PCORI's Problem: Reasons Are Not Causes

November 5th, 2012

by David Introcaso

Editor's note: *In another* Health Affairs *Blog post also published today, Joe Selby, the executive director of the Patient-Centered Outcomes Research Institute, responds to David Introcaso's post below. For more on the concept of patient centeredness, comparative effectiveness research, and the Patient-Centered Outcomes Research Institute, see* Health Affairs*' October issue, "Current Challenges In Comparative Effectiveness Research."*

This past May, the Patient-Centered Outcomes Research Institute (PCORI) approved five research priority areas: "assessment of prevention, diagnosis and treatment options"; "improving health care systems"; "communication and dissemination"; "disparities"; and, "patient-centered outcomes research and methodological research". In June PCORI announced 50 pilot project research awards totaling $30 million. PCORI is anticipated to spend $3 billion between now and 2019.

What does it mean to be "patient centered" and what does this then mean about ways of "improving "health care systems" and "communication and dissemination"? PCORI belies its "patient centered" mandate since it has not put a primacy on understanding and improving the interaction between the patient and the provider — the only way the quality of health care delivery is ultimately improved. None of the fifty PCORI pilot projects examine the quality of these interactions. Understanding and improving patient-provider interactions also explains how clinical evidence is produced or becomes meaningful.

What makes examining these interactions essential is that knowing and doing are not the same. Consider hand hygiene, likely *the* most basic health care practice. Though the benefits of hand washing have been well known since Semmelweis, studies show hand washing adherence rates are as low as 30 percent and frequently no better than 50 percent.

Why then is it that where evidence is well known, practice (or doing) does not follow, and what does this suggest about of PCORI's chances of filling the quality gap by producing and then disseminating ever more evidence?

The Accepted Paradigm

PCORI's problem or challenge, as hand hygiene compliance rates sadly demonstrate, is that reasons cannot be causes. Since the two cannot be conflated, or since knowing things and doing things do not fold together, research evidence never simply disseminates and transfers.

The accepted paradigm we use for improving health care delivery is one where we assume evidence or knowledge is first produced and then communicated and disseminated. This hypothetical linear process is what the World Health Organization terms the "know-do" gap. Similarly, the National Institutes of Health "Roadmap" program is designed to accelerate "from bench to bedside" the transfer of evidence to practice, and the Agency for Healthcare Research and Quality's "Office of Communication and Knowledge Transfer" is charged per the Affordable Care Act with creating tools to disseminate PCORI's evidence. These and other similar efforts assume an ability to disseminate discrete evidence or

knowledge in a linear and mechanistic fashion.

This method for improving health care quality means the use of "sound science" to produce "right" knowledge, or the use of explicit rules by which scientific propositions to improve health care can be obtained. This in turn means the generation of new evidence preferably by randomized control trials or at least by systematic reviews, e.g., the Cochrane Collaborative and now PCORI. The next step is to disseminate or transfer new knowledge, or have it travel along a "translational highway" to be brought to ever larger national scale. The effort becomes, to use the journal title, an "implementation science" challenge.

Understandably then, improving health care delivery takes on engineering or technical properties. Improvement means re-engineering or wedding engineering sciences with health care. This means largely driving out variation or bringing to ever larger "scale" uniform care delivery via processes such as Six Sigma or Toyota Lean Production. In the examination room it makes sense to talk about a clinician meeting a standard or a guideline and to then talk about a patient's compliance with that guideline.

The paradigm thus provides the rationale for numerous other activities intended to improve the quality of health care delivery. It underpins continuing medical education and related professional development activities, as well as many knowledge management and research dissemination programs and other quality improvement and organizational structure or system re-design efforts. It serves as the underlying rationale for evidence-based medicine, performance measures, quality indicators, report cards, check lists, pay-for-performance, and other approaches or programs all endeavoring to transfer evidence to practice.

But does this paradigm usefully explain how we know, and the know-how of improving health care quality? Does evidence creation actually precede the delivery of care to be then simply communicated and disseminated?

An Alternative Paradigm

Unfortunately, progress made in improving care delivery using the accepted paradigm has proven, to be overly polite, slow. Despite all the evidence on ways to improve quality and reduce medical errors and medication non-compliance, improving quality and outcomes remains largely a puzzle. Paradigm testing, Thomas Kuhn said, occurs when "persistent failure to solve a noteworthy puzzle has given rise to crisis" – or times like now. Rather than tinkering with the present paradigm, let's consider an alternative.

Let's assume knowledge does not arise singularly first in the mind of one individual, to be then transferred to the mind of another individual. Instead, let's assume the creation of knowledge or evidence actually begins with one's response to another's gesture and continues to build in the ongoing back and forth of the gesture and response of people communicating. Evidence or knowledge creation is then inherently a social act, the product of mutual adaptation. It is not meaning independent. No one individual owns knowledge. It cannot be stored or managed or simply be disseminated or transferred. In sum, people make sense of the world together. Reality, less brute facts, is a negotiated interpretation. As John Searle explains, it is a brute fact that the heart pumps blood, but defining its healthy functioning is not determined via evidence but by a value or goal to which we agree.

Reasons then are not causes. Evidence for an intervention is not by definition evidence for that intervention. Pronovost's checklist worked because clinicians in Michigan and elsewhere agreed the evidence "commodity" held value or was meaningful. Science determines only the strength of the evidence that exists for any particular hypothesis. It does not presuppose a purpose or end. That's teleology.

Under this paradigm, evidence creation and delivery improvement are not consecutive — they are more coherently understood as entwined, commonly occurring together in real time. Health care system improvement via systems thinking is actually a fantasy since systems are an abstraction of human interaction — and human interaction is all there is. Physicians and patients, being people, unlike objects in nature, are always self-interpreting (reflective and reflexive) entities. People are not planes. They cannot be engineered. There is ultimately no evidence or quality improvement, and moreover no meaning, absent one person's response to another person's gesture.

Consider this example. Dr. Warren Warwick, profiled in an Atul Gawande essay, is highly effective in treating his cystic fibrosis patients because his practice is primarily relational. Though Gawande illustrates Warwick's success for other purposes, he does finely detail Warwick's ongoing back and forth

InsureBlog
Managed Care Matters
NASHP State Health Policy Blog
Patient POV
Physicians for a National Health Program
Prescriptions (New York Times)
Reforming Health
Roper On Health
Running A Hospital
The California HealthCare Foundation's On Practical Progress
The Health Care Blog
The Incidental Economist
The Medical Professionalism Blog
Wall Street Journal Health Blog
Wing of Zock
Wright On Health

interactions with his patients. He focuses on Warwick's interaction with a particular young female, during which Warwick tries to make sense of the patient's reduced lung capacity by persisting in asking her about coughs, colds, treatment frequency, etc.

Eventually Warwick learns she has a new boyfriend and job and for these reasons she had been skipping her treatments. Learning this, Warwick is now able to work out an agreed-upon, meaningful treatment plan with his patient to reverse her functional decline. Not surprisingly we learn Warwick is disdainful of clinical guidelines, telling Gawande they are "a record of the past and little more."

Implications

Advancing evidence and improving practice are not separate activities but intertwined processes inherently social and occurring in real time. Practice, as Thomas Schwandt argues, is not assumed to stand "in subsidiary relationship to scientific knowledge." (See Schwandt's writings on this topic here and here.) The practice setting is a site for an unfolding of events, not a context for applying evidence.

This suggests the fundamental way to improve health care quality is by designing delivery that improves the interaction between the provider and the patient. Since health care (along with biological functioning) emerges from relationships, Paul Uhlig argues the key to transformation is in optimizing the patterns of patient-provider interaction. "Patterns of organization should reflect patterns of interaction," Uhlig says, "and patterns of organization should match patterns of patient need."

As the Warwick example illustrates, attention should be paid to what is actually going on in conversation between actors. Anthony Suchman terms this simply "relationship-centered care." In writing about improving health care effectiveness, Yaneer Bar-Yan makes an analogous point when he argues the fine-scale task of providing individualized care should not be corrupted by the health care industry's drive toward large-scale, undifferentiated functions.

The importance of the qualities of interactions cannot be over-stated. Focusing on what is "actually going on" is all that ultimately matters, since again improvement emerges only between people interacting in real time. We produce nothing of consequence outside these interactions. At present, the focus is instead on what should be. Attention becomes focused on how the evidence commodity is managed. This is reflected in the bulk of studies on improving health care quality that typically attempt to enumerate the properties or attributes of good care. Instead of learning what people are actually doing with one another, the literature is largely about practice or business improvement schemes or systems, measurement, or reimbursement. These activities are "as if" exercises, mere conceits since nothing is ultimately produced outside of the interaction between clinician and patient.

"As if" exercises constitute the sum of PCORI's fifty pilot projects. None among the fifty endeavor to provide evidence of the quality of interaction between the patient and provider. The projects instead propose developing aids, conceptual frameworks or models, databases, decision support, evaluation and gaming tools, frames, guidelines, instruments, lexicons, measures, mobile technologies, models, portals, scales, surveys, quality dimensions, tool-kits and other techniques. Similarly, in PCORI Executive Director Joe Selby's lengthy interview published in *Health Affairs* last winter, there was no mention made of improving how well a provider interacts with his or her patient (nor is there mention in PCORI's Draft Methodology Report).

Rhetorically, how is PCORI "patient centered" when patient interactions go ignored? This is all the more remarkable when you consider that our relation to the other, as Levinas noted, is the foundation of our knowing, not the reverse.

This entry was posted on Monday, November 5th, 2012 at 2:14 pm and is filed under All Categories, Comparative Effectiveness, Consumers, Health Reform, Physicians, Policy, Research, Science and Health. You can follow any responses to this entry through the RSS 2.0 feed. You can leave a response, or trackback from your own site.

Leave a Reply

You must be logged in to post a comment.

Enhancing Communication To Advance Research And Practice Is Among PCORI's Key Priorities

November 5th, 2012

by Joe Selby

Editor's note: *Joe Selby's post below responds to David Introcaso's* Health Affairs *Blog post, also published today. For more on the concept of patient centeredness, comparative effectiveness research, and the Patient-Centered Outcomes Research Institute, see* Health Affairs' *October issue, "Current Challenges In Comparative Effectiveness Research."*

The *Health Affairs* Blog post by David Introcaso, PhD, "PCORI's Problem: Reasons Are Not Causes," emphasizes the critical importance of the patient-provider relationship and communication in improving health outcomes. It also notes the challenges to changing practice through the creation and dissemination of new evidence, including the non-linear nature of the process of turning knowledge into action. But it argues that PCORI has missed the boat on these issues, specifically having failed to address them either in the $30 million in Pilot Project awards we issued earlier this year or in our current research priorities and funding announcements.

We genuinely appreciate the serious thought Dr. Introcaso has given to these matters and fully agree with him on their importance if we are to change practice and improve patient outcomes. And that's why both our work to date and the work we continue to pursue focuses directly on these questions in a number of concrete ways – a point he seems somehow to have missed.

The significance we at PCORI place on communication between patients and those who care for them is evident throughout our work: in our foundational research roadmap; in our initial round of research support (the Pilot Projects Program); and in the scope and detail of our primary research funding announcements' requirements.

PCORI's Ongoing Focus On Patient-Provider Relationships

One of our first actions was to adopt our National Priorities for Research and Research Agenda to set a framework for our research funding. "Communication and Dissemination Research" is one of our five priorities, and our funding announcement for this priority calls for studies of communication between patients and clinicians, of shared decision making, of considering patients values and preferences in the communication, and of training clinicians in the practice of shared decision making.

Our emphasis on better understanding patient-provider communication as a means of improving practice and outcomes also is evident in our Pilot Projects. One of the eight specific areas of interest listed in the solicitation for our $30 million Pilot Projects program was the study of the interactions of patients with their clinicians and healthcare systems.

Among the 50 projects funded, about two dozen include some type of focus on the patient-provider relationship and how it may be leveraged to improve care. This includes projects that are specifically developing decision-support tools, as well as others that note a goal or aim related to improving communication or interaction between patients and providers. Examples range from a project studying

ways to create more open and effective patient-provider communication generally to a series of studies aimed at developing targeted interventions to bring patient preferences and concerns fully into clinical decision making in the management of chronic and mental health conditions and improving care for the underserved, elderly and those with rare diseases. The full list of these initiatives can be seen here.

We clearly emphasize this approach further in the first round of our PCORI Funding Announcements (PFAs), which solicited $12 million in proposals specifically focused on communication and dissemination research. This work will generate greater insight into how we can improve the quality of care and improve outcomes by supporting communication to clinicians and patients, as well as among patients and their caregivers and providers.

An Emphasis On Quickly Affecting Practice And Outcomes

PCORI is also keenly aware that practice change does not necessarily follow the appearance of new evidence, no matter how sound. We are committed to addressing this gap and to having a prompt impact on practice and outcomes. Several specific steps reflect this commitment. Our pilot projects also called for studies of engaging patients in the translation of evidence into practice at the level of the health care system, and at least 19 of the funded projects address this directly. We agree with Dr. Introcaso that systems re-engineering will often be necessary. For this reason, improving health care systems is the second of our five national priorities for research.

Key requirements within all PCORI's funding announcements provide further evidence of our intention to link clinical research more tightly with dissemination and implementation of study findings. We require that both patients and other relevant stakeholders, often clinicians, be fully and meaningfully integrated into all aspects of the research process. We will not fund research that does not meet this requirement, because we believe that having the potential end users of the research present and involved throughout the process will increase chances that the findings of our studies are taken up, implemented and disseminated. Because we agree "good data" alone won't significantly change practice, we also require that applicants for PCORI funding assess the potential and the barriers to their study findings being implemented widely into practice.

PCORI's approach to patient-centered research values collaboration between stakeholder communities and strives to build a network of activated patients and engaged providers. Through this model, we are addressing the critical patient-provider communication issues that so often govern clinical practice. Our success is dependent on the level of input we receive, and the strength of the network we can create. We look forward to continued feedback from all members of the health and health care community as we advance toward our goal of bringing the patient's voice to comparative effectiveness research.

Email This Post Print This Post

This entry was posted on Monday, November 5th, 2012 at 2:24 pm and is filed under All Categories, Comparative Effectiveness, Consumers, Health Reform, Innovation, Policy, Quality, Research. You can follow any responses to this entry through the RSS 2.0 feed. You can leave a response, or trackback from your own site.

Leave a Reply

You must be logged in to post a comment.

InsureBlog
Managed Care Matters
NASHP State Health Policy Blog
Patient POV
Physicians for a National Health Program
Prescriptions (New York Times)
Reforming Health
Roper On Health
Running A Hospital
The California HealthCare Foundation's On Practical Progress
The Health Care Blog
The Incidental Economist
The Medical Professionalism Blog
Wall Street Journal Health Blog
Wing of Zock
Wright On Health

By Justin W. Timbie, D. Steven Fox, Kristin Van Busum, and Eric C. Schneider

DOI: 10.1377/hlthaff.2012.0150
HEALTH AFFAIRS 31,
NO. 10 (2012): 2168–2175

Five Reasons That Many Comparative Effectiveness Studies Fail To Change Patient Care And Clinical Practice

Justin W. Timbie (jtimbie@rand.org) is a health policy researcher at the RAND Corporation in Arlington, Virginia.

D. Steven Fox is a policy analyst at the RAND Corporation in Santa Monica, California.

Kristin Van Busum is a health project associate at the RAND Corporation in Boston.

Eric C. Schneider is a senior scientist and director at the RAND Corporation in Boston, Massachusetts.

ABSTRACT Despite widespread enthusiasm about the potential impact of new investments in comparative effectiveness research, recent history suggests that scientific evidence may be slow to change clinical practice. Reflecting on studies conducted over the past decade, we identify five causes that underlie the failure of many comparative effectiveness studies to alter patient care. These are financial incentives, such as fee-for-service payment, that may militate against the adoption of new clinical practices; ambiguity of study results that hamper decision making; cognitive biases in the interpretation of new information; failure of the research to address the needs of end users; and limited use of decision support by patients and clinicians. Policies that encourage the development of consensus objectives, methods, and evidentiary standards before studies get under way and that provide strong incentives for patients and providers to use resources efficiently may help overcome at least some of these barriers and enable comparative effectiveness results to alter medical practice more quickly.

The Patient-Centered Outcomes Research Institute is poised to develop and oversee a vast portfolio of new comparative effectiveness research. For this endeavor to transform patient care, new evidence must be disseminated to clinicians and patients, understood and considered relevant, and used in the decisions that inform clinical care.

Several treatment approaches have changed promptly in response to new evidence, often from randomized, placebo-controlled clinical trials. For example, hormone replacement therapy was widely adopted as a treatment for preventing heart disease in women. But it was abandoned by most physicians soon after a clinical trial found that the therapy increased rates of heart attacks and other adverse events.[1]

Similarly, after autologous bone marrow transplantation was found to be effective for patients with leukemia, its use increased among women with breast cancer. But after studies showed that the treatment was no better than conventional chemotherapy and had much higher risks of serious side effects, its use in treating breast cancer declined dramatically.[2]

However, decades of experience suggest that translating evidence into changes in clinical practice is rarely rapid and that treatments with weak evidence of effectiveness are often rapidly adopted. Despite mounting evidence that prostate specific antigen screening offers little or no benefit, the test is still widely used.[3] Moreover, cost-effective treatments such as thiazide diuretics for patients with hypertension are adopted far more slowly than heavily marketed treatments that are less effective.[4]

As policy makers seek to maximize the impact of new comparative effectiveness evidence on clinical practice, useful lessons might be drawn from prominent publications of the past decade. Examining the events that followed a selected

set of these publications, we found that five root causes appear to explain the failure of many comparative effectiveness studies to translate into changes in clinical practice. We offer policy solutions that address three key phases of the process by which evidence is translated into practice.

Conceptual Framework For Evidence Translation

To better understand the barriers and facilitators of evidence translation, we divided the traditional research translation process into five "steps" (see the online Appendix).[5]

First is the generation of comparative effectiveness results, which include both the design and the conduct of a study. Second is the interpretation of a study's results, when stakeholders assess the quality and relevance of the evidence and begin to formulate recommended changes to clinical practice. The third step is the formalization of results, at which point clinical experts, health information technology vendors, and other experts convert the recommended changes into clinical practice guidelines, performance measures, and clinical decision support tools. Fourth is dissemination, the process by which the formalized results and related tools are transmitted to local stakeholders. The last step is the implementation of new clinical practices by professionals and patients in local settings.

Each of these activities occurs within a context shaped by existing practice, professional expectations, financial incentives, local market demands, and regulation—all of which can promote or impede translation.

Root Causes For Incomplete Evidence Translation

With the aid of this framework, we reviewed selected studies from the past decade of comparative effectiveness research translation and identified a number of factors (root causes) that consistently predicted failure to translate evidence into clinical practice. For this analysis, we drew on carefully selected peer-reviewed publications and discussions with knowledgeable stakeholders.[6]

The root causes manifest themselves in distinct ways and often exert their effects in multiple steps. In the following sections of this article, we describe these factors and provide examples of comparative effectiveness studies in which they affected translation.

MISALIGNMENT OF FINANCIAL INCENTIVES Economic incentives, including the pervasiveness of both fee-for-service reimbursement and generous insurance coverage, are among the most commonly cited factors that lead providers and patients to select treatments that are inconsistent with evidence from comparative effectiveness research.[7] Two of the most widely cited comparative effectiveness studies of the recent past, COURAGE (Clinical Outcomes Utilizing Revascularization and Aggressive Drug Evaluation)[8] and SPORT (Spine Patient Outcomes Research Trial),[9] found that conservative treatment approaches for stable coronary artery disease and spinal stenosis, respectively, may be no less effective than more invasive treatments for many or even most patients (Exhibit 1).

Under fee-for-service payment systems, however, invasive treatments such as percutaneous coronary intervention and spinal surgery are reimbursed generously compared to their alternatives, while counseling patients about treatment options goes mostly unreimbursed. Meanwhile, most payers impose few coverage or payment obstacles for many invasive procedures, either because their appropriateness cannot be monitored easily or because payers are likely to face organized challenges from pharmaceutical or device manufacturers, professional societies, or patient advocates when they attempt to use comparative effectiveness evidence to modify coverage policies.

In sum, perverse financial incentives push both patients and providers to disregard the evidence and pursue aggressive treatments even if they are no more effective than more conservative treatment approaches.

Economic incentives influence all other steps of the translation pathway, including subsequent interpretation, formalization, and dissemination of the evidence. For example, pharmaceutical and device manufacturers have commonly used a variety of approaches—such as paying key opinion leaders to disseminate favorable messages to their peers (for example, through speakers' bureaus) or through the mass media, "detailing" programs to educate clinicians about medical technologies, and direct-to-consumer advertising—to influence interpretation of the evidence and adoption.[10]

All of these tactics were observed both before and after the release of results from CATIE (Clinical Antipsychotic Trials of Intervention Effectiveness)[11]—a landmark comparative effectiveness study that found that newer "atypical" antipsychotics were no more effective than conventional antipsychotics for patients with schizophrenia.

Once comparative effectiveness results are published, stakeholders may report them selectively, downplaying findings unfavorable to their economic interests and highlighting favorable

EXHIBIT 1

Recent Comparative Effectiveness Studies And Their Translation Outcomes

Study	Results	Dominant practice	Translation outcome
ALLHAT[a]	Thiazide diuretics were superior in preventing cardiovascular disease events and less expensive than calcium channel blockers, ACE inhibitors, and alpha-adrenergic blockers	Trend toward greater use of calcium channel blockers and ACE inhibitors	No change in practice
CATIE[b]	Conventional antipsychotics were as effective as atypical antipsychotics in patients with schizophrenia	Use of atypical antipsychotics	No change in practice
COMPANION[c]	Compared to optimal medical therapy, both cardiac resynchronization therapy (CRT) and CRT plus defibrillator use improved survival, reduced hospitalization rates, and improved functional status in patients with moderate to severe heart failure	Optimal medical therapy (underuse of CRT)	Some evidence of slow adoption of CRT (with and without defibrillator)
COURAGE[d]	Optimal medical therapy combined with percutaneous coronary intervention (PCI) had similar survival benefit and angina relief, compared to optimal medical therapy alone in patients with stable coronary artery disease	Use of PCI before optimal medical therapy	Little or no change in practice
SPORT[e]	Surgery for lumbar spinal stenosis had better outcomes than nonsurgical treatment, according to the cohort study results	Surgical treatment	No change in practice
Philadelphia Bone Marrow Transplant Group[f]	Compared with maintenance chemotherapy in conventional doses, high-dose chemotherapy plus bone marrow transplantation did not improve survival in women with metastatic breast cancer	Trend toward greater use of bone marrow transplantation	Rapid abandonment of bone marrow transplantation

SOURCE Authors' analysis of data from various sources; see below. **NOTE** ACE is angiotensin converting enzyme. [a]Antihypertensive and Lipid-Lowering Treatment to Prevent Heart Attack Trial. See Note 12 in text. [b]Clinical Antipsychotic Trials of Intervention Effectiveness. See Note 11 in text. [c]Comparison of Medical Therapy, Pacing, and Defibrillation in Heart Failure. See Note 14 in text. [d]Clinical Outcomes Utilizing Revascularization and Aggressive Drug Evaluation. See Note 8 in text. [e]Spine Patient Outcomes Research Trial. See Note 9 in text. [f]See Note 2 in text.

messages. According to one orthopedic surgeon, a widely disseminated and skewed interpretation of the SPORT trial—that "surgery works for back pain"—seemed to include complex spinal procedures that, although frequently used in practice, were rarely used in the trial itself.

Dissemination activities—such as academic detailing that rely on independent researchers or clinicians to disseminate unbiased information to practicing clinicians—are often inadequately financed. For example, the use of academic detailing to disseminate the results of ALLHAT (Antihypertensive and Lipid-Lowering Treatment to Prevent Heart Attack Trial),[12] which found that thiazide diuretics were more cost-effective than three other widely used classes of medications for the prevention of cardiovascular disease, may have been only modestly effective because of limited funding.[10]

Formalizing evidence through guidelines and quality measures is often prohibitively expensive without financial support from interested stakeholders. But professional societies may also resist formalization activities that adversely affect their members' financial or professional interests. Differences in the interpretation of the COURAGE results divided almost entirely along subspecialty lines: Most interventional cardiologists criticized the study's design, while cardiologists who did not specialize in percutaneous coronary interventions generally embraced the results.

It is not clear whether conflicts of interest on guidelines committees have led to liberal guidelines or appropriateness criteria for interventions like percutaneous coronary intervention. Nevertheless, the potential for introducing bias is a concern, considering that up to half of the guidelines relevant to this procedure are based on expert opinion.[13]

AMBIGUITY OF RESULTS Even the most carefully designed and conducted comparative effectiveness studies rarely produce definitive results. The number of outcomes assessed in a typical study and varying opinions about clinically meaningful effect sizes pose major barriers to reaching consensus on study results. For example, the COMPANION trial (Comparison of Medical Therapy, Pacing, and Defibrillation in Heart Failure),[14] which compared optimal medical therapy to cardiac resynchronization therapy with or without an implantable cardioverter defibrillator, found that optimal medical therapy produced poorer survival rates (the study's primary outcome).

However, a difference in a single secondary outcome between the two device arms of the trial led many cardiologists to conclude that cardiac resynchronization with an implantable defibrillator was the best approach of the three—

An explicit consensus on the goals of each study could minimize variability in the interpretation of results.

igniting a debate that persisted for years after the trial.[15] Without consensus on evidentiary standards prior to the release of comparative effectiveness results, ambiguous results become fuel for competing interpretations, making it difficult for providers, insurers, and policy makers to act on the evidence.

Methodological critiques of comparative effectiveness studies, whether or not they are justified, may also complicate interpretations of the results. In attempting to replicate real-world practice settings, these studies may sacrifice internal validity. For example, if patients who are assigned through randomization to one treatment choose instead to receive the other treatment in large numbers (as was observed in the COURAGE and SPORT studies), it can be difficult to interpret the results.

Comparative effectiveness studies that challenge existing practices are frequently criticized on a wide range of methodological issues in an effort to raise doubts about their results. This phenomenon was noted as early as three decades ago, following the publication of results from the Veterans Affairs Coronary Artery Bypass Graft Surgery Cooperative Study.[16] That study found a substantial benefit of surgery over medical therapy among patients with severe coronary artery disease. But when results for the subpopulation with less severe disease were subsequently released and showed no advantage for surgery, the quality of the study was roundly criticized. Providers may have been unduly influenced by similar criticisms of the CATIE and COURAGE trials, despite the rigorous methods used in both studies.

COGNITIVE BIASES IN INTERPRETING NEW INFORMATION Three common cognitive biases that affect the processing of new information may help explain why some research evidence fails to lead to changes in clinical practice.

The response to the CATIE trial exemplifies the problem of the first, confirmation bias,[17,18] the tendency to embrace evidence that confirms preconceived ideas and to reject contrary evidence. The CATIE results contradicted beliefs established by nearly a decade of evidence suggesting that atypical antipsychotics were superior to older, conventional antipsychotics,[19] beliefs reinforced through aggressive marketing campaigns that touted these drugs as "second generation." This favorable prior research and aggressive marketing, combined with high expectations for drugs that represented the first potential breakthrough in decades in the treatment for schizophrenia, led many psychiatrists to immediately reject the CATIE results.

Second is pro-intervention bias,[20] the tendency to choose action over inaction even if the marginal benefit of action is very small. In cardiology, this phenomenon is often referred to as the "oculostenotic reflex," the tendency to use percutaneous coronary intervention to open any coronary blockage, regardless of its clinical relevance.[21] Explaining this phenomenon, one cardiologist told us that by the time of angiography, a patient has become reduced to "an anatomic appearance on a picture"—implying that the presence of a blockage dwarfs consideration of the relative benefits and harms of the procedure or patient preferences. Pro-intervention bias may also be reinforced through referral pathways, if referral to specialists is interpreted as a request for aggressive intervention.

Third is pro-technology bias,[22,23] the uncritical tendency to believe that newer forms of technology are superior. This bias can undermine comparative effectiveness studies of long duration because of the appeal of new technologies that emerge before a study is completed. The COURAGE trial, for example, studied bare metal stents, but drug-eluting stents had become widespread by the time results were released, raising questions about the relevance of the trial's results.

FAILURE TO ADDRESS THE NEEDS OF END USERS Clinicians, patients, and policy makers may have different expectations about the goals and uses of comparative effectiveness research. Some clinicians might prefer studies that can support personalized medicine, while other clinicians (and payers) might prefer research whose results can be generalized to larger populations.

Although a single comparative effectiveness study cannot realistically meet the information needs of all stakeholders, an explicit consensus on the goals of each study could minimize variability in the interpretation of results. For example, the CATIE study primarily addressed differences in effectiveness, but many clinicians were seeking information about the relative

safety of the two antipsychotic classes. The study failed to address their needs because it was not sufficiently powered from a statistical perspective—meaning that not enough patients were enrolled—to detect differences in serious adverse effects.

Comparative effectiveness studies have tended to answer questions related to decision making at relatively "downstream" decision points in the referral process. The COURAGE trial, for instance, focused on the time just after the patient had undergone angiography—when he or she was conveniently fully prepped for percutaneous coronary intervention. At these later stages, financial incentives (and convenience) may exert stronger effects on decision making than at earlier stages—when primary care providers and patients may be less inclined to choose an intervention. Thus, results that influence "upstream" decisions may have greater leverage in changing clinical practice.

LIMITED USE OF DECISION SUPPORT Clinical decision support tools and aids for shared decision making may promote treatment approaches that are better aligned with evidence from comparative effectiveness research,[24,25] but these tools and aids are not widely used. The business case for clinicians and patients to use these technologies is lacking, and incentive programs that support their adoption—such as the Electronic Health Records Incentive Programs[26] of the Centers for Medicare and Medicaid Services—are still in the early stages. Decision support in the form of appropriateness feedback for the use of percutaneous coronary intervention and other procedures is advancing in a limited number of health care systems with advanced health information technology capabilities, such as the Department of Veterans Affairs health care system, but in few other settings.

Integrating clinical decision support tools seamlessly into clinical practice remains difficult because poorly designed tools can force clinicians to spend too much time engaging with computers rather than patients. For example, false alerts may trigger too frequently, disrupting attention to real alerts. Clinicians may also lack guidance on how to select an appropriate patient decision aid and may lack training in how and when to use the aids. Many providers that have experimented with decision aids have abandoned them because the providers had difficulty tailoring the aids to the work flow of their clinic.

Policy Options To Optimize Evidence Translation

In the face of such deeply rooted and intertwined barriers, policy makers hoping to speed the translation of comparative effectiveness results face a daunting challenge. Two encouraging trends, however, have appeared.

The Affordable Care Act realigns financial incentives to reward efficiency.

First, the Affordable Care Act realigns financial incentives to reward efficiency, which potentially will reverse the perverse financial incentives favoring costly intervention. Second, several professional societies and nonprofit research organizations have begun to actively manage the translation process. The following policies that support these nascent efforts may help improve the success of evidence translation in the coming decade.

DEVELOP CONSENSUS OBJECTIVES AND STANDARDS FIRST Recognizing that decisions made during the study design phase are key determinants of successful translation, the Patient-Centered Outcomes Research Institute has made multistakeholder engagement in study design the cornerstone of its new model for comparative effectiveness research. Building on these efforts, we believe that each large comparative effectiveness study would benefit from a formal consensus development process to determine the study's objectives and design, as well as the evidentiary standards that will guide the interpretation of its results. Maintaining a public record of these consensus decisions could reduce the potential for post hoc reframing of research objectives and guard against baseless criticisms of study designs and self-serving interpretations of results.

Such a process has already been pioneered by the Center for Medical Technology Policy, which convenes multistakeholder panels to generate consensus standards for comparative effectiveness studies. The Patient-Centered Outcomes Research Institute might consider mandating the use of such a consensus development process on a study-by-study basis for each of its grantees.

PROMOTE A BROADER PROFESSIONALISM In 2011 the Institute of Medicine issued a landmark report condemning the guideline development process in the United States. Citing abundant evidence of bias displayed by guideline committees that reflected a limited set of perspectives,

Federal agencies are already positioned to create the infrastructure for effective generation of comparative effectiveness evidence.

the institute proposed eight standards, one of which was that guideline development groups be "multidisciplinary and balanced."[27] Such broadly constituted committees might produce more balanced, consensus interpretations of comparative effectiveness results and resolve subspecialty differences of opinion with greater transparency. Other standards included the use of robust conflict-of-interest policies, more complete documentation of the strength of evidence supporting each guideline, and the use of more explicit statements regarding the recommended clinical actions to take in specific circumstances.

However, few specialty societies have openly embraced the Institute of Medicine's proposed standards because of the cost of implementing them and the potential for relevant experts to be excluded, among other concerns. In addition, few professional societies have voluntarily embraced an inclusive model in their development of quality measures. One notable exception is the American Medical Association's Physician Consortium for Performance Improvement,[28] which convenes a range of relevant specialists in its working groups to ensure that varied viewpoints are brought to bear and maintains a strict conflict-of-interest policy.

PROMOTE EMERGING PAYMENT AND COVERAGE POLICIES Emerging payment models, such as accountable care organizations, that may use global or bundled payment approaches to encourage efficient care delivery may drive the use of evidence-based clinical approaches. These include adherence to guidelines, use of performance feedback, and implementation of clinical decision support tools. Despite widespread enthusiasm for accountable care organizations, it is not yet clear how evidence translation will unfold within these entities, particularly for ones made up of loosely affiliated providers who have historically interpreted evidence and adopted new practices independently.

As new comparative effectiveness evidence emerges in the coming years, coverage policies such as value-based insurance design or reference pricing—both of which may rely on comparative effectiveness data—are likely to expand. In value-based insurance designs, employers or insurers typically waive or reduce cost sharing for more effective treatments or less costly and clinically comparable treatments within a class. Employers and insurers could lay the groundwork for the expansion of this insurance model by involving both physicians and members of the public in their design.[29] Insurers or employers that have already experimented with value-based insurance design might consider expanding these programs to address high-cost, low-value treatments—a more controversial policy, but one that has been recommended by many experts.[30,31]

These three policies—strengthening study design and facilitating interpretation of the evidence through consensus processes, promoting broader professionalism, and changing payment and insurance models—are intended to achieve greater balance among stakeholders in the design of comparative effectiveness studies, produce more actionable results, and reorient the incentives during the formalization and adoption steps of the translation process, which currently favor the adoption of clinical practices that are inconsistent with the evidence.

The Patient-Centered Outcomes Research Institute, the Agency for Healthcare Research and Quality, and other federal agencies are already positioned to create the infrastructure for effective generation of comparative effectiveness evidence. Similarly, a wide array of federal and state agencies, regulators, professional societies, and health care delivery organizations are poised to support the translation process.

Currently, however, these stakeholders are not sufficiently coordinated to overcome the key translation barriers we have outlined here. The limited set of policies we identified could, over time, modify the financial and other incentives that shape patient and clinician decisions, ensuring that those decisions are based on high-quality evidence and serve patients' best interests.

Conclusion

American taxpayers and other stakeholders have made a substantial investment in the production of new comparative effectiveness research, in the hope that the results will better inform clinician and patient decisions and thus improve the

quality and reduce the cost of care. To achieve these goals, policy makers must also recognize and address the root causes that have impeded translation of evidence into practice.

The root causes we describe played a role in many of the seminal comparative effectiveness studies of the past decade. Translating findings into best practices and promoting these recommendations remains largely ad hoc and post hoc, with few exceptions. The policies we recommend may provide the greatest leverage to speed the translation of evidence into practice. ■

Preliminary results from this study were presented at the AcademyHealth Annual Research Meeting in Orlando, Florida, June 24–26, 2012. The study was funded by the Office of the Assistant Secretary for Planning and Evaluation at the Department of Health and Human Services. The authors thank Caroline Taplin and Kate Goodrich of that office for their guidance and support.

NOTES

1 Hersh AL, Stefanick ML, Stafford RS. National use of postmenopausal hormone therapy: annual trends and response to recent evidence. JAMA. 2004;291(1):47–53.
2 Stadtmauer EA, O'Neill A, Goldstein LJ, Crilley PA, Mangan KF, Ingle JN, et al. Conventional-dose chemotherapy compared with high-dose chemotherapy plus autologous hematopoietic stem-cell transplantation for metastatic breast cancer. Philadelphia Bone Marrow Transplant Group. N Engl J Med. 2000;342(15):1069–76.
3 Etzioni R, Penson DF, Legler JM, di Tommaso D, Boer R, Gann PH, et al. Overdiagnosis due to prostate-specific antigen screening: lessons from U.S. prostate cancer incidence trends. J Natl Cancer Inst. 2002; 94(13):981–90.
4 Pollack A. The minimal impact of a big hypertension study. New York Times. 2008 Nov 28.
5 To access the Appendix, click on the Appendix link in the box to the right of the article online.
6 Schneider EC, Timbie JW, Fox DS, Van Busum K, Caloyeras JC. Dissemination and adoption of comparative effectiveness research findings when findings challenge current practices. Santa Monica (CA): RAND Corporation; 2011.
7 Robinson JC. Comparative effectiveness research: from clinical information to economic incentives. Health Aff (Millwood). 2010;29(10): 1788–95.
8 Boden WE, O'Rourke RA, Teo KK, Hartigan PM, Maron DJ, Kostuk WJ, et al. Optimal medical therapy with or without PCI for stable coronary disease. N Engl J Med. 2007; 356(15):1503–16.
9 Weinstein JN, Tosteson TD, Lurie JD, Tosteson ANA, Blood E, Hanscom B, et al. Surgical versus nonsurgical therapy for lumbar spinal stenosis. N Engl J Med. 2008; 358(8):794–810.
10 Avorn J. Transforming trial results into practice change: the final translational hurdle: Comment on "Impact of the ALLHAT/JNC7 Dissemination Project on thiazide-type diuretic use." Arch Intern Med. 2010;170(10):858–60.
11 Lieberman JA, Stroup TS, McEvoy JP, Swartz MS, Rosenheck RA, Perkins DO, et al. Effectiveness of antipsychotic drugs in patients with chronic schizophrenia. N Engl J Med. 2005;353(12):1209–23.
12 Major outcomes in high-risk hypertensive patients randomized to angiotensin-converting enzyme inhibitor or calcium channel blocker vs diuretic: the Antihypertensive and Lipid-Lowering Treatment to Prevent Heart Attack Trial (ALLHAT). JAMA. 2002;288(23):2981–97.
13 Tricoci P, Allen JM, Kramer JM, Califf RM, Smith SC Jr. Scientific evidence underlying the ACC/AHA clinical practice guidelines. JAMA. 2009;301(8):831–41.
14 Bristow MR, Saxon LA, Boehmer J, Krueger S, Kass DA, De Marco T, et al. Cardiac-resynchronization therapy with or without an implantable defibrillator in advanced chronic heart failure. N Engl J Med. 2004;350(21):2140–50.
15 Friedewald VE Jr., Boehmer JP, Kowal RC, Saxon L, Yancy CW, Roberts WC. The editor's roundtable: cardiac resynchronization therapy. Am J Cardiol. 2007;100(7):1145–52.
16 Murphy ML, Hultgren HN, Detre K, Thomsen J, Takaro T. Treatment of chronic stable angina. A preliminary report of survival data of the randomized Veterans Administration cooperative study. N Engl J Med. 1977;297(12):621–7.
17 Pines JM. Profiles in patient safety: confirmation bias in emergency medicine. Acad Emerg Med. 2006; 13(1):90–4.
18 Kaptchuk TJ. Effect of interpretive bias on research evidence. BMJ. 2003;326(7404):1453–5.
19 Carpenter WT, Buchanan RW. Lessons to take home from CATIE. Psychiatr Serv. 2008;59(5):523–5.
20 Ayanian JZ, Berwick DM. Do physicians have a bias toward action? A classic study revisited. Med Decis Making. 1991;11(3):154–8.
21 Topol EJ, Nissen SE. Our preoccupation with coronary luminology. The dissociation between clinical and angiographic findings in ischemic heart disease. Circulation. 1995;92(8):2333–42.
22 Owen R. Reader bias. JAMA. 1982;247(18):2533–4.
23 Emanuel EJ. What cannot be said on television about health care. JAMA. 2007;297(19):2131–3.
24 Bates DW, Pappius E, Kuperman GJ, Sittig D, Burstin H, Fairchild D, et al. Using information systems to measure and improve quality. Int J Med Inform. 1999;53(2–3):115–24.
25 Hunt DL, Haynes RB, Hanna SE, Smith K. Effects of computer-based clinical decision support systems on physician performance and patient outcomes: a systematic review. JAMA. 1998;280(15):1339–46.
26 CMS.gov. EHR Incentive Programs [Internet]. Baltimore (MD): Centers for Medicare and Medicaid Services; [last modified 2012 Aug 27; cited 2012 Sep 6]. Available from: http://www.cms.gov/Regulations-and-Guidance/Legislation/EHRIncentivePrograms/index.html?redirect=/ehrincentiveprograms/
27 Institute of Medicine. Clinical practice guidelines we can trust. Washington (DC): National Academies Press; 2011. p. 93.
28 American Medical Association. Physician Consortium for Performance Improvement [Internet]. Chicago (IL): AMA; [cited 2012 Sep 6]. Available from: http://www.ama-assn.org/ama/pub/physician-resources/physician-consortium-performance-improvement.page
29 Ginsburg M. Value-based insurance design: consumers' views on paying more for high-cost, low-value care. Health Aff (Millwood). 2010;29(11): 2022–6.
30 Robinson JC. Applying value-based insurance design to high-cost health services. Health Aff (Millwood). 2010;29(11):2009–16.
31 Fendrick AM, Smith DG, Chernew ME. Applying value-based insurance design to low-value health services. Health Aff (Millwood). 2010;29(11): 2017–21.

ABOUT THE AUTHORS: JUSTIN W. TIMBIE, D. STEVEN FOX, KRISTIN VAN BUSUM & ERIC C. SCHNEIDER

Justin W. Timbie is a health policy researcher at the RAND Corporation.

In this month's *Health Affairs*, Justin Timbie and coauthors, all affiliated with RAND, report on their examination of reasons why relatively recent and high-profile comparative effectiveness studies have been slow to change clinical practice. They identified five causes, ranging from financial incentives that dissuade clinicians from changing practice patterns to cognitive biases against new information. The authors recommend various steps to support the translation of new evidence, including the development of consensus objectives that can be promulgated and widely shared even before comparative effectiveness studies get under way.

Timbie is a health policy researcher at the RAND Corporation. His research focuses on the development of methods, including enhanced case-mix adjustments and aggregated performance indicators, to profile providers' use of resources and the quality of care they deliver.

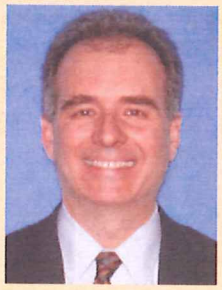
Timbie also participates in the evaluation of delivery system interventions, including medical homes and clinical decision support systems. He received a doctorate in health policy from Harvard University and completed a postdoctoral fellowship in the Veterans Affairs Health System.

D. Steven Fox is a policy analyst at RAND.

Steven Fox is a policy analyst at the RAND Corporation and an internist and pediatrician. His research focuses on global health and development, health care systems improvement, the implementation of new health and technology research, and domestic biosecurity. Before joining RAND, he practiced clinical medicine in the United States and overseas and worked as a financial analyst. He holds graduate degrees in policy analysis (Pardee RAND Graduate School), geoscience (University of Washington), and public health and tropical medicine (London School of Hygiene and Tropical Medicine). He is currently a policy analysis PhD candidate at the Pardee RAND Graduate School.

Kristin Van Busum is a health project associate at RAND.

Kristin Van Busum is a health project associate at the RAND Corporation. She has coordinated studies evaluating the effectiveness of workplace wellness programs and efforts to improve quality reporting and the patient experience. She was previously a research associate at the New York University School of Medicine, where she studied how vulnerable populations make decisions about health care. Van Busum has a master's degree in public administration, with a concentration in health policy, from New York University.



Eric C. Schneider is a senior scientist and director at RAND.

Eric Schneider is a senior scientist and director at the RAND Corporation, where he also holds the distinguished chair in health care quality. He is an associate professor of medicine at Harvard Medical School and in the Division of General Internal Medicine and Primary Care at the Brigham and Women's Hospital. He is also an associate professor at the Harvard School of Public Health.

Schneider is editor-in-chief of the *International Journal for Quality in Health Care*. His research addresses the quality of health care, particularly the development, specification, and impact of performance measures and reporting in health care. He is a fellow of the American College of Physicians and past chair of its Performance Measurement Committee. He earned a master's degree in health and medical sciences from the University of California, Berkeley, and a medical degree from the University of California, San Francisco.

By Harold Sox

DOI: 10.1377/hlthaff.2012.0171
HEALTH AFFAIRS 31,
NO. 10 (2012): 2176–2182

ANALYSIS & COMMENTARY

The Patient-Centered Outcomes Research Institute Should Focus On High-Impact Problems That Can Be Solved Quickly

Harold Sox (hsox@comcast.net) is a professor emeritus of medicine at the Geisel School of Medicine at Dartmouth and associate director for faculty at the Dartmouth Institute for Health Policy and Clinical Practice, in Lebanon, New Hampshire.

ABSTRACT **The Affordable Care Act created the Patient-Centered Outcomes Research Institute (PCORI) to help patients, clinicians, and policy makers make well-informed decisions about health care. Because its funding expires in 2019, the institute has little time in which to produce timely, practice-changing results that will build public support for comparative effectiveness research. PCORI should plan its research agenda strategically, so that it addresses research questions that comparative effectiveness research could answer quickly and decisively. To date, the institute has not chosen this path. In January 2012 PCORI's first research agenda described broad research priorities rather than specific clinical questions. The institute must drive the burgeoning discipline of comparative effectiveness research forward, starting with a research project agenda that conveys a sense of urgency and strategic direction.**

Comparative effectiveness research has received considerable attention from researchers, policy makers, and the public. This new branch of research is based on the disciplines of clinical epidemiology and medical decision making and the important concepts embodied in patient-centeredness. Researchers are interested in comparative effectiveness research because they hope it will answer important clinical questions about what works best for which patients. Policy makers hope that comparative effectiveness results will lead to less waste and better health outcomes. The research enjoys public support,[1] in part because of its focus on improving the evidence base that clinicians and patients need to make well-informed choices.

Comparative effectiveness research is not new. The medical literature of the past sixty years contains sporadic examples. Some well-known examples include the Gruppo Italiano per lo Studio della Streptochinasi nell'Infarto Miocardico (GISSI) randomized trial that compared thrombolysis to usual care in acute myocardial infarction;[2] the Antihypertensive and Lipid-Lowering Treatment to Prevent Heart Attack Trial (ALLHAT) randomized trial that compared three first-line antihypertensive drugs;[3] and the National Emphysema Treatment Trial (NETT) of lung reduction surgery versus aggressive medical therapy for emphysema.[4]

The leading US program of comparative effectiveness research has been, since its authorization by the Medicare Prescription Drug, Improvement, and Modernization Act of 2003, the Effective Health Care Program at the Agency for Healthcare Research and Quality. Section 1013 of the act authorizes research, demonstrations, and evaluations to improve the quality, effectiveness, and efficiency of the federally administered Medicare program, Medicaid, and the Children's Health Insurance Program. With that authority, the Effective Health Care Program produces systematic reviews, original re-

search, and evidence summaries pertaining to key questions of comparative effectiveness.

Then, in 2008, an Institute of Medicine committee recommended a national program of comparative effectiveness research.[5] Its report addressed the problems created by a proliferation of multiple, conflicting practice guidelines of widely varying quality. The committee envisioned a single national program that would set priorities for—and fund—systematic reviews and establish standards for the conduct of systematic reviews and practice guidelines. The Obama administration included $1.1 billion for comparative effectiveness research in the American Recovery and Reinvestment Act of 2009. The act also commissioned an Institute of Medicine study to set national priorities for the research. After a broadly based nominations process, the institute's committee, which was cochaired by this author, identified one hundred disease- and intervention-specific research topics.[6]

The most recent milestone in the path toward a national comparative effectiveness research program was the Affordable Care Act of 2010, which created the Patient-Centered Outcomes Research Institute (PCORI). In Section 6301 of the act, Congress charged PCORI with establishing national priorities and a research project agenda to develop evidence that will help people make informed health decisions.

This article focuses on barriers to fulfilling the promise of comparative effectiveness research. Some barriers are methodological, some are translational, and some are political. One political barrier, however, stands out: Section 6301 of the Affordable Care Act set a September 2019 sunset date for PCORI, giving the institute seven years to convince Congress to reauthorize it.

To achieve this goal, PCORI needs some early successes. The Affordable Care Act gave the institute the power to influence its own destiny, in part by specifying processes by which it could shape a research agenda of studies with a good chance of producing findings that will find their way into everyday medical practice. However, PCORI's initial national priorities and research agenda, published in May 2012, list broad research topics, not specific high-impact research questions.[7] On this evidence, PCORI is not using its research agenda strategically to address research questions that comparative effectiveness research could answer quickly and decisively.

This article argues for stronger action sooner. It describes the tools the Affordable Care Act provided to shape the PCORI research agenda, reviews the research methods that might yield actionable results quickly, and discusses agenda-setting strategies for managing the predicament posed by the 2019 PCORI sunset date.

'Comparative Effectiveness Research' Defined

The 2009 Institute of Medicine committee on national priorities for comparative effectiveness research devised the following definition: "The generation and synthesis of evidence that compares the benefits and harms of alternative methods to prevent, diagnose, treat, and monitor a clinical condition or to improve the delivery of care. The purpose of comparative effectiveness research is to assist consumers, clinicians, purchasers, and policy makers to make informed decisions that will improve health care at both the individual and population levels."[6(p41)]

Section 6301 of the Affordable Care Act used a similar definition: "The terms 'comparative clinical effectiveness research' and 'research' mean research evaluating and comparing health outcomes and the clinical effectiveness, risks, and benefits of 2 or more medical treatments, services, and items described in subparagraph (B)." Subparagraph B refers to a list of strategies, services, and items used in the treatment, diagnosis, and prevention of illness.

PCORI has defined the research that it will support using the term *patient-centered outcomes research*, rather than *comparative effectiveness research*. According to PCORI, patient-centered outcomes research "helps people and their caregivers communicate and make informed health care decisions, allowing their voices to be heard in assessing the value of health care options."[8] The PCORI definition describes the characteristics of patient-centered outcomes research in terms similar to those of the Institute of Medicine and the Affordable Care Act. This article uses the term *comparative effectiveness research*.

The key implications of these definitions are that comparative effectiveness research should directly compare active treatments; that study patients, clinicians, and interventions should be typical of those in usual practice; and that comparative effectiveness research should focus on helping patients, clinicians, and policy makers make informed choices that reflect their values. These goals appear to have broad support, but they do impose constraints on the comparative effectiveness research initiative, to be discussed below.

A Statutory Context For PCORI

Several provisions in the Affordable Care Act provide the context for this article.

FUNDING To pay for the research commissioned by PCORI, the statute created the Patient-Centered Outcomes Research Trust Fund within the Treasury Department. The trust fund's annual revenues are $150 million from

the Treasury in addition to receipts from a tax on public- and private-sector health insurance providers. Starting in fiscal year 2014, this tax will be two dollars per insured life per year.

Section 6301 of the Affordable Care Act states the following: "No amounts shall be available for expenditure from [the Trust Fund] after September 30, 2019, and any amounts in such Trust Fund after such date shall be transferred to the general fund of the Treasury." The implication is that the institute will close down in 2019 unless Congress reauthorizes it.

EVALUATION Section 6301 of the Affordable Care Act instructs the comptroller general to evaluate "the adequacy and use of the funding for the Institute and the activities conducted under section 937 of the Public Health Service Act, including a determination as to whether, based on the utilization of research findings by public and private payers, funding sources for the Patient-Centered Outcomes Research Trust Fund under section 9511 of the Internal Revenue Code of 1986 are appropriate and whether such sources of funding should be continued or adjusted."

This clause indicates that the evaluation of PCORI shall be based in part on the extent to which public and private payers use the findings of comparative effectiveness research, which positions payers to influence the future of comparative effectiveness research.

PRIORITIES Section 6301 of the statute decrees that the institute "shall identify national priorities for research, taking into account factors of disease incidence, prevalence, and burden in the United States (with emphasis on chronic conditions), gaps in evidence in terms of clinical outcomes, practice variations and health disparities in terms of delivery and outcomes of care, the potential for new evidence to improve patient health, well-being, and the quality of care, the effect on national expenditures associated with a health care treatment, strategy, or health conditions, as well as patient needs, outcomes, and preferences."

This language implies that the priority-setting process should identify specific diseases, tests, treatments, and care strategies.

RESEARCH PROJECT AGENDA Section 6301 of the statute states that the institute "shall establish and update a research project agenda for research to address the priorities identified under subparagraph (A), taking into consideration the types of research that might address each priority and the relative value (determined based on the cost of conducting research compared to the potential usefulness of the information produced by research) associated with the different types of research, and such other factors as the Institute determines appropriate."

Electoral politics will influence PCORI's future.

This description of the research project agenda takes its specificity partly from the national priorities that PCORI will establish and partly from the study methods that the research agenda will specify. The language seems to say that the agenda may specify the types of research needed to address each priority research question.

PCORI's Strategy And Survival

Electoral politics will influence PCORI's future. Reauthorization will probably face considerable opposition, and the balance of political power after upcoming presidential and congressional elections could render PCORI's performance entirely moot. H R 3827, introduced January 25, 2012, is a case in point: It calls for the repeal of PCORI.

The fledgling institute's leadership has a difficult task. The institute is a start-up that has one year to scale up to an organization that dispenses a half-billion dollars in research funding each year and only seven years to win reauthorization. The processes stipulated in the statute and described in the preceding section would help it maximize the impact of the research that it commissions.

Using the priority-setting criteria described in the Affordable Care Act, PCORI could choose research questions that address evidence gaps in clinical conditions that have a high burden of morbidity and mortality. Because the Affordable Care Act asks PCORI to specify the methods for studying high-priority research questions, the institute could specify methods that would produce results quickly.[9]

Key Characteristics Of Comparative Effectiveness Research

The purpose of comparative effectiveness research—to inform decision making in everyday practice—imposes several requirements that are difficult to meet, yet necessary.

APPLICABILITY TO EVERYDAY COMMUNITY PRACTICE The best way to obtain evidence that applies to a typical patient's chances of a successful outcome—so-called external validity—is to

Comparative effectiveness research must inform the often difficult choices between well-established active treatments.

enroll patients that are typical of community practice. Italy's GISSI-1 pragmatic randomized trial of heart attack treatments is a good example of high external validity.[2] The study had few exclusion criteria, and 88 percent of the coronary care units in Italy participated. Accordingly, the patients, hospitals, and physicians in the trial were typical of community practice. Enrolling patients and maintaining adherence to study protocols requires excellent cooperation from the clinical site in which study patients receive care—for example, a hospital or office practice. The GISSI investigators successfully coped with the logistical challenges of doing a large randomized study in a community setting.

INFORMING DECISIONS Comparative effectiveness research must inform the often difficult choices between well-established active treatments. Because many well-established active treatments have similar response rates, head-to-head trials may require thousands of patients to distinguish between them. The ALLHAT study, for example, compared the three most commonly used medications for patients beginning drug treatment for hypertension.[3] To be sure of detecting clinically important differences among the drugs, ALLHAT enrolled 33,357 patients and cost more than $100 million.

DETECTING DIFFERENCES IN SUBGROUPS' TREATMENT RESPONSIVENESS Large studies are also important for detecting differences in how subgroups of patients respond to treatments—a key element of individualized decision making. Comparative effectiveness research seeks to detect so-called treatment response heterogeneity and to identify the distinguishing characteristics of subgroups that respond differently. The Department of Veterans Affairs cooperative trial of surgery or medical treatment for chronic stable angina[10] is a case in point. At the beginning of the trial, researchers used clinical characteristics to identify each patient's five-year mortality risk. High-risk patients did better with surgery, low-risk patients did better with medical treatment, and patients at intermediate risk did equally well whether they received surgery or medical treatment.

OUTCOMES THAT MATTER TO PATIENTS The emerging model of patient-centered decision making is shared, informed decision making, in which patients participate actively in decisions about their health care. We know very little about what happens when patients and their physicians discuss the outcomes that matter most or how they make decisions together.

Strategic Agenda Setting

PCORI should use its research project agenda strategically to maximize the impact of the research that it commissions.

STRATEGIC PRIORITIES Strategic agenda setting begins with identifying the most important research questions. In its consensus-building process, the Institute of Medicine committee on national priorities for comparative effectiveness research used priority-setting criteria, preliminary rounds of voting, and discussion of each short-list topic before a final vote. Appendix Exhibit 1 shows a sample of high-priority research topics from the Institute of Medicine's list.[11] Each topic specifies a target condition, interventions, and a patient population. PCORI should identify specific high-priority research questions, perhaps using the Institute of Medicine list as a starting point. Then, for high-priority research questions, PCORI must recommend research methods that generate results relatively quickly, choosing from observational studies, pragmatic trials, systematic reviews, and cluster trials.

RANDOMIZED TRIALS AND OBSERVATIONAL RESEARCH The most difficult choice for PCORI is optimizing the balance of randomized trials and observational research. Observational research is inexpensive and fast, and it uses existing large data sets from representative clinical settings. Most researchers prefer randomized trials, which are usually expensive, slow, and selective but avoid the great drawback of observational research: Physicians tend to let prognostic factors influence the choice of treatments. Consider a hypothetical example: Sicker patients might receive radiation therapy for a cancer, while healthier patients would receive surgery, which might lead to a conclusion that radiation treatment is less effective than surgery when a subsequent randomized trial showed that the two treatments were equally effective. Therefore, factors other than a hypothesized effect of treat-

ment (confounders) could be responsible for outcome differences. Randomization avoids this hidden variable problem by distributing potential confounders equally among the compared treatment groups. Research to reduce confounding in observational research is ongoing, but progress is slow.

PRAGMATIC TRIALS Pragmatic randomized trials offer some of the virtues of randomized trials and observational studies while avoiding some of the shortcomings of these research techniques. Unlike observational studies, patients in pragmatic trials are randomly assigned to a treatment. Like observational studies, pragmatic trials take place in representative practice settings; involve typical patients, practitioners, and interventions; enroll many patients; and are faster and less expensive than typical randomized clinical trials. Italy's GISSI-1 pragmatic randomized trial of heart attack treatments, described in the preceding section, is a good example.[2] Pragmatic trials should have a substantial place in PCORI's near-term strategy.

SYSTEMATIC REVIEWS The aim of a systematic literature review is a clear, comprehensive, unbiased assessment of a body of evidence. The key elements are a prespecified research plan; an exhaustive search for evidence; a qualitative assessment of the body of evidence; and, often, a measure that summarizes the results of individual studies. A systematic review takes six to twelve months to complete.

EVALUATING THE IMPACT OF COMPARATIVE EFFECTIVENESS RESEARCH PCORI should fund studies to measure the impact of implementing comparative effectiveness research results. These studies would provide important evidence to help the public assess its investment in PCORI. Medical practices whose existing practice cultures are receptive to change would be ideal sites for such studies, which could compare different strategies for changing practice using a cluster randomized trial design. In this research design, practice sites are randomized to implement one or the other of the compared patient care strategies.

These implementation studies need not wait for the completion of PCORI-sponsored comparisons of clinical interventions. Instead, PCORI should fund studies to measure the impact of previously published comparative effectiveness research results in typical practice sites. For example, it might sponsor a cluster randomized trial of two approaches to maximizing adherence to annual screening with low-dose computed tomographic screening for lung cancer. Such screening was shown recently to reduce lung cancer death rates relative to screening chest radiographs.[12] In this way, PCORI can fund comparative effectiveness studies and implementation studies simultaneously, thereby making efficient use of the time remaining before 2019.

PCORI National Priorities For Research And Research Agenda

On January 23, 2012, PCORI presented for public comment its Draft National Priorities for Research and Research Agenda, which it adopted on May 21, 2012.[7] The first proposed national priority for research is comparative assessment of prevention, diagnosis, and treatment options. The stated research goal is to determine which interventions, in each of these categories, work best for distinct populations with specific health problems. The PCORI research agenda for this priority lists several cross-cutting topics (for example, a comparison of prognostic and risk-stratification methods with usual clinical approaches to treatment selection). However, it does not list any specific interventions, study populations, or health problems, as did the research questions chosen by the Institute of Medicine committee[6] or as the language of the Affordable Care Act appears to stipulate.

PCORI specified four additional research priorities that, collectively, would help support better decision making but would not necessarily compare interventions: improving health care systems, communication and dissemination research, health care disparities, and accelerating patient-centered and methodological research. For each of these priorities, the PCORI research agenda lists specific topics.

Improving health care systems would include interventions to improve access, decision making, care coordination, and other cross-cutting interventions. Communication and dissemination research would focus on improving decision makers' knowledge, patients' participation in decision making, use of decision aids, and dissemination of research results. Health care disparities research would focus on ways to reduce disparities in health outcomes and on measuring differences in the way populations experience the benefits and risks of interventions. Accelerating patient-centered and methodological research would focus on developing statistical methods; measuring patients' preferences; and developing an informed community of patients, researchers, and caregivers who participate in comparative effectiveness research.

These priorities focus on how to improve the broad infrastructure of health care so that it supports patient-centered decisions about the care of individual patients. This goal is important, but it does not directly address the most urgent priority for comparative effectiveness research:

to inform decisions about specific clinical problems.

PCORI classified the public comments about its research priorities and research agenda into ten recurring themes.[13] Most were supportive of PCORI's direction, but one theme recommended that PCORI focus on specific conditions or disease areas instead of proposing what PCORI called its "condition-neutral" research agenda.[13] In explaining its position, PCORI states that it did not want, at this early stage of its work, to "focus on a narrower set of questions or health care conditions" or to exclude any diseases from consideration.[7] Rather, PCORI expects that those applying for funding will choose the research question that they plan to address. The rationale is that researchers partnered with stakeholders can identify "compelling questions that form a truly patient-centered research agenda."[7]

Presumably, the PCORI study sections charged with reviewing grant proposals will judge an application in part by the research question it proposes. These policies and processes will put both grant applicants and study section members in the position of setting national priorities for comparative effectiveness research. In discussing its research agenda, PCORI stated that it "does not specify preferred study designs or analytic approaches." Rather, research questions could "be investigated by a variety of scientifically valid methods and approaches."[7]

The PCORI document provides some hope for critics who want the institute to take a proactive stance. It states that "over time it is anticipated that PCORI will develop a research portfolio that includes both broad calls for proposals as well as contracts targeted specifically at high-priority questions" identified by processes involving consultation with stakeholders and the public.[7] Moreover, the institute does not "preclude subsequent funding announcements that would specify a preferred or required methodological design for a specific research question."[7]

Can This Research Meet The Expectations Of Its Stakeholders?

Patients, researchers, payers, and policy makers expect much from comparative effectiveness research because it is trying to address unmet needs of patients who want to understand the choices that they face in an increasingly complicated health care system. These stakeholders hope that PCORI will launch a sustained program of comparative effectiveness research that informs patient-centered decision making. Because PCORI is their best hope for driving comparative effectiveness research forward, stakeholders fear for its well-being in today's partisan political climate.

This article makes two principal points. First, PCORI is in a difficult spot. It has until 2019 to convince a majority of members of Congress that comparative effectiveness research can improve the practice of medicine. To compound the problem, the pace of clinical research—from the request for applications to the completed and published work—is slow, and the pace of change in medical practice is even slower.

Second, PCORI's actions have not conveyed a sense of urgency or strategic direction. The institute's first substantive utterance—its National Priorities for Research and Research Agenda—does not list high-priority research questions or specify research methods to address them. It leaves these important tasks entirely to the research community, to patients, and to stakeholders. This approach could be effective in the long term, but it is not sufficiently responsive to the urgent circumstances dictated by the 2019 sunset date. PCORI must start now to implement a strategy to make the largest possible impact before its day of reckoning. ■

This article was adapted from the author's 2011 John M. Eisenberg Legacy Lecture, which was sponsored by the California HealthCare Foundation and delivered November 28, 2011, in San Francisco, California. The author cochaired the 2009 Institute of Medicine committee that defined *comparative effectiveness research* and set national priorities for clinical research questions for comparative effectiveness research to address.

NOTES

1 Gerber AS, Patashnik EM, Doherty D, Dowling CM. A national survey reveals public skepticism about research-based treatment guidelines. Health Aff (Millwood). 2010;29(10):1882–4.
2 Gruppo Italiano per lo Studio della Streptochinasi nell'Infarto Miocardico (GISSI). Effectiveness of intravenous thrombolytic treatment in acute myocardial infarction. Lancet. 1986;1(8478):397–402.
3 ALLHAT Collaborative Research Group. Major outcomes in high-risk hypertensive patients randomized to angiotensin-converting enzyme inhibitor or calcium channel blocker vs diuretic: the Antihypertensive and Lipid-Lowering Treatment to Prevent Heart Attack Trial (ALLHAT). JAMA. 2002;288(23):2981–97.
4 Fishman A, Martinez F, Naunheim K, Piantadosi S, Wise R, Ries A, et al. A randomized trial comparing lung-volume-reduction surgery with medical therapy for severe emphysema. N Engl J Med. 2003;348(21):2059–73.
5 Eden J, Wheatley B, McNeil B, Sox H, editors. Knowing what works in health care: a roadmap for the nation. Washington (DC): National Academies Press; 2008.

6 Institute of Medicine. Initial national priorities for comparative effectiveness research. Washington (DC): National Academies Press; 2009.
7 Patient-Centered Outcomes Research Institute. National priorities for research and research agenda [Internet]. Washington (DC): PCORI; 2012 May 21 [cited 2012 Sep 17]. Available from: http://www.pcori.org/assets/PCORI-National-Priorities-and-Research-Agenda-2012-05-21-FINAL.pdf
8 Patient-Centered Outcomes Research Institute. Patient-centered outcomes research [Internet]. Washington (DC): PCORI; 2012 Mar 5 [cited 2012 Sep 17]. Available from: http://www.pcori.org/what-we-do/pcor/
9 VanLare JM, Conway PH, Sox HC. Five next steps for a new national program for comparative-effectiveness research. N Engl J Med. 2010;362(11):970–3.
10 Detre K, Peduzzi P, Murphy M, Hultgren H, Thomsen J, Oberman A, et al. Effect of bypass surgery on survival in patients in low- and high-risk subgroups delineated by the use of simple clinical variables. Circulation. 1981;63(6):1329–38.
11 To access the Appendix, click on the Appendix link in the box to the right of the article online.
12 Aberle DR, Adams AM, Berg CD, Black WC, Clapp JD, Fagerstrom RM, et al. Reduced lung-cancer mortality with low-dose computed tomographic screening. N Engl J Med. 2011;365(5):395–409.
13 Patient-Centered Outcomes Research Institute. National priorities for research and research agenda: analysis and key findings from public comments [Internet]. Washington (DC): PCORI; 2012 May 16 [cited 2012 Sep 17]. Available from: http://www.pcori.org/assets/Draft-National-Priorities-and-Research-Agenda-Public-Comment-Analysis-Summary.pdf

ABOUT THE AUTHOR: HAROLD SOX

Harold Sox is a professor emeritus of medicine at the Geisel School of Medicine at Dartmouth.

In this month's *Health Affairs*, Harold Sox argues that the Patient-Centered Outcomes Research Institute created by the Affordable Care Act faces an urgent deadline: unless the institute is reauthorized, its funding will expire in 2019. This timeline demands that the institute move forward quickly on adopting a strategic research agenda that will lead to answering clinically important comparative effectiveness questions. So far, that hasn't happened, writes Sox, who offers a series of recommendations for how the institute should move forward to generate important results on a swift timetable.

Sox is a professor emeritus of medicine at the Geisel School of Medicine at Dartmouth and associate director for faculty at the Dartmouth Institute for Health Policy and Clinical Practice. Among his many accomplishments, Sox cochaired the Institute of Medicine (IOM) committee that in 2009 published a report listing the one hundred comparative effectiveness research questions that should receive the nation's highest priority for funding and attention.

Beginning in 1988, Sox served on the Dartmouth faculty for thirteen years as professor and chair of the Department of Medicine. Before that, he spent fifteen years on the faculty of the Stanford University School of Medicine, where he was chief of the Division of General Internal Medicine and director of ambulatory care at the Palo Alto Veterans Affairs Medical Center.

Sox practiced internal medicine until 2001, when he became editor of *Annals of Internal Medicine*, a post he held until 2009. He also served as president of the American College of Physicians and as chair of the US Preventive Services Task Force, the Centers for Medicare and Medicaid Services Medicare Coverage Advisory Committee, the IOM Committee to Study HIV Transmission through Blood Products, and the IOM Committee on Health Effects Associated with Exposures Experienced in the Gulf War. He was elected to the Institute of Medicine in 1993.

Sox is currently a member of the editorial board of the *Journal of the American Medical Association* and is a contributing writer for that journal. His books include *Medical Decision Making*, *Common Diagnostic Tests: Selection and Interpretation*, and *HIV and the Blood Supply: An Analysis of Crisis Decision Making*. Sox received his medical degree from Harvard University.

By Susan Dentzer and the Editorial Team of *Health Affairs*

Communicating About Comparative Effectiveness Research: A *Health Affairs* Symposium On The Issues

DOI: 10.1377/hlthaff.2012.0607
HEALTH AFFAIRS 31,
NO. 10 (2012): 2183–2187

Susan Dentzer (sdentzer@projecthope.org) is the editor-in-chief of *Health Affairs*, in Bethesda, Maryland.

ABSTRACT The United States is pursuing an agenda of increased comparative effectiveness research in an effort to identify high-value health care. One of many issues connected with the research is how findings will be communicated, particularly if they pertain to prescription drugs and if the findings could be useful for promotional purposes yet fall short of the standard of "substantial evidence" required by the Food and Drug Administration (FDA) under federal law. To surface and air the relevant issues, this article posits a hypothetical case study of comparative effectiveness research involving a fictional migraine drug. A symposium of analyses and responses follows from various stakeholders representing the FDA; the Patient-Centered Outcomes Research Institute, and pharmaceutical companies; an attorney for the industry; academic specialists in the area of pharmaceutical law and policy; a proponent of academic detailing, or university or non-commercial-based education of prescribers based on medical evidence; a payer; a bioethicist; and an official from an organization representing patients with chronic diseases and disabilities.

To identify high-value health care, the United States has stepped up efforts to conduct comparative effectiveness research: studies that compare the clinical outcomes of alternative approaches to preventing, diagnosing, or treating medical conditions. *Health Affairs* has chronicled this movement extensively,[1] focusing on, among other aspects, the Patient-Centered Outcomes Research Institute, the entity created by the Affordable Care Act to fund and further the research.[2] With the cluster of articles that follows, we now turn to another subset of issues: Once comparative effectiveness research has been carried out, who will interpret its results and communicate them to clinicians and patients, and how?

These questions are of critical importance to many stakeholders in this research—among them, patients, health care providers, and those who have developed and marketed new treatments. If a given comparative effectiveness study appears to yield evidence that treatment A is superior to treatment B, all of these stakeholders should have an interest in knowing and spreading those results. But a deeper examination identifies questions that should be resolved first. Was the research methodologically sound? Does it pertain to particular types of patients and not to others? What really should be done with this new evidence? Who should have the opportunity to convey the results, to whom, in what format, and with what caveats?

Manufacturers of prescription drugs are among those with vested interests in the answers to these questions, because federal laws and regulations limit what they can say about their products and in what circumstances. Manufacturers have to provide "substantial evidence" that their products are effective before they can obtain

approval from the Food and Drug Administration (FDA) to market drugs. The phrase "substantial evidence" is generally understood to mean at least one "adequate and well-controlled" study showing that a drug has a claimed effect, as Joseph Griffin and colleagues note.[3] By extension, any comparative claims of effectiveness or safety in promotional labeling and advertising also have to be supported by such "substantial evidence."

A classic well-designed, randomized clinical trial that compared one drug to another and was large enough to achieve statistically meaningful results could probably produce the "substantial evidence" likely to persuade the FDA to permit a manufacturer to cite the results in its promotional material. But such research is costly and unlikely to be done by manufacturers if they fear that the results may not cast their products in the best light.

So if other study methods are used—for example, observational research conducted by sifting through insurance claims and patients' health records to determine how they responded to medications after the fact—would the results qualify as "substantial evidence" that a manufacturer could cite for promotional purposes? Pharmaceutical companies can legally engage in "scientific exchanges" about such studies, but if they can't use them for promotional purposes, is that fair? More broadly, would the results of such studies be valid and applicable enough to share with patients? Would it be ethically appropriate to withhold them?

To surface the many nuances of these issues, *Health Affairs* attempted what for our journal is an innovative approach: creating a hypothetical case study and producing a written "symposium" crafted by various stakeholders. Our case study involves a fictional drug, Hemikrane, for the treatment of migraine, and a comparative effectiveness study that measured it against another fictional migraine drug, Cephalal.

We first sought agreement on the parameters of the hypothetical case study from participants in the symposium, and then written analyses of how its results could or should be used. Respondents were representatives of the FDA, the Patient-Centered Outcomes Research Institute, and pharmaceutical companies; an attorney for the industry; academic specialists in the area of pharmaceutical law and policy; a proponent of academic detailing, or university or non-commercial-based education of prescribers based on medical evidence; a payer; a bioethicist; and an official from an organization representing patients with chronic diseases and disabilities. Tevi Troy, who is profiled in this issue,[4] served as our symposium adviser.

The articles that follow illustrate the complexities of the issues around comparative effectiveness research that will need to be worked through before there is consensus on how the research should be conducted or used. As readers will find, the methodology of the hypothetical case study comes under withering attack in some of the symposium entries. Patients and the Patient-Centered Outcomes Research Institute conclude that study wasn't "patient-centered" at all and discuss the reasons why. The FDA concludes that with the Hemikrane and Cephalal studies, the "substantial evidence" test hasn't been met, although it acknowledges that the science base of comparative effectiveness research is evolving.

Pharmaceutical companies rise to the study's defense and point to other aspects of federal law that could offer a pathway for them to communicate the results. Our academic specialists set forth useful parameters that could shape valid observational research going forward. And our bioethicist points to a larger issue: Patients generally don't understand much of the health information that confronts them now—so why should we anticipate that comparative effectiveness research will make the situation any easier?

We hope the symposium and the issues raised serve as a jumping-off point for further discussions about a range of topics—from the principles that should underpin publicly funded comparative effectiveness research to the need for any policy changes around its use. Now, to the case study.

The Case Study: 'Hemikrane' Versus 'Cephalal'

MIGRAINE FACTS For the estimated 28–36 million Americans who suffer in some form from migraine, experiencing an attack can be sheer torment.[5] Far more than just a headache, migraine is "an extremely debilitating collection of neurological symptoms that usually includes a severe recurring intense throbbing pain" on one or both sides of the head, according to the Migraine Research Foundation.[6] Attacks can last from several hours to up to three days. They can be accompanied by "visual disturbances, nausea, vomiting, dizziness, extreme sensitivity to sound, light, touch and smell, and tingling or numbness in the extremities or face," the foundation explains.[6]

Historical figures as diverse as Julius Caesar and Ulysses S. Grant have been plagued by migraine, as was the artist Vincent Van Gogh. The English novelist Virginia Woolf long suffered from prolonged attacks now thought to have been migraine. After one, she recorded in her diary that the attacks left her "gazing into the

The articles that follow illustrate the complexities of the issues around comparative effectiveness research.

depths of the misery of human life."[7]

Today, studies and available data show that approximately 6 percent of American men and 17 percent of American women experience periodic migraines.[5] They also find that approximately 1–2 percent of Americans overall suffer from chronic migraine,[8] in which an individual experiences migraine symptoms at least 15 days each month.[9] Those who suffer from chronic migraine are more likely than others to suffer from a variety of comorbidities, including depression, anxiety disorders, epilepsy, asthma, and stroke.[10]

The financial burden associated with migraine is also large. Direct costs, including physician visits, diagnostic tests, and treatment and medication have been estimated at slightly more than $11 billion annually. Patients with migraine have significantly higher overall health costs than patients without migraine ($7,007 per patient per year versus $4,436 per patient per year; $p < 0.001$).[11] Indirect costs, including missed workdays and reduced productivity, push the toll even higher.[12] One analysis has estimated the indirect economic burden related to migraine at approximately $12 billion annually.[13]

THE MIGRAINE MEDICATIONS In our fictional case study, Hemikrane (from the Greek word *hemikrania*, or migraine) is a hypothetical FDA-approved drug that has recently come on the market. It was shown in clinical trials to reduce both the frequency and the severity of migraine headaches and therefore is a preventive agent, not a treatment for migraine once it has set in, such as the commonly used migraine drug sumatriptan (Imitrex). Hemikrane is taken once a week.

The FDA approved Hemikrane based on two randomized, double-blind, controlled clinical trials, each of which had three arms. In one arm, patients who experienced multiple migraine episodes each month took Hemikrane weekly. In another arm, a comparable group of patients received a different migraine drug, Cephalal (from *cephalalgia*, Latin for headache), taken daily. In a third arm, another equivalent group of patients received placebos. Enough patients were enrolled in each arm of the study that statistically, if Hemikrane were better than placebo and at least as good as Cephalal, the results would show that.

Each of the two randomized studies enrolled approximately 2,000 patients and lasted six months. They excluded patients with uncontrolled high blood pressure, diabetes, heart disease, or kidney dysfunction. The patients in the studies were cared for in a number of academic centers and clinical trial sites. All patients submitted daily diaries, recording their migraine symptoms and any side effects.

RESULTS FROM THE TRIALS The trials showed that the patients who took Hemikrane had a clinically significant reduction in the frequency, severity, and duration of headaches compared to placebo, but not to Cephalal. The trials were not designed to evaluate the comparative safety of the drugs, but Hemikrane, the once-weekly drug, demonstrated a good safety profile. Patients on the drug experienced one key side effect, nausea, but in relatively small numbers.

Although the above studies demonstrated the efficacy of Hemikrane in a controlled environment with a modest number of carefully selected patients, they did not assess patient experience in a real-world setting. The latter environment is complicated by a need to be adherent with a once-daily or once-weekly therapy when patients are not in a controlled trial, and the financial obligations to pay for the drug. The monthly cost of Hemikrane to insurers is $200, whereas alternative medications like Cephalal cost insurers $150 per month. (For the purposes of this hypothetical example, we assume that copayments paid by patients are the same for all of these drugs.) However, a number of clinical experts and Hemikrane's manufacturer, Aesculapion, hypothesized that in routine community care, patients' adherence to the once-weekly therapy might exceed that of a daily regimen. Thus, greater differences might be observed in effectiveness.

THE COMPARATIVE EFFECTIVENESS STUDY To test this hypothesis and assess real-world performance, a major philanthropic organization with an interest in advancing treatments for migraine sufferers funded a collaboration among researchers at Harvard; a regional health insurance company, Trident Health; and Hemikrane's manufacturer, Aesculapion. The insurance company, Trident Health, provided access to an ex-

tensive database of five million people, which included information on medication use, doctor visits, emergency department evaluations, and hospitalizations.

Using these records, the study identified a cohort of patients with migraine who made frequent visits to doctors or hospital emergency departments. The study compared information about patients receiving Hemikrane, the once-weekly drug, with two comparison groups: a group of patients who received the daily prophylactic regimen with Cephalal, and a group of patients receiving no prophylactic therapy at all.

The investigators attempted to replicate the original randomized trial results by assessing the frequency with which all patients in the study had migraine headaches. Because the database did not contain a diary of daily symptoms, which had been collected in the trials, the researchers substituted as a proxy the amount of medications such as codeine and sumatriptan (Imitrex) that patients had used each month for treatment of acute migraines.

The investigators found that the group receiving Hemikrane, the once-weekly drug, had lower use of these symptom oriented medications than those on Cephalal or on no prophylaxis. The investigators also found that patients taking Hemikrane had fewer emergency department visits than those taking Cephalal or on no prophylaxis. Although the medication costs were higher for patients taking Hemikrane because of its higher monthly drug cost, the overall episode-of-care costs were lower than for the comparison group taking Cephalal. As hypothesized, the medication adherence was higher in the once-weekly Hemikrane patients than in the daily Cephalal patients (80 percent and 50 percent, respectively, using the metric of medication possession ratio, the number of days of medication dispensed as a percentage of 365 days).

The investigators were concerned that the above findings might reflect the unique characteristics of Trident Health's population of covered patients, regional practice patterns, copayment designs for medications, the study's analytic approach, or unmeasured severity. They also worried that the results could be confounded by differences in the patients receiving Hemikrane, Cephalal, or no prophylaxis. One possibility, for example, was that patients who experienced the worst migraines might be more inclined to take the new drug, Hemikrane, since they had failed all previously available therapies. In that case, the results for a truly matched group of patients might have shown even more pronounced benefit for Hemikrane.

To see if the findings could be replicated, the investigators contacted the pharmacy benefit management company, BestScripts, that worked with Trident Health, and asked for access to additional data. To protect against data "fishing expeditions" that might yield a distorted view of the data, a research protocol was developed before any data were examined. Statistical adjustments were also made to balance the three groups of patients to be studied as well as possible—those taking Hemikrane, those taking Cephalal, and those not on prophylaxis. A so-called propensity score method (which estimates the probability of a person's being in one of the three groups) was used and balanced on age, sex, number of previous migraine emergency department visits, type and extent of prior medication use, and selected comorbidities.

The comparative effectiveness study's results corroborated the clinical trial results, and logically expanded on them, through the use of several databases.

The pharmacy benefit manager, BestScripts, had access to data covering more than fifty million lives. The findings in this second, much larger, database corroborated the earlier assessment: The once-weekly prophylactic therapy with Hemikrane clearly reduced the use of medications such as codeine to relieve symptoms, as well as emergency department visits compared to the daily prophylaxis and no prophylaxis groups. Similarly, the Hemikrane group had significantly better medication adherence than the Cephalal group.

In addition, BestScripts had data from a subset of employers that collected work loss information about their employees. These data showed that patients on Hemikrane were out of work for fewer days each month than patients taking Cephalal. This result seemed biologically plausible, because migraine sufferers often miss work when episodes occur; a reduction in frequency and severity of episodes would enable those patients to maintain greater productivity.

However, because this finding had emerged from an observational study, rather than a randomized clinical trial, its validity might be questioned.

In summary, the comparative effectiveness study constituted a seemingly well-constructed study funded by a philanthropy and executed by strong and knowledgeable partners: an academic institution, a payer, and a manufacturer. The study's results corroborated the clinical trial results, and logically expanded on them, through the use of several databases. It also used a carefully prespecified analytic protocol, and it employed statistical approaches to adjust for selected threats to validity, such as one group of drug recipients' being sicker than another.

The study showed that Hemikrane, the once-weekly prophylactic medication, reduced migraine headaches as measured by insurance claims for treatment of migraine in two real-world environments with very large populations of patients. The study expanded on the results of the randomized clinical trials by showing the impact on the use of drugs that treat migraine symptoms, emergency department use, overall direct medical costs, and days lost from work. These results could plausibly reflect the greater medication adherence observed for patients receiving once-weekly versus daily prophylactic therapy.

In the written symposium that follows, various authors dissect the hypothetical studies from their different perspectives. Readers will find far from universal consensus on the many communication issues raised. But there is clear agreement that much more thought must go into shaping workable approaches to communicating comparative effectiveness research if the nation is to achieve the goal of making the research truly useful to patients. ■

The editors of *Health Affairs* produced the case study presented here, based on a May 9, 2012, roundtable of the symposium participants hosted by *Health Affairs* and sponsored by the National Pharmaceutical Council on issues surrounding the communication of comparative effectiveness research results.

NOTES

1 New era of comparative effectiveness research (October 2010 special issue). Health Aff (Millwood). 2010; 29(10).

2 Dentzer S. The researcher-in-chief at the Patient-Centered Outcomes Research Institute. Health Aff (Millwood). 2011;30(12):2252–8.

3 Griffin J, Godfrey B, Sherman R. Regulatory requirements of the Food and Drug Administration would preclude product claims based on observational research. Health Aff (Millwood). 2012;31(10):2188–92.

4 People & Places: How can policy protect public health and still foster innovation? Health Aff (Millwood). 2012;31(10):2241.

5 Lipton RB, Stewart WF, Diamond S, Diamond M, Reed M. Prevalence and burden of migraine in the United States: data from the American Migraine Study II. Headache. 2001;41:646–57.

6 Migraine Research Foundation. Migraine fact sheet [Internet]. New York (NY): The Foundation; [cited 2012 Sep 17]. Available from: http://www.migraineresearchfoundation.org/fact-sheet.html

7 Woolf V. In: Nicolson N, Trautmann J, editors. The letters of Virginia Woolf, vol. 4: 1929–1931. San Diego (CA): Harcourt Brace Jovanovich; 1979. p. 78.

8 Headache Classification Subcommittee of the International Headache Society. The international classification of headache disorders: 2nd ed. Cephalalgia. 2004;24(Suppl 1): 9–160.

9 Natoli J, Manack A, Dean B, Butler Q, Turkel C, Stovner L, et al. Global prevalence of chronic migraine: a systematic review. Cephalalgia. 2010;30:599–609.

10 Katsarava Z, Buse DC, Manack AN, Lipton RB. Defining the differences between episodic migraine and chronic migraine. Curr Pain Headache Rep. 2012;16(1):86–92.

11 Hawkins K, Wang S, Rupnow M. Direct cost burden among insured US employees with migraine. Headache. 2008;48(4):553–63.

12 Hu XH, Markson LE, Lipton RB, Stewart WF, Berger ML. Burden of migraine in the United States: disability and economic costs. Arch Intern Med. 1999;159(8):813–8.

13 Hawkins K, Wang S, Rupnow M. Indirect cost burden of migraine in the United States. J Occup Environ Med. 2007;49:368–74.

The NIH Common Fund
WE ACCELERATE DISCOVERY.

About the NIH Roadmap

NIH COMMON FUND AND THE AMERICAN REINVESTMENT AND RECOVERY ACT (ARRA)

OVERVIEW

The NIH Roadmap for Medical Research was launched in September, 2004, to address roadblocks to research and to transform the way biomedical research is conducted by overcoming specific hurdles or filling defined knowledge gaps. Roadmap programs span all areas of health and disease research and boundaries of NIH Institutes and Centers (ICs). These are programs that might not otherwise be supported by the NIH ICs because of their scope or because they are inherently risky. Roadmap Programs are expected to have exceptionally high potential to transform the manner in which biomedical research is conducted. They are also expected to be short term, 5–10 year programs. This incubator space time frame is intended to allow the major roadblocks that were defined for each program to be overcome, thereby stimulating further research conducted through the ICs.

NIH COMMON FUND

Roadmap programs were initially funded by a 1 percent contribution from each of the NIH ICs. In 2006, Congress responded to the need for NIH to develop innovative and cross-cutting programs by authorizing and funding the **NIH Common Fund** within the Office of the Director. The NIH Common Fund, including the programs of the NIH Roadmap for Medical Research, and is coordinated by the Office of Strategic Coordination, one of the six offices of the Division of Program Coordination, Planning, and Strategic Initiatives (DPCPSI) within the Office of the Director. The annual Common Fund budget was $498 million in 2008, highlighting its importance in the overall NIH funding environment. To date, the Common Fund has been used exclusively to support the Roadmap.

SCIENTIFIC AREAS

Planning and implementation of Roadmap/Common Fund programs are highly dynamic to allow the NIH the flexibility to respond quickly to new ideas, challenges, gaps and advances in biomedicine. Roadmap/Common Fund programs are developed from a strategic planning process involving broad, representative input from multiple scientific and public sources. Through this process, 12 roadblocks were articulated that are being addressed through many initiatives. These 12 challenges have been categorized according to three themes: New Pathways to Discovery, Research Teams of the Future, and Reengineering the Clinical Research Enterprise. Initiatives funded through the Roadmap/Common Fund fit into one or more of these major themes and address specific roadblocks or gaps to:

- Foster high-risk/high-reward research
- Enable the development of transformative tools and methodologies
- Fill fundamental knowledge gaps
- Change academic culture to foster collaboration

Although the NIH Roadmap is still in its infancy, many of its programs have achieved significant research advances.

Subscribe to the NIH Roadmap E-mail list

Office of Communications and Knowledge Transfer
Mission Statement

The Office of Communications and Knowledge Transfer (OCKT) promotes the communication of information to both internal and external customers. It designs, develops, implements, and manages programs for disseminating and implementing the results of Agency activities with the goal of changing audience behavior. Specifically, OCKT:

1. Communicates the results and significance of health services research and other AHRQ initiatives to the health care industry, health care providers, consumers and patients, policymakers, researchers, and the media with particular emphasis on communicating AHRQ initiatives in the ways each of these constituencies are most interested and are likely to lead to behavior change.
2. Manages the publishing and information distribution processes of the Agency.
3. Manages the public affairs and media activities of the Agency.
4. Produces information products in a variety of print and electronic formats that are scientifically sound and appropriately targeted to various audiences.
5. Manages AHRQ's Web site, electronic dissemination program, and the Information Resources Center to provide the administrative support for reference services and the distribution of technical information to Agency staff.
6. Directs a knowledge transfer program to promote the implementation and use of AHRQ tools, products, and initiatives by various stakeholders.
7. Evaluates the effectiveness of Agency dissemination and implementation strategies and undertakes changes indicated by such evaluations.

Current as of March 2008

Internet Citation:

Mission Statement: Office of Communications and Knowledge Transfer. March 2008. Agency for Healthcare Research and Quality, Rockville, MD. http://www.ahrq.gov/about/ockt/ocktmiss.htm

Agency for Healthcare Research and Quality • 540 Gaither Road Rockville, MD 20850 • Telephone: (301) 427-1364

THE NEW YORKER

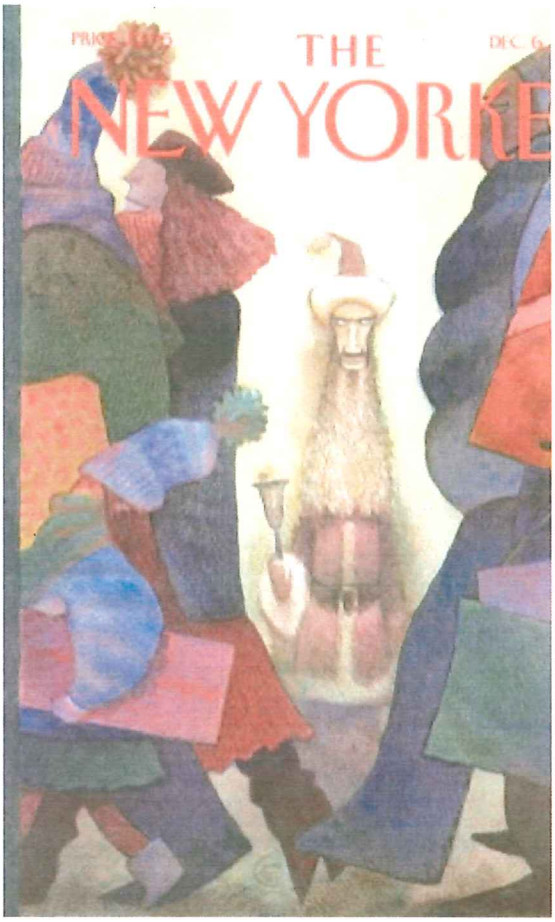
ANNALS OF MEDICINE

THE BELL CURVE

What happens when patients find out how good their doctors really are?

BY ATUL GAWANDE

DECEMBER 6, 2004



Every illness is a story, and Annie Page's began with the kinds of small, unexceptional details that mean nothing until seen in hindsight. Like the fact that, when she was a baby, her father sometimes called her Little Potato Chip, because her skin tasted salty when he kissed her. Or that Annie's mother noticed that her breathing was sometimes a little wheezy, though the pediatrician heard nothing through his stethoscope.

The detail that finally mattered was Annie's size. For a while, Annie's fine-boned petiteness seemed to be just a family trait. Her sister, Lauryn, four years older, had always been at the bottom end of the pediatrician's growth chart for girls her age. By the time Annie was three years old, however, she had fallen off the chart. She stood an acceptable thirty-four inches tall but weighed only twenty-three pounds—less than ninety-eight per cent of girls her age. She did not look malnourished, but she didn't look quite healthy, either.

"Failure to thrive" is what it's called, and there can be scores of explanations: pituitary disorders, hypothyroidism, genetic defects in metabolism, inflammatorybowel disease, lead poisoning, H.I.V., tapeworm infection. In textbooks, the complete list is at least a page long. Annie's doctor did a thorough workup. Then, at four o'clock on July 27, 1997—"I'll never forget that day," her mother, Honor, says—the pediatrician called the Pages at home with the results of a sweat test.

It's a strange little test. The skin on the inside surface of a child's forearm is cleaned and dried. Two small gauze pads are applied—one soaked with pilocarpine, a medicine that makes skin sweat, and the other with a salt solution. Electrodes are hooked up. Then a mild electric current is turned on for five minutes, driving the pilocarpine into the skin. A reddened, sweaty area about an inch in diameter appears on the skin, and a collection pad of dry filter paper is taped over it to absorb the sweat for half an hour. A technician then measures the concentration of chloride in the pad.

Over the phone, the doctor told Honor that her daughter's chloride level was far higher than normal. Honor is a hospital pharmacist, and she had come across children with abnormal results like this. "All I knew was that it meant she was going to die," she said quietly when I visited the Pages' home, in the Cincinnati suburb of Loveland. The test showed that Annie had cystic fibrosis.

Cystic fibrosis is a genetic disease. Only a thousand American children per year are diagnosed as having it. Some ten million people in the United States carry the defective gene, but the disorder is recessive: a child will develop the condition only if both parents are carriers and both pass on a copy. The gene—which was discovered, in 1989, sitting out on the long arm of chromosome No. 7—produces a mutant protein that interferes with cells' ability to manage chloride. This is what makes sweat from people with CF so salty. (Salt is sodium chloride, after all.) The chloride defect thickens secretions throughout the body, turning them dry and gluey. In the ducts of the pancreas, the flow of digestive enzymes becomes blocked, making a child less and less able to absorb food. This was why Annie had all but stopped growing. The effects on the lungs, however, are what make the disease lethal. Thickened mucus slowly fills the small airways and hardens, shrinking lung capacity. Over time, the disease leaves a child with the equivalent of just one functioning lung. Then half a lung. Then none at all.

The one overwhelming thought in the minds of Honor and Don Page was: We need to get to Children's. Cincinnati Children's Hospital is among the most respected pediatric hospitals in the country. It was where Albert Sabin invented the oral polio vaccine. The chapter on cystic fibrosis in the "Nelson Textbook of Pediatrics"—the bible of the specialty—was written by one of the hospital's pediatricians. The Pages called and were given an appointment for the next morning.

"We were there for hours, meeting with all the different members of the team," Honor recalled. "They took Annie's blood pressure, measured her oxygen saturation, did some other tests. Then they put us in a room, and the pediatrician sat down with us. He was very kind, but frank, too. He said, 'Do you understand it's a genetic disease? That it's nothing you did, nothing you can catch?' He told us the median survival for patients was thirty years. In Annie's lifetime, he said, we could see that go to forty. For him, he was sharing a great accomplishment in CF

care. And the news was better than our worst fears. But only forty! That's not what we wanted to hear."

The team members reviewed the treatments. The Pages were told that they would have to give Annie pancreatic-enzyme pills with the first bite of every meal. They would have to give her supplemental vitamins. They also had to add calories wherever they could—putting tablespoons of butter on everything, giving her ice cream whenever she wanted, and then putting chocolate sauce on it.

A respiratory therapist explained that they would need to do manual chest therapy at least twice a day, half-hour sessions in which they would strike—"percuss"—their daughter's torso with a cupped hand at each of fourteen specific locations on the front, back, and sides in order to loosen the thick secretions and help her to cough them up. They were given prescriptions for inhaled medicines. The doctor told them that Annie would need to come back once every three months for extended checkups. And then they went home to start their new life. They had been told almost everything they needed to know in order to give Annie her best chance to live as long as possible.

The one thing that the clinicians failed to tell them, however, was that Cincinnati Children's was not, as the Pages supposed, among the country's best centers for children with cystic fibrosis. According to data from that year, it was, at best, an average program. This was no small matter. In 1997, patients at an average center were living to be just over thirty years old; patients at the top center typically lived to be forty-six. By some measures, Cincinnati was well below average. The best predictor of a CF patient's life expectancy is his or her lung function. At Cincinnati, lung function for patients under the age of twelve—children like Annie—was in the bottom twenty-five per cent of the country's CF patients. And the doctors there knew it.

It used to be assumed that differences among hospitals or doctors in a particular specialty were generally insignificant. If you plotted a graph showing the results of all the centers treating cystic fibrosis—or any other disease, for that matter—people expected that the curve would look something like a shark fin, with most places clustered around the very best outcomes. But the evidence has begun to indicate otherwise. What you tend to find is a bell curve: a handful of teams with disturbingly poor outcomes for their patients, a handful with remarkably good results, and a great undistinguished middle.

In ordinary hernia operations, the chances of recurrence are one in ten for surgeons at the unhappy end of the spectrum, one in twenty for those in the middle majority, and under one in five hundred for a handful. A Scottish study of patients with treatable colon cancer found that the ten-year survival rate ranged from a high of sixty-three per cent to a low of twenty per cent, depending on the surgeon. For heartbypass patients, even at hospitals with a good volume of

experience, risk-adjusted death rates in New York vary from five per cent to under one per cent—and only a very few hospitals are down near the one-per-cent mortality rate.

It is distressing for doctors to have to acknowledge the bell curve. It belies the promise that we make to patients who become seriously ill: that they can count on the medical system to give them their very best chance at life. It also contradicts the belief nearly all of us have that we are doing our job as well as it can be done. But evidence of the bell curve is starting to trickle out, to doctors and patients alike, and we are only beginning to find out what happens when it does.

In medicine, we are used to confronting failure; all doctors have unforeseen deaths and complications. What we're not used to is comparing our records of success and failure with those of our peers. I am a surgeon in a department that is, our members like to believe, one of the best in the country. But the truth is that we have had no reliable evidence about whether we're as good as we think we are. Baseball teams have win-loss records. Businesses have quarterly earnings reports. What about doctors?

There is a company on the Web called HealthGrades, which for $7.95 will give you a report card on any physician you choose. Recently, I requested the company's report cards on me and several of my colleagues. They don't tell you that much. You will learn, for instance, that I am in fact certified in my specialty, have no criminal convictions, have not been fired from any hospital, have not had my license suspended or revoked, and have not been disciplined. This is no doubt useful to know. But it sets the bar a tad low, doesn't it?

In recent years, there have been numerous efforts to measure how various hospitals and doctors perform. No one has found the task easy. One difficulty has been figuring out what to measure. For six years, from 1986 to 1992, the federal government released an annual report that came to be known as the Death List, which ranked all the hospitals in the country by their death rate for elderly and disabled patients on Medicare. The spread was alarmingly wide, and the Death List made headlines the first year it came out. But the rankings proved to be almost useless. Death among the elderly or disabled mostly has to do with how old or sick they are to begin with, and the statisticians could never quite work out how to apportion blame between nature and doctors. Volatility in the numbers was one sign of the trouble. Hospitals' rankings varied widely from one year to the next based on a handful of random deaths. It was unclear what kind of changes would improve their performance (other than sending their sickest patients to other hospitals). Pretty soon the public simply ignored the rankings.

Even with younger patients, death rates are a poor metric for how doctors do. After all, very few young patients die, and when they do it's rarely a surprise; most already have metastatic cancer or horrendous injuries or the like. What one really wants to know is how we perform in typical circumstances. After I've done an appendectomy, how long does it take for my patients

to fully recover? After I've taken out a thyroid cancer, how often do my patients have serious avoidable complications? How do my results compare with those of other surgeons?

Getting this kind of data can be difficult. Medicine still relies heavily on paper records, so to collect information you have to send people to either scour the charts or track the patients themselves, both of which are expensive and laborious propositions. Recent privacy regulations have made the task still harder. Yet it is being done. The country's veterans' hospitals have all now brought in staff who do nothing but record and compare surgeons' complication rates and death rates. Fourteen teaching hospitals, including my own, have recently joined together to do the same. California, New Jersey, New York, and Pennsylvania have been collecting and reporting data on every cardiac surgeon in their states for several years.

One small field in medicine has been far ahead of most others in measuring the performance of its practitioners: cystic-fibrosis care. For forty years, the Cystic Fibrosis Foundation has gathered detailed data from the country's cystic-fibrosis treatment centers. It did not begin doing so because it was more enlightened than everyone else. It did so because, in the nineteen-sixties, a pediatrician from Cleveland named LeRoy Matthews was driving people in the field crazy.

Matthews had started a cystic-fibrosis treatment program as a young pulmonary specialist at Babies and Children's Hospital, in Cleveland, in 1957, and within a few years was claiming to have an annual mortality rate that was less than two per cent. To anyone treating CF at the time, it was a preposterous assertion. National mortality rates for the disease were estimated to be higher than twenty per cent a year, and the average patient died by the age of three. Yet here was Matthews saying that he and his colleagues could stop the disease from doing serious harm for years. "How long [our patients] will live remains to be seen, but I expect most of them to come to my funeral," he told one conference of physicians.

In 1964, the Cystic Fibrosis Foundation gave a University of Minnesota pediatrician named Warren Warwick a budget of ten thousand dollars to collect reports on every patient treated at the thirty-one CF centers in the United States that year—data that would test Matthews's claim. Several months later, he had the results: the median estimated age at death for patients in Matthews's center was twenty-one years, seven times the age of patients treated elsewhere. He had not had a single death among patients younger than six in at least five years.

Unlike pediatricians elsewhere, Matthews viewed CF as a cumulative disease and provided aggressive treatment long before his patients became sick. He made his patients sleep each night in a plastic tent filled with a continuous, aerosolized water mist so dense you could barely see through it. This thinned the tenacious mucus that clogged their airways and enabled them to cough it up. Like British pediatricians, he also had family members clap on the children's chests daily to help loosen the mucus. After Warwick's report came out, Matthews's treatment quickly

became the standard in this country. The American Thoracic Society endorsed his approach, and Warwick's data registry on treatment centers proved to be so useful that the Cystic Fibrosis Foundation has continued it ever since.

Looking at the data over time is both fascinating and disturbing. By 1966, mortality from CF nationally had dropped so much that the average life expectancy of CF patients had already reached ten years. By 1972, it was eighteen years—a rapid and remarkable transformation. At the same time, though, Matthews's center had got even better. The foundation has never identified individual centers in its data; to insure participation, it has guaranteed anonymity. But Matthews's center published its results. By the early nineteen-seventies, ninety-five per cent of patients who had gone there before severe lung disease set in were living past their eighteenth birthday. There was a bell curve, and the spread had narrowed a little. Yet every time the average moved up Matthews and a few others somehow managed to stay ahead of the pack. In 2003, life expectancy with CF had risen to thirty-three years nationally, but at the best center it was more than forty-seven. Experts have become as leery of life-expectancy calculations as they are of hospital death rates, but other measures tell the same story. For example, at the median center, lung function for patients with CF—the best predictor of survival—is about three-quarters of what it is for people without CF. At the top centers, the average lung function of patients is indistinguishable from that of children who do not have CF.

What makes the situation especially puzzling is that our system for CF care is far more sophisticated than that for most diseases. The hundred and seventeen CF centers across the country are all ultra-specialized, undergo a rigorous certification process, and have lots of experience in caring for people with CF. They all follow the same detailed guidelines for CF treatment. They all participate in research trials to figure out new and better treatments. You would think, therefore, that their results would be much the same. Yet the differences are enormous. Patients have not known this. So what happens when they find out?

In the winter of 2001, the Pages and twenty other families were invited by their doctors at Cincinnati Children's to a meeting about the CF program there. Annie was seven years old now, a lively, brown-haired second grader. She was still not growing enough, and a simple cold could be hellish for her, but her lung function had been stable. The families gathered in a large conference room at the hospital. After a brief introduction, the doctors started flashing PowerPoint slides on a screen: here is how the top programs do on nutrition and respiratory performance, and here is how Cincinnati does. It was a kind of experiment in openness. The doctors were nervous. Some were opposed to having the meeting at all. But hospital leaders had insisted on going ahead. The reason was Don Berwick.

Berwick runs a small, nonprofit organization in Boston called the Institute for Healthcare Improvement. The institute provided multimillion-dollar grants to hospitals that were willing to

try his ideas for improving medicine. Cincinnati's CF program won one of the grants. And among Berwick's key stipulations was that recipients had to open up their information to their patients—to "go naked," as one doctor put it.

Berwick, a former pediatrician, is an unusual figure in medicine. In 2002, the industry publication *Modern Healthcare* listed him as the third most powerful person in American health care. Unlike the others on the list, he is powerful not because of the position he holds. (The Secretary of Health and Human Services, Tommy Thompson, was No. 1, and the head of Medicare and Medicaid was No. 2.) He is powerful because of how he thinks.

In December, 1999, at a health-care conference, Berwick gave a forty-minute speech distilling his ideas about the failings of American health care. Five years on, people are still talking about the speech. The video of it circulated like samizdat. (That was how I saw it: on a grainy, overplayed tape, about a year later.) A booklet with the transcript was sent to thousands of doctors around the country. Berwick is middle-aged, soft-spoken, and unprepossessing, and he knows how to use his apparent ordinariness to his advantage. He began his speech with a gripping story about a 1949 Montana forest fire that engulfed a parachute brigade of firefighters. Panicking, they ran, trying to make it up a seventy-six-per-cent grade and over a crest to safety. But their commander, a man named Wag Dodge, saw that it wasn't going to work. So he stopped, took out some matches, and set the tall dry grass ahead of him on fire. The new blaze caught and rapidly spread up the slope. He stepped into the middle of the burned-out area it left behind, lay down, and called out to his crew to join him. He had invented what came to be called an "escape fire," and it later became a standard part of Forest Service fire training. His men, however, either thought he was crazy or never heard his calls, and they ran past him. All but two were caught by the inferno and perished. Inside his escape fire, Dodge survived virtually unharmed.

As Berwick explained, the organization had unravelled. The men had lost their ability to think coherently, to act together, to recognize that a lifesaving idea might be possible. This is what happens to all flawed organizations in a disaster, and, he argued, that's what is happening in modern health care. To fix medicine, Berwick maintained, we need to do two things: measure ourselves and be more open about what we are doing. This meant routinely comparing the performance of doctors and hospitals, looking at everything from complication rates to how often a drug ordered for a patient is delivered correctly and on time. And, he insisted, hospitals should give patients total access to the information. " 'No secrets' is the new rule in my escape fire," he said. He argued that openness would drive improvement, if simply through embarrassment. It would make it clear that the well-being and convenience of patients, not doctors, were paramount. It would also serve a fundamental moral good, because people should be able to learn about anything that affects their lives.

Berwick's institute was given serious money from the Robert Wood Johnson Foundation to offer those who used his ideas. And so the doctors, nurses, and social workers of Cincinnati Children's stood uncertainly before a crowd of patients' families in that hospital conference room, told them how poorly the program's results ranked, and announced a plan for doing better. Surprisingly, not a single family chose to leave the program.

"We thought about it after that meeting," Ralph Blackwelder told me. He and his wife, Tracey, have eight children, four of whom have CF. "We thought maybe we should move. We could sell my business here and start a business somewhere else. We were thinking, Why would I want my kids to be seen here, with inferior care? I want the very best people to be helping my children." But he and Tracey were impressed that the team had told them the truth. No one at Cincinnati Children's had made any excuses, and everyone appeared desperate to do better. The Blackwelders had known these people for years. The program's nutritionist, Terri Schindler, had a child of her own in the program. Their pulmonary specialist, Barbara Chini, had been smart, attentive, loving—taking their late-night phone calls, seeing the children through terrible crises, instituting new therapies as they became available. The program director, Jim Acton, made a personal promise that there would soon be no better treatment center in the world.

Honor Page was alarmed when she saw the numbers. Like the Blackwelders, the Pages had a close relationship with the team at Children's, but the news tested their loyalty. Acton announced the formation of several committees that would work to improve the program's results. Each committee, he said, had to have at least one parent on it. This is unusual; hospitals seldom allow patients and families on internal-review committees. So, rather than walk away, Honor decided to sign up for the committee that would reëxamine the science behind patients' care.

Her committee was puzzled that the center's results were not better. Not only had the center followed national guidelines for CF; two of its physicians had helped write them. They wanted to visit the top centers, but no one knew which those were. Although the Cystic Fibrosis Foundation's annual reports displayed the individual results for each of the country's hundred and seventeen centers, no names were attached. Doctors put in a call and sent e-mails to the foundation, asking for the names of the top five, but to no avail.

Several months later, in early 2002, Don Berwick visited the Cincinnati program. He was impressed by its seriousness, and by the intense involvement of the families, but he was incredulous when he learned that the committee couldn't get the names of the top programs from the foundation. He called the foundation's executive vice-president for medical affairs, Preston Campbell. "I was probably a bit self-righteous," Berwick says. "I said, 'How could you do this?' And he said, 'You don't understand our world.' " This was the first Campbell had heard about the requests, and he reacted with instinctive caution. The centers, he tried to

explain, give their data voluntarily. The reason they have done so for forty years is that they have trusted that it would be kept confidential. Once the centers lost that faith, they might no longer report solid, honest information tracking how different treatments are working, how many patients there are, and how well they do.

Campbell is a deliberate and thoughtful man, a pediatric pulmonologist who has devoted his career to cystic-fibrosis patients. The discussion with Berwick had left him uneasy. The Cystic Fibrosis Foundation had always been dedicated to the value of research; by investing in bench science, it had helped decode the gene for cystic fibrosis, produce two new drugs approved for patients, and generate more than a dozen other drugs that are currently being tested. Its investments in tracking patient care had produced scores of valuable studies. But what do you do when the research shows that patients are getting care of widely different quality?

A couple of weeks after Berwick's phone call, Campbell released the names of the top five centers to Cincinnati. The episode convinced Campbell and others in the foundation that they needed to join the drive toward greater transparency, rather than just react. The foundation announced a goal of making the outcomes of every center publicly available. But it has yet to come close to doing so. It's a measure of the discomfort with this issue in the cystic-fibrosis world that Campbell asked me not to print the names of the top five. "We're not ready," he says. "It'd be throwing grease on the slope." So far, only a few of the nation's CF treatment centers are committed to going public.

Still, after travelling to one of the top five centers for a look, I found I could not avoid naming the center I saw—no obscuring physicians' identities or glossing over details. There was simply no way to explain what a great center did without the particulars. The people from Cincinnati found this, too. Within months of learning which the top five centers were, they'd spoken to each and then visited what they considered to be the very best one, the Minnesota Cystic Fibrosis Center, at Fairview-University Children's Hospital, in Minneapolis. I went first to Cincinnati, and then to Minneapolis for comparison.

What I saw in Cincinnati both impressed me and, given its ranking, surprised me. The CF staff was skilled, energetic, and dedicated. They had just completed a flu-vaccination campaign that had reached more than ninety per cent of their patients. Patients were being sent questionnaires before their clinic visits so that the team would be better prepared for the questions they would have and the services (such as X-rays, tests, and specialist consultations) they would need. Before patients went home, the doctors gave them a written summary of their visit and a complete copy of their record, something that I had never thought to do in my own practice.

I joined Cori Daines, one of the seven CF-care specialists, in her clinic one morning. Among the patients we saw was Alyssa. She was fifteen years old, freckled, skinny, with nails

painted loud red, straight sandy-blond hair tied in a ponytail, a soda in one hand, legs crossed, foot bouncing constantly. Every few minutes, she gave a short, throaty cough. Her parents sat to one side. All the questions were directed to her. How had she been doing? How was school going? Any breathing difficulties? Trouble keeping up with her calories? Her answers were monosyllabic at first. But Daines had known Alyssa for years, and slowly she opened up. Things had mostly been going all right, she said. She had been sticking with her treatment regimen—twice-a-day manual chest therapy by one of her parents, inhaled medications using a nebulizer immediately afterward, and vitamins. Her lung function had been measured that morning, and it was sixty-seven per cent of normal—slightly down from her usual eighty per cent. Her cough had got a little worse the day before, and this was thought to be the reason for the dip. Daines was concerned about stomach pains that Alyssa had been having for several months. The pains came on unpredictably, Alyssa said—before meals, after meals, in the middle of the night. They were sharp, and persisted for up to a couple of hours. Examinations, tests, and X-rays had found no abnormalities, but she'd stayed home from school for the past five weeks. Her parents, exasperated because she seemed fine most of the time, wondered if the pain could be just in her head. Daines wasn't sure. She asked a staff nurse to check in with Alyssa at home, arranged for a consultation with a gastroenterologist and with a pain specialist, and scheduled an earlier return visit than the usual three months.

This was, it seemed to me, real medicine: untidy, human, but practiced carefully and conscientiously—as well as anyone could ask for. Then I went to Minneapolis.

The director of Fairview-University Children's Hospital's cystic-fibrosis center for almost forty years has been none other than Warren Warwick, the pediatrician who had conducted the study of LeRoy Matthews's suspiciously high success rate. Ever since then, Warwick has made a study of what it takes to do better than everyone else. The secret, he insists, is simple, and he learned it from Matthews: you do whatever you can to keep your patients' lungs as open as possible. Patients with CF at Fairview got the same things that patients everywhere did—some nebulized treatments to loosen secretions and unclog passageways (a kind of mist tent in a mouth pipe), antibiotics, and a good thumping on their chests every day. Yet, somehow, everything he did was different.

In the clinic one afternoon, I joined him as he saw a seventeen-year-old high-school senior named Janelle, who had been diagnosed with CF at the age of six and had been under his care ever since. She had come for her routine three-month checkup. She wore dyed-black hair to her shoulder blades, black Avril Lavigne eyeliner, four earrings in each ear, two more in an eyebrow, and a stud in her tongue. Warwick is seventy-six years old, tall, stooped, and frumpy-looking, with a well-worn tweed jacket, liver spots dotting his skin, wispy gray hair—by all appearances, a doddering, mid-century academic. He stood in front of Janelle for a moment,

hands on his hips, looking her over, and then he said, "So, Janelle, what have you been doing to make us the best CF program in the country?"

"It's not easy, you know," she said.

They bantered. She was doing fine. School was going well. Warwick pulled out her latest lung-function measurements. There'd been a slight dip, as there was with Alyssa. Three months earlier, Janelle had been at a hundred and nine per cent (she was actually doing better than normal); now she was at around ninety per cent. Ninety per cent was still pretty good, and some ups and downs in the numbers are to be expected. But this was not the way Warwick saw the results.

He knitted his eyebrows. "Why did they go down?" he asked.

Janelle shrugged.

Any cough lately? No. Colds? No. Fevers? No. Was she sure she'd been taking her treatments regularly? Yes, of course. Every day? Yes. Did she ever miss treatments? Sure. Everyone does once in a while. How often is once in a while?

Then, slowly, Warwick got a different story out of her: in the past few months, it turned out, she'd barely been taking her treatments at all.

He pressed on. "Why aren't you taking your treatments?" He appeared neither surprised nor angry. He seemed genuinely curious, as if he'd never run across this interesting situation before.

"I don't know."

He kept pushing. "What keeps you from doing your treatments?"

"I don't know."

"Up here"—he pointed at his own head—"what's going on?"

"*I don't know*," she said.

He paused for a moment. And then he began speaking to me, taking a new tack. "The thing about patients with CF is that they're good scientists," he said. "They always experiment. We have to help them interpret what they experience as they experiment. So they stop doing their treatments. And what happens? *They don't get sick.* Therefore, they conclude, Dr. Warwick is nuts."

"Let's look at the numbers," he said to me, ignoring Janelle. He went to a little blackboard he had on the wall. It appeared to be well used. "A person's daily risk of getting a bad lung illness with CF is 0.5 per cent." He wrote the number down. Janelle rolled her eyes. She began tapping her foot. "The daily risk of getting a bad lung illness with CF *plus treatment* is 0.05 per cent," he went on, and he wrote that number down. "So when you experiment you're looking at the difference between a 99.95-per-cent chance of staying well and a 99.5-per-cent chance of staying well. Seems hardly any difference, right? On any given day, you have basically a one-hundred-per-cent chance of being well. But"—he paused and took a step toward me—"it is a

big difference." He chalked out the calculations. "Sum it up over a year, and it is the difference between an eighty-three-per-cent chance of making it through 2004 without getting sick and only a sixteen-per-cent chance."

He turned to Janelle. "How do you stay well all your life? How do you become a geriatric patient?" he asked her. Her foot finally stopped tapping. "I can't promise you anything. I can only tell you the odds."

In this short speech was the core of Warwick's world view. He believed that excellence came from seeing, on a daily basis, the difference between being 99.5-per-cent successful and being 99.95-per-cent successful. Many activities are like that, of course: catching fly balls, manufacturing microchips, delivering overnight packages. Medicine's only distinction is that lives are lost in those slim margins.

And so he went to work on finding that margin for Janelle. Eventually, he figured out that she had a new boyfriend. She had a new job, too, and was working nights. The boyfriend had his own apartment, and she was either there or at a friend's house most of the time, so she rarely made it home to take her treatments. At school, new rules required her to go to the school nurse for each dose of medicine during the day. So she skipped going. "It's such a pain," she said. He learned that there were some medicines she took and some she didn't. One she took because it was the only thing that she felt actually made a difference. She took her vitamins, too. ("Why your vitamins?" "Because they're cool.") The rest she ignored.

Warwick proposed a deal. Janelle would go home for a breathing treatment every day after school, and get her best friend to hold her to it. She'd also keep key medications in her bag or her pocket at school and take them on her own. ("The nurse won't let me." "Don't tell her," he said, and deftly turned taking care of herself into an act of rebellion.) So far, Janelle was O.K. with this. But there was one other thing, he said: she'd have to come to the hospital for a few days of therapy to recover the lost ground. She stared at him.

"Today?"

"Yes, today."

"How about tomorrow?"

"We've failed, Janelle," he said. "It's important to acknowledge when we've failed."

With that, she began to cry.

Warwick's combination of focus, aggressiveness, and inventiveness is what makes him extraordinary. He thinks hard about his patients, he pushes them, and he does not hesitate to improvise. Twenty years ago, while he was listening to a church choir and mulling over how he might examine his patients better, he came up with a new stethoscope—a stereo-stethoscope, he calls it. It has two bells dangling from it, and, because of a built-in sound delay, transmits lung sounds in stereo. He had an engineer make it for him. Listening to Janelle with

the instrument, he put one bell on the right side of her chest and the other on her left side, and insisted that he could systematically localize how individual lobes of her lungs sounded.

He invented a new cough. It wasn't enough that his patients actively cough up their sputum. He wanted a deeper, better cough, and later, in his office, Warwick made another patient practice his cough. The patient stretched his arms upward, yawned, pinched his nose, bent down as far as he could, let the pressure build up, and then, straightening, blasted everything out. ("Again!" Warwick encouraged him. "Harder!")

He produced his most far-reaching invention almost two decades ago—a mechanized, chest-thumping vest for patients to wear. The chief difficulty for people with CF is sticking with the laborious daily regimen of care, particularly the manual chest therapy. It requires another person's help. It requires conscientiousness, making sure to bang on each of the fourteen locations on a patient's chest. And it requires consistency, doing this twice a day, every day, year after year. Warwick had become fascinated by studies showing that inflating and deflating a blood-pressure cuff around a dog's chest could mobilize its lung secretions, and in the mid-nineteen-eighties he created what is now known as the Vest. It looks like a black flak jacket with two vacuum hoses coming out of the sides. These are hooked up to a compressor that shoots quick blasts of air in and out of the vest at high frequencies. (I talked to a patient while he had one of these on. He vibrated like a car on a back road.) Studies eventually showed that Warwick's device was at least as effective as manual chest therapy, and was used far more consistently. Today, forty-five thousand patients with CF and other lung diseases use the technology.

Like most medical clinics, the Minnesota Cystic Fibrosis Center has several physicians and many more staff members. Warwick established a weekly meeting to review everyone's care for their patients, and he insists on a degree of uniformity that clinicians usually find intolerable. Some chafe. He can have, as one of the doctors put it, "somewhat of an absence of, um, collegial respect for different care plans." And although he stepped down as director of the center in 1999, to let a protégé, Carlos Milla, take over, he remains its guiding spirit. He and his colleagues aren't content if their patients' lung function is eighty per cent of normal, or even ninety per cent. They aim for a hundred per cent—or better. Almost ten per cent of the children at his center get supplemental feedings through a latex tube surgically inserted into their stomachs, simply because, by Warwick's standards, they were not gaining enough weight. There's no published research showing that you need to do this. But not a single child or teen-ager at the center has died in years. Its oldest patient is now sixty-four.

The buzzword for clinicians these days is "evidence-based practice"—good doctors are supposed to follow research findings rather than their own intuition or ad-hoc experimentation. Yet Warwick is almost contemptuous of established findings. National clinical guidelines for

care are, he says, "a record of the past, and little more—they should have an expiration date." I accompanied him as he visited another of his patients, Scott Pieper. When Pieper came to Fairview, at the age of thirty-two, he had lost at least eighty per cent of his lung capacity. He was too weak and short of breath to take a walk, let alone work, and he wasn't expected to last a year. That was fourteen years ago.

"Some days, I think, This is it—I'm not going to make it," Pieper told me. "But other times I think, I'm going to make sixty, seventy, maybe more." For the past several months, Warwick had Pieper trying a new idea—wearing his vest not only for two daily thirty-minute sessions but also while napping for two hours in the middle of the day. Falling asleep in that shuddering thing took some getting used to. But Pieper was soon able to take up bowling, his first regular activity in years. He joined a two-night-a-week league. He couldn't go four games, and his score always dropped in the third game, but he'd worked his average up to 177. "Any ideas about what we could do so you could last for that extra game, Scott?" Warwick asked. Well, Pieper said, he'd noticed that in the cold—anything below fifty degrees—and when humidity was below fifty per cent, he did better. Warwick suggested doing an extra hour in the vest on warm or humid days and on every game day. Pieper said he'd try it.

We are used to thinking that a doctor's ability depends mainly on science and skill. The lesson from Minneapolis is that these may be the easiest parts of care. Even doctors with great knowledge and technical skill can have mediocre results; more nebulous factors like aggressiveness and consistency and ingenuity can matter enormously. In Cincinnati and in Minneapolis, the doctors are equally capable and well versed in the data on CF. But if Annie Page—who has had no breathing problems or major setbacks—were in Minneapolis she would almost certainly have had a feeding tube in her stomach and Warwick's team hounding her to figure out ways to make her breathing even better than normal.

Don Berwick believes that the subtleties of medical decision-making can be identified and learned. The lessons are hidden. But if we open the book on physicians' results, the lessons will be exposed. And if we are genuinely curious about how the best achieve their results, he believes they will spread.

The Cincinnati CF team has already begun tracking the nutrition and lung function of individual patients the way Warwick does, and is getting more aggressive in improving the results in these areas, too. Yet you have to wonder whether it is possible to replicate people like Warwick, with their intense drive and constant experimenting. In the two years since the Cystic Fibrosis Foundation began bringing together centers willing to share their data, certain patterns have begun to emerge, according to Bruce Marshall, the head of quality improvement for the foundation. All the centers appear to have made significant progress. None, however, have progressed more than centers like Fairview.

"You look at the rates of improvement in different quartiles, and it's the centers in the top quartile that are improving fastest," Marshall says. "They are at risk of breaking away." What the best may have, above all, is a capacity to learn and adapt—and to do so faster than everyone else.

Once we acknowledge that, no matter how much we improve our average, the bell curve isn't going away, we're left with all sorts of questions. Will being in the bottom half be used against doctors in lawsuits? Will we be expected to tell our patients how we score? Will our patients leave us? Will those at the bottom be paid less than those at the top? The answer to all these questions is likely yes.

Recently, there has been a lot of discussion, for example, about "paying for quality." (No one ever says "docking for mediocrity," but it amounts to the same thing.) Congress has discussed the idea in hearings. Insurers like Aetna and the Blue Cross-Blue Shield companies are introducing it across the country. Already, Medicare has decided not to pay surgeons for intestinal transplantation operations unless they achieve a predefined success rate. Not surprisingly, this makes doctors anxious. I recently sat in on a presentation of the concept to an audience of doctors. By the end, some in the crowd were practically shouting with indignation: We're going to be paid according to our *grades*? Who is doing the grading? For God's sake, how?

We in medicine are not the only ones being graded nowadays. Firemen, C.E.O.s, and salesmen are. Even teachers are being graded, and, in some places, being paid accordingly. Yet we all feel uneasy about being judged by such grades. They never seem to measure the right things. They don't take into account circumstances beyond our control. They are misused; they are unfair. Still, the simple facts remain: there is a bell curve in all human activities, and the differences you measure usually matter.

I asked Honor Page what she would do if, after all her efforts and the efforts of the doctors and nurses at Cincinnati Children's Hospital to insure that "there was no place better in the world" to receive cystic-fibrosis care, their comparative performance still rated as resoundingly average.

"I can't believe that's possible," she told me. The staff have worked so hard, she said, that she could not imagine they would fail.

After I pressed her, though, she told me, "I don't think I'd settle for Cincinnati if it remains just average." Then she thought about it some more. Would she really move Annie away from people who had been so devoted all these years, just because of the numbers? Well, maybe. But, at the same time, she wanted me to understand that their effort counted for more than she was able to express.

I do not have to consider these matters for very long before I start thinking about where I would stand on a bell curve for the operations I do. I have chosen to specialize (in surgery for endocrine tumors), so I would hope that my statistics prove to be better than those of surgeons who only occasionally do the kind of surgery I do. But am I up in Warwickian territory? Do I have to answer this question?

The hardest question for anyone who takes responsibility for what he or she does is, What if I turn out to be average? If we took all the surgeons at my level of experience, compared our results, and found that I am one of the worst, the answer would be easy: I'd turn in my scalpel. But what if I were a C? Working as I do in a city that's mobbed with surgeons, how could I justify putting patients under the knife? I could tell myself, Someone's got to be average. If the bell curve is a fact, then so is the reality that most doctors are going to be average. There is no shame in being one of them, right?

Except, of course, there is. Somehow, what troubles people isn't so much being average as settling for it. Everyone knows that averageness is, for most of us, our fate. And in certain matters—looks, money, tennis—we would do well to accept this. But in your surgeon, your child's pediatrician, your police department, your local high school? When the stakes are our lives and the lives of our children, we expect averageness to be resisted. And so I push to make myself the best. If I'm not the best already, I believe wholeheartedly that I will be. And you expect that of me, too. Whatever the next round of numbers may say. ♦