

The Promise of Comparative Effectiveness Research

Paul Sullivan, MD

Don Goldmann, MD

THE AMERICAN RECOVERY AND REINVESTMENT ACT will provide an unprecedented stimulus for translational and health services research. A \$1.1 billion investment in comparative effectiveness research (CER)¹ should produce a torrent of new information about the effectiveness of drugs, technologies, and interventions. For this to result in better, more cost-effective health care, better evidence is needed to address the translational gap between clinical studies and everyday practice.² In essence, this is CER for implementation strategies (a type of CER seriously underrepresented in current discourse, but necessary to deliver on the Institute of Medicine's goals for improved health care quality).

Credit is due to the US Department of Health and Human Services for recognizing this need. Evaluation of the implementation and dissemination strategies for mainstream CER is embedded within several programs funded by the American Recovery and Reinvestment Act that share the broad objective of spreading CER findings widely (total funding: \$97 million; Karen Migdail, director of media relations, Agency for Healthcare Research and Quality, written communication, November 2010). For this investment to yield a timely return, researchers will need to adopt an array of pragmatic and context-sensitive study designs to determine which strategies are best for implementing evidence-based practices locally and at scale.

Traditional CER uses sophisticated data mining of longitudinal administrative databases to evaluate therapeutic and diagnostic interventions.³ These techniques could revolutionize the speed, population relevance, and context specificity of translational research. However, most databases currently do not include the rich clinical or organizational information needed to take the critical next step of comparing implementation strategies. Expansion of the electronic medical record and improvements in interoperability will help, especially if captured data include key information about organizational structures, clinical processes, risk factors and comorbidities, the true costs of diagnosis and treatment, and important outcomes such as functional status and quality of life. Obtaining and recording this information will require

closer collaboration among clinicians, health services researchers, systems improvement experts, and health information technologists.

In the meantime, new knowledge of strategies to implement the findings of CER is likely to come mainly from prospective studies. Randomized controlled trials (RCTs) are the criterion standard of clinical research, and designs such as cluster randomization are useful for implementation research in which the unit of analysis is a clinician, microsystem, organization, or community. However, such RCTs tend to be slow, expensive, and insensitive to the heterogeneous contexts in which their output will be deployed. Their restrictive entry criteria limit the rate of recruitment and generalizability of results, and their stringent protocols do not easily accommodate emerging new knowledge about the interventions themselves or changes in the environment in which they are being applied. Given these limitations, it is not surprising that RCTs have been criticized for failing to predict real-world outcomes.⁴

The pragmatic clinical trial attempts to address the limitations of conventional RCTs by seeking to reproduce conditions that the intervention will encounter in the real world.^{5,6} The entry criteria of pragmatic clinical trials aim to reflect the full range of patients and clinicians who will be using an intervention. This approach is well suited to research into systems of health care delivery that cannot be separated easily from everyday clinical practice. Pragmatic study designs also can take into account the local adaptations and amendments that often occur when new strategies are introduced.

Another flexible approach, the adaptive clinical trial, makes provision for planned, reactive changes as the trial progresses.^{7,8} This is especially important in evaluating the effects of a systems improvement intervention when observations reveal low compliance with care processes that are not the targets of the protocol but are linked to outcomes. In an adaptive trial, such unanticipated problems can be addressed in near real time by protocol adjustment. Bayesian data analysis can enhance efficiency and speed of adaptive

Author Affiliations: Institute for Healthcare Improvement, Cambridge, Massachusetts (Drs Sullivan and Goldmann); and Department of Pediatrics, Children's Hospital Boston and Harvard Medical School, Boston, Massachusetts (Dr Goldmann). Dr Sullivan is now with the Department of Quality Improvement, Salford Royal Foundation Trust, Salford, England.

Corresponding Author: Paul Sullivan, MD, Salford Royal Foundation Trust, Quality Improvement, Stolt Lane, Salford, LA M6 8HD, UK (paul.sullivan@srft.nhs.uk).

clinical trials by permitting early discontinuation of trial groups that are unlikely to succeed and allowing substantive results to be reported before the anticipated study completion date.⁹

Real-time, context-sensitive learning can also be achieved through quasi-experimental study designs, which are the staple of improvement and implementation science. While rapid cycle tests of change often are used to improve care processes and demonstrate progress over time in a specific organization, such projects may not generate transferable knowledge unless they are designed carefully and conducted rigorously. The addition of control groups improves the generalizability of such studies. Time series analyses are useful in evaluating the effects of planned interventions, and factorial and waiting time designs can increase confidence that observed improvements are not due solely to extraneous unmeasured interventions or secular trends. Improvement science is particularly suited for adapting interventions derived from more traditional CER to local contexts. For improvement science methods to realize their full potential to provide credible evidence about what works best, they must be applied with the same care and forethought as any other method of scientific inquiry, with clear descriptions of interventions, contexts, and evaluation methods.¹⁰ Quantitative evaluation alone is unlikely to yield a true understanding of factors influencing success or failure of an intervention, so input from other fields such as systems engineering, sociology, anthropology, behavioral science, and qualitative research is highly desirable.

Regardless of the study design, end points will have to be chosen carefully to provide usable information for policy makers. Aggregation of discrete outcomes is an appealing strategy to maximize statistical power and deliver a concise message, but it must be done with caution and rigor. Policy-relevant outcomes should include cost and cost-effectiveness data (including functional status and quality-of-life outcomes). Formal analysis of the economic consequences should be performed when appropriate.

Deploying these new research methods in real-world settings will be a challenge. Although a number of research groups are well equipped to perform field evaluations using multiple quantitative and qualitative methods, few have the interdisciplinary teams required to perform implementation research and evaluation, especially in the area of quality and systems improvement. Moreover, defined, well-organized test beds (organizations, health systems, and communities) are not widely available to enable the rapid deployment of such research. One approach might be to broaden the Clinical and Translational Science Awards model

of interdisciplinary research to focus on context-sensitive implementation research in diverse settings, both to develop new evidence regarding the effectiveness of high-priority CER interventions (eg, the use of patient navigators and care coordinators, medical home models, and strategies to reduce disparities) and translate the resulting evidence into practice. These research and field-testing units, which could be called interdisciplinary implementation and CER centers would include an interdisciplinary faculty capable of designing and executing multimodal implementation and evaluation research, as well as ready-to-roll field-testing sites and populations. They could leverage existing research networks, such as pediatric research in the office setting, to get as close as possible to the point of care. They would provide valuable inputs for comprehensive evidence sources, such as the Cochrane Effective Practice and Organization of Care Group, and serve as a framework in evaluating initiatives launched by the new Center for Medicare and Medicaid Innovation.

The expense and effort required to develop, support, and pilot-test such interdisciplinary research centers will not be trivial, but this investment is crucial to realize fully the potential of health care innovation and the enormous investments already made in Clinical and Translational Science Awards and CER to improve quality of care.

Conflict of Interest Disclosures: Both authors have completed and submitted the ICMJE Form for Disclosure of Potential Conflicts of Interest. Dr Sullivan reported receiving grant funding from the Health Foundation (London, England) as a Quality Improvement Fellow based at the Institute for Healthcare Improvement (Cambridge, Massachusetts). Dr Goldmann reported not receiving specific funding for his work.

REFERENCES

1. Steinbrook R. Health care and the American Recovery and Reinvestment Act. *N Engl J Med*. 2009;360(11):1057-1060.
2. Woolf SH. The meaning of translational research and why it matters. *JAMA*. 2008;299(2):211-213.
3. Schneeweiss S. Developments in post-marketing comparative effectiveness research. *Clin Pharmacol Ther*. 2007;82(2):143-156.
4. Horn SD, Gassaway J, Pentz L, James R. Practice-based evidence for clinical practice improvement: an alternative study design for evidence-based medicine. *Stud Health Technol Inform*. 2010;151:446-460.
5. Zwarenstein M, Treweek S, Gagnier JJ, et al; CONSORT Group; Pragmatic Trials in Healthcare (Practihc) Group. Improving the reporting of pragmatic trials: an extension of the CONSORT statement. *BMJ*. 2008;337:a2390.
6. Tunis SR, Stryer DB, Clancy CM. Practical clinical trials: increasing the value of clinical research for decision making in clinical and health policy. *JAMA*. 2003;290(12):1624-1632.
7. Chow SC, Chang M. Adaptive design methods in clinical trials: a review. *Orphanet J Rare Dis*. 2008;3:11.
8. Luce BR, Kramer JM, Goodman SN, et al. Rethinking randomized clinical trials for comparative effectiveness research: the need for transformational change. *Ann Intern Med*. 2009;151(3):206-209.
9. Berry DA. Clinical trials: is the Bayesian approach ready for prime time? yes! *Stroke*. 2005;36(7):1621-1622.
10. Davidoff F, Batalden P, Stevens D, Ogrinc G, Mooney SE; SQUIRE Development Group. Publication guidelines for quality improvement studies in health care: evolution of the SQUIRE project. *BMJ*. 2009;338:a3152.

By Steven D. Pearson and Peter B. Bach

DOI: 10.1377/hlthaff.2010.0623
 HEALTH AFFAIRS 29,
 NO. 10 (2010): 1796–1804
 ©2010 Project HOPE—
 The People-to-People Health
 Foundation, Inc.

TECHWATCH

How Medicare Could Use Comparative Effectiveness Research In Deciding On New Coverage And Reimbursement

Steven D. Pearson

(spearson@icer-review.org) is president of the Institute for Clinical and Economic Review, in Boston, Massachusetts.

Peter B. Bach is an associate attending physician at Memorial Sloan-Kettering Cancer Center, in New York City.

ABSTRACT Medicare must find new ways to achieve cost control without limiting access to beneficial services. We propose a payment model incorporating comparative effectiveness research to encourage Medicare to pay equally for services that provide comparable patient outcomes. The model would include higher payments for services demonstrated by adequate evidence to provide superior health benefits compared to alternative options. New services without such evidence would receive usual reimbursement rates for a limited time but then be reevaluated as evidence emerged. In spite of the substantial political hurdles to changing Medicare reimbursement, efforts should be made to use comparative effectiveness research to reward superior services, improve incentives for cost-effective innovation, and place Medicare on a more sustainable financial footing.

Just mentioning Medicare and comparative effectiveness research in the same sentence is enough to raise temperatures in Washington health policy circles. Those who see this research as a threat to patient choice or provider profits do not want it applied to Medicare. Those who see it as a remedy for the nation's health care ills do not want a politically explosive link to Medicare that might bring down the whole comparative effectiveness initiative.

Yet given that Medicare must find new approaches to help restrain the growth in health care costs, it seems inevitable that comparative effectiveness research will be considered an important potential tool in this effort.¹ But how can Medicare best use this evidence to guide coverage and reimbursement without limiting access to beneficial services? The uncertainty surrounding the answer to this question makes Medicare one of the most consequential and controversial “customers” of the comparative effectiveness re-

search initiative. That initiative, begun with the 2009 stimulus law and expanded under the Patient Protection and Affordable Care Act, is expected to direct hundreds of millions of dollars into comparative effectiveness research.

Seeking A Path Forward

In some respects, the prospects for applying comparative effectiveness research to Medicare's coverage or reimbursement policies appear quite limited. Under current law and because of years of precedent, Medicare generally covers any treatment that is deemed “reasonable and necessary,” regardless of the evidence on the treatment's comparative effectiveness or its cost in relation to other treatments.² Likewise, with only very rare exceptions, Medicare does not use comparative effectiveness information to set payment rates. Instead, it links reimbursement in one way or another to the underlying cost of providing services.³ To these established

precedents, the Affordable Care Act added several explicit limitations on Medicare's payment formulation and the use of comparative effectiveness research (Exhibit 1).

Despite these administrative and legislative constraints, comparative effectiveness research may be able to play a role in Medicare, particularly if a clear vision can be developed for the program's use of research data to help contain costs without restricting access to services. The most obvious strategy would be to use research to support decisions against covering new services. But this approach by itself would neither reap the greatest potential gain from the research nor avoid the political backlash that often follows decisions not to cover services.

The goal of this paper is to describe a different path forward. We believe that the best way for Medicare to benefit from the nation's new investment in comparative effectiveness research is to use it as a bridge—a conceptual and practical tool to link positive coverage decisions with evidence-based reimbursement levels.

Coverage And Reimbursement: Separate Silos

Medicare's processes for determining coverage and setting reimbursement rates are like computer programs that date all the way back to the 1960s. They demonstrate the arcane complexity of decades of ad hoc updates with no fundamental redesign. As a result of congressional mandate, determining coverage and setting reimbursement rates are separate and asynchronous, each containing multiple subprocesses.

COVERAGE Coverage of a service can start after a national decision or after a series of separate determinations made by the medical directors of independent contractors in different regions of the country. On rare occasions, coverage for specific services can be directly mandated by Congress; an example is the coverage of dialysis for kidney failure. And in some cases, national policies have linked coverage decisions to judgments by third parties, such as the decision to cover cancer drugs once they have been included in specific pharmaceutical compendia.^{4,5}

REIMBURSEMENT Medicare reimbursement rates are set centrally but are divided into nine-

EXHIBIT 1

Limits On Medicare's Use Of Comparative Effectiveness Research

Type of limitation/Affordable Care Act section	Specific limitation wording
CANNOT INCLUDE MANDATES FOR COVERAGE OR PAYMENT	
Section 6301(d)(8)(A)(iv)	The Institute shall ensure that the research findings not be construed as practice guidelines, coverage recommendations, payment, or policy recommendations.
Section 6301(j)(1)(A)	Nothing in this section shall be construed to permit the Institute to mandate coverage, reimbursement, or other policies for any public or private payer.
Section 937(a)(2)(B)	Materials, forums, and media used to disseminate the findings, informational tools, and resource databases shall not be construed as mandates, guidelines, or recommendations for payment, coverage, or treatment.
CAN USE EVIDENCE ONLY AS PART OF A LARGER PROCESS	
Section 1182(a)	The Secretary [of health and human services] may only use evidence and findings from research...to make a determination regarding coverage...if such use is through an iterative and transparent process which includes public comment and considers the effect on subpopulations.
Section 1182(b)(2)	Nothing in section 1181 shall be construed as authorizing the Secretary to deny coverage of items or services...solely on the basis of comparative clinical effectiveness research.
CANNOT USE EVIDENCE IN A MANNER THAT ASSIGNS A LOWER VALUE TO LIFE WITH A DISABILITY	
Section 1182(c)(1)	The Secretary shall not use evidence or findings from comparative clinical effectiveness research...in determining coverage, reimbursement, or incentive programs...in a manner that ...treats extending the life of an elderly, disabled, or terminally ill individual as of lower value than extending the life of an individual who is younger, nondisabled, or not terminally ill.
Section 1182(d)(1)	...precludes, or with the intent to discourage, an individual from choosing a health care treatment based on how the individual values the tradeoff between extending the length of their life and the risk of disability.
CANNOT DEVELOP OR USE A DOLLARS PER QUALITY-ADJUSTED LIFE-YEAR OR SIMILAR COST-EFFECTIVENESS THRESHOLD AS PART OF RECOMMENDATIONS OR TO DETERMINE COVERAGE OR PAYMENT	
Section 1182(e)	The [Institute]...shall not develop or employ a dollars-per-quality adjusted life year (or similar measure that discounts the value of a life because of an individual's disability) as a threshold to establish what type of health care is cost effective or recommended. The Secretary shall not utilize such an adjusted life year (or such a similar measure) as a threshold to determine coverage, reimbursement, or incentive programs.

SOURCE Authors' analysis of the Patient Protection and Affordable Care Act. **NOTE** The "Institute" refers to the Patient-Centered Outcomes Research Institute established in the Affordable Care Act.

teen major payment systems for different types or locations of services. Each system varies in the way it establishes and adjusts prices, and each has its own method of incorporating new tests or treatments. In addition, the reimbursement algorithms and equations differ widely among the systems, including prospective payments for acute care hospital services, fee-for-service payments made to individual physicians, and the complicated formulas used to pay private prescription drug plans for the Part D outpatient drug benefits.

Not infrequently, a single service, such as radiation therapy, will be reimbursed at quite different levels depending on whether the service is performed in a hospital or an outpatient setting, in the Northeast or the South, and in the current calendar year or the next.⁶

THE COMMON THREAD Despite the complicated web of rules for coverage and reimbursement, at a fundamental level the different strands constitute a basic approach that Medicare has followed for decades. Coverage is determined without any requirement for evidence demonstrating that the service in question is equally or more effective than other available options. Then, separately, after a decision is made to cover a particular service, Medicare sets the payment level with the primary goal of reimbursing hospitals or providers for their cost plus some profit margin. This cost-plus reimbursement approach is based largely on estimates of the material, time, and training involved in delivering the service.⁷

Policy analysts have long decried the way in which a payment system oriented primarily toward reimbursing costs can create perverse incentives to develop and then overuse expensive services.⁶ Despite this concern, little has changed over the years in Medicare's basic approach to setting reimbursement rates. Even those prospective payment approaches that have been adopted by Medicare, such as diagnosis-related group (DRG) payments for hospital services, are initially set and ultimately adjusted to reflect the underlying costs of providing these services.

POTENTIAL FOR REFORM In recent years, there have been many calls for reform of Medicare's coverage and reimbursement processes. The Affordable Care Act authorized several different types of pilots and other programs—including patient-centered medical homes, bundling of payments, and accountable care organizations—that attempt to reorganize the delivery of care while shifting some of the financial incentive to reducing use.

These new payment models may ultimately pave the way to a sweeping overhaul of the Medi-

Medicare has never considered the cost-effectiveness of a service to be part of coverage criteria.

care reimbursement system. However, testing, refining, and expanding the models will take years, and in the near term, Medicare is likely to rely heavily on its current coverage and payment systems for most services.⁷

Linking Comparative Effectiveness Research To Reimbursement

Many policy makers and academicians have recognized for decades the possible advantages of using evidence of effectiveness to help set Medicare's reimbursement levels.⁶ More recently, the new national comparative effectiveness initiative has stimulated further consideration of how research results could be applied to pricing or determining the best level of patient copayments.⁸⁻¹⁰

LINKING FROM THE OUTSET We believe that the time is ripe for Medicare to use comparative effectiveness research to reach a new paradigm of paying equally for services that provide equivalent results. To accomplish this goal, the program's coverage and reimbursement processes would need to be linked from the outset, when the evidence for or against a service's comparative clinical effectiveness would be weighed. Exhibit 2 shows how different levels of evidence could be linked to reimbursements.

This framework would not require that Medicare change the current threshold of evidence that it uses for deciding whether a new service is "reasonable and necessary." However, when Medicare determined that a service would be covered, it would also need to make a simultaneous determination of the service's comparative effectiveness. We propose that comparative effectiveness evidence be assigned to one of three categories—evidence of *superior* comparative clinical effectiveness; evidence of *comparable* comparative clinical effectiveness; or *insufficient evidence* to determine comparative clinical effectiveness.

► **SUPERIOR EFFECTIVENESS:** Medicare would assign a service to the first category if there were

adequate evidence to demonstrate that the service is more effective or has fewer side effects, or both, compared to the most relevant clinical standard. Payment for this kind of service would be set according to current Medicare formulas at a rate sufficient to reimburse providers for the cost of providing what is, demonstrably, a superior service.

► **COMPARABLE EFFECTIVENESS:** A service would go in the second category if there were sufficient evidence to determine that the service's clinical effectiveness was comparable to its most relevant alternative. Such a service would be assigned a payment level equal to that of the alternative. Payment along these lines would be a form of "reference pricing," sometimes used for pharmaceuticals, in which payers pay for brand-name drugs at the lower amount they pay for equally effective generic alternatives.

► **INSUFFICIENT EVIDENCE:** The third category would be for those services that meet the usual "reasonable and necessary" standard but for which there is insufficient evidence for whether the new service is comparable, superior, or inferior to relevant alternatives. We believe that the majority of services covered under Medicare have historically fallen into this category because the evidence has been insufficient to determine their comparative clinical effectiveness.

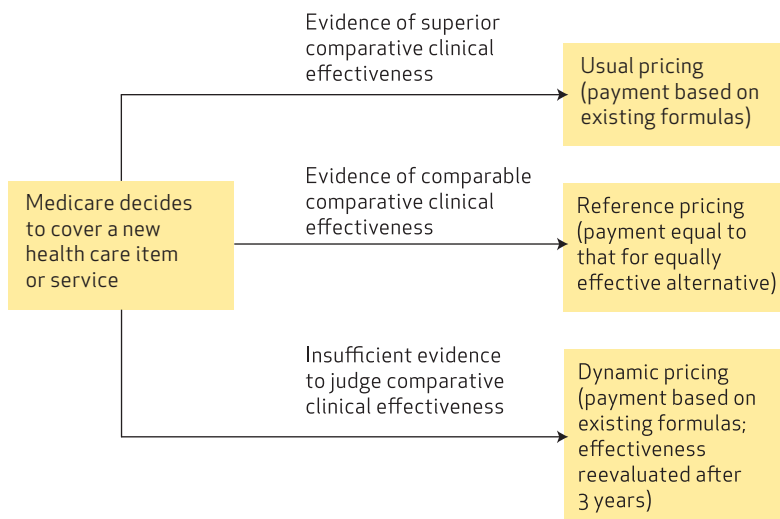
For services with insufficient comparative effectiveness evidence, we propose that Medicare use "dynamic pricing." In other words, the program would set payments according to the current cost-plus reimbursement formula for a period of three years. At the end of this period, Medicare would decide whether additional evidence was now available to determine if the service were superior, comparable, or inferior to alternatives.

Should Medicare decide that the service's clinical advantages had been demonstrated, the usual cost-plus reimbursement formula would continue indefinitely beyond the initial three-year period. However, if Medicare concluded that the evidence after the first three years demonstrated no comparative advantage or was still insufficient to make a clear determination, payment would be lowered to the reference price paid by Medicare for the most relevant alternative care option. Were the additional evidence to suggest that the new service was inferior to existing options, Medicare could reevaluate whether the service was reasonable and necessary.

EXAMPLE As an example of how this dynamic pricing approach would work, consider Medicare's coverage and reimbursement decisions

EXHIBIT 2

Using Comparative Effectiveness Research To Determine Medicare Coverage And Reimbursement



SOURCE Authors' proposal. **NOTES** Upon making a new coverage decision, Medicare would assign the service to one of three payment categories based on the evidence of comparative clinical effectiveness. After an initial three-year period, if services with insufficient evidence do not provide additional evidence demonstrating superior comparative clinical effectiveness, payment would drop to reference pricing levels.

concerning intensity-modulated radiation therapy when that treatment was introduced into practice in the early 2000s.¹¹ This therapy was a new form of "conformal" radiation therapy, which used computers to create three-dimensional pictures in order to target the highest possible dose of radiation to cancerous tumors while sparing normal tissue. Clinicians felt that intensity-modulated radiation therapy was an advance because of its new approach to focusing the radiation even more narrowly on the specific area presumed to harbor tumor cells. However, when Medicare decided to cover the treatment, there had been neither randomized trials nor contemporaneous cohort studies comparing the effectiveness and toxicities of intensity-modulated radiation therapy to traditional three-dimensional therapy.

► **EXISTING REIMBURSEMENT:** Nonetheless, following existing Medicare payment policies, the initial reimbursement for intensity-modulated radiation therapy was set in recognition of the increased cost of the necessary equipment and the complexity of its treatment planning process. For prostate cancer—the most common form of cancer for which this new therapy was used—reimbursement to providers for a single course of treatment was set at approximately \$42,000. For three-dimensional therapy, providers received only \$10,000.

This discrepancy led providers around the country to buy intensity-modulated radiation therapy machines and to abandon conventional three-dimensional therapy. One result was that there was no practical way to recruit clinicians or patients for trials comparing the two therapies. Another result was to add an estimated \$1.5 billion per year to Medicare's costs for prostate cancer alone.^{11,12}

► **REIMBURSEMENT USING PROPOSED FRAMEWORK:** In contrast, consider how coverage and reimbursement could have been managed for this new therapy according to our proposed framework. Medicare's decision to cover the treatment would have been accompanied by a determination that there was insufficient evidence with which to judge the comparative clinical effectiveness of the newer treatment against the standard three-dimensional therapy. Thus, the new treatment would have been slated to receive the higher—\$42,000 per course—reimbursement for only three years.

This higher payment level would have reimbursed physicians and hospitals for the higher costs of using intensity-modulated radiation in the short term, thereby not stifling innovation. However, the three-year time period would have created an incentive for manufacturers and clinicians to perform the research needed to evaluate the clinical performance of the new therapy in comparison to the standard three-dimensional treatment.

Not all of the important questions related to the comparative clinical effectiveness of intensity-modulated radiation could have been addressed in three years. Medicare would therefore have had to clarify in its initial decision what specific gaps in the evidence it expected to see addressed during the three-year period. But it is likely that three years would have been enough time for high-quality controlled trials to compare the rates of key adverse effects such as bowel damage and urinary dysfunction—outcomes of great importance to patients.

Over those three years, intensity-modulated radiation therapy could have been introduced into widespread use in a more measured manner and linked to efforts to develop further evidence of its comparative effectiveness. If the evidence had shown that the new treatment had lower risks of side effects than the old one, then reimbursement for the new service would have remained higher. Clinicians, patients, and payers all could have learned more about the relative advantages of intensity-modulated radiation therapy and made informed decisions accordingly.

If after three years the evidence had shown that the new therapy did not offer clinical advantages

Paying more for better results is the best way to spur innovation.

compared to the three-dimensional treatment, Medicare would have reduced its reimbursement for the new therapy to equal that paid for the traditional treatment. Intensity-modulated radiation therapy would still have been available to patients, but incentives for developing less expensive versions of the treatment would have been strong. And Medicare would not have been trapped into years of significantly higher payments for a new technology that might not improve patients' outcomes more than the older treatment.

Implications

The success or failure of this proposed approach to reimbursement would be judged by its impact on three factors: patients' access to care, innovation, and the increasing cost of Medicare.

EFFECTS ON ACCESS First, how would the approach affect access to care? New or improved treatments often are more expensive than older ones. Therefore, it is possible that any change to reimbursement that would threaten the ability of manufacturers and providers to recoup the higher costs required for a particular service would lead to much slower diffusion of more-expensive interventions. An implicit incentive to place new services in academic and other settings well suited to clinical research, in order to develop new evidence, might also delay diffusion throughout the broad health care system. If this were the case, patients in rural areas and others without ready access to academic sites might have less access to newly covered services than they do now.

We believe that this trade-off would be justifiable. It should be remembered that the services to be reimbursed at reference-price levels or at usual levels during a trial period are services lacking evidence to demonstrate that they are better than other options. In many ways, these services occupy a similar position to implantable cardio-defibrillators and cochlear implants—services that Medicare covered in the past under the rubric of “coverage with evidence development” policies.¹³

Medicare policies are the way they are because legislative, regulatory, and political processes have made them that way.

When there is insufficient evidence to assess comparative effectiveness, it is possible that further evidence would show that a new service was inferior to existing options, at least for some types of patients. Limiting the rapid dissemination of such a service is likely to be in the best interest of most patients. Therefore, it is neither unethical nor without precedent for Medicare to institute a coverage and reimbursement strategy that may limit access to some new services.

IMPACT ON INNOVATION A second standard by which to evaluate this proposed approach is its effect on innovation. It is almost certain that the incentives inherent in the approach would lead to a change in the focus and type of innovation in the US health care system. By dissociating reimbursement from considerations of comparative clinical effectiveness, Medicare's current payment systems have promoted with equal favor those new services that provide greatly improved patient outcomes, and services that are expensive, poorly studied, and either comparable to other options or only marginal improvements.

Would the proposed dynamic pricing model help tip the balance toward a better kind of innovation? There is no doubt that any change toward evidence-based reimbursement would appear threatening to many manufacturers, whose current business models are based on the existing payment structure. Ultimately, however, paying more for better results is the best way to spur the kind of innovation desired most by patients, clinicians, and payers.

The definition of "superior" comparative clinical effectiveness would have to be carefully crafted in order to support certain kinds of incremental innovation in challenging therapeutic areas such as oncology. But for most new services, reimbursement according to usual cost-based policies for a limited three-year period seems a reasonable way to help more-costly in-

novations enter the market even if adequate evidence of their comparative advantages does not yet exist. Under this set of incentives, focused, skilled innovators who know that their greatest reward will come from producing interventions of greatly enhanced effectiveness at a reasonable cost will dominate the market.

EFFECTS ON COSTS The final measure by which this proposed change to reimbursement would be evaluated is its ability to affect the overall cost trend in Medicare and, by extension, in private health care systems. We view this new payment model as part of a longer-term solution to Medicare's fiscal challenges, alongside efforts to move toward global payment models.

Even in many of the pilot programs of the health reform law, fee-for-service reimbursement is likely to remain the dominant approach to payment. Moreover, Medicare must function now with a complex set of different delivery systems, small group practices, and individual clinicians. As our example of the radiation treatments for prostate cancer suggests, we believe that modifying the program's payment structures to provide equal reimbursement for comparable results would achieve significant cost savings for Medicare while preserving patients' access to new interventions.

Challenges

Implementing the proposed approach would require overcoming many challenges, as follows.

A SINGLE TIME LIMIT First among these would be the difficulty of establishing a single time limit for further generation of evidence of comparative clinical effectiveness. Three years would probably be more than enough time to fill critical evidence gaps for some services, but for others it might be nearly impossible to complete and analyze appropriate comparative studies within that amount of time. An alternative approach would be to set a unique time interval for each service, but this strategy would bring the added complexity, uncertainty, and political pressure of ad hoc decisions.

We favor a common three-year period for all services because we believe that it would be possible for Medicare to frame the evidence requirements so that they can be met within that time frame. Moreover, the impetus of a three-year period would intensify the efforts of manufacturers, researchers, and the new Patient-Centered Outcomes Research Institute to develop methods and data infrastructure to help meet the needs of decision makers more rapidly.

TRANSPARENCY IN DEFINITIONS Another challenge would be defining with adequate transparency what constitutes "comparable" as opposed

to “superior” comparative clinical effectiveness. Medicare has long faced a similar challenge with its definition of “reasonable and necessary,” but it would need to strengthen its efforts to clarify how individual patients’ preferences and the likelihood of individual variation in response to treatments would be taken into consideration. Integrating the goals of personalized care with those of a payment approach based on evidence-based categories would probably require continuing adjustments.

DIFFERENT PATIENT SUBGROUPS In addition, Medicare would face the practical difficulty of assigning a payment category for a service when comparative effectiveness results differed across patient subgroups. Intensity-modulated radiation therapy, for example, may be judged comparable to three-dimensional therapy in treating patients with types of prostate cancer judged to be slow-growing and at low risk of metastasis, but superior for patients with higher-risk tumors. Ideally, Medicare would set two prices for intensity-modulated radiation therapy and pay the higher price only when diagnostic or other codes identify the patient as being at higher risk. But Medicare has never tried making such differential payments, and the procedure would be complex.

PREVIOUS COVERAGE DECISIONS Implementation of our proposed payment model would also raise the thorny question of whether previously covered services should be grandfathered out of the new system, or whether their prices should also be subject to determinations of comparative effectiveness evidence.

RECOGNIZING MINIMAL IMPROVEMENTS Ultimately, our proposed approach does nothing to address the dilemma posed by services that may be minimally better than existing options but that are far more costly to provide. Addressing this policy challenge will require other mechanisms.

The Way Forward

Despite these challenges and limitations, we believe that the possible advantages of using comparative effectiveness research to help set reimbursements for newly covered services are too important to let the approach languish. But is change possible? A shift in Medicare’s fundamental approach to coverage and reimbursement decisions would require new legislative authorities and would be highly contested by those with a vested interest in existing reimbursement systems. Current Medicare policies are the way they are because legislative, regulatory, and political processes have made them that way. The difficulty of changing these sys-

The model we suggest would allow equal payments without uprooting the entire incentive system for innovation.

tems is hard to overstate.

Nonetheless, we think that there is reason for some guarded optimism, largely because policy makers are feeling an increasing need to address the budgetary strain caused by Medicare’s ballooning costs. They are looking for solutions that are both politically palatable and administratively practical. We believe that our proposed three-prong reimbursement model can be both.

POLITICS Politically, the straightforward idea of paying equally for comparable results would make sense to most Americans. Powerful advocates—including purchasers, patients, and producers of cost-effective services—could be expected to form a natural coalition in support of an evidence-based approach to reimbursement. Private health plans and state Medicaid programs, both of which have greater flexibility than Medicare, might be early adopters of this kind of reimbursement approach. Lessons learned from such smaller-scale efforts could help pave the way for policy makers to overcome the traditional political inertia associated with Medicare.

PRAGMATISM From the practical perspective, our approach could be embedded in the existing administrative units that manage Medicare coverage and reimbursement. Skills and experience exist within the Centers for Medicare and Medicaid Services to allow this new model to be implemented without the need to create novel organizational units or other infrastructure.

It is also conceivable that our proposed payment model could be incorporated into Medicare processes without running afoul of the language in the Affordable Care Act that restricts Medicare’s use of comparative effectiveness research. The intent of that language was to keep national comparative effectiveness assessments from leading to mandatory decisions about coverage or payment. The secretary of health and human services might well determine that applying comparative effectiveness evidence in the way

Medicare needs new approaches to help it move to a more solid financial footing.

we describe would constitute the required “iterative and transparent process” (Exhibit 1) for setting reimbursement rates. But if Medicare efforts to adopt our proposed approach were judged to be inconsistent with the intent of the Affordable Care Act, we would recommend perseverance. New amendments to the act could be considered, but we would hope that minor changes to the reimbursement model would be all that would be needed to allow Medicare to move forward without sacrificing the basic goal of using comparative effectiveness evidence to link coverage and reimbursement.

CONCLUSION Paying equally for comparable results is a powerful principle, and the model

we suggest would allow equal payments to be implemented without uprooting the entire incentive system for innovation. Much like the patent period granted to new pharmaceuticals, a designated period of higher Medicare reimbursement—to cover the higher costs of new services, even if their comparative advantages have not yet been demonstrated—would allow manufacturers to recoup some of their investments in the development of new tests or treatments. Limiting the time that this higher payment rate was in effect would create a substantial incentive for manufacturers to conduct comparative effectiveness studies. For providers, this approach offers not only the prospect of better evidence with which to care for individual patients, but also the beneficial and sobering effect of removing perverse incentives to invest in and deliver services that add to the cost but not the quality of care.

Medicare needs new approaches to help it move to a more solid financial footing. Given the magnitude of the fiscal challenges ahead, using comparative effectiveness research to set reimbursement rates at the time of coverage is a promising option that we would argue the nation cannot afford to ignore. ■

The authors thank Joshua N. Mirkin for his input into the paper, as well as Gail Wilensky and two anonymous reviewers for their helpful comments.

NOTES

- 1 Orszag PR. Health care and the budget: issues and challenges for reform; statement before the U.S. Senate Committee on the Budget [Internet]. Washington (DC): Congressional Budget Office; 2007 Jun 21 [cited 2010 Aug 13]. Available from: <http://www.cbo.gov/ftpdocs/82xx/doc8255/06-21-HealthCareReform.pdf>
- 2 Tunis SR. Why Medicare has not established criteria for coverage decisions. *N Engl J Med*. 2004;350(21):2196–8.
- 3 Thorwarth WT. From concept to CPT code to compensation: how the payment system works. *J Am Coll Radiol*. 2004;1:48–53.
- 4 Raab GG, Parr DH. From medical invention to clinical practice: the reimbursement challenge facing new device procedures and technology—part 3: payment. *J Am Coll Radiol*. 2006;3:842–50.
- 5 Medicare Payment Advisory Commission. Medicare payment basics [Internet]. Washington (DC): MedPAC; [cited 2010 Jun 3]. Available from: http://www.medpac.gov/payment_basics.cfm
- 6 Newhouse JP. Pricing the priceless: a health care conundrum. Cambridge (MA): MIT Press; 2002.
- 7 Medicare Payment Advisory Commission. Report to the Congress: Medicare payment policy [Internet]. Washington (DC): MedPAC; 2010 Mar [cited 2010 Jun 3]. Available from: http://www.medpac.gov/documents/Mar10_EntireReport.pdf
- 8 Medicare Payment Advisory Commission. Report to the Congress: reforming the delivery system [Internet]. Washington (DC): MedPAC; 2008 Jun [cited 2010 Jun 30]. p. 107–37. Available from: http://www.medpac.gov/documents/Jun08_EntireReport.pdf
- 9 Wilensky GR. Cost-effectiveness information: yes, it's important, but keep it separate, please! *Ann Intern Med*. 2008;148(12):967–8.
- 10 Pearson S. From better evidence to better care: using comparative effectiveness research to guide practice and policy. In: Implementing comparative effectiveness research: priorities, methods, and impact. Washington (DC): Brookings Institution; 2009 Jun 9 [cited 2010 Sep 13]. Available from: http://www.brookings.edu/~media/Files/events/2009/0609_health_care_cer/0609_health_care_cer.pdf
- 11 Saul S. Profits and questions on prostate cancer therapy. *New York Times*. 2006 Dec 1.
- 12 Konski A, Watkins-Bruner D, Feigenberg S, Hanlon A, Kulkarni S, Beck JR, et al. Using decision analysis to determine the cost-effectiveness of intensity-modulated radiation therapy in the treatment of intermediate risk prostate cancer. *Int J Radiat Oncol Biol Phys*. 2006;66(2):408–15.
- 13 Tunis SR, Pearson SD. Coverage options for promising technologies: Medicare's “coverage with evidence development.” *Health Aff (Millwood)*. 2006;25(5):1218–30.

ABOUT THE AUTHORS: STEVEN D. PEARSON & PETER B. BACH



Steven D. Pearson is president of the Institute for Clinical and Economic Review.

In this issue of *Health Affairs*, Steven Pearson and Peter Bach take on a topic that became verboten in the national health reform debate: how Medicare could use comparative effectiveness research to make coverage and reimbursement decisions. The authors draw on their experience as physicians and specialists in the assessment of medical evidence and outcomes to build the case.

Pearson is the founder and president of the Institute for Clinical and Economic Review (ICER) at the Massachusetts General Hospital's Institute for Technology Assessment. ICER brings stakeholders together to evaluate the comparative effectiveness of medical interventions. He also serves as a lecturer in Harvard's Department of Population Medicine, a visiting scientist in the Department of Bioethics at the National Institutes of Health (NIH), and a member of NIH's Comparative Effectiveness Research Steering Committee.

Pearson previously was vice chair of the Medicare Evidence Development and Coverage Advisory Committee and a senior fellow at America's Health Insurance Plans. He received his medical degree from the University of California, San Francisco, School of Medicine, followed by a master of science degree in health policy and management from the Harvard School of Public Health. Earlier this year he was named to the board of directors of Health Technology Assessment International, an organization that supports and promotes the use of health technology assessment to inform decision makers and to make efficient use of health care resources.



Peter B. Bach is an associate attending physician at Memorial Sloan-Kettering Cancer Center.

Peter Bach is a pulmonary and critical care physician and member of the Health Outcomes Research Group at Memorial Sloan-Kettering Cancer Center. He focuses in part on developing approaches to creating new clinical evidence for

therapies and devices in typical clinical care settings. He serves on several national committees, including the Committee on Performance Management of the National Committee for Quality Assurance and the Institute of Medicine's National Cancer Policy Forum.

Bach knows the ways of Medicare well, having previously served as a senior adviser to the administrator of the Centers for Medicare and Medicaid Services. He earned his medical degree from the University of Minnesota, followed by a master of arts in public policy from the University of Chicago.

"How payers will use comparative effectiveness research has always been a flash point of controversy, given the possible use of evidence to limit access to certain interventions," Pearson observes. Yet on the plus side, he and Bach say that evidence could also allow payers to develop more-innovative benefit designs and reimbursement strategies to achieve better care at lower costs.

"The greatest potential for comparative effectiveness is in setting reimbursement levels," Bach agrees. In effect, data about what really works best in health care would then translate into health care prices, and broadly encourage more Americans to buy the most effective—and cost-effective—care.



US Government Kicks Off Program for Comparative Effectiveness Research

Mike Mitka

HEALTH REFORM LEGISLATION AND an earmarked \$1.1 billion have set the stage for the commencement of federally endorsed comparative effectiveness research. But it remains to be seen just how such research will be carried out—and how the results might change health care.

The engine for federally sponsored comparative effectiveness research is the new Patient-Centered Outcomes Research Institute, whose 21-member board of governors, which includes the directors of the Agency for Healthcare Research and Quality and the National Institutes of Health, was announced September 23. The institute, established by the Patient Protection and Affordable Care Act of 2010, is a non-profit public-private endeavor charged with disbursing funding for comparative effectiveness research.

According to the act, the purpose of the institute is “to assist patients, clinicians, purchasers, and policy-makers in making informed health decisions.” Institute staff are to apply research and evidence to improve methods for preventing, diagnosing, treating, monitoring, and managing health conditions. The institute is also charged with disseminating research findings on health outcomes, clinical effectiveness, and the appropriateness of medical treatments, services, and items.

To achieve these goals, the institute will create a standing methodology committee to develop and update scientifically based standards for research conducted through the institute. In addition, the institute will ensure peer review and make research findings publicly available within 90 days. The institute will also allow for public comment periods prior to such actions as the adoption of national priorities, the research project agenda, the methodological standards,

and the peer review process, and after the release of draft findings of reviews of existing research and evidence.

As clear as these expectations are, the institute must still navigate political waters roiled with charges that comparative effectiveness research is a tool for the continued takeover of health care by the federal government and a way to justify health care rationing. Such attitudes among certain members of Congress resulted in the passage of a law that severely restricts one of the perceived benefits of comparative effectiveness research: the possibility of cost savings in health care.

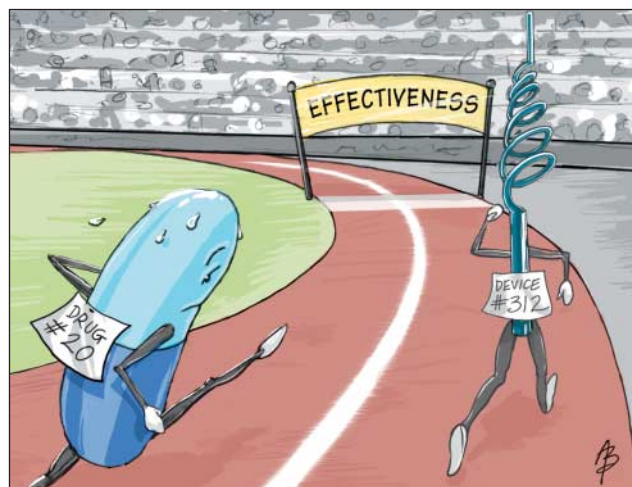
Savings would presumably follow from the identification of proven therapies that are less expensive than those in common use, although such research could also identify more effective treatments that are more expensive. The law stipulates that the institute “shall not develop or employ a dollars-per-quality adjusted life year (or similar measure that discounts the value of a life because of an individual’s disability) as a threshold to establish what type of health care is cost effective or recommended” under Medicare.

With such restrictions, little wonder that researchers and economists are looking at other ways to bring cost into the

comparative effectiveness research conversation. Some of those approaches are discussed in *Health Affairs*, which dedicated its October issue to a variety of articles exploring issues surrounding comparative effectiveness research.

LOOPHOLES

Harold C. Sox, MD, emeritus professor of medicine at Dartmouth Medical School in Hanover, NH, and editor emeritus of the *Annals of Internal Medicine*, points out that the law’s language only forbids discussing cost-effectiveness in terms of use by Medicare and that it does not specifically preclude weighing costs in a theoretical fashion. Sox stated his views in an article he coauthored with Alan M. Garber, MD, PhD, professor of medicine and of economics and health research and policy at Stanford University in Stanford, Calif (Garber AM and Sox HC. *Health Aff [Millwood]*. 2010;29[10]:1805-1811). “According to the letter of the law, the institute could actually commission cost-effective analysis where you use quality-adjusted life-years,” said Sox in an interview. “What it forbids is using that information to set a threshold below which something is cost-effective and above which it is not cost-effective.”



Comparative effectiveness research allows investigators to determine the superiority, inferiority, or equivalence of various interventions when pitted against each other.



Sox said the law's ambiguity regarding cost considerations reflects the debate seen across the nation.

"Comparative effectiveness research has gotten so much support because the cost of health care is affecting the country's economy, and yet there is a peculiar attitude from the federal government about considering cost-effectiveness," Sox said. "People are concerned about having the government say that it will pay for this and not pay for that. Patients and doctors do not want to have their hands tied by government rules about which health resources would be paid for under Medicare."

But Sox and Garber, after discussing the loophole allowing for cost analysis, do not call for the institute to commission such study. "Doing so might lead to the appearance, if not the reality, that the institute was attempting to define care standards for federal health insurance programs in the United States, which the Affordable Care Act discouraged," they wrote.

Instead, they recommend that the institute insist that studies it sponsors provide enough information to enable others to perform the analyses. That would allow analysts who are free of sanctions to develop cost-effective information.

"It seems like a pretty obvious idea that research sponsors would require authors to gather data that can be useful to people, but not necessarily those doing the original research," Sox said. "The [National Institutes of Health] has been requiring authors to get cost data on randomized clinical trials it sponsors, and the trialists may not be the ones who use that information."

INCENTIVES FOR INNOVATION

In another *Health Affairs* article, Steven Pearson, MD, MSc, and Peter B. Bach, MD, argue that Medicare must find ways to reduce costs. They suggest that a tweak in the existing law or an interpretation by the Secretary of the Department of Health and Human Services could give the agency the ability to use comparative effectiveness research to reward superior

services and to improve incentives for cost-effective innovation (Pearson SD and Bach PB. *Health Aff [Millwood]*. 2010; 29[10]:1796-1804). Pearson is president of the Institute for Clinical and Economic Review at the Massachusetts General Hospital's Institute for Technology Assessment in Boston; Bach is a pulmonary and critical care physician and member of the Health Outcomes Research Group at Memorial Sloan-Kettering Cancer Center in New York City.

The authors argue that the current method for reimbursing new medical devices or services too often allows for coverage without adequate rigorous study to determine whether the intervention is superior, or at least equivalent, to existing treatments. They cited Medicare's approval in the early 2000s for coverage of intensity-modulated radiation therapy, most often used to treat prostate cancer (at a coverage cost of about \$42 000), despite the absence of randomized trials or contemporaneous cohort studies comparing the effectiveness and toxic effects of the therapy with those of traditional 3-dimensional therapy (costing about \$10 000). The discrepancy in reimbursement led to hospitals and physicians buying the new machines and abandoning traditional therapy, costing Medicare \$1.5 billion annually for treating prostate cancer alone.

Pearson and Bach proposed that Medicare combine reimbursement and comparative effectiveness research from the beginning of any coverage plan. They would have Medicare assign a service to 1 of 3 categories: superior effectiveness, comparable effectiveness, and insufficient evidence. A device or service deemed superior would have adequate evidence that it was more effective or caused fewer adverse effects, or both, compared with the current clinical standard, and payment would be set in accord with the current Medicare cost-plus reimbursement formula. Use of a device or service rated as comparable with current clinical standards would be compensated with payments equal to the established treatment.

For therapies rated as having insufficient evidence, Pearson and Bach

would allow for payment at the current cost-plus reimbursement formula for 3 years to give the manufacturer time to show, in a more rigorous manner, whether the service was superior, comparable, or inferior to alternatives. If further time were needed, Medicare would then pay for the service at the comparable effectiveness rate. If the treatment were shown to be inferior, then Medicare could reevaluate whether the service was reasonable and necessary.

"We are trying to change the conversation from talking about reimbursement to discussing methods providing evidence of efficacy," Pearson said in an interview. "This is an attempt to think downstream when it will not feel as much as a Democratic or Republican issue, but a national issue to maintain access and change the cost patterns of Medicare."

A CAUTIONARY VIEW

A *Health Affairs* article by James C. Robinson, PhD, MPH, a professor of health economics at the University of California, Berkeley, provided a sobering reminder of just how difficult it will be to adopt comparative effectiveness research into clinical practice (Robinson JC. *Health Aff [Millwood]*. 2010;29[10]:1788-1795). "There are numerous examples of the very slow diffusion of new knowledge into practice due to inertia of all types," said Robinson, a former editor of *Health Affairs*. Slow adoption is particularly an issue "if the new knowledge should change the established patient preferences and established physician practice patterns and their potential revenue," he noted.

The link between comparative effectiveness research and cost-effectiveness is tenuous at best, Robinson said. "Comparative effectiveness research has been sold under the assumption that it will reduce the rate of health care cost increases, but that is not plausible on its face because some results will show expensive care is better care, and that will be adopted. And where it shows that less care is better and less expensive, that often means less revenues for physicians and hospitals, and they will fight it." □

Legislating against Use of Cost-Effectiveness Information

Peter J. Neumann, Sc.D., and Milton C. Weinstein, Ph.D.

The Patient-Centered Outcomes Research Institute . . . shall not develop or employ a dollars per quality adjusted life year (or similar measure that discounts the value of a life because of an individual's disability) as a threshold to establish what type of health care is cost effective or recommended. The Secretary shall not utilize such an adjusted life year (or such a similar measure) as a threshold to determine coverage, reimbursement, or incentive programs under title XVIII.

— The Patient Protection and Affordable Care Act¹

In 1996, after 2 years of deliberation, the U.S. Panel on Cost-Effectiveness in Health and Medicine, composed of physicians, health economists, ethicists, and other health policy experts, recommended that cost-effectiveness analyses should use quality-adjusted life-years (QALYs) as a standard metric for identifying and assigning value to health outcomes.² The recently enacted Patient Protection and Affordable Care Act (ACA) created a Patient-Centered Outcomes Research Institute (PCORI) to conduct comparative-effectiveness research (CER) but prohibited this institute from developing or using cost-per-QALY thresholds. The two events serve as revealing bookends to a long-standing debate over the role and shape of cost-effectiveness analysis in U.S. health care.

QALYs provide a convenient yardstick for measuring and comparing health effects of varied interventions across diverse diseases and conditions. They represent the effects of a health intervention in terms of the gains or losses in time spent in a series of “quality-weighted” health states. QALYs are used in cost-effective-

ness analyses (termed “cost-utility analyses” when QALYs are included) to inform resource-allocation decisions: the cost-per-QALY ratios of different interventions are compared in order to determine the most efficient ways of furnishing health benefits. In contrast, other health outcomes are generally expressed in disease-specific terms, such as incidence of cardiovascular events, cancer progression, intensity of pain, or loss of function. Though useful for measuring the effects of particular treatments, these outcomes do not permit comparisons among diseases and conditions or between treatment and prevention.³

Researchers have published thousands of cost-utility studies in leading medical and health policy journals. Health policymakers around the world have used such analyses to inform clinical guidelines and reimbursement decisions. The U.S. government, through agencies such as the Agency for Healthcare Research and Quality, the Centers for Disease Control and Prevention, and the National Institutes of Health, has sponsored cost-utility analyses. Medical specialty

societies have cited cost-utility studies in support of clinical guidelines.

The ACA specifically forbids the use of cost per QALY “as a threshold.” The precise intent and consequences of this language are unclear. One might interpret it to mean that the PCORI, or its contractors or grantees, can still calculate cost-per-QALY ratios as long as they are not compared with a threshold (e.g., \$100,000 per QALY) or used to make a recommendation based on such a threshold. Comparisons of cost-per-QALY ratios across interventions could still be useful to decision makers even without the invocation of an explicit threshold. However, the ACA suggests a broader ban on the use of cost-utility analyses — and this could have a chilling effect on the field.

The ACA's language might be seen as symptomatic of the legislation's aversion to policies that critics might see as enacting “big-government” health care or “death panels.” It may reflect a certain xenophobia toward the kinds of approaches used in Britain, where the National Institute of Health and Clinical Excellence makes recommendations about technologies and services on the basis of cost-per-QALY thresholds. Reflecting this sentiment, the ACA creates a new CER institute that it labels “patient-centered” and states that the findings of PCORI-sponsored research cannot be construed as mandates for practice guidelines, coverage recommendations,

payment, or policy recommendations.

The ban on using cost-per-QALY thresholds also seems to reflect long-standing concerns that the approach would discriminate on the basis of age and disability. The worry is that the metric unfairly favors younger and

ethical dilemmas. Taken literally, it means that spending resources to extend by a month the life of a 100-year-old person who is in a vegetative state cannot be valued differently from spending resources to extend the life of a child by many healthy years. Though the ACA may be

metric in the field of cost-effectiveness analysis is regrettable.

The ACA states that the PCORI is intended to assist patients, clinicians, purchasers, and policy-makers. Yet a ban on cost-utility analysis would leave decision makers with less information with which to compare the relative effects of interventions across diseases. The ACA states that PCORI-produced CER is intended to inform, not mandate, decisions. Why, then, be so prescriptive about costs per QALY? Better to simply develop and disseminate the information and let decision makers choose whether or not to use it and in what settings. Decision makers could consider cost-per-QALY ratios alongside other criteria, such as the priority of an intervention for vulnerable populations and concerns about equity and fairness. How can our market-driven health system work efficiently if participants lack information about the relationship between the costs and benefits of health interventions?

As the country searches for ways to curb health care spending, consideration of the cost-effectiveness of health interventions will unavoidably be part of the health care debate, alongside considerations of possible payment- and delivery-system reforms. The use of explicit, standard metrics such as cost-per-QALY ratios has the advantage of transparency and can help direct our resources toward the greatest health gains. These kinds of analyses will therefore endure as a rough benchmark of value and as a normative guide to resource-allocation decisions. It

QALYs simply give priority to interventions that offer the most health benefit in terms of measures people care about — more time spent in good health.

healthier populations that have more potential QALYs to gain.

To be sure, there are legitimate debates about the role of QALYs as the sole benchmark of health gains for purposes of allocating society's resources. However, acknowledging the measure's limitations, panels in the United States and Britain and at the World Health Organization have found QALYs preferable to alternative measures of health improvement.

QALYs simply give priority to interventions that offer the most health benefit in terms of measures people care about — more time spent in good health. In fact, populations with more impairment typically fare better in cost-effectiveness analyses, because they have more to gain from interventions; for example, it is generally less cost-effective to screen or treat healthier persons than persons who have poorer health at baseline or who are at greater risk for complications.

Moreover, a ban on valuing life extension presents its own

seeking to avert discrimination, it instead helps to perpetuate the current system of implicit rationing and hidden biases.

The antagonism toward cost-per-QALY comparisons also suggests a bit of magical thinking — the notion that the country can avoid the difficult trade-offs that cost-utility analysis helps to illuminate. It pretends that we can avert our eyes from such choices, and it kicks the can of cost-consciousness farther down the road. It represents another example of our country's avoidance of unpleasant truths about our resource constraints. Although opportunities undoubtedly exist to eliminate health care waste, the best way to improve health and save money at the same time is often to redirect patient care resources from interventions with a high cost per QALY to those with a lower cost per QALY.⁴ At a time when health care costs loom as the greatest challenge facing our country's fiscal well-being, legislating against the use of the standard

would be unfortunate if the ACA created a barrier to their development and use.

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

From the Center for the Evaluation of Value and Risk in Health, Institute for Clinical Re-

search and Health Policy Studies, Tufts Medical Center (P.J.N.), and the Department of Health Policy and Management, Harvard School of Public Health (M.C.W.) — both in Boston.

1. The Patient Protection and Affordable Care Act. PL 111-148. 3-23-2010.
2. Weinstein MC, Siegel JE, Gold MR, Kamlet MS, Russell LB. Recommendations of the Panel on Cost-effectiveness in

Health and Medicine. JAMA 1996;276:1253-8.

3. Neumann PJ, Greenberg D. Is the United States ready for QALYs? Health Aff (Millwood) 2009;28:1366-71.

4. Weinstein MC, Skinner JA. Comparative effectiveness and health care spending — implications for reform. N Engl J Med 2010; 362:460-5.

Copyright © 2010 Massachusetts Medical Society.



Comparative Effectiveness Research: The Need for a Uniform Standard

By Scott Gottlieb, M.D., and Coleen Klasmeier

Reconciling a new comparative effectiveness research (CER) agency with scientific standards established in existing regulations would enable government agencies to share a consistent framework for making decisions based on CER. It would provide the kind of level playing field and clear guidelines that are needed to spur medical product developers and private payers to generate more of their own comparative evidence.

Legislative proposals to create a new federal agency focused on CER are moving ahead in the Senate, but thus far the legislative drafts have not set out the scientific corroboration that should be necessary before the research is used in support of policy decisions. Just as troubling is the failure of these proposals to consider existing federal regulations governing how the Food and Drug Administration (FDA) regulates the dissemination of precisely these kinds of research results. While the FDA's regulations speak to the quality of information that private companies can share with doctors and patients, the risk is that other agencies—principally Medicare—will use instead a much lower standard when they act on the federally advanced CER. This sets up an asymmetric playing field; policy and reimbursement decisions will be made in one agency based on data that a sister agency judges too unreliable even for purposes of private-sector sharing.

Engaging these issues now serves multiple goals. It establishes agreement around the level of rigor that ought to govern the conduct of CER. It also helps to avert inevitable conflicts between different government agencies over the

Scott Gottlieb, M.D. (scott.gottlieb@aei.org), is a resident fellow at AEI. Coleen Klasmeier (cklasmeier@sidley.com) is a partner and head of the FDA regulatory practice at Sidley Austin LLP.

Key points in this Outlook:

- Policymakers are interested in establishing a formal framework for comparative effectiveness research.
- A science-minded agency, such as the FDA, should establish guidelines about when information is sufficiently rigorous to be actionable by other government entities.
- The FDA's "substantial clinical experience" standard is appropriate for judging the veracity of CER. It is a standard that can be shared across the FDA and a CER agency. It passes legal muster, too.
- Budget estimates suggest CER will not reduce the deficit. We can expect then that CER will be used in the future to make decisions about access and pricing of medical products.
- We need to invest in the creation of a clinical trial infrastructure that enables more rigorous CER.

appropriate standards for making decisions based on the results of these studies. It is especially important that we grapple with how federal programs like Medicare will use the CER data. It is unrealistic to think we can prohibit Medicare from considering these studies. But we can develop a standard for weighing this evidence that recognizes that Medicare is no ordinary payer, since it drives decisions made by the entire health care market, and that the new CER agency will not be an ordinary research origination either. It will carry the government's imprimatur, elevating the impact of its work, and it will be under immediate political pressure to show economic and political payoffs for the investments that are being made in these research studies.

Ideally, the federal regulatory and policy criteria for sharing and acting on CER that are applied to private actors should reflect the same standards and principles applied to public agencies. The FDA, the federal agency responsible for the safety, efficacy, and proper labeling of prescription drugs and medical devices, has developed objective criteria against which statements based on CER could readily be adjudicated. Even though the kinds of study designs contemplated by the new CER agency fall short of the rigor that the FDA requires to support medical product approvals, the FDA's regulatory scheme includes alternative substantiation standards, one of which—"substantial clinical experience"—was specifically designed to assist in the analysis of data from CER-type studies. The FDA's current regulations can thus provide uniform criteria for determining when CER is sufficiently rigorous to form the basis of policy and regulatory decisions across different government agencies.

If we do not develop a common standard for substantiating CER, the FDA could be put into conflict with the new agency, and the private sector will be held to a much higher standard than the government when it comes to the degree of reliability that evidence must have to support decisions based on its findings. Government would control the field for developing CER, communicating the results and making reimbursement decisions based on the findings. Sponsors, who are subject to a higher standard enforced by the FDA, might not only be prohibited from sharing similar CER, but also be unable to comment on the results of the studies generated by the federal agency. The risk is that the government alone may become the sole arbiter of CER, since it alone would be exempt from FDA regulation. This could hurt consumers, who ultimately benefit from a competitive market for clinical information.

Reconciling Conflicts

These conflicts could be addressed by reconciling the evidence standards that the FDA applies to medical product companies with the standards used by other agencies (including Medicare) to assess and act on information from studies undertaken by the CER agency. With the adoption of a uniform understanding of the level of clinical substantiation needed to make government decisions based on the results of CER research, conflicts between federal agencies could be avoided, physicians and patients could have a clear understanding of the level of scientific substantiation that guides government decisions, and companies would have greater incentives to sponsor their own research, all thanks to a clear path and level playing field for sharing and acting on this kind of information.

If we do not develop a common standard for substantiating CER, the FDA could be put into conflict with the new agency, and the private sector will be held to a much higher standard than the government when it comes to the degree of reliability that evidence must have to support decisions based on its findings.

There are other benefits to involving the FDA in establishing these standards. The FDA can provide a rigorous check on the new CER agency. The fear of CER critics is that even in cases in which the information from the new agency might be inconclusive, it could nonetheless be used in a political way to support government decisions about access and pricing.¹ A science-minded agency like the FDA can provide a dispassionate actor in the process of establishing a framework for when CER information is sufficiently rigorous to be actionable by Medicare and other government entities. Medicare—and even the new CER agency—might be politically motivated to over- or underinterpret the resulting CER to support narrow policy goals.

These policy efforts also must recognize that establishing an evidence standard is not the same thing as assigning a "grade" to different CER studies based on

their relative utility. The principle for substantiation should primarily consider the threshold level of underlying rigor and the reliability of the finding to establish a benchmark for when a study should be deemed sufficiently thorough to be an actionable piece of information for regulatory and policy decisions. Many “grading” systems seek to encompass a wider variety of considerations about the practicality and political utility of the data in addition to their precision.²

In short, the FDA could become an important influence in creating a more robust and rigorous exchange of comparative clinical information about medical products. FDA regulations already establish that “substantial clinical experience”—which is a standard that applies to the kinds of population-based, epidemiological data used to develop much of the comparative research envisioned by the new CER agency—is an appropriate standard for companies to rely on when they seek to develop and share CER. The FDA’s definition of “substantial clinical experience” can therefore, in turn, provide an ideal and consistent criterion to guide the level of substantiation needed for government agencies like Medicare to act on the results of studies issued by the new CER agency. To these ends:

- The FDA should develop a guidance document affirming that “substantial clinical experience” is an appropriate standard against which the FDA can evaluate comparative clinical studies that do not have the same design features as “adequate and well-controlled clinical trials” (which is the standard on which the FDA relies in making approval decisions). Although “substantial clinical experience” is a standard that currently exists in the FDA’s regulations governing prescription drug advertising, there is no FDA guidance explaining how the standard might be used in the CER context.
- The FDA should develop guidance on the specific issues that arise when the “substantial clinical experience” standard is applied not only to CER-based statements made by medical product companies, but also to such statements made by the proposed CER agency. Adopting a common definition would help form a clear, consistent, and clinically defensible standard for both public- and private-sector sharing of CER and for guiding regulatory and policy decisions based on the results of this research.

- The FDA should develop a guidance document interpreting the “competent and reliable scientific evidence” standard for health care economic information.³ The guidance document should establish a safe harbor enabling medical product companies to share a broad range of comparative information with sophisticated health care purchasers, such as health plans and Medicare. Ideally, the guidance would apply the “competent and reliable scientific evidence” standard not only to statements made in materials defined as “labeling” by the FDA, as the statute provides, but also to materials defined as “advertising” under the Food, Drug, and Cosmetic Act (FDCA).
- The public also needs to invest in the infrastructure for undertaking rigorous, prospective comparative studies that randomize treatment groups. We cannot rely solely on epidemiological data, and we cannot afford to commission prospective studies unless we have a more efficient process for undertaking them and more predictable standards governing when medical product companies in communications with payers, physicians, and patients can use data from those studies.

The Current FDA Paradigm for Comparative Research

The FDA standards governing medical product companies’ distribution of comparative research results are often ambiguous. They are made still murkier by the inconsistent guidance that the FDA sometimes gives sponsors in private communications, coupled with the agency’s reluctance to commit to any particular position in written guidance that could enable wider manufacturer communications of CER-type evidence held to a more applicable standard, rather than the agency’s preferred substantiation standard—“substantial evidence.”

Under current FDA regulations, companies can promote their products with claims of superiority over competing products—or over previous versions of a company’s own products—but the FDA scrutinizes these claims with particular care. The agency holds superiority (and other comparative) claims to the “substantial evidence” standard, which is the same standard of substantiation that applies to the determination of whether a new product is entitled to marketing authorization.

The increasing demand for different types of comparative information that cannot satisfy the “substantial evidence” standard has made it harder for the FDA to continue insisting on a single standard. But the relative ease of applying one rule to all manufacturer claims makes it likely that the FDA would resist proposals to issue guidance encouraging manufacturers to distribute different kinds of information (like CER) based on any alternative standard.

The FDA generally believes that superiority claims are often misleading to consumers and clinicians and that the claims themselves are seldom complete. The FDA has not issued regulations or guidance documents specifically defining what constitutes a false or misleading claim in the CER context. In the absence of this kind of general guidance, insights about the agency’s point of view come from reviewing FDA warning letters.⁴ Over the years, the FDA has issued warning and “untitled” letters to medical product manufacturers alleging that comparative claims are false, misleading, or otherwise violative. The FDA has long required “substantial evidence” to support specific superiority claims. Its evidence requirement of two adequate, well-controlled, head-to-head trials is, in many cases, difficult to enroll and very expensive—amounting to an effective ban on superiority claims.

In May 2007, for example, the FDA issued an untitled letter concerning a doctor brochure for GlaxoSmithKline’s Flonase (fluticasone propionate) nasal spray. Flonase is approved for seasonal allergic and perennial allergic rhinitis in certain patients. According to the letter, the brochure “misbranded” the drug because it made “unsubstantiated superiority claims that misleadingly imply” that Flonase was superior to a competing drug, Nasonex. The FDA found the presentations in the Flonase brochure misleading because the data did not constitute “substantial evidence” for two reasons.

First, the design of the study that compared Flonase to Nasonex did not contemplate a head-to-head comparison. So, the FDA said it is hard to rely on the data that were ultimately generated because the “study protocol” did not anticipate randomizing subjects to the two drugs. The FDA also said that the study was not replicated by a second study. Superiority claims, the FDA said, should be based on comparisons of the two drug products in “two adequate, well-designed, head-to-head clinical trials.” Almost needless to say, the FDA letter reflects a more exacting standard for making comparisons than what will be offered by the research that is envisioned in legislative proposals for a new CER agency.⁵ Few, if any,

of the studies developed by a new CER agency would meet this existing FDA standard.

“Substantial Clinical Experience” as a Common Standard

Indeed, proponents of CER acknowledge that it will be rare to have multiple prospective, randomized clinical trials comparing two products. We will rely instead on epidemiological data, reviews of databases, and registries (in which patients are not randomly assigned to the different treatment groups). This kind of practical evidence forms the core of what the new CER agency intends to pursue. To accommodate this, we need a different standard of substantiation more appropriately matched to this kind of evidence.

Adopting a uniform definition for substantiation that is consistent across the FDA and the proposed CER agency would allow the FDA to play its traditional role in defining standards for making recommendations on the basis of research.

The FDA’s “substantial clinical experience” standard encompasses a broader range of clinical data than the “substantial evidence” standard. “Substantial clinical experience” includes epidemiological data and registry data and is therefore an appropriate standard for judging the veracity of CER. Moreover, it is a standard that can be shared across the FDA and a CER agency, providing a uniform principle for weighing evidence.

Although some individuals inside the FDA’s Division of Drug Marketing, Advertising, and Communications and elsewhere in the agency have been reported to tell some sponsors privately that the FDA does not strictly respect “substantial clinical experience” as an appropriate standard for substantiating comparative claims (creating ambiguity and uncertainty), the regulations governing prescription drug advertisements do set forth this standard. More important than whether there may be arguments supporting the view that “substantial clinical experience” does not apply to most drugs for which CER would be conducted is the fact that the FDA has gone so far as to promulgate a regulation defining “substantial

clinical experience.” There are other evidentiary standards in the FDA’s regulatory scheme, but none both fits the type of evidence likely to be generated by the proposed CER agency and already has a regulatory definition.

Public health is best served when the FDA issues definitive guidance that provides clear definitions and boundaries.⁶ In this case, with so much policy interest in establishing a formal framework for CER, the FDA could play an important public health role by more clearly establishing how a definition of “substantial clinical experience” could assist in substantiating comparative research and deciding when sponsors should share this information with consumers and doctors and when government agencies should use this information for clinical and policy decision-making. Adopting a uniform definition for substantiation that is consistent across the FDA and the proposed CER agency would also allow the FDA to play its traditional role in defining standards for making recommendations on the basis of research.

Under current regulations, “substantial clinical experience” consists of experience “adequately documented in medical literature or by other data on the basis of which it can fairly and responsibly be concluded by qualified experts that the drug is safe and effective” for the claimed uses.⁷ The standard originally addressed the level of support a manufacturer had to provide for a claim about a drug that had not been evaluated by the FDA through the new drug application (NDA) process, which focuses on efficacy data from adequate and well-controlled clinical trials. But the language of the regulations is broad enough to apply to the substantiation of claims about any drug, including those that have been approved by the FDA.

There is credible legal support for an approach to CER that would rely on this interpretation of “substantial clinical experience.” The very cornerstone of the FDA’s legal authority to review the efficacy of new drugs recognizes the importance of information derived from “clinical experience.” Section 505 of the FDCA not only makes clear that data developed outside of adequate and well-controlled clinical trials is vitally important to the FDA’s assessment of the safety of a new drug, but also emphasizes the role of such data in the FDA’s continuing assessment of safety after approval.

Indeed, section 505 contains the phrase “clinical experience” in a number of places. For example, section 505(k)(1) requires the holder of an NDA to “establish and maintain such records, and make such reports . . . of data relating to clinical experience and other data or

information, received or otherwise obtained by such applicant with respect to such drug” that would enable the FDA to determine whether to invoke the withdrawal of approval provision in section 505(e).⁸

Pursuant to section 505(k)(1), the FDA has also issued a regulation requiring manufacturers to include in the NDA annual reports that they file with the agency “reports of clinical experience pertinent to safety (for example, epidemiological studies or analyses of experience in a monitored series of patients).”⁹ According to this provision, reports of clinical experience include epidemiological analyses and analyses derived from observation of patients. This is precisely the kind of clinical information that forms the basis of studies that will be developed by a new federal CER agency.

The only way that CER would contain costs would be if Congress were to limit access to some high-cost treatments on the basis of the data.

The FDA’s “substantial evidence” regulations explain what “substantial clinical experience” does *not* mean. The signal characteristic of “substantial evidence,” according to the FDA, is that it is generated from “adequate and well-controlled clinical trials.”¹⁰ A reasonable interpretation of “substantial clinical experience” is that it is not necessarily derived from such trials.

Although few companies, if any, have energetically embraced the “substantial clinical experience” standard due to the lack of a clear pathway set forth by the FDA in publicly available guidance, the regulations clearly contemplate the use of this alternative substantiation standard in the context of promotional claims. The FDA should make clear in guidance that any manufacturer is permitted to provide comparative effectiveness information directly to physicians and even to patients if the information satisfied the standard and was truthful and nonmisleading. This is consistent not only with the FDCA itself, but also with other FDA regulations that expressly provide for manufacturers to furnish price, in addition to benefit and risk, information directly to consumers in promotional communications.

In short, there are enough provisions in the existing FDA regulatory framework to guide the development of a clear, thorough, and consistent definition of “substantial

clinical experience” that could be used as a standard for substantiating claims based on CER data across both the FDA and a new CER agency. This would create a uniform standard and level playing field for the sharing of this kind of information.

The Argument for a Uniform Standard

It is unrealistic to believe that, over the long run, this federally generated CER will not be used by other federal agencies such as Medicare to inform decisions about access to and pricing of medical products. Notwithstanding efforts by political proponents of a CER agency to assuage critics by punting on this fundamental question, logic reveals the eventual outcome.

Budget estimates also betray the true intentions of CER proponents. Office of Management and Budget director Peter R. Orszag has cited the Obama administration’s support of CER as one of the principal ways to control health care spending.¹¹ Yet, on three separate occasions, the Congressional Budget Office (CBO) has said that there would not be any savings derived from the new federal CER effort. The first found that one CER proposal (H.R. 3162) would increase federal spending by \$1.1 billion over ten years.¹² The second also found that government funding for CER only increases the federal deficit for the first ten years.¹³ Then—CBO director Orszag issued a third estimate in testimony before Congress on June 12, 2007, when he stated that CER might not yield direct savings for at least ten years.¹⁴ The only way that CER would contain costs would be if Congress were to limit access to some high-cost treatments on the basis of the data. This is an unfashionable truth policymakers will not publicly acknowledge. As a result, under CBO rules, policymakers cannot claim savings in official budget estimates, even if cost containment—and rationing—remains their goal. The administration’s claims about savings, however, reveal the underlying intentions.

It is unlikely that we will receive more political candor about these goals. But it is reasonable to expect proponents to define the clinical and legal standards on which information could be shared and policy recommendations made based on the results of CER. It is hard to envision that the new CER agency will not be issuing press releases, policy papers, and other material that interprets and trumpets the results of its research. What will be the scientific basis for how the new agency interprets its findings and issues its recommendations?

Absent some agreement about the standards for substantiating CER, the proximate result is likely to be a plethora of debatable and uninterrupted research developed by a network of health systems researchers with a vested financial and intellectual interest in the promotion of this research. There will be political pressure to overinterpret the findings in order to justify the expenditure. There is also risk that the FDA evidentiary standards will become less relevant. The end result would be confusing drug information that does not correspond to clinical practice, let alone the way government health plans make reimbursement decisions.

Moreover, if payers were to begin to make reimbursement decisions based on data standards less rigorous than the FDA’s requirements, it would diminish incentives for sponsors to seek supplemental approvals for already-marketed products. The additional claims would not have economic value if government payers were willing to base reimbursement decisions on less rigorous and easier-to-execute research studies.

One of the benefits of existing regimes, such as the statistical standard that p must equal 0.05 in order for the result from a randomized clinical trial to be deemed “significant,” is that these standards provide a framework for making objective medical decisions. It provides a clear line for actionable results. It may not be an optimal standard for certain kinds of medical decision-making, and some argue that the widespread application of $p=0.05$ provides an unnecessarily binary basis for making decisions. But it does reveal the benefits of having some objective criteria to substantiate how decisions are made based on new information.

When it comes to CER, no similar standards exist. Truth will be in the eye of the researcher—or the agency with the most incentive and clout to shape the interpretation of a result. That exposes the fundamental risk in this scheme. Without agreement around the objective standards for substantiating CER, there will be an over-interpretation of the results by politically conflicted government agencies. It is true that standards to govern the substantiation of less rigorous epidemiological data—the kind of information generated by a new CER agency—will never have sharp boundaries like $p=0.05$. Nor is it desirable to create a grading system, since inherent in such a process is subjective interpretation that masks uncertainties. Instead, the standards embedded in FDA regulations provide consistent and applicable criteria for weighing this kind of clinical evidence. “Substantial clinical experience” provides a reasonable standard for

decision-making across the FDA, Medicare, and the new CER agency.

Based on the text of the FDCA and of FDA regulations, “substantial clinical experience” can be interpreted to mean experience in a monitored series of relatively heterogeneous patients who are administered a drug for treatment purposes outside the clinical trial context.¹⁵ To qualify as “substantial,” the monitoring must be conducted in a manner that helps assure the data generated from the patients are sufficiently reliable. By operation of the definition of “substantial clinical experience,” the information from the observed patients must be “adequately documented,” and it must be appropriate for “qualified experts” to conclude from the information that the drug has the attributes it is claimed to have.¹⁶

To adopt this definition as the standard used to substantiate CER across federal agencies, the FDA should first and foremost issue guidance that specifically addresses comparative efficacy claims about approved products based on data from CER-type studies. This guidance should explain clearly how manufacturers could use “substantial clinical experience” as an appropriate standard of substantiation for CER claims in promotional material.¹⁷ Once this standard is memorialized in guidance, the new CER agency can bridge easily to this definition as a basis for recommending when its resulting research reaches an adequate level of rigor and substantiation for purposes of being relied upon by other parties.

The FDCA also contains the “competent and reliable scientific evidence” standard, in section 502(a), which was amended by the Food and Drug Administration Modernization Act (FDAMA) in 1997 to state specifically that a manufacturer can provide “health care economic information” to selected managed-care organizations without risking that the materials the manufacturer uses to convey this information would be regarded by the agency as false or misleading. This standard is limited by the lack of FDA interpretive guidance and by the language of the provision itself, which refers to health care economic information provided to certain managed-care organizations and does not, therefore, currently provide a safe harbor for manufacturer statements about the comparative cost-effectiveness of medical products directed to health care practitioners or patients.¹⁸ In guidance, the FDA could affirm that “competent and reliable scientific evidence,” which is a standard used in a provision of the FDCA expressly allowing manufacturers to provide written materials containing CER information to managed-care organizations and public payers, includes the kinds

of study designs likely to be supported by a new CER agency, specifically epidemiological databases and non-randomized series.¹⁹

If the FDA were to provide guidance, the “competent and reliable scientific evidence” standard could provide a more limited but still important pathway for manufacturers to provide CER-type information to sophisticated managed-care entities and health care purchasers. Such guidance would establish a safe harbor for sponsor-generated CER as being appropriate for sharing with health plans, including Medicare.²⁰ Although a pathway for sharing this information is defined in the FDCA as amended by FDAMA, ambiguity remains around the permissibility of this sort of information exchange, and some sponsors are reluctant to share CER with payers.

While there is a place in medical research
for studies based on registries and
epidemiological data, we cannot rely on
these less rigorous and less precise clinical
trial constructs alone, especially to answer
difficult clinical questions.

The FDA had started addressing pharmacoeconomic claims’ substantiation issues before Congress enacted the “competent and reliable” standard in 1997, but there has been since then no meaningful guidance from the FDA regarding the meaning of that standard. One of the factors creating a disincentive for sponsors to develop information about the cost-effectiveness of various treatments is that their ability to share this information remains needlessly murky, owing to FDA reluctance to delineate clear guidelines.²¹

Finally, the kinds of clinical trial constructs being contemplated by a new CER agency (epidemiological databases, registries, nonrandomized simple large trials) are unreliable, especially when it comes to examining differential responses to treatments within smaller populations of patients. For many drug regimens, for example, the operative question is not whether one drug is best for everyone, but who should take which drug and under what circumstances. There are real clinical differences not only between similar patients, but also when

a drug gets started and stopped. Epidemiological studies of databases will not detect the clinical signals that provide answers to these questions. More rigorous trials are needed, often prospective trials that randomize similar patients to different treatment groups.

The question is how to develop more of this kind of research without spending tens of millions of dollars. The reason that less rigorous clinical trial constructs have become the sine qua non of a new CER agency is that they can be achieved with modest funding. Undertaking more rigorous clinical studies (which are also more definitive) can cost dollars on the pennies spent on currently proposed CER study designs. A single large, prospective, randomized trial can cost \$100 million or more. While there is a place in medical research for studies based on registries and epidemiological data, we cannot rely on these less rigorous and less precise clinical trial constructs alone, especially to answer difficult clinical questions.

The cancer cooperative groups maintained by the National Cancer Institute (NCI) have developed a model and a good track record for how we can build a clinical trial infrastructure in other therapeutic areas to enable more rigorous comparative research to be undertaken more efficiently. U.S. investments in CER should also include the creation of a clinical trial infrastructure that enables this kind of more rigorous research. We need to invest in our capacity to conduct rigorous CER. Most of the studies that the NCI sponsors through its network are, at their core, comparative trials, since they are mostly comparing current regimens to regimens in which a new agent is added to standard care. NCI has also had success at engaging more community physicians and academic researchers in the clinical trial process. This is another important goal for a new network created for undertaking rigorous CER, since many of the clinical questions that we want to answer involve decisions made in the community rather than at academic hospitals.

There are models for how rigorous clinical research can be conducted more efficiently using web-based data entry and centralized institutional review boards. These and other approaches can reduce the paperwork and compliance costs of enrolling subjects in clinical trials. Ultimately, the best way to translate the findings from the research process is to enlist community physicians in the conduct of the studies. There is no reason a new CER agency needs to rely solely or largely on less rigorous data constructs like databases and registries.

Conclusion

We should continue to press proponents of a CER agency on the sustainability of their own assumptions. They insist that the resulting CER data will save the health care system hundreds of millions of dollars, but they deny that Medicare reimbursement decisions will eventually be tied to the results. Most reasonable people will understand that these two objectives are incongruous. Most reasonable people know that CER will eventually be used to tweak coverage decisions.

But as the political effort to frame this prospective new agency's mandate takes shape, we also need to engage CER proponents directly in a serious discussion about the standards that should be used for making policy decisions based on the results of the research they espouse. To these ends, "substantial clinical experience" provides a good starting point for decisions about the substantiation needed for sponsors to share information from their own CER—and about the criteria government health programs like Medicare should be held to when the resulting information is ultimately used as support for their reimbursement policy decisions.

Notes

1. Scott Gottlieb, "Congress Wants to Restrict Drug Access," *Wall Street Journal*, January 20, 2009, available at www.aei.org/article/29219.
2. The government should have to follow a binary system of substantiation, rather than a graduated one, but it should recognize that companies, unlike the government, have First Amendment rights to make claims based on any level of substantiation, as long as they are truthful and nonmisleading.
3. See *Food, Drug, and Cosmetic Act*, section 502(a), as amended by the *Food and Drug Administration Modernization Act of 1997*.
4. Jeffrey K. Shapiro, "Comparative Claims: Legally Permissible, but Proceed with Care," *Medical Device & Diagnostic Industry* (September 2004), available at www.devicelink.com/mddi/archive/04/09/014.html (accessed June 4, 2009). Technically, statements in warnings and untitled letters do no more than to allege a legal violation; they do not necessarily reflect any official Food and Drug Administration (FDA) policy and do not bind the FDA to the views expressed in the letters. (*Code of Federal Regulations* 21, § 10.85[k].)
5. Congressional Budget Office (CBO), *Budget Options*, vol. 1: *Health Care* (Washington, DC: CBO, December 2008), available at www.cbo.gov/doc.cfm?index=9925 (accessed June 4, 2009).

6. FDA, Center for Drug Evaluation and Research, Division of Drug Marketing, Advertising, and Communications, "Promotional Material Review Process."

7. *Code of Federal Regulations* 21, § 202.1(e)(4)(ii)(c).

8. *U.S. Code* 21, § 355(k)(1).

9. *Code of Federal Regulations* 21, § 314.81(b)(2)(vi).

10. *Code of Federal Regulations* 21, § 202.1(e)(4)(ii)(b).

11. Gregory Twachtman, "Comparative Effectiveness in U.S. Will Not Be NICE, OMB's Orszag Says," *The Pink Sheet Daily*, March 5, 2009.

12. CBO, "Research on the Comparative Effectiveness of Medical Treatments: Issues and Options for an Expanded Federal Role" (CBO paper 2975, Washington, DC, December 2007), available at www.cbo.gov/doc.cfm?index=8891 (accessed May 17, 2009).

13. Gregory Twachtman, "CBO Holds Line on Comparative Effectiveness' Lack of Initial Savings," *The Pink Sheet Daily*, January 5, 2009.

14. Gregory Twachtman, "Savings from Expanded Comparative Effectiveness Could Be Decade Away—CBO," *The Pink Sheet*, June 18, 2007.

15. *Code of Federal Regulations* 21, § 202.1(e)(6)(ii).

16. *Code of Federal Regulations* 21, § 202.1(e)(4)(ii)(c).

17. FDA warning and untitled letters do not contain any commentary on the regulatory definition of "substantial clinical experience," and the FDA has not provided relevant guidance.

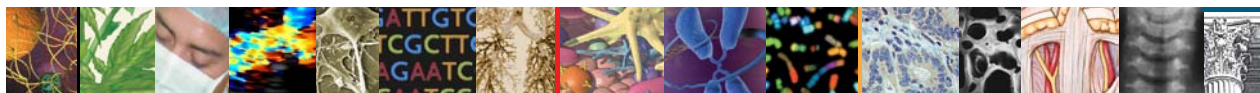
18. Other existing FDA regulations and policies—such as its

policy on responses to unsolicited requests, the regulations allowing promotion based on price (but not cost-effectiveness), and the scientific exchange provision of the investigational new drug regulations—are similarly limited.

19. *Food and Drug Administration Modernization Act of 1997*, section 114, codified at *U.S. Code* 21, § 352(a).

20. A similar approach, using enforcement discretion to create a safe harbor for distribution of certain information, is the basis for at least one recent guidance document. See FDA, "Guidance for Industry: Good Reprint Practices for the Distribution of Medical Journal Articles and Medical or Scientific Reference Publications on Unapproved New Uses of Approved Drugs and Approved or Cleared Medical Devices," January 2009, available at www.fda.gov/oc/op/goodreprint.html (accessed June 4, 2009).

21. Kate A. Stewart and Peter J. Neumann, "FDA Actions against Misleading or Unsubstantiated Economic and Quality-of-Life Promotional Claims: An Analysis of Warning Letters and Notices of Violation," *Value in Health* 5, no. 5 (2002): 390–97; P. J. Neumann, K. Claxton, and M. C. Weinstein, "The FDA's Regulation of Health Economic Information," *Health Affairs* 19, no. 5 (2000): 129–37; and Monica Schultz and Chris Lyle, "Health Economic Analysis: What Needs to Be Considered When Designing a Clinical Trial for Medical Technology," *Applied Clinical Trials* (May 2009), available at <http://appliedclinicaltrialsonline.findpharma.com/appliedclinicaltrials/Trial+Design+Articles/Health-Economic-Analysis/ArticleStandard/Article/detail/597463> (accessed June 4, 2009).



The NEW ENGLAND JOURNAL of MEDICINE

Perspective
DECEMBER 23, 2010

What Is Value in Health Care?

Michael E. Porter, Ph.D.

In any field, improving performance and accountability depends on having a shared goal that unites the interests and activities of all stakeholders. In health care, however, stakeholders have

myriad, often conflicting goals, including access to services, profitability, high quality, cost containment, safety, convenience, patient-centeredness, and satisfaction. Lack of clarity about goals has led to divergent approaches, gaming of the system, and slow progress in performance improvement.

Achieving high value for patients must become the overarching goal of health care delivery, with value defined as the health outcomes achieved per dollar spent.¹ This goal is what matters for patients and unites the interests of all actors in the system. If value improves, patients, payers, providers, and suppliers can all benefit while the economic sustainability of the health care system increases.

Value — neither an abstract ideal nor a code word for cost reduction — should define the framework for performance improvement in health care. Rigorous, disciplined measurement and improvement of value is the best way to drive system progress. Yet value in health care remains largely unmeasured and misunderstood.

Value should always be defined around the customer, and in a well-functioning health care system, the creation of value for patients should determine the rewards for all other actors in the system. Since value depends on results, not inputs, value in health care is measured by the outcomes achieved, not the volume of services delivered, and shifting focus from volume to

value is a central challenge. Nor is value measured by the process of care used; process measurement and improvement are important tactics but are no substitutes for measuring outcomes and costs.

Since value is defined as outcomes relative to costs, it encompasses efficiency. Cost reduction without regard to the outcomes achieved is dangerous and self-defeating, leading to false “savings” and potentially limiting effective care.

Outcomes, the numerator of the value equation, are inherently condition-specific and multidimensional. For any medical condition, no single outcome captures the results of care. Cost, the equation’s denominator, refers to the total costs of the full cycle of care for the patient’s medical condition, not the cost of individual services. To reduce cost, the best approach is often to spend more on some services to reduce the need for others.

Health care delivery involves numerous organizational units, ranging from hospitals to physicians' practices to units providing single services, but none of these reflect the boundaries within which value is truly created. The proper unit for measuring value should encompass all services or activities that jointly determine success in meeting a set of patient needs. These needs are determined by the patient's medical condition, defined as an interrelated set of medical circumstances that are best addressed in an integrated way. The definition of a medical condition includes the most common associated conditions — meaning that care for diabetes, for example, must integrate care for conditions such as hypertension, renal disease, retinal disease, and vascular disease and that value should be measured for everything included in that care.¹

For primary and preventive care, value should be measured for defined patient groups with similar needs. Patient populations requiring different bundles of primary and preventive care services might include, for example, healthy children, healthy adults, patients with a single chronic disease, frail elderly people, and patients with multiple chronic conditions.

Care for a medical condition (or a patient population) usually involves multiple specialties and numerous interventions. Value for the patient is created by providers' combined efforts over the full cycle of care. The benefits of any one intervention for ultimate outcomes will depend on the effectiveness of other interventions throughout the care cycle.

Accountability for value should be shared among the providers involved. Thus, rather than “fo-

cused factories” concentrating on narrow groups of interventions, we need integrated practice units that are accountable for the total care for a medical condition and its complications.

Because care activities are interdependent, value for patients is often revealed only over time and is manifested in longer-term outcomes such as sustainable recovery, need for ongoing interventions, or occurrences of treatment-induced illnesses.² The only way to accurately measure value, then, is to track patient outcomes and costs longitudinally.

For patients with multiple medical conditions, value should be measured for each condition, with the presence of the other conditions used for risk adjustment. This approach allows for relevant comparisons among patients' results, including comparisons of providers' ability to care for patients with complex conditions.

The current organizational structure and information systems of health care delivery make it challenging to measure (and deliver) value. Thus, most providers fail to do so. Providers tend to measure only what they directly control in a particular intervention and what is easily measured, rather than what matters for outcomes. For example, current measures cover a single department (too narrow to be relevant to patients) or outcomes for a whole hospital, such as infection rates (too broad to be relevant to patients). Or they measure what is billed, even though current reimbursement practices are misaligned with value. Similarly, costs are measured for departments or billing units rather than for the full care cycle over which value is determined. Faulty organizational

structure also helps explain why physicians fail to accept joint responsibility for outcomes, blaming lack of control over “outside” actors involved in care (even those in the same hospital) and patients' compliance.

The concept of quality has itself become a source of confusion. In practice, quality usually means adherence to evidence-based guidelines, and quality measurement focuses overwhelmingly on care processes. For example, of the 78 Healthcare Effectiveness Data and Information Set (HEDIS) measures for 2010, the most widely used quality-measurement system, all but 5 are clearly process measures, and none are true outcomes.³ Process measurement, though a useful internal strategy for health care institutions, is not a substitute for measuring outcomes. In any complex system, attempting to control behavior without measuring results will limit progress to incremental improvement. There is no substitute for measuring actual outcomes, whose principal purpose is not comparing providers but enabling innovations in care. Without such a feedback loop, providers lack the requisite information for learning and improving. (Further details about measuring value are contained in a framework paper, “Value in Health Care,” in Supplementary Appendix 1, available with the full text of this article at NEJM.org.)

Measuring, reporting, and comparing outcomes are perhaps the most important steps toward rapidly improving outcomes and making good choices about reducing costs.⁴ Systematic, rigorous outcome measurement remains rare, but a growing number of examples of comprehensive outcome measurement provide evidence of its feasibility and impact.

Determining the group of relevant outcomes to measure for any medical condition (or patient population in the context of primary care) should follow several principles. Outcomes should include the health circumstances most relevant to patients. They should cover both near-term and longer-term health, addressing a period long enough to encompass the ultimate results of care. And outcome measurement should include sufficient measurement of risk factors or initial conditions to allow for risk adjustment.

For any condition or population, multiple outcomes collectively define success. The complexity of medicine means that competing outcomes (e.g., near-term safety versus long-term functionality) must often be weighed against each other.

The outcomes for any medical condition can be arrayed in a three-tiered hierarchy (see Figure 1), in which the top tier is generally the most important and lower-tier outcomes involve a progression of results contingent on success at the higher tiers. Each tier of the framework contains two levels, each involving one or more distinct outcome dimensions. For each dimension, success is measured with the use of one or more specific metrics.

Tier 1 is the health status that is achieved or, for patients with some degenerative conditions, retained. The first level, survival, is of overriding importance to most patients and can be measured over various periods appropriate to the medical condition; for cancer, 1-year and 5-year survival are common metrics. Maximizing the duration of survival may not be the most important outcome, however, especially for older patients who may weight other outcomes more

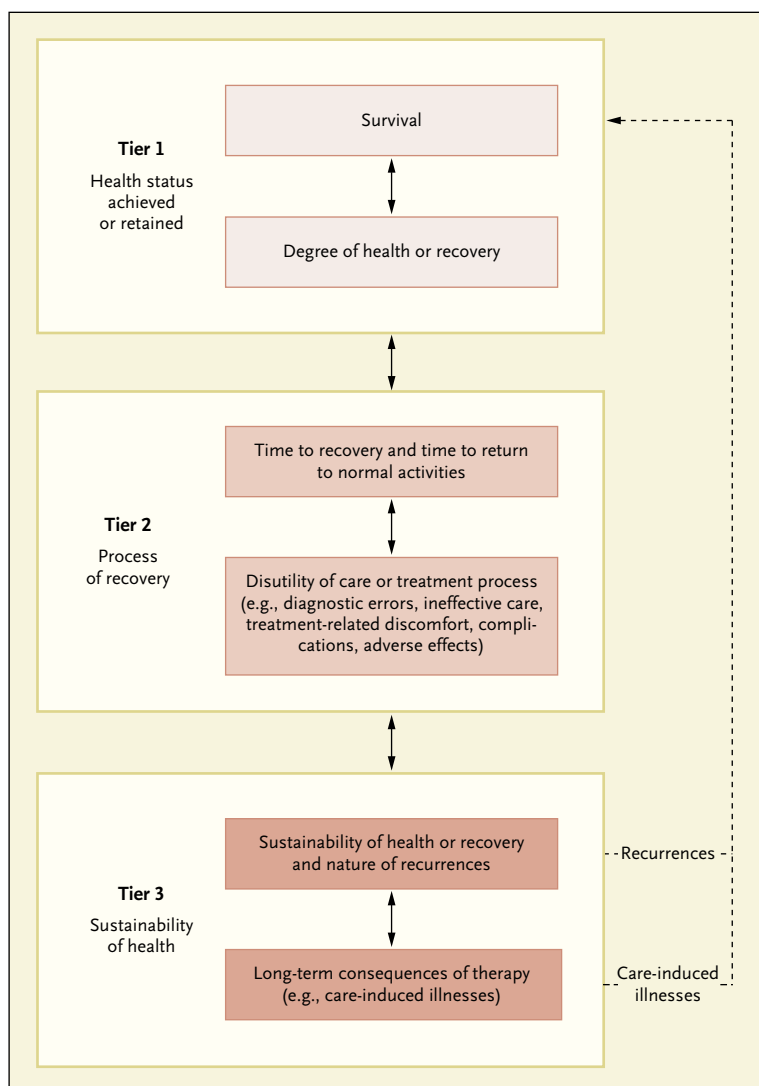


Figure 1. The Outcome Measures Hierarchy.

heavily. The second level in Tier 1 is the degree of health or recovery achieved or retained at the peak or steady state, which normally includes dimensions such as freedom from disease and relevant aspects of functional status.

Tier 2 outcomes are related to the recovery process. The first level is the time required to achieve recovery and return to normal or best attainable function, which can be divided into the time needed to complete various phases of care. Cycle time is a critical outcome for patients — not a

secondary process measure, as some believe. Delays in diagnosis or formulation of treatment plans can cause unnecessary anxiety. Reducing the cycle time (e.g., time to reperfusion after myocardial infarction) can improve functionality and reduce complications. The second level in Tier 2 is the disutility of the care or treatment process in terms of discomfort, retreatment, short-term complications, and errors and their consequences.

Tier 3 is the sustainability of health. The first level is recur-

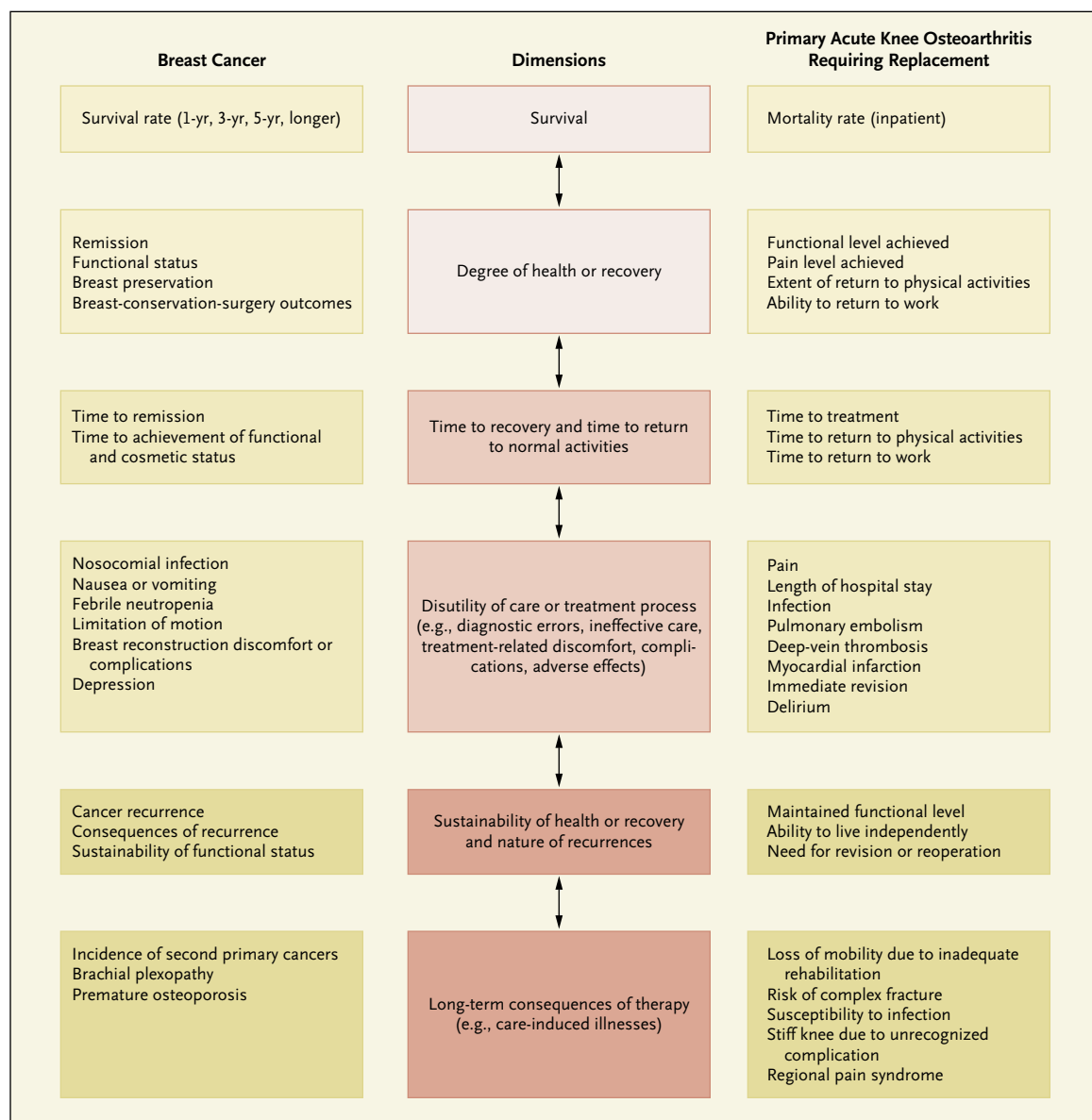


Figure 2. Outcome Hierarchies for Breast Cancer and Knee Osteoarthritis.

rences of the original disease or longer-term complications. The second level captures new health problems created as a consequence of treatment. When recurrences or new illnesses occur, all outcomes must be remeasured.

With some conditions, such as metastatic cancers, providers may have a limited effect on survival or other Tier 1 outcomes, but they can differentiate themselves in Tiers 2 and 3 by making care

more timely, reducing discomfort, and minimizing recurrence.

Each medical condition (or population of primary care patients) will have its own outcome measures. Measurement efforts should begin with at least one outcome dimension at each tier, and ideally one at each level. As experience and available data infrastructure grow, the number of dimensions (and measures) can be expanded.

Improving one outcome dimension can benefit others. For example, more timely treatment can improve recovery. However, measurement can also make explicit the tradeoffs among outcome dimensions. For example, achieving more complete recovery may require more arduous treatment or confer a higher risk of complications. Mapping these tradeoffs, and seeking ways to reduce them, is an essen-

tial part of the care-innovation process.

Figure 2 illustrates possible outcome dimensions for breast cancer and acute knee osteoarthritis requiring knee replacement. Most current measurement efforts fail to capture such comprehensive sets of outcomes, which are needed to fully describe patients' results. No organization I know of systematically measures the entire outcome hierarchy for the medical conditions for which it provides services, though some are making good progress. (Further details, including risk adjustment, are addressed in a framework paper, "Measuring Health Outcomes," in Supplementary Appendix 2, available at NEJM.org.)

The most important users of outcome measurement are providers, for whom comprehensive measurement can lead to substantial improvement.⁵ Outcomes need not be reported publicly to benefit patients and providers, and public reporting must be phased in carefully enough to win providers' confidence. Progression to public reporting, however, will accelerate innovation by motivating providers to improve relative to their peers and permitting all stakeholders to benefit fully from outcome information.

Current cost-measurement approaches have also obscured value in health care and led to cost-containment efforts that are incremental, ineffective, and some-

times even counterproductive. Today, health care organizations measure and accumulate costs around departments, physician specialties, discrete service areas, and line items such as drugs and supplies — a reflection of the organization and financing of care. Costs, like outcomes, should instead be measured around the patient. Measuring the total costs over a patient's entire care cycle and weighing them against outcomes will enable truly structural cost reduction, through steps such as reallocation of spending among types of services, elimination of non-value-adding services, better use of capacity, shortening of cycle time, provision of services in the appropriate settings, and so on.

Much of the total cost of caring for a patient involves shared resources, such as physicians, staff, facilities, and equipment. To measure true costs, shared resource costs must be attributed to individual patients on the basis of actual resource use for their care, not averages. The large cost differences among medical conditions, and among patients with the same medical condition, reveal additional opportunities for cost reduction. (Further aspects of cost measurement and reduction are discussed in the framework paper "Value in Health Care.")

The failure to prioritize value improvement in health care delivery and to measure value has

slowed innovation, led to ill-advised cost containment, and encouraged micromanagement of physicians' practices, which imposes substantial costs of its own. Measuring value will also permit reform of the reimbursement system so that it rewards value by providing bundled payments covering the full care cycle or, for chronic conditions, covering periods of a year or more. Aligning reimbursement with value in this way rewards providers for efficiency in achieving good outcomes while creating accountability for substandard care.

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

From Harvard Business School, Boston.

This article (10.1056/NEJMp1011024) was published on December 8, 2010, at NEJM.org.

1. Porter ME, Teisberg EO. Redefining health care: creating value-based competition on results. Boston: Harvard Business School Press, 2006.
2. Institute of Medicine. Performance measurement: accelerating improvement. Washington, DC: National Academies Press, 2006.
3. National Committee for Quality Assurance (NCQA). HEDIS and quality measurement: technical resources. NCQA Web site. (<http://www.ncqa.org/tabid/1044/Default.aspx>.)
4. Porter ME. Defining and introducing value in health care. In: Evidence-based medicine and the changing nature of health care: 2007 IOM annual meeting summary. Washington, DC: Institute of Medicine, 2008:161-72.
5. Porter ME, Baron JF, Chacko JM, Tang RJ. The UCLA Medical Center: kidney transplantation. Harvard Business School Case 711-410. Boston: Harvard Business School Publishing, 2010.

Copyright © 2010 Massachusetts Medical Society.

Putting the Value Framework to Work

Thomas H. Lee, M.D.

“Value” is a word that has long aroused skepticism among physicians, who suspect it of being code for “cost reduction.” Nevertheless, an increas-

ing number of health care delivery organizations, including my own, now describe enhancement of value for patients as a fundamental goal and are using con-

cepts developed by Michael Porter (see pages 2477–2480, and the framework papers in Supplementary Appendixes 1 and 2 of that article) to shape their strategies.



The NEW ENGLAND JOURNAL of MEDICINE



Perspective

Putting the Value Framework to Work

Thomas H. Lee, M.D.

N Engl J Med 2010; 363:2481-2483 | [December 23, 2010](#)

Article

“Value” is a word that has long aroused skepticism among physicians, who suspect it of being code for “cost reduction.” Nevertheless, an increasing number of health care delivery organizations, including my own, now describe enhancement of value for patients as a fundamental goal and are using concepts developed by Michael Porter (see pages 2477–2480, and the framework papers in Supplementary Appendixes 1 and 2 of that article) to shape their strategies. What has changed? And what are these organizations actually doing?

Practical motivations lie behind the interest in the value framework. Rising costs and a stagnant economy pose problems with no easy solution. Budgets cannot be planned responsibly by hoping for growth in volume. As all players try to protect their incomes, nerves are fraying. Physicians are pitted against hospitals, specialists against primary care physicians, academics against the community.

In this fractious context, value is emerging as a concept — perhaps the only concept — that all stakeholders in health care embrace. Providers, patients, payers, and policymakers all support the goal of improving outcomes and doing so as efficiently as possible. No one can oppose this goal and expect long-term success, just as no one in a for-profit company can resist decisions likely to enhance long-term shareholder value. The value framework thus offers a unifying orientation for provider organizations that might otherwise be paralyzed by constituents’ fighting for bigger pieces of a shrinking pie.

So how is the concept of value being translated into reality? As is often true in medicine itself, the critical first step is measurement. Provider organizations need to capture data on the outcomes that matter to patients, as well as the costs for a patient over meaningful episodes of care. These data are essential for assessing whether value is improving.

This work is not easy, because the collection of such data has not been encouraged by the fee-for-service system and is hindered by the silos in the current organizational structure of medicine. Current information systems are designed to support clinicians in performing individual services for individual patients and to collect their reimbursement. Outcomes as important as death are not routinely recorded; functional-status outcomes (e.g., whether a patient with head-and-neck cancer can swallow or talk) are buried in free text and are not captured in analyzable form. The physicians with whom I practice get sophisticated, unblinded productivity reports each month on how many visits and relative value units we have each “produced.” But we have never received reports on how many of our patients had emergency department visits or readmissions to the hospital.

At least not yet. Those reports are coming — some of my colleagues are already getting them. When they arrive, I expect that we will point out the inadequacy of risk adjustment, which makes comparison of rates among physicians meaningless. But we will look very carefully at the lists of names of patients and wonder how the visits and readmissions could have been avoided.

When measurement is oriented toward what happened to patients instead of what services were performed, interesting challenges and opportunities arise. For example, we are realizing that we need to expand our ability to measure and manage “cycle times” — the intervals between key moments in patients’ care. Some of these intervals have obvious implications for patients’ medical outcomes, such as door-to-balloon time for patients with myocardial infarction.¹

Calculating important intervals gets difficult when care is delivered in different parts of the health care system, but the clinical implications can be enormous. If patients who present to the emergency department with a transient ischemic attack are seen

promptly by clinicians in stroke clinics, the 90-day risk of stroke falls markedly (from 9.2% to 3.2% in one study²). If patients who have been hospitalized for a high-risk condition are seen within a week after discharge, their readmission rates are substantially reduced.

Measurement of such intervals and the outcomes that they influence is in its infancy in my organization, as in most others. And as the saying goes, if you can't measure it, you can't manage it. We are finding that just the collection of such data requires organizational change and the weakening of walls between our silos.

Which brings me to the “bad news” that goes with orientation toward improvement of value. Making progress in the value framework requires real teamwork, which sometimes seems an unnatural act in health care. It means capturing data in different parts of the delivery system, which means that we all have to use the exact same terminology. And it means sharing accountability for performance. Who should be held responsible if a patient with heart failure is not seen within 7 days after discharge? The hospital? The primary care physician? The specialist?

The answer, of course, is “all of the above.” Improvement in outcomes or reduction in costs of care cannot be achieved without active cooperation among providers, which is difficult to achieve if they're all functioning as separate business units. The value framework thus makes enormous demands for cultural and organizational change among health care providers. It pushes them toward functioning as one organization focused on delivering excellent outcomes as efficiently as possible.

Which brings us to the good news: difficult though they may be, these changes feel like the right thing to do. To improve outcomes and efficiency for patients with specific conditions, providers must organize interdisciplinary teams around those conditions. In my organization, teams focused on stroke, colon cancer, diabetes, and other diagnoses are currently developing “value dashboards.” They are identifying “pause points” in patient care and defining what steps should happen routinely at those points. An example might be ensuring that palliative care consultations are offered to patients with newly diagnosed lung cancer, a strategy that was recently shown to improve both the quality of life and survival.³ Each item on these “checklists”⁴ is being chosen because of the expectation that reliable performance should lead to better outcomes, greater efficiency, or both — in other words, improvement in value.

My colleagues appreciate that the value framework is not primarily a tool for competition or comparison among providers. Porter's outcome hierarchy makes clear that there are multiple outcomes that matter for any patient condition (see “Measuring Health Outcomes” in Supplementary Appendix 2 of the article by Porter), and they all have different units of measurement. There is no useful way to weight them, add them, and divide them by a dollar figure to derive value ratios to be compared among providers. But you can detect change — one hopes in the right direction.

The goal of the value framework is to create a context for improvement, for every physician and provider group to try to be better this year than it was last year. Value can be enhanced by improving one or more outcomes without compromising others or by reducing the costs required to achieve the same levels of outcomes. The competition is with oneself. That feels like a fair fight — and worthwhile work.

In the effort to improve the value of their own care, providers look at data on the outcomes and costs associated with other providers through a different lens. These data offer the opportunity to learn. If a certain provider group has a much lower readmission rate than others, the value framework should drive the other providers to ask what that group is doing right, not worry about the adequacy of the risk adjustment.

No one should expect the value framework to be easy to implement. The measurement of outcomes and costs, the organization of clinicians into teams focused on improving care for patient populations, the evolution of a payment system that rewards providers who are more effective in improving the value of their care — these are all formidable tasks. That work is under way in my own organization and many others, but it will require many years. Indeed, it will never end.

Furthermore, a focus by providers on improving the value of care is unlikely to be sufficient to address completely the economic challenges facing health care. Almost surely, patients will be forced to bear more costs, benefits will be limited, and increases in providers' rates will be restrained in a variety of ways. But the need for such cruder strategies and the damaging effects of their unintended consequences can be reduced if providers orient themselves toward higher-value care.

Nothing can be considered guaranteed about the future for physicians and other health care providers except that there will always be patients who need care. In these uncertain times, health care providers need a path forward. The value framework provides one.

This article (10.1056/NEJMp1013111) was published on December 8, 2010, at NEJM.org.

[Disclosure forms](#) provided by the author are available with the full text of this article at NEJM.org.

Dr. Lee is an associate editor of the *Journal* and network president for Partners Healthcare System, Boston.

References

- 1 Cannon CP, Gibson CM, Lambrew CT, et al. Relationship of symptom-onset-to-balloon time and door-to-balloon time with mortality in patients undergoing angioplasty for acute myocardial infarction. *JAMA* 2000;283:2941-2947
[CrossRef](#) | [Web of Science](#) | [Medline](#)
- 2 Wasserman J, Perry J, Dowlathshahi D, et al. Stratified, urgent care for transient ischemic attack results in low stroke rates. *Stroke* 2010;41:2601-2605
[CrossRef](#) | [Web of Science](#) | [Medline](#)
- 3 Temel JS, Greer JA, Muzikansky A, et al. Early palliative care for patients with metastatic non-small-cell lung cancer. *N Engl J Med* 2010;363:733-742
[Full Text](#) | [Web of Science](#) | [Medline](#)
- 4 Gawande A. *The checklist manifesto: how to get things right*. New York: Henry Holt, 2009.

SOUNDING BOARD

Equipoise and the Dilemma of Randomized Clinical Trials

Franklin G. Miller, Ph.D., and Steven Joffe, M.D., M.P.H.

In the 1980s, bioethics scholars defined the dilemma of randomized clinical trials as a central problem in clinical research ethics.^{1,2} How can physicians offer their patients optimal medical care at the same time that their treatment is selected by chance in the context of a randomized clinical trial? The solution that gained widespread acceptance appeals to “equipoise,” which has assumed canonical status in research ethics. Physicians can ethically randomly assign patients to treatments provided that equipoise — a state of professional uncertainty about their relative therapeutic merits — exists. If equipoise exists, no participant in a randomized clinical trial is knowingly given inferior treatment.

A major problem with the equipoise solution to the dilemma is that it narrowly locates the ethical concern about randomized clinical trials within the orbit of the doctor–patient relationship. The proponents of equipoise have characterized randomized clinical trials solely as tools to guide clinicians in making decisions about optimal medical care. In particular, they have argued that randomized trials are consistent with physicians’ ethical duties to their patients if, and only if, equipoise exists.³ This “therapeutic orientation” to clinical research ignores the wider societal interest in evidence-based health policy, as reflected in regulatory decisions to approve new treatments for licensing and in health coverage decisions by national health systems and other payers.⁴ The requisite knowledge to make decisions about individual patient care, however, is not necessarily sufficient to guide the health policy decisions about licensing approval and coverage.

Clinicians make the best decisions they can for individual patients here and now, notwithstanding more or less uncertainty regarding the benefits and risks of alternative options. In contrast, regulatory authorities and health plans adopt a population perspective. In the United States, the Food and Drug Administration (FDA)

is charged with determining that new drugs are safe and effective for defined groups of patients before they are approved. Public and private insurers make decisions about health coverage on the basis of “medical necessity,” which generally rests on convincing evidence that a treatment offers a net health benefit for a defined patient population.

WHY APPEAL TO EQUIPOISE IS MISGUIDED

Regardless of whether equipoise, as it is traditionally understood, works to resolve the dilemma of randomized clinical trials,⁵ it fails to offer sound ethical guidance regarding the appropriateness of randomized clinical trials as tools to generate the knowledge needed for drug approval or coverage decisions. Five considerations that militate against appeal to equipoise as the arbiter of the ethical legitimacy of randomized trials to evaluate new treatments, even for life-threatening or highly debilitating conditions, are the following: the imprecision in defining the concept of equipoise, the reliance on expert opinion, the limitations of determining efficacy on the basis of surrogate outcomes, the high costs of new treatments, and the tendency toward premature termination of randomized clinical trials.

The reigning conception of equipoise is known as “clinical equipoise.” According to Freedman’s classical formulation, clinical equipoise exists when “there is no consensus within the expert clinical community about the comparative merits of the alternatives to be tested.”³ Despite widespread endorsement of equipoise as an ethically necessary condition for randomized clinical trials, its proponents have not clarified how to determine when it exists. Assuming that the relevant expert community can be identified, what is the minimal proportion of members who must favor treatment A over treatment B as an appro-

appropriate therapy for patients with a given medical condition, thus justifying a randomized, clinical trial comparing the two? No authoritative answer has been provided. An approximately 50–50 split in expert opinion, which would best reflect the underlying idea of (collective) indifference reflected in the term “equipoise,” is unlikely. However, if only 1% of expert clinicians favor treatment A, it is difficult to see how the community is in equipoise. Where, between these extremes, do the boundaries of equipoise lie? Furthermore, however the presence of equipoise might be specified, systematic data are rarely available or developed to define the degree of consensus within the expert community to guide decisions about commencing or designing randomized clinical trials. Thus, operationally, equipoise fails as a guide to conduct.

More importantly, it is ironic that under the equipoise standard the permissibility of undertaking randomized clinical trials, which are intended to provide the most rigorous basis for clinical evidence, rests on mere expert opinion about the relative value of treatment options. Overwhelming support among experts for treatment A over treatment B, signifying a lack of equipoise, may or may not be grounded in sufficient evidence to guide policy, or even individual treatment, decisions. The well-known fallibility of expert opinion in support of the therapeutic value of treatments, without evidence from well-designed randomized clinical trials, is reflected in notable examples of widely used treatments that were subsequently proved to be ineffective or harmful. These treatments include antiarrhythmia drugs that were adopted on the basis of surrogate outcomes and proved to increase mortality as compared with placebo controls⁶; high-dose chemotherapy with bone marrow transplantation for metastatic breast cancer, which produced high rates of response in phase 2 trials but proved no more effective and more toxic than standard chemotherapy⁷; arthroscopic surgery for osteoarthritis of the knee, which was found to be no better than a sham intervention in relieving pain⁸; and hormone-replacement therapy, which was shown to lack benefit in promoting cardiovascular health and to be associated with multiple serious adverse outcomes.^{9,10} Moreover, the perception among clinicians that trial enrollment conflicted with equipoise probably delayed the recruitment of patients for some

of these and other important randomized clinical trials.¹¹

In the case of new treatments, expert opinion often rests on data from early-phase trials, which typically evaluate the agent’s effect on surrogate end points. The limitations of drawing conclusions about efficacy on the basis of surrogate end points deserve emphasis.^{12,13} In the case of cancer treatments, single-group phase 1 or 2 trials may produce valid evidence of tumor response; however, the causal connection between this surrogate outcome and the clinical outcomes of improved survival and enhanced quality of life over time is open to question.^{14–18} When a lack of equipoise relating to a promising new treatment derives from such response data, it generally represents a weak consideration against conducting a rigorous randomized clinical trial to evaluate definitive measures of clinical benefit.

The high cost of new treatments for life-threatening or highly debilitating conditions reinforces the importance of rigorous evaluation of therapeutic value. Whether or not cost considerations are considered relevant to decisions about regulatory approval, decisions about health coverage should reflect judgments of cost-effectiveness. Failure to honestly face the challenge of treatments that provide insufficient therapeutic value to justify their expense is a principal reason for the burgeoning cost of health care in the United States — a level of spending far in excess of other countries, without commensurate benefits in terms of improved health outcomes. Even if decisions about approval and coverage are made without any explicit consideration of cost-effectiveness, high cost ought to be relevant to the assessment process. When treatments are likely to be very expensive, and their clinical benefits are uncertain based on current knowledge, it becomes all the more important to develop sufficiently rigorous evidence about their risks and benefits. However, the traditional understanding of the dilemma of randomized clinical trials and equipoise makes no reference to the costs of treatments, owing to the exclusive ethical focus on decisions about patient care in light of current knowledge. When the specter of cost receives due consideration, randomized clinical trials may be considered ethical to generate the rigorous knowledge needed to guide health policy decisions despite a lack of equipoise. Otherwise, there will be accelerated access to new treatments that may

prove either to have unfavorable risk–benefit ratios as compared with available alternatives or to offer only marginal net benefits that do not justify their costs.

Finally, in addition to serious deficiencies in determining the ethical legitimacy of randomized clinical trials, equipoise promotes premature discontinuation of trials based on interim data relating to treatment benefit.^{19,20} According to the equipoise doctrine, trials should be terminated when equipoise has been disturbed.³ However, data-monitoring committees may determine that equipoise has been disturbed by interim trial results before these data are sufficient to guide health policy decisions. Systematic reviews have documented an increasing incidence of early termination of randomized clinical trials, resulting in overestimates of treatment benefit.^{21–23} Early discontinuation also impedes the development of rigorous evidence regarding adverse treatment effects. Together, these equipoise-driven consequences bias the evidence base relevant to risk–benefit assessment.

APPLICATION TO A CONTROVERSIAL CLINICAL TRIAL

The problem with using equipoise to determine whether a randomized clinical trial is ethically appropriate is vividly illustrated by the recent controversy surrounding the development of a new agent for patients with metastatic melanoma, a uniformly fatal condition. The proportion of patients who have a response to dacarbazine, the current standard of care, averages 15%, with most responses associated with only partial tumor shrinkage.²⁴ PLX4032 is an experimental targeted intervention that has undergone early-phase testing.²⁵ On the basis of impressive rates and durations of response in a phase 1 trial among patients with melanoma that harbors a particular genetic mutation, the pharmaceutical company that developed the experimental agent has undertaken a randomized clinical trial designed as an open-label, head-to-head comparison between PLX4032 and dacarbazine.²⁶ Because the primary end point of this trial is overall survival, crossover from dacarbazine to PLX4032 after disease progression is not allowed.

Is there equipoise between the two treatments of this randomized clinical trial? A recent *New*

York Times article featuring this study suggests that equipoise is lacking.²⁷ In light of current knowledge, if PLX4032 were clinically available outside the trial, it is reasonable to suppose that virtually all clinicians and informed patients would opt for this treatment over the marginally effective and toxic standard chemotherapy. The fact that some physician-investigators support the conduct of this trial does not prove that equipoise exists. These persons may represent only a very small minority of experts, and they may favor conducting the trial for various reasons despite the lack of equipoise. Assuming that there is no equipoise, does it follow that this trial, which is aimed at developing the knowledge needed for regulatory assessment of effectiveness, is unethical?

To be sure, one might argue that there are ethically preferable designs for randomized clinical trials that are likely to confer greater benefits for — or impose less toxicity on — study participants, while still answering the relevant scientific, clinical, and policy questions. One ethical requirement for clinical research is to maximize benefits to trial participants, provided that this requirement is consistent with promoting the social value and scientific validity of the research.^{28,29} Perhaps crossover to the experimental agent, in a trial designed to evaluate progression-free rather than overall survival as a primary end point, should be permitted for patients with disease that progresses while they are receiving dacarbazine. Or perhaps, given the toxicity and limited efficacy of dacarbazine, a straightforward placebo-controlled design is appropriate. Although the planned trial offers the best opportunity to assess survival benefit, the relative merits of alternative designs for randomized clinical trials are debatable. None of them, however, satisfy equipoise. Clinicians will not be indifferent to randomly assigning patients between the promising experimental agent and the current, marginally effective, standard treatment, even if crossover were an option for those with disease progression,³⁰ nor would they be indifferent about the possibility of randomization to placebo.

Some observers might argue that PLX4032 should be approved without evaluation in a randomized clinical trial, based on the impressive evidence from the phase 1 trial. Additional, though less rigorous, evaluation of effectiveness might be accomplished by requiring that all patients re-

ceiving the new drug are enrolled in a data registry, with careful evaluation of outcomes to be compared with historical data from previous trials. If the marginal usefulness of the evidence that will result from a randomized clinical trial of PLX4032 for the purposes of making both clinical and policy decisions is limited, as compared with further observational research, then FDA approval without a randomized clinical trial would be justified, thus speeding access to the potential benefits of the new agent. PLX4032 is a genuinely difficult case, and reasonable people may differ over whether evaluation in a randomized clinical trial to clarify whether the drug offers a survival benefit is needed. For physicians and ethicists who adopt a therapeutic orientation to clinical trials and espouse equipoise, the answer is clear: a randomized clinical trial comparing PLX4032 with either the standard treatment or placebo is unethical because equipoise does not exist. From a health policy perspective, however, the answer depends on judging the knowledge value added by a suitably designed randomized clinical trial. In making this determination, equipoise is irrelevant.

CONCLUSIONS

Equipoise is fundamentally flawed as a criterion for determining whether a randomized clinical trial is justified to evaluate new treatments, even in the context of life-threatening or seriously debilitating conditions with marginally effective therapeutic options. As a rule, to inform regulatory and coverage decisions, rigorous evaluation of a new treatment before it is made available in clinical practice must be pursued beyond the point at which physicians and informed patients would choose it over the current standard treatment based on initial efficacy data. However, the examples of cisplatin for testicular cancer and bortezomib for multiple myeloma show that randomized clinical trials are not always required before new treatments are approved and covered.³¹⁻³³ Several scientific and clinical criteria may provide support for validation of new treatments without randomized clinical trials (Table 1). Nevertheless, evaluation in randomized clinical trials should be the default, with a heavy burden of proof before new treatments are approved and covered solely on the basis of evidence from uncontrolled

Table 1. Criteria Supporting the Approval of Agents for Life-Threatening Diseases on the Basis of Nonrandomized Evidence.

Compelling, usually mechanism-based rationale favoring the efficacy of the new agent
Evidence of large effect sizes on the basis of early clinical studies
Well-understood, typically poor outcome with limited interpatient variability given current therapy or supportive care
Availability of one or more concurrent or historical control groups with characteristics similar to those of the patients to be enrolled in the proposed study
Use of a clinical or validated surrogate primary end point in the uncontrolled trial

trials. The expense of new treatments (e.g., between \$4,000 and \$8,000 per month for targeted therapies such as bevacizumab for metastatic breast cancer) augments the burden of proof.^{34,35}

The resort to equipoise to guide decisions about evaluation of new treatments rests on a flawed intuition that study participants are harmed or wronged by being denied access to a promising but partially evaluated treatment. Participants are not harmed because they are not knowingly made worse off than they otherwise would be outside the trial, where presumably they would be offered standard treatment. They are not wronged because their right to evidence-based medical care is not violated: they are not entitled to experimental treatment that has yet to be adequately evaluated. Though they are psychologically and interpersonally challenging, trials of new treatments for life-threatening diseases that violate equipoise are both ethical and necessary for the development of evidence to support health policy decisions made on behalf of populations of patients.

The opinions expressed are those of the authors and do not necessarily reflect the position or policy of the National Institutes of Health, the Public Health Service, or the Department of Health and Human Services.

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

From the Department of Bioethics, Clinical Center, National Institutes of Health, Bethesda, MD (F.G.M.); and the Department of Pediatric Oncology, Dana-Farber Cancer Institute, and the Department of Medicine, Children's Hospital Boston — both in Boston (S.J.).

1. Marquis D. Leaving therapy to chance. *Hastings Cent Rep* 1983;13(4):40-7.
2. Gifford F. The conflict between randomized clinical trials and the therapeutic obligation. *J Med Philos* 1986;11:347-66.
3. Freedman B. Equipoise and the ethics of clinical research. *N Engl J Med* 1987;317:141-5.

4. Miller FG, Rosenstein DL. The therapeutic orientation to clinical trials. *N Engl J Med* 2003;348:1383-6.
5. Miller FG, Brody H. A critique of clinical equipoise: therapeutic misconception in the ethics of clinical trials. *Hastings Cent Rep* 2003;33(3):19-28.
6. Echt DS, Liebson PR, Mitchell LB, et al. Mortality and morbidity in patients receiving encainide, flecainide, or placebo — the Cardiac Arrhythmia Suppression Trial. *N Engl J Med* 1991;324:781-8.
7. Stadtmauer EA, O'Neill A, Goldstein LJ, et al. Conventional-dose chemotherapy compared with high-dose chemotherapy plus autologous hematopoietic stem-cell transplantation for metastatic breast cancer. *N Engl J Med* 2000;342:1069-76.
8. Moseley JB, O'Malley K, Petersen NJ, et al. A controlled trial of arthroscopic surgery for osteoarthritis of the knee. *N Engl J Med* 2002;347:81-8.
9. Rossouw JE, Anderson GL, Prentice RL, et al. Risks and benefits of estrogen plus progestin in healthy postmenopausal women: principal results from the Women's Health Initiative randomized controlled trial. *JAMA* 2002;288:321-33.
10. Manson JE, Hsia J, Johnson KC, et al. Estrogen plus progestin and the risk of coronary heart disease. *N Engl J Med* 2003;349:523-34.
11. Rettig RA, Jacobson PD, Farquhar CM, Aubry WM. False hope: bone marrow transplantation for breast cancer. Oxford, England: Oxford University Press, 2007.
12. Miller FG, Joffe S. Benefit in phase 1 oncology trials: therapeutic misconception or reasonable treatment option? *Clin Trials* 2008;5:617-23.
13. Pazdur R. Response rates, survival, and chemotherapy trials. *J Natl Cancer Inst* 2000;92:1552-3.
14. Buyse M, Thirion P, Carlson RW, Burzykowski T, Molenberghs G, Piedbois P. Relation between tumour response to first-line chemotherapy and survival in advanced colorectal cancer: a meta-analysis. *Lancet* 2000;356:373-8.
15. Shanafelt TD, Loprinzi C, Marks R, Novotny P, Sloan J. Are chemotherapy response rates related to treatment-induced survival prolongations in patients with advanced cancer? *J Clin Oncol* 2004;22:1966-74.
16. Johnson KR, Ringland C, Stokes BJ, et al. Response rate or time to progression as predictors of survival in trials of metastatic colorectal cancer or non-small-cell lung cancer: a meta-analysis. *Lancet Oncol* 2006;7:741-6.
17. Tang PA, Bentzen SM, Chen EX, Siu LL. Surrogate end points for median overall survival in metastatic colorectal cancer: literature-based analysis from 39 randomized controlled trials of first-line chemotherapy. *J Clin Oncol* 2007;25:4562-8.
18. Bruzzi P, Del Mastro L, Sormani MP, et al. Objective response to chemotherapy as a potential surrogate end point of survival in metastatic breast cancer patients. *J Clin Oncol* 2005;23:5117-25.
19. Gifford F. Community-equipoise and the ethics of randomized clinical trials. *Bioethics* 1995;9:127-48.
20. Buchanan D, Miller FG. Principles of early stopping of randomized trials for efficacy: a critique of equipoise and an alternative nonexploitation ethical framework. *Kennedy Inst Ethics J* 2005;15:161-78.
21. Montori VM, Devereaux PJ, Adhikari NK, et al. Randomized trials stopped early for benefit: a systematic review. *JAMA* 2005;294:2203-9.
22. Bassler D, Briel M, Montori VM, et al. Stopping randomized trials early for benefit and estimation of treatment effects: systematic review and meta-regression analysis. *JAMA* 2010;303:1180-7.
23. Wilcox RA, Djulbegovic B, Moffitt HL, Guyatt GH, Montori VM. Randomized trials in oncology stopped early for benefit. *J Clin Oncol* 2008;26:18-9.
24. Lui P, Cashin R, Machado M, Hemels M, Corey-Lisle PK, Einarson TR. Treatments for metastatic melanoma: synthesis of evidence from randomized trials. *Cancer Treat Rev* 2007;33:665-80.
25. Flaherty KT, Puzanov I, Kim KB, et al. Inhibition of mutated, activated BRAF in metastatic melanoma. *N Engl J Med* 2010;363:809-19.
26. Livingstone E, Zimmer L, Piel S, Schadendorf D. PLX4032: does it keep its promise for metastatic melanoma treatment? *Expert Opin Investig Drugs* 2010;19:1439-49.
27. Harmon A. New drugs stir debate on rules of clinical trials. *New York Times*. September 18, 2010:A1.
28. Joffe S, Miller FG. Bench to bedside: mapping the moral terrain of clinical research. *Hastings Cent Rep* 2008;38(2):30-42.
29. Emanuel EJ, Wendler D, Grady C. What makes clinical research ethical? *JAMA* 2000;283:2701-11.
30. Mishra R. Placebos break taboo in cancer drug tests: study seeks hope for desperately ill. *Boston Globe*. July 4, 2004:A1.
31. Tsimberidou A-M, Braiteh F, Stewart DJ, Kurzrock R. Ultimate fate of oncology drugs approved by the US Food and Drug Administration without a randomized trial. *J Clin Oncol* 2009;27:6243-50.
32. Richardson PG, Barlogie B, Berenson J, et al. A phase 2 study of bortezomib in relapsed, refractory myeloma. *N Engl J Med* 2003;348:2609-17.
33. Einhorn LH, Donohue J. Cis-diamminedichloroplatinum, vinblastine, and bleomycin combination chemotherapy in disseminated testicular cancer. *Ann Intern Med* 1977;87:293-8.
34. Berenson A. A cancer drug shows promise, at a price that many can't pay. *New York Times*. February 15, 2006.
35. Dedes KJ, Matter-Walstra K, Schwenkgenks M, et al. Bevacizumab in combination with paclitaxel for HER-2 negative metastatic breast cancer: an economic evaluation. *Eur J Cancer* 2009;45:1397-406.

Copyright © 2011 Massachusetts Medical Society.

RECEIVE IMMEDIATE NOTIFICATION WHEN AN ARTICLE
IS PUBLISHED ONLINE FIRST

To be notified by e-mail when *Journal* articles
are published Online First, sign up at **NEJM.org**.