

Value-Based Insurance Design

By abandoning the archaic principle that all services must cost the same for all patients, we can move to a high-value health system.

by **Michael E. Chernew, Allison B. Rosen, and A. Mark Fendrick**

ABSTRACT: When everyone is required to pay the same out-of-pocket amount for health care services whose benefits depend on patient characteristics, there is enormous potential for both under- and overuse. Unlike most current health plan designs, Value-Based Insurance Design (VBID) explicitly acknowledges and responds to patient heterogeneity. It encourages the use of services when the clinical benefits exceed the cost and likewise discourages the use of services when the benefits do not justify the cost. This paper makes the case for VBID and outlines current VBID initiatives in the private sector as well as barriers to further adoption. [*Health Affairs* 26, no. 2 (2007): w195–w203 (published online 30 January 2007; 10.1377/hlthaff.26.2.w195)]

ONE OF THE FUNDAMENTAL TENETS of clinical medicine is *primum non nocere*: “First do no harm.” In today’s complex health care environment, this principle should be extended beyond the clinician-patient relationship to health care financing. Implementing it is a challenging task in both clinical and financial settings for a number of reasons.

On the clinical side, most if not all interventions intended to improve health entail some risk of an adverse event. Clinicians must weigh these risks against the benefits when determining the appropriate course of treatment. In health care financing, there is often a similar yet underappreciated trade-off between cost containment initiatives and access to effective medical services. Efficiency would promote the use of “valuable” interventions whose expected net clinical benefits justify the associated expenditure and limit access to those services whose costs exceed the expected clinical gain. This is the fundamental paradigm of cost-effectiveness analysis.

In the status quo, cost-sharing amounts are generally constant for each specific service, although the clinical values of these services are extremely disparate and likely depend upon who receives them. With some exceptions for preventive and screening services, the level of cost sharing is seldom related to the potential benefit each service might provide.

Ideally, uniform patient copayments would discourage use of low-value care

.....
Michael Chernew (chernew@hcp.med.harvard.edu) is a professor of health care policy at Harvard University in Boston, Massachusetts. Allison Rosen is an assistant professor in the Division of General Medicine at the University of Michigan in Ann Arbor. Mark Fendrick is a professor of internal medicine there.

only. This assumes, however, that patients can distinguish between high- and low-value therapies and respond to copayments accordingly. Yet a large body of evidence demonstrates that higher copayments reduce the use of both highly valuable and marginally valuable health care services and may result in worse health outcomes.¹ In fact, the literature demonstrates that the adverse consequences of higher copayments can arise at even relatively modest levels.²

In response to the likely adverse clinical effects of the current trend toward higher copayments, Mark Fendrick and colleagues have proposed the Value-Based Insurance Design (VBID) approach, which advocates that copayment rates be set based on the value of clinical services (benefits and costs)—not exclusively the costs.³ In this setting, cost sharing is still put to use, but a clinically sensitive approach is explicitly adopted to mitigate the adverse health consequences of high out-of-pocket spending.⁴ Recognizing that the value of an intervention varies across patients, more-efficient resource allocation can be achieved when the amount of patient cost sharing is a function of the value that the specific service provides to the specific patient. Subsequent literature supports this basic idea.⁵

VBID: Economic Theory

Economic theory suggests that the value of insurance arises because it allows people to alleviate the financial risk associated with the risk of illness and because it allows those who become ill to afford care they would otherwise not be able to purchase.⁶ However, by lowering the cost of care to patients at the point of service, insurance encourages use of services whose clinical benefits might not justify the total cost. This excess consumption is commonly termed “moral hazard” and reduces the value provided by the health care system.⁷

The motivation behind the use of cost sharing to allocate medical services and contain costs follows standard economic theory, which presumes that consumers will use only those services whose benefit exceeds the cost to them. By increasing costs at the point of service, moral hazard can be reduced and value increased. The optimal amount of cost sharing reflects a balance between the risk and income-transfer effects of insurance against the moral hazard costs.

VBID relaxes the questionable assumption that when faced with cost sharing, consumers will balance costs and clinical value optimally. The underuse of valuable clinical services when a person is faced with even modest copayments likely represents a range of information issues, including how people understand their medical care, how they make decisions amid uncertainty, and how they make trade-offs over time.⁸

Because consumers' behavior might not follow standard assumptions, targeted reductions in the level of cost sharing can increase value by reducing underuse (for example, reducing cost sharing for beta-blocker therapy for patients with congestive heart failure [CHF] can increase beta-blocker adherence and therefore value in the health care system).

Experience With VBID

Although the theory of VBID argues for cost sharing that varies by individual, the administrative costs of implementing such a system, communication issues, and current information requirements make such a system impractical for widespread adoption. However, employers are actively experimenting with variations of VBID, and these initial efforts merit further consideration.⁹

■ **Two approaches.** In practice, there are two general approaches to VBID targeting. The first approach simply targets clinically valuable services for copayment reduction (for example, beta-blockers). Although these services provide substantial benefit for some users (such as patients with CHF or myocardial infarction [MI]), they provide less value for other patients (such as those with performance anxiety), and the system does not attempt to differentiate between these patients. The second approach targets patients with select clinical diagnoses (for example, CHF) and lowers copayments for specific high-value services (for example, beta-blockers and angiotension-converting enzyme [ACE] inhibitors).

The second approach, although requiring more-sophisticated data systems to implement, creates a differential copayment based on patients' characteristics. Programs using this approach typically identify patients with specific diseases, such as diabetes or coronary heart disease (CHD), and reduce copayments for only high-value services for these patients. Both the targeting of high-value services only and high-value services for specific groups of patients are examples of VBID, because they both use assessment of value to determine copayment rates.

■ **Experimentation with first approach.** Several firms are experimenting with one of these two forms of VBID. Pitney Bowes (Stamford, Connecticut) uses the first approach, reducing copayments for all users of drugs commonly prescribed for diabetes, asthma, and hypertension. A second program, implemented by ActiveHealth Management (an integrated care management company that is an independent subsidiary of Aetna), focuses on drugs as well, lowering copayments for ACE inhibitors and angiotensin-receptor blockers (ARBs), beta-blockers, medications for glucose control, statins, and inhaled steroids (used largely to treat asthma). In these initiatives, all users of these classes of drugs pay lower copayments, regardless of their level of benefit from them. The ActiveHealth program goes two steps further by excluding patients with contraindications from the copayment relief and by informing those who would benefit from, but are not using, the targeted services of the lower copayment.

Similar programs have been incorporated into some health savings account (HSA) products, which provide first-dollar coverage for medications used to treat important chronic diseases. For example, Aetna's HSA defines *preventive care* to include services that are important for chronic disease patients and therefore gives these services first-dollar coverage.¹⁰

■ **Experimentation with second approach.** Use of the second approach, which targets patients, is less common. Two examples are the municipality of Asheville,

North Carolina, and the University of Michigan. Both of these employers implemented a program that lowered copayments for selected medications for employees with diabetes. The Asheville program is pharmacist-led and includes coached self-management. It has since expanded to include other employers.

Barriers To VBID

Despite these examples of VBID, the national trend in health insurance design does not use value in setting cost-sharing parameters. We believe that this reflects several barriers to VBID implementation.

■ **Concern over costs of increased use.** With health care costs rising rapidly, purchasers are looking for ways to constrain cost growth. VBID typically involves lowering copayments for some underused, high-value services. Lower copayments are associated with higher costs and concerns that VBID will increase spending—at least in the short term—and dampen enthusiasm for VBID. Moreover, the employer might not capture any long-term savings accruing as a result of improved health status because of employee turnover.

■ **Cost of implementation.** Implementation of VBID involves identification of high-value services and, in cases in which the system targets specific patient groups, identification of which groups would be eligible for lower copayments. Systems that target patients will be more costly to implement, because the eligibility data must then be transferred from the payers to the point of service, often requiring data transfers and cooperation across organizations.

■ **Data issues.** It is not surprising that current patient-targeted VBID programs focus on diabetes, because patients with diabetes can easily be identified using existing pharmaceutical data sets. Integrated claims data would facilitate progress in other disease areas but would likely be more costly.

Additional challenges include absence of risk factors in claims data (for example, past heart attack and smoking status) and lack of data for new enrollees. VBID programs that target specific patient groups need alternative processes to deal with these data issues, which might add cost.

Electronic medical records and health assessment data—increasingly available as part of disease management programs—will expand capabilities and add further efficiencies. In fact, integration of VBID with disease management could offer a powerful program that might be more effective than either of these programs would be alone, while leveraging existing information systems. Some companies, such as ActiveHealth Management, have developed such information systems and are marketing patient-targeted VBID support systems.¹¹

■ **Insufficient research.** Another concern about VBID is that it will only succeed if research can differentiate between high- and low-value services. More-sophisticated systems that target patient groups will require more-detailed evidence than now exists in many disease areas.¹² However, existing evidence is sufficient to support VBID in selected disease areas.

■ **Human resource concerns.** Some stakeholders have expressed concern that people will object to some patients' being charged less than others for certain services. Explaining the program to employees could be complex, particularly if programs differentiate by patient group. Employees would also need to be informed of their eligibility for the program, which could change over time. Moreover, where workers are unionized, employers might need to get approval from the union.

■ **Fraud.** VBID programs that differentiate among patients will inevitably require algorithms that define which patients are eligible for the lower copayment. One concern is that patients or providers might be encouraged to misreport information to qualify for the reduced copayment. To minimize this concern, programs must be limited to areas where identifiable information exists to classify patients. As discussed above, some disease areas are more amenable to this than others.

■ **Legal barriers.** An additional concern is that legal and regulatory barriers might impede implementation of VBID programs. However, existing programs, such as those discussed above, demonstrate that these concerns can be overcome. In some cases, regulatory concerns are relevant. For example, there is ambiguity regarding the legality of inclusion of preventive services for chronic diseases in the definition of *preventive services* for HSAs. In government programs, other policies are relevant. The Medicare Health Support programs, which serve patients with chronic diseases, are limited in their ability to give patients financial incentives to encourage the use of high-value services.

■ **Privacy concerns.** Another concern, particularly in programs that vary by patient group, is that VBID requires identification of employees with specific conditions. It is important that the transfer of data and communication activities surrounding VBID be sensitive to this information and that they comply with the Health Insurance Portability and Accountability Act (HIPAA) privacy regulations. Similar issues arise with disease management programs.

■ **Unintended incentives.** Two types of unintended incentives associated with VBID are of concern. First, if copayments are lowered for all products, incentives to use more-efficient delivery settings or services might be reduced (for example, VBID might discourage use of generic medications if the copayments for important brand-name medications are lowered). The magnitude of this effect is an empirical issue, but the concern can be addressed by maintaining a cost advantage for favored products or by use of other programs to encourage use of favored products.

Second, because certain risk factors are associated with behavior such as smoking, VBID could be interpreted as encouraging such behavior. This concern can be addressed without abandoning the underlying VBID design by adjusting the employee share of premiums or integrating the program with a disease management program.

■ **Adverse selection.** Since VBID favors patients with specific diseases, either because the patients are targeted or because the services they use are targeted, VBID plans might attract a disproportionate number of patients with chronic condi-

tions.¹³ This selection issue is similar to that which could arise any time a plan offered high-quality services for patients with chronic diseases through mechanisms such as disease management. The concern is more salient for small employers or employers that offer multiple plan options; it can be surmounted by risk adjustment or by implementing the VBID design for all employees in a firm.

Despite these barriers, VBID programs need not incorporate all possible details and degrees of sophistication. Many barriers can be surmounted by simplifying the system. Programs that do not differentiate by patient group clearly face fewer barriers but will likely have less favorable financial profiles. The appropriate degree of targeting will depend on the trade-off between the cost of overcoming these barriers relative to the possible gain from better targeting. As the experiences of the existing programs illustrate, benefit packages in the VBID spirit can be implemented with success.

Lessons From The Field: VBID At The University Of Michigan

The evolution of a VBID program implemented at the University of Michigan (UM) might prove instructional for future VBID efforts. On 1 July 2006, UM implemented M-Healthy: Focus on Diabetes Program for its 2,200 employees and dependents with a diagnosis of diabetes mellitus.¹⁴ This program provides copayment reductions to targeted patients (diabetics) for targeted interventions deemed from the medical evidence as highly beneficial. The targeted services include several drugs that affect blood sugar, blood pressure, cholesterol, and depression and that help prevent or reduce the long-term complications of diabetes. Copayments for annual eye exams were also reduced for enrollees in the UM health plan. Only people with diabetes, identified by pharmaceutical claims, are eligible for copayment reductions.¹⁵

Because of contract language with the three unions representing UM employees, implementation of the pilot program required agreement by the unions. The university's pharmacy benefit management (PBM) firm provided the targeted copayment reductions at the point of service. All UM employees were notified about the pilot program by letter and e-mail. To maintain the tiered formulary incentives for use of less expensive medications (such as generics), the VBID intervention lowers copayments in a graded fashion. For the medications of interest, tier 1 copays decreased by 100 percent (from \$7 to \$0); tier 2 copays, by 50 percent (from \$14 to \$7); and tier 3 copays, by 25 percent (from \$24 to \$18). The program received overwhelming employee support through numerous e-mail testimonials and virtually no dissent, which suggests that human-resource concerns can be overcome.

Financial Effects Of VBID

The goal of the health care system is to improve health, not to save money. Dropping coverage completely could save money, at least in the short run, yet it would

not be socially desirable. The driving idea behind VBID is that the use of high-value services should be encouraged. Yet given the concern about health care cost growth, it is imperative to assess the financial consequences of a VBID design. Because there is no single VBID intervention, it is difficult to provide an answer to the question regarding the “bottom line” effects of such a plan.

■ **Direct costs plus added value.** The basic accounting identity that describes the financial effects of lowering copayments (or maintaining low copayments) for any given service is straightforward. Specifically, the cost to the payer of lower copays is the extra share of spending for the services that would have been used anyway and the purchaser share of the costs of increased consumption resulting from the copay reductions. This additional expense of extra consumption is assumed to add value because VBID targets high-value services.

■ **Savings from improved health.** Offsetting the direct costs are the savings due to the improved health generated by the extra service use. For example, the direct costs of lower copayments for cholesterol-lowering medication would be offset, at least partially, by savings attributable to fewer heart attacks. The net financial benefit will be greater if the underlying risk of an adverse outcome is high, if the cost of that adverse outcome is high, if consumers are very responsive to lower copayments, and if the service is very effective at preventing the adverse outcome.

■ **The targeting factor.** Because these factors vary across the population, the financial impact of a VBID program will depend on the level and precision of targeting. Most services provide significant value for a subset of patients. The better the system is at identifying those patients, and the more responsive those patients are to copayment changes, the more likely the system will be to achieve a high financial return. Employers with more-targeted programs incur lower costs because fewer services are eligible for lower copayments, and most of the financial and clinical gains still accrue because the patients who benefit the most get the lower copayments. In deciding whether to limit VBID to targeted patient groups (as opposed to just targeting high-value services), purchasers will need to weigh added implementation costs against the better financial profile from more-targeted programs.

Simulation exercises suggest that well-targeted VBID programs could save money. Allison Rosen and colleagues provide an example of where VBID could save money, reporting that cost savings are possible when selected drug classes are provided free of charge to Medicare enrollees with diabetes mellitus.¹⁶ However, Medicare beneficiaries are at greater risk of costly adverse events in a shorter window of time, so these results might not generalize to a commercially insured population.

One could design a VBID system to achieve any cost target by financing the costs of lower copays for high-value services through higher copays on less valuable services. Dana Goldman and colleagues provide the best available analysis of such an approach by examining the impact of financing lower copays for high-benefit statin users by increasing copays for lower-benefit statin users.¹⁷ If the

clinical benefits of statins provided to those low-risk patients were cost-effective (which we believe to be so), it would be preferable to implement a broader VBID program financed by raising copays for other services unrelated to statins, or even unrelated to cardiovascular disease, that are determined to be of lesser value.

Estimates from the Pitney Bowes and Asheville experiences suggest that VBID can save money. One year after Pitney Bowes lowered medication copays for asthma and diabetes medications in 2001, the company reported in the *Wall Street Journal* a one-year savings of \$1 million, although more rigorous controlled evaluations of this program would be needed to definitively assess its impact.¹⁸ An evaluation of the Asheville project (which included more than copay reduction) reported five-year outcomes that include marked increases in medication adherence, a two- to threefold increase in achieving diabetes performance measures, approximately a 50 percent decrease in average annual sick leave, and a trend in overall medical costs that was 58 percent below expected levels.¹⁹ However, it is unclear how sensitive this finding is to the methods used to estimate expected costs.

Concluding Comments

VBID is a clinically sensitive form of cost sharing because it recognizes that services vary in the value they provide to patients and that not all patients with a specific clinical condition receive the same level of benefit from a specific intervention. If different cost-sharing provisions are allowed for different services, value can be increased without eliminating the role of cost sharing in the system.

In this way, VBID can address several important inconsistencies in the current system and work synergistically with other initiatives. For example, current disease management programs and pay-for-performance (P4P) systems devote resources to improving the quality of care for targeted patients in selected clinical areas. Financial aspects of benefit design should support such efforts, but existing cost-sharing arrangements often discourage the use of the high-value services encouraged by P4P and disease management.²⁰ Through an alignment of incentives based on overall value of clinical services, not just cost, VBID could ameliorate this concern. By using our knowledge wisely and abandoning the archaic principle that all services must cost the same for all patients, regardless of clinical situation, we can move toward a high-value health care system for all.

.....
This research was conducted as part of work for the Center for Value-Based Insurance Design, which receives support from Pfizer, Merck, and the University of Michigan. Michael Chernew and Mark Fendrick have a consulting agreement with ActiveHealth Management, which is mentioned in the paper.

NOTES

1. J.P. Newhouse and the Insurance Experiment Group, *Free for All? Lessons from the RAND Health Insurance Experiment* (Cambridge, Mass.: Harvard University Press, 1994); T.B. Gibson, R.J. Ozminkowski, and R.Z. Goetzel, "The Effects of Prescription Drug Cost Sharing: A Review of the Evidence," *American Journal of Managed Care* 11, no. 11 (2005): 730-740; T. Rice and K.Y. Matsuoka, "The Impact of Cost-Sharing on Appro-

- appropriate Utilization and Health Status: A Review of the Literature on Seniors,” *Medical Care Research and Review* 61, no. 4 (2004): 415–452; M. Heisler et al., “The Health Effects of Restricting Prescription Medication Use because of Cost,” *Medical Care* 42, no. 7 (2004): 626–634; and A.L. Siu et al., “Inappropriate Use of Hospitals in a Randomized Trial of Health Insurance Plans,” *New England Journal of Medicine* 315, no. 20 (1986): 1259–1266.
2. For example, Goldman and colleagues identify adverse effects of copayments when the average nonpreferred brand-name copayment increased from about \$12 to about \$20. D.P. Goldman et al., “Pharmacy Benefits and the Use of Drugs by the Chronically Ill,” *Journal of the American Medical Association* 291, no. 19 (2004): 2344–2350.
 3. The VBIID concept was originally referred to as the “Benefit Based Copay.” See A.M. Fendrick et al., “A Benefit-Based Copay for Prescription Drugs: Patient Contribution Based on Total Benefits, Not Drug Acquisition Cost,” *American Journal of Managed Care* 7, no. 9 (2001): 861–867.
 4. A.M. Fendrick and M.E. Chernew, “Value Based Insurance Design: A ‘Clinically Sensitive’ Approach to Preserve Quality and Contain Costs,” *American Journal of Managed Care* 12, no. 1 (2006): 18–20.
 5. J.D. Kleinke, “Access versus Excess: Value-Based Cost Sharing for Prescription Drugs,” *Health Affairs* 23, no. 1 (2004): 34–47; A.M. Garber, “Cost-Effectiveness and Evidence Evaluation as Criteria for Coverage Policy,” *Health Affairs* 23 (2004): w284–w296 (published online 19 May 2004; 10.1377/hlthaff.w4.284); J.C. Robinson, “Managed Consumerism in Health Care,” *Health Affairs* 24, no. 6 (2005): 1478–1489; and J.P. Newhouse, “Reconsidering the Moral Hazard–Risk Avoidance Tradeoff,” *Journal of Health Economics* 25, no. 5 (2006): 1005–1014.
 6. K. Arrow, “Uncertainty and the Welfare Economics of Medical Care,” *American Economic Review* 53, no. 5 (1963): 941–973; and J.A. Nyman, “The Value of Health Insurance: The Access Motive,” *Journal of Health Economics* 18, no. 2 (1999): 141–152.
 7. M.V. Pauly, “The Economics of Moral Hazard,” *American Economic Review* 58, no. 3 (1968): 531–537.
 8. Newhouse, “Reconsidering the Moral Hazard–Risk Avoidance Tradeoff.”
 9. Health plan designs are often created by self-insured employers. Therefore, much of the discussion of VBIID relates to initiatives in plan designs driven by employers.
 10. J.C. Robinson, “Consumer-Directed Health Insurance: The Next Generation,” *Health Affairs* 24 (2005): w583–w590 (published online 13 December 2005; 10.1377/hlthaff.w5.583).
 11. J.C. Robinson and J.M. Yegian, “Medical Management after Managed Care,” *Health Affairs* 23 (2004): w269–w280 (published online 19 May 2004; 10.1377/hlthaff.w4.269).
 12. R.A. Hayward et al., “Reporting Clinical Trial Results to Inform Providers, Payers, and Consumers,” *Health Affairs* 24, no. 6 (2005): 1571–1581.
 13. J.P. Newhouse and A.D. Sinaiko, “What We Know and Don’t Know about the Effects of Cost Sharing on Demand for Medical Care—and So What?” (Paper presented at Health Economics Conference, Oberlin College, September 2006).
 14. Michigan Healthy Community, “M-Healthy: Focus on Diabetes,” 2005, <http://www.umich.edu/~hrra/mhealthy/improve/diabetes.html> (accessed 8 January 2007).
 15. The program will be expanded to UM employees with ischemic heart disease once clinical and pharmacy data are linked to allow for real-time identification of patients with that condition.
 16. A.B. Rosen et al., “Cost-Effectiveness of Full Medicare Coverage of Angiotensin-Converting Enzyme Inhibitors for Beneficiaries with Diabetes,” *Annals of Internal Medicine* 143, no. 2 (2005): 89–99.
 17. D.P. Goldman, G.F. Joyce, and P. Karaca-Mandic, “Varying Pharmacy Benefits with Clinical Status: The Case of Cholesterol-Lowering Therapy,” *American Journal of Managed Care* 12, no. 1 (2006): 21–28.
 18. S. Hensley, “From ‘One Size Fits All’ to Tailored Co-Payments,” *Wall Street Journal*, 16 June 2004; and J.J. Mahoney, “Reducing Patient Drug Acquisition Costs Can Lower Diabetes Health Claims,” *American Journal of Managed Care* 11, no. 5 Supp. (2005): S170–S176.
 19. C.W. Cranor, B.A. Bunting, and D.B. Christensen, “The Asheville Project: Long-Term Clinical and Economic Outcomes of a Community Pharmacy Diabetes Care Program,” *Journal of the American Pharmaceutical Association* 43, no. 2 (2003): 173–184.
 20. M.E. Chernew, A.B. Rosen, and A.M. Fendrick, “Rising Out-of-Pocket Costs in Disease Management Programs,” *American Journal of Managed Care* 12, no. 3 (2006): 150–154.

Cost-Effectiveness And Evidence Evaluation As Criteria For Coverage Policy

Cost-effectiveness analysis could shift from being an academic curiosity to an essential tool for health care decision making.

by **Alan M. Garber**

ABSTRACT: Private health plans and government health insurance programs in the United States base their coverage decisions on evidence criteria, rather than explicit cost-effectiveness criteria. As health spending continues to grow rapidly, however, approaches to coverage policy that ignore costs fail to meet the needs of consumers, employers, health plans, and federal and state governments. I describe the role of evidence-based criteria in formal coverage decision making and contrast the ways that these criteria differ from cost-effectiveness criteria. Finally, I discuss options for incorporating considerations of cost-effectiveness into coverage policy and other aspects of benefit design.

RESURGENT HEALTH SPENDING GROWTH and the continuing erosion of private health insurance have renewed U.S. debates over health care reform. Absent from these debates, however, is any systematic discussion of processes to choose the medical goods and services that health insurance should cover. Policymakers may instinctively sidestep the topic as a narrowly technical issue, to be decided by physicians and others with the patience and interest to evaluate a mass of information about medical treatments and diagnostic tests. They may also see little incentive to pursue it, knowing the political risk that comes with any public stand on coverage policy.

Their reticence is unfortunate, though, because coverage policy is so tightly linked to the affordability of health insurance, and hence the rate of uninsurance. When the cost of purchasing a private health insurance plan rises, the number of Americans with commercial health insurance falls: Employers stop offering their employees health insurance, and employees stop paying their share of premiums when their employers continue to offer insurance. Coverage policy also influences the types of medical care Americans receive, because health insurance coverage is the gateway to the availability of medical innovations. It is difficult to imagine

Alan Garber (garber@stanford.edu) is staff physician, Veterans Affairs Palo Alto Health Care System. He is also the Henry J. Kaiser Jr. Professor and professor of medicine, director of the Center for Primary Care and Outcomes Research, and director of the Center for Health Policy at Stanford University in Stanford, California.

how therapies that cost thousands of dollars per patient, such as left ventricular assist devices for severe congestive heart failure, could be adopted if health insurance did not cover them.

Although they did not arise from an explicit legislative process, de facto principles for coverage decision making have emerged. They are the product of historical practices, legal decisions, and insurance contract language. Coverage policy under both Medicare and most commercial health insurance plans is based upon a determination that a medical product has proved to be effective. That is, most explicit processes for making coverage decisions in the United States are based on evidence, not on cost-effectiveness or any other direct measure of value.¹

This paper discusses the similarities and differences in using evidence and value criteria as bases for coverage decision making. It also addresses the complementary roles of coverage policy and other approaches to limiting the cost of health insurance, such as increased cost sharing, and the implications for the design of health insurance.

Coverage Policy And The Costs Of Health Insurance

Economists have long identified “moral hazard,” the overuse of health services that occurs because the insured person bears only a fraction of the cost of covered services, as the chief cause of excessive health spending.² Moral hazard raises the level of health spending. By stimulating the development of new technologies, it also increases the rate of spending growth. The expectation that health insurance will boost demand and revenues in the future is a powerful incentive to invest in the development of new medical technologies. The high rate of innovation in medical products and services leads to better health but also higher spending.³

The simplest instrument insurers have to keep costs down is to negotiate favorable reimbursement rates with providers. Such strategies, of course, have little ability to offset the growth in spending that results from heavier use of costly services. Thus, it may be more important to control moral hazard.⁴ Their main tools for discouraging excessive use are supply-side incentives, direct utilization controls, and copayments and deductibles that expose patients to the financial consequences of their use of health care. In addition, all health insurance plans place boundaries on the products and services that qualify for reimbursement. These boundaries constitute the insurer’s coverage policy.

Coverage is described in health insurance contracts, which list entire categories of products and services that are excluded from coverage, such as cosmetic surgery, as well as categories of inclusion, such as hospitalization for medical emergencies. The contracts cannot provide detailed descriptions of every service that will and will not be covered within a category of services eligible for inclusion, so they usually state that the insurer will reimburse all “medically necessary” goods and services. The interpretation of the term “medically necessary” has varied over time and across health plans, but today it generally rests upon the application of

an evidence standard.⁵

Such standards have two critical components: a determination about whether enough evidence is available to support conclusions about the effectiveness of the intervention in question (adequacy of evidence), and a determination about what that evidence implies about effectiveness (magnitude of benefit). This approach promises to reduce waste and improve safety by avoiding payment for products and services that are likely to be harmful or of no benefit.⁶ The emphasis on high-quality evidence represents a marked change from an earlier era of medicine, when the doctor's beliefs about the value of an intervention, especially if they were widely shared, were sufficient to establish medical necessity.

Although different groups do not always reach the same conclusion about a particular technology, and the specific processes that they use to evaluate evidence vary, there has been a remarkable convergence in the acceptance of the principle that coverage determinations—and indeed, medical practice itself—should be guided by the results of rigorously designed studies, rather than expert opinion or the most common forms of practice. Particularly for processes that are intended to inform coverage decisions for large numbers of people, such as the Blue Cross Blue Shield Association's Medical Advisory Panel and the federal Medicare Coverage Advisory Committee (MCAC), these evidence processes typically place great weight on information from well-designed clinical trials.⁷ The approach to evidence evaluation is similar to approval processes used by the U.S. Food and Drug Administration (FDA) and to the evidence ratings pioneered by Canadian and U.S. task forces on preventive services.⁸

Randomized clinical trials have great influence because they are less susceptible to bias than are studies with less rigorous design. Observational studies sometimes accurately predict the results of randomized trials, particularly in areas such as the treatment of heart disease. However, observational studies in many other areas, such as cancer treatment, are highly susceptible to bias.⁹ Treatments that appeared to provide large benefits in well-designed observational studies, such as bone marrow transplantation for advanced breast cancer, were found to be ineffective in randomized trials, perhaps because the women who received more aggressive treatment in observational settings were healthier at the outset.

For health plans, rigorous evidence-based processes have a powerful appeal: It is difficult to argue that an ineffective test or treatment should be covered by insurance or even administered by a physician. Cutting waste without eliminating effective care would seem a painless way to begin to limit medical spending. But the adoption of an evidence standard does not represent solely a commitment to avoid ineffective care. Since no intervention is assumed to be effective until it has been proved effective, the burden of proof for a new medical intervention is placed on its advocates. Examining the evidence requirement from their point of view is an important step toward understanding its consequences.

The Burden Of Establishing Effectiveness

Meeting an evidence standard is costly. The drug approval process offers a view into how an evidence-based process works and what it costs. Of course, evaluations of diagnostic tests and surgical procedures differ in important respects from the evaluation of drugs for FDA approval, and drugs are a relatively small fraction of the mix of products and services that health plans cover. Nevertheless, the most critical issues faced by any group or individual seeking to demonstrate effectiveness are common to all medical interventions.

Tests of safety and effectiveness in humans are believed to be responsible for more of the cost of drug development than basic drug discovery research. In recent years, the cost of large-scale clinical trials for a successful drug, according to a study based on industry-supplied data, averaged about \$86.3 million (in 2000 dollars).¹⁰ Per patient costs of trials in the United States are estimated to fall in the range of \$10,000–\$50,000.

These high costs make it essential to keep trials as small as possible. However, a trial that enrolls too few patients will be unable to demonstrate conclusive (statistically significant) evidence of benefit. The magnitude of the intervention's health improvement, the characteristics of the patients enrolled in the trial, and the variability in their health outcomes are among the factors that determine how large a trial is needed to achieve statistical significance. Investigators can improve a study's prospects of success by enrolling participants who are likely to show the greatest benefit most quickly, or by increasing either the duration of follow-up or the number of patients studied. The importance of sample size can hardly be overstated. A pooled analysis of clinical trials conducted in the early 1990s noted that the vast majority of trials reporting negative results did not have adequate statistical power, despite large effect sizes (relative changes in the health outcome) of 25 percent or even 50 percent.¹¹

Although they cannot eliminate the risk and uncertainty inherent in evaluating a new intervention—if there were no uncertainty, there would be no reason to conduct a trial—researchers can prevent many of the pitfalls by increasing the size or duration of the trial.¹² The combination of high per patient costs and the need to have adequate sample sizes is responsible for the high costs of clinical trials.

Although these costs can be a daunting obstacle, the prospect of revenues from monopoly in the sale of the intervention is a strong enough incentive for both pharmaceutical companies and device manufacturers to fund trials.¹³ Lacking well-defined, enforceable intellectual property rights, developers of innovative care processes and medical procedures have little prospect of gaining a monopoly. They cannot expect future payments large enough to offset the cost of studies that would establish effectiveness. Perhaps that is one reason why many surgical innovations are tied to the use or implantation of a patented device (for example, left ventricular assist device, implantable cardioverter-defibrillator, or coronary stent). Interventions that do not lead to monopoly products are sometimes stud-

ied with the support of the National Institutes of Health (NIH), the Department of Veterans Affairs (VA), and other government agencies, but federal funds only support trials of a fraction of promising forms of care.

Thus, evidence-based processes, which usually build upon explicit, statistically based criteria, are subject to the important qualification that someone had to have conducted a convincing study. Because monopoly rewards are often the chief incentive to fund research, evidence standards tend to favor monopoly products over other approaches to improving health outcomes, such as a new use for a generic drug, a better diagnostic strategy, or an improvement in delivering care. A bias toward such products, in turn, has important implications for spending.¹⁴

Applying Cost-Effectiveness Analysis To Coverage Decisions

Advocates for quality improvement remind us that evidence-based processes reduce spending by discouraging the use of ineffective medical care. Cutting waste is an attractive way to cut the level of health spending, but it may not slow its rate of growth. Most innovation represents improvements in care, and it is the growth in the volume and intensity of care, not disproportionate growth in wasteful care, that drives medical spending.¹⁵ Cost-effectiveness analysis can complement strategies to eliminate waste, since it can be used to guide utilization away from procedures that produce little benefit at high cost—in other words, to improve the efficiency of health care.¹⁶

Ideally, health insurance would promote the use of cost-effective medical services. It might do so by covering only services whose cost-effectiveness ratio is equal to or less than a cutoff (threshold) value.¹⁷ Under specific assumptions, the cutoff can be inferred from individual preferences, but the limited literature on this topic has not led to a consensus about how such thresholds should be determined and used.¹⁸ For example, if the cost-effectiveness threshold were based upon a person's willingness to pay for an improvement in health, the threshold would vary from one person to another. But many proponents of using cost-effectiveness analysis for health care decision making would apply a single threshold to an entire population.

Another approach would avoid selecting a threshold cost-effectiveness ratio and would instead compare the cost-effectiveness of various widely used interventions, giving an idea of the value of the intervention relative to other familiar health interventions. The ranking of the cost-effectiveness of various interventions is presented in a "league table." This approach has been criticized in part because the tables often report results that have been obtained from studies using different, and often incompatible, methods.¹⁹ The problem is particularly severe when studies use different measures of health effects: One may use changes in life expectancy; another, changes in quality-adjusted life years (QALYs); and yet another, changes in cholesterol or blood pressure. Most importantly, although a league table ranks interventions by their cost-effectiveness ratios, it does not tell

the reader where to draw the line between acceptable and unacceptable interventions—indeed, that would be equivalent to selecting a cost-effectiveness cutoff.²⁰

Setting the cutoff at a level that would lead to the rejection of potentially life-saving procedures is controversial among those who expect that all effective care will be available to everyone. Furthermore, rigid application of a specific cutoff cost-effectiveness ratio is rarely possible—if only because effectiveness varies from one person to another—nor would it guarantee socially acceptable outcomes. Awareness of the incompleteness of the threshold as a decision criterion has led expert panels to conclude that it should be combined with other information to guide clinical and policy decisions. For example, they would consider whether other treatments are available for the disease in question. They might also modify standards to shift care toward underserved racial or ethnic groups. This is similar to the approach that Oregon adopted in its attempt to distribute Medicaid funds to a broader population of uninsured people. Oregon started with a ranking of procedures based principally on cost-effectiveness but developed a very different list after extensive public discussion.²¹

Once the details of such a process are determined, how do the resulting choices differ from those based on an evidence-based approach? We begin by asking which interventions that are highly effective for their cost will readily pass an evidence standard, and which will not.

Both methods are likely to “pass” an intervention that is inexpensive and highly effective. A relatively small sample size or a short-duration trial, or both, would be sufficient to establish a statistically significant benefit. If the intervention is extremely expensive, it will pass an evidence but not a cost-effectiveness criterion.²²

Beyond these general points, the higher costs of establishing effectiveness in a trial—largely driven by statistical power, but also by considerations such as the difficulty of identifying and enrolling patients suitable for the trial and the burdens placed on patients who decide to enroll—tend to make an evidence hurdle higher. A cost-effectiveness criterion will be harder to pass when the intervention is very expensive.

Recent deliberations of the MCAC highlight differences between a purely evidence-based approach to the evaluation of medical interventions and one based on cost-effectiveness. The MCAC concluded that there was adequate evidence of effectiveness for implantable cardioverter-defibrillators; left ventricular assist devices; and verteporfin, a drug used to treat age-related macular degeneration. However, there were substantial questions about the appropriate role of each of these technologies. Published studies showed that cardioverter-defibrillators represented relatively good values for some but not all potential candidates for treatment. The cost-effectiveness of verteporfin and the left ventricular assist device had never been studied, and there was much doubt about whether the benefits were worth the high costs. The MCAC had no mechanism by which it could consider costs or value in making its coverage recommendations.

Is It Time To Rethink Evidence Evaluation As The Basis For Coverage Determinations?

Physicians, hospitals, and health plans find, as does Medicare, that evidence-based processes are not fully adequate for designing care or reimbursement policies. The idea that health plans should only pay for care that is of known effectiveness is no longer controversial. But health care providers and plans increasingly question whether medical innovations that provide genuine but modest benefits at high cost should be adopted.

Cost-effectiveness analysis has long been the preferred method to explicitly address value in medical care, yet it is not a common feature of formal coverage decision making by private U.S. health plans. My colleagues and I conducted a survey of medical directors of 228 managed care plans nationwide in 2001, representing 119 million covered lives. This survey revealed that 90 percent of the plans consider costs in some form when evaluating new interventions.²³ However, only 40 percent use formal cost-effectiveness analysis (Exhibit 1). Effectiveness appears to trump cost as an influence on coverage decisions: According to medical directors, 93 percent of all plans and 98 percent of large plans will cover a more effective intervention, even if it is more costly. If a new intervention is more expensive but no more effective than an existing one, only 16 percent will cover it, while only 8 percent will cover a less costly new intervention if it is also less effective (Exhibit 2).

Concerns about the interpretation of existing insurance contracts and about potential litigation may have discouraged the explicit use of cost-effectiveness analysis in coverage decisions.²⁴ In addition, providers and health plans may have doubts about the soundness of cost-effectiveness methods. Notwithstanding widely cited standards for the conduct of cost-effectiveness studies, questions remain about technical aspects of the methods and the ways they should be implemented.²⁵ The medical profession is not nearly as familiar with cost-effectiveness analysis as with clinical trials, and to many nonspecialists, cost-effectiveness anal-

EXHIBIT 1

How Health Plans Take Cost Into Consideration When Evaluating New Interventions

	Formal CE analysis (%)	Selectively apply preauthorization (%)	Establish explicit coverage policies (%)	Require less costly interventions first (%)	Consider cost in any of these ways (%)
Small plans	36	48	53	62	92
Large plans	44	51	56	55	88
All plans	40	49	54	58	90

SOURCE: Adapted from L.A. Bergthold et al., "Using Evidence and Cost in Managed Care Decision-Making" (Stanford, Calif.: Center for Health Policy/Center for Primary Care and Outcomes Research, Stanford University, 2002), available as a supplemental document online at content.healthaffairs.org/cgi/content/full/hlthaff.w4.284v1/DC2.

NOTE: CE is cost-effectiveness.

EXHIBIT 2
Likelihood That Plan Will Cover A New Intervention Compared With A Standard Intervention

	Equal effectiveness for equal cost (%)	Equal effectiveness for greater cost (%)	Less effectiveness for equal cost (%)	Less effectiveness for less cost (%)	Greater effectiveness for equal cost (%)	Greater effectiveness for greater cost (%)
Small plans	92	10	2	3	99	87
Large plans	96	21***	4	13**	99	98
All plans	94	16	3	8	99	93

SOURCE: Adapted from L.A. Bergthold et al., "Using Evidence and Cost in Managed Care Decision-Making" (Stanford, Calif.: Center for Health Policy/Center for Primary Care and Outcomes Research, Stanford University, 2002), available as a supplemental document online at content.healthaffairs.org/cgi/content/full/hlthaff.w4.284v1/DC2.

** $p < .05$ *** $p < .01$

ysis is neither transparent nor easily understood. This strikes specialists as ironic, since the technique can highlight otherwise implicit assumptions and make it easier to appreciate their implications.

Another reason for its limited adoption is the difficulty in conveying the magnitude and implications of uncertainty in the findings. There are numerous sources of uncertainty in such analyses: sample variability of outcomes observed in clinical trials; uncertainty about health events occurring after the end of a trial; uncertainty about nearly every component of costs; and uncertainty about the structure of models used in the analyses. Several techniques are available to measure and present these sources of uncertainty, including sensitivity analysis and probabilistic cost-effectiveness analysis.²⁶ The methodology and presentation of uncertainty in cost-effectiveness analysis has grown more sophisticated in recent years, but there is not a consensus about how the uncertainty should influence decision making. For all of these reasons, cost-effectiveness analysis is not poised to replace evidence evaluation in the near future.

However, it is likely that cost-effectiveness analysis will complement evidence evaluation. Although they report limited use of formal cost-effectiveness analysis and are not sure how to use it, many medical directors believe that it can and should play a greater role.²⁷ They believe that evidence evaluation should remain an important component of the decision to cover a medical product or service. However, because it does not incorporate cost considerations and because it is an imperfect proxy for cost-effectiveness, it is no longer an adequate basis for coverage decision making.²⁸ They could build upon current approaches while incorporating costs simply by assessing the cost-effectiveness of interventions that pass an evidence criterion but whose value is in question, and they could use cost-effectiveness analysis to help decide what to do when there is suggestive but not compelling evidence of effectiveness.

What action should plans take when they conclude that an intervention is not cost-effective? They could deny coverage entirely in limited circumstances, such

as a procedure that costs more and is clearly less effective (in the language of cost-effectiveness analysis, strictly dominated) than an alternative. It would be harder to deny coverage for a unique treatment for a life-threatening disease solely on the basis of poor cost-effectiveness.

There may be a broader scope for application of cost-effectiveness analysis in other aspects of benefit design. Several years ago, Mark Pauly and Philip Held argued that future cost savings from some interventions approached or even exceeded their immediate costs. For example, pneumococcal vaccine in a high-risk patient costs less than the resulting decline in spending for future care of pneumonia. A health plan could improve health outcomes and lower overall health spending by waiving the copayment—or even providing a subsidy—to ensure that such patients received the vaccine.²⁹ Tiered copayments for prescription drugs became common after Pauly and Held's paper appeared, and today it seems obvious that a similar copayment design could be applied to other medical interventions. In typical three-tier copayment arrangements, small copayments are required for the generic drugs; high copayments for brand-name, nonpreferred drugs; and intermediate copayments for brand-name, preferred drugs. The tiers, which are primarily based on drug acquisition costs, shift use toward lower-cost drugs.³⁰

A drawback of tiered copayment, however, is that the low-cost drugs it promotes are not necessarily high-value drugs; sometimes the most cost-effective drug is in the second tier, not the first, despite a higher acquisition cost. Mark Fendrick and colleagues have argued that health plans should set the copayment level (which could vary from one patient to another based on clinical characteristics) based on the benefit the intervention provides, not solely on its cost.³¹ Copayments for other forms of health care might also be adjusted for benefits. Procedures that are effective but not cost-effective in any identifiable patient population might be subject to high copayments or fixed, high coinsurance rates (percentage payments rather than fixed-dollar amounts), with no individual adjustments.

Any approach that requires different copayments for different interventions or for different patients may seem too complex to administer and understand today. Not long ago, tiered copayments for medications were criticized on the same grounds, yet they are ubiquitous today. As spending continues to rise, employers, consumers, and health plans will become more willing to explore alternatives to traditional health insurance.

In the absence of a return to heavily managed care or the adoption of novel approaches to coverage policy, commercial health plans are expected to continue to shift more costs to the insured, giving individuals a larger stake in the costs of the care they use. According to the 2000 and 2003 Henry J. Kaiser Family Foundation/Health Research and Educational Trust Surveys of Employer-Sponsored Health Benefits, out-of-pocket spending grew dramatically during the study years. Deductibles in preferred provider organization (PPO) plans grew by 57 percent (pre-

“As they bear more of the cost of care out of pocket, consumers may become the main audience for information about value.”

.....

ferred provider) and 65 percent (nonpreferred provider) in that time frame, while prescription drug copayments grew by 46 percent (preferred drugs) and 71 percent (nonpreferred drugs). The limits on out-of-pocket payments are also rising, and fixed coinsurance rates for hospitalizations and other costly forms of care will give patients further incentives to take costs into account. Cost sharing is an even more prominent feature of “self-directed” or “consumer-directed” health plans. As they bear more of the cost of care out of pocket in both absolute and relative terms, consumers may become the main audience for the information about value that cost-effectiveness analysis can provide.

Commercial health plans incorporate value considerations into benefit design in different ways as they compete for subscribers. One person could choose a plan that consistently applies cost-effectiveness criteria to coverage and other aspects of benefit design, providing a high-value, low-cost package. Another could choose a plan that either uses less restrictive cost-effectiveness criteria (that is, reimburses care with a less favorable cost-effectiveness ratio) and is broader in its coverage, or perhaps uses more traditional benefit design, charging higher premiums and using cost sharing more heavily. Although state and federal regulation limit the scope of plan variation, the market might help sort out which approaches have the greatest appeal to consumers.

Most government health insurance programs—Medicare Advantage plans are an important exception—do not compete in the same way that commercial health insurance plans do. Many, like Medicaid, serve a diverse and often vulnerable population, so extensive cost sharing is not feasible. Cost-effectiveness analysis will likely have a different role in these settings. State Medicaid programs use mechanisms such as capitation, low reimbursement rates, and restrictive coverage to control costs. Cost-effectiveness calculations undoubtedly enter into some of their benefit decisions, as happened so explicitly in Oregon. Medicaid programs will be more likely to pursue cost-effectiveness analysis as a basis for new approaches to benefit design if they face severe financial stresses. They might then conclude that their current formulas cannot control costs and yield acceptable health outcomes, and that formal cost-effectiveness analysis would provide useful guidance, particularly in setting limits on covered products and services.

Medicare is also a government program whose features are determined by legislation and regulatory interpretation. The use of cost-effectiveness in benefit design will be determined, therefore, by what is politically acceptable. Medicare differs from Medicaid in a very important respect: Because they are a large and politically powerful constituency, Medicare beneficiaries have a powerful voice in deliberations over any major change. The opposition of some Medicare beneficia-

ries, as well as several other influential constituencies, stymied Medicare officials' repeated attempts to introduce cost-effectiveness or even explicit consideration of cost in their coverage decision making. Past failures do not mean that every future effort of this kind will fail, though. The Medicare Prescription Drug, Improvement, and Modernization Act (MMA) of 2003 will have diverse effects, many of them unknowable, but among the certainties is its commitment of half a trillion dollars in additional federal funds to Medicare over the next ten years. By the end of that period, large numbers of baby boomers will have become eligible for Medicare. As the repercussions of this demographic phenomenon are felt and it becomes untenable to claim that costs can or should be ignored, the terms of debate about reform to Medicare benefit design may shift dramatically.

COST-EFFECTIVENESS ANALYSIS is a decades-old technique that has been studied more than it has been applied. Although it is not without flaws, it was never widely applied to U.S. coverage decisions because there was neither a consensus about how it should be used nor strong enough incentives to adopt it. The erosion of commercial health insurance and the growing burden of public health insurance programs may transform it from an academic curiosity to an essential tool for health care decision making.

.....
This paper is based in part on work supported by Robert Wood Johnson Foundation Grant no. 039396, National Institute on Aging Grant no. AG 17253, and the Laughlin Endowment. The author is grateful to Linda Bergthold and Jay Bhattacharya for their helpful comments.

NOTES

1. See A.M. Garber, "Evidence-Based Coverage Policy," *Health Affairs* 20, no. 5 (2001): 62–82.
2. Classic descriptions are in K.J. Arrow, "Uncertainty and the Welfare Economics of Medical Care," *American Economic Review* 53, no. 5 (1963): 941–973; and M.V. Pauly, "The Economics of Moral Hazard: Comment," *American Economic Review* 58, no. 3 (1968): 531–537.
3. See B.A. Weisbrod, "The Health Care Quadrilemma: An Essay on Technological Change, Insurance, Quality of Care and Cost Containment," *Journal of Economic Literature* 29, no. 2 (1991): 523–532; and S.T. Burner and D.R. Waldo, "National Health Expenditure Projections, 1994–2005," *Health Care Financing Review* 16, no. 4 (1995): 221–242.
4. See I. Ehrlich and G.S. Becker, "Market Insurance, Self-Insurance, and Self-Protection," *Journal of Political Economy* 80, no. 4 (1972): 623–648.
5. See discussions in L.A. Bergthold, "Medical Necessity: Do We Need It?" *Health Affairs* 14, no. 4 (1995): 180–190; S.J. Singer and L.A. Bergthold, "Prospects for Improved Decision Making about Medical Necessity," *Health Affairs* 20, no. 1 (2001): 200–206; and L.A. Bergthold et al., "Using Evidence and Cost in Managed Care Decision-Making" (Stanford, Calif.: Center for Health Policy/Center for Primary Care and Outcomes Research, Stanford University, 2002), available online at content.healthaffairs.org/cgi/content/full/hlthaffw4.284v1/DC2.
6. See D.M. Eddy, "Benefit Language: Criteria That Will Improve Quality While Reducing Costs," *Journal of the American Medical Association* 275, no. 8 (1996): 650–657; and D.M. Eddy, "Investigational Treatments: How Strict Should We Be?" *Journal of the American Medical Association* 278, no. 3 (1997): 179–185.
7. The processes Blue Cross Blue Shield uses are described in S. Gleeson, "Blue Cross and Blue Shield Association Initiatives in Technology Assessment," in *Adopting New Medical Technology*, ed. A.C. Gelijs and H.V. Dawkins (Washington: National Academies Press, 1994). The Medicare Coverage Advisory Committee (MCAC) is described in Health Care Financing Administration, "Procedures for Making Coverage Deci-

sions,” *Federal Register* 64, no. 80 (1999): 22619–22625. There are undoubtedly many reasons for the acceptance of evidence-based processes. Among them are the recognition that there are widespread variations in practice patterns that cannot be explained by patient characteristics alone and that clinical trials and other high-quality clinical studies are now common, so it seems more feasible than in the past to meet an evidence standard.

8. See U.S. Preventive Services Task Force, *Guide to Clinical Preventive Services*, 2d ed. (Baltimore: Williams and Wilkins, 1996); and Canadian Task Force on the Periodic Health Examination, “The Periodic Health Examination: Canadian Task Force on the Periodic Health Examination,” *Canadian Medical Association Journal* 121, no. 9 (1979): 1193–1254.
9. See J. Concato, N. Shah, and R.I. Horwitz, “Randomized, Controlled Trials, Observational Studies, and the Hierarchy of Research Designs,” *New England Journal of Medicine* 342, no. 25 (2000): 1887–1892; K. Benson and A.J. Hartz, “A Comparison of Observational Studies and Randomized, Controlled Trials,” *New England Journal of Medicine* 342, no. 25 (2000): 1878–1886; and M.A. Hlatky et al., “Comparison of Predictions Based on Observational Data with the Results of Randomized Controlled Clinical Trials of Coronary Artery Bypass Surgery,” *Journal of the American College of Cardiology* 11, no. 2 (1988): 237–245.
10. Cost estimates are from J.A. DiMasi, R.W. Hansen, and H.G. Grabowski, “The Price of Innovation: New Estimates of Drug Development Costs,” *Journal of Health Economics* 22, no. 3 (2003): 151–185.
11. See D. Moher, C.S. Dulberg, and G.A. Wells, “Statistical Power, Sample Size, and Their Reporting in Randomized Controlled Trials,” *Journal of the American Medical Association* 272, no. 2 (1994): 122–124.
12. Increasing the number of patients enrolled is only one of the mechanisms to ensure a large enough number of observed events, which drive the power of the trial. For example, investigators can make great efforts to improve the completeness of reporting of all health events, and they can work to minimize the number of people who drop out of a trial or are lost to follow-up. Investigators also try to enroll only those patients who are likely to adhere to all aspects of demanding protocols for participation in the trial, improving the chances that the treatment will be used properly and its effects observed. These and other aspects of trial design that tend to increase statistical power, while increasing the credibility of study results, are labor-intensive.
13. For many devices, the evidence barrier (both to approval and to the entry of new competitors) has been much lower than for pharmaceuticals, so large, well-designed randomized trials are more common for drugs than for devices.
14. Evidence from the past ten to fifteen years suggests that team care—or “disease management”—is often the most effective approach to the management of chronic diseases. Chronic disease management typically requires selecting a portfolio of diagnostic, monitoring, and treatment strategies, tailored to the individual patient, rather than simply dispensing a medication and obtaining occasional laboratory tests. Although some programs use proprietary software or are provided by dedicated disease management companies, the key features of disease management are matters of public knowledge. Because the benefits of research in these strategies are difficult for any individual firm to capture, randomized trials of disease management are less common than trials of drugs and medical devices. Furthermore, reimbursement for disease management was slow to develop, especially among fee-for-service insurers. According to the McKinsey Global Health Care Productivity study, disease management for diabetes reduced costs of care and improved outcomes. Such programs were adopted earlier in the United Kingdom than in the United States; slower U.S. adoption seemed to reflect the absence of reimbursement for components of diabetes team care. See M.N. Baily and A.M. Garber, “Health Care Productivity,” *Brookings Papers on Economic Activity: Microeconomics* (1997): 143–202.
15. See M.V. Pauly, “Should We Be Worried about High Real Medical Spending Growth in the United States?” *Health Affairs*, 8 January 2003, content.healthaffairs.org/cgi/content/abstract/hlthaff.w3.15 (7 April 2004); and Burner and Waldo, “National Health Expenditure Projections, 1994–2005.”
16. See M.C. Weinstein and W.B. Stason, “Foundations of Cost-Effectiveness Analysis for Health and Medical Practices,” *New England Journal of Medicine* 296, no. 13 (1977): 716–721; and D.M. Eddy, “Cost-Effectiveness Analysis: A Conversation with My Father,” *Journal of the American Medical Association* 267, no. 12 (1992): 1669–1675.
17. See A.M. Garber et al., “Theoretical Foundations of Cost-Effectiveness Analysis,” in *Cost-Effectiveness in Health and Medicine*, ed. M.R. Gold et al. (New York: Oxford University Press, 1996); and C.E. Phelps and A.I. Mushlin, “On the (Near) Equivalence of Cost Effectiveness and Cost Benefit Analysis,” *International Journal of Technology Assessment in Health Care* 7, no. 1 (1991): 12–21.
18. A.M. Garber and C.E. Phelps, “Economic Foundations of Cost-Effectiveness Analysis,” *Journal of Health Eco-*

- nomics 16, no. 1 (1997): 1–31.
19. S. Birch and A. Gafni, “Cost-Effectiveness Ratios: In a League of Their Own,” *Health Policy* 28, no. 2 (1994): 133–141; and M. Drummond, J. Mason, and G. Torrance, “Cost-Effectiveness League Tables: Think of the Fans,” *Health Policy* 31, no. 3 (1995): 231–238.
 20. A comprehensive listing of cost-effectiveness ratios, with comments on characteristics of the studies used to generate the numbers, can be found at the Harvard Center for Risk Analysis Web site, www.hsph.harvard.edu/cearegistry/ (6 May 2004).
 21. See J.A. Kitzhaber, “Prioritising Health Services in an Era of Limits: The Oregon Experience,” *British Medical Journal* 307, no. 6900 (1993): 373–377; D.M. Eddy, “Oregon’s Methods: Did Cost-Effectiveness Analysis Fail?” *Journal of the American Medical Association* 266, no. 15 (1991): 2135–2141; and T.O. Tengs et al., “Oregon’s Medicaid Ranking and Cost-Effectiveness: Is There Any Relationship?” *Medical Decision Making* 16, no. 2 (1996): 99–107.
 22. The definition of “cost” often determines whether an intervention is considered “expensive,” and to whom. Most pharmaceutical products have prices that are very high compared with the marginal cost of production. If “cost” in the cost-effectiveness analysis refers to the retail price, such a drug will often pass an evidence criterion more readily than a cost-effectiveness criterion. If the cost of production is high relative to the price, as would often be the case for a complex surgical procedure, it may be relatively difficult to pass an evidence criterion, since there would be so little return to an investment in studies demonstrating effectiveness. This would even be true for a procedure that was highly cost-effective.
 23. The survey was mailed to the medical directors of 346 eligible managed care plans in 49 states and the District of Columbia; the 66 percent of plans that responded were responsible for the care of 77 percent of the members of the 346 plans in the sample. The survey instrument was a closed-ended mail questionnaire consisting of forty-two questions divided into seven topic areas, including evaluation of clinical effectiveness and evaluation of cost and cost-effectiveness. Details of the survey and its methods are in Bergthold et al., “Using Evidence and Cost.”
 24. D.M. Eddy, “The Use of Evidence and Cost-Effectiveness by the Courts: How Can It Help Improve Health Care?” *Journal of Health Politics, Policy and Law* 26, no. 2 (2001): 387–408; and P.D. Jacobson and M.L. Kanna, “Cost-Effectiveness Analysis in the Courts: Recent Trends and Future Prospects,” *Journal of Health Politics, Policy and Law* 25, no. 2 (2001): 291–326.
 25. The recommendations of the federally sponsored Panel on Cost-Effectiveness in Health and Medicine appear in Gold et al., eds., *Cost-Effectiveness in Health and Medicine*.
 26. Leading studies of methods for valuing uncertainty appear in J. Mullahy and W.G. Manning, “Statistical Issues in Cost-Effectiveness Analysis,” in *Valuing Health Care: Costs, Benefits, and Effectiveness of Pharmaceuticals and Other Medical Technologies*, ed. F. Sloan (New York: Cambridge University Press, 1994); B.J. O’Brien et al., “In Search of Power and Significance: Issues in the Design and Analysis of Stochastic Cost-Effectiveness Studies in Health Care,” *Medical Care* 32, no. 2 (1994): 150–163; P. Wakker and M.P. Klaassen, “Confidence Intervals for Cost/Effectiveness Ratios,” *Health Economics* 4, no. 5 (1995): 373–381; A. Briggs and M. Sculpher, “Sensitivity Analysis in Economic Evaluation: A Review of Published Studies,” *Health Economics* 4, no. 5 (1995): 355–371; and A. Briggs, M. Sculpher, and M. Buxton, “Uncertainty in the Economic Evaluation of Health Care Technologies: The Role of Sensitivity Analysis,” *Health Economics* 3, no. 2 (1994): 95–104.
 27. Bergthold et al., “Using Evidence and Cost.”
 28. See N. Daniels and J.E. Sabin, *Setting Limits Fairly: Can We Learn to Share Medical Resources?* (New York: Oxford University Press, 2002).
 29. M.V. Pauly and P.J. Held, “Benign Moral Hazard and the Cost-Effectiveness Analysis of Insurance Coverage,” *Journal of Health Economics* 9, no. 4 (1990): 447–461.
 30. B. Motheral and K. Fairman, “Effect of a Three-Tier Prescription Copay on Pharmaceutical and Other Medical Utilization,” *Medical Care* 39, no. 12 (2001): 1293–1304.
 31. A.M. Fendrick et al., “A Benefit-Based Copay for Prescription Drugs: Patient Contribution Based on Total Benefits, Not Drug Acquisition Cost,” *American Journal of Managed Care* 7, no. 9 (2001): 861–867.

Developing A Center For Comparative Effectiveness Information

High-level consideration of a new U.S. entity to assist in developing evidence for decision making based on effectiveness.

by **Gail R. Wilensky**

ABSTRACT: Interest in objective, credible comparative clinical effectiveness information has been growing in the United States, both by those who support competitive behavior in health care and by those who support administered pricing. The Medicare drug benefit has heightened interest in better information, although the potential payoff is even greater for medical procedures than for drugs, since procedures account for more of the health care dollar. Careful consideration needs to be given regarding the appropriate structure, placement, financing, and function of an agency devoted to comparative effectiveness if it is to achieve its objective: a mechanism to support better decision making in health care. [*Health Affairs* 25 (2006): w572–w585 (published online 7 November 2006; 10.1377/hlthaff.25.w572)]

AMONG THE MANY CONTROVERSIAL FEATURES of the Medicare Prescription Drug, Improvement, and Modernization Act (MMA) of 2003, none has caused more dissension than the provision stating that the secretary of health and human services (HHS) “may not interfere” in the negotiations between drug manufacturers and plans regarding prices or formularies. With the full drug benefit just being implemented, it is far too early to know whether this provision will be able to withstand the inevitable future pressures to rein in Medicare Part D spending. The prohibition on government from using administered pricing or the full weight of Medicare’s purchasing power to drive down prescription drug prices in the context of the new drug benefit has renewed interest in having better information available on the relative clinical effectiveness and cost-effectiveness of alternative therapeutic treatments as a strategy to moderate spending. What has been particularly promising, at a political level, is that interest in clinical effectiveness data is present both among those who support administered pricing and among those who oppose it.

Interest in better comparative information is not new.¹ Indeed, it seems to in-

Gail Wilensky (gwilensky@projecthope.org), an economist and a former administrator of HCFA, is a senior fellow at Project HOPE in Bethesda, Maryland.

crease whenever there is renewed interest in competitive behavior in health care. But interest in objective, credible comparative effectiveness information has not been limited to those with free-market interests. Several European Union (EU) health systems rely on comparative effectiveness information in their decision making regarding drug coverage.²

Most of the interest in comparative effectiveness information has focused on pharmaceuticals, although it has sometimes been directed toward device coverage decisions as well. This focus on pharmaceuticals has also occurred in the United States, probably because the larger share of out-of-pocket spending has produced a greater public awareness of spending on drugs, although it might also be because there is greater uniformity among drugs than among other health services.

Although drugs and medical devices are important areas of health care, they are not the only areas that could benefit from comparative effectiveness information. In fact, since drug spending accounts for only about ten cents of each health care dollar, the potential payoff for better decision making is even greater in other areas of health care, particularly medical procedures. But because of the relatively rapid rise in prescription drug spending earlier in the decade and the political “third rail” that pharmaceutical manufacturers have long represented, the interest in good comparative data is especially strong for prescription drugs.

Finding mechanisms that will help the United States make better coverage and spending decisions is critical. The United States spends far more per capita than other developed countries: It spent \$5,267 per capita in 2002, compared with only \$3,446 per capita in the next-highest-spending country, Switzerland.³ There are many reasons that explain the higher U.S. levels of spending, such as higher incomes and greater system capacity. However, the increased spending does not appear to be producing uniformly better outcomes than other countries experience. More importantly, the long-term U.S. spending growth on health care will present many challenges if it continues indefinitely. On average, spending on health care has increased about 2.5 percent faster than the economy. If this growth rate were to continue until 2045, the federal share of spending on health care would claim the same share of the economy as the total federal budget today (not counting payments on interest on the national debt). Finding politically acceptable ways to reduce the long-term growth rate in health care spending will be difficult. Within this context, learning how to “spend smarter,” rather than relying on arbitrary mechanisms to limit spending, begins to look very appealing.

The focus of this paper is to assess the various options regarding the structure, placement, financing, and functions of an agency devoted to comparative (clinical) effectiveness assessment. Pros and cons of the major options are presented, along with a judgment about which strategies would be most likely to be acceptable to the most important stakeholders. A brief discussion of how other countries have handled decisions about the placement and financing of comparative effectiveness centers is also included.

Current Practice/Current Law

Under current law, drugs and devices need to obtain approval from the Food and Drug Administration (FDA) with regard to safety and efficacy before they can be marketed. Although data from clinical trials serve the needs of the FDA approval process, they generally do not provide information that is useful for comparative effectiveness purposes. FDA clinical trials typically focus on efficacy relative to placebo, whereas analyses of comparative effectiveness would require information on the relevant alternatives to the new therapy, device, or procedure.

■ **Private-sector efforts.** Private-sector entities have attempted to assess comparative clinical effectiveness as part of their coverage decision processes. One of the pioneers is the Technology Evaluation Center (TEC) established by the Blue Cross Blue Shield Association (BCBSA) in 1985. Its technology assessments rely on comprehensive reviews of existing clinical evidence and focus on the clinical effectiveness and appropriateness of a specific medical procedure, device, or drug. Its clients include other private-sector payers as well as the Centers for Medicare and Medicaid Services (CMS), although the CMS is prohibited by law from making coverage decisions based on drugs' or devices' relative effectiveness.

Health plans and hospitals have implicitly performed such assessments as part of their formulary deliberations, although they have been criticized for using procedures that lack transparency and rigor. The multipayer U.S. system has probably contributed to slower growth in the use of formularies compared with single-payer countries, although formularies have now become an important part of the pharmaceutical benefit management (PBM) industry. Late in the 1990s, Regence BlueShield (Seattle) began asking pharmaceutical companies to submit standardized packages of clinical and economic evidence as part of their formulary design deliberations.⁴ The Academy of Managed Care Pharmacy (AMCP) has endorsed a set of guidelines for providing similar information and has encouraged health plans to use its guidelines. Unlike the clinical information associated with FDA studies, these guidelines provide detailed information on a drug's economic value relative to alternative therapies, in addition to the drug's safety and efficacy. According to a 2003 Bruckner Group survey, managed care organizations, representing approximately 65 percent of covered lives, had "officially adopted" the AMCP guidelines, and "nearly all players" were using them "to some extent."⁵

■ **Public-sector efforts.** Several federal agencies are also involved to some degree in promoting or assessing clinical effectiveness, although most of them rely on systematic reviews of existing research rather than funding new prospective studies of comparative effectiveness. The agency that has been specifically directed to address issues of comparative effectiveness as part of MMA is the Agency for Healthcare Research and Quality (AHRQ). AHRQ is the only federal agency whose primary mission is both to support and to conduct health services research, including comparative effectiveness, although it is not the only agency that does health services research. AcademyHealth, the professional association of health services research-

ers, recently estimated that AHRQ accounted for only a little more than 20 percent of the \$1.5 billion in federal funds spent for health services research.⁶ The agency with the largest health services research spending is thought to be the National Institutes of Health (NIH), although its funds for health services research represent only a small fraction of its budget.⁷

Section 1013 of MMA authorized \$50 million and appropriated \$15 million in fiscal year 2004 for AHRQ to conduct research and set priorities relating to improving outcomes as well as the clinical effectiveness and appropriateness of health services, including prescription drugs. There is no provision for the use of cost-effectiveness information in MMA, which presumably reflects continued sensitivity to the use of that type of analysis in Medicare's decision making. The law also requires that the secretary of HHS establish an initial list of priorities, complete the evaluation of the initial priorities, and disseminate the research findings within eighteen months, and then develop strategies. Thus, in existing law, AHRQ is clearly envisioned as the site of future research and funding for a center of comparative effectiveness information. However, many questions remain as to whether AHRQ would be the best placement for this effort, and what the alternatives would be. I now turn to an examination of other countries' experiences, in an effort to assist U.S. policymakers in their deliberations.

Experiences In Other Countries

Many countries have centralized the process for performing comparative clinical and economic assessments. These agencies typically exist as part of their governments, which is not surprising, since these are all countries with centralized payer systems. They do differ in important respects, however, particularly with regard to the mandatory nature of the guidelines.

■ **Australia.** Australia was an early adopter of cost-effectiveness as a requirement for a drug's inclusion on the national formulary. At some level, there has been centralized review since the Pharmaceutical Benefits Scheme (PBS) was first established in the 1950s and, with it, the Pharmaceutical Benefits Advisory Committee (PBAC). The health minister is directly responsible for coverage decisions but cannot list a drug without a positive recommendation from the PBAC. A separate organization negotiates the listing price with the manufacturer. There is no formal process for appeal. The final decision is made public, but not the rationale for the decision or the relevant clinical or cost-effectiveness data.⁸

■ **United Kingdom.** The U.K. National Institute for Health and Clinical Excellence (NICE) initiates and conducts its own evaluations, unlike in Australia, where the government body reviews and interprets the data and economic analyses submitted by the drug companies. NICE reviews all types of medical technologies, including drugs, that are likely to have a sizable health or budgetary impact or otherwise to be controversial. The actual evaluation and assessment of the technology is done by a technical committee called the Technology Appraisal Committee (TAC),

which includes a large group of academic experts, clinicians, patient advocates, and industry representatives. An academic group does the actual assessment; the TAC reviews it and publishes a recommendation, which can be appealed. The recommendation is then submitted to NICE. The NICE appraisal process has been estimated to take a year or more. NICE is not bound by the TAC's recommendations; however, drugs recommended by NICE are required to be funded by the government.⁹

■ **Canada.** Canada only recently (2003) introduced a coordinated process for reviewing drug coverage applications, the Common Drug Review (CDR). The CDR reviews only new chemical entities and new combination products, unlike NICE, which reviews some existing entities under limited circumstances. The reviews, which are not binding, are done for government drug plans in all provinces other than Quebec. An advisory committee of experts, appointed by the deputy ministers of health from each province, makes recommendations to the CDR based on assessments by reviewers, who can be either internal or external to the CDR. The advisory committee sends the initial recommendation to the manufacturer, which can appeal the decision. A summary of the recommendation and the rationale is posted, although neither the data nor the assessment is made public.¹⁰

■ **Germany.** Germany adopted a different model in 2003 when it established its Institute for Quality and Efficiency (IQWiG). The Federal Joint Committee that administers health services in Germany established the institute, which is federally funded but governed by a private foundation. The institute's governance structure involves a twelve-member foundation board, a five-member board of directors, and a thirty-member board of trustees that is reflective of its stakeholders, which acts as an advisory committee. The board also has a scientific advisory board comprising up to a dozen members. The institute is responsible for evaluating the use, quality, and efficiency of drugs and services in Germany and also evaluates clinical practice guidelines for the epidemiologically most important diseases.¹¹

The Federal Institute for Drugs and Medical Devices (BfArM) is responsible for authorizing pharmaceuticals, but authorizing them doesn't necessarily mean having them reimbursed by the statutory health insurance companies. That depends on the decision taken by the Federal Joint Committee after evaluating reports by the IQWiG. The evidence usually requires data from randomized controlled trials and a demonstrated impact of patient-relevant outcomes. The committee defines uniform pharmaceutical reimbursement for agents with similar effects—that is, within the same reference class.

Function, Placement, And Financing Options

The U.S. reliance on a multipayer health system makes the function, placement, and financing of a center for comparative effectiveness more complex, at least politically, than in the preceding countries. Unless all major payers regard the placement and financing of such a center as being consistent with the production of objective and unbiased data, the information it produces will be of little use.

The primary function of this center would be to provide an independent assessment of the comparative effectiveness of alternative therapies and procedures for use by various payers and to provide supporting information so that both patients and providers can improve their decision making. Unlike the work being done by NICE or centers in other countries, or by the BCBSA and other private-sector U.S. organizations, this center would fund prospective trials on key questions for which comparative effectiveness evidence was found missing, in addition to funding systematic reviews of existing research. This feature also distinguishes it from the work being done by AHRQ, which primarily involves systematic reviews done by its funded Evidence-based Practice Centers (EPCs) and retrospective analysis of administrative electronic health record (EHR) data. The review of existing research is an important function for the various private-sector organizations to continue doing. However, it is the production of new information, done in house by the center or by contract with various academic or clinical institutions, and the assembly and availability of known information about comparative effectiveness that will be the focus of the center being envisioned.

The placement of such a center should be judged by whether the data produced will be perceived as objective and credible, represent minimal or no conflict of interest, and be perceived as being insulated from stakeholder pressures. Financing options should be judged primarily in terms of financial sustainability and stability and, perhaps to a lesser extent, equity and political acceptability.

■ **AcademyHealth report.** AcademyHealth released a report last year on the placement, coordination, and funding of health services research within the federal government, which also includes a discussion about the establishment and placement of a comparative effectiveness center.¹² This report focuses on strategies that will strengthen health services research as a field, which is not regarded as a relevant criterion for evaluating a comparative effectiveness center. It makes several recommendations, including the formation of a separate agency to serve as the lead agency for health services research (which is the function AHRQ now serves), increased funding for health services research, and increased coordination of health services research within HHS and across the federal government.

The AcademyHealth report lays out several options for the placement of a comparative effectiveness center: placing it within AHRQ; having AHRQ oversee the comparative effectiveness studies and establishing a Federally Funded Research and Development Center (FFRDC) to undertake research syntheses of comparative effectiveness findings; creating a new quasi-governmental entity for comparative effectiveness research; and reconstituting AHRQ as a quasi-governmental entity that would include comparative effectiveness research.

Although AcademyHealth does not take a position on which option is preferable, the report emphasizes the importance of maintaining a strong linkage with the lead agency for health services research and the need for the findings to be based on scientific evidence and to be shielded from political or budgetary factors.

The latter is consistent with the criteria used in this paper to judge the various placement options.¹³ The importance of improving coordination within government of health services research in general and between the lead agency for this type of research and a comparative effectiveness center seems clear and obvious.

■ **Coordination.** Linkages and coordination between AHRQ and a new comparative effectiveness center could be accomplished in a variety of ways, such as by having the center be part of AHRQ or by having a formal or informal reporting relationship between the two. Better coordination of health services research both within HHS and between HHS and the rest of the federal government is important in its own right. The creation of a comparative effectiveness center, which by its nature will involve relationships with the FDA and perhaps the National Center for Health Statistics (NCHS) as well, will only increase the importance of such coordination. FDA approval, for example, is likely to trigger consideration for a drug's inclusion in a comparative effectiveness analysis.

■ **Preserving AHRQ.** Early on, some of the informal discussions and interviews involved in the preparation of the AcademyHealth report, in which I participated as a member of the committee that prepared it, focused on the potential risk that a comparative effectiveness center separate from AHRQ could present to AHRQ's integrity and stability. Although this might well be the case if the comparative effectiveness center were of a very modest size—as, for example, if it were funded at the \$50 million level provided for comparative effectiveness research in MMA and made separate from and therefore rival to AHRQ—it seems less likely to be an issue if a separate center were funded at a more appropriate level.

A multibillion-dollar comparative effectiveness center would make it clear that the new center's purpose is to provide credible, objective information on comparative effectiveness, allowing AHRQ to maintain its role as the place of traditional health services research, including analyses that might make use of data provided by the center. The center's size would reflect the need to sponsor new research and produce new data on comparative clinical effectiveness for the many new and existing technologies that have come on the market over the past several decades. Even at a multibillion-dollar annual level, research efforts on comparative effectiveness would need to be prioritized according to some agreed-upon principles.

Four Placement Options

■ **Option 1: placement within AHRQ.** For reasons already outlined, the most obvious choice for the placement of a comparative effectiveness center is AHRQ. AHRQ, as currently configured, could be augmented by the establishment of an independent external board, along with a panel of experts to advise on research priorities and to provide oversight for the monitoring of research contracts and the dissemination of results.¹⁴ An advantage of this approach is that it would provide a mechanism for the private sector to participate in establishing a comparative effectiveness research agenda. An independent, external board might also improve the

credibility of the findings. It would make AHRQ a strong partner for other federal agencies that would be interested in such research and would also obviously increase AHRQ's prominence and visibility, an issue that is important to some researchers but not central to the current consideration.

However, the increased prominence and visibility that a comparative effectiveness function would bring to AHRQ is a disadvantage as well, both to the agency and to the concept of a comparative effectiveness center. The center's findings might anger various stakeholders affected by the findings, who, in turn, could use the political process to threaten the continued existence of the agency that produces the "threatening" material. There is also a question of whether information produced by a governmental agency will be perceived as being objective and credible. To some extent, credibility will be affected by the even-handedness with which the process is carried out—that is, whether the areas chosen for evaluation are a good reflection of disease burden, financial burden, and scientific opportunity. Informal discussions with members of medical academe suggest that they might find it difficult to regard findings produced by a governmental organization as being other than political, whether or not that is the case. The distrust would be especially strong if the governmental agency were also the payer, like the CMS; even so, there appears now to be substantial mistrust of government's motives, with questions being raised about appointments to various scientific and medically related committees. Finally, placement within AHRQ would limit opportunities for the private sector to participate in funding. This might not only take away a possible funding source but also might mean less of a commitment by the private sector to the success of such a center.

■ **Option 2: placement elsewhere within HHS, as a new or existing entity.** A second alternative is to establish a new center or board elsewhere within HHS. In principle, such an entity could also be established outside of HHS, elsewhere within the executive branch, but there is little obvious advantage to such an arrangement.

Discussions with people inside and outside of government suggest little enthusiasm for placing a comparative effectiveness center within the NIH. Although it conducts a large dollar amount of health services research in its various centers, such research is clearly not a primary focus of any of its centers, and there is little reason to believe that comparative effectiveness would be regarded as comparable in importance to other NIH center activities. The NIH, however, has enjoyed a reputation as being highly objective; to the extent that this spilled over to work on comparative effectiveness, it would make the NIH attractive as a site.

Another alternative is to establish a moderate-size board (five to fifteen members) within HHS, or else an entirely new agency within HHS, that would be responsible for comparative effectiveness information. One advantage of a new entity is that it could be created on the model of the Federal Reserve Board—with commissioners or members having fixed, multiyear, staggered terms. Its membership therefore would not be under the control of any one president and would be

less likely to be regarded as political. Discussions with people within the industry and also within the academic and not-for-profit worlds suggest that this concept is somewhat attractive.

It has some disadvantages as well. The most obvious is that it would require the establishment of a new entity. Placing it within government also would limit any opportunities for private-sector funding, which would affect both the potential funding and, possibly, the private sector's commitment to the center. The independence of the board might lessen suspicions associated with the center's findings, although the information produced would still be associated with a government entity. To the extent that such an association results in distrust, it might still be present in this altered format. Alternatively, the more separate the entity is in government, the easier it is to target for pressure. Given prior experiences in health services research in particular, any entity that is part of government is likely to be hugely pressured by industry.

■ **Option 3: placement within a quasi-governmental entity.** A variety of quasi-governmental structures could house a comparative effectiveness center. The Institute of Medicine (IOM) is frequently thought of when the term "quasi-governmental structure" is mentioned, but there are a variety of other models to consider as well. Among two particularly interesting models, in part because they tend to be more closely associated with federal agencies, are Federally Funded Research and Development Centers (FFRDCs) and Public Foundations.

IOM/NRC. The IOM, either by itself or in conjunction with the National Research Council (NRC), is one obvious model; in fact, the IOM has expressed a willingness to serve as a clearinghouse for comparative effectiveness information and has created a senior-level Roundtable on Evidence-based Medicine to issues and feasibility.¹⁵ A rationale for adding the NRC is that having engineering assistance would be desirable and also that it would avoid any appearance of having the entity responsible for comparative effectiveness oversight being captive to the physician community.

A primary advantage of having the IOM serve this function is that it would provide for a trusted and independent intermediary to supervise the use of funds as well as the reporting and translation functions while making use of existing capacity in government for research contract management. In addition, the IOM has generally been highly regarded by both industry and government, and it might also be able to generate private funds, from industry and foundations.

Several disadvantages could be associated with housing this activity in the IOM. There is some question as to whether the IOM can act in a timely way. Although the IOM has produced several reports within a matter of months, mostly on very narrow and focused topics, its consensus process can be cumbersome.¹⁶ There is also a question of whether all administrations would be equally comfortable having this function housed in the IOM. Finally, it is unclear whether Congress would be willing to fund most of the cost of this enterprise if it were not

housed directly in government, in part because clear accountability is lacking when an activity is housed outside of government. This is not now an issue for the IOM because it is mostly funded on a project basis, albeit heavily by government, which provides its own type of accountability.

FFRDC. A different type of quasi-governmental entity that circumvents some of the issues raised by the IOM model is the FFRDC. FFRDCs are generally linked to a federal agency; in this case, the most obvious would be AHRQ. AHRQ would commission new research on comparative effectiveness. The FFRDC would synthesize existing research, including the newly generated research resulting from AHRQ contracts; could make recommendations and assessments concerning the findings; and could determine how the findings would be disseminated.

FFRDCs usually receive most of their funding from federal agencies and need to be sponsored by an executive-branch agency, which monitors their funds. They typically operate as private, not-for-profit organizations and by law can only accept 30 percent of their funding from private sources, although depending on the size of the FFRDC, 30 percent could represent a sizable contribution from the private sector.

There are several advantages to the FFRDC model. FFRDCs clearly involve the private sector and therefore could provide some additional private buy-in. They also allow the private sector to finance work on comparative effectiveness. At the same time, they are directly linked to the federal government, which might be important if the government is assumed to be providing most of the funding.

Some of the disadvantages of FFRDCs are those associated with the uncertainties of a mechanism that is used only occasionally and under limited circumstances. Supposedly they are only used when they are associated with work that cannot be accomplished by existing government or contractor resources. It is also unclear whether the private sector would be assured that the research being commissioned by the federal agency was without political influence and, in general, whether there is enough arm's-length distance between the FFRDC and its agency sponsor to provide assurance of objectivity.

The FFRDC linked to AHRQ represents an interesting compromise between maintaining the comparative effective function in government and the use of a more independent type of quasi-government entity, such as the IOM. However, to the extent that the FFRDC and AHRQ were viewed as being too closely related, one could imagine this combination representing the worst of all worlds: complications of a separate organization responsible for some of the comparative effectiveness activities, suspicions about the independence and credibility of the material produced, and increased exposure and potential threats to AHRQ.

Public Foundation. An alternative to the FFRDC model is the Public Foundation (PF). PFs are not-for-profit organizations that act as a type of public charity. Their primary purpose is to make grants. Part of the requirement for being established as a public charity and for receiving tax-exempt status from the Internal Revenue

Service (IRS) is that PFs have to seek money from diverse sources, and at least one-third of their money has to be from the general public. Also, unlike FFRDCs, PFs can act completely independently from their parent organizations. In contrast to FFRDCs, which are more limited in scope and supposedly restricted to work that cannot be done by existing government agencies or contractor resources, PFs usually have broader missions.¹⁷

The advantages and disadvantages of PFs in general are the same as for the FFRDC, but the PF seems a slightly less relevant model. FFRDCs are more likely to be associated with research, such as the National Defense Research Institute, which is part of RAND. PFs have been used by the Centers for Disease Control and Prevention (CDC) and the NIH, but they are mostly used to raise money and make grants to other institutions.

■ **Option 4: placement within the private sector.** Maintaining the comparative effectiveness center function within HHS or elsewhere in the executive branch represents one extreme. Locating such a center within the private sector represents the other. In principle, a comparative effectiveness center could be a freestanding institution or one affiliated with a university or other entity. Presumably, it would be a not-for-profit institution, committed to following certain federal guidelines regarding transparency and availability of data.

The advantage of locating the center in the private sector is that this would minimize any concern that the outcomes reflected political pressure from the government. It would not be subject to any of the personnel or contract constraints of government and would provide maximum opportunities for private-sector participation, in both funding and substantive involvement.

There are also a variety of disadvantages to this approach. Perhaps the most serious is that the government might be unwilling to be the primary funder if the center were located in the private sector because of concerns about control and accountability. A second, potentially serious disadvantage is that the lack of government involvement or oversight might raise questions about the objectivity of the findings and leave the center subject to charges of being captured by industry. This concern might be alleviated by the use of an external board of government and academic experts providing governance to the center. A related disadvantage involves issues of transparency and whether it would be possible to require the same level of transparency in a private-sector activity as in a governmental or quasi-governmental activity and whether, without transparency, the findings would be regarded as objective.

The notion of a comparative effectiveness center in the private sector has not been regarded as a serious option to date, probably because the complete removal of government would raise too many questions regarding objectivity or capture by industry. It is not surprising that government or academe would not find this type of structure interesting, but it is surprising that there does not seem to be strong support in industry, either.

Financing A Comparative Effectiveness Center

The most obvious and direct way to finance at least the public portion of a comparative effectiveness center is through a direct appropriation by Congress. The rationale for public funding is that information is a public good, in the most classic sense: that is, it is not excludable and nonrival in its consumption. Economic theory argues that goods or services that meet this definition will be underproduced by the private sector and should therefore be financed by government.

The theoretical argument for traditional public financing of a comparative effectiveness center seems clear enough. However, the vagaries of relying on an adequate annual appropriation suggest that a different financing mechanism might be preferable. The most frequently considered alternative is financing by the Medicare Trust Funds. This would present a potentially more stable funding source, since it would not be subject to annual review, but it is difficult to imagine political agreement on this approach in the current political environment. However, it might be possible to rely on the Medicare Trust Funds to finance a portion of the costs, since Medicare would obviously be an important beneficiary and could allow decision making about reimbursement that could result in future Medicare savings. This, of course, assumes that Medicare would be empowered to use comparative effectiveness as one of the criteria in setting reimbursement, although now that private-sector entities are administering the Part D drug benefit, better information on comparative effectiveness could produce savings even without additional authority being granted.

The private sector might also be willing to underwrite some of the costs. One option is to rely on voluntary contributions, although since information can be regarded as a public good, economic theory suggests that this is not likely to succeed. One of the attributes of a public good is that once the good or service—or in this case, the information—is available, it will or should be available to all without charge. However, that encourages “free riders.”

A small charge or fee could be assessed on all users, providers, or suppliers of health care services or on health plans. This charge should be broadly based, rather than limited to one sector or a portion of the payers and definitely not limited to a small sector like pharmaceuticals or devices. A user-fee system of financing has supplemented FDA funding and helped reduce the time required for the approval process; this could be regarded as a precedent for partial funding of a center on comparative effectiveness. In sum, while the preferred financing is general-fund financing, it is possible to imagine a combined funding strategy that would be less susceptible to the vagaries of the appropriations process.

Concluding Comments

Better information about the comparative effectiveness of various medical strategies and procedures might not, in itself, lead to better decision making in health care unless there is also a major change in financial incentives. However,

better information is an important component of such an outcome. Most countries have limited their focus in comparative effectiveness information to drugs and medical devices, but since the majority of health care spending occurs for medical and surgical procedures, these areas also must be included in the design of a comparative effectiveness center.

Other countries that have established cost- or clinical effectiveness as a requirement for coverage or reimbursement, or both, have centralized the process. Countries differ in terms of whether the recommendations that come out of such groups are mandatory or advisory, the transparency of the process, and whether the results are subject to appeal. The appropriate function, structure, placement, and financing of a comparative effectiveness center in the United States will need to reflect this country's political sensitivities and the unique public/private structure that has developed here. The function of a comparative effectiveness center in the United States would be to provide credible, objective information on the comparative effectiveness of alternative therapies and technologies, not to make centralized coverage decisions. The information would be available to any payer and also could be used for better decision making by patients and providers. To date, the United States has been unwilling to include statutory language that would allow cost-effectiveness information to be used in making coverage decisions even in large public programs such as Medicare, nor has there been any inclination to make all payers use the same coverage or reimbursement decisions. Although continued increases in health care spending that are much greater than the economy's growth could change this history, it does not appear likely anytime soon.

The obvious choices of where to place a comparative effectiveness center are to locate it in HHS, either as a part of AHRQ, which already carries out some of these functions although to a rather limited extent, or elsewhere in the department. Alternatively, a variety of quasi-governmental or even private-sector options are possible. On balance, the placement of a comparative effectiveness center within a quasi-governmental entity seems the most attractive. The idea of establishing an FFRDC, perhaps linked to AHRQ or to a newly established board in HHS, is intriguing. FFRDCs are still relatively unknown, particularly in the health services research world, and more attention would need to be given to their administrative complexities or limitations to their use. The notion of attaching the FFRDC to a new board within HHS, which had a membership appointed in staggered terms, is also very appealing, although its appeal would have to be weighed against the advantage of linking to AHRQ, which could be generating some of the research needed by the FFRDC.

If the FFRDC model proved too complex or an otherwise undesirable administrative entity, more exploration of the IOM, particularly paired with the NRC, should be undertaken. Some believe that the IOM is too cumbersome, largely because of its review process, but it has shown itself to be capable of being time-sensitive in at least some circumstances. Having a new entity that reports to the

IOM but is not of the IOM/NRC, per se, might allay some of these concerns.

Despite many different views, there is widespread agreement on the attributes that need to be associated with a comparative effectiveness center: objectivity in the selection of what is studied, credibility in the findings, and independence—from political pressures generated either by government or by private-sector stakeholders. How best to achieve this set of outcomes, not surprisingly, differs in the eyes of different beholders.

.....
This paper was supported by a grant from the Commonwealth Fund. The author thanks Barbara Cooper and Stuart Guterman of the Commonwealth Fund for their support, Philip Musgrove and Andrea Zuercher of Health Affairs for their editing assistance, and Sean Tunis for his substantive advice; she also recognizes the thoughtful comments of the anonymous reviewers.

NOTES

1. U.E. Reinhardt, "An Information Infrastructure for the Pharmaceutical Market," *Health Affairs* 23, no. 1 (2004): 107–112; and P.M. Ellwood, A.C. Enthoven, and L. Etheredge, "The Jackson Hole Initiatives for a Twenty-first Century American Health Care System," *Health Economics* 1, no. 3 (1992): 149–168.
2. See, for example, A. Mitchell, "Update and Evaluation of Australian Guidelines: Government Perspective," *Medical Care* 34, no. 12 Supp. (1996): DS216–DS225; and P.J. Neumann, "Evidence-based and Value-based Formulary Guidelines," *Health Affairs* 23, no. 1 (2004): 124–134.
3. G.F. Anderson et al., "Health Spending in the United States and the Rest of the Industrialized World," *Health Affairs* 24, no. 4 (2005): 903–914.
4. Regence Group, "Guidelines for Submission of Clinical and Economic Data Suggesting Formalizing Considerations," Version 1.2 (Seattle: Regence Group, 1997); Regence Group, "Familiarizing Submission Guidelines," Version 2.0 (Seattle: Regence Group, 2002); and Neumann, "Evidence-based and Value-based Formulary Guidelines."
5. D. Balekdjian and M. Russo, "Managed Care Mandate: Show Us the Value," *Pharmaceutical Executive*, September 2003, <http://www.highbeam.com/doc/1G1-108450162.html> (accessed 17 October 2006).
6. AcademyHealth, *Placement, Coordination, and Funding of Health Services Research within the Federal Government*, September 2005, <http://www.academyhealth.org/publications/placementreport.pdf> (accessed 19 September 2006).
7. *Ibid.*
8. S.G. Morgan et al., "Centralized Drug Review Processes in Australia, Canada, New Zealand, and the United Kingdom," *Health Affairs* 25, no. 2 (2006): 337–347.
9. *Ibid.*
10. *Ibid.*
11. *Ibid.*
12. AcademyHealth, *Placement, Coordination, and Funding*.
13. *Ibid.*
14. *Ibid.*
15. Institute of Medicine, "Roundtable on Evidence-based Medicine," 27 July 2006, <http://www.iom.edu/?id=28186&redirect=0> (accessed 19 September 2006).
16. Examples of quick-turnaround reports from the IOM include the following: (1) Vaccinations and Sudden Unexpected Death in Infancy—meeting date: 28–29 October 2002, release date: 6 October 2003; (2) Vaccine and Autism—meeting date: 9 February 2004, release date: 17 May 2004; and (3) Review of the HIVNET 012 Perinatal HIV Prevention Study—meeting date: September 2004, release date: 7 April 2005, Sponsor: NIH.
17. AcademyHealth, *Placement, Coordination, and Funding*.