Phase 1, 2, and 3 clinical trials to demonstrate safety and efficacy. Following preliminary chemical discovery, toxicology, and animal studies, Phase 1 studies examine the safety of new drugs in healthy individuals. Phase 2 studies involve varying drug doses in individuals with the disease to explore efficacy and responses and adverse effects. Based on the dose or doses identified in Phase 2, a Phase 3 study examines drug response in a larger number of patients to again determine safety and efficacy in the hope of getting a new drug approved for sale by regulatory authorities. However, because fewer than 10,000 individuals have usually received the drug during all of these trials, uncommon adverse outcomes may not become apparent until usage is broadened and extended. For example, depending on dosage, between 1 in 24,200 and 1 in 40,500 patients who received the antibiotic chloramphenicol¹⁸⁹ developed fatal aplastic anemia (in which the bone marrow no longer produces any blood cells). This adverse effect was discovered only in the 1960s after chloramphenicol was initially considered safe and had been widely used during the 1950s.¹⁹⁰

For all approved drug and therapeutic biological products, FDA has managed postmarketing safety surveillance since 1969 through the Adverse Event Reporting System. Health care professionals, including physicians, pharmacists, nurses, and others, and consumers, including patients, family members, lawyers, and others, are expected to report adverse events and medication errors. It is a voluntary system with the following limitations: (1) uncertainty that the drug caused the reported event, (2) no requirement for proof of a causal relationship between product and event, (3) insufficient detail to evaluate events, (4) incomplete reporting of all adverse events, and (5) inability to determine the incidence of an adverse event because the actual number of patients receiving a product and the duration of use of those products are unknown.

In 1999, rofecoxib (Vioxx), a Cox-2 selective nonsteroidal anti-inflammatory drug, was approved for pain relief in part on the basis of studies that suggested that it induced less gastrointestinal bleeding than other nonsteroidal anti-inflammatory drugs. In 2004, the manufacturer announced a voluntary worldwide withdrawal

188. *Preliminary Report: Effect of Encainide and Flecainide on Mortality in a Randomized Trial of Arrhythmia Suppression After Myocardial Infarction: The Cardiac Arrhythmia Suppression Trial (CAST) Investigators*, 321 *New Eng. J. Med.* 406–12 (1989).

189. Two pre-*Daubert* cases from the Fifth Circuit dealt with product liability suits against the manufacturer: *Christophersen v. Allied-Signal Corp.*, 939 F.2d 1106 (5th Cir. 1991); *Osburn v. Anchor Labs.*, 825 F.2d 908 (5th Cir. 1987). Clarke, *supra* note 152, at 515.

190. Clarke, *supra* note 152, at 815.

of rofecoxib when a prospective study confirmed that the drug increased the risk of myocardial infarctions (heart attacks) and stroke with chronic use.¹⁹¹

This section demonstrates some of the issues that physicians grapple with in treatment decisions. Some generally avoid using new drugs until sufficient experience with the medication provides an opportunity for unknown adverse effects to emerge following drug approval. Others may be quick to adopt new drugs, especially drugs perceived to have improved safety or efficacy such as through a novel mechanism of action. By withholding use of new drugs, more conservative physicians may avoid the occurrence of unforeseen adverse consequences, but they may also delay the use of new drugs that may benefit their patients. The converse may occur, of course, with physicians who are early adopters of new drugs, tests, or technologies.

Even in a randomized trial in which a drug is found to be beneficial, some patients who received the drug may have been harmed, emphasizing the need to individualize the balancing of risks and benefits and explaining in part why some physicians may not adhere to guideline recommendations. The fundamental dilemma articulated by Bernard in 1865 still haunts the clinician: The response of the “average” patient to therapy is not necessarily the response of the patient being treated.¹⁹² Indeed, the average results of clinical trials do not apply to all patients in the trial. Even with well-defined inclusion and exclusion criteria, variation in outcome risk and, therefore, treatment benefit exists so that even “typical” patients included in the trial may not be likely to get the average benefits.

The Global Utilization of Streptokinase and tPA for Occluded Coronary Arteries Trial is a case in point. The trial suggested that accelerated tissue plasminogen accelerator (tPA) reduced mortality from acute myocardial infarction, with the tradeoff being an increased risk of bleeding from tPA.¹⁹³ In a reanalysis of this study, most (85%) of the survival benefit of tPA accrued to half of the patients (those at highest risk of dying from their heart attack). Some patients with very low risk of dying from their heart attack who received tPA likely were harmed because their risk of intracranial hemorrhage exceeded the benefit.¹⁹⁴ In practice then, even in a randomized controlled trial demonstrating survival benefit, *on average*, those benefits may not accrue to every patient in that trial that received treatment. Therefore, to optimize treatment decisions, physicians attempt to individualize treatment decisions based on their assessment of the patient’s risk versus benefit. Even then, physicians may be reluctant to administer a medication such

191. See generally *In re Vioxx Prods. Liab. Litig.*, 360 F. Supp. 2d 1352 (J.P.M.L. 2005).

192. Salim Yusuf et al., *Analysis and Interpretation of Treatment Effects in Subgroups of Patients in Randomized Clinical Trials*, 266 JAMA 93–98 (1991) (hereinafter “Yusuf”).

193 *An International Randomized Trial Comparing Four Thrombolytic Strategies for Acute Myocardial Infarction. The GUSTO Investigators*, 329 New Eng. J. Med. 673–82 (1993).

194. David M Kent et al., *An Independently Derived and Validated Predictive Model for Selecting Patients with Myocardial Infarction Who Are Likely to Benefit from Tissue Plasminogen Activator Compared with Streptokinase*, 113 Am. J. Med. 104–11 (2002).

as tPA that can cause severe harm such as an intracranial hemorrhage. A single clinical experience with a patient who bled when given tPA might well color their judgment about the benefits of the treatment.

A fundamental principle of evidence-based medicine is that “Evidence alone is never sufficient to make a clinical decision.”¹⁹⁵ Nearly all medical decisions involve some tradeoff between a benefit and a risk. Besides the options and the likelihood of the outcomes, patient preferences about the resulting outcomes should affect care choices, especially when there are tradeoffs such as a risk of complications or dying from a procedure or treatment versus some benefit such as living longer (provided the patient survives the short-term risk of the procedure) or improving their quality of life (relieving symptoms). Besides individualizing risk and benefit assessments, physicians may also deviate from guideline recommendations (“warranted variation”) because of a particular patient’s higher risk of adverse events or lower likelihood of benefit or because of patient preferences for the alternative outcomes, such as when risks occur at different times. For example, given a hypothetical choice between living 25 years for certain or a 50:50 chance of living 50 years or dying immediately, most individuals choose the 25 years for certain. Although both options yield, on average, 25 years, most individuals are risk averse and prefer to avoid the near-term risk of dying. When interviewed, some patients with “operable” lung cancer were quite averse to possible immediate death from surgery, and so, based on their preferences, these patients probably would opt for radiation therapy despite its poorer long-term survival.¹⁹⁶

Besides risk aversion, some treatments may improve quality of life but place patients at risk for shortened life expectancy, and some patients may be willing to trade off quality of life for length of life. When presented with laryngeal cancer scenarios, some volunteer research subjects chose radiation therapy over surgery to preserve their voices despite a reduced likelihood of future survival. “These results suggest that treatment choices should be made on the basis of patients’ attitudes toward the quality as well as the quantity of survival.”¹⁹⁷

To illustrate this principle, a National Institutes of Health Consensus Conference recommended breast-conserving surgery when possible for women with Stage I and II breast cancer¹⁹⁸ because well-designed studies with long-term followup on thousands of women demonstrated equivalence of lumpectomy and radiation therapy or mastectomy for survival and disease-free survival (being alive without breast cancer recurrence). In one study, lumpectomy and radiation appeared to have a lower risk of breast cancer recurrence with 5 women reported to have had breast cancer recurrences following lumpectomy and radiation versus

195. Guyatt, *supra* note 150, at 8; *see also supra* Section IV.C.3.

196. McNeil, *supra* note 63, at 986.

197. *Id.* at 982.

198. *NIH Consensus Conference: Treatment of Early-Stage Breast Cancer*, 265 *JAMA* 391–95 (1991).

10 women after mastectomy.¹⁹⁹ However, breast cancer that recurred in the breast that had been operated on was censored (i.e., deliberately not considered in the statistical analysis).²⁰⁰ When including these censored cancer recurrences, 20 breast cancer recurrences occurred after lumpectomy versus 10 after mastectomy, and so lumpectomy actually had a higher overall risk of recurrence.²⁰¹ As expressed by one woman, “The decision about treatment for breast cancer remains an intensely personal one. The mastectomy I chose . . . felt a lot less invasive than the prospect of six weeks of daily radiation, not to mention the 14% risk of local recurrence.”²⁰² In such a case, patient preferences²⁰³ regarding tradeoffs involving breast preservation and increased risk of breast cancer recurrence or the need for radiation therapy associated with lumpectomy may play an important role in determining the optimal decision for any particular patient.²⁰⁴

D. Informed Consent

1. Principles and standards

Medical informed consent is an ethical, moral, and legal responsibility of physicians.²⁰⁵ It is guided by four ethical principles: autonomy, beneficence, non-maleficance, and justice.²⁰⁶ Autonomy refers to informed, rational decisionmaking after unbiased and thoughtful deliberation. Beneficence represents the moral obligation of physicians to act for the benefit of patients.²⁰⁷ These two principles place physicians in conflict because they wish to provide the care they believe is best for the patient, but because that care usually involves some risk or cost, physicians also recognize that patient preferences may affect their recommendation. In a study examining the incidence of erectile dysfunction with use of a beta blocker medication known to be beneficial, heart disease patients were (1) blinded

199. Joan A. Jacobson et al., *Ten-Year Results of a Comparison of Conservation with Mastectomy in the Treatment of Stage I and II Breast Cancer*, 332 *New Eng. J. Med.* 907–11 (1995) (hereinafter “Jacobson”).

200. Bernard Fisher et al., *Eight-Year Results of a Randomized Clinical Trial Comparing Total Mastectomy and Lumpectomy With or Without Irradiation in the Treatment of Breast Cancer*, 320 *New Eng. J. Med.* 822–28 (1989); Jacobson, *supra* note 199, at 998.

201. Jacobson, *supra* note 199, at 999.

202. Karen Sepucha et al., *Policy Support for Patient-Centered Care: The Need For Measurable Improvements In Decision Quality*, *Health Affairs Supp. Web Exclusives VAR 54, VAR 62* (2004).

203. Proctor & Gamble Pharm., Inc. v. Hoffman-LaRoche, Inc., 2006 WL 2588002, at *10 (S.D.N.Y. 2006) (detailing the testimony of a physician stating that, in addition to efficacy, he considers patient preferences when determining treatment for osteoporosis).

204. Jerome P. Kassirer, *Adding Insult to Injury. Usurping Patients' Prerogatives*, 308 *New Eng. J. Med.* 898–901 (1983) (hereinafter “1983 Kassirer”).

205. Timothy J. Paterick et al., *Medical Informed Consent: General Considerations for Physicians*, 83 *Mayo Clinic Proc.* 313–19 (2008) (hereinafter “Paterick”).

206. Jaime S. King & Benjamin W. Moulton, *Rethinking Informed Consent: The Case for Shared Decision Making*, 32 *Am. J.L. & Med.* 429–501 (2006) (hereinafter “King & Moulton”).

207. *Id.* at 435.

to the drug, (2) informed of the drug name only, or (3) informed about its erectile dysfunction adverse effect. Among those blinded, 3.1% developed erectile dysfunction compared with 15.6% of those given the drug name and 31.2% of those informed about adverse effects, showing that being informed increased the risk for adverse effects and might deprive patients of benefit from a drug because they stop taking it.²⁰⁸ Physicians must balance the desire to provide beneficial care with the obligation to promote autonomous decisions by informing patients of potential adverse effects or tradeoffs.

State jurisdictions differ in their standards for disclosure, with half adopting the physician or professional standard (the information that other local physicians with similar skill levels would provide) and the other half adopting the patient or materiality standard (the information that a reasonable patient would deem important in decisionmaking).²⁰⁹ The informed consent process involves the disclosure of alternative treatment options including no treatment and the risks and benefits associated with each alternative. Discussion should include severe risks and frequent risks, but the courts have not provided explicit guidance about what constitutes sufficient severity or frequency. Patients should be considered by the court to be competent and should have the capacity to make decisions (understanding choices, risks, and benefits). The decision should be voluntary—of free mind and free will, without coercion or manipulation. The language used should be understandable to the patient, and treatment should not proceed unless the physician believes the patient understands the options and their risks and benefits.

Patients may withdraw consent or refuse treatment. Such an action should engender additional discussion, and documentation may include the completion of a withdrawal-of-consent form. In certain situations, exceptions to medical consent may arise in emergencies, when the treatment is recognized by prudent physicians to involve no material risk to patients and when the procedure is unanticipated and not known to be necessary at the time of consent.²¹⁰

The Merenstein case described an unpublished trial in which, during his residency, Dr. Merenstein examined a highly educated man. The examination included a discussion of the relevant risks and benefits regarding prostate cancer screening using the prostate-specific antigen (PSA) test based on recommendations from the U.S. Preventive Services Task Force, the American College of Physicians–American Society of Internal Medicine, the American Medical Association, the American Urological Association, the American Cancer Association, and the American Academy of Family Physicians. Dr. Merenstein testified that the patient declined the test because of the high false-positive rate, the risk of treatment-related adverse effects, and the low risk of dying from prostate cancer.

208. Antonello Silvestri et al., *Report of Erectile Dysfunction After Therapy with Beta-Blockers Is Related to Patient Knowledge of Side Effects and Is Reversed by Placebo*, 24 *Eur. Heart J.* 1928, 1928 (2003).

209. King & Moulton, *supra* note 206, at 430.

210. Paterick, *supra* note 205, at 315.

Another physician seeing the same patient subsequently ordered a PSA without any patient discussion. The PSA was high and the patient was diagnosed with incurable advanced prostate cancer. The plaintiff's attorney argued that despite the guidelines above, the standard of care in Virginia was to order the blood test without discussion, based on four physician witnesses. The jury ruled in favor of the plaintiff.²¹¹

To illustrate the importance of patient preferences, a woman with breast cancer described her experience: "But as the surgeon diagramed incision points on my chest with a felt-tip pen, my husband asked a question: Is it really necessary to transfer this back muscle? The doctor's answer shocked us. No, he said, he could simply operate on my chest. That would cut surgery and recovery time in half. He had planned the more complicated procedure because he thought it would have the best cosmetic result. 'I assumed that's what you wanted.'"²¹² Instead the woman preferred the less invasive approach that shortened her recovery time.

In the research setting, a randomized trial with and without informed consent demonstrated that the process of getting informed consent altered the effect of a placebo when given to patients with insomnia. The first patient of each pair was randomized to no informed consent and the second to informed consent. Out of 56 patients randomized to informed consent, 26 declined to participate in the study (the patients without informed consent had no choice and were unaware of their participation in a study). The informed consent process created a "biased" group because the age and gender for those who declined participation differed significantly from those who did agree to be included in the study. The hypnotic activity of placebo was significantly higher without informed consent, and adverse events were found more commonly in the group receiving informed consent. The study suggests that the process of getting informed consent introduced biases in the patient population and affected the efficacy and adverse effects observed in this clinical trial, thereby potentially affecting the general applicability of any findings involving informed consent.²¹³

Besides physicians, patients may get health information from the Internet, family, friends, and the media (newspapers, magazines, television). Among Internet users, 80% had searched for information on at least 1 of 15 major health topics but use varied from 62% to 89% by age, gender, education, or race/ethnicity.²¹⁴ Conducted between November 2006 and May 2007, a cross-sectional national survey of U.S. adults who had made a medical decision found that Internet use

211. King & Moulton, *supra* note 206, at 432–34; Daniel Merenstein, *A Piece of My Mind: Winners and Losers*, 291 JAMA 15–16 (2004).

212. Julie Halpert, *Health: What Do Patients Want?* Newsweek, Apr. 28, 2003, at 63–64.

213. R. Dahan et al., *Does Informed Consent Influence Therapeutic Outcome? A Clinical Trial of the Hypnotic Activity of Placebo in Patients Admitted to Hospital*, 293 Brit. Med. J. Clin. Res. Ed. 363–64 (1986).

214. Pew Internet, Health Topics, <http://pewinternet.org/Reports/2011/HealthTopics.aspx> (last visited Feb. 12, 2011).

averaged 28% but varied from 17% for breast cancer screening to 48% for hip/knee replacement among those 40 years of age and older.²¹⁵ However, even among Internet users, health care providers were felt to be the most influential source of information for medical decisions, followed by the Internet, family and friends, and then media.

2. Risk communication

Multiple health outcomes may result from alternative treatment choices, and how patients feel about the relative importance of those outcomes varies.²¹⁶ When patients with recently diagnosed curable prostate cancer were presented with 93 possible questions that might be important to patients like themselves, 91 of the questions were cited as relevant to at least one patient.²¹⁷ Communication skills should include patient problem assessment (appropriate questioning techniques, seeking patient's beliefs, checking patient's understanding of the problem); patient education and counseling (eliciting patient's perspective, providing clear instructions and explanations, assessing understanding); negotiation and shared decisionmaking (surveying problems and delineating options, arriving at mutually acceptable solutions); relationship development and maintenance (encouraging patient expression, communicating a supportive attitude, explaining any jargon, and using nonverbal behavior to enhance communication).²¹⁸

Certain forms of risk communication, however, may be confusing and should be avoided: "single event probabilities, conditional probabilities (such as sensitivity and specificity), and relative risks."²¹⁹ An example of a single-event probability would be the statement that a particular medication results in a 30% to 50% chance of developing erectile dysfunction.²²⁰ Although physicians are referring to patients, patients may misinterpret this as referring to their own sexual encounters and having an erectile dysfunction problem in 30% to 50% of their sexual encounters. The preferred natural frequency statement would be "out of 100 people like you taking this medication, 30 to 50 of them experience erectile dysfunction." The natural frequency statement specifies a reference class, thereby reducing misunderstanding.²²¹

215. Mick P. Couper et al., *Use of the Internet and Ratings of Information Sources for Medical Decisions: Results from the DECISIONS Survey*, 30 *Med. Decision Making* 106S–14S (2010).

216. 1983 *Kassirer*, *supra* note 203, at 889.

217. Deb Feldman-Stewart et al., *What Questions Do Patients with Curable Prostate Cancer Want Answered?* 20 *Med. Decision Making* 7–19 (2000).

218. Michael J. Yedidia et al., *Effect of Communications Training on Medical Student Performance*, 290 *JAMA* 1157–65 (2003).

219. Gerd Gigerenzer & Adrian Edwards, *Simple Tools for Understanding Risks: From Innumeracy to Insight*, 327 *BMJ* 741–44 (2003).

220. Gigerenzer, *supra* note 87, at 4.

221. *Id.* at 4; *see also* Section IV.A.2.

Regarding relative risk, consider a statement that taking a cholesterol-lowering medication reduces the risk of dying by 22%.²²² This may be misinterpreted as saying that out of 1000 patients with high cholesterol, 220 of them can avoid dying by taking cholesterol-lowering medications. The actual data show that 32 deaths occur among 1000 patients taking the medication, and 41 deaths occur among 1000 patients taking the placebo. The relative risk reduction equals 9 divided by 41. A preferred way to express the benefit would be the absolute risk reduction (the difference between 41 and 32 deaths in 1000 patients), or to say that in 1000 people like you with high cholesterol, taking a cholesterol medication for 5 years helps 9 of them avoid dying.²²³ Calculating an odds ratio, the cholesterol-lowering medication reduces the odds of dying by 23%; notice that neither the relative risk nor the odds ratio characterizes the number of events without treatment and that the odds ratio always magnifies the risk or benefit when compared with the relative risk. To illustrate further, a relative risk reduction of 20% has very different absolute risk reductions depending on the number of events without treatment. If 20 of 100 patients without treatment would die, then the absolute risk reduction is 4 of 100 or 4% (20% times 20), but if 20 of 100,000 patients without treatment would die, then the absolute reduction is 4 of 100,000 or 0.004%. The number needed to treat is an additional form of risk communication popularized as part of evidence-based medicine to account for the risk without treatment. It is the reciprocal of the absolute risk difference or 1 divided by the quantity 9 lives saved per 1000 ($1 \div (9/1000)$) treated with cholesterol medications in the above example. Therefore 111 patients need to be treated with a cholesterol medication for 5 years to save one of them, or in the illustrative example, with a relative risk reduction of 20%, either 25 or 25,000 would need to be treated to save 1 patient.

In the analysis of mammography for the U.S. Preventive Services Task Force, the number needed to be invited (NNI) for screening to avoid one breast cancer death was 1904 for 39- to 49-year-olds, 1339 for 50- to 59-year-olds, and 377 for 60- to 69-year-olds.²²⁴ To account for possible harm, there is a corresponding determination of the number needed to harm (NNH) that is calculated in the same manner. Considering breast biopsy as a morbidity, 5 women need to undergo breast biopsy for every one woman diagnosed with breast cancer for 39- to 49-year-olds, and the corresponding numbers are 3 for women ages 50 to 59 and 2 for women ages 60 to 69 years old.²²⁵ Estimates of overdiagnosis ranged mostly from 1% to 10%, and so, out of 100 women diagnosed with breast cancer from screening, 1 to 10 of them undergo treatment for a cancer that would never have caused any mortality.²²⁶ Clearly no one can tell if any particular woman has

222. Gigerenzer, *supra* note 87, at 34.

223. *Id.* at 34–35.

224. Heidi D. Nelson et al., *Screening for Breast Cancer: An Update for the U.S. Preventive Services Task Force*, 151 *Annals Internal Med.* 727–37 (2009).

225. *Id.* at 732.

226. *Id.* at 731–732

been overdiagnosed because this is unobservable.²²⁷ The estimated extent of overdiagnosis requires estimating mortality reductions in a screened population compared with an unscreened population over a long period. The difference between the two groups provides an estimate of the extent of overdiagnosis.

To summarize the evidence, “Mammography does save lives, more effectively among older women, but does cause some harm. Do the benefits justify the risks? The misplaced propaganda battle seems to now rest on the ratio of the risks of saving a life compared with the risk of overdiagnosis, two very low percentages that are imprecisely estimated and depend on age and length of followup.”²²⁸ In the USPSTF recommendations for mammography in 40- to 49-year-olds, the focus has been on the first part of their statement “The USPSTF recommends against routine screening mammography in women aged 40 to 49 years.” Although screening has demonstrated benefits, in their view, the benefits of screening do not sufficiently and clearly outweigh the potential harms to make a recommendation that all women 40 to 49 years old have routine screening mammography from a public health or population perspective. Oft neglected, the USPSTF in their immediately subsequent sentence recognizes that individual preferences should affect the care that patients receive: “The decision to start regular, biennial screening mammography before the age of 50 years should be an individual one and take patient context into account, including the patient’s values regarding specific benefits and harms.”²²⁹ The recommendation recognizes that depending on their experiences, values, and preferences, some women may seek the benefit in reducing breast cancer deaths and others may prefer to avoid possible morbidity (breast biopsy and worry) and potential overdiagnosis and overtreatment.

3. Shared Decisionmaking

The “professional values of competence, expertise, empathy, honesty, and commitment are all relevant to communicating risk: Getting the facts right and conveying them in an understandable way are not enough.”²³⁰ Shared and informed decision-making has emerged as one part of patient care. It distinguishes “problem solving” that identifies one “right” course that leaves little room for patient involvement from “decisionmaking” in which several courses of action may be reasonable and in which patient involvement should determine the optimal choice. In such cases, health care choices depend not only on the likelihood of alternative outcomes resulting from each strategy but also on the patient preferences for possible outcomes and their attitudes about risk taking to improve future survival or quality

227. Klim McPherson, *Screening for Breast Cancer—Balancing the Debate*, 341 *BMJ* 234–35 (2010).

228. *Id.* at 234.

229. U.S. Preventive Services Task Force, *Screening for Breast Cancer: U.S. Preventive Services Task Force Recommendation Statement*, 151 *Annals Internal Med.* 716, 716 (2009).

230. Adrian Edwards, *Communicating Risks*, 327 *BMJ* 691–92 (2003).

of life and the timing of that risk whether the risk occurs now or in the future.²³¹ Informed decisionmaking occurs

when an individual understands the nature of the disease or condition being addressed; understands the clinical service and its likely consequences, including risks, limitations, benefits, alternatives, and uncertainties; has considered his or her preferences as appropriate; has participated in decision making at a personally desirable level; and either makes a decision consistent with his or her preferences and values or elects to defer a decision to a later time.²³²

Shared decisionmaking occurs “when a patient and his or her healthcare provider(s), in the clinical setting, both express preferences and participate in making treatment decisions.”²³³

To assist with shared decisionmaking, health decision aids have been developed to help patients and their physicians choose among reasonable clinical options together by describing the “benefits, harms, probabilities, and scientific uncertainties.”²³⁴ In 2007, the legislature in the state of Washington became the first to establish and recognize in law a role for shared decisionmaking in informed consent.²³⁵ The bill goes on to encourage the development, certification, use, and evaluation of decision aids. The consent form provides written documentation that the consent process occurred, but the crux of the medical consent process is the discussion that occurs between a physician and a patient. The physician shares his or her medical knowledge and expertise and the patient shares his or her values (health goals) and preferences. It is an opportunity to strengthen the patient–physician relationship through shared decisionmaking, respect, and trust.

V. Summary and Future Directions

Having sequenced the human genome, medical research is poised for exponential growth as the code for human biology (genomics) is translated into proteins (proteomics) and chemicals (metabolomics) to identify molecular pathways that lead to disease or that promote health. With advances in medical technologies in diagnosis and preventive and symptomatic treatment, the practice of medicine will be profoundly altered and redefined. For example, consider lymphoma, a blood cancer that used to be classified simply by appearance under the microscope as

231. Michael J. Barry, *Health Decision Aids to Facilitate Shared Decision Making in Office Practice*, 136 *Annals Internal Med.* 127–35 (2002).

232. Peter Briss et al., *Promoting Informed Decisions About Cancer Screening in Communities and Healthcare Systems*, 26 *Am. J. Preventive Med.* 67, 68 (2004).

233. *Id.* at 68.

234. Annette M. O’Connor et al., *Risk Communication in Practice: The Contribution of Decision Aids*, 327 *BMJ* 736, 736 (2003).

235. Bridget M. Kuehn, *States Explore Shared Decision Making*, 301 *JAMA* 2539–41 (2009).

either Hodgkin's or non-Hodgkin's lymphoma. As science has evolved, it is now further classified by cellular markers that identify the underlying cancer cells as one of two cells that help with immunity (protecting the body from infection and cancer): T cells or B cells. Current research is attempting to characterize those cells further by identifying underlying genetic and cellular markers and pathways that may distinguish these lymphomas and provide potential therapeutic targets. The growth in the research enterprise, both basic science and clinical translational (the translation of bench research to the bedside or basic science research into novel treatments or diagnostics), has greatly expanded research capacity to generate scientific research of all types.

With greatly expanded knowledge, research and specialization, judgments about admissibility and about what constitutes expertise become increasingly difficult and complex. The sifting of this research into sufficiently substantiated, competent, and reliable evidence, however, relies on the traditional scientific foundation: first, biological plausibility and prior evidence and, second, consistent repeated findings. The practice of medicine at its core will continue to be a physician and patient interaction with professional judgment and communication central elements of the relationship. Judgment is essential because of uncertainties in the underlying professional knowledge or because even if the evidence is credible and substantiated, there may be tradeoffs in risks and benefits for testing and for treatment. Communication is critical because most decisions involve tradeoffs, in which case individual patient preferences for the outcomes that may be unique to patients and that may affect decisionmaking should be considered.

In summary, medical terms shared in common by the legal and medical professions have differing meanings, for example, differential diagnosis, differential etiology, and general and specific causation. The basic concepts of diagnostic reasoning and clinical decisionmaking and the types of evidence used to make judgments as treating physicians or experts involve the same overarching theoretical issues: (1) alternative reasoning processes; (2) weighing risks, benefits, and evidence; and (3) communicating those risks.

Glossary of Terms

- adequacy.** In diagnostic verification, testing a particular diagnosis for its adequacy involves determining its ability to account for all normal and abnormal findings and the observed time course of the disease.
- attending physician.** The physician responsible for the patient's care at the hospital in which the patient is being treated.
- Bayes' theorem (rule).** A mathematical approach to integrating suspicion (pretest probability) with additional information such as from a test result (posttest probability) by using test characteristics (sensitivity and specificity) to demonstrate how well the test performs in individuals with and without the disease.
- causal reasoning.** For physicians, causal reasoning typically involves understanding how abnormalities in physiology, anatomy, genetics, or biochemistry lead to the clinical manifestations of disease. Through such reasoning, physicians develop a "causal cascade" or "chain or web of causation" linking a sequence of plausible cause-and-effect mechanisms to arrive at the pathogenesis or pathophysiology of a disease.
- chief complaint.** The primary or main symptom that caused the patient to seek medical attention.
- coherency.** In diagnostic verification, testing a particular diagnosis for its coherency involves determining the consistency of that particular diagnosis with predisposing risk factors, physiological mechanisms, and resulting manifestations.
- conditional probability.** The probability or likelihood of something given that something else has occurred or is present, for example, the likelihood of disease if a test is positive (posterior probability) or the likelihood of a positive test if disease is present (sensitivity). See Bayes' theorem or rule.
- consulting physician.** A physician, usually a specialist, asked by the patient's attending physician to provide an opinion regarding diagnosis, testing, or treatment or to perform a procedure or intervention, for example, surgery.
- diagnostic test.** A test ordered to confirm or exclude possible causes of a patient's symptoms or signs (distinct from screening test).
- diagnostic verification.** The last stage of narrowing the differential diagnosis to a final diagnosis by testing the validity of the diagnosis for its coherency, adequacy, and parsimony.
- differential diagnosis.** A set of diseases that physicians consider as possible causes for patients presenting with a chief complaint (hypothesis generation). As additional symptoms with further patient history, signs found on physical examination, test results, or specialty physician consultations become available, the likelihood of various diagnoses may change (hypothesis refinement) or new ones may be considered (hypothesis modification) until the diagnosis is nearly final (diagnostic verification).

- differential etiology.** Term used by the court or witnesses to establish or refute external causation for a plaintiff's condition. For physicians, etiology refers to cause.
- external causation.** External causation is established by demonstrating that the cause of harm or disease originates from outside the plaintiff's body, for example, defendant's action or product.
- general causation.** General causation is established by demonstrating, usually through scientific evidence, that a defendant's action or product causes (or is capable of causing) disease.
- heuristics.** Quick automatic "rules of thumb" or cognitive shortcuts often involving pattern recognition that facilitate rapid diagnostic and treatment decisionmaking. Although characteristic of experts, it may predispose to known cognitive errors. See Hypothetico-deductive.
- hypothesis generation.** A limited list of potential diagnostic hypotheses in response to symptoms, signs, and lab test results. See differential diagnosis.
- hypothesis modification.** A change in the list of diagnostic hypotheses (differential diagnosis) in response to additional information, e.g., symptoms, signs, and lab test results. See differential diagnosis.
- hypothesis refinement.** A change in the likelihood of the potential diagnostic hypotheses (differential diagnosis) in response to additional information, e.g., symptoms, signs, and lab test results. As additional information emerges, physicians evaluate those data for their consistency with the possibilities on the list and whether those data would increase or decrease the likelihood of each possibility. See differential diagnosis.
- hypothetico-deductive.** Deliberative and analytical reasoning involving hypothesis generation, hypothesis modification, hypothesis refinement, and diagnostic verification. Typically applied for problems outside an individual's expertise or difficult problems with atypical issues, it may avoid known cognitive errors. See Heuristics.
- individual causation.** See specific causation.
- inductive reasoning.** The process of arriving at a diagnosis based on symptoms, signs, and lab tests. See differential diagnosis.
- inferential reasoning.** See inductive reasoning.
- overdiagnosis.** Screening can lead to "pseudodisease" or "overdiagnosis," e.g., the identification of slow-growing cancers that even if untreated would never cause symptoms or reduce survival because the screening test cannot distinguish the abnormal-appearing cells that would become cancerous from those that would never do so. See overtreatment.
- overtreatment.** The treatment of patients with pseudodisease whose disease would never cause symptoms or reduce survival. The treatment may place

patients at risk for treatment-related morbidity and possibly mortality. See overdiagnosis.

parsimony. In diagnostic verification, testing a particular diagnosis for its parsimony involves choosing the simplest single explanation as opposed to requiring the simultaneous occurrence of two diseases to explain the findings.

pathogenesis. See causal reasoning.

pathology test. Microscopic examination of body tissue typically obtained by a biopsy or during surgery to determine if the tissue appears to be abnormal (different than would be expected for the source of the tissue). The visual components of the abnormality are typically described (e.g., types of cells, appearance of cells, scarring, effect of stains or molecular markers that help facilitate identification of the components) and, on the basis of visual pattern, the abnormality may be classified, e.g., malignancy (cancer) or dysplasia (precancerous).

posttest probability. See predictive value.

predictive value or **posttest probability.** The suspicion or probability of a disease after additional information (such as from a test) has been obtained. The predictive value positive or positive predictive value is the probability of disease in those known to have a positive test result. The predictive value negative or negative predictive value is the probability of disease in those known to have a negative test result.

pretest probability. The suspicion or probability of a disease before additional information (such as from a test) is obtained.

prior probability. See pretest probability.

screening test. A test performed in the absence of symptoms or signs to detect disease earlier, e.g., cancer screening (distinct from diagnostic test).

sensitivity. Likelihood of a positive finding (usually referring to a test result but could also be a symptom or a sign) among individuals known to have a disease (distinct from specificity).

sign. An abnormal physical finding identified at the time of physical examination (distinct from symptoms).

specific causation or **individual causation.** Established by demonstrating that a defendant's action or product is the cause of a particular plaintiff's disease.

specificity. Likelihood of a negative finding (usually referring to a test result but could also be a symptom or a sign) among individuals who do not have a particular disease (distinct from sensitivity).

syndrome. A group of symptoms, signs, and/or test results that together characterize a specific disease.

symptom. The patient's description of a change in function, sensation, or appearance (distinct from sign).

References on Medical Testimony

Lynn Bickley et al., *Bates' Guide to Physical Examination and History Taking* (10th ed. 2008).

Gerd Gigerenzer. *Calculated Risks. How to Know When Numbers Deceive You* (2002).

Trisha Greenhalgh. *How to Read a Paper: The Basics of Evidence-Based Medicine* (4th ed. 2010).

Gordon Guyatt et al., *Users' Guides to the Medical Literature: Essentials of Evidence-Based Clinical Practice* (2d ed. 2009).

Jerome P. Kassirer et al., *Learning Clinical Reasoning* (2d ed. 2009).

Harold C. Sox et al., *Medical Decision Making* (2006).

Sharon E. Straus et al., *Evidence-Based Medicine* (4th ed. 2010).