Celebrating 200 Years at the Forefront of Medicine

A special collection of review and perspective articles published in 2012 to commemorate the NEJM 200th anniversary.
DEAR COLLEAGUE:

In 2012, the New England Journal of Medicine celebrated 200 years of publishing practice-changing medical advances. Throughout the anniversary year, we published special content to commemorate this landmark event. The content that was published in print, a series of Review and Perspective articles that each explore a story of progress, is now collected in this book.

We hope that this collection continues to provide value and interest. The 200th anniversary celebrated the NEJM community — from researcher to clinician reader — that works together to bring discovery to practice every day. Thank you to all of those who contribute to this mission.

Jeffrey M. Drazen, M.D.
Editor-in-Chief
The New England Journal of Medicine
## CONTENTS

<table>
<thead>
<tr>
<th>Page</th>
<th>Title</th>
<th>Author(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>NEJM@200 — Two Centuries at the Journal</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>A Reader’s Guide to 200 Years of the New England Journal of Medicine</td>
<td>Allan M. Brandt, Ph.D.</td>
</tr>
<tr>
<td>10</td>
<td>A Tale of Coronary Artery Disease and Myocardial Infarction</td>
<td>Elizabeth G. Nabel, M.D., and Eugene Braunwald, M.D.</td>
</tr>
<tr>
<td>20</td>
<td>The Perpetual Challenge of Infectious Diseases</td>
<td>Anthony S. Fauci, M.D., and David M. Morens, M.D.</td>
</tr>
<tr>
<td>29</td>
<td>Patients and Doctors — The Evolution of a Relationship</td>
<td>Robert D. Truog, M.D.</td>
</tr>
<tr>
<td>35</td>
<td>A Patient with Asthma Seeks Medical Advice in 1828, 1928, and 2012</td>
<td>Erika von Mutius, M.D., and Jeffrey M. Drazen, M.D.</td>
</tr>
<tr>
<td>43</td>
<td>Major Trends in the U.S. Health Economy since 1950</td>
<td>Victor R. Fuchs, Ph.D.</td>
</tr>
<tr>
<td>48</td>
<td>What We Don’t See</td>
<td>Margaret Kendrick Hostetter, M.D.</td>
</tr>
<tr>
<td>55</td>
<td>The Evolving Roles of the Medical Journal</td>
<td>Scott H. Podolsky, M.D., Jeremy A. Greene, M.D., Ph.D., and David S. Jones, M.D., Ph.D.</td>
</tr>
<tr>
<td>60</td>
<td>Two Hundred Years of Surgery</td>
<td>Atul Gawande, M.D., M.P.H.</td>
</tr>
<tr>
<td>69</td>
<td>The Evolving Primary Care Physician</td>
<td>Susan Okie, M.D.</td>
</tr>
<tr>
<td>75</td>
<td>Two Hundred Years of Cancer Research</td>
<td>Vincent T. DeVita, Jr., M.D., and Steven A. Rosenberg, M.D., Ph.D.</td>
</tr>
<tr>
<td>83</td>
<td>The Burden of Disease and the Changing Task of Medicine</td>
<td>David S. Jones, M.D., Ph.D., Scott H. Podolsky, M.D., and Jeremy A. Greene, M.D., Ph.D.</td>
</tr>
<tr>
<td>89</td>
<td>Two Centuries of Assessing Drug Risks</td>
<td>Jerry Avorn, M.D.</td>
</tr>
<tr>
<td>95</td>
<td>Doctors, Patients, and Lawyers — Two Centuries of Health Law</td>
<td>George J. Annas, J.D., M.P.H.</td>
</tr>
<tr>
<td>101</td>
<td>Unfinished Journey — A Century of Health Care Reform in the United States</td>
<td>Jonathan Oberlander, Ph.D.</td>
</tr>
<tr>
<td>107</td>
<td>Tuberculosis, Drug Resistance, and the History of Modern Medicine</td>
<td>Salmaan Keshavjee, M.D., Ph.D., and Paul E. Farmer, M.D., Ph.D.</td>
</tr>
<tr>
<td>113</td>
<td>Therapeutic Evolution and the Challenge of Rational Medicine</td>
<td>Jeremy A. Greene, M.D., Ph.D., David S. Jones, M.D., Ph.D., and Scott H. Podolsky, M.D.</td>
</tr>
<tr>
<td>119</td>
<td>The Past 200 Years in Diabetes</td>
<td>Kenneth S. Polonsky, M.D.</td>
</tr>
<tr>
<td>128</td>
<td>Two Hundred Years of Progress in the Practice of Midwifery</td>
<td>Michael F. Greene, M.D.</td>
</tr>
<tr>
<td>137</td>
<td>A Glimpse of the Next 100 Years in Medicine</td>
<td>Isaac S. Kohane, M.D., Ph.D., Jeffrey M. Drazen, M.D., and Edward W. Campion, M.D.</td>
</tr>
<tr>
<td>139</td>
<td>200th Anniversary Special Edition Poster Illustrations</td>
<td></td>
</tr>
</tbody>
</table>
Two hundred winters ago, in January 1812, the New England Journal of Medicine and Surgery was born. In the centuries since then, the Journal has chronicled the evolution of medicine and bioscience. Immunization strategies have eradicated smallpox and have protected millions from other formerly fatal infectious diseases. Thousands of new medicines, some targeted with precision, are now available to clinicians. Novel, minimally invasive techniques have changed the way we perform procedures. Numerous human genomes have been sequenced, opening the door to an even better understanding of diseases and their treatments. And although health care delivery today is vastly different from that of prior eras, the core precept of medicine — one person helping another — remains unchanged.

Now, beginning in this Northern Hemisphere winter of 2012, the Journal starts its third century of publication. We have special features that highlight the remarkable progress that has taken place since the publication was launched. Our focus continues to be on the entire medical community, with its scientific insights and advances in clinical care. We will emphasize the contributions of the many clinicians, scientists, and study participants whose dedication, hope, and willingness to take risks have led to profound changes in the health of the nation and of the world.

Our anniversary series begins in this issue with an article by Nabel and Braunwald that describes the first clinical descriptions of angina pectoris; in 1812, the first issue of the Journal opened with an article by John Warren on this same topic. The current article takes us from those early insights to our current understanding of coronary artery disease and the remarkable improvements in its treatment, as well as the increased survival that has ensued. Offerings celebrating the Journal’s 200th anniversary will appear throughout 2012.

Thanks to our readers and contributing authors, the Journal has been fortunate enough to publish important work in many fields and is now considered to be one of the premier journals in the field. As we continue to evolve, our goal is to serve the needs of the clinical and scientific communities. For example, our website, launched in 1996, has become the preferred means of access to the Journal for most of our readers. It facilitates the rapid online publication of written articles and offers rich multimedia content. Our special anniversary website, NEJM200.NEJM.org, includes a timeline of important discoveries across the medical spectrum, historical images with an entertaining Image Challenge, videos that convey the history of the Journal and its editors, and other exciting features.

We want you to become involved in our anniversary celebration. Please post a story, video, or comment about your own path as a physician or health care professional. Tell us who or what inspires your work in medicine, or share an experience that has influenced your work. How does the Journal help you with your practice, patients, and peers? How does it help you with your research? And how could the Journal do a better job of meeting your needs?

The Journal may be viewed as a mirror of our times and, we hope, a force for change as well. During the next century, we intend to continue publishing important new studies, discoveries, and inventions in medicine, as well as to provide current and innovative reviews and features. We welcome your feedback, whether criticism or praise, as we continue to provide the very best information so that you can provide the best care to your patients.
continuity and success. (See Fig. 1, from an 1812 issue.) John Collins Warren, the renowned Boston surgeon, his colleague James Jackson, a founder of Massachusetts General Hospital, and the small group of distinguished colleagues who joined them in starting the New England Journal of Medicine and Surgery, and the Collateral Branches of Science expressed modest and largely local aspirations for the enterprise. Boston, a growing urban center, and the wider New England environs had no medical journal of their own, although much medical knowledge and practice was considered region-specific. Although the name and format of the Journal would vary until 1928, 7 years after its ownership passed to the Massachusetts Medical Society, it remains the longest continuously published medical periodical in the world. The prospectus for the Journal, a call for papers issued in late 1811, explained the goals of Warren and his collaborators: “The editors have been encouraged to attempt this publication by the opinion, that a taste for medical literature has greatly increased in New England within a few years past. New methods of practice, good old ones which are not sufficiently known, and occasional investigations of the modes in common use, when thus distributed among our medical brethren in the country, will promote a disposition for inquiry and reflection, which cannot fail to produce the most happy results.”

At a time of intense debate and controversy regarding the causes...
of disease, the nature of therapeutics, and the basis of professional authority, the young Journal worked to steer a middle course. This was certainly advisable from a commercial point of view, since it could easily alienate diverse medical readers by endorsing a particular therapeutic system or medical wisdoms. As the editors explained in 1837, “It has been a point of ambition with us . . . to make these pages the vehicle of useful intelligence, rather than the field of warfare. . . . The Journal is to all intents and purposes, designed to be a record of medical and surgical facts. It is the medium through which the profession may interchange sentiments and publish the results of their experience” (see box for cited Journal articles).

Given the breadth of the Journal’s interests and contributors over these past two centuries, it serves today as a remarkable resource for understanding the profound changes that have occurred in medicine. The Journal (now available electronically from its first issue onward) is not just a window on clinical medicine and scientific advance; it serves as a basis for investigating the history of medicine in all its complexity: it reflects the relationships of culture, society, economy, and politics to medical knowledge, practice, and the organization of health care. (An interactive timeline providing access to the Journal archives is available with the full text of this article at NEJM.org.) Encyclopedic in its breadth, the Journal has covered virtually every aspect of medical science and its evolution. Indeed, if it were the only source available to the medical historian, much could be recovered to reconstruct this period of seismic change and revolutionary shifts in knowledge and practice.

But despite this emphasis on the growing power of medical science to define and treat disease effectively, there are deep continuities within medicine that a review of the Journal reveals as well. A dive into the digital archive exposes a world of medicine and science radically different from today’s, as well as a stability of orientation and approach to fundamental problems of disease in patients and populations.

THE ENDURING PROBLEM OF DISEASE

The observation and investigation of disease is perhaps the most salient consistent feature of the Journal. From the meticulous description of angina pectoris in the first issue to the early descriptions of AIDS in the early 1980s, there has been an ongoing recognition that therapeutic approaches must await the sharp articulation of symptoms. The first decades of the Journal’s history...
reflected the intensive concern with the epidemics affecting New England and the new nation, and it was not unusual during the early years for authors to direct attention to the environment as a critical variable in the production of disease. John Gorham, an editor writing in 1828, offered a “Medical Report of the Weather and Prevalent diseases for the last Three months.” Such articles may appear both quaint and humorous from our contemporary scientific perch, but they reveal a serious commitment to understanding more fully the vagaries of epidemic disease that could devastate town and country in short order. Furthermore, they offer a complex notion of causality that characterized much 19th-century medicine, in which disease was seen as the result of interactions of the patient’s individual “constitution” with an ever-changing and often dangerous environment. By the late 20th century, many observers would renew concerns voiced more than a century earlier about the environment’s relationship to disease.

In 1832, as cholera raged in New York City, the Journal published an article advocating immediate treatment upon diagnosis with 100 drops of laudanum “mixed with nearly as much of the spirit of essence of peppermint into a wineglass, and filled with brandy.” The author cautioned against the use of bloodletting and cathartics (showing impres- 
sive therapeutic restraint, given their popularity). By the early 20th century, as epidemics of such infectious diseases as tuberculosis and smallpox receded, the Journal began to emphasize studies of the systemic chronic diseas- es, including cancer, heart dis- ease, and diabetes, that would become so characteristic of disease patterns in the developed world. Thus, the Journal reflected the shifts in the burden of disease caused by forces typically beyond the reach of medical intervention.

**DOCUMENTING THERAPEUTIC INNOVATION**

The Journal provides a powerful record of the course taken by medical science and its applications over a 200-year period. It quickly became a conduit for reporting new investigations and findings and for summarizing and disseminating evolving medical knowledge across the widest range of practice. After issuing favorable reports on bloodletting, herbal treatments, and other “he- roic” practices of the early 19th century, the Journal began to reflect a growing skepticism toward such approaches. Authors increasingly pointed to the benefits of the healing powers of nature — *vis medicatrix naturae* — as physicians came to recognize some of the iatrogenic effects of their interventions that had previously been difficult to differen- tiate from the course of serious disease. Therapeutics based on ancient notions of humoral excess and depletion gave way to a renewed emphasis on empirical ob- servation and experiment. The first demonstration of surgical anesthesia, conducted at Massa- chusetts General Hospital in 1846 in an amphitheater soon to be re- named the “Ether Dome,” was first reported in the Journal (Fig. 2). Others quickly began using ether in their practices. One surgeon wrote in the Journal, “I performed the amputation of an arm, the second under the use of ether, while the patient was dreaming of her harvest labors in Ireland, and felt grat- ing but not painful sensations, ‘as if a reaping-hook was in her arm’” (1850).

The rise of the germ theory was vigorously debated in the Journal in the late 19th century. As one author noted, “Whether these organisms are of vegetable or animal origin, whether they are really the cause of the dis- eases they accompany, either by the activity which they exercise as living organisms, or by the products they give rise to, whether they are actually the contagious power, is a question still sub judice” (1871). But the idea of specific causes of specific diseases would come to dominate the pages of the Journal by the end of the century. By then, the decided advan- tages of “aseptic surgery,” “aseptic midwifery,” and other approaches to reducing infection had been demonstrated by the prevailing anecdotal logic of repeated case reports (Fig. 3).

The diagnostic opportunities inherent in the germ theory were apparent in a discussion of Paul Ehrlich’s techniques for identifying the tubercle bacillus in sputum (1891). Early in the 20th century, Ehrlich’s introduction of Salvarsan to treat syphilis heralded a new age of “magic-bullet” med- 
icine, in which therapies would be identified and designed to target specific pathogens. Although the full therapeutic implications of these insights — frequently reported in the Journal — would be delayed until the introduction of antibiotics — notably, penicil- 
in — in the 1940s, the revolu- 
tionary aspects of this approach to infection would not be lost on Journal readers. With germ theory, the scientific foundation for the use of vaccines, new and old, was at last demonstrated. Compulsory vaccination was a constant topic
of debate in the Journal from its earliest years, as it remains in contemporary societies, conveying ongoing tensions between social mandates and individual liberties, the good of the many and the risks to the few.

The Journal would report many “firsts” in subsequent years, including the first major quantitative study linking smoking to lung cancer (1928), the introduction of the pulmonary artery catheter (1970), and early clinical descriptions of AIDS (1981). And breakthroughs reported elsewhere quickly found their way to the Journal. Insulin, for example, first described in the Journal of Laboratory and Clinical Medicine in 1922, received extensive review and discussion in the Journal later that year, and many articles analyzing its optimal use in diabetes followed.

Myriad new diagnostic technologies accompanied changes in the theory and treatment of disease. The Journal offers a window onto the rise of new medical technologies, from stethoscopes to improved tourniquets, from Wilhelm Roentgen’s x-rays to magnetic resonance imaging and beyond. Technologies that probe and visualize the body represented a critical aspect of the development of modern medical practice and the conceptualization of pathologies. The focus on disease specificity and causal mechanism that emerged with the germ theory would ultimately drive research at the molecular and genetic level that continues to be reflected in the Journal.

The consistent editorial goal was to bring cutting-edge knowledge to a wide audience of clinicians in a timely way. As a result, the Journal followed a “synthetic” approach, incorporating sophisticated findings (from observation, experimentation, and the laboratory) into pragmatic clinical and population-based approaches to the amelioration of disease and its associated morbidity and mortality. But viewing the Journal merely as a “record” of medical “progress” would be to diminish its value in recovering a complex view of our medical past, often filled with conflicts, debates, and indicators of values and beliefs that transcended scientific developments.

EDUCATION AND THE DISSEMINATION OF MEDICAL KNOWLEDGE

From the beginning, the Journal has critically covered essential debates about the character and
quality of medical education. The editors considered one of their primary goals to be educating the profession, so assessment of medical school programs was in harmony with their mission; after all, these schools produced their readers. In the late 19th century, the Journal frequently noted the great inconsistencies in educational standards and quality. A decade before the publication of the Flexner reforms, prominent Boston physician Henry Bowditch anticipated many key aspects of the report when he called for linking medical education to universities, lengthening the course of study, and demanding deeper preparation in the sciences and wider domains of knowledge (1900). He argued for active learning to replace didactics, a theme that would echo through the debates about medical education. As late as 1900, when Bowditch proposed his reforms in the Journal, less than half the students at Harvard Medical School had completed a college education. After the publication of the Flexner Report in 1910 and the massive changes that followed, the Journal applauded the consolidation of medical education on a new scientific foundation.

But controversies about the relationship of scientific and clinical expertise, generalism and specialization, and the medical curriculum continued unabated in the Journal's pages. The recognition that new knowledge crowded the curriculum intensified the debates about medical education. “No individual can grasp all the facts of scientific interest. . . . The attempt to teach everything has been abandoned,” noted the Journal in 1926. Given this reality, concerns frequently arose that clinical insight and the professional values of caregiving might be marginalized in the battles for time and authority in the curriculum. With increased emphasis on the basic sciences, some physicians lamented the loss of time at the bedside. “The number of formal lectures could be reduced and the teaching revitalized by more frequent contact with clinical material,” one physician noted in the Journal in 1928.

Physicians frequently insisted that medicine was both an art and a science. In part, this claim reflected concerns already articulated in the late 19th and early 20th centuries that medicine's turn to a new science had alienated deeper humanistic values and concerns about the patient. These apprehensions would be reflected in debates about how best to prepare new generations of physicians. Even while the Journal voiced such anxieties, it continued to advocate aggressively for evidence-based practice long before the nomenclature “evidence-based medicine” came into vogue. “Only thus can medicine progress; only through observation and experiment can the world grow in wealth of knowledge,” explained one editorial (1919). Simultaneously, however, the Journal repeatedly directed attention to the vagaries and values of the doctor–patient relationship (1935).

There was never a great distance between the Journal's interest in developments in medical education and its commitment to professional education. In 1895, Dr. Richard Cabot of Massachusetts General Hospital began using case records centered on post-mortems, operative findings, and diagnostic uncertainty “in private quiz exercises at [his] office” (1939). At the turn of the 20th century, the Journal started publishing these “exercises” in clinical thinking under his editorial supervision. In 1923, the “Case Records of the Massachusetts General Hospital” became a special feature (Fig. 4) — and it continues today (1948). These “teaching cases” remained at the heart of the Journal's ongoing commitment to “continuing” medical education, even as the rigors of pub-
TOO MUCH TO KNOW

With the radical expansion and shifting of the scientific basis of medicine at the turn of the 20th century, the *Journal* recorded growing interest in and concern about specialization. From a largely undifferentiated notion of medical training and expertise, many new and specific divisions of the medical profession developed.⁶ Where as the *Journal* came to view specialization as the inevitable result of exploding medical knowledge, the creation of medical “specialism” was viewed with considerable skepticism and lamentation, if not outright hostility. Much ink was spilled in attempts to determine the relationship of general knowledge and practice to increasingly specific (and limited) areas of expertise. How would the “whole patient” be treated when specialties had divided the body into organ systems and medicine into categories of disease and authority over various technologies and techniques?

Although “general practitioner” was becoming a term of nostalgia, if not derision, commentators frequently pointed to the dilemmas of coordination implicit in the growth and division of medical knowledge. “With this specialization . . . are we not losing to a considerable degree our professional competence?” the *Journal* asked in 1924. “Are we not losing sight of that fundamental thread of truth that gives us a perspective of the real value of our work; that enables us to consider our patient as an individual and not a pathological unit of a human body or a representative of an age group?” Such concerns would reverberate through the *Journal’s* second century; systematic expertise and notions of standardization displaced to a considerable degree the personal and intimate connections of local practice and continuity of care. “How much can the specialist know of home conditions, of family difficulties, and their relation to the case?” lamented a physician in 1923.

The simultaneous explosion of new knowledge and its fracturing into specialized fields posed particular issues for the *Journal* itself. With more and increasingly specialized papers arriving at its doorstep, the traditional board of editors could no longer critically evaluate their quality without wider consultation. Peer review now required identifying independent experts and soliciting their assessment of submissions.

THE PERMEABLE BOUNDARIES OF SCIENCE AND MEDICINE

Despite the *Journal’s* deep commitment to empirical reasoning and scientific rationality, cultural and political beliefs and values are ever apparent in its pages. In some instances, professional prerogatives and social assumptions are exposed. For example, when the introduction of women students at Harvard Medical School was debated in 1878, the *Journal* expressed concern: “It would . . . be impossible to avoid an indiscriminate mingling of the sexes in the dissecting or autopsy rooms, and in the amphitheatres, to witness exercises which justly have hitherto been thought of a character to be witnessed by one sex alone.” Harvard would ultimately admit women in 1945, when the war caused a shortage of male candidates. In the 1950s, the *Journal* expressed regret that some women physicians with children “have found it impossible to carry on their practices” (1954).

Nowhere, perhaps, is the porous membrane of medicine and science seen more clearly than in the *Journal’s* coverage of eugenics and mandatory sterilization in the early 20th century. Advocating for the sterilization of the “unfit,” the *Journal* argued in 1928, “Sterilization of mental defectives seems, on first thought, like taking decided liberties with the individual. Viewed as a public health measure, however, it becomes apparent that it is a practical method of reducing the incidence of defectiveness and eventually preventing further deterioration of the race. It is more humane and practical than permanent segregation of the individual, it is simple and it is effective.”

Most alarming was the admiration expressed for developments early during the Third Reich. The *Journal* announced in 1934, “Germany is perhaps the most progressive nation in restricting fecundity among its unfit. . . . In America it is probable that the sentiment of the people is not ready for the adoption of the German
plan, and will be inclined to restrict compulsory sterilization to a small proportion of those who might properly be regarded as especially fit subjects of this treatment."

Of course, such pronouncements can only be fully understood through deeper investigation of their context, but they remind us that medicine and journals disseminating scientific knowledge are not immune from deeply held values or dangerous social and political forces. By the end of World War II, the Journal would utterly condemn Nazi science and medicine, noting the horrors inflicted on behalf of medical science in a totalitarian regime (1949). After the Holocaust, medical ethics would be radically reframed with a new emphasis on patient autonomy and informed consent. In 1966, anesthesiologist Henry Beecher published an exposé in the Journal of unethical research using human subjects that had previously appeared in major American medical journals (including the Journal). That article was an important contribution to the larger process of recognizing patients’ rights — as research subjects and in clinical care. Beecher concluded, “It is absolutely essential to strive for [consent] for moral, sociologic and legal reasons.”

REFLECTIONS ON THE JOURNAL AT 200

While the Journal embraced new science and the critical apparatus of peer review, it rejected a narrow notion of specialization, continuing to cover the widest range of contributions to medical knowledge. In an increasingly atomized medical world, the commitment to publish on cross-cutting educational, professional, ethical, and policy issues pulled together diverse readers, including physicians and other health care providers, public health experts, and policymakers, around issues that were often beyond their immediate expertise. The radical growth of teaching hospitals, federal funding for basic science and clinical research, and academic medical centers (all developments reflected in the Journal) have been crucially linked to the Journal’s growth, stability, and success.

During the Journal’s first 200 years of publication, medicine and health care moved from the social periphery to become dominant aspects of our science, culture, and economy. The Journal unquestionably owes its success and stability to this monumental shift in the status, authority, and impact of biomedicine. But the Journal has also played a critical role in these developments. By combining a commitment to publishing papers of scrupulous scientific merit across wide-ranging domains, with a recognition of the central questions and values uniting the profession, the Journal has remained true to its founders’ vision. It has recognized that advances in medical science can finally be assessed only in the context of delivery, care, and outcome. The Journal reflects today, as at its inception, a view that medical science and its applications are fundamentally tied to patient care and public health. It therefore continues to draw a range of readers wider than Warren could have imagined.

Today, the ability to disseminate publications so widely through digital technologies promises to bring innovations in medical knowledge to a new set of global constituents. The first hundred issues of Warren’s journal were, of course, distributed on horseback.

No one having looked at the last 200 years of medicine — in which changes came so quickly and dramatically — would hazard a prediction about the next two decades, let alone the next two centuries. Nonetheless, as vast as these changes have been, there are substantial continuities in the nature of scientific inquiry, the care of the patient, and powerful questions concerning the public’s health. Today we say that medicine is a “public good.” But though their society, culture, skills, and science differed so profoundly from ours, those who started this Journal in 1812 undoubtedly understood that basic truth.

Disclosure forms provided by the author are available with the full text of this article at NEJM.org.

From the Department of Global Health and Social Medicine, Harvard Medical School, Boston; and the Department of the History of Science and the Office of the Dean, Graduate School of Arts and Sciences, Harvard University, Cambridge, MA.


Copyright © 2012 Massachusetts Medical Society.
A Tale of Coronary Artery Disease and Myocardial Infarction

Elizabeth G. Nabel, M.D., and Eugene Braunwald, M.D.

The remarkable facts, that the paroxysm, or indeed the disease itself, is excited more especially upon walking up hill, and after a meal; that thus excited, it is accompanied with a sensation, which threatens instant death if the motion is persisted in; and, that on stopping, the distress immediately abates, or altogether subsides; have . . . formed a constituent part of the character of Angina Pectoris.¹

“Remarks on Angina Pectoris” by John Warren, M.D., appeared in 1812 as the first article in the first issue of The New England Journal of Medicine and Surgery.¹ Warren’s description of angina pectoris (derived from the Latin angina, “infection of the throat”; from the Greek ἄγχωνη, “strangling”; and from the Latin pectus, “chest”) is equally apt for physicians and medical students today. At the time, the pathogenesis was unknown, and treatment consisted of bloodletting, a tincture of opium, bed rest, or a combination thereof. In 1799, Caleb H. Parry speculated that Syncope Anginosa was related to coronary-artery ossification (i.e., calcification), occurring predominantly in men at about 50 years of age and rarely in women or children.²

Medical knowledge in the 18th and 19th centuries was grounded in clinical observation and anatomical dissection. Cardiovascular science emerged in the physiological era of the late 19th and early 20th centuries, first in Europe and subsequently in North America. To celebrate the 200th anniversary of the New England Journal of Medicine, our essay focuses on the themes of coronary artery disease and myocardial infarction to highlight the interplay between science and medicine, emphasizing how the remarkable advances in our understanding of the pathogenesis of heart disease have produced life-saving and life-extending therapies.

The Emergence of Coronary Artery Disease

After Heberden’s clinical description of angina³ in 1772, it took almost a century for pathologists to focus their attention on the coronary arteries and describe thrombotic occlusions in addition to “ossification.” However, for decades thereafter, these observations were not related to the symptoms of myocardial ischemia, which had become well known to physicians. Near the end of the 19th century, cardiovascular physiologists noted that occlusion of a coronary artery in the dog caused “quivering” of the ventricles and was rapidly fatal.⁴,⁵ These three great branches of medical knowledge — clinical medicine, pathology, and physiology — advanced in separate yet parallel universes. In 1879, the pathologist Ludvig Hektoen concluded that myocardial infarction is caused by coronary thrombosis “secondary to sclerotic changes in the coronaries.”⁶ In 1910, two Russian clinicians who were trained in pathology described five patients with the clinical picture of acute myocardial infarction, which was confirmed at postmortem examination.⁷ Two years later,
James B. Herrick emphasized total bed rest as the treatment for this condition and by 1919 had used electrocardiography to diagnose it. These approaches were the standard of care for patients with myocardial infarction until the mid-20th century.

CORONARY RISK FACTORS

Two seminal developments in the 1960s radically changed our understanding and management of acute myocardial infarction, which struck down and killed or greatly impaired apparently healthy men in their 40s or 50s, during their most productive years. One of the first acts of the National Heart Institute, later renamed the National Heart, Lung, and Blood Institute (NHLBI), was to establish the Framingham Heart Study in 1948, which involved the close collaboration of professionals from three disciplines: clinical cardiology, biostatistics, and epidemiology. Their goal was to understand how heart disease developed by studying the lifestyles of the residents of Framingham, Massachusetts. The first description of their findings, “Factors of Risk in the Development of Coronary Heart Disease,” indicated that elevations in blood pressure and cholesterol levels were associated with an increased incidence of ischemic heart disease and acute myocardial infarction. The study also showed a high frequency of myocardial infarction among women, which often occurred later in life than it did in men. The identification of elevated blood pressure and cholesterol levels as risk factors and the institution by the NHLBI of national programs to educate clinicians and the public about the importance of controlling these risk factors have contributed to dramatic improvements in age-adjusted cardiac death rates (Fig. 1). (See the timeline in the Supplementary Appendix, available with the full text of this article at NEJM.org.) With the identification of these coronary risk factors and others that followed, the veil that masked the underlying mechanisms in angina and myocardial infarction was lifted, and the concept that coronary heart disease and its complications could be prevented was introduced. Increasingly large multicenter clini-
cal trials subsequently showed that both primary and secondary prevention was possible when steps were taken to lower blood pressure and serum total cholesterol. Fortunately, drugs to reduce these risk factors safely became available as a result of a series of productive collaborations between industry and academic medicine.

**CORONARY CARE UNITS**

Until 1961, patients with acute myocardial infarction — if fortunate enough to survive until they reached a hospital — were placed in beds located throughout the hospital and far enough away from nurses’ stations that their rest would not be disturbed. Patients were commonly found dead in their beds, presumably from a fatal tachyarrhythmia. Indeed, the risk of death occurring in the hospital was approximately 30%. The development of the coronary care unit, which provided continuous monitoring of the electrocardiogram, closed-chest cardiac resuscitation, and external defibrillation, reduced in-hospital mortality by half among patients admitted with acute myocardial infarction.

**PHYSIOLOGY, CARDIAC CATHETERIZATION, ANGIOPLASTY, AND SURGERY**

The publication of De Motu Cordis in 1628, William Harvey’s seminal description of the circulation and the function of the heart, set the stage for the physiological era several centuries later. The 19th-century French physiologist Claude Bernard catheterized animals and measured the pressures in the great vessels and cardiac chambers. This experiment led to the first human cardiac catheterization, performed by Werner Forssman — on himself — in 1929, which in turn led to the exploration of cardiac hemodynamics by André Frédéric Cournand and Dickinson W. Richards. All three of these investigators were awarded the Nobel Prize in Physiology or Medicine in 1956.

Cardiac catheterization paved the way for the development of coronary arteriography in 1958. When combined with left ventriculography, the use of this imaging technique allowed clinicians to elucidate the natural history of coronary artery disease. Coronary arteriography and left ventriculography became the standard diagnostic tool for defining pump function and vessel anatomy and provided the foundation for surgical treatment by means of coronary revascularization. The development and refinement of the technique of open-heart surgery required close collaborations among surgeons, engineers, cardiologists, anesthesiologists, and hematologists. The field of invasive cardiology soon emerged, built on the pioneering work of Dotter and Judkins, although Andreas Grünzig is considered the father of percutaneous interventional cardiology (Fig. 2). The initial technique of balloon angioplasty was followed by the insertion of bare-metal stents, and today, drug-eluting stents are used to prevent coronary restenosis. Once again, cross-disciplinary collaborations, this time among engineers, cardiologists, radiologists, and pathologists, forged remarkable advances in terms of improved vascular devices and techniques. Obstructions in the heart and circulation can now be successfully opened, and abnormal openings successfully closed, in the catheterization laboratory.

**MODERN THERAPY**

By the 1970s, in-hospital mortality from acute myocardial infarction was approximately 15%, and in the first year after hospital discharge, roughly 10% of patients died from left ventricular failure associated with large infarctions. Studies in laboratory animals suggested that infarct size could be reduced by rectifying the imbalance between myocardial oxygen supply and demand. In 1976, cardiologists were able to open acutely occluded coronary arteries by intracoronary infusion of the fibrinolytic agent streptokinase. The Italian Group for the Study of Streptokinase in Myocardial Infarction (Gruppo Italiano per lo Studio della Streptochinasi nell’Infarto Miocardico) (GISSI) trial, one of the first cardiac “mega-trials” (involving more than 10,000 patients), showed that intravenous streptokinase reduced early mortality in patients with acute myocardial infarction. The Second International Study of Infarct Survival (ISIS-2) showed that the addition of aspirin (an antiplatelet drug) led to further reductions in mortality. Coronary angioplasty and stenting, together with newer, more potent platelet inhibitors (e.g., P2Y12 and glycoprotein IIb/IIIa platelet–receptor blockers), further reduced in-hospital mortality to about 7%. The efficacy of
these treatments, including ventricular defibrillation, depends on a short interval between the onset of symptoms and the patient's arrival at the hospital. Considerable progress has been achieved since the 1970s through massive public and professional education programs led by partnerships among the NHLBI, the American Heart Association, and the American College of Cardiology. It was also in this era that randomized, controlled clinical trials became the paradigm for the advancement of clinical cardiovascular therapeutics.

Based on studies in animals showing the benefits of angiotensin-converting–enzyme inhibitors in experimentally induced myocardial infarction, the Survival and Ventricular Enlargement (SAVE) trial showed that long-term administration of these inhibitors reduced mortality among patients with left ventricular dysfunction after infarction. The use of beta-adrenergic blockers and aldosterone blockers in these patients further reduced mortality. Despite these notable advances, however, life-threatening heart failure still occurs late in patients with extensive ventricular scarring as a consequence of large infarcts. Implantable defibrillators, cardiac resynchronization therapy with pacemakers, and left ventricular assist devices have improved the prognosis for such patients. Cardiomyocytes from patients with severe heart failure have been found to be deficient in sarcoplasmic reticulum Ca\(^{2+}\) ATPase (SERCA2a). In a pilot study, an adeno-associated virus has been used to deliver the gene for SERCA2a by intracoronary infusion, with seemingly beneficial results.

In the late 1930s, alert clinicians called attention to what we now refer to as unstable angina and non–ST-segment elevation acute coronary syndrome. Patients with this disorder have severe anginal pain, usually at rest, often with biochemical evidence of some myonecrosis and severe, multivessel, obstructive coronary artery disease. These patients now outnumber those with ST-segment elevation myocardial infarction by about 3 to 1 and account for about 1 million hospital admissions yearly in the United States. Patients with non–ST-segment elevation acute coronary syndrome have improvement with prompt coronary revascularization and require inhibition of the two clotting-system pathways with aspirin and a platelet P2Y\(_{12}\)-receptor antagonist (e.g., clopidogrel), together with an anticoagulant (low-molecular-weight heparin). Their course after hospital discharge is improved by an intensive reduction in low-density lipoprotein (LDL) cholesterol levels and administration of an anticoagulant. The latter advance is reported in this issue of the *New England Journal of Medicine*, high-

---

**Figure 2. Percutaneous Coronary Angiography and Intervention in the Treatment of Arterial Stenosis.**

Panel A shows the technique of percutaneous transluminal coronary angioplasty, as pioneered by Andreas Grüntzig in 1979, and the catheter used in the procedure. To treat stenosis of the coronary artery (top image), the catheter is introduced over a guidewire, passed across the lesion (middle image), and then inflated, dilating the artery (bottom image). Panel B shows severe stenosis of a left anterior descending artery as revealed on coronary angiography (arrow in upper left image). After angioplasty was performed (upper right and lower left images), repeat coronary angiography 4 weeks later showed improved patency of the artery (arrow in lower right image). (Reprinted from Grüntzig et al.)

---

**UNSTABLE ANGINA AND NON–ST-SEGMENT ELEVATION MYOCARDIAL INFARCTION**

Panel A shows the technique of percutaneous transluminal coronary angioplasty, as pioneered by Andreas Grüntzig in 1979, and the catheter used in the procedure. To treat stenosis of the coronary artery (top image), the catheter is introduced over a guidewire, passed across the lesion (middle image), and then inflated, dilating the artery (bottom image). Panel B shows severe stenosis of a left anterior descending artery as revealed on coronary angiography (arrow in upper left image). After angioplasty was performed (upper right and lower left images), repeat coronary angiography 4 weeks later showed improved patency of the artery (arrow in lower right image). (Reprinted from Grüntzig et al.)
lighting that after 200 years, the clinical problems of coronary artery disease and myocardial infarction are still being actively investigated and reported in the Journal.

CORONARY ATHEROSCLEROSIS

The ability to access vascular and cardiac tissue rapidly led to the development of animal models of vascular disease, as well as clinical studies in humans. Two lines of investigation in the 1970s and 1980s forged the field of vascular biology: the observations that thrombotic occlusion of a ruptured or eroded atherosclerotic plaque led to acute myocardial infarction and 1980s forged the field of vascular biology: the observations that thrombotic occlusion of a ruptured or eroded atherosclerotic plaque led to acute myocardial infarction.

Thrombi that interrupt blood flow on either a temporary basis (causing unstable angina) or a permanent one (causing myocardial infarction). Physical disruption (rupture) of the plaque exposes procoagulant material within the core of the plaque to coagulation proteins and platelets, triggering thrombosis.

Evidence of the causative role of LDL cholesterol in atherosclerosis is threefold: first, genetic mutations that impair receptor-mediated removal of LDL cholesterol from plasma cause fulminant atherosclerosis; second, animals with low LDL-cholesterol levels have no atherosclerosis, whereas increasing these levels experimentally leads to disease; and third, human populations with low LDL-cholesterol levels have minimal atherosclerosis, and the process increases in proportion to the level of LDL cholesterol in the blood.

A remarkable victory for patients with coronary artery disease came when the LDL-cholesterol pathway was delineated and the use of 3-hydroxy-3-methylglutaryl coenzyme A (HMG-CoA) reductase inhibitors (statins), discovered by Akira Endo, was developed to lower LDL-cholesterol levels. Brown and Goldstein’s discovery of the LDL-receptor pathway, for which they were awarded the 1985 Nobel Prize in Physiology or Medicine.
Normal endothelial surface

Denuded endothelial surface

B

Endothelial cell

Monocyte

T cell

Intima

Dendritic cell

Macrophage

Foam cell

Media

Smooth-muscle cells

Adventitia

Fibroblast

Mast cell

Collagen

Migrating smooth-muscle cells

C

Foam cell

Apoptotic bodies

Apoptotic macrophage

Cholesterol crystal

Dividing smooth-muscle cell

Vasa vasorum

D

Platelet

Thrombus formation

Fibrin-
cap rupture

Lipd coe

2 gauge 5 min

ACh ~8.0 ~7.5 ~7.0 ~6.5 ~6.0 Chamber washout

NA ~8.0
FIGURE 4. The LDL-Receptor Pathway and Treatment with LDL Cholesterol–Lowering Drugs, which Improves Cardiovascular Outcomes.

Panel A shows the sequential steps in the LDL cholesterol–receptor pathway. After binding to the LDL receptor, LDL cholesterol undergoes receptor-mediated internalization, degradation by lysosomal hydrolases, and generation of LDL-derived cholesterol, which regulates the cell’s cholesterol content by suppressing the activity of 3-hydroxy-3-methylglutaryl coenzyme A (HMG-CoA) reductase, activating a cholesterol-esterifying enzyme (acyl CoA:cholesterol acyltransferase [ACAT]), and inhibiting transcription of the LDL-receptor gene. (Adapted from Goldstein and Brown.43) Panel B shows the beneficial effects of LDL cholesterol–lowering therapy on the risks of fatal coronary artery disease and nonfatal myocardial infarction (left) and on the need for coronary-artery bypass surgery or angioplasty (right) in patients with average cholesterol levels, as shown in the Cholesterol and Recurrent Events (CARE) trial.45 Kaplan–Meier estimates of the incidence of coronary events in patients treated with either a statin (pravastatin) or a placebo are shown. (Adapted from Sacks et al.45)
Prize in Physiology or Medicine, provided a genetic cause for myocardial infarction in persons with familial hypercholesterolemia and introduced three general concepts to cell biology: receptor-mediated endocytosis, receptor recycling, and feedback regulation of receptors. This last concept is the mechanism by which statins selectively lower LDL-cholesterol levels in plasma, reducing the risk of myocardial infarction and prolonging life, as shown in multiple, definitive clinical trials (Fig. 4B).45,47

However, statin therapy does not eliminate cardiovascular risk.48,49 Levels of high-density lipoprotein (HDL) cholesterol correlate inversely with cardiovascular risk, but despite considerable improvements in our understanding of HDL cholesterol and its metabolism, none of the pharmacologic agents that raise HDL cholesterol that have been tested so far have had a significant effect on cardiovascular morbidity and mortality. Ongoing clinical trials of agents that raise HDL-cholesterol levels and that have other antiinflammatory and antithrombotic effects are currently under way.50

**GENOMICS, CELL-BASED THERAPIES, AND MOLECULAR TARGETING — THE NEXT FRONTIERS**

Several active areas of investigation hold promise for future advances in cardiovascular science and medicine, including genetics and genomics, molecular targeting, pharmacogenomics, and stem-cell biology and regenerative medicine.

Genetic investigations have led to discoveries of the heritable components of cardiovascular risk factors and coronary artery disease, including studies of families with inherited genetic mutations51 and genomewide association studies across populations.52 Multiple chromosomal loci associated with coronary artery disease harbor protein-coding genes known to be important in variations in lipid levels. In addition, associations of single-nucleotide polymorphisms with chemokines suggest that an inflammation pathway may regulate the process of coronary atherosclerosis.52 To date, the major contribution of these genomewide association studies has been new insights into biologic pathways that were often unsuspected and that underlie the development of cardiovascular disease. These insights have in turn led to hypothesis-driven research in which molecular, genetic, biochemical, and cellular techniques are used to investigate pathways. Knowledge of molecular pathways is essential to the development of therapeutics, defined conceptually as “molecular targeting.”

Pharmacogenomics applies our understanding of genetic variability in patients’ responsiveness to a drug in order to inform clinical decisions about dosing and selection. The anticoagulant warfarin is a case in point. Genetic variation in CYP2C9 and VKORC1, the two genes that encode the liver proteins required for warfarin metabolism, explains up to 40% of the differences observed among patients in their responses to the same dose of warfarin. The Food and Drug Administration has used this information to revise warfarin labeling in order to allow for genotype-specific dose ranges.53 In patients with gene variants in the cytochrome P-450 enzyme, CYP2C19, the antiplatelet drug clopidogrel is less efficacious and the risk of coronary artery disease is increased.54 Deep sequencing of the genes related to drug absorption, distribution, metabolism, and excretion may identify specific variants that contribute to the heterogeneity of patients’ responsiveness to cardiovascular drugs.

Cell-based therapies ranging from autologous noncardiac cells (e.g., bone marrow, skeletal muscle, fat, and endothelial progenitors) to allogeneic mesenchymal cells and putative resident cardiac progenitors have been studied in preclinical animal models and in early trials in humans, with mixed, yet promising, results.55-57 A subset of progenitors is mobilized in vivo by paracrine signals in cases of cardiac injury, suggesting that the delivery of such signals to the heart or vasculature may stimulate regenerative tissue.58

**GLOBAL CARDIOVASCULAR DISEASE**

Cardiovascular disease, including heart disease and stroke, is the leading cause of death worldwide, including low-income and middle-income countries.59 Several factors account for the increasing burden of cardiovascular diseases, including a longer average life span, tobacco use, decreased physical activity, and increased con-
assumption of unhealthful foods.\textsuperscript{60} New collaborations are under way to address cardiovascular and other noncommunicable diseases by building capacity in health care delivery, research, and training and developing low-cost interventions.\textsuperscript{3,62}

**CONCLUSIONS**

From John Warren’s description of angina pectoris in 1812 as a straining of the chest vaguely related to ossification of the coronary arteries to our current understanding of the genetic and molecular basis of coronary artery disease, the pathways of discovery, innovation, and therapeutic advancement in cardiovascular science and medicine over the past two centuries have been truly remarkable. We are now poised to take advantage of scientific opportunities, fueled by the results of rich epidemiologic studies of populations and large, randomized clinical trials evaluating science-based therapeutics, and thus further refine the cardiovascular care of patients around the globe.

Disclosure forms provided by the authors are available with the full text of this article at NEJM.org.

**REFERENCES**

2. Parry CH. An inquiry into the symptoms and causes of the syncope anginoso, commonly called angina pectoris. Bath, United Kingdom: Cruttwell, 1799.
28. Burkhardt JD, Wilkoff BL. Interven
tional electrophysiology and cardiac re-synchronization therapy: delivering elec
34. Furchtott RF, Zawadzki JV. The obliga
tory role of endothelial cells in the relaxa
35. Libby P, Ridker PM, Hansson GK. Progress and challenges in translating the


The Perpetual Challenge of Infectious Diseases

Anthony S. Fauci, M.D., and David M. Morens, M.D.

Among the many challenges to health, infectious diseases stand out for their ability to have a profound impact on the human species. Great pandemics and local epidemics alike have influenced the course of wars, determined the fates of nations and empires, and affected the progress of civilization, making infections compelling actors in the drama of human history. For 200 years, the Journal has captured the backdrop to this human drama in thousands of articles about infectious diseases and about biomedical research and public health efforts to understand, treat, control, and prevent them.

The Uniqueness of Infectious Diseases

Infections have distinct characteristics that, when considered together, set them apart from other diseases (Table 1). Paramount among these characteristics is their unpredictability and their potential for explosive global effect, as exemplified by the bubonic–pneumonic plague pandemic in the 14th century, the 1918 influenza pandemic, and the current pandemic of human immunodeficiency virus (HIV) infection and the acquired immunodeficiency syndrome (AIDS), among others.

Infectious diseases are usually acute and unambiguous in their nature. The onset of an infectious illness, unlike the onset of many other types of disease, in an otherwise healthy host can be abrupt and unmistakable. Moreover, in the absence of therapy, acute infectious diseases often pose an all-or-nothing situation, with the host either quickly dying or recovering spontaneously, and usually relatively promptly, often with lifelong immunity to the specific infecting pathogen.

Not only are some infectious diseases transmissible to others, a unique characteristic among human diseases, but their transmission mechanisms are relatively few (including inoculation and airborne and waterborne transmission), well understood, and comparatively easy to study, both experimentally and in the field. In addition, such transmission is generally amenable to medical and public health interventions. Unlike many chronic and lifestyle-associated diseases resulting from multiple, interacting risk cofactors, most infectious diseases are caused by a single agent, the identification of which typically points the way not only to general disease-control measures (e.g., sanitation, chemical disinfection, hand washing, or vector control) but also to specific medical measures (e.g., vaccination or antimicrobial treatment).

Given their nature, infectious diseases are potentially preventable with personal protection, general public health measures, or immunologic approaches such as vaccination. As preventive measures have become more effective and efficient, history has shown that certain infectious diseases, particularly those with a broad global health impact and for which there is no nonhuman host or major reservoir, can be eliminated. Such diseases include poliomyelitis, which has been eliminated in the Western Hemisphere, and smallpox, which has been eliminated globally.

Another unique aspect is that the extraordinary adaptability of infectious pathogens (i.e., their replicative and mutational capacities) provides them with a temporary evolutionary advantage against pressures aimed at their destruction. These pres-
Sures include environmental factors and antimicrobial drugs, as well as the human immune response. At the same time, such adaptations provide us with opportunities to respond with new vaccine antigens, such as annually updated influenza vaccines, or new or different anti-infective agents. This back-and-forth struggle between human ingenuity and microbial adaptation reflects a perpetual challenge.

Infectious diseases are closely dependent on the nature and complexity of human behavior, since they directly reflect who we are, what we do, and how we live and interact with other people, animals, and the environment. Infectious diseases are acquired specifically and directly as a result of our behaviors and lifestyles, from social gatherings, to travel and transportation, to sexual activity, to occupational exposures, to sports and recreational activities, to what we eat and drink, to our pets, to the environment — even to the way we care for the ill in hospitals and other health care environments. Moreover, microbial colonizing infections that lead to long-term carriage without disease (e.g., within the endogenous human microbiome) may influence the development of infections with exogenous microbes and also have an effect on general immunologic and physiologic homeostasis, including effects on nutritional status. Human microbiomes seem to reflect, and may even have helped to drive, human evolution.

In this struggle, infectious diseases are intimately and uniquely related to us through our immune systems. The human immune system, including the primitive innate system and the specific adaptive system, has evolved over millions of years from both invertebrate and vertebrate organisms, developing sophisticated defense mechanisms to protect the host from microbes. In effect, the human immune system evolved as a response to the challenge of invading pathogens. Thus, it is not by accident that the fields of microbiology and immunology arose and developed in close association long before they came to be considered distinct disciplines.

### Disease Emergence and Reemergence

Because infectious pathogens are evolutionarily dynamic, the list of diseases they cause is ever-changing and continually growing. Since newly emerging infectious agents do not arise spontaneously, they must recently have come from somewhere else, usually from animal infections, as occurred with HIV infection, influenza, and the severe acute respiratory syndrome. This interspecies transmission underscores the importance of interdigitating the study of human and animal diseases and recognizing the central role that microbial reservoirs, including those in animals, vectors, and the environment, play in human infectious diseases. Preexisting or established infectious diseases also may reemerge in different forms, as in extensively drug-resistant tuberculosis, or in different locations, as in West Nile virus infection in the United States, to cause new epidemics (Table 2). Indeed, many human infectious diseases seem to have patterns of evolution, sometimes played out over thousands of years, in which they first emerge and cause epidemics or pandemics, undergo periodic resurgences, and eventually become endemic with the potential for future outbreaks (Fig. 1).

### Table 1. Characteristics of Infectious Diseases That Set Them Apart from Other Human Diseases.

<table>
<thead>
<tr>
<th>Characteristic</th>
</tr>
</thead>
<tbody>
<tr>
<td>Potential for unpredictable and explosive global impact</td>
</tr>
<tr>
<td>Frequent acquisition by host of durable immunity against reinfection after recovery</td>
</tr>
<tr>
<td>Reliance of disease on a single agent without requirement for multiple cofactors</td>
</tr>
<tr>
<td>Transmissibility</td>
</tr>
<tr>
<td>Potential for becoming preventable</td>
</tr>
<tr>
<td>Potential for eradication</td>
</tr>
<tr>
<td>Evolutionary advantage over human host because of replicative and mutational capacities of pathogens that render them highly adaptable</td>
</tr>
<tr>
<td>Close dependence on the nature and complexity of human behavior</td>
</tr>
<tr>
<td>Frequent derivation from or coevolution in other animal species</td>
</tr>
<tr>
<td>Possibility of treatment for having multiplying effects on preventing infection in contacts and the community and on microbial and animal ecosystems</td>
</tr>
</tbody>
</table>

### Historical Perspectives and Current Status

Just over a decade before the publication of the first issue of the Journal, President George Washington died of an acute infectious disease believed to have been bacterial epiglottitis. Wash-
An interactive timeline on selected infectious diseases from 1812 to the present is available at NEJM.org

# Table 2. Broad Categories of Infectious Diseases.*

<table>
<thead>
<tr>
<th>Type of Disease</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Established infectious diseases</td>
<td>Endemic diseases that have been prevalent for a sufficient period of time to allow for a relatively stable and predictable level of morbidity and mortality (e.g., viral and bacterial respiratory and diarrheal diseases, drug-susceptible malaria and tuberculosis, tropical diseases such as helminthic and other parasitic diseases, nosocomial infections)</td>
</tr>
<tr>
<td>Newly emerging infectious diseases</td>
<td>Diseases that are recognized in the human host for the first time (e.g., HIV/AIDS, Nipah virus, severe acute respiratory syndrome)</td>
</tr>
<tr>
<td>Reemerging infectious diseases</td>
<td>Diseases that historically have infected humans but continue to reappear either in new locations (e.g., West Nile virus in the United States) or in resistant forms (e.g., influenza, methicillin-resistant Staphylococcus aureus, drug-resistant malaria) or reappear after apparent control or elimination (e.g., polio in parts of Africa, cholera in Haiti, dengue in Florida) or under unusual circumstances (e.g., deliberately released agents, including the anthrax release in 2001)</td>
</tr>
</tbody>
</table>

* Categories of infectious diseases include those that are newly emerging, those that have become established and may periodically reemerge, and those that have become stably endemic. Modern concepts that relate to emerging infections are more fully described in an influential 1992 report by the Institute of Medicine.

Washington’s life reflects the history of his era and provides both a window into infectious diseases two centuries ago and a benchmark for measuring our remarkable progress since then. Washington was born in 1732, just before the deadliest diphtheria epidemic on the North American continent. He was scarred by smallpox, survived multiple debilitating bouts of malaria, suffered wound infections and abscesses, nursed his brother on a tropical island as he died of tuberculosis, and even had an influenza pandemic named after him (the Washington influenza of 1789–1790). During his presidency, he stayed in the then-capital city of Philadelphia while most of the government fled during the nation’s deadliest yellow fever epidemic. At the time of Washington’s birth, there was no well-defined concept of infection or immunity, no vaccines, almost no specific or effective treatments for infectious diseases, and little idea that any treatment or public health measure could reliably control epidemic diseases.

During Washington’s lifetime, infectious diseases were the defining challenges of human existence. No one alive then could have imagined the astonishing breakthroughs that lay ahead. In this regard, it is noteworthy that almost all the major advances in understanding and controlling infectious diseases have occurred in the past two centuries (Table 3 and interactive timeline). Experimental animal-transmission studies that were conducted soon after the War of 1812 were followed by the development of better microscopes, which linked fungi to skin diseases and protozoa to mucosal diseases — for example, Alfred Donné’s 1836 work with Trichomonas vaginalis and David Gruby’s studies of Candida albicans in the early 1840s. The breakthroughs in the late 1800s, which taken together provided the compelling unifying principle of infectious diseases and must surely rank among the most important advances in the medical sciences, were the characterization of specific cultivatable microorganisms and proof of their association with specific diseases. This triumph was led by the work of Davaine and Koch in establishing anthrax as the first fully characterized infectious disease. This seminal process was facilitated by the development of defined criteria for establishing causality (Koch’s postulates).

Additional breakthroughs followed quickly, including the discovery and characterization of pathogen-specific immune responses; the demonstration that when inactivated by heat or chemicals or grown under limiting conditions that changed certain biologic properties (e.g., attenuation), organisms or their products could safely stimulate protective responses in a host; and development of anti-infective serums and chemicals to destroy pathogens. Over the next 135 years, a wide array of vaccines and antibiotics and, more recently, antiviral agents have saved hundreds of millions of lives, greatly extended the human life span, and reduced untold suffering. Undeniably, these countermeasures against infectious disease rank among the greatest achievements in public health and medicine.

History reminds us that new challenges in infectious diseases will continue to emerge and re-
emerge. We must be prompt in identifying them and devising new countermeasures. In this effort, we still follow the familiar pathway that was set down in the late 1800s for the identification and characterization, both clinical and epidemiologic, of the causative agent; the characterization of the human immune response to the pathogen; and the development of pathogen-specific diagnostic tests, treatment strategies, and public health prevention strategies such as vaccinations.

**Diagnosis and Characterization of Pathogens**

In the late 1800s, the realization that identifiable microbes caused specific diseases led to pathogen-specific medical diagnosis. Although the time-honored techniques of growing bacteria in broth or solid cultures and staining and examining them under microscopes are still important today, newer technologies have transformed the field of microbial diagnosis. Among the first emerging epidemic diseases to be identified by one such method was the hantavirus pulmonary syndrome, a centuries-old disease caused by an unknown phlebovirus (Sin Nombre) that was discovered unexpectedly in 1993 by the application of a then-novel molecular genetic technique, polymerase chain reaction (PCR). This followed quickly on the 1992 discovery of the previously unknown agent causing an infectious chronic condition, Whipple's disease. Less than a year later, PCR-related subtraction techniques solved a century-old mystery of the cause of Kaposi's sarcoma, human herpesvirus 8. Now, less than two decades later, sophisticated, high-throughput, rapid sequencing of the genomes of pathogens not only dramatically hastens initial identification but also detects individual genetic variants, facilitating identification of the genetic basis of drug resistance. Additional gene-based diagnostic tools include microchips and other technologies that detect short sequences of many different genes or their proteins, allowing simultaneous diagnosis or diagnostic elimination of multiple pathogens. New serologic techniques such as enzyme-linked immunosorbent assay can be many times more sensitive than traditional techniques in detecting and measuring antibodies to pathogens. Furthermore, monoclonal antibody techniques, which involve the use of cellular clones to produce antibodies against specific pathogen epitopes, have been adapted for the purposes of diagnosis, identification of the molecular structures of pathogens, elucidation of the natural history and pathogenesis of infectious diseases, development of conformationally accurate immunogens to be used as vaccine candidates, and even treatment. Many of these data-rich approaches require sophisticated bioinformatics systems (e.g., phylogenetic comparisons and genome construction analyses).

**Vaccine Development**

Vaccines against infectious diseases such as anthrax and rabies have been produced since the late 1870s. Only in the past half century, however, have technological advances in vaccination led to dramatic changes in the field of disease prevention. The World Health Organization now estimates that each year more than 120 different types of vaccines save 2.5 million lives and with optimal uptake could save an additional 2 million. Trivalent combined inactivated and live attenuated poliomyelitis vaccines were licensed in 1955 and 1962, respectively; a live attenuated trivalent vac-
<table>
<thead>
<tr>
<th>Disease</th>
<th>1812</th>
<th>1912</th>
<th>2012</th>
</tr>
</thead>
<tbody>
<tr>
<td>Malaria</td>
<td>Cause unknown; wealthy persons flee to higher elevations in malaria season; &quot;Jesuit’s bark&quot; (containing quinine) used to treat symptoms</td>
<td>Causative organism identified in 1880 (Laveran; Nobel Prize awarded 1907); anophelines mosquito identified as principle vector (Sir Ronald Ross; Nobel Prize awarded 1902); vector control attempts under way</td>
<td>Genomes of the host (human), two of the principal parasites (Plasmodium falciparum and vivax), and the principle vector (Anopheles gambiae) sequenced; drugs for treatment and prophylaxis; vaccines under development; successes with public health control, but malaria still causes &gt;800,000 annual deaths</td>
</tr>
<tr>
<td>Variola (smallpox)</td>
<td>Cause unknown; control begins in the developed world with Jenner’s 1798 publication on vaccination</td>
<td>Variola greatly controlled by vaccination in the developed world; developing world still has deadly epidemics</td>
<td>Variola eradicated in 1980 through aggressive global vaccination campaign</td>
</tr>
<tr>
<td>Plague</td>
<td>Cause and mode of transmission unknown; frightening disease for millennia; no good control measures; global quarantine systems not completely effective</td>
<td>1890s pandemic brings plague to the U.S. for the first time (1900); disease becomes enzootic and endemic, but fears of Black Death pandemic begin to subside</td>
<td>Plague a minor disease in U.S.; sporadic outbreaks still occur in the developing world, but fear of pandemics has subsided</td>
</tr>
<tr>
<td>Yellow fever</td>
<td>Cause and mode of transmission unknown; the “American Plague” is most frightening U.S. disease after deadly epidemics in 1793–1798, which led to forerunner of the U.S. Public Health Service (1798)</td>
<td>U.S. Public Health Service forerunner sets up Hygienic Laboratory (1887) to study the microbiology of infectious diseases, eventually becoming the National Institutes of Health; transmission by Aedes aegypti shown by Walter Reed team (1900); vector control efforts by Gorgas in Cuba and Panama soon lead to substantial control</td>
<td>Effective live attenuated vaccine developed in 1936; yellow fever is largely gone from the developed world and greatly reduced in the developing world</td>
</tr>
<tr>
<td>Tuberculosis</td>
<td>Cause unknown; consumption an old and feared disease; on the rise in the industrial age</td>
<td>Recognized as the deadliest infectious disease of the 19th century; organism discovered by Robert Koch in 1882; beginning to be controlled by public health measures and sanatorium movement</td>
<td>Bacille Calmette–Guérin vaccine (marginally effective against transmission of pulmonary tuberculosis) in 1921; antituberculosis chemotherapy initiated in 1950s; control in the developed world upset by HIV pandemic and in the developing world is never achieved; 1.3 million still die annually; emergence of multidrug-resistant and extremely drug-resistant tuberculosis; better vaccines actively pursued</td>
</tr>
<tr>
<td>Wound infections and puerperal fever</td>
<td>Causes unknown; amputations without anesthesia are often ineffective; maternal postpartum deaths common</td>
<td>General anesthesia introduced in 1840s; puerperal streptococcal infections controlled by hand washing in mid-19th century; aseptic technique introduced in 1867; causative organisms isolated in late 1900s</td>
<td>Antibiotics, first used in the late 1930s, are common by the early 1950s in the developed world</td>
</tr>
<tr>
<td>Diphtheria</td>
<td>“Throat distemper” (caused by diphtheria and streptococci, sometimes in combination) is a major cause of childhood deaths</td>
<td>Causative organism discovered in 1884; diphtheria antitoxin (1894) is the first passive immunotherapy</td>
<td>Vaccine produced in 1913; disease controlled in developed world</td>
</tr>
<tr>
<td>Poliomyelitis</td>
<td>Not well described but probably endemic</td>
<td>First U.S. epidemic in Vermont in 1894; recurring epidemics cause fear in the U.S. during next 60 years</td>
<td>Inactivated Salk and live Sabin vaccines introduced in 1955 and 1962, respectively; effective global control has led to near eradication</td>
</tr>
<tr>
<td>HIV/AIDS</td>
<td>Unknown</td>
<td>May have emerged in Africa but was not recognized</td>
<td>First reported in 1981; causative virus identified in 1983–1984 by Luc Montagnier and Robert Gallo; combination antiretroviral therapy greatly prolongs the lives of infected patients</td>
</tr>
</tbody>
</table>
Effective live attenuated vaccine developed in 1931;
genomes of the host (human), two of the principal Plasmodium falciparum.

Antibiotics, first used in the late 1930s, are commonly by the early 1950s in the developed world.

Anopheles gambiae. Variola eradicated in 1980 through aggressive Bacille Calmette–Guérin vaccine (marginally effective against transmission of pulmonary tuberculosis) in 1921; antituberculosis chemotherapy initiated in 1950s; control...

Plague a minor disease in U.S.; sporadic outbreaks still occur in the developing world, but fears of pandemics have subsided.

Nobel Prize awarded 1907); anopheles mosquito identified as the principal vector (Sir Ronald Ross; Nobel Prize awarded 1902); vector-control attempts under way.

Hygienic Laboratory (1887) to study the microbiology of infectious diseases, eventually becoming the National Institutes of Health; transmission by...

Vaccine produced in 1913; disease controlled.

Tuberculosis Cause unknown; consumption an old and feared (containing quinine) used to treat symptoms. Caused organism discovered in 1884; diphtheria.....

Yellow fever Cause and mode of transmission unknown; the anti-toxin (1894) is the first passive immunotherapy.

Inactivated Salk and live Sabin vaccines introduced.

"American Plague" is the most frightening U.S. disease after deadly epidemics in 1793–1798, which led to forerunner of the U.S. Public Health Service (1798).

Streptococci, sometimes in combination, is a major cause of childhood deaths.

HIV/AIDS Unknown May have emerged in Africa but was not recognized First reported in 1981; causative virus identified.

Vaccines against three unrelated diseases (measles, mumps, and rubella) was licensed in 1971; and a variety of vaccine approaches and platforms have been introduced since then. It is now possible to determine high-resolution crystallographic structures of pathogens and use this information to design vaccines directed at the most relevant epitopes in the microbe's complex structure, an approach known as structure-based vaccine design. 57

**TREATMENT**

Successful treatment with pathogen-immune serum was another critical breakthrough of the late 19th century. 55 This approach to therapy also encouraged scientists to develop chemicals to kill the specific pathogens that they were regularly identifying. Ehrlich succeeded first in 1910 with his magic bullet against syphilis (arsphenamine, or salvarsan) 58. Within two decades, a new generation of scientists was working on what eventually be called antibiotics. As a result of these efforts, sulfa drugs were developed in 1936, and penicillin in 1943. 59,60 In the United States, tuberculosis had been only partially controlled by public health measures and incompletely effective vaccines. 61

It was not until the introduction of specific antituberculosis therapy in the 1950s 62 that sanatoriums were emptied and cases of active disease were substantially reduced. Antibiotics have revolutionized the treatment of many other important bacterial infections and have saved many millions of lives since their introduction.

When antiviral drugs were first developed in the 1960s, they did not seem to be particularly promising, with a few exceptions. In response to the HIV/AIDS pandemic, however, the development of antiretroviral drugs markedly expanded the arsenal of available antiviral agents and invigorated the research-and-development pathway for these important drugs. Effective combinations of powerful antiretroviral drugs have led to substantial prolongation of the lives of millions of persons with previously almost invariably fatal HIV infection, a true landmark in therapies for infectious diseases. 15,63

All antibiotic and antiviral drugs, however, share an inherent weakness: the organisms against which they are directed almost invariably evolve mechanisms of resistance. Bacteria become resistant by a variety of mechanisms. 64 The evolution of antimicrobial resistance is enhanced by overuse of antibiotics in animals and by inappropriate use in humans. Many viruses, particularly RNA viruses such as influenza virus, rapidly develop mutations even in a single brief replication cycle. A number of approaches have been pursued to meet the ever-present challenge of antimicrobial resistance. The development of new classes of antibiotic, antiviral, and antiparasitic agents aimed at diverse microbial targets, often with the use of high-throughput screening of compounds. 65 is strengthening and broadening the therapeutic armamentarium. In addition, combination therapies (e.g., antiretroviral agents for HIV infection and multidrug approaches to tuberculosis) have proved to be successful in slowing the emergence of resistance.

**PUBLIC HEALTH ACHIEVEMENTS**

Breakthroughs in the field of infectious diseases have had far-reaching effects, including the realization of the critical importance of clean water and basic sanitation and hygiene for the prevention of a great number of infectious diseases. In addition, disease-specific approaches to prevention and treatment have led in many cases to the widespread control of diseases that historically have caused substantial morbidity and mortality.

The treatment of infectious diseases is in itself a prevention measure, limiting or preventing transmission to others. Eradication, the ultimate goal in facing the threat of an established or emerging infectious disease, is no longer unrealistic. Specifically, in addition to the millions of lives saved by vaccines and antibiotics, certain infectious diseases have been eliminated from large regions of the world or even completely eradicated, an accomplishment rarely, if ever, seen in other medical disciplines. In 1980, smallpox became the first eradicated disease, 9 making this among the most momentous achievements in human disease control. In May 2011, the veterinary morbillivirus disease rinderpest was declared eradicated, and its presumed descendant, human measles virus, is now being targeted for eradication. 67 Poliomyelitis has been eliminated from several regions of the world, and it is hoped that within a reasonable period, it will be eradicated globally. 68 Dracunculiasis (guinea worm disease) is also almost completely eradicated. 69 These are just a few examples of what has been and can be accomplished by aggressive and concerted public
health measures using the tools provided by basic and clinical research.

NEW VISTAS

An unanticipated outcome of the explosion of information concerning the microbial world is the recognition that a growing number of chronic diseases that were once attributed to host, environmental, or lifestyle factors or to unknown causes are actually directly or indirectly caused by infectious agents that potentially can be controlled through prevention and treatment. For example, liver cancer and cirrhosis are complications of hepatitis B and C infections. Cervical cancer is a complication of human papillomavirus (HPV) infection, and gastric and duodenal ulcers may result from Helicobacter pylori infection.70-72 Vaccines against two of these agents, hepatitis B and HPV, are already in use, exemplifying the concept of cancer-preventing vaccines. H. pylori infection can be cured with antibiotics, and chronic hepatitis B and C infections are being treated by means of antiviral regimens with growing success rates. Certain autoimmune conditions have also been attributed to infections. For example, enteric microbes have been associated with inflammatory arthritides, and Campylobacter jejuni and certain viruses have been associated with the Guillain–Barré syndrome.73 In addition, with new technologies and approaches, scientists are exploring new facets of microbiology, including the role of the human microbiome in maintaining homeostasis in the ecosystems of our bodies and its possible relationship to conditions such as obesity and inflammatory bowel disease.74

REFERENCES


THE PERPETUAL CHALLENGE

We are living in a remarkable era. Almost all the major advances in understanding and controlling infectious diseases have occurred during the past two centuries, and momentous successes continue to accrue. These breakthroughs in the prevention, treatment, control, eradication, and potential eradication of infectious diseases are among the most important advances in the history of medicine. Nevertheless, because of the evolutionary capacity of infectious pathogens to adapt to new ecologic niches created by human endeavor, as well as to pressures directed at their elimination, we will always confront new or reemerging infectious threats. Our successes in meeting these threats have come not just from isolated scientific triumphs but also from broad approaches that complement the battle against infectious diseases on many different fronts, including constant surveillance of the microbial landscape, clinical and public health efforts, and efficient translation of new discoveries into disease-control applications. These efforts are driven by the necessity of expecting the unexpected and being prepared to respond when the unexpected occurs. It is a battle that has been well fought for more than two centuries but that will almost certainly still be raging, in now-unimagined forms, two centuries from now. The challenges are truly perpetual. Our response to these challenges must be perpetual as well.

Disclosure forms provided by the authors are available with the full text of this article at NEJM.org.
The relationship between patients and doctors is at the core of medical ethics, serving as an anchor for many of the most important debates in the field. Over the past several decades, this relationship has evolved along three interrelated axes — as it is defined in clinical care, research, and society. Many of the pivotal discussions of these issues have appeared in the pages of the Journal (see box).

**Clinical Care**
The relationship between patients and doctors in the clinical realm has historically been framed in terms of benevolent paternalism. Until about 1960, most codes of medical ethics relied heavily on the Hippocratic tradition, framing the obligations of physicians solely in terms of promoting the welfare of the patient, while remaining silent about patients’ rights. The past several decades have seen tectonic societal shifts that have resulted in increasing empowerment of individuals against the authority of government and institutions, creating a surge of rights-based movements, with patients’ rights emerging as a societal demand alongside women’s rights, minority groups’ rights, consumers’ rights, and others. This dramatic shift appeared to move the locus of authority in decision making from the physician to the patient. And indeed the emergence of the Internet, with its myriad health-related websites and other sources of medical information, has given many patients the impression that they can largely manage their own medical affairs, with physicians serving primarily as consultants. But the reality is more complex: the wealth of information available to patients has proved to be as dangerous as it is helpful, and today patients and physicians are beginning to find a healthier balance of power through a process of shared decision making. With this approach, physicians are seen as having expertise and authority over matters of medical science, whereas patients hold sway over questions of values or preferences. This division of labor reflects a recognition of the naturalistic fallacy, the erroneous notion that one can derive ethical conclusions from scientific facts; in truth, an “ought” cannot be deduced from an “is.” Although physicians may be experts on the medical facts of a patient’s condition, those facts are never sufficient to specify a course of treatment; clinical decisions must always include consideration of the values and preferences of the patient. This approach has many implications — for example, in recognizing the right of a competent adult to refuse a lifesaving blood transfusion on the basis of his or her religious beliefs, or the right of a patient...
to refuse mechanical ventilation for a treatable and reversible cause of respiratory failure. Stated succinctly, today we acknowledge that competent patients have a virtually unlimited right to refuse unwanted medical care, even when physicians correctly claim that it would be medically effective and indeed lifesaving. More important, however, is the way that physicians present patients with what they see as reasonable medical options and then help them to incorporate personal values and preferences to arrive at decisions that make the most sense for them in terms of both the medical facts and their unique personal perspective. This approach to engaging patients has other benefits as well, such as promoting their sense of self-efficacy and improving their adherence to treatment recommendations.

Despite this vigorous consensus about the rights of patients to refuse unwanted care, equally strong disagreement persists about patients’ rights to demand care that physicians regard as medically inappropriate. Consider situations in which the data clearly show that the likelihood of successful resuscitation would be less than 1%. Does this fact allow us to conclude that resuscitation should not be performed, regardless of the patient’s preference? Or is the threshold for what should count as a reasonable likelihood of success fundamentally a question of value that should be defined by the patient? When the medical evidence overwhelmingly suggests that a patient is permanently unconscious, does this imply that further medical treatment is inappropriate and should not be provided (as stated in the guidelines of many professional societies), or is what counts as a “life worth living” a personal choice that should be respected?

In recent years, the Texas Advance Directives Act has defined one very concrete approach to addressing these dilemmas. When families demand treatments that have an exceedingly low likelihood of success or that sustain life of such low quality that one might reasonably say it is of no benefit to the patient, Texas law allows physicians to refuse to provide such treatments. Under the Texas legislation, demands by families for treatments that appear to meet these criteria are adjudicated by a hospital-based committee, and if the committee agrees with the
clinicians, and if other providers cannot be located who are willing to provide such care, then treatment may be withdrawn without the permission of the patient's surrogate. Although Texas has the most experience with this approach, other states are showing interest in similar proposals that address both the financial implications of providing allegedly inappropriate care and the concerns of clinicians who must endure the moral burdens and burnout associated with being compelled to provide treatments they believe are ethically wrong.

CLINICAL RESEARCH
The Nuremberg trials in 1946 marked the beginning of modern discussion of the ethics of clinical research. Although Nuremberg showed how physicians could be led astray by a corrupt political regime, it was not until Henry Beecher’s alarming exposé in 1966 that U.S. physicians had to confront the fact that blatantly unethical research — such as injecting patients with malignant cells without their knowledge or permission — was prevalent even on the wards of prestigious academic medical centers in the United States. These and other revelations led to the development of the principles outlined in the Belmont Report, federal regulations governing the conduct of clinical research, and the creation of institutional review boards charged with applying these guidelines to individual research protocols.

Despite the strong safeguards that are now in place to ensure that patients are fully informed and provide their consent before being enrolled in research trials, important tensions remain between the ethical obligations of the physician–patient relationship and those of the researcher–subject relationship. Both physicians and patients have a psychological tendency to minimize these tensions; neither wants to recognize the important ethical conflicts that may exist between clinical care and research. This phenomenon, known as the therapeutic misconception, has been shown by Paul Appelbaum and others to be ubiquitous among both researchers and research subjects, manifesting as the false and often implicit belief that the primary aim of research is to benefit the patient. Although patients often do benefit from their involvement in research, nearly all clinical research includes procedures that carry risks to subjects that are not compensated for by corresponding benefits. By definition, research protocols are designed to answer scientific questions, not to optimize the medical care of the patient.

The concept of clinical equipoise was developed with the promise of easing the ethical tension between clinical care and research. Clinical equipoise exists when there is genuine uncertainty in the medical community about which of two treatments is better in a given situation. When this condition is satisfied, a researcher enrolling his or her patients in a randomized trial can honestly say to the patient, “Although I may personally prefer treatment X, if you were being seen by another, equally competent physician, you could be given treatment Y. Hence, your medical care will not be compromised if you agree to enroll in this trial and have your treatment determined by chance.” Under clinical equipoise, there appears to be no conflict between the ethics of clinical care and those of research.

Over the past several years, however, clinical equipoise and other attempts to harmonize the ethics of clinical care and research have become less tenable, particularly in the context of placebo-controlled trials. Consider, for example, placebo-controlled trials of new antidepressants. Given that some antidepressants are known to be efficacious, clinical equipoise cannot be used to justify such trials, since a physician could never defend prescribing a placebo for a patient outside of the trial.

These problems extend beyond placebo-controlled trials, however. For example, oncologists have recently been engaged in an anguish debate about the ethics of a randomized, controlled trial of a new drug for metastatic melanoma in the face of impressive preliminary evidence supporting the efficacy of the new drug, in combination with the known dismal prognosis associated with the standard therapy given to patients in the control group. Trials like this one strain the concept of clinical equipoise beyond the breaking point: most oncologists would agree that these trials must be performed, but few would say that the physician researchers believe the treatments are in equipoise.

Given the scientific and ethical rationale for these trials and the failure of the clinical equipoise paradigm to provide justification for them, over the past several years a new way of looking at the ethics of clinical research has developed — one that regards the ethical principles governing clinical care and research as fundamentally distinct and indeed often in tension. If physicians forthrightly inform patients that the goal of clinical research is not primarily to optimize their clinical care but to advance knowledge for the benefit of future pa-
Patients, both patients and physicians may be guided to a more transparent view of research that is not distorted by the therapeutic misconception.

Separating the ethics of research from clinical care has other advantages. With this approach, patients become more active participants in the research enterprise. They expect to be informed about research results and engaged in setting research priorities. Parallel to the evolution of the relationship between physicians and patients in clinical care, this new vision of the ethics of clinical research moves research subjects out from under the paternalistic umbrella of clinical investigators and empowers them to have a more equal and active role in the process of advancing medical knowledge.

**Populations and Health Care Systems**

One of the most revered principles in medical ethics has been that physicians should be exclusively devoted to the best interests of their patients. As Norman Levinsky put it, “Physicians are required to do everything that they believe may benefit each patient without regard to costs or other societal considerations. In caring for an individual patient, the doctor must act solely as that patient’s advocate, against the apparent interests of society as a whole, if necessary.” In reality, this has never been more than a lofty ideal, since physicians have always had competing pressures that prevent them from providing everything that might be of medical benefit to a patient, beginning with the fact that no physician can spend an unlimited amount of time with any one patient.

But this ideal is itself coming under increasing scrutiny, as both physicians and society come to recognize that the benefits of a solely patient-centered focus to care must be balanced against the value of offering entire populations of patients equitable access to necessary health care.

The issue first arose in U.S. politics in 1962, when dialysis became available for only a limited number of patients and the public was horrified to learn that a small committee of anonymous citizens was tasked with deciding who should live and who should die. Rather than face this difficult problem head on, Congress eventually passed legislation mandating that renal-replacement therapies be fully funded by the government, a decision that stands to this day. Clearly, however, taking this approach to solving every rationing decision would lead to financial disaster.

More recently, the patient-versus-population dilemma played out over new recommendations for screening mammography. Although analysis shows that current screening practices are exceptionally cost-inefficient, there is no doubt that they have saved the lives of many women. Yet given the seeming impossibility of having a public debate about when it may be permissible to forgo some beneficial care that is very expensive in favor of providing other benefits that are a much better value, those who supported the new screening recommendations were forced to justify them solely in terms of what would be best for an individual woman, regardless of cost. Not surprisingly, by acceding to the taboo that renders open discussion of cost-effectiveness off limits, proponents of the new recommendations appear to have lost the debate.

The United States already spends more on health care than any other country on earth yet does not attain the health benefits that are achieved in many countries with much more limited resources. By refusing to bring the issues of cost-effectiveness and rationing into the political discourse, we allow the myth to persist that there is some yet-to-be-discovered alternative to a thoughtful and systematic approach to the allocation of resources. Despite Levinsky’s seductive view that the relationship between physicians and patients should be isolated from any external pressures, we must recognize that population-based factors such as justice, efficiency, and fairness are also ethically relevant. Overcoming our inability to muster the political will and courage to acknowledge the necessity of rationing and to grapple with the best way to use the tremendous resources currently being devoted to health care is likely to be the greatest challenge in the evolving relationship between physicians and patients in the decades to come.

Although the relationship between patients and doctors is often idealized in terms of universal and timeless principles, it has not been immune from the larger social and cultural forces surrounding it. This relationship has been profoundly shaped by the human rights movement of the past several decades, and clinical care today is guided by norms of shared decision making rather than benevolent paternalism. Clinical research is no longer regarded as a side benefit of providing patients with clinical care, but rather as a compatible but distinct activity that requires us to view patients as partners in the process of advancing medical knowledge. And finally, the greatest challenge still lies largely be-
fore us, as we will struggle in the years to come to balance the personal advocacy that all patients rightfully expect from their physicians with the equally compelling obligation of physicians to see that health care resources are used wisely in ways that are efficient and fair.

Disclosure forms provided by the author are available with the full text of this article at NEJM.org.

From the Departments of Anesthesia and of Global Health and Social Medicine, Harvard Medical School; and the Division of Critical Care Medicine, Children’s Hospital Boston — both in Boston.


Copyright © 2012 Massachusetts Medical Society.
A Patient with Asthma Seeks Medical Advice in 1828, 1928, and 2012

Erika von Mutius, M.D., and Jeffrey M. Drazen, M.D.

PEOPLE HAVE SUFFERED FROM ASTHMA FOR MILLENNIA. ALTHOUGH THE clinical presentation of asthma has probably changed little, there are many more people who now bear its consequences than there were 200 years ago. As a result of an intense interest in the condition, our understanding of its pathobiology, how to diagnose it, and — most important — how to treat it has evolved dramatically over the past two centuries. To illustrate this change, we provide three fictional reports of consultations performed for essentially the same patient, who has what we in 2012 would refer to as asthma. (A timeline of the major advances in the treatment of asthma from 1812 through 2012 is available with the full text of this article at NEJM.org.)

The first report is from 1828, the year that the New England Journal of Medicine and Surgery and Collateral Branches of Science joined with the Medical Intelligencer to form the Boston Medical and Surgical Journal. The second is from 1928 when the title of the publication was changed to the New England Journal of Medicine, and the third report is from the present.

The three accounts reflect the way in which care was delivered at the time. The first account is in the voice of a general practitioner who was contacted for consultation about a woman with intermittent episodes of dyspnea. The second is in the voice of a generalist who works in a private practice and has an interest in asthma; the patient has been referred to this physician by her own general physician. The third account is in the voice of a sub-specialty physician whose practice is limited to the care of patients with asthma. The contemporary patient identified this physician as a specialist in asthma through an Internet search and is consulting him for a second opinion about the appropriateness of her asthma care. She brings to the consultation a detailed history that she wrote, as well as notes from her primary care physician and an allergist.

Our three views of this medical consultation for a patient with asthma are not meant to provide a history of asthma but rather to offer a set of snapshots of the care that the same patient might have received had she sought medical advice in these distinct epochs. There are many diagnostic and therapeutic techniques that we do not mention; this does not mean that they are not important; it simply means that their use does not fit the time frame of our fictitious consultations. Finally, since this article is meant to contribute to the celebration of the Journal’s 200th anniversary, we have largely, but not exclusively, used literature from the Journal; our apologies to others who claim primacy.

OFFICE NOTE ON MRS. A. SMITH

I attended at the home of a woman aged 35 years who had just moved with her family to Boston. Her household includes herself and her husband of 17 years, four chil-
children, a cook, two maids, a stable boy, and a footman. She sent for me with a complaint of repeated shortness of breath.

THE HISTORY OF HER ILLNESS

When a fit of dyspnea occurs, the patient hears a musical noise in her chest, and she must labor to draw and expel a full breath. When she is stricken, it is her custom to stop all her activities and to inhale the steam coming from the spout of a kettle that her cook keeps always at the ready. With such treatment, she usually recovers within one or two days. Once or twice a year she has a severe fit, which may last for a week, and she is confined to her sick bed. She has suffered such fits of laborious breathing since her childhood. They occur at any time of the year but are more common in the spring, when the trees bloom, and in the late summer than at other times. In the winter she reports that it is common for her to be so stricken when she walks from the harbor to her home, a distance of nearly a mile along a path that ascends steeply. This difficulty of respiration has become such a frequent occurrence that she now routinely calls for her coach even for very short journeys outside her home. During each of her periods of confinement for childbearing, the fits were far less numerous and severe in character, but within a few months after she had given birth, they returned.

Often, even when she is not suffering from laborious breathing, she will arise in the dark of the night and stand at the open window, gasping for air. By the time that dawn arrives she has usually regained control of her breathing and returns to sleep.

Her difficulty of respiration is accompanied by itchy eyes and a runny nose. She has a cough with these fits, but she does not produce phlegm. She does not have hemoptysis. No one among her family or close acquaintances has died from consumption. Her weight has been stable, and when she is not suffering from laborious breathing, her strength is good. She has not had rheumatic fever.

Her mother, now deceased, also suffered from difficulty of respiration; her father did not. Of her four children, ages 14, 12, 9, and 7, her two eldest, both boys, have suffered from the same symptoms, although her oldest son has not had a fit of laborious breathing for more than a year.

MY EXAMINATION

Observation of her breathing on the occasion of my consultation revealed nothing far out of the ordinary. Her speech was full and normal. The movements of her chest were full. I could palpate nothing abnormal in her heart motion. There was no swelling of her liver or her legs. I used a newly acquired stethoscope to examine her chest. Although the patient could not hear the musical sounds that have been termed “wheezes,” I was able to hear them.

MY OPINION

The patient clearly suffers from an asthma; she may also have what has been described as “hay fever” in the spring and fall. I believe her fits of laborious breathing are similar to the asthmatic fits that Sir John Floyer suffered from and described in his “Treatise of the Asthma.” He describes this type of asthma as follows: “[T]he expiration is very slow and leisurely and wheezing, and the asthmatic can neither cough, sneeze, spit nor speak freely; and in the asthmatic fit, the muscular fibres of the bronchia and vesiculae of the lungs are contracted, and that produces the wheezing noise which is most often observable in expiration.” I have no concern that she suffers from consumption or from conditions of the heart that may lead to dropsy.

I think that she may benefit from smoking the leaf of \textit{Datura stramonium}, also known as the thorn-apple plant. Many asthma sufferers have tried this remedy, and it seems to provide relief from a fit, even though it will not prevent a recurrence. Several years ago, Dr. Bree reported in the \textit{New England Journal of Medicine and Surgery} that such smoking had a deleterious effect on a number of patients suffering from difficulty of respiration. However, in my experience, patients such as this woman will derive benefit from such treatment in that it shortens the duration of their indisposition from an asthmatic fit. I recommended this treatment to my patient, and she tells me that she has benefited from it.

Comment: In the early 1800s, there were many “asthmatics,” since this was the term for any episodic shortness of breath. The physician needed to be sure that the primary cause was not tuberculosis or cardiac disease (e.g., mitral stenosis); both were very common at the time. Once a diagnosis of asthma (as we know it now) was established, the number of effective treatments was quite limited; inhalation of smoke from burning \textit{Datura stramonium}
was probably the best. This agent had anticholinergic properties and was the forerunner of the currently used antimuscarinic agents, such as ipratropium and tiotropium.5 There were numerous other treatments, such as inhalation of the fumes of hydrocyanic acid6 or inflation of the lungs with a bellows.7 Fortunately, such treatments and many others that produced no benefit and probably caused harm are no longer used.

**1928**

**LETTER REGARDING MRS. A. SMITH**

Dear Dr. Jones,

Thank you for referring your patient, Mrs. A. Smith, for evaluation concerning a possible diagnosis of asthma. I found the patient’s history, as recounted in your office notes, to be complete and accurate.

**HISTORY**

The critical feature of her case is that Mrs. Smith, age 35, has been having “asthma attacks” since her early childhood. Her attacks are characterized by the relatively sudden onset of dyspnea; they are more frequent in the spring and fall, when they are often preceded by symptoms of rhino-conjunctivitis. If untreated, an attack will last for a few days, but if she is treated with a subcutaneous injection of adrenaline, as you have administered at your office, she often has relief from acute symptoms, and the attack may or may not recur. Recently, her attacks have been more frequent, and she does not feel that her breathing is improved to the point where she can carry out her responsibilities as a wife and mother.

Her mother carried a diagnosis of asthma, as do two of her children. She is currently not using any medications.

**PHYSICAL EXAMINATION**

Her physical examination, at a time when she was not having acute asthmatic symptoms, showed normal body temperature, blood pressure, and pulse. She had no rashes. Her nasal passages were closely examined and showed inflammation and edema but no polyps. Her respirations were 24 and slightly labored. She had diminished tactile fremitus. Expiratory wheezing of modest profusion was audible in all lung fields. Her cardiac examination was normal. There was no clubbing, cyanosis, or edema.

**LABORATORY STUDIES**

I examined the radiograph of the chest that she brought with her, which was taken within the last month. It showed hyperinflation of the lungs, but there were no abnormal shadows; there were no findings that would suggest tuberculosis. Her cardiac silhouette did not show any abnormalities.

A blood smear was made, showing 14 per cent eosinophils; in a normal person this is most often less than 5 per cent. A sputum sample was also examined, and all the polymorphonuclear leukocytes observed were eosinophils. Specialized skin testing was performed. She had positive reactions to extracts of ragweed and horse dander.

**MY OPINION**

Your diagnosis of asthma is correct. The episodes are characteristic, and there is no other likely cause suggested by her medical history or the physical examination and laboratory findings. In fact, the presence of eosinophils in the blood and sputum makes the diagnosis virtually certain. The positive skin tests make this case one of extrinsic asthma. Hypersensitivity to proteins is the likely physiological basis of asthma, although the exact mechanisms leading to sensitization are not clear.

Treatment is difficult. Your use of adrenaline injections for acute attacks is appropriate; there is reason to believe that treatment with oral ephedrine may also help with her asthmatic episodes. The relief is of longer duration than with injected adrenaline and the patient can administer it herself. Ephedrine is not a substitute for injections of adrenaline when the patient is in extremis.

The critical factor in treatment is removing the patient from exposure to the proteins to which she is sensitive. Her positive skin test to ragweed pollen extract is in agreement with the clinical history of worsening disease in the autumn. However, there may be proteins to which she is allergic that were not included in our skin test panel. In my experience, removing a protein from a patient’s exposure is very hard to accomplish. One strategy, which I am loath to suggest unless there is no other hope, is a move to a climate where there are fewer proteins in the air to which the patient would be exposed.

**Comment:** By 1928, the differential diagnosis of asthma was well established, and diagnostic techniques were
available that made it possible to be reasonably certain that patients did not have heart disease or pneumonia when they were labeled as asthmatic. Physicians of the time often used the term “asthma” to refer to episodic dyspnea, but qualifiers such as “cardiac” were used. By 1928, eosinophils in the blood and sputum were known to be characteristic of asthma. Skin tests for allergies had been developed and were used clinically to help clinicians identify specific offending environmental proteins. The issues that plague us today — allergies to multiple allergens and difficulty in interpreting skin tests — were of concern to physicians in 1928.

Figure 1. Asthma Cigarettes.
Asthma cigarettes made from the leaves of Datura stramonium (thorn apple) were widely sold in the 1800s and into the early 1900s. These cigarettes provided a means of delivering an inhaled treatment; we now know that the active component of this smoke was antimuscarinic alkaloid. Antimuscarinic treatment of asthma has recently been studied with the use of chemically synthesized moieties, such as tiotropium bromide and ipratropium bromide. Images courtesy of Mark Sanders, www.inhalatorium.com.

E-mail Message to Ms. Smith
Dear Ms. Smith,
Thank you for asking me to provide you with a direct personal consultation concerning your asthma and your asthma care. I will summarize the salient facts from the detailed written history and physician’s note you kindly provided.

As pointed out in your written history, you have had asthma since childhood. Among your earliest recollections is receiving injection treatments and later inhalation treatments for asthma in an emergency room. In your early teenage years you started treatment with inhaled Vanceril (beclomethasone), two puffs twice a day; 10 years ago, you switched to inhaled Qvar (beclomethasone driven by a hydrofluoroalkane [an ozone-layer–friendly] propellant), and Singulair (montelukast) was added to your regimen. Over the past 10 years, you have tried two different “combination inhalers,” containing both inhaled glucocorticoids and long-acting $\beta_2$-agonists — namely, Advair (fluticasone propionate and salmeterol) and Symbicort
(budesonide and formoterol fumarate dehydrate). These medications did not improve your symptoms or lung function as compared with inhaled beclomethasone alone, and you switched back to Qvar.

Even with this regimen, however, your asthma symptoms are still present and bothersome. For example, two to three times a month you are awakened from your sleep between 3 a.m. and 4 a.m. by shortness of breath and cough; you can hear yourself wheeze. If you use your rescue albuterol inhaler, you are usually able to get back to sleep by 5 a.m.

Two years ago, skin tests were performed, and your total IgE level was measured. Your only positive skin tests were for house-dust mites and ragweed. Your total IgE level was 75 IU per milliliter. The allergist who did the testing suggested that you add a nonsedating antihistamine, such as loratadine, to your treatment during the times of year when you are most susceptible to symptoms; the loratadine was of some small help in controlling your runny nose, but there was no change in your asthma symptoms. Your allergist also referred you to a gastroenterologist, who performed 24-hour esophageal pH monitoring and found no abnormalities.

In the past decade, you have required treatment with oral prednisone on three occasions; the last instance was in 2009. Each of these exacerbations occurred during your allergy season. You have a peak-flow meter, which you use occasionally. Your best reading is 500 liters per minute; on most days, your peak-flow values are between 350 and 400 liters per minute.

You work in an office. You live with your husband and two children in a single-family home heated and air-conditioned with forced air. You
have taken extensive measures to remove allergens from your home, including having the air ducts cleaned and tested for allergens. You have no pets. You have never smoked, and the same is true for your husband and your children. Smoking has not been allowed in your workplace for more than a decade. Your mother had asthma.

Your current medications are Qvar, 80 μg per puff, two puffs twice a day; Singulair, 10 mg per day, taken at night; and one multivitamin per day.

You would like to see a single consultation and confidential second opinion as to how your asthma has been managed and how to improve your asthma control.

On physical examination today, you looked well. Your weight was 135 lb [61.2 kg]. Your blood pressure was 110/75 mm Hg, and your pulse was 77 beats per minute according to the pulse oximeter, which also indicated that your hemoglobin saturation while you were breathing ambient air was 95%. Your physical examination was largely normal. No abnormalities were noted in your eyes, nose, or ears. Your chest examination was normal except for the presence of scattered expiratory wheezes, which were heard best during rapid, shallow breathing. There were no abnormalities in your extremities. Your neurologic examination was normal as well. Lung-function testing was performed in our laboratory; the results are attached to this letter (Fig. 3).

I think that the diagnosis of asthma is well established. You have a long history of asthma and have had salutary symptomatic responses to asthma treatments, your lung-function tests still show reversibility of airway obstruction of more than 15% with albuterol, and no other competing diagnosis has emerged over many years. The major issue now is to determine whether there are additional treatments that could help suppress your asthmatic symptoms without increasing the treatment burden.

You and your physicians have done an excellent job of managing your asthma. The treatments you are using now are well established and known to be effective. There are three treatments that could be added to your regimen, but it is difficult to be certain that they would be effective. First, oral theophylline could be added to your regimen. Although you cannot recall having received treatment with theophylline, given your age and asthma history, it is likely that you were treated with this agent as a child. This therapy could be of value, but it is necessary to monitor blood levels of the drug to obtain an optimum response, and some patients find testing to be burdensome. There is a small chance that theophylline could make your asthma worse by relaxing the muscle that separates your stomach from your esophagus; if this occurred, the treatment would be stopped.

Second, Singulair could be replaced with Zyflo CR (zileuton, controlled release). The active ingredient in Singulair is montelukast, which blocks the action of the cysteinyl leukotrienes at the CysLT1 receptor, whereas zileuton prevents the synthesis of both cysteinyl leukotrienes and dihydroxy leukotrienes. There are theoretical reasons to believe that controlled-release zileuton would yield a clinical benefit, but there are no compelling data to support this approach. Monitoring of liver function is required during initiation of treatment with zileuton.

Third, Xolair (omalizumab) could be added to your regimen. This anti-IgE monoclonal antibody is given once a month by injection. There is clearly an allergic component of your disease; your total IgE level is elevated, but it is not so high as to preclude the use of omalizumab.

As we discussed, I think your primary care physician has done an excellent job in designing your asthma treatment. You should discuss our consultation with her and decide what is in your best interest.

Comment: There have been three major changes in our understanding of asthma between 1928 and 2012. First, spirometry, which had been invented in the 1840s, was refined by adding time to volume output, and between the late 1940s and early 1950s, measurements made from forced exhalations were used in the diagnosis and treatment of asthma. Other lung-function tests were developed and used, and the relationships between clinical physiology and symptoms were delineated. Second, glucocorticoids were identified as an effective and useful asthma treatment. They were first used systemically in the early 1950s and were subsequently made available in inhaled form; these agents remain the standard of care today. Third, our understanding of the immunobiology of asthma progressed beyond the view that the essential mechanism was an immediate hypersensitivity reaction. Unfortunately, these advances in understanding the cell biology of asthma have not yet been translated into new therapies, although new therapies have been derived from our improved understanding of immediate hypersensitivity reactions—notably, the use of leukotriene modifiers and anti-IgE antibodies.
Our patient is current in her medical knowledge and is using medical information widely available on the Internet to help in the management of her chronic condition. The consultant used measures of lung function to quantify her physiological deficit. The consultant also measured the patient’s IgE level, which was consistent with allergic asthma, and provided the information needed for anti-IgE treatment, should the patient elect this approach. The patient has used all the standard asthma therapies but has residual symptoms. The consultant outlines other asthma treatments that the patient could try, highlighting the need to try different treatments to see whether one or another will work. Sadly, we still do not have a way to predict a given patient’s response to therapy.

**Figure 3. Spirometric Results for Ms. Smith.**

Although the forced expiratory volume in 1 second (FEV₁) is within the normal range of the predicted value, the ratio of FEV₁ to forced vital capacity (FVC) is low. The patient’s FEV₁ increases to almost 700 ml with inhaled albuterol, indicating that she has substantial reversible airway obstruction. The tracings and data shown are similar to the data displays provided by many spirometers that are currently available. FEF₂₅–₇₅ denotes forced expiratory flow between 25 and 75% of FVC, FET forced expiratory time, IFR inspiratory flow rate, PEFR peak expiratory flow rate, and PIFR peak inspiratory flow rate.

**CONCLUSIONS**

These three case histories illustrate that asthma as a disease has not changed for two centuries. We have made real progress in identifying patients with asthma and in understanding its biologic basis and its treatment. Progress has also been made in diagnostic testing, which has been refined to measure lung function with great accuracy and repeatability. In addition, we can measure the lung’s responsiveness to triggering agents and thereby obtain objective indications of disease activity, in addition to the patient’s history. We have come to realize that allergic responses often sub-
stantially contribute to both chronic persistent asthma and acute exacerbations of asthma symptoms, although this association may have been overemphasized. The current focus is on immune responses in affected patients; we believe that airway inflammation plays a critical role in asthma, but the precise nature of that inflammation remains a mystery. The knowledge that asthma runs in families has become the basis for very sophisticated studies of the genetics of asthma, but not much of its heritability can be explained by all the genes identified to date. In spite of all the progress achieved over the past two centuries, we still do not understand the fundamental causes of asthma. Hence, we do not have therapies that address the underlying mechanisms; we have no cure for asthma. The medications at hand provide relief of symptoms and improve lung function and airway responsiveness, but they do not prevent exacerbations or progression of disease, and the most desirable accomplishment, the primary prevention of asthma, is a vision that has not yet become a reality.

Disclosure forms provided by the authors are available with the full text of this article at NEJM.org.

REFERENCES
6. Granville AB. Further observations on the internal use of the hydrocyanic (Prussic) acid, in pulmonary complaints; chronic catarrhs; spasmodic coughs; asthma; hooping cough; and some other diseases, with full directions for the preparation and administration of that medicine. N Engl J Med Surg 1820;9:193-7.

Copyright © 2012 Massachusetts Medical Society.
Major Trends in the U.S. Health Economy since 1950
Victor R. Fuchs, Ph.D.

Rapid advances in medical science and technology, substantial gains in health outcomes attributable to medical care, and budget-busting increases in health care expenditures fueled by private and public insurance have marked the past six decades of health care in the United States. As the country struggles to emerge from a multiyear financial and economic crisis, policymakers and the public have increasingly homed in on those skyrocketing health care expenditures. What lessons can be drawn from the evolution, since 1950, in the sources of payment and objects of expenditures in the health care arena?

HEALTH EXPENDITURES
The rapid growth of health expenditures is one of the most important economic trends in the United States in the post–World War II era. It has implications for the financial viability of federal and state governments and has resulted in stagnation of wages in most industries. In 1950, health expenditures accounted for only 4.6% of the gross domestic product (GDP). In 2009, they accounted for more than 17%, a larger share than all manufacturing, or wholesale and retail trade, or finance and insurance, or the combination of agriculture, mining, and construction. According to public finance experts such as Alan Blinder and Alice Rivlin, control of health care expenditures is the greatest fiscal policy challenge facing the United States.

From 1950 through 2009, there was an almost continuous increase in annual real per capita health expenditures, with the exception of one 2-year pause in the mid-1990s, when the effect of managed care was at its peak1 (see line graph). The absolute rate of growth has been increasing over time, as evidenced by the concave shape of the curve in the graph. The relative rate of increase was greater between 1950 and 1980 than between 1980 and 2009 — 4.6% versus 4.1% per year — primarily because of the introduction of Medicare and Medicaid in 1965.

Unfortunately, the slight slowing in the rate of growth of health expenditures since 1980 was accompanied by even greater slowing in the growth of the GDP (per capita adjusted for inflation), from 2.6% per year in 1950–1980 to 1.8% per year in 1980–2009. Thus, the gap between the rate of growth of health expenditures and that of GDP increased from 2.0% to 2.3% per year between the two periods. Most experts be-
lieve that such a gap is not sustainable over the long term, because health expenditures would cut too drastically into the availability of other essential goods and services.

The most important explanation for the increase in real per capita health expenditures is the availability of new medical technology and the increased specialization that accompanies it. Between 1974 and 2010 alone, the number of U.S. patents for pharmaceutical and surgical innovations increased by a factor of six. Second in importance is the spread of public and private health insurance, which diminishes the effect of health care prices on demand. There is a positive-feedback loop between new technology and the spread of health insurance: new technology stimulates the demand for insurance, and the spread of insurance stimulates the demand for new technology. Finally, a small portion of the increase, typically 0.1 or 0.2 percentage points per year, is attributable to the aging of the population. It’s not possible to estimate how much of the increase in expenditures reflects higher health care prices and how much reflects greater quantities of care, because the content of a day in the hospital or a visit to a physician keeps changing. No doubt some of the increase in expenditures reflects an increase in the quantity of medical care, if quantity is adjusted for improvements in the quality of care.

**Sources of Payment**
The sources of payment for medical care have changed significantly since 1950 (see table). The most important trends have been a decline in out-of-pocket payment and a rise in third-party payment (both private and public), an increase in government’s share of payment and a decrease in the private share, and an increase in the federal government’s share as compared with that of state and local governments.

Third-party payment has grown partly because of expensive interventions that expose individuals to large financial risk and partly because employers’ contributions to employee health insurance are not considered part of employees’ taxable income. Since World War II, there has been a large increase in the number of workers who must pay income tax and an even greater increase in the number who must pay payroll taxes. These increases have made tax-exempt employer-based health insurance more attractive. A shift from individual to group insurance has also contributed to the spread of coverage by reducing marketing and administrative costs and, thanks to compulsory participation within firms, limiting the risk of adverse selection for insurance companies.

The growth of government’s share, and especially the federal share, can be explained by the public’s desire to cover more of the public with insurance and private insurers’ difficulty in providing coverage for the elderly and the poor. Federal legislation also substantially extended public coverage for children.

**Objects of Expenditures**
Throughout the period since 1950, health expenditures have gone primarily to hospitals, physicians, and drugs. Moreover, the rate of growth of expenditures in each of these categories between 1950 and 2009 has been fairly close to the rate of growth of total health expenditures (see bar graph). Drug expenditures may appear to have grown more slowly, but that’s probably due to a data mismatch: the 1950 figure includes sundries, whereas the 1980 and 2009 figures are for prescription drugs only. Such stability in the share of these categories is remarkable, given the great changes that have occurred in medical technologies, sources of payment, and health policy since 1950. As a rule of thumb, the ratio 3:2:1 does a fairly good job of describing the relative importance (in dollar terms) of hospitals, physicians, and drugs. The “other” expenditures are divided among many categories, the most important of which are public administration
and the net cost (premiums minus benefits paid) of private health insurance, nursing homes, and dental services.

There have been periods in the past 60 years when individual categories accounted for greater or lesser proportions of expenditures. Spending for hospital care and physicians received a boost between 1950 and 1980 from the introduction of Medicare and Medicaid. Spending for drugs accelerated sharply after 1980 following the introduction of a host of new products for treating heart diseases, mental illness, gastrointestinal disorders, and cancer and a large increase in private and public insurance coverage for drugs.

The ability of hospitals to maintain their high share is particularly noteworthy, because between 1950 and 2009 the industry had several large shocks. Psychiatric hospitals virtually emptied out. Admission rates to acute care hospitals (“community” hospitals) dropped precipitously after 1970, as did the average length of stay. As a result, the average daily census, adjusted for population growth, has decreased by almost 50% over the past four decades. Hospitals have maintained and increased their revenues in part through more intensive treatment of inpatients. Despite shorter stays, the cost per case (in 2009 dollars) jumped from $6,600 in 1997 to $9,200 in 2009. Hospitals’ total incomes were also preserved through expansion of outpatient services, including same-day surgery, magnetic resonance imaging and computed tomography, and outpatient clinics for diagnosing and treating cancer, heart disease, and other illnesses.

Community hospitals (including academic medical centers), the recipients of the largest share of health expenditures, have seen dramatic shifts in demand for and supply of inpatient care since 1950. During the first three decades of this period, the number of inpatient days per 1000 population increased by more than a third, driven by Medicare and Medicaid, the spread of employer-based insurance, and lax utilization controls by public and private payers (see Table 1 in the Supplementary Appendix, available with the full text of this article at NEJM.org). A slight decline in the average length of stay was more than offset by a 50% increase in the number of admissions per 1000 population. The industry’s 31% increase in the number of beds per 1000 population, abetted by consultants’ predictions of ever-growing demand, proved to be an expensive mistake. In the late 1960s and early 1970s, there was mounting evidence that many hospital admissions were ill-advised and that lengths of stay for many patients were overly long (see the Supplementary Appendix).

Between 1980 and 2009, the number of inpatient days per 1000 population fell by almost half, with declines in admissions and average length of stay contributing almost equally. The decline in length of stay was particularly spectacular in some major categories of patients. For example, stays for uncomplicated myocardial infarction dropped from 3 weeks to 3 days; for uncomplicated vaginal delivery, from 1 week to 1 day; and for herniorrhaphy, from 6 days to same-day surgery. The average decrease among all patients, however, was smaller than those for individual causes of admission, because the aver-

<table>
<thead>
<tr>
<th>Variable</th>
<th>Year or Period</th>
<th>1950</th>
<th>1980</th>
<th>2009</th>
</tr>
</thead>
<tbody>
<tr>
<td>Per capita expenditures</td>
<td>2009 dollars</td>
<td>407</td>
<td>2,050</td>
<td>6,807</td>
</tr>
<tr>
<td>Source of payment (%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Out-of-pocket (%)</td>
<td></td>
<td>56</td>
<td>27</td>
<td>14</td>
</tr>
<tr>
<td>Third-party (%)</td>
<td></td>
<td>44</td>
<td>73</td>
<td>86</td>
</tr>
<tr>
<td>Private (%)</td>
<td></td>
<td>73</td>
<td>60</td>
<td>53</td>
</tr>
<tr>
<td>Public (%)</td>
<td></td>
<td>27</td>
<td>40</td>
<td>47</td>
</tr>
<tr>
<td>Federal (%)</td>
<td></td>
<td>13</td>
<td>26</td>
<td>35</td>
</tr>
<tr>
<td>State and local (%)</td>
<td></td>
<td>14</td>
<td>14</td>
<td>12</td>
</tr>
<tr>
<td>Average annual rate of change</td>
<td>(%) in 2009 dollars</td>
<td>3.0</td>
<td>1.9</td>
<td>2.4</td>
</tr>
<tr>
<td>Out-of-pocket (%)</td>
<td></td>
<td>7.1</td>
<td>4.7</td>
<td>5.9</td>
</tr>
<tr>
<td>Third-party (%)</td>
<td></td>
<td>4.7</td>
<td>3.7</td>
<td>4.2</td>
</tr>
<tr>
<td>Private (%)</td>
<td></td>
<td>6.7</td>
<td>4.7</td>
<td>5.7</td>
</tr>
<tr>
<td>Public (%)</td>
<td></td>
<td>7.8</td>
<td>5.0</td>
<td>6.4</td>
</tr>
<tr>
<td>State and local (%)</td>
<td></td>
<td>5.2</td>
<td>3.8</td>
<td>4.6</td>
</tr>
</tbody>
</table>

* The percentage of payments by the federal government was calculated on the basis of National Health Care Expenditure data. Data are from the Department of Health and Human Services and the U.S. Census Bureau.
age severity of patients’ conditions on admission increased. The hospital industry responded to the drop in demand by closing some hospitals (net decrease of 18%) and closing off some beds as unavailable, but even so, the average occupancy rate fell by 10 percentage points to the inefficient level of 65.5%.

**Physicians**

The number of active physicians in the United States increased by a factor of approximately four between 1950 and 2009 (see Table 2 in the Supplementary Appendix). As the population grew, the number of active physicians per 1000 population increased from 1.41 to 2.73, an annual growth rate of 1.1%. That figure may overstate the growth of physicians’ availability, however, since the number of hours the average physician worked probably decreased appreciably between 1950 and 2009. Major trends in the physician supply that had important implications for the health economy were large increases in the percentages of female physicians, specialist physicians, and hospital-based physicians.

Because women, even professional women, still bear a disproportionate share of domestic responsibilities, female physicians tend to differ from their male peers in preferences regarding annual hours of work, night coverage, self-employment, specialty choice, and other aspects of practice.

The increase in the proportion of physicians who are specialists and subspecialists has resulted in a considerable increase in the number of years the average physician spends in training, although a restructuring of medical education could change that. There has been a large increase in the number of specialists and an even larger increase in the number of specialties and subspecialties, from a few dozen 50 years ago to more than 150 now.

The shift away from office-based practice, along with possible changes in payment systems, may portend a time when most medical care will be delivered by teams of physicians and other health care providers (e.g., nurse practitioners and physician assistants) working in accountable care organizations.

**Changes in Organization and Delivery**

An important recent trend affecting hospitals and physicians is a sharp division between physicians who treat outpatients and others, called hospitalists, who treat only inpatients. The number of hospitalists has grown rapidly, from no more than 1000 15 years ago to 7000 10 years ago to approximately 30,000 in 2011, according to physician-economist David Meltzer of the University of Chicago. Hospitalists are said to improve both the efficiency of care (mostly through reducing lengths of stay) and its quality. Though primary care physicians initially resisted this change in professional responsibilities, many now prefer the new system because they perceive that hospital visits were not an efficient use of their time.

Another trend attracting wide attention is the use of electronic medical records (EMRs) in physicians’ offices. Opinions vary regarding the effects of EMRs on the efficiency and quality of care. I believe a well-organized health care system can benefit substantially from EMRs, but the fragmented nonsystem of U.S. medical care is not likely to derive enough benefit to justify the cost.

During this period, another change that affected hospitals and physicians was the development of managed care. Until about 1990, most insured patients could choose freely among providers, physicians’ decisions were not subject to frequent questions by insurers, and payment was typically fee for service. The rapid growth of health care expenditures in the late 1980s, combined with sluggish growth of the GDP, fueled a demand for change. In the 1990s, insurers selectively contracted with providers, fees and prices were negotiated in advance, physicians’ decisions became subject to insurance-company review, and patients faced financial penalties for obtaining out-of-plan care. The effect on health care expenditures was
PAST AND FUTURE

The six decades since 1950 have been remarkable for the U.S. health economy in many ways, especially the extraordinary increase in health care expenditures. Future historians may, with some irony, refer to this period as a golden age for U.S. medicine because health care's share of the GDP quadrupled from 4.6% in 1950 to more than 17% in 2009; in most peer countries, the share is 9 to 11%. Other noteworthy trends in the health economy have been the spread of private and public health insurance to the point where almost 90% of the total bill for care is paid by third parties; the increased role of the federal government in funding health care; the decline in inpatient use of hospitals (fewer admissions and shorter stays) and the expansion of hospital outpatient services; the shift in the physician workforce toward more women, more specialists, and more hospital-based physicians; and the deluge of new medical technologies confronting clinicians with a menu of 6000 drugs and 4000 procedures to choose from.

It is difficult to see how the health sector can continue to expand rapidly at the expense of the rest of the economy, but every past prediction of a sustained slowing of the growth of health expenditures has been proved wrong. Rapid growth may continue as a result of political gridlock regarding the form that curbs on expenditures should take. There is no public consensus about how much care should be provided for the poor and sick or how it should be done. Similarly, there's no public consensus regarding efforts to increase the efficiency of care. A rational approach to the financing, organization, and delivery of care seems politically impossible. However, the observation by de Tocqueville that in the United States “events can move from the impossible to the inevitable without ever stopping at the probable” may prove to be prescient.

Disclosure forms provided by the author are available with the full text of this article at NEJM.org.

From Stanford University, Stanford, CA.

6. Fuchs VR. Alan Gregg Lecture: The structure of medical education — it’s time for a change. Presented at the Annual Meeting of the American Association of Medical Colleges, Denver, November 6, 2011.

Copyright © 2012 Massachusetts Medical Society.
What We Don’t See

Margaret Kendrick Hostetter, M.D.

SIXTY-EIGHT YEARS AFTER THE INAUGURAL ISSUE OF THE NEW ENGLAND JOURNAL OF MEDICINE AND SURGERY, Sir William Osler introduced the term “pediatrics.” Although “diseases peculiar to children” had figured in Benjamin Rush’s lectures at the University of Pennsylvania since 1789, most physicians in the early 19th century did not recognize children as a distinct population with particular medical needs. Indeed, in most medical journals of this period, the words “infant,” “child,” and “children” figured only in case reports of obstetrical complications or in accounts of epidemic-related mortality. Osler’s use of the term “pediatrics” not only differentiated physicians “specially connected with pediatrics” from other physicians but also drew attention to the creation of a “special” discipline. After Osler’s introduction of the term, articles entitled “Progress in Pediatrics” began to appear sporadically in the Journal from 1904 forward, and the specialty of pediatrics was accorded its own section in the Journal in 1954.

However, charting the influence of the Journal on two centuries of pediatric medicine is perhaps more thoughtfully addressed by means of a computational algorithm to search the Journal’s archives for Osler’s designation — pediatrics — so as to gain insight into the increasing emphasis on childhood health over the past 200 years (Fig. 1). This approach allows us to demarcate four periods, albeit somewhat arbitrarily defined, that reflect the emergence and discoveries of pediatrics in the pages of the Journal — the recognition of children as a particular population benefiting from medical practice (1812 through 1880), the introduction of public health programs to reduce childhood mortality (1881 through 1930), the development of vaccines (1931 through 1980), and the global dissemination of pediatric practice (1981 through 2012) (see timeline, available with the full text of this article at NEJM.org). Each of these periods also coincides with particular advances against infectious disease — the single leading cause of childhood death yesterday and today (Fig. 2) — that were heralded in the pages of the New England Journal of Medicine: the concept of social intervention, the battle against summer gastroenteritis, the defeat of smallpox, and the drive to reduce childhood mortality from diarrhea and the human immunodeficiency virus (HIV).

Seminal articles in the Journal have also advanced the fields of pediatric surgery, oncology, neonatology, and a host of other pediatric disciplines over the past two centuries. Transformative approaches to noninfectious causes of childhood death — congenital heart disease and leukemia — appeared first in the Journal. Landmark articles describing functional asplenia in sickle cell disease, legal redress for the battered child, the genetic foundations of population-based screening for phenylketonuria and cystic fibrosis, molecular and therapeutic advances in surfactant deficiency, and the relationship between prone sleeping position and the sudden infant death syndrome have blazed the trail of progress in pediatrics.

But the fact remains that infectious diseases were the predominant cause of childhood death at the time of the Journal’s inception and remain so, globally, today (Fig. 2). Therefore, this review emphasizes how articles in the New England Journal of Medicine have directed and illuminated the course of progress against the infec-
tious diseases of childhood across these four eras. By reviewing some of the Journal’s most celebrated contributions, we can recognize what we don’t see today and realize what we must see in the future.

1812 THROUGH 1880 — SEE IT

For centuries, the cumulative weight of experience fostered the conclusion that childhood deaths were inevitable. Many families, even those of plentiful means, lost half or more of their children. Seemingly nothing could be done — medically, politically, or economically — save to let nature take its course. Indeed, the mind of the public had changed little since the 2nd century, when the emperor Marcus Aurelius wrote, “One man prays, ‘How I may not lose my little child’, but you must pray, ‘How I may not be afraid to lose him’” (Meditations 9.40).

One has but to view the tombstones in colonial cemeteries to understand that death in childhood represented a grievous but seemingly inexorable trajectory. The death toll from infection among the very young was often obscured in colonial epidemics, when smallpox, diphtheria, cholera, dysentery, and measles typically killed without respect to age. In November 1713, for example, the wife of the Puritan minister Cotton Mather died in a measles epidemic, along with her newborn twins, a 2-year-old daughter, and a servant. Two sons and four daughters, all older than 7 years of age, survived. Apart from these individual tragedies, however, there was little recognition of the special susceptibility of children, particularly those under 5 years of age, until the diphtheria epidemic.
in New England (1735 through 1740), in which 80% of its 5000 victims — almost 2.5% of the population — were children.13

By the middle of the 19th century, a child’s death, far from intolerable, was frequently viewed as blessed, a release from the torment of hectic infection or the lingering complications of diseases such as rheumatic fever. A contemporary physician’s description of death from croup reflects this view:

his inarticulate appeals and beseeching looks for relief . . . constitute one of the most touching scenes which we are called upon to witness in the practice of medicine. Happily the extreme suffering usually, though not always, subsides towards the close of life, and death takes place at last with comparative ease14

and prefigures, in its resignation, the death of the fictional character Beth in Louisa May Alcott’s novel Little Women:

As Beth had hoped, the ‘tide went out easily’, and in the dark hour before dawn, on the bo-

som where she had drawn her first breath, she quietly drew her last, with no farewell but one loving look and a little sigh. . . . When morning came . . . the spring sunshine streamed in like a benediction. . . .

But gradually, the attitude of helplessness changed, first to inquiry and then to responsibility. The recognition that social, as well as divine, intervention could influence the life and death of children took hold. Notably, the Journal’s annual summaries of childhood deaths in Massachusetts not only tallied the deaths but also began to include prescriptions for change. Reporting on infant mortality in the Journal in 1873, J.O. Webster concluded with the “point[s] that strike us most forcibly . . . that sanitary reform, as a means of reducing our infant, as well as general mortality, demands our earnest attention.”15

1881 THROUGH 1930 — FIX IT

Reforms surged. Photographs such as those by Jacob Riis in How the Other Half Lives16 (Fig. 3) pricked the societal conscience to see children as a group to be protected. The visible symbol of this new recognition was the children’s hospital, opening first in Philadelphia (1855) and then in Boston (1869) and Cincinnati (1887). Many of these institutions arose from the efforts of ladies’ auxiliaries, which typically provided food, clothing, nursing, and free care for all patients in the early days of each hospital.17 The Harriet Lane Home, a much larger institution, was endowed in Baltimore in 1903 with a $400,000 gift from its benefactress, and the end of the 19th century marked the establishment of more than two dozen children’s hospitals.

An editorial in the Journal drew on the charter statement for Children’s Hospital Boston to emphasize not only the charitable nature of these enterprises but also their importance for scientific progress and national stature:

We believe that, apart from the actual medical treatment of sick and injured chil-
dren, there is a want in our community which has long been felt in our medical schools, though provided for in foreign cities; namely, an opportunity to study infantile disease. . . . For people forget that on the health of the growing up generation hangs that of generations more.18
Attendant protections extended to the prohibition of child labor and the guarantee of education. The *Journal* reported on French legislation prohibiting boys less than 13 years of age and girls less than 14 years of age from working more than 6 hours per day and published lectures addressing the necessity and content of compulsory education for girls as well as boys. Child labor laws, first enacted in New England in 1832, gained momentum throughout this era, as did standards of compulsory childhood education that were patterned on those introduced in the Massachusetts Bay Colony in 1647. The *Journal* did not hesitate to give voice to these political initiatives, and their coverage in a leading medical journal underscored their profound medical implications.

But perhaps the most important movement during this period was the emphasis on public health, as exemplified by the fight against cholera infantum, or summer gastroenteritis, the cause of 15 to 22% of childhood deaths across New England and New York. By the middle of the 19th century, breast-feeding rates had declined considerably, especially among the urban working poor, who were forced to avail themselves of unstandardized cow’s milk powders that necessitated mixing with milk or water, which frequently came from contaminated sources. Abraham Jacobi, a New York pediatrician, advocated the universal pasteurization of milk, and New York merchant Nathan Straus implemented this principle in his free “milk stations,” the first opening in 1893. Providing guidance, the *Journal* promoted scholarly analyses of the ideal composition of milk-based products for infant nutrition, which definitively stated which additives were needed for milk from a Jersey cow (5% fat) as compared with the milk from a Holstein (3% fat).

Nevertheless, as the 20th century dawned, infant deaths due to summer gastroenteritis failed to decline, despite mandatory pasteurization and the best efforts of the milk stations. It remained for S. Josephine Baker, a physician and the first director of the New York City Bureau of Child Hygiene, to demonstrate the profound effect of maternal education on breast-feeding, home hygiene, and infant care. Over a period of just 10 years, between 1907 and 1917, the infant mortality rate in New York decreased from 144.4 to 88.8 deaths per 1000 and the continued implementation of practical principles of household hygiene and child protection — what pediatricians today call “anticipatory guidance” — led to a dramatic fall in childhood deaths from diphtheria 10 years before the vaccine was in widespread use (Fig. 4).

**1931 THROUGH 1980 — PREVENT IT**

Progress in public health fixed many problems that had been present for decades. Moving to the next stage required a paradigm shift: the belief that some problems could actually be stopped before they occurred. The concept of immunity and the development of vaccine technology expanded dramatically the list of diseases we don’t see. Beginning with von Behring’s diphtheria vaccine in 1913, the *Journal* reported on trials of vaccines that became the conquerors of common childhood diseases — the field trial of pertussis vaccine in Michigan by Drs. Pearl Kendrick and Grace Eldering from 1934 through 1937, the diphtheria–pertussis–tetanus vaccine trials in the 1940s, and the intramuscular and oral poliomyelitis vaccine trials in the 1950s and 1960s.

In the mid-1950s, twice as many children died from measles as from poliomyelitis, and the immunologic principles and clinical efficacy of the attenuated measles virus vaccine in a variety of childhood populations were thoroughly detailed in a series of eight landmark articles, all published in the July 28, 1960, issue of the *Journal*. The combined measles, mumps, and rubella vaccine, developed by Maurice Hilleman in 1971, was a boon for pediatrics and obstetrics alike. The principles underlying the development of successful protein-based vaccines, in turn, opened the door to large-scale trials of polysaccharide and conjugate vaccines for *Haemophilus influenzae*...
Figure 4. Diphtheria as a Fraction of All Deaths in the United States, 1900–1956.

Diphtheria declined as a cause of death in the United States over the period from 1900 through 1956. The large panel shows the raw data and a logistic curve fitted to the data. The inset shows the same data and a Fisher–Pyt transform that renders the S-shaped curve linear and normalizes the process to 1. \( F \) denotes the fraction of the process completed. The time it took for the reduction of diphtheria to go from 10% to 90% is 40 years, and 1911 is the midpoint. Data from the U.S. Bureau of the Census.23 Adapted from Ausubel et al.24

The capstone of this era was the worldwide eradication of smallpox by means of vaccination campaigns, trumpeted in the Journal in 1980.43 The subject of 13,968 articles in the Journal since 1812, the perils of smallpox and the benefits of vaccination had long been recognized. An article from 1814 states, “In the report of the National Vaccine Establishment of Great-Britain, it appears that the prejudices against vaccination have greatly declined, and that in the city of London three-fourths of the children born are submitted to that salutary operation.”44

Mandatory vaccination, however, incited fervid public debate in both the United Kingdom and the United States, until the publication of Charles Dickens’s Bleak House. That novel seared into the public consciousness the meaning of the death of a child — the orphaned street sweeper Jo, whose illness, unnamed but readily recognizable as smallpox, arose from the miasma of impoverished London and enveloped both the middle-class heroine and her aristocratic mother:

\[
death = \text{probability of death} = \frac{\text{number of deaths}}{\text{total population}} \]

Dead! Dead, your Majesty. Dead, my lords and gentlemen. Dead, Right Reverends and Wrong Reverends of every order. Dead, men and women, born with Heavenly compassion in your hearts. And dying thus around us every day.

“And dying thus around us every day.” Even the House of Lords listened. Within the year, the British Parliament mandated that every child born in England and Wales after August 1, 1853, be vaccinated for smallpox within 3 months after birth.45 It took 50 years for the United States to follow suit.

In 1967, when the World Health Organization (WHO) established its smallpox eradication unit, 131,000 cases were reported worldwide, but 10 million to 15 million cases were believed to have occurred, with a 15 to 20% fatality rate.43 Three years later, correspondence in the Journal from leaders of the WHO initiative addressed the requisites for progress, and subsequent improvements in surveillance, communication, and coordination led to the elimination of the disease in 1980 — the first (and only) infectious disease to have been eradicated from the face of the earth. It is not overly optimistic to hope that polio and malaria will vie for second.46

As these practices transformed childhood life expectancy throughout industrialized nations, the ease of intercontinental travel and the rise of global economies brought the conditions of developing countries into sharp focus. Thirteen years after the first report of HIV infection, in 1981, Pediatric AIDS Clinical Trials Group Protocol 076 established that the use of zidovudine ante partum and intrapartum in the mother and post partum in the newborn child reduced the mother-to-child transmission of HIV by 67.5%.47 The ability to prevent vertical transmission of HIV infection spurred the widespread adoption of state legislation requiring the testing of pregnant women, their infants, or both,48 and the numbers of new cases of vertically transmitted HIV have plummeted in virtually all U.S. cities. In adapting these approaches for global use, highly effective regimens of single-dose antiretroviral agents were eventually developed to meet the needs of economically disadvantaged countries.

\[
\text{Percent of Process Remaining} = \frac{1}{F} - 1
\]

\[
\text{Percent of Process Remaining} = \frac{1}{F} - 1
\]
Despite commendable progress in controlling mother-to-child transmission of HIV, the leading causes of death in children under 5 years of age worldwide are not HIV infections but pneumonia and diarrheal diseases, each of which killed more than 1 million children in 2008 (Fig. 2).13 Ironically, the 15% global fatality rate from diarrhea today differs not one whit from its toll in Massachusetts in 1873.14 The *New England Journal of Medicine* has championed the dissemination of vaccines from industrialized countries by publishing critical analyses of their efficacy in resource-poor countries.49 Concerted efforts from public and private partnerships may soon ensure that the major killers of children—pneumonia and diarrhea—will enter the archives of diseases we don’t see.

As we celebrate the impact of the *Journal* during the past 200 years, we must applaud its expanding influence on the practice of pediatrics—from the children of Massachusetts to those of the United States, and now to the children of the world. As each era of progress amply demonstrates, the best place to begin a healthy life is at the beginning. In the coming century, by publishing premier research on such topics as fetal programming and the fetal—maternal interface or the genetics and treatment of birth defects and chronic illnesses, the Journal must continue to heighten the visibility of children to physicians and philanthropists, statesmen and statisticians, educators and economists. As we look back, the diseases we don’t see represent a triumph; as we look ahead, what we don’t see will require the *Journal* to light the way.

I thank Tim Schwartz of San Diego, who developed the algorithm for Figure 1, and Texas Children’s Hospital, Children’s Hospital Boston, the University of Minnesota Amplatz Children’s Hospital, Yale–New Haven Children’s Hospital, and Cincinnati Children’s Hospital Medical Center for their continued inspiration throughout the past four decades.

Disclosure forms provided by the author are available with the full text of this article at NEJM.org.

**REFERENCES**


16. Riis JA. How the other half lives. New York: Charles Scribner’s Sons, 1890.


33. Enders JF, Katz SL, Milovanovic MV, Holloway A. Studies on an attenuated measles-virus vaccine. I. Development and

Copyright © 2012 Massachusetts Medical Society.
The Evolving Roles of the Medical Journal
Scott H. Podolsky, M.D., Jeremy A. Greene, M.D., Ph.D., and David S. Jones, M.D., Ph.D.

As Chicago physician J.H. Salisbury remarked in 1906, the influence of the medical journal on the life of the physician is unparalleled: “Medical school is attended, as a rule, but once in a lifetime; the meetings of the medical society are usually infrequent, but the medical journal, like the newspaper, is an ever-present friend whose influence and advice are potent for good or evil.” Yet medical journals have often had a tenuous existence. Thousands have come and gone over the past 200 years, and many continue to struggle to define their role.

The New England Journal of Medicine has persisted, and its history provides a window on the changing functions of both medical journals and the medical profession. Journals don’t simply disseminate new knowledge about medical theory and practice. They also define the scope of medical concerns and articulate norms for physicians’ professional and social roles. Simultaneously, they work to preserve their reputation, financial stability, and editorial independence in a constantly changing publishing environment, amid an avalanche of medical information.

THE ORIGINS AND EPIDEMIOLOGY OF MEDICAL JOURNALS

Launched in 1812, the New England Journal of Medicine and Surgery and the Collateral Branches of Science was a latecomer to medical publishing in the new republic: medical journals had already been established in New York, Philadelphia, and Baltimore. The first issue included essays on clinical problems, reviews of general progress in the sciences, a celebration of François Xavier Bichat, and an account of an effort to decipher a secret French remedy for gout. Initially published quarterly, the Journal merged in 1828 with the Boston Medical Intelligencer to become the weekly Boston Medical and Surgical Journal, a name it kept for a century. Survival was a challenge. As prominent journals struggled and folded during the Civil War, the Journal asked readers for financial support and submissions “which will show that the light of pure medical science still burns among us undimmed by the heavy clouds of national trial and adversity” (1864; see box for cited Journal articles).

Throughout the century, hundreds of new journals appeared, as medical societies, medical colleges, and therapeutic sects planted their flags. In 1879, the same year he established the Index Medicus to tame the growing medical literature, Surgeon General John Shaw Billings complained that “it is as useless to advise a man not to start a new journal as it is to advise him not to commit suicide” (1879a). By 1882, nearly three quarters of the 509 journals founded in the United States since 1797 had reportedly “passed over to the silent land.”

Nor did survival ensure quality. According to an 1879 Journal article, the Chicago Medical Journal and Examiner had surveyed the survivors and deemed many “absolutely worthless” and others “undeniably worse than worthless — they are dangerous and disgusting parasites upon the body medical.” In 1876, the Journal’s editors opined that it “would be an immense gain if the medical journalism of our whole country could be compressed into not more than half a dozen weeklies,” plus a few monthlies, quarterlies, and specialty journals. Such dreams of rationalized medical communications yielded to an awareness of the commercial realities of a Darwinian “struggle for existence” (1913).

Although the Journal endured through the 19th century and beyond, it initially lived in the shadow of the Journal of the American Medical Association (JAMA, founded in 1883), the dominant medical journal in North America well into the 20th century. Nonetheless, the Journal celebrated its centennial twice, in 1912 and 1928. In 1921, it was sold to the Massachusetts Medical Society for $1 — though the purchase provided more stability than the price suggests. Competition remained keen in 1921, with more than 5000 medical journals “descending like locusts on the land” (1952). And it intensified over the century, as general journals begat specialty journals, subspecialty journals, single-subject journals . . . and then, to link them up again, interdisciplinary journals, all of which “were fruitful and multiplied” (1981). Although this journal, renamed the New England Journal of Medicine in 1928, achieved global prominence, its editors remained anxious — in particular, Franz Ingelfinger admitted in 1972, about the role of a general medical journal in a world of subspecialized knowledge and practice.

THE MANY ROLES OF A MEDICAL JOURNAL

Speaking before the American Association of Medical Editors in 1884, president Leartus Connor defined an ambitious vision for the medical journal. It ought to be a medical school, a residency program, a clinical preceptor, a set of textbooks, and a medical society unto itself. “In short,” Connor concluded, “it is the great unifier of the past and present, the diffuser of all new facts, new thoughts, all new and better appliances for the study of the hu-
man body and for the relief of its derangements.” Although this may seem an impossible mandate, nearly all these goals remain recognizable in journals today.

The dissemination by journals of new research results in pathophysiology and therapeutics demonstrates the changing ways in which medical knowledge is produced. In the Journal’s early decades, original research contributions from American physicians couldn’t fill even a quarterly, according to Joseph Garland (1952). The editors supplemented research with reviews of books, conferences, and other journals and presented French and German scholarship (conducting literally “translation” medicine). They also published reviews of clinical practice, case reports, readers’ letters, and their own editorials.

Over the years, journals have not simply responded to submissions, disseminating pieces of knowledge like cargo on a train. They have shaped both the tracks and the trains, using their perspective and authority to articulate and enforce standards for the forms research should take — promoting the use of case studies in the early 20th century and lamenting “weak research designs” and the slow adoption of randomized, controlled trials in the late 20th century (1979b). Editors have begun aggressively policing conflicts of interest among researchers (2000) and have banded together to call for transparency in clinical trials, requiring public trial registration as a precondition for consideration for publication (2004).

Journals have also played a more tacit yet crucial role: delineating the appropriate domain of medical knowledge and practice. In choosing which topics to cover, editors make judgments about what doctors and health policymakers ought to know. Reports about botany and natural history, once essential to medical practice, have disappeared, replaced by topics such as epigenetics and pharmacogenomics. Sometimes editors exclude topics. The Journal, for instance, repeatedly refused to grant homeopathy and other therapeutic alternatives the “recognition which would be extended cheerfully to a legitimate enterprise” (1877, 1979a).

More often, they follow medicine far afield, from the climatology data that appeared regularly in 1812, to concerns about degeneracy and race suicide in 1912 (1912b), and exposés on the health effects of environmental hazards (1966a), nuclear war (1986), and global climate change (1989). Joseph Garland, who oversaw the Journal’s expansion into an international forum, believed that journals and their editorials need not “necessarily be confined to topics related to medicine, so long as they are ones in which the reader has or might have an interest” (1952). Subsequent editors have concurred, arguing that editorial decisions must “take into account philosophy, politics, economics, pedagogy and other social aspects of health care” (1977) and that medical journals must include “exposure and discussion of important issues that involve, even indirectly, health and medicine” (1999a). This ambition to participate in contemporary social and political debates is reflected in the Journal’s engagement with health care reform today.

The Journal’s mediation between biomedical science and the social context of health care demonstrates another role of medical journals: their defining of the medical profession as a social and moral community. The Journal’s early editors recognized that general medical journals “furnish a bond of union and sympathy between the members of our profession which nothing else can supply” (1865). Journals have provided a forum for community news, announcements, and obituaries of exemplars of professional conduct. Such ephemera often convey important norms for practicing physicians. When the Titanic sank, editorials celebrated the ship’s surgeons, who died at their posts: “The heroism of these two physicians, though no more than their recognized duty, is worthy cause for gratitude and gratification to the profession whose representatives they were and to whose ideals they were so loyally true” (1912a).

Sometimes the Journal published travelogues, from Jacob Bigelow’s ascent of Mount Washington in 1816 to Cecil Austin’s Norwegian vacations in 1909. Such accounts stressed that doctors,
even at leisure, were men of science, attentive to the geology, botany, zoology, and anthropology of places they visited. Vestiges of this genre can be seen in the photographs by physicians published today. Though used primarily “to fill otherwise empty space” (1999c), they reflect the identities and aspirations of a community of readers. When taken by physicians living in distant locales, they also reflect the globalization of the Journal and medical science.

Medical journals have occasionally explicitly focused on community norms. The Journal has famously taken stands on research ethics, from Henry Beecher’s influential 1966 exposé of unethical research (1966b) to debates about standards for conducting clinical trials in developing countries (1997a, 1997b). Journals have also addressed the changing ethics of clinical practice. The Journal published a 1906 plea to reconsider reflexive prohibitions against euthanasia, Timothy Quill’s moving 1991 account of his decision to assist in the suicide of a patient dying from leukemia, and continuing coverage of debates about physician-assisted death (2008). It has provided a forum for debates about the merits of truth telling and withholding information from patients, whether in cases of medical uncertainty (1914) or medical error (2007). Indeed, the moral responsibility to discuss errors — within the medical community — motivated the 1923 decision to publish the Case Records of the Massachusetts General Hospital. These were presented in part to provide “comfort” to physicians struggling in isolation with difficult cases. “Harbors are made safer for mariners not by records of prosperous voyages,” a 1923 editorial noted, “but by buoying the dangerous reefs and sunken ledges that have caused disasters. If for nothing else, these Case Records are of exceptional value because of their honest acknowledgment of mistakes.”

CHALLENGES THEN AND NOW
To survive, medical journals have relied on sponsors (medical societies or colleges) and advertisers (pharmaceutical and otherwise), forming relationships sometimes seen as Faustian bargains. Although medical society sponsorship might seem benign, editors have often bridled at compromising their autonomy. In 1884, Connor dreamed of a time when journals could be free from sponsors who “cripple their perfect independence to work, and think, and speak for the great masses of the medical profession.”

Yet many prominent medical journals remain under the control of medical societies. In the past 15 years, tensions between editors and overseers have been implicated in editor turnover at JAMA (1999b), the New England Journal of Medicine (1999d), and the Canadian Medical Association Journal (2006).

The relationship with advertisers has been no easier. Connor decried the “prostitution of the reading pages to the supposed interests of the advertising columns.” Garland, in 1957, lamented that the medical journal remained dangerously dependent on pharmaceutical advertisements for its financial lifeblood. Nevertheless, Garland understood that the problem could not be resolved by abstinence: “We as physicians recognize and appreciate our ancient partnership and our friendly relations with our dynamic friends, the manufacturing apothecaries.” He continued, more pointedly, that “as editors we delight in the revenue from their adventures in competitive advertising, even as we seek to put restraints on its ultrapersuasiveness and keep it within the bounds that medical propriety and a sense of service to humanity have set.” Yet constant vigilance has been required to maintain firewalls between promotion and education (1992).

Even when survival seemed sure, journals have struggled to make themselves heard amid the competing literature. Ever since Billings assembled the Index Medicus, writers have used the language of pathology to critique the vastness of the journal literature. Even before medical research dramatically expanded after World War II, the Lancet reported in 1935 on the growing “journalistic blastoma.” Four years later, Sir Robert Hutchinson described the “enormous proliferation, this pullulating fungus-like growth,” that threatened...
the medical community’s vitality. Rather than “see science suffocated in its own secretions,” Hutchinson prescribed journalistic “birth control.” Journals remained nonadherent.

By the late 20th century, journals needed to compete not just with each other but with newspapers and other media. As Franz Ingelfinger noted in 1977, “Medicine has become the stuff of headlines.” New topics drew further attention to medical journals — debates about health policy and national medical insurance, malpractice, special-interest lobbying regarding particular diseases, and interest in health education. The growing market in medical news attuned journal editors to their content’s “newsworthiness.” In 1969, the Journal articulated this relationship in its Ingelfinger Rule, a policy against publishing anything that had already appeared elsewhere. Other journals followed suit. This rule, combined with embargo policies, has led to a carefully choreographed production in which medical journals and the popular press work cooperatively and competitively to influence the news cycle.

New media, too, have presented challenges. In the 1960s, the National Institutes of Health (NIH) experimented with Information Exchange Groups, letting scientists communicate directly with one another by sending mimeographs to colleagues. From a group of 32 membrane-biology researchers in 1961, the program expanded to 3600 members 5 years later. Many saw in this mechanism the inevitable demise of medical journals. A former Lancet editor predicted, “A day will come when journals will be superseded as a means of publishing new research.” But the NIH closed the program in 1966, owing to insufficient funds for mimeographs and mailings and concerns about the absence of refereeing and excessive emphasis on priority.

The Internet and other social media technologies have enabled new experiments. Open-access journals shift publication costs to authors so that research results may be freely shared globally. Blogs, listserves, and Twitter feeds permit everyone an authoritative voice and increasingly influence what counts as relevant medical information. Journalists and bloggers routinely cover major medical conferences, broadcasting promising discoveries before research has been submitted for peer review and publication. Many of these journal-bypassing models assert a morality that contrasts the virtues of open access with the centralized power and profits of traditional publishers. The optimal role of the medical journal in this environment remains undetermined.

THE FUTURE

For two centuries, medical journals have mediated the production and dissemination of medical knowledge, their histories inextricably entwined with that of the medical system. Despite the emergence of rivals that were unimaginable a generation ago, the Journal and its peers maintain substantial influence on both the medical profession and society. But anxiety remains: What does the future hold?

Navigating many obstacles, medical journals have persisted, adapting to changing environments and embracing new opportunities. Elaborate websites now offer Web-only content, including audio, video, and other once-impossible formats. Journals must continue to manage not just the shifting landscape of the production and publishing of medical knowledge, but also the broader currents of their social, economic, and political contexts. This responsibility cannot be borne by editors alone. Readers’ decisions — which subscriptions they keep, which free content they browse — will shape the decisions of journal editors, sponsors, and advertisers. Priorities and prospects for journals in the future depend on how, and how well, physicians use them today.

Disclosure forms provided by the authors are available with the full text of this article at NEJM.org.

From the Department of Global Health and Social Medicine, Harvard Medical School (S.H.P., J.A.G., D.S.J.); the Center for the History of Medicine, Countway Library of Medicine (S.H.P.); and the Division of Pharmacoepidemiology and Pharmacoeconomics, Brigham and Women’s Hospital (J.A.G.) — all in Boston; and the Department of the History of Science, Harvard University, Cambridge, MA (J.A.G., D.S.J.).


Copyright © 2012 Massachusetts Medical Society.
Two Hundred Years of Surgery
Atul Gawande, M.D., M.P.H.

Surgery is a profession defined by its authority to cure by means of bodily invasion. The brutality and risks of opening a living person's body have long been apparent, the benefits only slowly and haltingly worked out. Nonetheless, over the past two centuries, surgery has become radically more effective, and its violence substantially reduced — changes that have proved central to the development of mankind's abilities to heal the sick.

Surgery Before the Advent of Anesthesia

The first volume of the New England Journal of Medicine and Surgery, and the Collateral Branches of Science, published in 1812, gives a sense of the constraints faced by surgeons, and the mettle required of patients, in the era before anesthesia and antisepsis. In the April issue for that year, John Collins Warren, surgeon at the Massachusetts General Hospital and son of one of the founders of Harvard Medical School, published a case report describing a new approach to the treatment of cataracts. Until that time, the prevalent method of cataract treatment was “couching,” a procedure that involved inserting a curved needle into the orbit and using it to push the clouded lens back and out of the line of sight. Warren's patient had undergone six such attempts without lasting success and was now blind. Warren undertook a more radical and invasive procedure — actual removal of the left cataract. He described the operation, performed before the students of Harvard Medical School, as follows:

The eye-lids were separated by the thumb and finger of the left hand, and then, a broad cornea knife was pushed through the cornea at the outer angle of the eye, till its point approached the opposite side of the cornea. The knife was then withdrawn, and the aqueous humour being discharged, was immediately followed by a protrusion of the iris.

Into the collapsed orbit of this unanesthetized man, Warren inserted forceps he had made especially for the event. However, he encountered difficulties that necessitated improvisation:

The opaque body eluding the grasp of the forceps, a fine hook was passed through the pupil, and fixed in the thickened capsule, which was immediately drawn out entire. This substance was quite firm, about half a line in thickness, a line in diameter, and had a pearly whiteness.

A bandage was applied, instructions on cleansing the eye were given, and the gentleman was sent home. Two months later, Warren noted, inflammation required “two or three bleedings,” but “the patient is now well, and sees to distinguish every object with the left eye.”

The implicit encouragement in Warren's article, and in others like it, was to be daring, even pitiless, in attacking problems of an anatomical nature. As the 18th-
century surgeon William Hunter had told his students, “Anatomy is the Basis of Surgery, it informs the Head, guides the hand, and familiarizes the heart to a kind of necessary inhumanity.” That first volume of the Journal provided descriptions of a remarkable range of surgical techniques, including those for removing kidney, bladder, and urethral stones; dilating the male urethra when strickent by the passage of stones; tying off aneurysms of the iliac artery and infrarenal aorta; treating burns; and using leeches for bloodletting. There were articles on the problem of “the ulcerated uterus” and on the management of gunshot and cannonball wounds, not to mention a spirited debate on whether the wind of a passing cannonball alone was sufficient to cause serious soft-tissue injury.

Surgery, nonetheless, remained a limited profession. Pain and the always looming problem of infection restricted the extent of a surgeon’s reach. Entering the abdomen, for instance, was regarded with reproach — attempts had proved almost uniformly fatal. The chest and joints were also out of reach. The primary remit of surgery was therefore the management of external conditions, and medicine dealt with the internal ones (hence the term “internal medicine,” which persists to this day). Even for those conditions that appeared to be externally accessible, surgical accounts often spoke of failure more than derring-do. For example, in an article on spina bifida that appeared in the January 1812 issue of the Journal, a surgeon noted the uniform fatality of the condition and recounted an effort to repeatedly lance, drain, and bandage an infant’s meningocele, which proved to be utterly futile. The skin “had become thickened, and as inelastic . . . as the upper leather of a shoe; it also ulcerated,” the author wrote. “Pus was formed in the sac, and the infant died.” Such reports often maintained an almost defiant optimism. (“We have no doubt,” this surgeon concluded, “that if performed with due caution, a technique of draining meningoceles will be engineered and “the disease of Spina Bifida may cease to be an opprobrium of medicine.”) Nonetheless, breakthrough surgical successes were, for a long time, few and far between.

They were also often illusory. In 1831, for instance, a Mr. Preston reported in the Journal his treatment of a man with an acute stroke that had resulted in left hemiparesis and speech difficulties. He did not use the usual, ineffective method of bloodletting and applying leeches but instead decided to take the curious approach of ligating the patient’s right common carotid artery. Preston conjectured that by diminishing the supply of blood to the affected side of the brain, the treatment would reduce congestion and inflammation. By luck, the man survived. He was discharged 1 month later, walking with the aid of a stick and speaking normally, leading Preston to propose that surgeons might consider tying both carotids in future cases. Fortunately, his case notwithstanding, the procedure failed to catch on.

The crucial spark of transformation — the moment that changed not just the future of surgery but of medicine as a whole — was the publication on November 18, 1846, of Henry Jacob Bigelow’s groundbreaking report, “Insensibility during Surgical Operations Produced by Inhalation” (Fig. 1). The opening sentences crisply summarized the achievement: “It has long been an important problem in medical science to devise some method of mitigating the pain of surgical operations. An efficient agent for this purpose has at length been discovered.” Bigelow described how William Morton, a Boston dentist, had administered to his own patients, and then to several more who had undergone surgery at the Massachusetts General Hospital, a gas he called “Letheon,” which successfully rendered them insensible to pain. Morton had patented the composition of the gas and kept it a secret even from the surgeons. Bigelow revealed, however, that he could smell ether in it. The news burst across the world. The Letters to the Editor pages were occupied for months with charges and countercharges over Bigelow’s defense of Morton’s secrecy and credit for the discovery. Meanwhile, ether anesthesia rapidly revolutionized surgery — how it was practiced, what could be attempted with its use, and even what it sounded like.

Consider, for instance, amputation of the leg. The procedure had long been recognized as lifesaving, in particular for compound fractures and other wounds prone to sepsis, and at the same time horrific. Before the discovery of anesthesia, orderlies pinned the patient down while an assistant exerted pressure on the femoral artery or applied a tourniquet on the upper thigh (Fig. 2A, upper drawing). Surgeons using the circular method proceeded through the limb in layers, taking a long curved knife in a circle through the skin first, then, a few inches higher up, through the muscle, and finally, with the assistant retracting the muscle to expose the bone a few inches higher still, taking an amputation saw smoothly through the
The horror of great darkness, and the sense of desertion by God and man, bordering close on despair, which swept through my mind and overwhelmed my heart, I can never forget, however gladly I would do so. During the operation, in spite of the pain it occasioned, my senses were preternaturally acute, as I have been told they generally are in patients in such circumstances. I still recall with unwelcome vividness the spreading out of the instruments: the twisting of the tourniquet: the first incision: the fingering of the sawed bone: the sponge pressed on the flap: the tying of the blood-vessels: the stitching of the skin: the bloody dismembered limb lying on the floor.

Before anesthesia, the sounds of patients thrashing and screaming filled operating rooms. So, from the first use of surgical anesthesia, observers were struck by the stillness and silence. In London, Liston called ether anesthesia a “Yankee dodge” — having seen fads such as hypnotism come and go — but he tried it nonetheless, performing the first amputation with the use of anesthesia, in a 36-year-old butler with a septic knee, 2 months after the publication of Bigelow’s report. As the historian Richard Hollingham recounts, from the case records, a rubber tube was connected to a flask of ether gas, and the patient was told to breathe through it for 2 or 3 minutes. He became motionless and quiet. Throughout the procedure, he did not make a sound or even groan. “When are you going to begin?” asked the patient a few moments later. He had felt nothing. “This Yankee dodge beats mesmerism hollow,” Liston exclaimed.

It would take a little while for surgeons to discover that the use of anesthesia allowed them time to be meticulous. Despite the advantages of anesthesia, Liston, like many other surgeons, proceeded in his usual lightning-quick and bloody way. Spectators in the operating-theater gallery would still get out their pocket watches to time him. The butler’s operation, for instance, took an astonishing 25 seconds from incision to wound closure. (Liston operated so fast that he once accidentally amputated an assistant’s fingers along with a patient’s leg, according to Hollingham. The patient and the assistant both died of sepsis, and a spectator reportedly died of shock, resulting in the only known procedure with a 300% mortality.)

bone so as not to leave splintered protrusions (Fig. 2A, lower drawing). Surgeons using the flap method, popularized by the British surgeon Robert Liston, stabbed through the skin and muscle close to the bone and cut swiftly through at an oblique angle on one side so as to leave a flap covering the stump (Fig. 2B).

The limits of patients’ tolerance for pain forced surgeons to choose slashing speed over precision. With either the flap method or the circular method, amputation could be accomplished in less than a minute, though the subsequent ligation of the severed blood vessels and suturing of the muscle and skin over the stump sometimes required 20 or 30 minutes when performed by less experienced surgeons. No matter how swiftly the amputation was performed, however, the suffering that patients experienced was terrible. Few were able to put it into words. Among those who did was Professor George Wilson. In 1843, he underwent a Syme amputation — ankle disarticulation — performed by the great surgeon James Syme himself. Four years later, when opponents of anesthetic agents attempted to dismiss them as “needless luxuries,” Wilson felt obliged to pen a description of his experience:

Figure 1. Operation Being Performed with the Use of Ether Anesthesia.
This daguerreotype was taken in the spring of 1847 by Josiah Hawes in the Operating Room (now known as the Ether Dome) of the Massachusetts General Hospital. The first public demonstration of surgical anesthesia occurred in the same room on October 16, 1846, presided over by the surgeon John Collins Warren, seen here touching the patient. Although it is believed that a photographer was present during the first event as well, he took no pictures because the sight of blood made him nauseated. Courtesy of the Massachusetts General Hospital, Archives and Special Collections.
A NEW ERA OF ANESTHESIA AND ANTISEPSIS

Surgeons soon found, however, that anesthesia allowed them to perform more complex, invasive, and precise maneuvers than they had dared to attempt before. Within a decade, for instance, the first successful hysterectomy and bilateral ovariotomy — removal of massive ovarian cysts weighing several pounds — proved that the abdomen could be safely penetrated. Further experiments revealed other effective anesthetics: nitrous oxide, chloroform, and eventually halothane and other nonvolatile agents. Narcotics such as laudanum were found to relieve postoperative suffering. Suddenly, pain was no longer a barrier to surgical capability.

A second major barrier persisted, however: sepsis. The mortality associated with ovariotomy and other types of major abdominal surgery, repair of open fractures, and limb amputation commonly remained at 50% or higher owing to infection. One therefore might have thought that the news of Joseph Lister’s landmark series of articles in The Lancet in 1867 describing the effectiveness of his new system of antisepsis with the use of carbolic acid would be received with the same fanfare as the report on ether anesthesia had been (Fig. 3). Instead, it was regarded with overwhelming skepticism. The Journal first mentions Lister’s breakthrough as a method that was neither original nor beneficial. Nearly a decade later, a surgeon writing in the Journal on the dressing of wounds could still insist “that there is good reason to believe that the theory of M. Pasteur, upon which Lister bases his treatment, is unsound.” Ignaz Semmelweis, the Viennese obstetrician who in 1847 had found that hand washing by birth attendants eliminated puerperal sepsis, the leading cause of maternal death, was not even mentioned in the Journal until the end of the 19th century. J.M.T. Finney recalled his experience as a house officer at the Massachusetts General Hospital in the 1880s: “The operating surgeon was usually garbed in a black Prince Albert coat, kept hanging in a closet for the occasion and showing numerous evidences of previous operations in the way of dried blood, wound secretions, etc.” For decades, hand washing and skin cleansing remained routinely perfunctory.

Some surgeons, however, especially younger ones, began accepting the diligence required for aseptic and antiseptic practice. Such practice, along with effective anesthesia, led them to hitherto unimagined treatments and discoveries. In 1868, for instance, John Stough Bobbs reported on a 30-year-old woman with a large, painful right abdominal mass presumed to be an ovarian cyst. Chloroform allowed him to explore her abdomen through a 4-in. incision. Sweeping the omental adhesions aside with a finger, he encountered a 5-in.—wide, smooth-walled, oval tumor. When he cut through the wall of the tumor, “a perfectly limpid fluid escaped, propelling, with considerable force, several solid bodies about the size of ordinary rifle bullets.” He drained the sac of its fluid, extracted some 50 concretions that ranged in size “from that of a mustard seed to that of a bullet,” and then closed the incision, uncertain what the concretions were. The patient recovered uneventfully with the help of laudanum and lem-
onade. Only later, when he carefully examined the smooth, mahogany-colored, irregularly spherical objects that he had extracted, did Bobbs understand what he had encountered. They were gallstones. The absence of green–yellow bile in the sac had confused him — the clear, mucoid fluid would come to be known as “white bile” — but he had, in fact, performed the world’s first successful gallbladder operation.

A slew of “firsts” rapidly followed, each more daring than the last. In 1880, Tait performed the first transabdominal resection of a gangrenous appendix, and Rehn performed the first subto-tal thyroidectomy for Graves’ disease. In 1884, Bennett and Godlee reported the first successful removal of a brain tumor. Methods for supra-pubic prostatectomy, total gastrectomy, chest surgery, and joint repair were worked out. Alexis Carrel devised methods for suturing blood vessels and performing surgical grafts that became the foundation for the field of vascular surgery and won him a Nobel Prize in 1912. Surgeons developed such skill and confidence that they began performing exploratory laparotomies simply for the purpose of diagnosis. Indeed, articles raising concern that there were perhaps too many laparotomies began to appear by the turn of the century. The key barriers to surgical knowledge and imagination were gone.

Until this time, surgical discoveries had provided only halting contributions to the capabilities of the medical community. In the early part of the 19th century, just one fifth of the Journal’s scientific articles were surgical in nature (if one takes as a guide the review and classification of each article in the first volume for each decade, beginning with 1812) (Fig. 4). Surgery had been, one might politely say, a modest contributor to medical progress. Between the mid-1800s and the 1920s, however, the coverage of surgical advances took up half the Journal. Physicians in the Victorian era had few effective drugs, but surgeons began reporting new treatments almost monthly, and the breakneck pace of innovation continued for nearly a century. Surgery became a dominant force in medical advancement.

Surgery also began to progress through an increasingly important process of refinement and professionalization. William Halsted introduced and popularized the use of rubber gloves to help prevent infection. Care of burns and other wounds was made radically simpler and less traumatizing. Anesthetic techniques and apparatus were being made more reliable and sophisticated. And in 1917, the American College of Surgeons founded the Hospital Standardization Program (later renamed the Joint Commission for the Accreditation of Hospitals) to shift the role of hospitals from serving primarily as a place for the convalescence of the sick poor to providing safe and effective care for patients undergoing surgery.

Specialization was likewise an important force. Historians continue to debate whether the growth in knowledge drove specialization or specialization led to the growth in knowledge. (There are numerous examples of each.) Nonetheless, in 1905, the Long Island Society of Anesthetists was formed, which would evolve into the American Society of Anesthesiologists. After World War I, national associations were formed for neurologic surgeons, orthopedic surgeons, urologists, and other specialists, and dedicated training programs were established. Surgery — the invasion of people’s bodies for cure — was becoming normalized.

Much of the story appeared only obliquely in the Journal. But that too reflected the changing nature of progress. The milestone reports of the new era often seemed obscure when they first ap-
appeared. Werner Forssmann, a 25-year-old surgical intern in Eberswalde, Germany, published his report on successful catheterization of the heart in a German medical journal in 1929.\(^{30}\) Forbidden by his professors from attempting the experiment on either animals or patients (they thought the idea preposterous and dangerous), he performed it on himself (Fig. 5). His investigation would eventually lead to the creation of the field of cardiology and win Forssmann the Nobel Prize in Medicine. Yet it was more than a decade before anyone recognized the significance of his report. Likewise, anesthetist resident David Massa’s ingenious creation of the plastic intravenous catheter probably seemed a minor innovation to many at the time. A description of his device appeared in 1950 in the Proceedings of the Staff Meetings of the Mayo Clinic under the modest title “A Plastic Needle,” and it was not until the 1960s that this type of catheter became well known.\(^{30}\) Eventually, however, it transformed the approach to patient resuscitation.

The field of surgery, with its ethos of radical action and perfectionism refinement, defined much of medical culture in the early 20th century. By midcentury, however, surgery’s outsize role and influence began to subside. Whereas its discoveries had taken up half the space in the journal in 1922, the proportion fell to one third during the next decade. By the 1950s, reports of new diagnostic tests, vaccines, antibiotics, and other innovations from the wet laboratory dominated the pages of the journal. Scientists had found an even more prolific source of discovery than the operating room: the laboratory bench. With the advent of chemotherapeutics, molecular medicine, and other technologies, surgery was no longer the driving force behind medical breakthroughs. Since 1972, just a tenth of the journal’s articles have been devoted to surgical advances.

To be sure, the field of surgery continued to register a steady stream of seminal breakthroughs. This was the era in which the heart was conquered, after all. In 1948, Dwight Harken and colleagues published an astonishing report describing the successful surgical treatment of mitral-valve disease\(^{31}\); in 1945, Blalock and Taussig designed their shunt operation for “blue babies”\(^{32}\); Robert Gross and colleagues reported in 1952 on open-heart surgery to close atrial septal defects in children\(^{33}\); and the development of cardiopulmonary-bypass technology allowed open-heart procedures to be carried out. Similarly, transplantation of organs between human beings — first kidneys, then livers, then hearts and lungs, and most recently, even faces — altered basic concepts about ourselves and led us to redefine death. Implantation of organs engineered in the laboratory from a person’s own cells is now being reported.\(^{34}\) Surgeons are still traversing remarkable frontiers.

But the most striking story of surgery in recent decades is how firmly it has become established as an essential tool for helping people live long and healthy lives. Virtually no one escapes having a condition for which effective treatment requires surgery — a serious orthopedic injury, a cataract, a tumor, obstructed labor, joint failure, severe cardiac disease. Today, surgeons have in their arsenal more than 2500 different procedures. Thus, the focus of recent advances in the field has been less on adding to the arsenal than on ensuring the successes of the treatments we have.

Minimization of the invasiveness of surgical procedures is an advance that is arguably as significant as the discovery of anesthesia. In recent decades, the advent of laparoscopy and thoracoscopy reduced the debilitating, half-meter-long abdominal and chest incisions to a half centimeter. The subsequent introduction of endoscopic and percutaneous techniques has turned incisions into mere puncture wounds. Gallbladder surgery, resection of colonic polyps and ovarian tumors, and

![Figure 4: Changes in the Proportion of Articles on Surgery Published in the Journal since 1812.](image-url)
Figure 5. Radiograph of the Heart Self-Catheterized by a Surgical Intern in 1929.

The radiograph shows successful self-catheterization of the heart, performed by Werner Forssmann, at the time a 25-year-old surgical intern in Eberswalde, Germany.

lungs biopsies have become outpatient procedures. We are now in an era in which a teenage boy can undergo reoperation for repair of a severe coarctation of his thoracic aorta percutaneously on a Thursday and be well enough to sprain his ankle playing sports the following Saturday (as my son did not long ago). The technological refinement of our abilities to manipulate the human body has been nothing short of miraculous.

The increased safety and ease of surgery have produced an explosion in the volume of operations being performed — to at least 50 million annually in the United States alone. At the present rate, the average American can expect to undergo seven operations during his or her lifetime. This profound evolution has brought new societal concerns, including how to ensure the quality and appropriateness of the procedures performed, how to make certain that patients have access to needed surgical care nationally and internationally, and how to manage the immense costs. As early as the 1970s, researchers began documenting substantial rates of fatal errors in surgical care, wide differences in outcomes among institutions, and large disparities in access to care both within the United States and between countries. The science of effectively routinizing surgery for mass populations is still in its infancy, as it is for all areas of medicine. The Journal is entering its third century of publication, yet we are still unsure how to measure surgical care and its results. Experiments in the delivery of care will probably provide the next major advancement in the field of surgery.

Meanwhile, the practice of surgery itself will continue to change. Prognostication is a hazardous enterprise. But if the past quarter century has brought minimally invasive procedures, the next may bring the elimination of invasion. One feels foolish using terms like nanotechnology — I haven’t the slightest idea what it really means or can do — but scientists are already experimenting with techniques for combining noninvasive ways of seeing into the body through the manipulation of small-scale devices that can be injected or swallowed. Surgical work will probably even become fully automated.

The possibilities are tantalizing. A century into the future, a surgeon will tell the tale — that is, if the world still makes such people.

Disclosure forms provided by the author are available with the full text of this article at NEJM.org.

I thank Ami Karlage for assistance with the historical research.

REFERENCES

29. Forssmann W. Die Sondierung des rechten Herzens. (Probing the right heart.) Klin Wochenschr 1929;8:2085-7. (In German.)
Copyright © 2012 Massachusetts Medical Society.
The primary care doctor is a rapidly evolving species — and in the future could become an endangered one. As the United States grapples with the dual challenges of making health care more widely available and reducing the national price tag, it’s hard to say how primary care physicians will fit into the delivery models that emerge. Will they be increasingly replaced by nurse practitioners and physician assistants? Will they become partners or leaders on multidisciplinary teams, spending more time supervising others and less interacting with patients? Will most become employees of large health systems, as solo and small-group practices disappear? Will having a primary care physician become a luxury, available chiefly to people who can pay a premium to enroll in a concierge practice?

Even the question of whether the country faces an impending doctor shortage is debatable: groups of experts have reached opposite conclusions depending on their assumptions about who will be delivering care in the future and how. The report of a conference held in May 2011, sponsored by the Josiah Macy Jr. Foundation, cited estimates predicting a shortage of more than 100,000 physicians by the middle of the next decade, with primary care specialties most affected.¹ Meanwhile, another report, published last year by the Robert Wood Johnson Foundation, concluded, “Data do not support the suggestion that the United States is currently experiencing or facing an imminent shortage of primary care providers; numbers of physicians, nurse practitioners and physician assistants have grown in recent years relative to the general population.”² That report, however, stated that such providers are profoundly maldistributed, resulting in severe shortages in rural areas and among underserved populations. It noted that nurse practitioners are the fastest-growing group of primary care practitioners, their numbers having grown by an average of more than 9% per year relative to the population in the 6 years ending in 2005.

Changes in Doctors’ Work

The doctor’s duties are changing, influenced by advances in medical knowledge and technology; the increasing use of computers, handheld devices, and electronic...
Despite the shifting professional landscape, national physician surveys conducted by the Center for Studying Health System Change, a nonpartisan, nonprofit organization in Washington, D.C., show no significant increase during the past decade in the percentage of doctors who say they’re dissatisfied with medicine, nor are there significant differences in levels of satisfaction between primary care physicians and specialists. In 2008, 81.5% of physicians surveyed said they were somewhat or very satisfied with their work, and 14.7% said they were somewhat or very dissatisfied. The surveys also fail to show any consistent change since the mid-1990s in the percentage of physicians indicating that they don’t have enough time to spend with patients.

David Mechanic, a professor of behavioral sciences and director of the Institute for Health, Health Care Policy, and Aging Research at Rutgers University, reported a decade ago that despite the widespread impression among physicians and patients that medical visits had gotten progressively shorter, the average visit length increased by 1 to 2 minutes between 1989 and 1998. However, in recent years, the pressure on physicians to move quickly and accomplish multiple goals during a visit has intensified, Mechanic said. In surveys, many patients describe their doctors as hurried and unresponsive, whereas nurse practitioners are viewed as “much more willing to take time, talk to the patient, and answer questions,” he added. In one study, researchers who videotaped outpatient physician–patient visits and audiotaped patients afterward found that the most highly satisfied patients perceived the visit as having lasted 3 minutes longer, on average, than it really had.

Observers of medical practices report striking changes in how doctors work. Physicians Christine Sinsky and Thomas Bodenheimer visited 23 primary care practices during the past year for a project, sponsored by the American Board of Internal Medicine Foundation, focused on ways of sustaining primary care doctors’ joy in practice. Sinsky, a primary care internist at a large multispecialty group practice in Iowa, commented that the electronic health record, despite its obvious benefits, has been “an enormously disruptive innovation,” altering the practice environment and the daily tasks performed by physicians and other health care workers.

“What I’ve really seen is a lot of waste within the health care system at the level of utilization of physician skills,” she said. “I think two thirds of many physicians’ days are spent on documentation, administrative tasks, paperwork completion, rote inbox management, data gathering, and data entry. . . . It’s something that’s hard to recognize when you’re the one doing it.”

Sinsky noted that some innovative practices have responded by assigning much of the responsibility for data entry to other staff members. For example, in one more collaborative model of care, a nurse or medical assistant accompanies the physician during each visit, enters the findings and treatment plan into the computer, and prepares prescriptions and instructions for the patient.

Abraham Verghese, a professor of medicine at Stanford University, believes that the ubiquity of computers in clinical settings has contributed to a decrease in face-to-face interactions between doctors and patients and an erosion of physicians’ skill in the physical examination. Teams of residents may conduct rounds by checking test results on a computer and entirely omit examining patients. Often during patient encounters, “the physician’s back is turned because he or she is entering data into the computer,” he noted. Yet for patients, talking with a doctor and being examined has both ritual and therapeutic value, in addition to its importance for gathering information.

“I love technology, but I think
we’re discovering that the physician–patient relationship is timeless,” Verghese said. “It cannot be abandoned because we have better tests and can do away with the human interface.”

Verghese has been campaigning to expand the teaching of bedside diagnostic skills during medical school’s clinical years and argues that newly trained physicians should have to undergo a “high-stakes clinical examination,” similar to those required in some other countries, as a condition of board certification in internal medicine or family practice. Medical schools now routinely evaluate students’ communication skills. A Chicago couple recently donated $42 million to the University of Chicago Medical Center to establish an institute to improve the doctor–patient relationship. Even when encounters are brief, “you can actually train physicians to do interpersonal things that make patients feel they’re more committed to their welfare,” said Rutgers’ Mechanic.

Some evidence suggests that primary care doctors’ scope of practice is narrowing. In a recent survey of family practice physicians by the American Board of Family Medicine (ABFM), 80 to 90% reported spending no time on preoperative, postoperative, or maternity care; 60% said they spent no time on office surgery; 55% said they spent no time on mental health; and more than 25% said they did not see children. In a letter e-mailed to members of the specialty to report these findings, Warren Newton, the ABFM chair and vice dean for education at the University of North Carolina School of Medicine, noted that the data were preliminary but prompted the question, “Why is it happening and what should we do about it?”

Newton explained in an interview that the data are drawn from online surveys that the country’s 70,000 board-certified family physicians must regularly complete to maintain their certification. He noted that the proportion of family physicians who report caring for children has declined sharply in recent years, a trend that may be related to the increase in the number of doctors employed by integrated health systems. Since general internists are in shorter supply than pediatricians, health systems are more likely to deploy family physicians to care for adults, he said. He suggested that the survey’s findings may underestimate the proportion of family doctors who treat mental health problems, because physicians may assume that the question doesn’t include the prescribing of drugs to treat depression.

With the proliferation of hospitalists and the rapid expansion of medical knowledge, the scope of practice of many internists has also narrowed, Sinsky noted. “In my own practice and others, there may be early referrals for things that could be managed locally with stronger channels of communication,” she said. Sinsky added that she has tried to maintain breadth in her practice by learning to perform procedures, such as joint injections and endometrial biopsies, and by cultivating relationships with specialists whom she can call for advice, enabling her to manage more conditions without referral.

**Emphasis on Team-Based Care**

The health care reform law has focused national attention on the patient-centered medical home and the concept of team-based care. For several years, Colorado’s family practice residencies have been training young physicians in team-based care, especially by having them collaborate with mental health professionals and pharmacists, said Frank deGruy, professor and chairman of family medicine at the University of Colorado School of Medicine. More recently, the state’s family practice programs have also expanded their focus on practice design and...
management, teaching “how you put together your care team so the people who you have available are practicing at the top of their license, communicate well, and collectively create a practice that continuously measures and improves its level of quality,” he said. Trainees have responded enthusiastically to the new approach.

“I would say the morale of the family doctors out there in the field is pretty much in the toilet, whereas the morale of the residents . . . and the newly minted family doctors is good,” deGruy said. “We’ve got people coming out who think they can change the world.”

Team-based care is the norm in some of Colorado’s large urban health systems, such as Kaiser Permanente and Denver Health, deGruy said. However, in private practice settings, the economics of the reimbursement system and a shortage of medical specialists and other services in rural areas make it difficult for physicians to apply the team-based model even when they have been trained to do so. “The biggest problem” for primary care doctors entering practice is that “it’s still really hard to make it financially,” deGruy said. Primary care physicians are “trying to do the right thing in the face of very strong incentives to do the wrong thing.”

Faced with an aging patient population with mounting medical needs, North Carolina internist Douglas Kelling radically reinvented his own practice in 1994, becoming a pioneer of the team-based model. At that time, Kelling was a solo practitioner in Concord, North Carolina, employed by the local hospital, which supports his practice. Today, he leads a greatly expanded practice that includes two other physicians, six physician assistants, and two doctors of pharmacy, as well as a case worker, a discharge planner, and other staff. In the past, it has also included nurse practitioners. He sees patients 4½ days a week in his office, manages the care of hospitalized patients, and spends his evenings and weekends developing and updating the written protocols that specify the practice’s standard treatment plans for various medical conditions. When accepted national standards exist, he applies them. When they don’t, he does the research necessary to devise an evidence-based protocol.

“We’ve found that 98 or 99% of all the patients, with our systems and pathways, can be managed by the physician assistants and nurse practitioners,” he said. “They’re empowered to go ahead and do the obvious. That frees me up to do what I enjoy most — . . . to take care of patients where there’s a gray area, where there’s no best way to handle it.”

Whether a patient sees Kelling or a physician assistant as a primary care provider, each visit is also likely to include interactions with other team members. For example, a patient taking warfarin would see a doctor of pharmacy for monitoring and adjustment of that medication, and patients with hypertension might speak with a nurse who’s trained to counsel patients on following the Dietary Approaches to Stop Hypertension (DASH) diet. Another nurse assists uninsured patients in filling out enrollment paperwork for programs that provide free prescription medications. “It’s an encounter with the system,” Kelling said. “Although obviously seeing me is important, there are certain other aspects” of each patient’s care “that other staff members can handle better than I can.”

A fourth-generation physician, Kelling is proud of his practice’s effectiveness in caring for a large population, including patients with multiple chronic conditions. He said most residency training is “training to be the Lone Ranger, where one person goes in and is supposed to be all things to all people.” His own approach is “to make medical care more like NASCAR,” with the doctor as the driver and other team members responsible for the fuel and the tires. “You as the doctor are in charge, but unless you allow other people to do what they do best, you can never be successful.”

REFORMS IN EDUCATION
Producing the number and kind of primary care physicians that the country needs will require recruiting more diverse medical students — including more from underserved areas and populations — and reforming the system and expanding the settings for training health professionals, predicted George Thibault, president of the Macy Foundation. Thibault, a former professor of medicine and medical education at Harvard Medical School, recalled that last year a resident in primary care medicine at Massachusetts General Hospital told him that she wouldn’t be comfortable going into outpatient practice because she didn’t feel she’d learned the skills needed for ambulatory care. “We probably should be using the Kaisers and the health centers of the world more than we are now” as training settings, Thibault said. If residents “never leave the ICU . . . they’re never going to have
a comfort level to even imagine” working in a community setting.

He added that tracking of students who are interested in different career pathways should begin much earlier than it currently does. “Somebody who wants a rural setting needs to have experiences that prepare him for that — not working in a big inner-city clinic,” he said. “The training experiences need more to match the career pathway.” Both the duration and the content of medical school and residency training could be varied to reflect physicians’ career goals. The number and type of accredited programs for training physicians in various specialties should also more closely match the country’s needs, Thibault suggested.

The Macy Foundation is also focusing on the development of interprofessional education, funding grants at about 20 universities and health systems to pilot programs for teaching medical students, nursing students, and other health professionals how to work together in teams, beginning early in their education. Currently, “we professionalize everybody separately, and only when they’re fully formed do we do the mixing,” Thibault said. Each profession has its own culture, so “it’s not surprising that they don’t work well” together.

He believes regular training and experience working collaboratively with other professionals should be incorporated throughout medical school and residency. “You need to learn both to be a leader and to be a member of a team, because we’re all going to play this whole gradient of roles,” Thibault said. “I really do believe that we’ll never have the health care system we want and need unless we pay a lot more attention to how we’re training people to enter it.”

Disclosure forms provided by the author are available with the full text of this article at NEJM.org.

Dr. Okie is a medical journalist and a clinical assistant professor of family medicine at Georgetown University School of Medicine, Washington, DC.

Two Hundred Years of Cancer Research

Vincent T. DeVita, Jr., M.D., and Steven A. Rosenberg, M.D., Ph.D.

In the 200 years since the New England Journal of Medicine was founded, cancer has gone from a black box to a blueprint. During the first century of the Journal's publication, medical practitioners could observe tumors, weigh them, and measure them but had few tools to examine the workings within the cancer cell. A few astute observers were ahead of their time, including Rudolf Virchow, who with the benefit of a microscope deduced the cellular origin of cancer in 1863, and Stephen Paget, who in 1889 wisely mused about the seed-and-soil hypothesis of metastatic disease, a theory that is coming into its own today (Table 1). Other key advances were the discovery of a viral cause of avian cancer by Peyton Rous in 1911 and the proposal by Theodor Boveri in 1914 that cancer can be triggered by chromosomal mutations.

But the lid of the black box was not seriously pried open until 1944, when a retired scientist at Rockefeller University, Oswald Avery, reported the results of his beautifully clear experiments with the pneumococcal bacillus, which showed that cellular information was transmitted not by proteins but by DNA. His work led directly to the important discovery of the structure of DNA by Watson and Crick in 1953. Eight years later, the genetic code was broken by Nirenberg and colleagues, and the central dogma of biology was established; that information was transmitted from DNA to RNA and resulted in the synthesis of proteins. Then, the first of a series of totally unexpected discoveries disrupted this thinking, and we were reminded that things are not always what they seem in dealing with Mother Nature. The discovery of reverse transcriptase by Temin and Mizutani and Baltimore, which showed that information could be transmitted the other way, from RNA to DNA, had a profound influence on medicine but most particularly on cancer medicine.

Early investigators discovered that DNA is a very large molecule that was difficult to study in the laboratory. In 1970, Smith and Wilcox solved this problem by identifying enzymes that bacteria used defensively to cleave DNA at specific restriction sites. These discoveries gave birth to the molecular revolution and the biotechnology industry. They also paved the way for the sequencing of the genome.

This kind of science was expensive. The U.S. Congress partially addressed the problem by passing the National Cancer Act, which expanded the role of the National Cancer Institute (NCI), the first disease-oriented agency at the National Institutes of Health (NIH). The act, which was signed into law on December 23, 1971, by President Richard Nixon, created a new mandate for an NIH institute: “to support research and the application of the results of research to reduce the incidence, morbidity and mortality from cancer.” The emphasis on the application of the results of research was new; it had not been in the mission statement of the NIH. The act would quintuple the budget of the NCI by the end of the decade and provide the fuel for the revolution in molecular biology.

Although the enthusiasm in Congress for eradicating cancer was largely derived from excitement over a few clinical advances, about 85% of these new funds went to support basic research. At its peak in the early 1980s, the NCI accounted for 23%
of the budget of the NIH, yet it supported 53% of the research in molecular biology in the United States. And the results have been explosive. The discovery of genes that drive or suppress cellular growth and the complex regulation of signaling systems used by both normal cells and cancer cells to communicate with each other and their environment have brought the blueprint of cancer-cell machinery into bold relief (Table 1). The association of specific abnormalities with specific cancers has allowed scientists to identify persons who are at increased risk for common cancers, such as breast and colon cancer.

### Table 1. Singular Discoveries and Major Events in the Cancer Field and Changing Relative Survival Rates for Patients with Cancer in the United States, 1863–2006.*

<table>
<thead>
<tr>
<th>Year</th>
<th>Discovery or Event</th>
<th>Relative Survival Rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>1863</td>
<td>Cellular origin of cancer (Virchow)</td>
<td>1863</td>
</tr>
<tr>
<td>1889</td>
<td>Seed-and-soil hypothesis (Paget)</td>
<td>1889</td>
</tr>
<tr>
<td>1914</td>
<td>Chromosomal mutations in cancer (Boveri)</td>
<td>1914</td>
</tr>
<tr>
<td>1937</td>
<td>Founding of NCI</td>
<td>1937</td>
</tr>
<tr>
<td>1944</td>
<td>Transmission of cellular information by DNA (Avery)</td>
<td>1944</td>
</tr>
<tr>
<td>1950</td>
<td>Availability of cancer drugs through Cancer Chemotherapy National Service Center</td>
<td>1950</td>
</tr>
<tr>
<td>1953</td>
<td>Report on structure of DNA</td>
<td>1953</td>
</tr>
<tr>
<td>1961</td>
<td>Breaking of the genetic code</td>
<td>1961</td>
</tr>
<tr>
<td>1970</td>
<td>Reverse transcriptase</td>
<td>1970</td>
</tr>
<tr>
<td>1971</td>
<td>Restriction enzymes and Passage of National Cancer Act</td>
<td>1971</td>
</tr>
<tr>
<td>1975</td>
<td>Hybridomas and monoclonal antibodies tracking of cancer statistics by SEER program</td>
<td>1975</td>
</tr>
<tr>
<td>1976</td>
<td>Cellular origin of retroviral oncogenes</td>
<td>1976</td>
</tr>
<tr>
<td>1979</td>
<td>Epidermal growth factor and receptor</td>
<td>1979</td>
</tr>
<tr>
<td>1984</td>
<td>G proteins and cell signaling</td>
<td>1984</td>
</tr>
<tr>
<td>1986</td>
<td>Retinoblastoma gene</td>
<td>1986</td>
</tr>
<tr>
<td>1990</td>
<td>First decrease in cancer incidence and mortality</td>
<td>1990</td>
</tr>
<tr>
<td>1991</td>
<td>Association between mutation in APC gene and colorectal cancer</td>
<td>1991</td>
</tr>
<tr>
<td>1994</td>
<td>Genetic cancer syndromes and Association between BRCA1 and breast cancer</td>
<td>1994</td>
</tr>
<tr>
<td>2000</td>
<td>Sequencing of the human genome</td>
<td>2000</td>
</tr>
<tr>
<td>2002</td>
<td>Epigenetics in cancer and MicroRNAs in cancer</td>
<td>2002</td>
</tr>
<tr>
<td>2005</td>
<td>First decrease in total number of deaths from cancer</td>
<td>2005</td>
</tr>
<tr>
<td>2006</td>
<td>Tumor stromal interaction</td>
<td>2006</td>
</tr>
</tbody>
</table>

* Data are from the National Cancer Institute (NCI) Survival, Epidemiology, and End Results (SEER) program. APC denotes adenomatous polyposis coli.

Experiments that can be done in hours in the laboratory take months and years to replicate in the clinic, so clinical advances, though plentiful, develop slowly. Figures 1 and 2 depict the pace of change for the past two centuries in four areas: cancer treatment, chemoprevention, viruses and cancer-vaccine development, and tobacco control.

In the treatment of cancer, surgery was the first tool available. In 1809, Ephraim McDowell removed an ovarian tumor without the use of anesthesia, the first abdominal surgery performed in the United States, and provided evidence that tumor masses could be cured by surgery. The first public use of anesthesia, as reported by John Collins Warren in the Journal in 1846, and the introduction of antisepsis by Joseph Lister in 1867 paved the way for a cascade of surgical firsts in cancer treatment in the 19th and early 20th centuries. These innovative surgeons showed that any organ that was affected by cancer could be dealt with surgically.

The most profound influence on cancer surgery occurred in 1894, when William Halsted introduced radical mastectomy for breast cancer. Halsted based his operation on the supposition that breast cancer spread in a centrifugal fashion from the primary tumor to adjacent structures. He recommended en bloc resection of all surrounding tissue to remove all cancer cells, even the head of the humerus if it was involved. En bloc resection became known as “the cancer operation,” and it was applied to the removal of all other cancers, despite scant evidence supporting its use. It would be 74 years before the use of radical mastectomy and en bloc resection was questioned by another surgeon, Dr. Bernard Fisher. On the basis of experiments in rodent tumors, Fisher proposed that breast cancer had early access to the bloodstream and lymphatic tissues. Lymph-node involvement, he hypothesized, was merely an indication of generalized spread of disease. Radical mastectomy was both too much and too little: too much for small tumors and too little for large tumors that had already metastasized. In a series of clinical trials conducted by what is now called the National Surgical Adjuvant Breast and Bowel Project...
(NSABP), which Fisher led, he clearly showed that radical en bloc removal of tissue did nothing more than could be accomplished by removing the tumor mass itself, if surgery was supplemented by chemotherapy, radiation therapy, or both. Fisher also showed that less radical surgery plus chemotherapy or radiation therapy accomplished the goal with much less morbidity. These studies revolutionized the treatment of breast cancer. Since then, most other surgical procedures have been tailored to the availability of other treatments, and cancer surgery has become more effective, with less morbidity. In the first half of the 20th century, however, surgery was the only option, and a minority of patients could be cured by surgical removal of their tumors alone.

Figure 1. Timeline of Pivotal Events in Cancer Treatment.
CML denotes chronic myeloid leukemia.
The era of radiation treatment began in 1895, when Roentgen reported on his discovery of x-rays, and accelerated in 1898 with the discovery of radium by Pierre and Marie Curie. In 1928, it was shown that head and neck cancers could be cured by fractionated radiation treatments, a milestone in the field. The modern era of radiation therapy began in 1950 with the introduction of cobalt teletherapy. Since then, aided by advances in computing, the field has been driven by advances in technology that have allowed the therapeutic radiologist to deliver beam energy precisely to the tumor and to spare the normal tissue in the path of the radiation beam.

Figure 2. Timeline of Pivotal Events in Cancer Prevention.
BCG denotes bacille Calmette–Guérin, DCIS ductal carcinoma in situ, FDA Food and Drug Administration, and HPV human papillomavirus.
Like surgery, radiation therapy has become more effective, with less morbidity, and can be used in combination with other treatments.

By the 1950s, it had become apparent that no matter how complete the resection or how good the radiation therapy or how high the dose delivered, cure rates after surgery, radiation therapy, or the two combined had flattened out. Only about a third of all cancers could be cured by the use of these two treatment approaches, alone or together.

It was Paul Ehrlich at the turn of the 20th century who first made a concerted effort to develop chemicals to cure cancer. He coined the word “chemotherapy.” After animal models of transplantable tumors were developed in the early 20th century, researchers devoted the first half of the century to establishing screening systems that would reliably predict antitumor activity in humans on the basis of data from murine models. However, these efforts were largely unsuccessful. Part of the problem was the limited capability for testing new agents in humans. Two events provided optimism about the future of anticancer drugs: the use of nitrogen mustard in lymphomas at Yale in 1943 and Farber’s report in 1948 that folic acid antagonists could induce temporary remission in childhood leukemia. In 1955, these discoveries led to a national screening effort to develop and test anticancer drugs. Then the use of cancer chemotherapy, although shrouded in controversy, began in earnest. Missing was proof of principle, already established for surgery and radiation therapy, that drugs could cure any cancer. Major advances came in the mid-1960s with firm evidence that childhood leukemia and advanced Hodgkin’s disease in adults could be cured by combination chemotherapy.

Proof of cure by chemotherapy had a permissive effect on the use of drugs as an adjuvant to surgery and radiation therapy. Doctors started to be willing to consider using chemotherapy. In the mid-1970s, two landmark studies of adjuvant chemotherapy in breast cancer were published: one from the NSABP, which tested a single drug and was reported by Fisher and colleagues in 1975, and one from Italy, which tested a drug combination and was reported by Bonadonna et al. in 1976. The latter study evaluated a combination regimen (cyclophosphamide, methotrexate, and fluorouracil) developed by the NCI but was performed under contract with the Milan Cancer Institute, despite large populations of patients with operable breast cancer in the United States, because no major U.S. center was willing to test combination chemotherapy as an adjuvant. The results of both studies were positive, and the race was on. By 1991, thanks to the availability of multiple effective chemotherapeutic agents and hormone treatments, improved diagnostic tools for early diagnosis, and intelligently designed clinical trials, the rate of death from breast cancer began to fall, a trend that has continued. Early diagnosis and lumpectomy coupled with systemic therapy have greatly reduced the morbidity associated with breast-cancer treatment, with good cosmetic effects. Such advances have fulfilled the mandate of the war on cancer “to support research . . . to reduce the incidence, morbidity and mortality from cancer.”

The success of adjuvant treatment of breast cancer, in turn, had a permissive effect on the use of drugs in the postoperative treatment of other major cancers, such as colorectal cancer. As a consequence of early diagnosis, prevention, and adjuvant treatment, the rate of death from colorectal cancer has fallen by 40% during the past four decades. Another paradigmatic change in cancer treatment occurred in 2006, when Druker et al. showed the efficacy of a drug (imatinib) that targeted the unique molecular abnormality in chronic myeloid leukemia. This work provided proof of principle that treatments targeting specific molecular abnormalities that are unique to certain cancers could convert them into manageable chronic illnesses. Since then, chemotherapy has become targeted therapy, and the literature has been dominated by the search for drugs to inhibit unique molecular targets, with recent success in the treatment of some very difficult-to-treat tumors, such as melanoma and lung cancer.

Until recently, cancer treatment was a three-legged stool sitting on a base of surgery, radiation therapy, and chemotherapy. In the past 25 years, immunotherapy has been added as an important component of cancer treatment. Antibodies were first described in the 1880s and dominated studies of immunology for almost 100 years but had little effect on cancer treatment. In 1975, Köhler and Milstein developed methods for producing antibodies by fusing cultured myeloma cells with normal B cells from immunized mice. The availability of large amounts of an-
tibodies with a single specificity led to the successful development of therapeutic antibodies for cancer, starting with approval by the Food and Drug Administration (FDA) of rituximab for the treatment of B-cell lymphomas in 1997 and followed by the approval of many other antibodies, most of which act by inhibiting growth factor receptors on the surface of cancer cells.

In the early 1960s, it became clear that cellular rather than humoral immunity played a major role in the immune destruction of experimental cancers, although the inability to manipulate T cells outside the body severely hampered studies of tumor immunity. The description of T-cell growth factor (subsequently called interleukin-2) in 1976 was a seminal discovery that stimulated extensive studies of the cellular immune reaction to experimental and human cancers. The durable regression of metastatic melanoma and renal cancers in humans after the administration of interleukin-2, described in 1985, represented the first clear demonstration that immune manipulations could cause the regression of invasive metastatic disease. Interleukin-2 was approved for the treatment of metastatic renal cancer in 1992 and for metastatic melanoma in 1998. The subsequent development of immunomodulatory agents such as ipilimumab, the development of cell-transfer therapies, and the use of genetically engineered lymphocytes to treat cancer have provided additional evidence of the ability of immunotherapy to mediate cancer regression. With the increasing use of these agents, the cancer-treatment platform sits firmly on four legs.

CANCER PREVENTION

No matter how easy cancer treatment may become, it is preferable to prevent cancer. But prevention has been an elusive goal. Figure 2 illustrates three notable pathways to success, with discoveries of the connection between viruses and cancer, methods of chemoprevention, and the role of tobacco in cancer. When the cause of cancer is known, its prevention becomes a problem in modifying human behavior. Nicotine is one of the most addicting substances known, and exposure to tobacco smoke is by far the best known and most frequent cause of cancer, causing an estimated 40% of all deaths from cancer. It was suggested as early as 1912 that smoking might be related to lung cancer, with the epidemiologic evidence becoming solid in the 1950s. These findings led to the Surgeon General’s report on smoking and cancer that was issued in 1964, the use of warning labels on cigarette packages in 1965, and a ban on tobacco advertising in 1970. These and other aggressive, well-publicized public health measures, which were strongly pursued by the American Cancer Society with support from the NCI, have led to a steady reduction in the rate of smoking, which has decreased to half the 1950 level in the United States. It takes time for the deleterious effects of the thousands of carcinogenic chemicals in tobacco to dissipate, and it was not until 1990 that the incidence of lung cancer in men began to decline, followed by a decline in lung-cancer mortality beginning in 1991.

To date, the historic goal of creating a cancer vaccine has been realized only for cancers that are caused by viral infections. Even when the causal virus has been identified, the elapsed time from discovery to prevention has been long. The human papillomavirus was discovered in 1907, but it was not linked to cervical cancer until 1976, and a vaccine to prevent infection by the virus in young girls was not approved by the FDA until 2000. Hepatitis B virus was discovered in 1967 and was linked to liver cancer in 1974. In 1984, it was shown that both hepatitis B and liver cancer could be prevented by vaccination against hepatitis B. Since then, in some parts of the world, vaccination of newborns against the hepatitis B virus has become routine. Since it is estimated that 20% of all cancers are caused in some way by viruses, further development of vaccines holds much promise.

The use of chemicals to prevent cancer (chemoprevention) can be effective. Antiestrogens can prevent ductal carcinoma in situ and reduce the incidence of breast cancer, finasteride can prevent prostate cancer, and plain old aspirin can prevent colorectal cancer. However, this approach is not widely used because large numbers of otherwise normal persons would need to be exposed to potentially toxic materials in order to prevent some cancers.

SURVIVAL NOW AND IN THE FUTURE

Table 1 shows the changes in relative cancer survival rates related to events in science, and Figures 1 and 2 show changes in cancer incidence and mortality, with notations in the charts when the rates of death from a specific cancer began to
fall. Soon after the development of successful treatments in the 1970s, disease-specific death rates began to fall dramatically for childhood leukemia and Hodgkin’s disease. The incidence of these diseases was too low to affect overall rates of death from cancer. Overall rates began to decline soon after the introduction of better early diagnosis and preventive measures and effective adjuvant treatment of common cancers, such as cancer of the breast and colon. The 5-year relative survival rate for all cancers, which was 38% in the late 1960s, just before the passage of the National Cancer Act, is now 68%. Straight-line projections indicate that the survival rate will rise to 80% by 2015.53,54 Overall rates of death from cancer, which began to decline in 1990 in the United States, have decreased by 24% overall since then.53,54 Straight-line projections to the year 2015 indicate that the overall absolute reduction in cancer mortality will be about 38 percentage points.

However, these projections are almost certainly underestimate, since they are based on the assumption that there will be little change in the management of cancer between now and 2015. Most of the current declines are the result of the widespread implementation of old technology for diagnosis, prevention, and treatment, stimulated by funds provided by the war on cancer. However, the biggest payoff from that investment — the clinical application of the fruits of the extraordinary molecular revolution initiated by the National Cancer Act — is yet to come and cannot be measured with the use of current statistics.

**THE FUTURE**

The sequencing of the human genome in 2000 has had a profound effect on all of medicine. The cost of sequencing is reminiscent of Moore’s law, with the cost halving every 2 years. It is not difficult to foresee a time when a person’s individual genome can be sequenced for as little as $100, putting genetic studies in the realm of a routine laboratory test. Starter companies with this aim already exist.

Second- and third-generation deep sequencing is revealing the complexity of the cancer blueprint and no doubt will reveal networks not yet imagined. Nonetheless, we are clearly facing a future in which patients with cancer or those at increased risk will have their genome sequenced as a matter of routine, with comparisons between the premalignant tissue and the malignant tissue. Detected abnormalities will become targets of relatively simple drug therapies, and if the effects mirror what we have seen in recent years with targeted therapy, the ability to prevent or treat cancers in the future will be impressive. The economic and social consequences of converting cancer into a curable or chronic disease will be both gratifying and daunting. This overview of 200 years of the cancer field provides support for the principle of the value of patience and investment in research.

Disclosure forms provided by the authors are available with the full text of this article at NEJM.org.

---

**REFERENCES**

18. Fisher B, Redmond C, Fisher ER. Ten-year results of a randomized clinical trial comparing radical mastectomy and total...

Copyright © 2012 Massachusetts Medical Society.
At first glance, the inaugural 1812 issue of the New England Journal of Medicine and Surgery, and the Collateral Branches of Science seems reassuringly familiar: a review of angina pectoris, articles on infant diarrhea and burns. The apparent similarity to today’s Journal, however, obscures a fundamental discontinuity (1812a, b, c; see box). Disease has changed since 1812. People have different diseases, doctors hold different ideas about those diseases, and diseases carry different meanings in society. To understand the material and conceptual transformations of disease over the past 200 years, one must explore the incontrovertibly social nature of disease.

Disease is always generated, experienced, defined, and ameliorated within a social world. Patients need notions of disease that explicate their suffering. Doctors need theories of etiology and pathophysiology that account for the burden of disease and inform therapeutic practice. Policy-makers need realistic understandings of determinants of disease and medicine’s impact in order to design systems that foster health. The history of disease offers crucial insights into the intersections of these interests and the ways they can inform medical practice and health policy.

Epidemiologic Transitions
In addition to angina, diarrhea, and burns, early Journal issues examined gunshot wounds, spina bifida, tetralogy of Fallot, diabetes, hernia, epilepsy, osteomyelitis, syphilis, cancer, croup, asthma, rabies, and urethral stones. Although some case reports describe patients who might walk into a clinic today, others are nearly unrecognizable. Apoplexy, a syndrome of fainting spells that might mean stroke, seizure, or syncope today, was understood to arise from a “nervous sympathy” by which the stomach influenced the head (1812d). Doctors agreed that even a near miss by a cannonball — without contact — could shatter bones, blind people, or even kill them (1812f). Reports of spontaneous combustion, especially of “brandy-drinking men and women,” received serious, if skeptical, consideration (1812g). And physicians were obsessed with fevers — puerperal, petechial, catarrhal, and even an outbreak of “spotted fever” in which some patients were neither spotted nor febrile (1812e). The bill of mortality from 1811 (see figure) contains both the familiar and the exotic (1812h). Consumption, diarrhea, and pneumonia dominated the mortality data, but teething, worms, and drinking cold water apparently killed as well.
A century later, the infections that filled the Journal had been re-defined according to specific microbial causes. The Journal ran reviews of tuberculosis (1912b), gonorrhea (1912e), and syphilis (1912i). Diphtheria, measles, pneumonia, scarlet fever, and typhoid made frequent cameos, and Massachusetts still maintained a leper colony on Penikese Island (1912d). Tropical infections also fascinated authors, whether worms in immigrants or outbreaks of plague, yellow fever, and malaria in the nation’s new tropical empire.

Doctors in 1912 did have some reasons to celebrate. By any account, the previous year had been “the healthiest of which there is any record” (1912a). Nearly every Journal issue that year mentioned another centenarian, and coverage of the “overwhelming success” of U.S. athletes at the Stockholm Olympics celebrated American racial vigor (1912f). One editorial, describing progress made since the Journal’s early years, rhapsodized about what another century of medical discovery might bring: “Perhaps in 1993, when all the preventable diseases have been eradicated, when the nature and cure of cancer have been discovered, and when eugenics has superseded evolution in the elimination of the unfit, our successors will look back at these pages with an even greater measure of superiority” (1912c).

Such paens to progress, however, were accompanied by fear of the diseases of modernization. One article described a new problem, “automobile knee,” and decried the prevalence of “persons of extremely indolent habits of life” who no longer walked more “than the few steps that are needed from the chamber to the elevator, from the elevator to the dining-room, or lounging-room, and then to the automobile” (1912j). Long-standing concern about epilepsy, alcoholism, and feeblemindedness took on new relevance in a society increasingly preoccupied by fears of race suicide and the promise of eugenics (1912g, 1912h). Doctors struggled with cancer, eclampsia, impotence, heart disease (chiefly infectious or valvular rather than atherosclerotic), and arthritis.

During the 20th century, heart disease, cancer, and other chronic conditions assumed more dominant roles (see bar graph), although outbreaks of infectious disease — from eastern equine encephalitis (1938) and kuru...
(1957) to legionnaires’ disease (1977), AIDS (1981), and multidrug-resistant tuberculosis (1993) — necessitated ongoing vigilance against microbes. New concerns also came to medical attention, from the terrifying consequences of thermonuclear war (1962) to the indolent but devastating effects of environmental pollution (1966) and climate change (1989). Optimism about prospects for the health of future populations persisted but remained tempered by concern about the pathologies of civilization. An obesity epidemic, feared in 1912, has come to pass. Our previously steady increase in life expectancy has stalled and may even be reversed (2005).

DEFINITIONS AND CONSEQUENCES

The material and conceptual dynamism of disease poses challenges: how do we define disease meaningfully, and how do we measure our burden of disease and set health policy priorities? These are deceptively simple questions. The definition of disease in Merriam-Webster’s Medical Dictionary as “an impairment of the normal state of the living animal or plant body” raises questions: What is normal? What is impaired? We cannot answer by referencing biology alone; the line between the normal and the pathological requires value judgments. As physicians know, not every symptom constitutes a disease. Nor, as anthropologists have shown, is it feasible simply to contrast “disease,” as diagnosed by doctors, with “illness,” as experienced by patients. As contemporary disputes over the definitions of alcoholism, chronic fatigue syndrome, and attention-deficit disorder make clear, physicians are never the sole arbiters of disease.

Any responsible attempt to define disease must account for the phenomenon’s complexity. A disease has characteristic signs and symptoms, afflicts particular groups of people, and follows a characteristic course. Doctors name diseases and work to identify their causes and develop ways to prevent and treat them. But patients also ascribe meaning to their suffering and assign responsibility for what went wrong. And diseases have utility, with concrete consequences for patients, doctors, and their institutions. They mediate patients’ claims to the sick role and adjudicate access to health care resources. Disease definitions structure the practice of health care, its reimbursement systems, and our debates about health policies and priorities. These political and economic stakes explain the fierce debates that erupt over the definition of such conditions as chronic fatigue syndrome and Gulf War syndrome. Disease is a deeply social process. Its distribution lays bare society’s structures of wealth and power, and the responses it elicits illuminate strongly held values.

The complexities and consequences of disease extend to its measurement. Even after a disease has been clearly defined, measuring its frequency, intensity, and relevance is not simple. Since the 17th century, polities have compiled causes of death into annual bills of mortality. Successive generations of demographers and epidemiologists have transformed such statistics into age-adjusted measures of disease-specific mortality and developed...
measures of morbidity and of the impact disease has on people’s ability to lead meaningful, productive lives. But such measures, including disability- and quality-adjusted life-years, reduce the complex experience of disease to a single coefficient.

A population’s disease status can also be gauged by lists of common diagnoses at clinics or hospitals, but no single measure definitively characterizes a population’s burden of disease. Choosing among metrics is as much about values and priorities as about science, and it directly affects health policy. Whereas advocates of clinical and research funding for cardiovascular disease might use mortality data to support their claims, mental health advocates can cite morbidity measures in seeking greater resources. Data on causes of childhood mortality would justify certain priorities; analyses of health care spending would justify others. An ideal, sophisticated health policy would integrate all measures to form a holistic map of the burden of disease, but in practice competing interests use different representations of disease burden to recommend particular policies.

ACCOUNTING FOR THE BURDEN OF DISEASE

Regardless of the metric chosen, any map of the burden of disease exposes disparities within and among populations. Two aspects of the burden of disease have remained particularly vexing: changes over time in the prevailing diseases and the persistence of health inequalities.

By examining the many new diseases that have appeared over the past two centuries, historians have categorized the ways in which diseases emerge. New causes (e.g., severe acute respiratory syndrome, motor vehicle accidents, radiation poisoning), new behaviors (cigarette smoking, intravenous drug use), and even the consequences of new therapies (insulin transforming the course and manifestations of diabetes) can produce new diseases. Changing environmental and social conditions can increase the prevalence of once-obscure ailments (myocardial infarction, lung cancer, kuru, and “mad cow” disease). New diagnostic technologies and therapeutic capacity can unmask previously unrecognized conditions (hypertension). New diagnostic criteria can expand a disease’s boundaries (hypercholesterolemia, depression). Changing social mores can redefine what is or is not a disease (homosexuality, alcoholism, masturbation). New diseases can emerge as the result of conscious advocacy by interested parties (chronic fatigue syndrome, sick building syndrome). HIV–AIDS alone demonstrates many of these modes of emergence. The emergence, recognition, and impact of disease are never just a biologic process; the advent of a new disease always involves social, economic, and political processes that shape its epidemiology and influence our understanding and response.

The interaction between the biologic and the social is equally apparent in the decline of a disease. Cannonball injuries, a frequent cause of concern in 1812, disappeared from the Journal, only to be replaced by a new and more terrible litany of war-related injuries. Neurasthenia, a widespread phenomenon of depleted nervous energy in the late 19th century, has disappeared, but traces of it have remained recognizable in other diagnoses throughout the past century. In some cases, a disease’s decline clearly resulted from medical action. Immunizations eradicated smallpox and may someday eradicate polio. Genetic screening has led to dramatic reductions in Tay–Sachs disease, thalassemia,
and familial dysautonomia (2009). But often the potential for eradication has been incompletely realized — witness the continued prevalence of AIDS and tuberculosis in low-income countries and of atherosclerotic heart disease globally.

Even as prevailing diseases have changed, health disparities have endured. Inequalities in health status have always existed, regardless of how health has been measured or populations defined. When Europeans arrived in the Americas, they witnessed stark disparities in the fates of European, American, and African populations. During the ravages of 19th-century industrialization, physicians grew familiar with health disparities between rich and poor. Health inequalities remain ubiquitous, not just among races and ethnic groups but also according to geography, sex, educational level, occupation, income, and other gradients of wealth and power.5

The persistence of health inequalities challenges our scientific knowledge and political will. Can we explain them and alleviate them? Genetic variations don’t explain why mortality rates double as you cross Boston Harbor from Back Bay to Charlestown or walk up Fifth Avenue from midtown Manhattan into Harlem. Nor do they explain why Asian-American women in Bergen County, New Jersey, live 50% longer than Native American men in South Dakota.6 Although we know something about the relationships among poverty, stress, allostatic load, and the hypothalamic–pituitary–adrenal axis, doctors and epidemiologists need more precise models that sketch in the steps between social exposure, disease, and death.

Accounting for the history of disease also requires us to examine why some disparities in disease are seen as proof of a natural order while others are considered evidence of injustice. The 4.3-year life-expectancy gap between blacks and whites in the United States provokes outrage, but the 4.9-year gap between men and women does not. It is tempting to assume that differences between the sexes are natural and those between races are not. But a 19th-century Journal reader might be skeptical of this explanation: men then lived at least as long as women. The survival advantage of women that appeared in the 20th century owed as much to changes in childbearing, improvements in obstetrical practice, and a new epidemic of heart disease disproportionately affecting men as to differences between the X and Y chromosomes. Disparities in health and disease are outcomes that are contingent on the ways society structures the lives and risks of individuals.

Recognition of the contingency of health inequalities should make them a target for intervention, yet the opposite has frequently happened: the ill health of impoverished or marginalized groups has been used against them — as evidence of their inferiority or as an argument that they’re unworthy of assistance. In the late 19th and early 20th centuries such sentiments produced government policies with tragic outcomes for blacks and Native Americans. They may underlie current policies that would limit health care access for mentally ill, impoverished, and immigrant populations.

THE ROLES OF MEDICINE
Medical practice and health policy rely on the assumption that the solution to the problem of disease is to be found in physicians and their therapies. Physicians tend to credit biomedical science with 20th-century improvements in health and longevity. The history, however, is complex and contested.

For example, after Robert Koch’s 1882 discovery of Mycobacterium tuberculosis and the advent of antibiotics in the 1940s, physicians claimed responsibility for the decline of tuberculosis in Europe and North America. But closer examination revealed that this decline had begun before Koch’s discovery and had substantially run its course before effective antibiotics became available. Medicine’s critics instead credited improvements in the standard of living, especially diet. A similar debate has emerged about coronary artery disease. Heart disease, like tuberculosis, followed a century-long epidemic wave, peaking in the United States in the 1960s before beginning 50 years of decline. Researchers have struggled to determine how much credit should be given to health care providers and how much to risk-factor reduction (2007). This debate has now been complicated by recent increases in coronary disease elsewhere, notably Russia and China, and by signs of a plateau and possible reversal of decline in the United States, Australia, and Western Europe (2005). The stakes of this debate are substantial, with implications for the allocation of contested health care resources.

Is there a best health policy? Our goal should be an integrated policy under which health care and public health programs together fully address the disease burden. But the details depend on how we conceptualize and mea-
sure disease. And disease is never static. Just as organisms evolve to keep up with changing environmental conditions (the “Red Queen Effect”), medicine struggles to keep up with the changing burden of disease. Since therapeu
tic innovation takes time, the burden shifts even as solutions appear. By the time antibiotics and vaccines began combating infectious diseases, mortality had shifted toward heart disease, cancer, and stroke. Great progress has been made to meet these challenges, but the burden of disease will surely shift again. We already face an increasing burden of neuropsychiatric disease for which satisfying treatments do not yet exist.

In many respects, our medical systems are best suited to diseases of the past, not those of the present or future. We must continue to adapt health systems and health policy as the burden of disease evolves. But we must also do more. Diseases can never be reduced to molecular pathways, mere technical problems requiring treatments or cures. Disease is a complex domain of human experience, involving explanation, expectation, and meaning. Doctors must acknowledge this complexity and formulate theories, practices, and systems that fully address the breadth and subtlety of disease.

Disclosure forms provided by the authors are available with the full text of this article at NEJM.org.

From the Department of Global Health and Social Medicine, Harvard Medical School (D.S.J., S.H.P., J.A.G.); the Center for the History of Medicine, Countway Library of Medicine (S.H.P.); and the Division of Pharmacoepidemiology and Pharmacoeconomics, Brigham and Women’s Hospital (J.A.G.) — all in Boston; and the Department of the History of Science, Harvard University, Cambridge, MA (D.S.J., J.A.G.).


Copyright © 2012 Massachusetts Medical Society.
Two Centuries of Assessing Drug Risks

Jerry Avorn, M.D.

The history of medicine is largely the story of medicines — a continuing tale of unfolding risks and benefits. Yet the medical world into which the Journal was born in 1812 did not systematically assess the side effects of treatments in relation to the good they did. Often, there was no understanding of the causal linkage between adverse events and the therapies that led to them. The mixed bag that was the era’s pharmacopoeia is illustrated in the first article of the Journal’s first issue, “Remarks on Angina Pectoris,” in which John Warren describes a patient with acute cardiac ischemia who “was ordered to take opium and aether, or the fetid gums; to bathe the feet in warm water; and under the direction of a physician, to lose a little blood. . . . The nitrate of silver was prescribed in solution.” Tobacco rounded out the regimen (1812; see box for cited Journal articles). There was no systematic approach for determining which treatments could be effective with an acceptable level of risk and which were merely toxic.

This confusion is illustrated by an 1814 article on arsenic noting that in the treatment of “herpetic affections,” “the beneficial effects of the remedy are not apparent until after its use has been sometime discontinued” — a perfect method for confounding assessment of efficacy (in this case nil) with side effects. The author describes a patient in whom arsenic treatment was stopped when it caused the skin eruptions that are a hallmark of its toxicity (1814a). The lesions subsided after the arsenic was discontinued, a primitive but commonsensical basis for attributing causality to drug-induced pathology. Not so fast, warned the expert, a Dr. Kinglake. “Under such circumstances, it is natural enough for the patient to deny the medicine any share in the cure, attributing the benefit rather to the discontinuance than the efficacy of the remedy.” In reality, he explained, the arsenic had had a “high stimulating effect” that “prevent[ed] its beneficial operation from being observable.” This “arsenical excitement” was “mistook for an unaltered continuance of the original affliction.” Treat on through the signs and symptoms of toxic effects for up to 3 months, Kinglake recommended. The Journal’s own commentator wisely noted, “we cannot divest ourselves of apprehension that [such continuing treatment] may lay the foundation of subsequent irreparable mischief, and should by no means recommend its general employment to such an extent.”

That same year, a case report...
from Dublin commented on the difficulty of determining whether a particularly florid syndrome was an adverse drug effect, a new unrelated disease, or — most problematic — the consequence of stopping medication too early. The patient, who had been treated with mercury, developed a dramatic skin eruption, fever to 108°F, tachycardia to 130 beats per minute, headache, nausea, convulsions, blisters that discharged an acrimonious lymph, and desquamation so severe that large pieces “of the hand will come off, so entire as to resemble a glove.” Appropriately, the article was titled “Description of a disease produced by the use of mercury.” Helpfully, the report pointed out that “The cure of this disease is very simple. It consists, first, in removing the exciting cause and then its effects.” But the author “thought, in some cases, that continuance of mercury [is] proper.” For although this devastating syndrome could be caused by “an incautious use of mercury,” it could also “be produced in an aggravated form by a too early removal of the mineral” (1814b).

The Journal’s clinical descriptions of adverse drug effects were often detailed and apt, but there was little capacity to weigh these downsides of therapies against their usefulness. When the Journal was launched, the systematic use of randomized, controlled trials was still more than 130 years in the future; it did not become standard until after World War II. Attribution of drug-induced syndromes to a cause was therefore often confused and occasionally counterproductive. At the end of the Journal’s first century of publication, the dean of Harvard Medical School summed up the risk-benefit situation neatly: “I firmly believe that if the whole materia medica could be sunk to the bottom of the sea, it would be all the better for mankind, and all the worse for the fishes.”

Appreciation of this imbalance between toxicity and benefit was not widely shared by his medical contemporaries, but it caught the attention of their patients. By the turn of the 20th century, the proliferation of “patent medicines” had become a national scandal. Manufacturers were not even required to reveal their products’ contents, and certainly not to ensure their safety and efficacy. In the Journal, the debate became a battle between those concerned with patient safety and benefit and a powerful industry intent on protecting its profitability and autonomy from government control. A 1906 issue reported the view of H.W. Wiley, who headed a com-
More than 100 children had been fatally poisoned by the Massengill company’s preparation of the new antimicrobial sulfanilamide (see photo) dissolved in diethylene glycol, a solvent known to be lethal. An enraged public demanded that the FDA be given the right to require that manufacturers prove their products were not toxic before they could be sold.

A Journal editorial placed the sulfa disaster in the context of other drug-induced tragedies: deaths and blindness caused by dinitrophenol, fatal hepatotoxic effects from cinchophen, and “acute and chronic poisoning . . . from the improper use of thyroid and radium preparations.” Perhaps, the editorialist suggested, “the boyish enthusiasm with which we accept nearly anything new, and swallow hook, line, and sinker of the printed page, may be again tempered with that calm appraisal of the facts that has so long been considered a heritage of the medical fraternity” (1937b).

As reaction to the “Massengill massacre” grew, G. Philip Grabfield of Peter Bent Brigham Hospital and Harvard Medical School noted that “Under the present law the only liability of this firm lies in connection with use of the word ‘elixir,’ which is defined as an alcoholic solution. In other words, the firm cannot be indicted for manslaughter but only for misbranding.” Selling a poisonous medication was not yet illegal. He continued, “It was shown in 1930 that the toxicity of diethylene glycol was approximately that of wood alcohol. Apparently, this mixture was distributed for sale not only without being tested but without even a casual investigation of the literature on the part of its makers. This ghastly experience indicates . . . the crying need for adequate legislation to control the marketing of all medicinal substances” (1938c).

Grabfield argued that deficits in physicians’ knowledge contributed to the drug-safety problem, noting that the teaching of pharmacology and preventive medicine “is deficient in most schools. After graduation it should become the concern of the legally constituted health authorities to keep the physicians under their jurisdiction continually conscious of these pitfalls of therapeutics . . . . Education, continuous and unremitting, is the only practicable method of breaking down the hold that proprietary medicine has upon the medical and lay public.”

**The Lethal Sulfanilamide Preparation That Killed More Than 100 Children in 1937.**
Control of pharmacologic toxicity was to be joined in the new legislation with control of informational toxicity — the outrageous and often completely untenable claims made in promoting drugs. But industry pushback was intense. In a 1938 letter to the editor, Henry Christian of Peter Bent Brigham Hospital exhorted physicians to agitate for stronger laws to protect patients: “Remember that powerful interest in the drug and food trades fight against restrictive laws; they are well organized; so there is great need that every physician express himself to those who make our laws in Washington” (1938a). This concern was echoed in an editorial criticizing the new drug-safety law: “Why was such a bill written and approved? The proper answer seems to be that powerful business interest of the trade in drugs and cosmetics saw in this method an escape from the more effective provisions of a different bill” that would have given the FDA stronger authority to monitor promotional claims (1938b).

The next major drug-safety event, the thalidomide disaster, came nearly a quarter-century later, when pregnant women who took the heavily promoted sedative–antinauseant gave birth to children with crippling limb-reduction malformations (1962). Before the drug–defect association was understood, this epidemic of congenital anomalies afflicted more than 10,000 children worldwide, but very few in the United States, where an astute FDA medical officer had prevented the drug’s approval. Like the sulfanilamide tragedy, the thalidomide debacle led to major legislative reforms, this time giving the FDA the power to require drug manufacturers to demonstrate effectiveness as well as nontoxicity.

The Journal played a central role in two prominent drug-safety developments of the 21st century, involving the cyclooxygenase-2 inhibitor rofecoxib (Vioxx) and the oral hypoglycemic agent rosiglitazone (Avandia). According to the pivotal study promoting rofecoxib, the Vioxx Gastrointestinal Outcomes Research (VIGOR) trial (2000), patients given that drug had significantly fewer episodes of gastrointestinal bleeding than those taking naproxen. The authors also reported an incidental finding: naproxen users had one fourth as many myocardial infarctions as patients in the rofecoxib group. They speculated that the cause was a cardioprotective antiplatelet effect of naproxen.

The VIGOR article became notorious, for several reasons. In 2004, another randomized trial of rofecoxib was stopped early because the drug, as compared with placebo, nearly doubled the risk of myocardial infarction and stroke (2005a). Accumulating evidence clarified that the VIGOR findings on myocardial infarction were not attributable to a cardioprotective effect of naproxen but to a cardiotoxic effect of rofecoxib. Follow-up publications drew attention to other important problems with the depiction of adverse events in that trial, including the selective omission of key adverse events caused by the sponsor’s drug and an earlier cutoff date for recording side effects that cast the product in an unfavorable light (myocardial infarctions) than for those revealing its clinical advantage (gastrointestinal bleeding events). This selective reporting prompted the editors to write two “Expressions of Concern” about the reporting of risk data in that trial (2005b, 2006).

Recent years have seen the growth of a new mechanism for the study of adverse drug events: court action forcing the release of raw clinical trial data held by the companies that funded the studies. Years after publication of the original Vioxx studies, such re-analyses of their underlying trial reports have found clear evidence, from 3 years before the drug’s withdrawal, of an increased risk of cardiovascular death, as well as evidence that its gastroprotective advantage for most patients was greatly overstated. The advent of mandatory adverse-event reporting at ClinicalTrials.gov is likely in the coming years to provide an additional means of ensuring more timely detection of drug-safety problems (2011a).

Rofecoxib became another milestone in the punctuated evolution of drug-safety science and
policy. In the wake of its withdrawal, influential reports from the Institute of Medicine and the Government Accountability Office, along with Congressional hearings, questioned how a drug that nearly doubled the risk of myocardial infarction or stroke could have been used by more than 20 million Americans over 5 years without that risk being widely appreciated. The debate highlighted the inadequacy of the FDA’s reliance on spontaneously reported adverse events as a main method of ongoing drug-safety surveillance. In the resulting FDA Amendments Act of 2007, Congress required the FDA to develop a near-real-time surveillance system capable of scanning the electronic records of more than 100 million Americans by 2012. This “Sentinel System” is now operational (2011b).

The past two decades have also seen the widespread application of newer quantitative methods for analyzing drug risks. In a meta-analysis of publicly available data from the clinical trials of rosiglitazone, Nissen and Wolski reported a 43% increase in the incidence of myocardial infarction among patients randomly assigned to receive that drug versus several comparators (2007). This finding, initially contested by rosiglitazone’s manufacturer, was similar to results of analyses conducted by both the company and the FDA, which had not previously been made public. The drug was effectively removed from the market in late 2010 (2010).

The latest development affecting our understanding of drug risks is the use of a computer-based review of records from thousands (or millions) of patients, combined with advanced pharmacoepidemiologic methods, to accurately quantify the rates of specific adverse effects. These tools permit assessment of risk in relation to a drug’s benefits. For example, in 2008, the Journal published two observational studies designed to determine whether using the procoagulant aprotinin (Trasylol) in cardiac surgery increased the risk of thrombotic events. Scanning the records of more than 88,000 patients given aprotinin or a comparator agent, the two teams of investigators found a significant increase in the risk of death associated with aprotinin, after adjustment for underlying differences between the groups (2008a, 2008b). The larger of the two studies had been funded by the drug’s manufacturer to inform an FDA advisory committee meeting, but the company didn’t provide its findings in time for inclusion in those deliberations, and the committee determined that there wasn’t enough information to warrant withdrawing the drug from the market. However, when these articles and findings from a randomized trial with similar results were published (2008c), aprotinin was withdrawn from routine use; it has since been reintroduced in Europe and Canada.

Thus, two centuries after the Journal cited experts recommending that physicians “treat through” the side effects of arsenic and mercury therapy, the assessment of drug risks has become considerably more sophisticated. Throughout this history, the study and management of this inevitable aspect of therapeutics have involved a complex interplay of clinical practice, pharmacology, epidemiology, policy, and politics. Greater access to data and the application of modern information technology and sophisticated epidemiologic approaches are finally providing a valuable and potentially lifesaving way to balance the good that medications can do against the harm that they sometimes cause.

Disclosure forms provided by the author are available with the full text of this article at NEJM.org.

From Harvard Medical School and the Division of Pharmacoepidemiology and Pharmacoeconomics, Department of Medicine, Brigham and Women’s Hospital — both in Boston.


DOI: 10.1056/NEJMz1206652
Copyright © 2012 Massachusetts Medical Society.
Doctors, Patients, and Lawyers — Two Centuries of Health Law

George J. Annas, J.D., M.P.H.

Medical care in 2012 is unrecognizable as compared with what it was in 1812, and no 19th-century physician would be at home in a modern hospital. A 19th-century lawyer, however, would be completely at home in a contemporary courtroom, as would a present-day lawyer transported back to the early 19th century. Although slavery was still legal and women did not yet have the right to vote, the U.S. Supreme Court was the highest court in the land and the U.S. Constitution and its Bill of Rights would be familiar, as would the jury and the common law system adopted from England.

Physicians and lawyers did not necessarily get along better in 1812 than they do today, primarily because of medical malpractice litigation. Herman Melville’s 1851 metaphoric Massachusetts masterpiece, *Moby-Dick*, symbolizes the view of many physicians, then and now, that medical malpractice litigation is the white whale: evil, ubiquitous, and seemingly immortal (Fig. 1). Medicine and law were nonetheless often viewed as the two major professions, and for the leading physicians at that time, including Walter Channing (Fig. 2), editor-in-chief from 1825 to 1835 of what is now the *New England Journal of Medicine*, the relationship between medicine and law was of great intellectual and practical interest.

Over the past two centuries, the discipline of medical jurisprudence — the application of medical knowledge to the needs of justice — has been renamed legal medicine (including forensic science), and applying the law to medicine has expanded from medical law to health law. Legal procedures and courtrooms have changed little, but there have been almost as many changes in the application of law to medicine over the past 200 years as there have been changes in the practice of medicine. Health law’s intimate relationship with medical ethics also has a strong precedent. Thomas Percival’s original title for his 1803 *Medical Ethics* text, medical malpractice litigation, in the eyes of many physicians, is like the white whale in Melville’s *Moby-Dick* (1851).
which has been described as “the most influential treatise on medical ethics in the past two centuries,” was *Medical Jurisprudence.* More than half of Percival’s text specifically addresses “professional duties . . . which require a knowledge of law.”

Walter Channing’s almost-poetic academic title was Professor of Midwifery and Medical Jurisprudence. He lectured on the subject at what would become Harvard Medical School, and his interest in midwifery was based on his belief that physicians should know enough law to be useful and credible witnesses in court. He believed that medicine and law, “two of the most diverse callings may act in perfect harmony, and for the equal benefit of both.” He also quoted medicolegal expert David Paul Brown: “A doctor who knows nothing of law, and a lawyer who knows nothing of medicine, are deficient in essential requisites of their respective professions.” Two cases dealt with in some detail by Elwell illustrate the standards to which courts held physicians (and quacks) in the late 18th and early 19th centuries.

The major areas of medical jurisprudence in the early and mid-19th century were forensic pathology (determination of the cause of death in criminal cases, especially when poisoning was suspected) and forensic psychiatry (determination, for example, of whether a defendant was “sane” at the time he committed a crime). In 1854, the year Channing retired from teaching, his course was entitled “Obstetrics and Medical Jurisprudence.” Insight into the medical jurisprudence of Channing’s times can be found in a remarkable three-part book review, spanning 24 journal pages, which he wrote 6 years later. The book he reviewed, by physician–lawyer John J. Elwell, *A Medico-Legal Treatise on Malpractice and Medical Evidence: Comprising the Elements of Medical Jurisprudence,* was also published in 1860. Both the book and Channing’s review can help us see how medical jurisprudence evolved into health law in much the way that midwifery evolved into obstetrics.

**PHYSICIANS AND THE LAW**

Apart from the many areas of the law that directly affected the practice of obstetrics in the 19th century (most notably, abortion, feticide, and infanticide), medical jurisprudence was not Channing’s main subject. Nonetheless, primarily on the basis of the importance of medical testimony in both civil and criminal cases and on the basis of his own courtroom experiences as an expert witness, Channing strongly believed that physicians should know enough law to be useful and credible witnesses in court. He made this conviction a core of his medical school lectures on the subject. Channing believed that medicine and law, “two of the most diverse callings may act in perfect harmony, and for the equal benefit of both.”

Elwell reasonably objected to the court’s conclusion that if a physician is engaging in a unique experiment, then that fact alone makes the physician “guilty of rashness and recklessness.” He noted that the “recklessness” standard “points strongly to criminal intent or of foolhardiness and culpable rashness [which would make the physician] actually guilty of a crime.” Later in his text, Elwell described just such a case, which he termed “the leading American case on criminal malpractice” and which I call “the case of the coffee quack.”

The coffee quack was charged with murder in the death of his patient. He had come to Beverly, Massachusetts, in 1807 and announced himself as...
a physician with “the ability to cure all fevers.” He used several concoctions, including drugs he called “coffee,” “well-my-gristle,” and “ram-cats.” He administered these drugs, together with heat and blankets, for approximately 1 week to a patient who had employed him to cure a severe cold. The patient vomited frequently, became exhausted, and within days suffered a series of convulsions from which he died. There was testimony that in high doses the “coffee” drug could act as a poison. The jury was instructed that to find the coffee quack guilty of murder they must find that the killing was done with malice, and there was no evidence of this. A finding of manslaughter required that the killing be “the consequence of some unlawful act,” but there was no legal requirement at the time for either licensure or education in order to call oneself a physician. The judge summed up his instructions to the jury:

It is to be exceedingly lamented, that people are so easily persuaded to put confidence in these itinerant quacks . . . If this astonishing infatuation should continue, there seems to be no adequate remedy by a criminal prosecution, without the interference of the legislature, if the quack . . . should prescribe, with honest intentions and expectations of relieving his patients.8

The jury accordingly found the defendant not guilty. At least partially as a result of this verdict, the Massachusetts legislature passed its first physician-licensing law in 1818. That law prohibited unlicensed healers from using the courts to collect payment. It was not until the end of that century that practicing medicine without a license was made a crime.9

**MEDICAL MALPRACTICE AND LAY JURIES**

Historian Michael Bliss argues that in the 19th century, “much of the therapeutic power of medicine stemmed from surgery.”10 Whether or not it was therapeutic, Channing noted in the mid-19th century that malpractice was “almost exclusively charged on surgical practice.”15 Elwell catalogued and provided specific examples of the most common surgical malpractice cases, those involving amputation and the treatment of fractures.6 Beck also appropriately devoted, in Channing’s words, “much of his work” (15 of 42 chapters, and 232 of 582 pages) to the issue of medical malpractice, which “gives to his volume a great value, and makes him a large benefactor to the profession.”4,5

Although Channing thought the jury a wonderful institution, he did not think it was appropriate for medical malpractice cases. He argued that medicine was inherently difficult to understand and not suited to lay juries, which he thought were mostly influenced by dueling expert witnesses whose testimony they could not fathom.5 Channing asked, in words that find common expression today, “What shall be done to remedy so glaring a defect in our jurisprudence — a defect involving so much evil to the accused, and to a profession?”5 His own response was to suggest that, like military officers, physicians should be tried by their “peers” because...
“there is no other way it is possible for them to get justice.”

His view was not unique at the time. It has been independently reported that “between 1845 and 1861 physicians were truly alarmed at the increase of malpractice claims,” and an 1850 communication to the Massachusetts Medical Society referred to the “alarmingly frequent” prosecutions for malpractice and the belief that some surgeons were closing their practices because of this.11 The Massachusetts Medical Society “recommended that a disinterested physician be engaged to adjudicate a threat of malpractice by a disgruntled patient.”11

A century and a half of “malpractice reforms” has not changed the medical profession’s views on medical malpractice litigation, which is still seen as unnecessarily adversarial, shaming, and unfair.12-14 To many physicians, medical malpractice litigation remains the dangerous white whale. Lawyers themselves are not uncommonly viewed as sharks or vultures, bringing to mind Melville’s description of the sharks that harass the whale boats, “seemingly rising from out the dark waters . . . maliciously snapping at the oars . . . following them in the same prescient way that vultures hover.”

CONTEMPORARY HEALTH LAW AND THE SUPREME COURT

Law and medicine have been intimately associated for at least the past two centuries, but it was not until 1964 that the Journal inaugurated a regular feature on the subject (then called “medico-legal relations”) and William J. Curran began writing his “Law–Medicine Notes.”15 Like Elwell and Beck before him, Curran devoted a significant number of his articles to medical malpractice (including hospital liability), forensic medicine (including abortion), and forensic psychiatry, but he also addressed new topics, including the physician’s changing roles in capital punishment, torture, care of the dying, fetal research, and determining death according to brain criteria.15

In 1991, I began writing a Journal feature called “Legal Issues in Medicine” (now “Health Law, Ethics, and Human Rights”). Of the 60 articles that I have written under these two rubrics, approximately 20% have dealt with the power of government over physicians and medical practice; 20% with abortion, pregnancy, and childbirth; 20% with public health issues; and the remainder with research, care of the dying, patient rights, forensic medicine, and forensic psychiatry. What is perhaps most noteworthy, however, is the number of health law cases that have been decided by the U.S. Supreme Court.

Health law — that is, law applied to the health care field — has expanded far beyond anything Channing could have imagined. The recognition of patients’ rights and the expansion of regulatory-oversight rules and mechanisms, for both medical practice and financing, has vastly enlarged the field. Patients’ rights, especially the doctrine of informed consent, were furthered by such judgments as that at the trial of the Nazi doctors at Nuremberg (1946–1947)16 and the Supreme Court’s decision on abortion in Roe v. Wade (1973).17 Informed consent is the core of the Nuremberg Code, as it could have been the core of Slater v. Baker nearly 250 years ago. On its face, Roe v. Wade overturned most state laws that made abortion a crime, but its impact on medical care goes far beyond abortion. The Court ruled that the rights of both the physician and the patient have a constitutional dimension that limits the state’s power to interfere in the physician–patient relationship.17 The politics of abortion have led the Court to decide more than 3 dozen cases on state abortion laws in the past 40 years. The evolving structures of health care financing and practice would also be unrecognizable to 19th-century medical practitioners, including private health insurance plans, Medicare and Medicaid, managed care, the health insurance exchanges and accountable care organizations encouraged by the Affordable Care Act, antitrust regulations, measures to prevent fraud and abuse, and financial disclosure requirements.

A third development is also noteworthy — the application of health law to the field of international human rights, including the right to health, the regulation of research on human subjects, and the physician’s role in war and civil conflict. Physicians and lawyers now work together in U.S.-based organizations such as Physicians for Human Rights and Global Lawyers and Physicians. Working separately, medical associations, including the British Medical Association and the World Medical Association, rather than legal associations, deserve much of the credit for the growth of the international “health and human rights” arena.18 Both law and medicine are criti-
cal tools for improving health and well-being on a global level, and each profession is more effective when the two work together.

Law remains interwoven with the practice of medicine, as it was in the 19th century. Physicians who do not have a basic understanding of the law are, as Channing recognized, at a distinct disadvantage when practicing medicine. The evolution of medical jurisprudence into health law over the past two centuries has been dramatic (Table 1). But equally consequential are the ways in which health law issues are framed and the legal forums in which they are resolved. State laws governing medical practice (including abortion and end-of-life care) are now challenged as unconstitutional infringements of individual rights, with the final determination made by the Supreme Court. The Court has also become active in determining the constitutionality of federal health-related legislation and in interpreting the meaning of federal statutes in the health field, ranging from regulation of tobacco and drugs to gun control. The fate of the Affordable Care Act, the major “health law” of the past decade, has also been decided by the Supreme Court — unthinkable in Channing’s day. The changes in substance and emphasis in health law from the publication of Moby-Dick can be appreciated by reading a contemporary non-fiction best seller about an event that occurred in 1951, which was 100 years after Melville published his masterpiece: the taking of cells that would later be called “HeLa” cells from Henrietta Lacks.19 Although malpractice remains a concern, more central legal issues in contemporary medical practice include the fiduciary nature of the doctor–patient relationship, patient rights and patient safety, informed consent, privacy, commercialization, the regulation of medical research and biobanking, the patenting of genes and cell lines, the application of genomic information to medical practice, racial disparities, and equitable access to quality medical care.20,21

The author of The Immortal Life of Henrietta Lacks, Rebecca Skloot, opens her book with the words of Elie Wiesel that almost all physicians and lawyers would agree should apply to all patients, not least because of the “fiduciary duty” that physicians owe patients under the law (and medical ethics): “We must not see any person as an abstraction. Instead, we must see in every person a universe with its own secrets, with its own sources of anguish, and with some measure of triumph.”19 Disclosure forms provided by the author are available with the full text of this article at NEJM.org.

I thank my health law colleagues Leonard Glantz and Wendy Mariner for their thoughtful comments on early drafts of this article.

Table 1. Some Health Law Highlights.

<table>
<thead>
<tr>
<th>Year</th>
<th>Event</th>
</tr>
</thead>
<tbody>
<tr>
<td>1767</td>
<td>Slater v. Baker and Stapleton, CB Eng Rptr (UK) (medical experimentation)</td>
</tr>
<tr>
<td>1803</td>
<td>Percival’s Medical Ethics published (original title, Medical Jurisprudence)</td>
</tr>
<tr>
<td>1809</td>
<td>Commonwealth v. Thompson, 6 Mass. 134 (wrongful death, quackery)</td>
</tr>
<tr>
<td>1818</td>
<td>First medical licensure statute enacted in Massachusetts</td>
</tr>
<tr>
<td>1823</td>
<td>Theodorick Beck’s Elements of Medical Jurisprudence published</td>
</tr>
<tr>
<td>1840</td>
<td>Medical malpractice litigation appears in the United States</td>
</tr>
<tr>
<td>1860</td>
<td>John J. Elwell’s A Medico-Legal Treatise published</td>
</tr>
<tr>
<td>1905</td>
<td>Jacobson v. Massachusetts, 197 U.S. 11 (no right to refuse smallpox vaccination)</td>
</tr>
<tr>
<td>1946–1947</td>
<td>Doctors’ Trial at Nuremberg (Nuremberg Code set forth in the judgment)</td>
</tr>
<tr>
<td>1955</td>
<td>American College of Legal Medicine founded</td>
</tr>
<tr>
<td>1966</td>
<td>Medicare and Medicaid enacted</td>
</tr>
<tr>
<td>1972</td>
<td>American Society of Law and Medicine founded</td>
</tr>
<tr>
<td>1973</td>
<td>Roe v. Wade, 410 U.S. 113 (right to terminate pregnancy)</td>
</tr>
<tr>
<td>1990</td>
<td>Cruzan v. Director, Missouri Department of Health, 497 U.S. 261 (right to refuse life-sustaining treatment)</td>
</tr>
<tr>
<td>2010</td>
<td>Patient Protection and Affordable Care Act enacted</td>
</tr>
<tr>
<td>2012</td>
<td>National Federation of Independent Business v. Sebelius (upheld all of the Patient Protection and Affordable Care Act as constitutional except the penalty for states that do not expand their Medicaid programs)</td>
</tr>
</tbody>
</table>

REFERENCES


Copyright © 2012 Massachusetts Medical Society.
In 1915, reformers issued the first major proposal for national health insurance in the United States (see timeline). They believed that America should follow European countries such as Germany and England in securing access to medical care for workers and protecting them against the economic burdens of illness. The leadership of the American Medical Association (AMA) initially agreed, and the prospects for reform appeared promising.

Yet by 1920, the health care reform campaign had failed, the victim of intense opposition (from businesses and the insurance industry, among others), bad timing (the American entry into World War I), demagoguery, and xenophobia (charges that the health care proposals were “Made in Germany,” “Bolshevik,” and “un-American”). After an internal revolt, the AMA became a steadfast opponent of national health insurance. The issue briefly disappeared from the agenda.1

Nearly 100 years after that first proposal, Americans are still debating health care reform, the perils of “socialized medicine,” and the tensions between individual liberty and government aid. What have been the major developments in U.S. health policy over the past century? And what challenges lie ahead? I focus here on two critical issues, health insurance coverage and cost containment.

**COVERAGE**

Political struggles over expanding access to insurance have long defined U.S. health policy. Although proposals focused at first on industrial workers, by the 1940s reformers were seeking a universal health insurance program for all Americans. But universal coverage remained elusive during the 20th century. The same forces that initially stalled national health insurance — resistance from powerful interest groups bent on preserving the status quo, demagoguery, and fear of socialized medicine — endured to undercut subsequent reform efforts. A parade of presidents — including Harry Truman,
Richard Nixon, and Bill Clinton — pursued universal coverage. They all failed.

That failure is often attributed to a political culture suspicious of centralized power and enamored of individual responsibility. There is no question that the anti-government strain in U.S. politics made the reformers’ task extraordinarily difficult. However, U.S. political institutions represented an equally important — or perhaps even more important — barrier to reform. In the fragmented U.S. system, health care legislation died in Congress even when it enjoyed support from the president and the president’s party had majorities in the House and Senate. If we had a parliamentary system, the United States probably would have adopted universal insurance decades ago.

The failure of early proposals for national health insurance crucially shaped U.S. health policy. Instead of a single insurance system organized by the government, the United States developed a patchwork of public and private coverage. Employer-sponsored private insurance emerged to cover working Americans and their families. It spread widely in the 1940s and 1950s, as unions pressed for health benefits. Linking insurance to employment provided insurers with a convenient risk pool and a reliable source of premium payments. It also gave opponents of government insurance a viable alternative that embodied “the American way.” Yet private insurance benefited from government largesse: the federal government subsidized employer-sponsored coverage by excluding from taxable income premium payments made by employers on behalf of workers.

Even as it grew, employer-sponsored insurance remained beyond the reach of many Americans. Having failed to secure national health insurance, reformers switched strategies midcentury. They promoted less controversial policies, such as federal funding of hospital construction and medical research, and they decided to build a federal health insurance system incrementally, group by group. Government programs would cover politically sympathetic, deserving populations — beginning with the elderly — who had trouble obtaining private insurance, as well as certain categories of low-income people who couldn’t afford it. The 1965 enactment of Medicare and Medicaid established this pattern of demographic incrementalism, while transforming the government role in U.S. medical care. Thereafter, policymakers would focus on expanding public insurance coverage of pregnant women, children, and persons with disabilities and specific illnesses (such as end-stage renal disease).

Despite the rise of employer-sponsored insurance and the advent of Medicare and Medicaid, the U.S. health insurance system has long had serious gaps and inequities. Many working Americans, particularly those at small firms, do not have access to employer-based coverage and have found it difficult to purchase af-
fordable, comprehensive policies in the nongroup insurance market. The revelation that private insurers have targeted pregnant women and patients with cancer for coverage rescissions[3] — contriving reasons to cancel insurance for persons whose medical circumstances made them “bad” actuarial risks — perfectly captures the economic imperatives and moral illogic of the individual market. In this market, the sickest persons who most need insurance have had the hardest time obtaining it.

There are gaps in public insurance, too. Medicare beneficiaries face substantial cost sharing, and the program does not cover long-term nursing home stays. Medicaid enrollees often have trouble finding doctors who will see them, largely a consequence of low reimbursement rates for a population that lacks the political clout to ensure adequate payment. Cash-strapped states have at times cut Medicaid benefits and eliminated coverage for optional populations during economic downturns.

U.S. insurance arrangements are also bedeviled by complexity: Medicaid has about 50 different eligibility pathways, low-income Medicare beneficiaries are also covered by Medicaid, Medicare’s benefits are sufficiently limited that most enrollees carry secondary insurance, and most uninsured children are eligible for Medicaid or the Children’s Health Insurance Program (CHIP) but are not enrolled. Americans “churn” across different insurance programs depending on their age, parental status, employment, income, and disease. It’s no wonder that U.S. medical care is often characterized as a “nonsystem.”

As the costs of medical care increased, Americans’ access to health insurance eroded. Between 1987 and 2010, the uninsured population grew from 31 million (12.9% of the population) to 50 million (16.3%). The incremental policies adopted to expand access to insurance could not keep pace with the large number of Americans who were losing employer-based coverage. Neither could the safety net of community health centers, hospitals, and other providers who care for the uninsured. Uninsured patients are financial losers for health care institutions, and they consequently face serious barriers to care — a reality underscored by a 1986 law that sought to stop hospitals from dumping patients who lacked coverage. Providers who do see many uninsured patients are, in effect, punished financially for their compassion.

Despite their growing numbers, the uninsured often faded from public view in recent decades. Changing political alignments, the sobering legacy of previous failed reform efforts, and the limited electoral power of the uninsured pushed health care reform down the congressional agenda. However, federal inaction spurred state efforts. No state achieved universal coverage, but some made significant coverage gains during the 1980s and 1990s. And the landmark 2006 Massachusetts law provided a
political and policy blueprint for national health care reform. In 2010, President Barack Obama and Democratic majorities in Congress drew on that blueprint, and lessons from previous reform failures, to win passage of the Patient Protection and Affordable Care Act (ACA) — a historic achievement. The ACA’s sweeping scope — encompassing subsidies for the uninsured, a Medicaid expansion, new insurance exchanges, individual and employer mandates, insurance-market regulations, and much more — broke with the incrementalism of recent decades.\(^4\) When the ACA is fully implemented, an estimated 30 million people will gain insurance coverage, and insured Americans will receive important new protections, such as the prohibition of lifetime dollar caps on insurance benefits. The ACA moves the United States closer to the ideal that all persons, regardless of health status and income, should have access to health insurance.

Still, the ACA underscores the limits of U.S. health policy. Even if the ACA’s projected enrollment targets are met, 30 million persons will remain uninsured a decade from now. That this landmark law will leave half of the uninsured without coverage reveals just how difficult the politics of U.S. health care reform are and how far we still have to go to reach universalism.

**COSTS**

During most of the 20th century, health care costs were not a public policy issue. Spending more on medical care was seen as an investment in the country’s health. Private insurance plans — which largely catered to physicians’ and hospitals’ interests — had few restraints on costs. Medicare, too, initially implemented generous payment policies, partly to curry favor with the health care industry and thereby ensure the program’s successful start.

Investing in medical technologies has produced substantial benefits, such as reduced mortality from heart disease. But since 1970, excessive rates of health care spending have been viewed as a serious problem that threatens government budgets and employers’ bottom lines. In response, U.S. policymakers have formulated a wide array of responses.\(^5\) President Richard Nixon imposed price controls on the health care industry and promoted health maintenance organizations (HMOs). During Gerald Ford’s presidency, Congress advanced health planning, including health systems agencies and certificate-of-need requirements, which aimed to rationalize source use and restrain expansion of medical facilities. President Jimmy Carter tried and failed to win passage of a plan to contain hospital costs; the hospital industry instead launched a short-lived “voluntary effort” at restraint. The Reagan administration supported prospective payment of hospitals by Medicare, and during President George H.W. Bush’s administration, Congress enacted a Medicare fee schedule for physicians. The Clinton administration proposed managed competition within a budget. The Obama administration has emphasized delivery- and payment-system reform.

Some of these proposals were designed to curtail spending across the health care system. However, cost control has been defined largely as a budgetary problem, meaning that presidential administrations and Congress often focus only on reducing federal spending. Medicare savings have been a regular feature of deficit-reduction legislation since the 1980s.

Absent systemwide controls, it has fallen to private payers to contain spending for Americans not covered by government programs. Indeed, much of U.S. health policy is effectively ceded to private actors, who help drive the direction of change. Employers have pursued a variety of cost-containment strategies over the years, ranging from moving workers into HMOs and relying on selective contracting with providers to secure lower payment rates to adopting high-deductible plans and requiring greater cost sharing.

Increasing cost sharing and moving from comprehensive to catastrophic coverage rest on the dubious idea that patients can and should act as consumers do in other markets. In a country with a vast uninsured population, the belief that Americans are overinsured has oddly taken root. Various cost-containment measures — including managed-care limits — have also eroded physicians’ clinical autonomy.

U.S. health policy, in both the public and private sectors, has been highly innovative in producing new organizations and payment methods. Currently, employers, insurers, and state and federal governments are embracing value-based payment- and delivery-system reforms, such as accountable care organizations, that seek to
reverse the traditional financial incentives to provide more services. These reforms are central to the ACA’s vision of cost containment.

Americans used to reassure themselves that although the United States failed to provide universal coverage and affordable care, at least the quality of our health care system was superb. Since the 1970s, research has increasingly challenged that assumption, showing that the quality of care in the United States is inconsistent, often inadequate, and varies by geographic location — problems that other countries struggle with as well. By highlighting the potential for saving money by cutting down on wasteful services, these discoveries have strengthened policymakers’ interest in containing health care costs. Enthusiasm for delivery- and payment-system reform embodies the politically appealing aspiration that the United States can moderate spending by improving quality.

Yet for all the innovation, Americans have been singularly unsuccessful in restraining health care spending. The United States has moved through fads at a dizzying pace in recent decades — from managed to consumer-driven to accountable care — but they have thus far failed to produce reliable cost control. Rising health care costs are an issue throughout the industrialized world, though other countries manage to spend much less while insuring their entire populations. Still, lessons from international experience are largely ignored by U.S. policymakers and analysts intent on fashioning a “uniquely American solution.” The United States has not adopted the cost-containment policies that work in other countries: global budgeting, systemwide fee schedules and payment rules, monopsony purchasing, and supply-side controls on expensive technologies. Instead, America continues to abide high prices and the staggering administrative costs imposed by our byzantine insurance system.

U.S. health policy is a story of progress — but also abject failure, having produced an inequitable, inefficient system that is the most expensive in the world and that leaves 20% of the nonelderly population uninsured.

THE FAILURES OF U.S. HEALTH POLICY
U.S. health policy is a story of progress, with substantial gains in health insurance coverage over the past century, culminating in the ACA’s enactment. But U.S. health policy has also been an abject failure, having produced an inequitable, inefficient system that is the most expensive in the world and that leaves 20% of the nonelderly population uninsured. Health insurance should be a source of security and reassurance. The U.S. insurance system is too often a source of suffering, anxiety, economic insecurity, and frustration.

Too many Americans who fall ill are forced to worry about how to pay their medical bills and the threat of medical bankruptcy, rather than focusing on getting well or coping with maladies that won’t improve. Too many Americans who are eligible for Medicaid and CHIP fall between the cracks. Too many insured Americans are only one illness away from discovering they have inadequate coverage that leaves them with overwhelming bills. Too many Americans have to fight their insurance companies to obtain covered benefits.

That these and other indignities have persisted so long is an indictment of U.S. health policy and its moral quality. If there is one thing we should learn from the experiences of other countries that have universal coverage, it is that it doesn’t have to be this way. None of these problems are natural or inevitable — they are all the result of policy choices that the United States has made.

FUTURE CHALLENGES
In coming years, U.S. health policy will be shaped and perhaps transformed by fiscal pressures and deficit politics. The size of Medicare and Medicaid and their
projected spending growth make them likely targets for plans to reduce the federal deficit. The question is whether health care providers or Medicare and Medicaid beneficiaries will bear the brunt of spending cuts. Tax policy will also have a vital impact, since both programs will require additional revenues to absorb growing populations and finance rising medical costs. Meanwhile, the search for stronger cost control and improved quality will continue.

The most crucial issue, though, is what happens to the ACA after the 2012 elections. Barack Obama's reelection would ensure that the ACA moves forward, albeit with continued conflicts over its implementation at both the state and federal levels. If Mitt Romney wins the presidency, however, and Republicans secure majorities in the House and Senate, major provisions of the law could be overturned.

The ACA will not remedy all that ails U.S. medical care. Much can be done to strengthen its coverage and cost-containment foundations. But the ACA will dramatically improve the health care circumstances of tens of millions of Americans, making coverage more accessible and affordable for uninsured Americans and more secure for those who are insured. After a century of struggle, the ACA's enactment provides strong grounds for optimism about the future of the American health care system. Yet with implementation of the ACA uncertain, U.S. health policy stands at a crossroads: will we continue down the path of reform or move backward?

Disclosure forms provided by the author are available with the full text of this article at NEJM.org.

From the University of North Carolina at Chapel Hill.


DOI: 10.1056/NEJMtp1202111
Copyright © 2012 Massachusetts Medical Society.
Tuberculosis, Drug Resistance, and the History of Modern Medicine

Salmaan Keshavjee, M.D., Ph.D., and Paul E. Farmer, M.D., Ph.D.

Tuberculosis is a treatable airborne infectious disease that kills almost 2 million people every year. Multidrug-resistant (MDR) tuberculosis — by convention, a disease caused by strains of *Mycobacterium tuberculosis* that are resistant to isoniazid and rifampin, the backbone of first-line antituberculosis treatment — afflicts an estimated 500,000 new patients annually. Resistance to antituberculosis agents has been studied since the 1940s; blueprints for containing MDR tuberculosis were laid out in the clinical literature and in practice, in several settings, more than 20 years ago.1,2 Yet today, barely 0.5% of persons with newly diagnosed MDR tuberculosis worldwide receive treatment that is considered the standard of care in the United States.3 Those who have not received appropriate treatment continue to fuel a global pandemic that now includes strains resistant to most — and by some accounts all — classes of drugs tested.4,5 Despite the enormity of the threat, investments to contain the epidemic and to cure infected patients have been halting and meager when compared, for example, with those made to address the acquired immunodeficiency syndrome (AIDS) pandemic. In this essay we seek to elucidate the reasons for the anemic response to drug-resistant tuberculosis by examining the recent history of tuberculosis policy.

Research in Tuberculosis — Midwife of Modern Biomedicine

On the evening of March 24, 1882, when Robert Koch completed his presentation on the infectious cause of tuberculosis, silence enveloped the crowded room at the Berlin Physiological Society.6 A means of combating tuberculosis — a disease that in the 19th century caused, by some accounts, about 25% of all deaths in Massachusetts and New York and claimed the lives of one fourth of Europe’s population — was now within reach.7 Koch summarized the importance of his findings, for which he received the 1905 Nobel Prize, in a manuscript published in the *Berliner Klinische Wochenschrift* shortly after his announcement: “In the future the fight against this terrible plague of mankind will deal no longer with an undetermined something, but with a tangible parasite, whose living conditions are for the most part known and can be investigated further.”8

But therapy lagged. It was not until 60 years later, in 1943, that the first effective antituberculosis agent, streptomycin, was isolated in the laboratory of Selman Waksman at Rutgers University (see timeline, available with the full text of this article at NEJM.org). In November 1944, a patient with tuberculosis received streptomycin and was declared cured of the disease.6 Other cases of successful treatment soon followed.9,10 The British Medical Research Council conducted the first large-scale clinical trial of streptomycin in 1948.11 This study, said to be the world’s first published drug trial that involved the randomization of participants, set the meth-
Celebrating 200 Years

The new drug led to the selection of mutations conferring resistance to it. Resistance to rifampin was observed soon after it was first administered. Laboratory data from trials revealed the rapid onset of isoniazid resistance among patients receiving monotherapy and the suppression of resistance when isoniazid was given in combination with streptomycin or para-aminosalicylic acid. These observations led to the use of multidrug treatment regimens — a strategy widely used today to treat a variety of infectious diseases and cancers. Ultimately, through a series of multicountry clinical trials led by the British Medical Research Council, a four-drug regimen was recommended for use in patients with newly diagnosed tuberculosis. The backbone of such empirical regimens was the combination of isoniazid and rifampin, the most effective and reasonably well-tolerated oral agents, given for 6 to 8 months. Thus, short-course chemotherapy was born.

Drug resistance, however, has remained a challenge. The early hypothesis that resistance always conferred a loss of bacterial fitness, and hence led to lower case fatality rates and decreased transmission of such strains, had been disproved by the 1950s. The first national drug-resistance survey in the world, which involved 974 clinical isolates cultured from newly diagnosed cases of tuberculosis in Britain (1955–1956), showed strains that were resistant to streptomycin (2.5%), para-aminosalicylic acid (2.6%), and isoniazid (1.3%). Similarly, data from the United States showed that isoniazid resistance increased from 6.3% (between 1961 and 1964) to 9.7% (between 1965 and 1968) among patients with newly diagnosed tuberculosis. Between 1970 and 1990, there were numerous outbreaks of drug-resistant tuberculosis involving strains resistant to two or more drugs. As early as 1970, an outbreak in New York City of highly virulent tuberculosis that was resistant to multiple drugs proved to be a grim reminder that resistance did not necessarily reduce a microbe's fitness: the index patient died; 23 of 28 close contacts had evidence of new infection, and active, drug-resistant disease developed in 6 of these 23 contacts, 5 of whom were children.

Tuberculosis, whether caused by drug-susceptible or drug-resistant strains, rarely made even medical headlines, in part because its importance as a cause of death continued to decline in areas in which headlines are written. In such settings, where many of the social determinants of tuberculosis — extreme poverty, severe malnutrition, and overcrowded living conditions — became the exception rather than the norm, some public health experts declared that “virtual elimination of the disease as a public health problem” was in sight. In the United States, federal funding for tuberculosis research was cut; consequently, drug discovery, development of diagnostics, and vaccine research ground almost to a halt.

Optimism that tuberculosis would soon be eliminated was not restricted to wealthy countries. At the 1978 International Conference on Primary Health Care in Alma-Ata (now called Almaty), Kazakhstan, delegates from around the world endorsed the goal of “health for all by the year 2000.” The eradication of smallpox had been announced the previous year, and the future of international public health looked promising to many who were gathered there.

But it was not to be. By the mid-20th century, tuberculosis outcomes had diverged along the fault lines of the global economy: while tuberculosis became rare in countries where income was high, epidemics of the disease raged on in low-income settings. In 1982, the Mexican government defaulted on many of its loan payments,
triggering a debt crisis in many countries with weak economies. Increasing numbers of international health donors and policymakers, slow to contribute resources toward the ambitious Alma-Ata agenda, embraced the idea of selective primary health care: discrete, targeted, and inexpensive interventions.\textsuperscript{25,26} Bilateral assistance withered, and poor countries became increasingly reliant on loans from international financial institutions such as the World Bank, which based its health agenda on the principles of “cost-effectiveness” and “affordable health for all” — the latter concept a nod to the Alma-Ata Declaration.\textsuperscript{27}

Selective primary health care offered clear targets, measurable outcomes, and a high return on health investments, all of which appealed to donors worried about investing in countries that were on the brink of default.\textsuperscript{28,29} But several leading causes of disability and death, including tuberculosis, were deemed too costly and complex to address in resource-poor settings and were largely excluded from the emerging, constricted agenda for effective health investments.

“Leprosy and tuberculosis require years of drug therapy and even longer follow-up periods to ensure cure,” wrote two of the architects of selective primary health care in 1979. “Instead of attempting immediate, large-scale treatment programs for these infections, the most efficient approach may be to invest in research and development of less costly and more efficacious means of prevention and therapy.”\textsuperscript{25}

But tuberculosis, which persisted in settings of poverty, could not be hidden away for long. In 1993, the World Bank began to use disability-adjusted life-years — a means of measuring the “cost-effectiveness” of a given health intervention that took into account morbidity, mortality, and age — to determine which health interventions to support.\textsuperscript{30} As a result of this new economic calculus, short-course chemotherapy for tuberculosis was declared a highly “cost-effective” intervention and gained momentum.\textsuperscript{31} Seizing the opportunity, the World Health Organization (WHO) shaped and promoted the DOTS (directly observed therapy, short-course) strategy, an approach that conformed to the selective primary health care agenda: simple to treat, algorithmic, and requiring no expensive inputs. According to this strategy, the diagnosis was to be made with the use of smear microscopy alone — in spite of the insensitivity and inability of this technique to detect drug resistance — and the treatment approach was to be based on the empirical use of first-line antituberculosis agents only.\textsuperscript{32} Facility-based infection control was not part of the DOTS strategy. Despite these exclusions, DOTS was an important development in global tuberculosis policy. Increasingly, poor countries began implementing the DOTS approach; many lives were saved and many new cases averted. However, for children with tuberculosis, people with both tuberculosis and advanced disease from the human immunodeficiency virus (HIV), and the increasing proportion of patients infected with strains of tuberculosis that were already drug-resistant, the DOTS strategy provided limited options for prompt diagnosis and cure.

\addvspace{10pt}

\begin{table}[h]
\centering
\caption{The Emergence of MDRTB Globally}
\begin{tabular}{|c|c|}
\hline
\textbf{Year} & \textbf{Event} \\
\hline
1993 & \textit{DOTS} introduced by WHO \\
1996 & First cases of MDR-TB reported in US \\
2003 & WHO launches Stop TB Initiative \\
2012 & Global prevalence of MDR-TB reaches 110,000 cases \\
\hline
\end{tabular}
\end{table}

These shifts in tuberculosis policy — linked to the reconceptualization of this leading infectious killer of young adults and children from a disease deemed to be costly and difficult to treat to a disease deemed to be “cost-effective” to treat and slated for eradication — convey precisely what is meant by the “social construction of disease.”\textsuperscript{33} \textit{M. tuberculosis} did not conform to the regnant disease-control strategy, and resistant strains continued to emerge and to be transmitted because empirical treatment with first-line antituberculosis drugs was ineffective for those sick with strains resistant to these drugs. HIV infection fanned epidemics of tuberculosis. In the late 1980s and early 1990s, outbreaks of MDR tuberculosis were again reported in the United States.\textsuperscript{17} Genetic analysis of drug-resistant strains showed that airborne transmission of undetected and untreated strains played a major role in these outbreaks, disabusing practitioners of the notion that resistance stemmed solely from “sporadic pill taking.”\textsuperscript{20,34} Public health officials developed a national action plan to combat drug-resistant tuberculosis and to increase funding for relevant research.\textsuperscript{17,35-37} The experience in New York City offered a blueprint that was quite different from the DOTS strategy; it consisted of diagnosis with the use of mycobacterial culture and fast-track drug-susceptibility testing, access to second-line antituberculosis medications, proper infection
control, and delivery of medications under direct observation.1

Outbreaks of MDR tuberculosis in the United States were a harbinger of the coming global pandemic. By the early-to-mid-1990s, MDR tuberculosis had been found wherever the diagnostic capacity existed to reveal it. But in contrast to the U.S. strategy, the WHO — the principal standard-setting body for many countries — continued to advocate the use of sputum-smear microscopy and first-line antituberculosis treatment alone for combating epidemics in resource-poor settings. Some international policymakers thought that treating MDR tuberculosis would be too expensive and complex — claims similar to those made about treating drug-susceptible tuberculosis before this approach was found to be “cost-effective” — and would distract attention from the newly branded (and often successful) DOTS strategy.38 Contemporaneous experience in the United States and in several countries in the former Soviet Union suggested, however, that short-course chemotherapy was ineffective against strains shown to be resistant to precisely those drugs on which such therapy was based.1,17,39,40

THE LIMITS OF SHORT-COURSE CHEMOTHERAPY

The failure of short-course chemotherapy against MDR tuberculosis, though unsurprising clinically, was difficult politically. In Peru, for example, a campaign to promote the DOTS strategy had been so successful in making short-course chemotherapy available that the country’s leaders elevated it as a point of national pride. Peru emerged as a crucible for debates about the treatment and management of MDR tuberculosis in poor countries.2 In 1995, an outbreak in a shantytown in the northern reaches of Lima was identified.41 Many patients were infected with strains found to have broad-spectrum resistance to first-line drugs. Nongovernmental organizations worked with the Peruvian Health Ministry to apply the standard-of-care treatment used in New York City and elsewhere in the United States. The strategy was modified to provide community-based care, with good results.42 After arguing that the DOTS strategy alone could rein in the mutant bacteria, the WHO and other international public health authorities advised the Peruvian government to adopt a low-cost, standardized regimen for the treatment of MDR tuberculosis rather than protocols based on the results of drug-susceptibility testing. In the absence of tailored therapy, many hundreds of deaths occurred among some of Lima’s poorest people.43 As expected, amplification of drug resistance was documented.44,45

By the end of the 1990s, facing mounting evidence that MDR tuberculosis could be treated effectively in resource-poor settings,46,47 a multi-institutional mechanism — the Green Light Committee — was created to encourage and learn from pilot projects for treating MDR tuberculosis.2,17,48 This coincided with a grant from the Bill and Melinda Gates Foundation to scale up treatment of MDR tuberculosis in Peru and elsewhere and to change global policy.

TUBERCULOSIS POLICY AND GLOBAL HEALTH EQUITY

Drug resistance is well established as an inevitable outcome of antibiotic use; the fault lines of the MDR tuberculosis pandemic are largely man-made. The contours of global efforts against tuberculosis have always been mediated by both biologic and social determinants, and the reasons for the divergence in the rates of tuberculosis and drug resistance between rich and poor countries are biosocial.49 As case rates dropped in wealthy countries, funding for research and implementation programs dried up, even though tuberculosis remained the world’s leading infectious killer of young adults throughout the 20th century. Tuberculosis “control” in the 1990s was defined by the legacy of selective primary health care: targeted, “cost-effective” interventions packaged together, in the case of tuberculosis, as the DOTS strategy. Such protocols helped standardize tuberculosis treatment around the world — a process that was sorely needed — but they hamstring practitioners wishing to address diagnostic and therapeutic complexities that could not be addressed by the use of sputum-smear microscopy and short-course chemotherapy or other one-size-fits-all approaches. These complexities, which now range from pan-resistant tuberculosis to undiagnosed pediatric disease, account for more than a trivial fraction of the 9 million new cases of tuberculosis and the almost 2 million deaths from this disease that occur around the globe each year.

The history of divergent policies for combat-
ing drug-resistant tuberculosis shows that decades of clinical research and effective programs in high-income settings did not lead to the deployment of similar approaches in settings of poverty. Achieving that goal demands a commitment to equity and to health care delivery. The U.S. response to the outbreaks of MDR tuberculosis in New York City and elsewhere was bold and comprehensive; it was designed to halt the epidemic. A similar response has not yet been attempted in low- and middle-income countries. Instead, selective primary health care and “cost-effectiveness” have shaped an anemic response to the ongoing global pandemic.

New diagnostics and therapeutics are urgently needed; most of the methods used currently were developed decades ago. Today, we have rapid nucleic acid–based tests for drug-resistant tuberculosis, sound models for laboratory expansion and for treatment delivery, and several drug candidates in the pipeline. To tackle tuberculosis, we also need an equity plan that takes seriously the biosocial complexity of a lethal airborne infection that has stalked us for centuries. The global AIDS effort of the past decade has shown how much can be accomplished in global health when effective diagnosis and care are matched with funding and political will. Stunting on investments or on bold action against tuberculosis—in all its forms—will ensure that it remains a leading killer of people living in poverty in this decade and the next.

Dr. Keshavjee reports receiving grant support from Eli Lilly. No other potential conflict of interest relevant to this article was reported.

Disclosure forms provided by the authors are available with the full text of this article at NEJM.org

We thank Mercedes Becerra, Nadza Durakovic, Edward Nardell, Haun Saussy, and Jon Weigel for their assistance and for editing an earlier draft of the manuscript.

REFERENCES

31. World Bank. World development re-

Copyright © 2012 Massachusetts Medical Society.
Medical School. Warren’s clinical descriptions should still sound familiar to anyone who has treated coronary artery disease. His therapeutic strategies, in contrast, appear downright bizarre. He treated one patient, a “plethoric” clergyman, with stimulants, bloodletting (see Fig. 1), and topical ether, then with more bloodletting, opium, powerful laxatives, and caustic agents that blistered the skin over his sternum. As the patient’s anginal attacks increased in frequency and intensity, Warren tried asafetida (see Fig. 2) — a botanical resin known as “Devil's dung” for its sulfuric, excremental smell — and additional caustics such as silver nitrate to provoke draining blisters on his thighs and arms. With the clinical picture worsening, Warren sent his patient on a therapeutic voyage to Georgia, “where he passed the winter, and suffered less violent attacks than in a more northern climate” (1812a; see box for cited Journal articles). When the minister returned to Boston and his attacks again intensified, Warren added arsenic and bled him vigorously, to no avail. Before his patient's death, Warren noted that the minister's condition improved somewhat with the use of tobacco.

Seen at a remove of two centuries, Warren’s treatments seem excessive, even futile. Apart from opiates — which still have a role in treating severe angina — they have nothing in common with today’s cardiovascular therapeutics. Thrombolytic agents, antiplatelet drugs, beta-blockers, stents, and bypass surgery — the mechanisms of which are understood in many cases at a molecular level — have demonstrably improved the patient's odds of surviving and leading a productive life, even after a major heart attack. Yet an examination of the history of therapeutic practice can do more than simply chart our progress over the past two centuries. It can also demonstrate how change occurs in medicine, revealing what has been gained and what opportunities have been lost along the way.

As generations of physicians have sought more rational bases for medical practice, they have swung between the poles of enthusiasm and skepticism. They have sought therapeutic power
and confidence by reducing their scope of vision toward more precise targets of intervention and measures of success, sometimes losing sight in the process of the broader significance of therapy within the lives of patients and populations. A historical approach to therapeutics, as examined through the pages of the \( \text{Journal} \), can help to redirect our attention toward the practical context in which medicine has evolved.

**Therapeutics in Context**

Although many practices of 19th-century physicians sound macabre to us today, it is important to understand that their therapeutics *actually worked* — within the context of a very different way of thinking about disease and therapeutic efficacy.\(^1\) Both patients and doctors in 1812 generally believed that health and disease were related to the balance and free flow of the four humors: blood, phlegm, black bile, and yellow bile. They also shared expectations about therapeutics: a remedy should provoke powerful symptoms to restore balance and flow. A patient who was feverish, flushed, and delirious from malaria could be calmed and cooled, at least to the touch, by bleeding. Patients who were convinced that their suffering stemmed from intestinal obstructions were gratified by the voluminous vomiting and diarrhea that emetics and cathartics produced. Moreover, treatments were tailored to individual characteristics, such as age, habits, occupation, and locale.

Humoral therapeutics took a distinctly American turn in the young republic. Warren’s seemingly buckshot therapeutic approach evokes caricatures of the practitioner of “heroic medicine,” an approach commonly associated with Philadelphia’s Benjamin Rush. Heroic medicine employed dramatic interventions to “shock” the body back into a state of humoral balance and health. The more dire the disease, the more heroic the intervention.

An 1812 article in the \( \text{Journal} \) advised “copious bleeding” of patients with gunshot wounds — a therapeutic strategy that seems oxymoron until we recall that physicians’ principal concern (once the initial hemorrhage was stayed) lay in preventing suppuration and gangrene. Since these processes were known to follow inflammation and fever, and bloodletting reduced visible signs of both, physicians had a moral imperative to bleed as much as was tolerable in order to save life and limb (1812d). Therapeutic rationality took many forms.

**Skepticism, Enthusiasm, and the Therapeutic Imperative**

The \( \text{Journal} \)’s inaugural issue featured a largely favorable review of Rush’s teachings, along with the lament that “were our knowledge of diseases and their treatment as definite as our acquaintance with the forms and laws of matter; there would be neither doubt nor diversity in medical practice, and mankind would be entitled to reach the allotted period of three score years and ten” (1812c).

Yet doubt and diversity were on the rise. Boston soon became home to a skeptical practice style that directly disparaged Rush’s heroic approach. Even in the \( \text{Journal} \)’s first issue, Jacob Bigelow critiqued the varied rationales justifying existing treatments for burns and appealed for empirical evidence to support the “negative mode of treating burns, which should...
Many factors fostered the spread of skepticism about therapeutics during the first half of the 19th century. American doctors admired the work of Pierre Louis and the “numerical method” taught at La Charité Hospital in Paris, where Louis tallied up outcomes in patients with pneumonia who were treated with or without bloodletting and found no measurable difference. The local marketplace played a role as well: “regular” physicians faced competition from homeopaths, hydropaths, naturopaths, eccentrics, and other sectarians who lampooned the traditional devotion to the lancet and offered less painful alternatives. By the Journal’s 50th anniversary, the vis medicatrix naturae had become so central to U.S. therapeutic practice that Harvard’s John Ware devoted the first 2 parts of a 21-part series on “General Therapeutics” to explicating the concept (1861).

The philosophy of therapeutic skepticism was perhaps most famously articulated by Oliver Wendell Holmes, who remarked in 1860 that “if the whole materia medica, as now used, could be sunk to the bottom of the sea, it would be all the better for mankind — and all the worse for the fishes” (2009).

Though therapeutic skepticism was an animating force in American medicine, its more extreme incarnation, therapeutic nihilism, was never a viable solution for physicians. Doctors could not abandon heroic medicine overnight simply on the basis of numerical “proof” that bloodletting didn’t work — for no doctor worthy of the title could morally countenance doing nothing when confronted with suffering patients. Therapeutics, embedded in both matters of proof and mat-
ters of practice, could change only as much as medical theory and patient expectations allowed. As Holmes observed, “there is a changeable as well as a permanent element in the art of healing; not merely changeable as diseases vary, or as new remedies are introduced, but changeable by the going out of fashion of special remedies, by the decadence of popular theory from which their fitness was deduced, or other cause not more significant” (2009).

Heroic therapies faded only as physicians shifted their enthusiasm to new interventions — notably, quinine, alcohol, and other purported stimulants — in the mid-to-late 19th century. Quinine, for instance, became popular as both a specific treatment for malaria and a general “tonic.” Just as cathartics and bleeding made sense to doctors concerned about fever, obstruction, and humoral balance, quinine and other stimulants made sense in a medical world increasingly dominated by consumption and other diseases characterized by loss of vital energies. And mid-19th-century physicians didn’t abandon the iconic forms of heroic therapeutics — mercury and the lancet — without a fight.

Even as Union Army physicians used less and less calomel in coping with Civil War casualties, they rallied to court-martial the Surgeon General in 1863 after he moved to ban the use of this mercury compound. And even after physicians had tempered their heroic therapies, they remained committed to tailoring remedies to patients’ idiosyncrasies.

**THERAPEUTIC REVOLUTIONS**

By the mid-19th century, however, the focus on patients’ particularities began to give way to interest in the specific causes of disease. Motivated by breakthroughs in cellular pathology, pathophysiology, and especially bacteriology, doctors increasingly came to see diseases as specific entities, each with its own specific causes, manifested as characteristic syndromes. This new model prompted doctors to seek therapies tailored to the disease and not the patient. This transformation, like other “therapeutic revolutions,” took a complex course. Old ideas about therapeutic skepticism and individualization endured, and the promise of new therapies often didn’t materialize for decades.

Consider the “revolution” launched by William Morton’s 1846 demonstration of ether anesthesia at Massachusetts General Hospital. Described in the *Journal* on November 18, 1846, ether anesthesia was one of the first significant medical discoveries to cross the Atlantic from west to east and transform medical practice in both North America and Europe (1846). The transformation was, however, neither rapid nor smooth. Anesthesia enabled dramatic innovation in surgery, but it also increased the dangers of surgery. Before the use of antisepic and aseptic techniques, operative mortality and postoperative infections took a staggering toll, as any Civil War surgeon could recount. Many surgeons, long inured to the pain they inflicted, wondered whether pain relief justified the unknown risks from the new anesthetic agents. Surgical decision making required a delicate “calculus of suffering” in which the surgeon weighed the factors in each case, and the anesthetic “revolution” followed a more halting course than one might imagine.

By the *Journal’s* centennial in 1912, however, surgeons had mastered aseptic techniques and the rituals of the modern operating room. The *Journal* abounded with accounts of innovations in abdominal surgery (1912b, 1912c), vascular surgery (1912d), orthopedic surgery (1912e), obstetric and gynecologic surgery (1912g), and thoracic surgery (1912f) that had previously been inconceivable. The revolution in surgery required not just ether but a careful articulation of diverse processes — anesthesia and asepsis, but also the choreography of surgeons, anesthetists, scrub nurses, linens, autoclaves, and redesigned hospitals. Even those revolutions that in retrospect seem most obvious followed a complicated course.

A similar story played out in the realm of pharmacotherapy in the early 20th century. During the *Journal’s* centennial year, there were effusive reports on the innovations in antibacterial chemotherapy emerging from the Berlin laboratory of Paul Ehrlich, who sought a “magic bullet” — a specific therapeutic that would selectively poison a pathogenic microbe while leaving the host unharmed. After 605 failures, the antisyphilitic Compound 606 — Salvarsan — was widely hailed.
Salvarsan also showed the conceptual limits of a reductionist approach to medicine. Syphilis was not simply a collection of signs and symptoms that followed infection by a particular pathogen. It was a complex social phenomenon, involving shame, stigma, and other moral complications associated with sexually transmitted infection. Although Salvarsan provided relief to some patients, it offered only a partial solution to a complex disease. Preserved in the archives of the *Journal* for that year are the voices of physicians who worried that too much of traditional practice had been lost in focusing on treating diseases and not patients. Should not students, one author worried, “also be taught the art of relieving, of soothing and comforting those who suffer, and of steadying and supporting those who walk in the valley of the shadow?” (1912a).

**THERAPEUTIC SKEPTICISM REVISITED**

Pharmaceutical progress accelerated dramatically between the 1940s and the 1960s. Even in the context of other 20th-century therapeutic revolutions — such as psychoanalysis and cardiac surgery — the midcentury surge in pharmaceutical therapy stands out. More than 4500 new drug products entered the U.S. market in the 1950s as industry churned out new classes of therapeutic agents: broad-spectrum antibiotics, antidiabetic agents, antihypertensives, antipsychotics, antidepressants, and cholesterol-lowering medications. Of every dollar spent on pharmaceuticals in 1961, 70 cents went to drugs that had been unavailable just 10 years earlier (1962b).

This new enthusiasm provoked new forms of skepticism. By the *Journal’s* sesquicentennial year, Louis Goodman and other clinical pharmacologists echoed Holmes in bemoaning the “therapeutic jungle” of the 1950s wonder drugs. Other critics were concerned by what they saw as the “brainwashing” of clinicians by pharmaceutical marketing (see Fig. 4). These concerns were reflected in televised hearings on the marketing practices of the prescription-drug industry orchestrated by Senator Estes Kefauver from 1959 to 1962. The *Journal* offered blow-by-blow coverage of the hearings, focusing on the need for the Food and Drug Administration to formally adjudge therapeutic efficacy and transform the research and development process (1960a, 1960b, 1961a, 1961b, 1962a).

However, passage of the Kefauver–Harris Amendments of 1962, which gave rise to the structure of phase 1, 2, and 3 clinical trials for demonstrating therapeutic efficacy, owed as much to the thalidomide tragedy as to Kefauver’s efforts. The horrors of thalidomide, the sedative–antinauseant that caused limb-reduction malformations in children of women who took the drug while pregnant, extended beyond the drug itself (2011b); as a *Journal* editorial noted, given the furious pace of pharmaceutical development, marketing, and consumption, “only continued and increasing vigilance can prevent the experience from being repeated” (1962c). Outside the medical profession, thalidomide would inspire even more nihilistic perspectives, embodied in popular works such as Morton Mintz’s *The Therapeutic Nightmare* and Ivan Illich’s *Medical Nemesis*. The specter of iatrogenesis these books invoked continues...
to haunt practice, from thalidomide to Vioxx, from DES to Avandia.

Extending this renaissance of skepticism, some questioned the overall role of medicine itself in improving public health. In 1962, the physician-demographer Thomas McKeown published an analysis of the decline of tuberculosis in England and Wales. Noting that the decline had begun before the bacillus was discovered and had nearly concluded before streptomycin was developed, McKeown argued that modern therapeutics had been falsely credited with public health improvements that could be better explained by secular changes in nutrition and standards of living. Similarly, those attempting to bring the benefits of modern tuberculosis drugs to impoverished populations in the 1960s realized that drugs were necessary but not sufficient for transforming health — a lesson that would be relearned through global efforts to treat malaria, tuberculosis, and HIV infection in the 21st century (2006).  

RECONTEXTUALIZING THERAPEUTICS

From the leeches, lancets, and purgatives of the early 1800s to today’s targeted molecular medicines, doctors have constantly sought new and better therapies. Yet the evolution of the field of therapeutics has not been linear, and none of the therapeutic revolutions of the past two centuries have been immediate or complete. Rather, our field’s progress owes as much to changing forms of therapeutic skepticism as to changing forms of therapeutic enthusiasm.

As the locus of disease has narrowed from the afflicted person to the molecular mechanism, and the target of magic bullets has followed suit, physicians have faced regular reminders of the limits of the reductionist approach. The history of therapeutics offers a space to reflect on these more subtle logics of medical knowledge and practice, restoring our appreciation for the breadth of the physician’s task and the complexity of our mission.

Disclosure forms provided by the authors are available with the full text of this article at NEJM.org.

From the Department of the History of Medicine and the Department of Medicine, Johns Hopkins University School of Medicine, Baltimore (J.A.G.); the Department of Global Health and Social Medicine, Harvard Medical School (D.S.J., S.H.P.), and the Center for the History of Medicine, Countway Library of Medicine (S.H.P.) — both in Boston; and the Department of the History of Science, Harvard University, Cambridge, MA (D.S.J.).

DOI: 10.1056/NEJMp1113570
Copyright © 2012 Massachusetts Medical Society.
The Past 200 Years in Diabetes

Kenneth S. Polonsky, M.D.

Diabetes was first recognized around 1500 B.C.E. by the ancient Egyptians, who considered it a rare condition in which a person urinated excessively and lost weight. The term diabetes mellitus, reflecting the fact that the urine of those affected had a sweet taste, was first used by the Greek physician Aretaeus, who lived from about 80 to 138 C.E. It was not until 1776, however, that Matthew Dobson actually measured the concentration of glucose in the urine of such patients and found it to be increased.1

Diabetes was a recognized clinical entity when the New England Journal of Medicine and Surgery was founded in 1812. Its prevalence at the time was not documented, and essentially nothing was known about the mechanisms responsible for the disease. No effective treatment was available, and diabetes was uniformly fatal within weeks to months after its diagnosis owing to insulin deficiency. In the intervening 200 years, major fundamental advances have been made in our understanding of the underlying causes of diabetes and the approach to its prevention and treatment (see timeline, available with the full text of this article at NEJM.org). Although diabetes is still associated with a reduced life expectancy, the outlook for patients with this disease has improved dramatically, and patients usually lead active and productive lives for many decades after the diagnosis has been made. Many effective therapies are available for treating hyperglycemia and its complications. The study of diabetes and related aspects of glucose metabolism has been such fertile ground for scientific inquiry that 10 scientists have received the Nobel Prize for diabetes-related investigations since 1923 (Table 1). Thus, as a result of the efforts of the past 200 years, there is much good news to report regarding diabetes.

Ironically, although scientific advances have led to effective strategies for preventing diabetes, the pathway to cure has remained elusive. In fact, if one views diabetes from a public health and overall societal standpoint, little progress has been made toward conquering the disease during the past 200 years, and we are arguably worse off now than we were in 1812. Two centuries ago, severe insulin deficiency dominated the clinical presentation of diabetes. Although it is possible that some people had milder forms of hyperglycemia at that time, they largely escaped clinical detection. In 2012, the commonly encountered spectrum of diabetes is quite different. Although severe insulin deficiency still occurs, it now accounts for only about 10% of cases overall and can be readily treated with insulin. The vast majority of patients with diabetes are overweight and have a combination of insulin resistance and impaired insulin secretion. The prevalence of this form of diabetes has been increasing dramatically, particularly in the past three to four decades, resulting in a worldwide epidemic that has made diabetes one of the most common and most serious medical conditions humankind has had to face.
Table 1. Nobel Prizes for Diabetes-Related Research.

<table>
<thead>
<tr>
<th>Year</th>
<th>Category</th>
<th>Recipient</th>
<th>Contribution</th>
</tr>
</thead>
<tbody>
<tr>
<td>1923</td>
<td>Medicine</td>
<td>F.G. Banting and J.J.R. Macleod</td>
<td>Discovery of insulin</td>
</tr>
<tr>
<td>1947</td>
<td>Medicine</td>
<td>C.F. Cori and G.T. Cori</td>
<td>Discovery of the course of the catalytic conversion of glycogen</td>
</tr>
<tr>
<td>1947</td>
<td>Medicine</td>
<td>B.A. Houssay</td>
<td>Discovery of the role of hormones released by the anterior pituitary lobe in the metabolism of sugar</td>
</tr>
<tr>
<td>1958</td>
<td>Chemistry</td>
<td>F. Sanger</td>
<td>Work on the structure of proteins, especially insulin</td>
</tr>
<tr>
<td>1971</td>
<td>Medicine</td>
<td>E.W. Sutherland</td>
<td>Discoveries concerning the mechanisms of action of hormones</td>
</tr>
<tr>
<td>1977</td>
<td>Medicine</td>
<td>R. Yallow</td>
<td>Development of radioimmunoassays for peptide hormones</td>
</tr>
<tr>
<td>1992</td>
<td>Medicine</td>
<td>E.H. Fischer and E.G. Krebs</td>
<td>Discoveries concerning reversible protein phosphorylation as a biological regulatory mechanism</td>
</tr>
</tbody>
</table>

THE SCIENTIFIC BASIS OF CURRENT TREATMENT APPROACHES

STUDIES OF GLUCOSE METABOLISM
In the past 200 years, we have made dramatic advances in our understanding of the regulation of normal glucose metabolism. Beginning in the mid-19th century, Claude Bernard showed that blood glucose levels are regulated not just by the absorption of dietary carbohydrate but also by the liver, which plays a central role in producing glucose from nonglucose precursors.2 Other investigators built on this discovery to identify the enzymes responsible for the synthesis and breakdown of glycogen,3 the role of anterior pituitary hormones in glucose metabolism and the onset of diabetes,4 the role of reversible protein phosphorylation by a protein kinase,5 and the discovery of cyclic AMP and its role in hormonal action, particularly that of epinephrine and glucagon, both of which elevate the blood glucose concentration and contribute to diabetic hyperglycemia.6

THE ROLE OF THE PANCREAS AND THE DISCOVERY OF INSULIN
In 1889, Joseph von Mering and Oskar Minkowski found that removing the pancreas from dogs resulted in fatal diabetes, providing the first clue that the pancreas plays a key role in regulating glucose concentrations.7,8 In 1910, Edward Albert Sharp-Schafer hypothesized that diabetes was due to the deficiency of a single chemical produced by the pancreas; he called this chemical insulin, from the Latin word insula, meaning island and referring to the pancreatic islet cells of Langerhans. In 1921, Frederick Banting and Charles Best actually discovered insulin when they reversed diabetes that had been induced in dogs with an extract from the pancreatic islet cells of healthy dogs.9,10 Together with James Collip and John Macleod, they purified the hormone insulin from bovine pancreases and were the first to use it to treat a patient with diabetes. The production of insulin and its therapeutic use quickly spread around the world. This series of events may be the most dramatic example of the rapid translation of a discovery in basic science into a benefit for patients. Once insulin injections became available, young people with insulin deficiency who had previously faced almost certain, painful death within weeks to months were able to survive for prolonged periods of time. Figure 1 shows a patient before and after she was treated successfully with insulin in 1922.11

INSULIN CHEMISTRY, BIOLOGY, AND PHYSIOLOGY
The dramatic discovery of insulin and the rapid demonstration that it is essential for human health stimulated intense interest in its chemistry and biology. A number of landmark discoveries resulted, some of which reached beyond diabetes research. For example, Frederick Sanger was awarded the Nobel Prize in Chemistry for developing methods to sequence the amino acids of proteins, and he used insulin as an example of his approaches.12 Insulin was the first hormone for which the three-dimensional crystal structure was determined (by Dorothy Hodgkin, who had previously received the Nobel Prize in Chemistry for determining the structure of vitamin B12). Donald Steiner’s demonstration in 1967 that the two-polypeptide insulin molecule is derived from a
single-chain precursor proinsulin\textsuperscript{13} was important not only for our understanding of the biochemistry of insulin but also because it applies to other peptide hormones that are transcribed as single-chain precursors. Insulin was the first hormone to be cloned\textsuperscript{14} and then produced for therapeutic use by means of recombinant DNA technology, which provided an unlimited supply of this important molecule and laid the foundation for the biotechnology industry. Figure 2 shows the structure of insulin.

The development of the radiimmunoassay for insulin by Rosalyn Yalow and Solomon Berson in 1959 permitted the quantitative measurement of pancreatic beta-cell function in animals and humans and established the radiimmunoassay as a powerful tool for measuring proteins, metabolites, and other chemicals present in very low concentrations.\textsuperscript{15} Much of our current understanding of diabetes has resulted from the ability to measure serum insulin levels.

PATHOGENESIS OF DIABETES

INSULIN RESISTANCE AND INSULIN DEFICIENCY

Over the past two centuries, we have learned that diabetes is a complex, heterogeneous disorder. Type 1 diabetes occurs predominantly in young people and is due to selective autoimmune destruction of the pancreatic beta cell, leading to insulin deficiency. Type 2 diabetes is much more common, and the vast majority of people with this disorder are overweight. The increase in body weight in the general population, a result of high-fat, high-calorie diets and a sedentary lifestyle, is the most important factor associated with the increased prevalence of type 2 diabetes. Older adults are most likely to have type 2 diabetes, although the age at onset has been falling in recent years, and type 2 diabetes is now common among teenagers and young adults.

Harold Himsworth first proposed in 1936 that many patients with diabetes have insulin resistance rather than insulin deficiency.\textsuperscript{16} We now know that insulin resistance is essential in the pathogenesis of type 2 diabetes and that the disease results from both insulin resistance and impaired beta-cell function.\textsuperscript{17} A clinical phenotype widely called the metabolic syndrome, which includes insulin resistance, upper-body obesity, hypertension, hypertriglyceridemia, and low levels of high-density lipoprotein cholesterol,\textsuperscript{18} identifies persons at high risk for glucose intolerance and diabetes. Such persons are also at high risk for cardiovascular disease and should be targeted for preventive strategies.

GENETIC FACTORS

Genetic factors play an important role in the development of diabetes. Type 1 and type 2 diabetes are polygenic disorders, and multiple genes and environmental factors contribute to the development of the disease. A few forms of diabetes (e.g., maturity-onset diabetes of the young and neonatal diabetes) are single-gene disorders that affect the pancreatic beta cell\textsuperscript{19,20} but account for only 1 to 2% of cases. In type 1 diabetes, alleles at the human leukocyte antigen locus on the short arm of chromosome 6 appear to explain up to 50% of the cases of familial clustering.\textsuperscript{21,22} In contrast, a predominant genetic susceptibility lo-
cus for type 2 diabetes has not been found. Genetic studies have identified over 40 genetic variants that increase the risk of type 2 diabetes, but in the aggregate these variants account for only about 10% of the heritability of the disorder.\textsuperscript{23,24} Individually, persons with these variants have an increased risk of diabetes of 10 to 15%, as compared with persons without the variants. The multiplicity of genes that contribute to the risk of type 2 diabetes makes it difficult to determine this risk precisely or to develop selective preventive or therapeutic strategies based on the genetic profile.

\section*{Prevention and Treatment of Diabetes}

The approach to the prevention and treatment of diabetes has been transformed since the discovery of insulin, which led to the rapid development of a widely available and lifesaving new treatment and initiated a series of advances that have fundamentally enhanced the daily lives of patients with diabetes and dramatically extended their life expectancy. Many advances have resulted from important clinical trials that were reported in the Journal and elsewhere.\textsuperscript{25-49} Some highlights of these studies include the use of biosynthetic human insulin, which has virtually eliminated local reactions at the injection site; insulin syringes and needles that are small and convenient to use and have reduced the pain of injections; home glucose monitoring,\textsuperscript{25} which together with measurements of glycated hemoglobin,\textsuperscript{26} allows therapy to be altered on the basis of accurate assessments of glucose control; and insulin pumps\textsuperscript{27} driven by computer algorithms\textsuperscript{28} that adjust insulin doses on the basis of the continuous measurement of glucose levels to achieve glucose concentrations within the physiologic range (Fig. 3). Preventive strategies and treatments for diabetic complications have undergone impressive improvements. The beneficial effects of angiotensin-receptor blockade, angiotensin-converting-enzyme inhibition, and protein restriction in preventing diabetic nephropathy have been shown.\textsuperscript{29-34} Advances in kidney transplantation have extended the lives of patients with advanced diabetic kidney disease, and laser photocoagulation has preserved the vision of millions of patients with diabetic retinopathy.\textsuperscript{35} Advances in islet-cell and pancreas transplantation have also been impressive.\textsuperscript{36,37} Recent evidence exemplified by the results of two randomized, controlled clinical trials reported this past spring in the Journal suggests that bariatric surgery to induce weight loss in patients with

\begin{figure}
\centering
\includegraphics[width=\textwidth]{structure_of_human_proinsulin.png}
\caption{The Structure of Human Proinsulin.}
\begin{quote}
Proinsulin is converted to insulin by proteolytic converting enzymes that remove the connecting peptide (C-peptide) and the lysine-arginine (Lys-Arg) and arginine-arginine (Arg-Arg) sequences of dibasic amino acids, leaving the mature insulin molecule, which consists of A and B chains connected by disulfide bonds.
\end{quote}
\end{figure}
**Detection devices**
Detection of diabetes has progressed from use of the saccharometer in the 1800s to measure urine density (a proxy for urinary glucose content) to instruments that monitor blood glucose levels at home.

**Insulin syringes**
Insulin syringes were initially glass and were used on multiple occasions, with needles that were also reused. Insulin pens, which became available in the 1990s, allow patients to vary the injected dose and to administer insulin discreetly.

**Insulin preparations**
The first highly refined form of insulin was extracted from porcine or bovine pancreas. Recombinant human insulin is now readily available.

**Insulin pumps**
The first insulin pumps, such as the Mill Hill infuser (near right), were invented in 1976 and weighed more than 0.5 kg. Current pumps are much smaller and more portable. Pumps that simultaneously infuse insulin and monitor glucose, allowing instantaneous feedback, are currently under investigation.

*Figure 3. Milestones in Diabetes Diagnosis and Management.* Photographs of the saccharometer and the early insulin preparation are from the Science Museum collection at the Science & Society Picture Library.
type 2 diabetes is much more effective than either standard or intensive medical therapy alone in lowering glucose levels and even in achieving disease remission.\textsuperscript{36,39} Advances in technology have thus profoundly improved our ability to monitor diabetic control (from urine testing to home glucose meters to continuous glucose monitoring) and to treat this disease and its complications (laser therapy for diabetic retinopathy, kidney transplantation for diabetic renal disease, and bariatric surgery to induce disease remission).

Diabetes care has been at the forefront of efforts to develop team-based approaches to patient care that involve physicians, nurses, nutritionists, social workers, podiatrists, and others and in developing models of care delivery for chronic illness. Using such an approach, the Diabetes Prevention Program showed that physical activity and weight loss can reduce the risk of diabetes in predisposed persons by 58%.\textsuperscript{40} Major effects are also seen after treatment with metformin\textsuperscript{40} or pioglitazone.\textsuperscript{41} The Diabetes Control and Complications Trial showed that improved glucose control reduces microvascular complications in type 1 diabetes,\textsuperscript{42} and the United Kingdom Prospective Diabetes Study showed the same for type 2 diabetes.\textsuperscript{43} Intensive insulin therapy to prevent hyperglycemia improves outcomes in critically ill patients.\textsuperscript{44,45}

The effect of diabetes treatment on cardiovascular outcomes and mortality is a critical issue. The Steno-2 Study showed that a multifactorial intervention aimed at improving control of glucose levels, lipid levels, and blood pressure led to a 50% reduction in cardiovascular mortality among patients with type 2 diabetes.\textsuperscript{46,47} Among patients with type 1 diabetes, improved glucose control leads to a reduction in macrovascular disease, an effect that becomes apparent only many years after the improvement has been achieved.\textsuperscript{48} The recent Action to Control Cardiovascular Risk in Diabetes (ACCORD) trial showed that aggressive glycemic control of type 2 diabetes reduced the risk of nonfatal myocardial infarction but increased overall mortality.\textsuperscript{49} The reasons for these differences between studies are not clear, but in type 2 diabetes, multiple factors increase the predisposition to cardiovascular disease. Indeed, treatment of hyperlipidemia and hypertension appears to be more effective in reducing cardiovascular events than does treatment to lower glucose levels. As a result of these and other findings, the treatments available for patients with diabetes have improved dramatically, particularly over the past 30 to 40 years.

**Prevalence of Diabetes — A Worldwide Epidemic**

Unfortunately, the improvement in outcomes for individual patients with diabetes has not resulted in similar improvements from the public health perspective. The worldwide prevalence of diabetes has continued to increase dramatically. The difficulty in applying the principles of diabetes care from the individual patient to the population reflects the unique challenges of implementing research findings and effecting behavioral change. Figure 4 shows the number and percentage of persons in the U.S. population with diagnosed diabetes between 1980 and 2010 (http://www.cdc.gov/diabetes/statistics/prevalence_national.htm). During this period, the number of diagnosed cases of diabetes increased from 5.6 million to 20.9 million, representing 2.5% and 6.9% of the population, respectively. Nearly 27% of persons over 65 years of age have diabetes. If current trends continue, 1 in 3 U.S. adults could have diabetes by 2050. The American Diabetes Association estimated that the cost of diagnosed diabetes in the United States was $174 billion in 2007,\textsuperscript{50} and efforts to prevent and treat diabetes threaten to overwhelm health systems throughout the world.

**Future Challenges**

Given the surge in the prevalence of diabetes, timely prevention of this disease at the population level is essential. Opportunities abound for the implementation of preventive public policies. Rigorous scientific methods will be needed to evaluate the effects of policy and legislative initiatives to eliminate trans fat from the diet; require restaurants to provide the caloric content of items on their menus; reduce the availability of high-calorie, high-fat foods in school cafeterias; and impose a tax on sugar-sweetened beverages. Lifestyle modification will undoubtedly play a key role in the ultimate solution to the problem of diabetes, but the necessary modifications have not been easy to implement, and more definitive solutions will depend on the ability of basic science to point prevention and treatment in new directions. Advances in basic immunology — in particular, the transformation of primitive stem cells into pancreatic beta cells — offer promise for the preven-
Figure 4. Number of Persons and Percentages of the Population with Diagnosed Diabetes in the United States, 1980–2010.
Panel A shows the number of cases of diagnosed diabetes between 1980 and 2010 among U.S. adults 18 to 79 years of age. During this period the number increased from 5.6 million to 20.9 million. Panel B shows the crude and age-adjusted percentages of the U.S. population with diagnosed diabetes for this same period of time. Data are from the Centers for Disease Control and Prevention (http://cdc.gov/diabetes/statistics/prevalence_national.html).
tion and treatment of autoimmunity in patients with type 1 diabetes. Advances in the identification of diabetes-susceptibility genes should clarify the relative role of insulin resistance and beta-cell dysfunction and identify molecular pathways and new drug targets, leading to more effective approaches to the prevention and treatment of type 2 diabetes. Although the challenges are still substantial, if we build on past accomplishments, there is every reason for optimism that another breakthrough as dramatic as the discovery of insulin will occur in the foreseeable future, with a similarly dramatic impact.

Disclosure forms provided by the author are available with the full text of this article at NEJM.org.

REFERENCES

Two Hundred Years of Progress in the Practice of Midwifery

Michael F. Greene, M.D.

No specialty shows better than obstetrics the tremendous progress that has been made in medicine in the past 200 years (see timeline, available with the full text of this article at NEJM.org). Progress in that ancient craft has occurred not in leaps and bounds or flashes of inspiration but incrementally by the incorporation of myriad scientific advances into obstetrical practice. These advances include improvements in general medicine such as aseptic technique, anesthesia, the ability to measure blood pressure and recognize its association with eclampsia, the understanding and technical developments that permit the safe transfusion of blood, and imaging technology. Equally important have been changes in attitudes toward women, which have resulted in a more central role for women in society, autonomy in controlling their own reproductive destinies, and an evolution in the nature of the relationship between the physician and patient.

To illustrate these changes, two “lectures” follow — one hypothetically given in 1912, the other in 2012. Each is faithful to the facts of the day, but neither was actually ever delivered by these fictitious speakers.

A Lecture in 1912

On a cold, snowy evening in January 1912, Dr. Walter B. Franklin, assistant visiting physician at the Boston Lying-In Hospital, steps up to the lectern to address the Obstetrical Society of Boston in its 51st year.

“Gentlemen, I would like to discuss with you this evening the current status of our specialty, including our experience at the Boston Lying-In Hospital with the first 100 cesarean sections performed at the hospital and reported in considerable detail in the Boston Medical and Surgical Journal.¹

“The first cesarean section in the Boston Lying-In Hospital was performed on July 15, 1894, and the 100th on June 29, 1907. In the present series of cases, the operation has been performed almost exclusively for a pelvic indication, there being only four cases in which pelvic deformity or the repeated loss of children in previous labors has not been the definite indication. The cases can be divided into three groups according to the classification system described by Dr. Reynolds [Table 1].² Primary sections were those that were performed before the beginning of labor. Secondary sections were performed after labor had begun but before exhaustion had set in, and late sections were, according to Dr. Reynolds, ‘performed after definite arrest of the head at the pelvic brim.’

“In the series of 100 cesarean sections, there were eight maternal deaths. This is an unduly high mortality for any abdominal operation that is usually not absolutely necessary and in which the saving of the life of the child is the chief indication for the operation. But, when we study the results according to the classification of the procedures, interesting conclusions can be drawn.

“There were 43 primary operations in this series, with the death of only one mother. The cause of death was general peritonitis, probably due to some slip in
Table 1. Classification of Cesarean Sections as Described in 1907 by Reynolds in the Boston Medical and Surgical Journal.*

<table>
<thead>
<tr>
<th>Section</th>
<th>Performance</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Primary</td>
<td>“Before the beginning of labor, or with the advent of the first pains”</td>
<td></td>
</tr>
<tr>
<td>Secondary</td>
<td>“After a certain amount of labor had demonstrated its probably unsatisfactory character”</td>
<td>“Performed before exhaustion set in”</td>
</tr>
<tr>
<td>Late</td>
<td>“After definite arrest of the head at the [pelvic] brim”</td>
<td>“Often involved a precarious situation”</td>
</tr>
</tbody>
</table>

* Classifications are from Reynolds.²

the operative technique. One newborn died from congenital heart disease a few minutes after delivery. Thus, in this group, 2.3% of mothers and 2.3% of babies, mortality of 16%. After attempts at pelvic delivery, 50% of our patients were more or less seriously compromised and one third of the babies died. Thus, today, given the mortal consequences of failed labor and failed forceps delivery, it must be recognized that the obstetrician who allows a patient to go into labor without a careful study to assess as definitely as possible the probable outcome for both mother and child, and who assumes that everything will probably go well, has failed in his professional duty. When a patient has had a previous cesarean section, some danger always exists that the scar of the uterine wound may rupture during the pregnancy, and it is very liable to rupture during labor. Therefore, it is not advisable to allow a patient who has had a cesarean section to go into labor.

“The conditions that render cesarean section unduly dangerous are previous attempts at pelvic delivery and infection of the uterus as shown by increasing temperature or as rendered probable by repeated vaginal examinations without perfect asepsis. In certain rare cases, it may be found advisable to remove the uterus after cesarean section. When the patient with the uterus already infected undergoes an operation for other conditions that render pelvic delivery impossible, a complete hysterectomy should always be performed to provide the patient with the only chance to survive, since the closing of a septic uterus and replacing it in the clean peritoneal cavity practically amounts to signing the patient’s death warrant.

“Of the 100 sections in this series, 25 were repeat operations. The reason for this large proportion is that it is the policy of the Boston Lying-In Hospital never to remove or impair the function of healthy organs to prevent subsequent pregnancy. This policy contrasts with the advice of J. Whitridge Williams.³ He questions the recommendations of some authorities who consider that
sterilization should form an integral part of every cesarean section. He recommends that, if the patient is intelligent, the decision should be left to her or her family; whereas with the ignorant it is incumbent on the physician to do what he thinks is best under the circumstances.

“Traditionally, placenta previa has been managed as early in pregnancy as possible with examination of the os uteri to determine how well it is dilated and the degree to which it is covered by the placenta. If the os is adequately dilated and the placenta can be palpated to completely cover it, then detachment of the placenta as far as the fingers can easily reach will relieve the os from the restriction that the adhesion of the placenta caused, allowing the os to dilate. If this procedure results in heavy bleeding, then the quickest and surest plan is to turn the fetus according to the method described by Braxton Hicks: bring down one of the legs of the fetus to plug the os to stop the bleeding, and labor pains will generally begin shortly [Fig. 1].

“A great deal has been written in the past few years about the advisability of performing a cesarean section in cases of placenta previa. Tait reported the first such case of treatment of ‘unavoidable hemorrhage’ by means of Porro cesarean section in Great Britain 13 years ago in the Lancet. [The Porro cesarean section is performed through a vertical midline incision that is 16 to 18 cm in length with the umbilicus at its midpoint. The uterus is incised longitudinally in its midline, and the baby is removed feet first. The cord is clamped, and the child is transferred away from the operative field. The procedure, from skin incision to delivery, should take no more than 90 seconds. The placenta is left in place. As soon as the baby is delivered, an elastic ligature is tied tightly around the upper portion of the mother’s cervix. The infundibulo-pelvic ligaments are tied and cut, and the uterus is amputated just above the rubber ligature. The cervical stump is then sewn into the lower end of the abdominal incision. The remainder of the abdominal wound is closed in the usual fashion. The stump and elastic ligature slough off, leaving a wound that granulates closed.]

“The second case of cesarean section for placenta previa was reported by W.J. Gillette from Toledo, Ohio, in the Boston Medical and Surgical Journal in 1901, despite the fact that his colleague decried the operation as ‘a very unnecessary mutilation.’ Although the maternal mortality rate from placenta previa is approximately 50% and infant mortality 66%, my own feeling is that an abdominal operation on a more or less exsanguinated patient for the sake of a premature child is not advisable, except in the rare cases in which a contracted pelvis or a rigid cervix would render a pelvic delivery difficult and dangerous.

Figure 1. A “New Method of Version in Abnormal Labor” for Managing Placenta Previa, Described by Braxton Hicks. Panel A shows manual dilation of the cervix and sweeping of the examiner’s finger around the cervix and lower uterus to separate the placenta from the cervix to permit it to further dilate. Panel B shows the use of both of the examiner’s hands, one on the mother’s abdomen and one in the uterus, to turn the fetus into breech presentation. Panel C shows the examiner grasping one leg of the fetus to bring the leg through the cervix to initiate labor and achieve tamponade of bleeding from the placental bed. Illustrations are from Braxton Hicks.
"At the present time, the successful performance of a large number of operations for various conditions has led to a gradual extension of the indications for operation until we are now standing on the threshold of one of the most marked changes that has ever to be considered in obstetrical practice. When a definite obstruction to labor, even of minor degree, is regarded as probable, is it not wiser to resort to a primary cesarean section for the benefit of both mother and child rather than to allow the patient to undergo a more or less exhausting labor, to be followed almost inevitably by a difficult pelvic operation, which is doubtful in its outcome, both to mother and child? "Gentlemen, thank you for your kind attention this evening."

**Progress in the 20th Century**

After the era in which this lecture took place, tremendous improvements in the practice of obstetrics occurred. Riva-Rocci developed a practical sphygmomanometer during the 1890s, and the availability of this device led to the widespread practice of blood-pressure measurement. Although the association between proteinuria and eclampsia had been known for centuries, the association with elevated blood pressure was not appreciated until the early 1900s. Between 1905 and 1907, case series involving women with eclampsia began to describe their elevated blood pressures. The concept of routine prenatal care was developed during the 1920s, when it seemed that eclampsia might be predicted by routinely measuring blood pressure in pregnant women. Initiation of prenatal care for this purpose was immediately rewarded with reductions in maternal mortality.

The major blood groups were defined during the first 40 years of the century, and the harsh realities of World War II brought major advances in blood banking and transfusion. Pregnant women benefited tremendously from the ready availability of transfusion to treat obstetrical hemorrhage. By the early 1930s, the pathophysiology of

---

**Figure 2. Spectrophotometric Analysis of Amniotic Fluid Showing Absorbance of Light by Bilirubin.**

Panel A shows a minimal level of bilirubin in the amniotic fluid from a normal pregnancy at 35 weeks of gestation. Panel B shows deviation in the level of bilirubin from the normal fluid curve in an affected pregnancy at 31, 35.5, and 38 weeks of gestation. Data are from Liley.
erythroblastosis fetalis was understood to involve massive destruction of fetal red cells, fetal anemia, extensive extramedullary erythropoiesis, and the release of immature nucleated red cells into the circulation. By the late 1930s, immune destruction of fetal erythrocytes with an antibody resulting from maternal–fetal blood-group incompatibility was suspected as the likely cause. There was consternation, however, over the fact that many of the cases seemed to involve mothers and babies with the same ABO group. By 1941, shortly after Levine and Stetson and Landsteiner described the rhesus red-cell antigen, there was the strong suspicion that much of the disease was based on incompatibility between an Rh-negative mother and an Rh-positive fetus. This understanding, in turn, resulted in the rapid abandonment of the previously routine practice of the use of the husband’s blood in a woman undergoing transfusion. Care of affected fetuses by appropriate timing of early delivery was improved tremendously by a 1961 study reporting spectroscopic analysis of amniotic fluid to detect bilirubin [Fig. 2]. It was not until the late 1960s, however, that anti-Rh(D) immune globulin was developed as an effective prophylaxis against Rh isoimmunization.

Undoubtedly, the most important technical innovation of the last 30 years of the 20th century was the development of high-resolution real-time diagnostic ultrasonography. It has permitted early diagnosis of potentially life-threatening ectopic gestations. It has also led to prompt, definitive diagnosis of fetal death, thus eliminating the potentially lethal problem of disseminated intravascular coagulation resulting from a retained dead fetus. It has permitted accurate diagnosis of placental position and fetal lie and presentation to inform decisions about safe methods of delivery. It has brought unprecedented opportunities to examine the fetus in utero and to treat the fetus as a patient in its own right, complete with a whole new set of diagnostic and moral dilemmas. The past 25 years have brought a wave of progress in medical genetics. There have been steps forward from the molecular genetic diagnosis of single-gene disorders in relatively small numbers of phenotypically abnormal persons to routine screening of large numbers of phenotypically normal persons to inform them about their risk of conceiving abnormal offspring. Correlations between genotype and phenotype remained challenging.

On a March morning in 2012, Dr. Katherine J. Reagan, professor of reproductive medicine at the University of California, San Diego, School of Medicine steps to the podium to address second-year medical students and brings her PowerPoint presentation onto the screen.

“Good morning. As you prepare to see patients in your Ob/Gyn rotations, let me introduce you briefly to the scope of care that you will see provided to patients who wish to have healthy families.
“First, despite a variety of available contraceptive methods, approximately half of all pregnancies in the United States are unintended, with no appreciable change in that rate in several decades. Approximately 43% of those pregnancies end in induced abortions. In 1972, the last calendar year before the Roe v. Wade U.S. Supreme Court decision, which paved the way for legalized abortion, there were 39 maternal deaths due to illegal abortions in the United States. In the most recent decade for which we have data (1998–2007), there were 2 maternal deaths due to illegal abortions—a 99.5% reduction.22

“Approximately 10% of couples will not conceive after having unprotected intercourse on a regular basis for 1 year. Those couples will increasingly turn to in vitro fertilization, or IVF. The technique of microinjection of a single sperm pronucleus into an ovum, which is known as intracytoplasmic sperm injection, or ICSI, will also be used in patients with male-factor infertility as part of their IVF [Fig. 3]. Collectively known as assisted reproductive technology, or ART, these techniques are associated with the highest per-cycle success rates among subfertile couples.23 Per-cycle live birth rates decrease with increasing maternal age. Among younger women, the rate is 50% and increases cumulatively to 70% after several cycles.24 Collectively, infertility treatments are responsible for more than 1% of all babies born in the United States today25 and more than 2% of those born in Sweden.26 Associated with the success of all types of treatments for subfertility has been a 74% increase in the rate of multiple births in the United States, from 1.9% in 1980 to 3.3% in 2008.27

“As women get older, their risks of conceiving and delivering aneuploid fetuses increase progressively.28 In 1980, women 30 years of age or older delivered 20% of the babies in the United States; between 2000 and 2008, that proportion was 35%.29 In the 1980s, the only options available to women who wished to avoid delivering aneuploid fetuses were invasive diagnostic procedures: amniocentesis or chorionic villus sampling followed by pregnancy termination of affected fetuses.30 Because of the invasive nature of the testing and the risk of procedure-related complications,31–33 testing was offered only to women at highest risk—those 35 years of age or older. The past three decades have seen several developments in noninvasive, and therefore risk-free, prenatal screening based on fetal ultrasonographic imaging [Fig. 4] and maternal serum biochemical markers.34–36 As compared with a population-based maternal age-specific risk, these techniques derive an estimate of risk of fetal aneuploidy that is specific to a pregnancy in an individual woman. Women at high risk for single-gene disorders or aneuploidies who want to avoid these births but find pregnancy termination unacceptable may choose preimplantation genetic diagnosis. In this process, a single blastomere is removed from each of the embryos created by means of IVF and ICSI and studied with the appropriate molecular genetic techniques to enable transfer into the uterus of only genetically normal embryos.37 Recognition that measurable quantities of cell-free fetal DNA appear in the maternal circulation very early in gestation38 has resulted in successful noninvasive prenatal diagnosis of single-gene disorders in fetuses of women known to be at risk.39,40

Figure 5. Radiographic Image of the First Successful Intrauterine Fetal Transfusion Performed under Fluoroscopic Guidance by Liley, in 1963.

As described when the procedure was reported, the radiograph shows “contrast medium and the coiled catheter [arrowhead] in the [fetal] peritoneal cavity. The Tuohy needle [arrow] has been withdrawn and lies on the mother’s abdominal skin.”44
It also promises the potential that batched, or ‘multiplexed’ massively parallel sequencing could accomplish highly sensitive and specific mass population screening for aneuploidies at an affordable price. New techniques identify new genetic differences that may or may not be responsible for phenotypic consequences but are most certainly responsible for difficult conversations between doctors and patients and agonizing decisions for patients. Most recently, the entire genome of a human fetus has been sequenced noninvasively from cell-free fetal DNA in the maternal circulation.

“Sir William Liley was the first person to treat a fetus in utero when he performed an intrathecal transfusion into the peritoneal cavity under fluoroscopic guidance for severe isoimmune hemolytic disease in New Zealand in 1963 [Fig. 5]. Enthusiastic reports of noncontrolled case series of ‘fetal surgery’ in the 20th century ultimately resulted in randomized, controlled trials in the 21st century, with a more sober assessment of mixed results.

“You will note careful attention to screening for and treatment of infections, especially human immunodeficiency virus (HIV) infection in pregnancy. Among women with untreated HIV infection in pregnancy, the probability of transmitting the infection vertically to their fetuses and neonates is 25%. Zidovudine treatment in the mother during pregnancy, labor, and delivery and in the newborn for several weeks after birth was first shown in 1994 to reduce the risk of vertical transmission of HIV to 8%. Treatment of the mother with highly active antiretroviral therapy that is successful in reducing the maternal viral load to less than 50 copies per milliliter can reduce the risk of vertical transmission to less than 0.2%.

“In the labor-and-delivery suite, you will encounter physicians and patients engaged in conversations about the most appropriate route of delivery for the individual patient. Such conversations are a relatively new phenomenon. For the past 100 years, it has been the patient’s assumption and the obstetrician’s practice that women would labor and deliver vaginally, with cesarean delivery reserved for clinical circumstances that made vaginal delivery immediately dangerous for the mother or fetus. This change in the standard of practice has resulted in a progressive increase in the cesarean delivery rate in the United States from single-digit percentages in the first decade of the 20th century to 32.8% in 2011 [Fig. 6]. Currently, maternal mortality is one 100th and newborn mortality one 100th of the corresponding rates at the start of the 20th century. Dramatic regional and other practice-based differences in cesarean delivery rates without similar differences in maternal or newborn outcomes make it clear that current high cesarean delivery rates are not absolutely necessary to achieve the observed benefits in maternal and newborn outcomes. Furthermore, as risks of all types of adverse outcomes decrease in modern obstetrical practice and care providers respect the autonomy and decisions made by prospective parents, many of our traditional perspectives and practices are changing. It would be the worst form of hubris for clinicians to think that we can anticipate the details of obstetrical practice 100 years from now, any more than an obstetrical care provider 100 years ago could have predicted where we stand today.”

The Future

Future generations will judge the success of changes introduced into obstetrical care in the next 100 years by a more expansive, complex, and multidimensional set of criteria than have generally been used to date. They will expect that clinicians will continue to reduce morbidity and mortality while increasing the options and autonomy...
of patients in their care. But they will also demand that, as those goals are achieved, clinicians will assume more responsibility and accountability for the financial burden that their decisions and practices place on society. While those of us who provide care assume a greater stewardship role over costs and value for patients who currently have access to care, we must negotiate for adequate societal resources to bring care to those who are currently without it. As necessary and important as these achievements will be, they will seem somewhat selfish and parochial if we do not also find ways to reduce the burden of perinatal morbidity and mortality in areas of the world that presently have outcomes that are similar to those in the United States a century ago. Science and technology will undoubtedly help extend and deliver care in new ways; however, more fundamentally, progress in global obstetrical health will depend on developing the infrastructure, political will, and culture that value the health of women and their pregnancies.

Disclosure forms provided by the author are available with the full text of this article at NEJM.org.

REFERENCES


A Glimpse of the Next 100 Years in Medicine

Isaac S. Kohane, M.D., Ph.D., Jeffrey M. Drazen, M.D., and Edward W. Campion, M.D.

Over the past year, the Journal has commemorated 200 years of publication and the astonishing progress made since 1812 in the science and practice of medicine. Thanks to digital technology, our anniversary celebration has had a wide reach, with more than 1 million visits to the NEJM200.NEJM.org site; many viewers of our documentary video, Getting Better; and large numbers of viewers of our simulcast symposium, Dialogues in Medicine. As the Journal’s 200th year comes to a close, we want to think about the changes and challenges that medicine faces in the decades ahead. Although it is foolish to attempt specific predictions about the future, it would be unwise not to think about the emerging trends, new opportunities, and the principles that should guide the medicine of the future.

In the decades ahead, the pace of biomedical discovery will accelerate. The state of an individual person will be characterized with increasing precision from the molecular level to the genomic level to the organ level and by interactions with medications, nutrients, the microbiome, therapeutic devices, and the environment. This precision medicine will become possible because of huge data sets on large populations, with millions of characterizations of each person. Study populations will grow to millions, which will allow observational studies with novel statistical methods that will allow discovery of useful, reproducible patterns and relationships from these data. This will be possible because virtually all the data will be in an advanced infrastructure of electronic health records (EHRs) that includes input from physiological monitoring, which is already starting to become part of the management of chronic diseases and of guidelines for prevention and fitness. Therapeutic and monitoring instruments will continue to become smaller, smarter, more interactive, and more connected to the health information infrastructure. However, the quantum leaps will come not from the devices but from inferences drawn from the data.

The size and complexity of this multidimensional characterization of patients will lead to far more complex diagnostic and prognostic categories than are currently in use. The multivariate descriptors of large populations will allow stratification of a kind seen only in the most recent genomically informed clinical trials. Massive data crunching will yield analytic or algorithmic formulas that will be useful for clinical purposes even though they defy easy summary in a language most of us can understand. Complex but empirically validated algorithms will be embedded in EHR systems as decision support tools to assist in everyday patient care. Those management algorithms will evolve and be modified continuously in accordance with inputs from ongoing clinical observations and from new research. Clinical decision support algorithms will be derived entirely from data, not expert opinion, market incentives, or committee consensus. The huge amount of data available will make it possible to draw inferences from observations that will not be encumbered by unknown confounding.

Both patients and payers will demand increased transparency, particularly for new therapies that will have to be monitored in ongoing studies of comparative efficacy. This will increase the pressure on regulatory authorities for greater plasticity that will allow them to adapt rapidly, accurately, and decisively to the evolving understanding of the merits and risks of different therapies. Scientists, physicians, and the public
will demand that all the primary data be made public, along with the analytical tools necessary to reanalyze, test, refine, and build on them. Data security will have to evolve and thereby win the public's trust with new techniques that will do what now seems impossible: guarantee protection of privacy while providing detailed information about each person. Societies will come to accept that comprehensive knowledge of disease, prevention, and effective treatment is an essential public good.

Biomedical research, data technologies, and clinical care all require resources, but the era of shifting more and more economic resources toward health care is going to end. The medicine of the future will focus on more efficient use of resources to prevent disease, with the goal of delivering what provides the best value for the patient who needs treatment. The future of medicine also depends on reducing the enormous disparities in health, particularly those between the richest and the poorest countries of the world. A basic standard of sound medical care will become an expectation of every society. Research-rich countries may come to see that achieving basic health care throughout the world is a strategy to promote stability and peace. The increasing power of information and communication technologies can help find ways to improve global health. However, that goal also requires the educational and economic development that are essential for societies to achieve a reasonable standard of health. The moral mandate here only becomes stronger as clinical progress continues to accelerate in developed societies.

The high-technology, information-rich medicine of the future will provide powerful and useful tools for clinical medicine. The medicine of the future will not, of course, solve all problems, and it cannot prevent violent or self-destructive human behaviors. Patients will continue to rely on physicians and the medical community for the guidance, support, and help that only a skilled and caring health professional can deliver. The medical community must provide direction to ensure that powerful new technologies are used to benefit the health of all. As advances in science and technology continue to bring disruptive changes, the Journal must continue to evolve creatively in order to continue in its mission of inspiring discovery and advancing care. As we head into this medicine of the future, the Journal should remain true to the principles that were set down by its founding physicians two centuries ago: “The Journal will always be open to the accurate observer of nature, the useful experimenter, and the rational therapist.”

Disclosure forms provided by the authors are available with the full text of this article at NEJM.org.

From Boston Children’s Hospital and Harvard Medical School — both in Boston (I.S.K.).


DOI: 10.1056/NEJMe1213371
Copyright © 2012 Massachusetts Medical Society.
Celebrating 200 Years
at the Forefront
of Medicine

A Special Collection of Review and Perspective
Articles Published in 2012 to Commemorate
The NEJM 200th Anniversary.