ENGELBERG CENTER for Health Care Reform at BROOKINGS

THE HAMILTON PROJECT

JUNE 2009

Implementing Comparative Effectiveness Research: Priorities, Methods, and Impact

BROOKINGS

The Engelberg Center for Health Care Reform provides data-driven, practical policy solutions that will foster high-quality, innovative care – care that is both more affordable and more effective at actually improving patient health.

Our nation may face no more important domestic policy challenge than the much-needed reform of our health care system. Meaningful improvements in health care will depend on dramatically reforming the structure of the delivery system at large – a challenge that is made more difficult by current fiscal realities. Still, circumstances should not encourage inaction; they should instead underscore the need for strategies that both address gaps in quality and efficiency and further the advances that have already been made in health status and medical innovation.

The Engelberg Center goes beyond merely studying the issues and making policy recommendations. We promote the broad-based exchange of ideas to develop consensus around practical steps, and then go one step further by providing technical support for collaborative work among a wide range of health care stakeholders and actual implementation. Our focus is on key priority areas that are critical to the kind of reform that will improve not just the health care system, but the health of individual patients.

The Hamilton Project seeks to advance America's promise of opportunity, prosperity, and growth. The Project's economic strategy reflects a judgment that long-term prosperity is best achieved by making economic growth broad-based, by enhancing individual economic security, and by embracing a role for effective government in making needed public investments. Our strategy – strikingly different from the theories driving economic policy in recent years – calls for fiscal discipline and for increased public investment in key growth-enhancing areas. The Project will put forward innovative policy ideas from leading economic thinkers throughout the United States – ideas based on experience and evidence, not ideology and doctrine – to introduce new, sometimes controversial, policy options into the national debate with the goal of improving our country's economic policy.

The Project is named after Alexander Hamilton, the nation's first treasury secretary, who laid the foundation for the modern American economy. Consistent with the guiding principles of the Project, Hamilton stood for sound fiscal policy, believed that broad-based opportunity for advancement would drive American economic growth, and recognized that "prudent aids and encouragements on the part of government" are necessary to enhance and guide market forces.

ENGELBERG CENTER for Health Care Reform at BROOKINGS

THE HAMILTON PROJECT

Implementing Comparative Effectiveness Research: Priorities, Methods, and Impact

NOTE: These discussion papers are proposals from the authors. The authors are invited to express their own ideas in discussion papers, whether or not the Engelberg Center for Health Care Reform staff and The Hamilton Project's staff or advisory board agrees with the specific proposals. These discussion papers are offered in that spirit.

BROOKINGS

JUNE 2009

CONTENTS

Background on Comparative Effectiveness Research

In the American Recovery and Reinvestment Act of 2009 (ARRA), Congress and the President invested $1.1 billion in the conduct and dissemination of comparative effectiveness research (CER). A total of $700 million was allocated to the Agency for Healthcare Research and Quality (AHRQ), $400 million of which was to be transferred to the National Institutes of Health (NIH) to conduct CER. The remaining $400 million was allocated to the Secretary of the Department of Health and Human Services (HHS) for discretionary investments in CER.

The bill also established a basic structure to govern the use of these resources. The new Federal Coordinating Council for Comparative Effectiveness Research (FCC-CER) is comprised of federal employees in health-related agencies appointed by the HHS Secretary. Its purpose is to advise Congress and the President on necessary infrastructure and federal spending for CER, coordinate the CER activities of various federal agencies, and produce annual reports on these topics. Finally, ARRA also directed the HHS Secretary to enter into a contract with the Institute of Medicine to produce recommendations on national priorities for CER by June 30, 2009. The Secretary will consider these recommendations and others put forward by the FCC-CER when determining how resources should be allocated.

Implementation of the CER provisions under ARRA is underway. The Institute of Medicine will deliver its report by the end of June 2009, and the FCC-CER has been appointed. It recently released its draft definition of CER:[1]

Comparative effectiveness research *is the conduct and synthesis of systematic research comparing different interventions and strategies to prevent, diagnose, treat and monitor health conditions. The purpose of this research is to inform patients, providers, and decision-makers, responding to their expressed needs, about which interventions are most effective for which patients under specific circumstances. To provide this information, comparative effectiveness research must assess a comprehensive array of health-related outcomes for diverse patient populations. Defined interventions compared may include medications, procedures, medical and assistive devices and technologies, behavioral change strategies, and delivery system interventions. This research necessitates the development, expansion, and use of a variety of data sources and methods to assess comparative effectiveness.*

Research meeting this definition is far from unprecedented. Considerable research already being conducted with support from AHRQ, NIH, Veterans Administration, Centers for Medicare and Medicaid Services, Food and Drug Administration, and the private sector is intended to provide evidence on which treatments work best for particular types of patients, and which policies can promote those treatments. Still, large gaps remain both in the availability and the use of relevant comparative effectiveness research.

1. Department of Health and Human Services"Draft Definition of Comparative Effectiveness Research for the Federal Coordinating Council." Available http://www.hhs.gov/recovery/programs/cer/draftdefinition.html. Accessed 29 May 2009.

Comparative Effectiveness Research: Will it Bend the Health Care Cost Curve and Improve Quality?

Mark McClellan, MD, PhD and Joshua Benner, PharmD, ScD

The Need for Health Care Reform that Increases the Value of Health Care

In the debate about health care reform, one area of agreement is the need to address the gaps in quality and efficiency in health care in the United States. While America leads the world in many measures of health care innovation, it lags behind many developed nations in important health outcomes like mortality rates for conditions amenable to medical care and has much higher health care costs. Spending on health care will consume approximately 18 percent of GDP in 2009, or $2.5 trillion—and at current rates of growth, health care will exceed one-fourth of GDP by 2025.[2,3] Federal spending accounts for about one-third of those totals, and federal outlays for Medicare and Medicaid alone are projected to nearly double from $720 billion in 2009 to $1.4 trillion in 2019. Over the longer term, the Congressional Budget Office has determined that health care costs represent the single greatest challenge to balancing the federal budget.[4]

Policymakers are hopeful that health care spending growth can be reduced to a more sustainable level without reducing access to care that improves health. Substantial evidence on the variations in medical care from area to area around the country suggests that as much as 30 percent of spending reflects medical care of uncertain or questionable value. Investigators at Dartmouth have documented significant geographic variations in the intensity of services for colorectal cancer, hip fracture, acute myocardial infarction, and end-of-life care.[5] Intensity of discretionary services such as lumbar surgery, hysterectomy, and bypass surgery can vary by as much as a factor of 20, depending simply on where one lives. For example, in Idaho Falls, Idaho, 4.6 lumbar fusions were reported per 1,000 Medicare enrollees annually compared to 0.2 in Bangor, Maine, with no difference in the outcomes.[6]

Many of these medical treatments in common use, as well as many emerging therapies, are not backed by strong empirical evidence. Overall, the Institute of Medicine has estimated that less than 50 percent of treatments delivered today are supported by evidence.[7] A recent review of practice guidelines developed by the American College of Cardiology and the American Heart Association found that relatively few recommendations were based on high-quality evidence.[8] A similar study revealed that most

1. Schoen C, et al. U.S. Health System Performance: A National Scorecard. *Health Affairs*, Web Exclusive (September 20, 2006):w457–w475.
2. Sisko A, et al. Health Spending Projections Through 2018: Recession Effects Add Uncertainty to the Outlook. *Health Affairs*, Web Exclusive (February 24, 2009), pp. w346–w357.
3. Congressional Budget Office. January, 2008. Technological change and the growth of health care spending. Available at: http://www.cbo.gov/ftpdocs/89xx/doc8947/01-31-TechHealth.pdf.
4. Congressional Budget Office, The Long-Term Outlook for Health Care Spending (November 2007). Available at http://www.cbo.gov/ftpdocs/87xx/doc8758/11-13-LT-Health.pdf.
5. Fisher, ES and Wennberg, JE, *Health care quality, geographic variations, and the challenge of supply-sensitive care*. Perspectives in Biology and Medicine, 2003. 46(1): p. 69-79.
6. Sung, NS, et al., Central challenges facing the national clinical research enterprise. JAMA, 2003. 289(10): p. 1278-87.
7. Institute of Medicine. 2007. Learning What Works Best: The Nations Need for Evidence on Comparative Effectiveness in Health Care . http://www.iom.edu/ebm-effectiveness.
8. Tricoci P, Allen JM, Kramer JM, Califf RM, and Smith SC. 2009. Scientific Evidence Underlying the ACC/AHA Clinical Practice Guidelines. *Journal of the American Medical Association* 301(8):831-841.

guidelines for treating lung cancer were not based on adequate evidence.[9] A major reason for the gap is limited investment in comparative effectiveness research: of the nation's more than $2 trillion annual health expenditure, currently less than 0.1 percent is invested in assessing the comparative effectiveness of available interventions.[10,11]

All of this suggests that developing better evidence on which treatments work best is essential for effective health care reform.

The absence of timely and relevant evidence appears headed toward becoming an even larger problem as the amount of information potentially relevant to decisions for particular patients explodes. Indeed, the 21st century is promised to be the era of personalized medicine.[7] New technologies based on advances in genomics, proteomics, and molecular biology hold the potential to prevent, diagnose and treat diseases on a more targeted basis. But actual use of highly targeted treatments based on information related to individualized aspects of diseases has only occurred on a limited basis. A critical reason is that demonstrating the effectiveness of these personalized diagnostics and therapies—showing that they reliably work significantly better in particular types of patients—has proven very difficult to do on a targeted basis. If it has been difficult to identify which major treatments work best for broader populations, the challenges for identifying the most effective use of complex, customized treatments for small subgroups of patients seems far more difficult using existing methods. Thus, the lack of evidence to support the comparative advantages of more personalized care is an increasingly prominent part of the "evidence gap."

All of this suggests that developing better evidence on which treatments work best is essential for effective health care reform. With this goal in mind, the American Recovery and Reinvestment Act of 2009 (ARRA) provided a "down payment" of $1.1 billion to the Agency for Healthcare Research and Quality (AHRQ), the National Institutes of Health (NIH), and the Secretary of the Department of Health and Human Services (HHS) to support the development and dissemination of evidence on comparative effectiveness research (CER). The law also provided for the formation of a Federal Coordinating Council on Comparative Effectiveness Research to help direct and coordinate the use of the new funding. The Institute of Medicine was commissioned to develop an initial list of "national priorities" for CER which will be completed by June 30, 2009. While the high level of policy interest and new funding mean that much more CER activity is coming, how to implement CER so that it has the most beneficial impact on these challenges is not clear.

High Hopes—but Differing Expectations—for CER

Supporters of expanded CER believe that compelling evidence will lead patients and their doctors to make better choices among various medical treatments, leading to better outcomes and in cases where an equally or more effective treatment may cost less lower health care spending. CER could help usher in the era of personalized medicine – an era that has been slow in coming despite enormous progress in basic sciences – by identifying the most effective strategy for individuals, through conducting studies

9. Harpole LH, Kelley MJ, Schreiber G, Toloza EM, Kolimaga J, and McCrory DC. 2003. Assessment of the scope and quality of clinical practice guidelines in lung cancer. Chest 123(1 SUPPL.):7S-20S.
10. Moses, H, 3rd, et al., Financial anatomy of biomedical research. JAMA, 2005. 294(11): p. 1333-42.
11. AcademyHealth, Placement, Coordination, and Funding of Health Services Research within the Federal Government, in AcademyHealth Report. September 2005.

of the different treatments in "subgroups" within broader populations. The Director of the Office and Management and Budget, Peter Orszag, has noted that better evidence is a key step in reducing cost growth and achieving a more sustainable health care system.[12]

On the other hand, some critics argue that CER results are likely to be misused, and the evidence may be outdated by the time it is available. The technical capabilities of an intervention, particularly a device, may evolve as providers become more experienced with it; drugs may be dosed differently or used in different combinations than in CER studies. Finally, if the results of a CER study of alternative treatments are strictly applied to a broad population—for example, through a decision not to cover a treatment based on the CER results—then outcomes may worsen for particular patients who, for various reasons such as comorbidities, race and ethnicity, genetics, or preferences, may have responded better. There are also concerns about comparative *cost-effectiveness* research. Some critics suggest that CER will not only identify treatments that are more clinically effective, but would also identify treatments that have benefits but are very costly (e.g., an expensive biologic drug that increases life expectancy by a month but costs $100,000), and coverage of these treatments would be precluded based on the research, as is the case in the United Kingdom and other countries. To address these concerns, ARRA clarified that none of the reports from the entities involved in overseeing CER should be construed as recommendations on the translation of CER findings into practice guidelines or coverage policies. However, that clarification does not answer the question of how CER evidence will be used by those who make decisions about the coverage of and reimbursement for medical services and products.

Finally, many have expectations for CER that are much more modest. Indeed, some of the same reasons cited by critics may actually be reasons why patients, their providers, and policymakers will be reluctant and/or slow to apply the results of CER. Moreover, many argue that CER studies may well be much more expensive than traditional clinical trials, because sufficient samples from specific groups of patients who may plausibly respond differently need to be included, because some important long-term outcomes take a long time to observe, and because studies must be "powered" to detect differences in outcomes that are important but not large—for example, even a 10 percent difference in outcomes between alternative treatments may be clinically important. All of these factors were presumably considered by the Congressional Budget Office (CBO) when it recently evaluated the impact of a national investment in CER on health care costs over the 2010-2019 period. The program was estimated to generate modest savings in government health care costs, but these were not enough to offset the cost of the research; the net effect was well short of the break-even point.[13] CBO did note that effects on quality of care may be larger, since the new evidence may result in net increases as well as net decreases in the use of health care services, both of which could improve outcomes. But the impression is one of a modest impact as long as incentives within the health care system remain unchanged. CBO speculated that the impact of CER might be larger if there were incentives and processes for more rapid and effective transfer of CER findings into the practice of medicine.

These diverse views suggest that the ultimate impact of CER for better or worse is very uncertain. The key unresolved questions deal with whether CER *as it will actually be implemented*—and it will—can reduce costs and improve outcomes.

12. Peter R. Orszag, "Health Costs Are the Real Deficit Threat," *Wall Street Journal*, 15 May 2009. http://online.wsj.com/article/SB124234365947221489.html. Accessed 29 May 2009.
13. Congressional *Budget Office*, *Budget Options*, *Volume 1: Health Care*, "Option 45: Fund Research Comparing the Effectiveness of Treatment Options," December 2008. Available: http://www.cbo.gov/ftpdocs/99xx/doc9925/12-18-HealthOptions.pdf.

Will CER have a substantial impact on medical practices and health care costs?

Can CER lead to lower-cost care and thus "bend the curve" of increasing health care expenditures?

Will CER improve health outcomes?

Even if savings are achieved, do these savings come at the expense of better health, or do they promote it? In particular, will CER reduce the use of treatments that are beneficial to some patients, as well as those that are not effective, and will it reduce the returns and thus the incentives for biomedical innovation?

The answer to both questions is that it depends, and the key questions—even the key issues—involved in the implementation of a large-scale CER strategy for the United States remain unresolved.

CER Done Right

The positive expectations of CER may be achievable if the research provides timely, relevant evidence for individual patient care decisions, and for policy decisions that affect how CER is applied to populations of patients. The application of CER findings must be reinforced by health care system incentives that drive the behavior of stakeholders including consumers and patients, providers, payers, and the medical products industry.

What evidence will be most relevant to treatment decisions? As the evidence related to potential overuse of medical care implies, CER related to costly treatments that are used at very different rates from area to area may provide information that can lead to reductions in the variations in the use of such procedures. Such geographic variation can be partially attributed to varying rates of illness, differences in the prices that Medicare pays for the same service (which are adjusted on the basis of local costs for labor and equipment in the health sector), differences in income or the stated preferences of individuals for specific types of care, and unmeasured differences in the demand for care.[14] However, a growing body of evidence suggests there is greater geographic variation in treatment patterns when there is less consensus within the medical community about the best treatment to use, suggesting that an absence of evidence is partly to blame. For example, patients who have fractured their hip clearly need to be hospitalized, and there is relatively little variation in admission rates for that diagnosis. Hip and knee replacements, however, are discretionary procedures, and the surgery rates vary more widely. Another example is back surgery—a treatment whose benefits have been the subject of substantial questions. Because many of these technologies may be beneficial in some patients, such studies might try to focus on the particular types of patients where benefits are most questionable.

However, the evidence on overuse shows that most of the variation in health care costs is related not to variations in the use of major, costly treatment alternatives, but to more pervasive differences in the use of a broad range of medical resources. Rather, it is attributable to differences in the treatment of common conditions like diabetes and heart disease in terms of frequency of seeing a physician, likelihood of referral to a specialist, use of laboratory tests and imaging procedures, and frequency of admission and intensity of use of hospital care. CER must find ways to address these more subtle but widespread variations in medical practice, which seem to be difficult to evaluate in clinical trials.

14. John E. Wennberg, Elliott S. Fisher, and Jonathan S. Skinner, "Geography and the Debate Over Medicare Reform," *Health Affairs*, Web Exclusive (February 13, 2002), pp. w96–w97.

Further, the problems in quality of care do not simply involve overuse of treatments; in many cases where strong evidence on effectiveness of particular treatments does exist, it is not applied. For example, one recent study found that Medicare enrollees frequently did not receive care that was recommended or deemed appropriate.[15] Even preventive services proven effective at disease prevention—such as influenza vaccines and mammograms—are underused by one in four beneficiaries.[16] Another study that examined adherence to 439 indicators of health care quality for 30 acute and chronic conditions as well as preventive care, found that Americans received only 55 percent of recommended care.[17] Furthermore, there is little correlation between the amount of care delivered and care quality. High utilization of health care services, resulting in higher overall health care costs, may even be detrimental to quality.[18]

Big gaps in the quality of care, and substantial avoidable health care costs, are also related to the misuse of medical technologies. Misuse is a different type of failure to apply existing evidence—incorrect diagnoses and thus treatments, as well as medical errors and other sources of preventable complications (such as infections that patients acquire during a hospital stay). The Institute of Medicine has issued several reports documenting the extent of medical errors and their consequences.[19] Recently, Medicare has stopped paying for what are termed "never events"—mistakes such as operating on the wrong body part, but system-wide estimates of the extent and costs of medical errors do not suggest a decline over the past decade.

The problems of failing to apply evidence, or more generally of not achieving the best outcomes for patients without unnecessary costs, suggests another way to increase the impact of CER: research should focus on comparisons not just of alternative treatments in particular kinds of patients, but on alternative *policies* that clearly influence how those treatments are used. Indeed, compared to a randomized trial of alternative treatments for prostate cancer, it may be relatively easy to conduct studies of decision support tools to present the available evidence to patients and providers, or different coverage policies (many health plans now have authorization rules or copayment requirements), or different provider payment policies (e.g., payment based on quality of care or results, rather than fee-for-service). Such evidence would clearly be more directly relevant to evaluating the impact of coverage decisions and other payment reforms, and would address the concerns raised by CBO and many other experts that better incentives and support are needed to encourage the effective use of CER. It may also reduce the need for some CER studies: instead of a costly head-to-head trial of alternative drugs for controlling blood pressure, for example, a comparison of different formulary strategies could identify an optimal design that improves the use of blood pressure medicines, rates of blood pressure control, and lowers costs for a population of patients. Such evidence on how policies affect care within a population of patients may also provide some insights relevant to the care of particular types of patients who are treated differently as a result.

15. Stephen F. Jencks, Edwin D. Huff, and Timothy Cuerdon, "Change in the Quality of Care Delivered to Medicare Beneficiaries, 1998–1999 to 2000–2001," *Journal of the American Medical Association*, vol. 289, no. 3 (January 15, 2003), pp. 305–312
16. Medicare Beneficiary Use of Clinical Preventive Services. U.S. General Accounting Office. 2002
17. McGlynn EA, Asch SM, Adams J, et al. Quality of health care delivered to adults in the U.S. *NEJM* 2003;348:2635-45.
18. "Figure 1.The Relationship Between Medicare Spending and Quality of Care, by State, 2004" in Douglas W. Elmendorf, "Options for Controlling the Cost and Increasing the Efficiency of Health Care," testimony before the Subcommittee on Health Committee on Energy and Commerce U.S. House of Representatives, Washington, D.C., 10 March 2009. Available: http://www.cbo.gov/ftpdocs/100xx/doc10016/03-10-Health_Care.pdf. Accessed 28 May 2009.
19. See, for example, Committee on Identifying and Preventing Medication Errors, *Preventing Medication Errors: Quality Chasm Series*, ed. Philip Aspden, Julie Wolcott, J. Lyle Bootman, Linda R. Cronenwett, (Washington, D.C.: Institute of Medicine, 2006), or Committee on Quality of Health Care in America, *To Err Is Human: Building a Safer Health System*, ed. Linda T. Kohn, Janet M. Corrigan, and Molla S. Donaldson , (Washington, D.C.: Institute of Medicine, 2000.)

Three Guideposts for Delivering on the Promise of CER

The Engelberg Center for Health Care Reform at Brookings and The Hamilton Project have asked experts in CER and health economics to identify the key guideposts for achieving the maximum beneficial impact of CER on health while avoiding unnecessary health care costs, and to propose specific steps for achieving these guideposts as CER is implemented. Each paper is intended to both raise a critical issue in implementing CER effectively, and to provide specific recommendations for a path to CER "done right."

Guidepost 1: CER funds must be prioritized carefully to assure that studies address the most important clinical and health policy decisions.

A key step in maximizing the impact of CER is to develop a process for identifying priority research topics. The predominant process of setting research priorities for federal funding, which involves awarding budgets to granting organizations who in turn give research funding to the most meritorious proposals from investigators, helps assure that research meets technical and relevance standards, but probably does not maximize the value of CER. Moreover, this process often does not acknowledge or coordinate with ongoing research in the private sector. The resulting body of evidence contains both gaps and overlaps. National priorities for CER could be more efficiently set and achieved if a more systematic process were developed with criteria that directly reflect the needs of decision makers in both patient care and policies that affect patient care.

In their paper, "Setting Priorities for Comparative Effectiveness Research," Alan Garber and David Meltzer propose the principles and process for establishing national CER priorities. They suggest that CER studies should maximize the "value of information," a quantitative term that is higher for studies of interventions where the best choice is very uncertain and the consequences of making the wrong choice are very large, both in clinical and economic terms. Another consideration when setting priorities for federally funded research is the natural incentives that may exist for studies to be conducted in the private sector. Because it can be difficult to calculate the value of research systematically over a range of potential topics on a variety of clinical conditions and interventions, Garber and Meltzer propose a more practical approach that employs the same principles, developed by a working group of industry and academic leaders, facilitated by McKinsey & Company. This approach requires decision makers to identify the interventions that are going to be compared, create outcome measures that can determine the impacts of the selected interventions, characterize how study findings might change what is known about the interventions, and account for how these findings might alter treatment decisions. The authors go on to suggest how a priority-setting body should be constructed to minimize bias and what voting procedures it should follow to generate a list of priorities. Finally, they recommend that national priorities be re-evaluated periodically by repeating the process in light of new information.

Guidepost 2: CER requires investment in methodological guidance and a more efficient infrastructure for applied research on both alternative treatments and alternative policies for promoting the use of the best treatments without unnecessary costs.

Investigators, study sponsors, and patient advocates alike agree that the current process of conducting clinical studies fails to meet the modern need for rapid, relevant information to support clinical decisions. Prospective trials can take too long, cost too much, and the pace of technological change often makes results obsolete before they are even known. On the other hand, analysis of administrative and clinical databases is rapid and inexpensive, yet fraught with the potential for biased findings due to limitations in the data and methods. Moreover, the absence of a research infrastructure to support

population-level studies of systems to support complex decisions, of alternative financing mechanisms, and of other policy choices means that most policies are implemented today without clear evidence of their effectiveness. Consequently, the essential foundation for CER is a 21st century research infrastructure that provides for more efficient experimental trials and observational studies – and guidance on how and when each should be used. The new research infrastructure must identify and promote best practices for the conduct and reporting of CER studies. New kinds of training and expertise are required to support this capacity.

In our second discussion paper, Sean Tunis outlines a series of recommendations for "Improving Methods and Infrastructure for Comparative Effectiveness Research." He points out the importance of involving those with a stake in the decision under study—especially consumers and patients—in the research and recommends that federal grants for CER studies require investigators to develop a stakeholder advisory committee. Next, he revisits the traditional "hierarchy of evidence" and recommends that it be replaced by a new, decision-focused evidentiary framework be developed to guide the selection of CER methods. With such a framework in place, best practices should be identified for defined categories of CER studies. Tunis's Center for Medical Technology Policy has developed the concept of an Effective Guidance Document that could be developed by relevant stakeholders to guide research sponsors and investigators in the development of CER evidence for specific health care interventions. For CER trials, he emphasizes the need to grow and link practice-based research networks, to create capacity for interventions to be evaluated in the real world. There is also a need to address sources of inefficiency in protocol approvals by ethics committees and contracting with research centers. To enable a true "learning health care system," Tunis calls for the development of a distributed data network to securely link public and private-sector administrative and clinical databases – including Medicare and Medicaid. In sum, Tunis proposes that at least 30 percent of CER funding over the next decade be allocated to developing the needed methods and infrastructure.

Guidepost 3: CER must be implemented in conjunction with other reforms that provide stronger incentives and support to use evidence to improve quality and lower costs.

Clearly, even when relevant evidence from CER has been developed, it may not have a significant or timely impact on medical practice. What steps could change that? Considerable evidence indicates that a wide range of interventions may affect the awareness and use of evidence in decisions by patients and their health care providers. This includes steps to assist patients and providers in identifying and interpreting the available evidence, such as relevant and up-to-date reviews and electronic decision support systems. Even more broadly, it includes a wide range of financial and other incentives to use evidence to improve outcomes and avoid unnecessary costs. For example, the current fee-for-service reimbursement environment provides limited incentives at best to use effective treatments that cost less: virtually all of the treatments, big and small, that vary substantially from area to area receive higher payments when they are used more often, not necessarily when they contribute to better outcomes. Payment reforms such as paying for better outcomes and lower overall costs may change these incentives.[20,21] Benefit reforms can also have an impact, such as lower copays for treatments that are cost-effective (but continuing to provide coverage with higher copays for other treatments that may be preferable in certain patients) or copays that are

20. Elliott S. Fisher, Mark B. McClellan, John Bertko, Steven M. Lieberman, Julie J. Lee, Julie L. Lewis, and Jonathan S. Skinner, "Fostering Accountable Health Care: Moving Forward In Medicare," *Health Affairs*, March/April 2009; 28(2): w219-w231.
21. Alan M. Garber and Mark B. McClellan, "Satisfaction Guaranteed — 'Payment by Results' for Biologic Agents," *New England Journal of Medicine*, 18 October 2007; 357 (16): 1575-1577.

targeted to the characteristics of individual patients. Consequently, CER is likely to have larger effects on costs and health when implemented in a coordination with a broader range of health care reforms that promote the use of evidence to achieve higher-value care.[22]

While these topics collectively are much broader than the scope of this volume, one potentially important piece of applying CER results is reviewing, summarizing, and helping various audiences interpret the array of CER-related evidence. In his paper, "From Better Evidence to Better Care: Using Comparative Effectiveness Research to Guide Practice and Policy," Steve Pearson explains in more detail why research often fails to have an impact on practice, and derives from these barriers a potential strategy for meeting the needs of these diverse audiences. Among his recommendations are development of a rating system to judge the strength of CER evidence, potentially based on rating systems developed by himself and others in organizations such as the U.S. Preventive Services Task Force, Blue Cross Blue Shield Association Technology Evaluation Center, the Institute for Clinical and Economic Review. Pearson calls for the use of cost-effectiveness analysis to measure an intervention's overall value, and for these findings to be translated into actionable recommendations for payers, providers, and patients alike. To ensure that CER evidence has appropriate influence on health care decisions, Pearson recommends that clinician societies be closely involved in both the development and translation of CER evidence, and he supports the use of CER by payers in the forms of evidence-based benefit designs, patient incentives, and physician reimbursement. Finally, he proposes models for payers to structure reimbursement of new technologies contingent on the development of new CER evidence, as a method of rewarding privately sponsored CER and stimulating innovation.

Moving Forward with CER to Better Care and Slower Spending Growth

Even if arguments remain about these specific recommendations, the proposals have a clear implication: for CER to make a substantial, positive contribution to reforming health care, the critical implementation issues of prioritizing CER spending to "high-value" studies, creating an efficient research infrastructure, and creating an environment that promotes the effective use of evidence from CER must be addressed. CER is likely to have a significant impact on health and unnecessary health care spending if:

1. It is conducted on questions where the "value of research" is high;

2. A robust research infrastructure is developed, including an evidentiary framework, methodological best practices, data infrastructure, and a well-trained workforce of investigators; and

3. Mechanisms are developed to promote the appropriate use of new evidence in clinical practice and health policy in a timely way.

Achieving these objectives will be challenging, but doing so could enable CER to play an essential role in achieving the goal of bending the cost curve while improving health in America.

22. Engelberg Center for Health Care Reform, "Real Health Care Reform in 2009: Getting to Better Quality, Higher Value, and Sustainable Coverage," November 2008. Available at: http://www.brookings.edu/events/2008/~/media/Files/events/2008/1117_realhealthcare/1117_strategicreview.pdf. Accessed 28 May 2009.

Setting Priorities for Comparative Effectiveness Research

Alan M. Garber, MD, PhD and David O. Meltzer, MD, PhD

Introduction

The $1.1 billion allocated to comparative effectiveness research (CER) in the American Recovery and Reinvestment Act (ARRA) is a significant first installment in what many believe will become a large and sustained commitment to learn more about what forms of health care work best. The hope is that we can abandon those forms of care that are ineffective, and adopt those diagnostic tests, treatments, and approaches to prevention that do the most to improve health. But even this substantial increase in resources devoted to CER can seem small in relation to the goals of the effort. A randomized trial that compares effective (compared to placebo) interventions can cost many tens of millions of dollars, and investments in the information technology that can serve as infrastructure for such studies can cost billions of dollars. Resources for CER will be limited relative to the needs, even with a much larger investment in the activity.

Prioritization is the selection of the specific projects that comprise the comparative effectiveness research portfolio. Its goal is to ensure that the research investment in CER will have the greatest long-term payoff to the health of the public. Because it can take so long for the research to yield results — randomized clinical trials, for example, frequently last for seven or more years — a successful prioritization strategy must not only produce results that can be generated and used quickly, but also be guided by long-term considerations. In this chapter, we discuss the principles that should guide prioritization, and strategies for implementing them.

In our view, the group of persons making decisions about prioritization should be broadly representative of the people who will be applying the results of CER, and should also have relevant methodological and clinical expertise. But selecting the right group will not, by itself, guarantee that the prioritization effort will be successful. Any such group will need appropriate information with which to make decisions, and will need to follow procedures that enable them to select a portfolio with the greatest potential for impact. They will need to follow procedures, in other words, that build upon a set of principles about how to maximize the value of the comparative effectiveness research investment.

Other features of the processes are also important. For example, when a public or public-private agency that conducts or sponsors comparative effectiveness research makes decisions, prioritization will need to be a transparent process, and should be subject to ongoing public scrutiny.

In the next section, we describe "value of information" principles that can inform the development of procedures that can guide the selection of a research portfolio with the greatest potential to produce value. We then describe the composition of the group or groups that make decisions about prioritization, and offer general comments about the procedures that they might follow. We close with a set of specific recommendations about how prioritization should be conducted in order to maximize the overall health impact of CER, while addressing concerns that might not be adequately addressed by a crude application of a value of information approach.

Our discussion focuses on research about specific clinical interventions. The scope of CER is much broader, and priority-setting is not limited to decisions about which drugs, surgical procedures, or diagnostic tests to study. Studies of the performance

of alternative health care delivery arrangements, of insurance benefit design, and of public health measures are considered by many to be within the scope of CER as well. The principles and approaches we discuss apply equally well to such "interventions." In addition, the CER effort will require investments in research infrastructure, such as the development and aggregation of data from electronic health records and methodological research to support CER. There will also be investments in communication of CER results to diverse audiences. Such investments will be crucial to the ultimate utility of CER, and are best viewed as ways to deliver better information to those who will use it. In each of these areas, we believe that resources should be directed toward uses that will maximize the overall value of the CER portfolio, and the procedures we describe will be helpful here too.

Guiding Principle: The Value of Information

Prioritization can be viewed as an activity designed to maximize the value of a comparative effectiveness research effort. Prioritization is needed because resources are limited and it is not possible to conduct every potentially valuable project. The key issue in understanding how to maximize the value of specific CER efforts is to identify the mechanism by which such research will produce value. There may be many reasons to conduct research – basic research, for example, can provide insights with consequences that are often difficult to predict, and can lead to tangible developments such as new therapeutic approaches many years later. Nevertheless, the chief purpose of CER is pragmatic: to produce information that changes clinical decisions for the better. This is a key element in the medical application of the principle of the "value of information" (VOI).[1]

The value of information (also known as the "value of research") is the difference between the value of the outcome given the decision one would make in the absence of additional information and the value of the outcome of the decision that would be made as additional information became available as a result of research. This may appear to be puzzling since the main reason we want to conduct the research is that we don't know what the outcomes will be with the intervention and with the alternatives to which it is compared. But to calculate the expected value of the information it is necessary to start with a best guess about the possible effects of the treatment and the alternatives. Sometimes the guess will be based on very good information, such as preliminary studies in human beings that suggest a strong treatment effect but are too small to support a definitive conclusion. Sometimes the information will be less direct, such as knowledge of the molecular mechanisms of action, which can help predict side-effects or efficacy, or there might have been evidence of effectiveness in animal studies. Deciding on some sort of assessment of expected effectiveness based on this information is needed as a first step in calculating the value of more information. Even a decision maker with strong beliefs about the effectiveness of the intervention and the alternatives will recognize that more information could cause those beliefs to change greatly.

The value of research is estimated by building upon these beliefs about the outcomes from an intervention and outcomes when alternative interventions are used. One can calculate the average (expected) health outcomes that will occur under alternative results of the research — compared to the average (expected) health outcomes if the research were not done.[2] This is the essential idea behind the "value of research" approach. Expected health outcomes change only if the research stimulates changes in clinical decisions.

1. Raiffa, H., Schlaifer, R., 1961. Applied Statistical Decision Theory. Harvard Business School, Colonial Press.
2. This approach to formulating the value of information assumes that decision-makers prefer decisions that produce the best decision on average, also known as the expected value. This is calculated by multiplying the differences in outcomes times the probability of those outcomes. Because decision-makers may understandably be averse to risks, especially large ones, this may be an unacceptable assumption in some settings. Even when such concerns are present, expected outcomes may provide a useful benchmark.

A simple example demonstrates how the value of research can be calculated. Suppose that the relevant clinical choice is between treatment A (TxA) and treatment B (TxB), and that current evidence suggests that TxA produces better outcomes for patients. While TxA appears to the better option now, there is a probability p that further research would show that TxB was better than TxA. The expected value of conducting further research on these treatments would be the probability p that TxB was better than TxA multiplied by the difference in true benefits between treatments A and B. This can expressed mathematically as:

> Expected value of research = p * (Value of Outcome (Tx B) – Value of Outcome (Tx A)).

To calculate the value of information requires developing measures of both the probability that the best decision would change and the difference in benefit that would arise from the change in the decision that will occur in response to the study findings.

To determine whether one intervention is more beneficial than another requires a comprehensive measure of health outcomes, such as quality-adjusted life years (QALYs).[3] These measures of benefit should be as complete and relevant to the individual patient as possible; indeed, one of the goals of comparative effectiveness research is not only to determine whether an intervention produces a benefit on average or in a narrowly defined population, but also in different groups of patients, distinguished by such characteristics as age, gender, race, severity of illness, the presence of co-morbidities, and the presence of specific genetic markers. Comparative effectiveness research can also determine when a less costly therapy offers equivalent or better health outcomes than an alternative or to support economic evaluations that seek to determine whether a more effective and more expensive alternative is worth the extra cost.[4] In cases where an intervention is beneficial but costly and a decision maker wishes to consider costs, it is possible to calculate the net benefits of an intervention using either cost-benefit analysis, which values health in dollar terms and subtracts off costs, or a newer framework called net health benefits, which value health benefits in QALYs and subtracts off QALYs that could have been produced at the same cost as the intervention if they were used for some other, cost-effective, use. These aggregate measures of net benefits are useful in thinking about value of research calculations because they can be used to calculate the net benefit that comes from research on the comparative effectiveness of interventions once the costs of the interventions are accounted for.

A second aspect of the value-of-research calculation is to determine what (clinical) decisions would be made absent the research, and how often and in what ways the decisions would change in response to the research findings. In most studies of the value of research, the "pre-research" deci-

3. Quality-adjusted life years (QALYs) are measures of life expectancy in which years of life are weight by quality of life (QOL) weights between 0 and 1 in which 0 is equal to death and 1 is equal to perfect health. If QOL life and QALYs are measured correctly, people would be expected to think that an intervention that increased life expectancy by two years at QOL of 0.5 would be equally attractive as an intervention that increased life expectancy by one year at a QOL of 1.
4. The most frequently used approach to economic evaluation is cost-effectiveness analysis, which calculates the ratio of added costs to added effectiveness. Because cost-effectiveness is reported as a ratio of costs to effectiveness, a limitation is that those ratios do not provide data on the magnitude of the total size of the benefits and costs of the intervention so a highly cost-effective intervention that would only provide small health benefits for a few people does not look different than a highly cost-effective intervention that would provide larger benefits for many people. To address this concern, decision makers interested in the magnitude of the total net benefits, can construct a comprehensive measure of the net benefits for the intervention using either cost-benefit analysis (CBA), which puts a dollar value on health benefits and subtracts off any costs, or a similar measure called net health benefits (NHB), which calculates health benefits from the intervention in some comprehensive measure such as QALYs and then subtracts off the number QALYs that could have been obtained if the financial resources required for the intervention had been directed towards another (cost-effective) intervention that was not being done. Just as when, in the cost-benefit framework, a cost-effective intervention produces positive net benefits and an intervention that is not cost-effective produces negative net benefits, in the net-health benefits framework, a cost-effective intervention produces positive net health benefits and an intervention that is not cost-effective produces negative net health benefits. For a more general discussion of these and other methods in cost-effectiveness analysis, see *Methods for the Economic Evaluation of Health Care Programmes*, Third Edition by Drummond, Sculpher, Torrance, O'Brien, and Stoddart, Oxford University Press, 2005.

sion is assumed to be the best decision that could be made with the information available before the research is conducted. An alternative and arguably more meaningful baseline would be the *actual* decisions that are made before the research. Then the shift in decisions is based on a comparison to real rather than ideal practice. For example, if (non-ideal) treatment A was the baseline approach and research revealed that treatment B provided greater benefit, the assumption would be that the choice would change from A to B.

To estimate the likelihood that treatment decisions will change in this way, it is necessary either to make assumptions about effectiveness or to use existing data on effectiveness. For example, if prior studies have suggested that there is a 70 percent chance that a treatment is beneficial, this information would be used for the calculation. This is, of course, a simplification; treatment B typically has an entire probability distribution of benefits compared to treatment A. Comparative effectiveness research narrows and potentially shifts the distribution of relative benefits, generally offering greater assurance but not absolute certainty that treatment that appears more effective will lead to better health outcomes. By this method, one can build up an estimate of the expected value of research. This is an estimate of the average benefit from improved clinical decisions, calculated based on the likelihood of each potential study outcome. And although this language suggests that the focus will be on an "average" patient, similar calculations can be performed for closely defined groups, such as female heart attack survivors aged 70-75 who have type II diabetes mellitus.

A Stylized Example of Value of Information (VOI)

A simple example in the context of comparative effectiveness research might be studies of a new chemotherapy regimen that costs $50,000 and is thought to have a 25 percent chance of increasing life expectancy by 1 year and a 75 percent chance of offering no increase in life expectancy. Imagine also that this is a treatment that is not fully covered by insurance, so that patients need to consider the costs to them of the treatment. Without research, decision-making for any patient must weigh the $50,000 cost versus the 25 percent chance of a one year increase in life expectancy. Imagine also that people have the $50,000 available to them to get access to this treatment if they feel it is worth the cost to them and that they all value a year of life at $100,000 per life year, so that if the treatment were found to be effective (producing an extra year of life expectancy), its cost of $50,000 would make it cost-effective for these persons to choose the treatment.

Assume that in the absence of information, half the people choose the treatment and half do not, perhaps simply because they do not know whether it offers benefits that are worth the costs to them and must guess. Now imagine that the planned research is done and provides a definite answer to the question of the effectiveness of the treatment. Based on our assumptions above, there is a 25 percent chance the study would find the treatment is effective. In that case, the half of people who would have chosen it anyway would continue to choose it. The people who would not have chosen it would now change their decision so that they select the therapy and gain 1 year of life expectancy (worth $100,000) while paying $50,000 for the treatment, for a net benefit of $50,000 in dollar terms. Thus, with 25 percent likelihood, there would be a benefit worth $50,000 for half of the population, or an average per capita value of $25,000 across the whole population in terms of potential health benefits. The above assumes the treatment turns out to be effective (i.e. to increase life expectancy). By our assumptions above, there is, however, a 75 percent chance that the treatment turns out not to be effective. In this case, results of the research would not change the decision for the half of people who would not have chosen the treatment, but would cause the half who would have chosen the treatment to now reject it, saving the $50,000 they would have paid for the treatment. Thus there is a 75 percent chance that 50 percent of the population would save $50,000, for an average per-capita benefit across the whole population of $37,500 in terms of cost savings. Combining these two halves of the population who would and would not have chosen the treatment in the absence further research would yield a total per capita benefit of $25,000+$37,500 = $62,500 per person.

The benefits of the research can also be aggregated up the population level. For example, in 1000 persons who would be eligible for the treatment each year, research that answered this question ten years earlier than it might be otherwise would produce a total value of $62,500 per person*1000 persons/year*10 years = $625 million minus the costs of doing the research. Viewed differently, one could say that the research was worth doing as long as it cost less than $625 million, though clearly any study that cost so much might affect the resources available for other studies unless more resources could be obtained for research and the value of other studies foregone would have to be considered.

In practice, application of these concepts can range from formal value of information calculations to much more informal, easily applied, approaches.

Implementation of VOI Principles

Although this basic approach can be illustrated with a stylized example such as that above, in practice, application of these concepts can range from formal value of information calculations to much more informal, easily applied, approaches. Table 1 lists the concepts, and a set of questions that decision makers might find helpful in applying the VOI framework. Although the table refers to treatments, these approaches can be applied to decisions regarding preventive care, diagnostic strategies, and other health interventions as well.

TABLE 1

Concepts, Measures, and Questions for Decision Makers for VOI Approaches to Prioritizing Research

VOI Concept	Measures	Questions for Decision Makers
Choice to be Made treatments	List of relevant alternatives	What treatment options are being studied? Are these the right treatment options to study?
Value of Outcomes	Distribution and expected value of comprehensive outcome measures (e.g. QALYs, costs, Net Benefits, or Net Health Benefits)	What health and/or cost outcomes might these treatments affect? Are measures of these outcomes available? If not, is there data on effects on the most important outcomes? How much uncertainty is there about these effects?
Potential Findings of Research	Potential findings of the research about the (expected) value of the treatment, and the likelihood of finding that information	What are the possible findings of the research and how likely are they? How much uncertainty about outcomes might there be after the research findings are known? Should other study designs be considered? Will this or another study be done anyway? Would other studies be likely to change the new information provided by this study?
Probability of Change in Choice	Probability that research findings will change treatment choice	Could the information generated by this research change treatment choices? How likely would it be for that information to change treatment choices?

Choice to be Made: Comparative effectiveness research must begin by defining the choice that is going to be studied by identifying the relevant set of interventions to be compared. Although the choice of alternatives will often be obvious, in many instances it will be difficult. For example, simply comparing treatments A and B might not be appropriate if there is a treatment C that might offer some advantages over either of these. Even if all the broad classes of options have been considered, there might be modest, but important, alterations in how treatments might be used that could change how one would think about the value of research. For example, the choice of a treatment could be linked with the generation of additional evidence about its effectiveness or its continuing refinement. These alternatives become especially salient in the context of recent initiatives such as Medicare's "coverage with evidence development" (CED) program. As a practical matter, only a subset of all possible choices can be considered in the formal analysis, so initial discussions often focus on the scope of alternatives to be considered in detail.

Value of Outcomes: Second, to apply value-of-information methods, one needs to be able to construct a single comprehensive outcome measure, such as QALYs (or net health benefits), that can capture the value of all health (and perhaps cost) effects of multiple interventions and therefore can be used to compare benefits across interventions. Indeed, because comparative effectiveness research is done specifically to address uncertainty about the outcomes of a medical intervention, comprehensive outcomes data are generally not available. Estimates of such quantities are often generated by developing decision models. Such models build upon a characterization of all the outcomes of a treatment, estimates of the probabilities of each outcome, and construction of a single measure of value that encompasses all of those outcomes.. This is often done through structured reviews of the literature, including techniques such as meta-analysis, which formally combines the results of multiple studies. Often the set of possible outcomes, the likelihood of those outcomes, or the value placed on those outcomes, is not precisely known. A large literature examines how to address these sources of uncertainty. For example, available information may be sufficient to bound an effect size. Sensitivity analysis can help identify when a decision is likely to change if more data about an uncertain outcome variable became available.

For some conditions, a single outcome measure might be enough. For example, in the treatment of a rapidly fatal disease, survival might be a sufficient outcome measure, especially if quality of life and cost are not major concerns. But often more than one outcome is important. There are several possible approaches for aggregating across those measures. Imagine, for instance, a treatment that benefited patients with Alzheimer's disease but increased caregiver burden; different people might weight the benefits and adverse effects differently. In those cases, the most valuable information to convey to decision makers might be a characterization of these outcomes along with an assessment of their uncertainty, and an estimate of how additional research would change these measures. Although this process would be inherently subjective, a major advantage of VOI is that it forces the careful examination of these factors even when a single comprehensive outcome measure cannot be developed. The same is probably true about concerns about the distributional effects of choices; there are several ways to integrate such concerns into a decision, and so approaches that describe distributional consequences without choosing among them may be most useful to decision makers. Often the sheer number of persons affected by a condition can be a very powerful determinant of the value of research, though surely one should not assume that this is always the case.

Potential Findings of Research: Third, one needs to be able to characterize the potential findings of research, and how they might affect what is known about the outcomes of the treatment. Data to characterize how additional research might change uncertainty about the outcomes of an intervention can often be drawn from observational studies or small

randomized studies. In addition, multiple study designs need to be considered. For example, even if a randomized trial has a sufficiently large expected benefit relative to the current state of knowledge to justify its cost, it might be a poor investment if a smaller, less costly randomized trial or observational study could provide similar information. Because research is generally part of an iterative process, the consequences of a given study for future research and decision-making may also be important to consider; for example, an observational study might be a logical precursor to a randomized trial.

When prior information on the value of information on outcomes is especially limited, it may be possible to bound estimates of the value of research by using burden-of-illness calculations to measure the potential benefits of successfully treating the condition.[5] While this approach can only provide an upper bound on the value of research, it can be informative when the upper bound suggests a very low value of research.

Value-of-information analysis can be very helpful as a tool for prioritization of research, and it can focus discussions by highlighting critical information and areas of uncertainty.

Probability of Changing Clinical Decisions: For the research to change the outcomes of care, it must change the choices that are made, so considering the probability that choices will change is essential in applying a VOI approach. In formal VOI analyses, it is generally assumed that when decision makers are presented with information, they choose the option that is superior given the information they have been presented with. Thus, if treatment A has higher survival (QALYs, net health benefits, etc,) than treatment B, then A would be chosen. The critical calculation for predicting the probability that the choice will change in response to research results is the probability that the measured outcome of the treatment not currently being chosen is better than the outcome of the treatment currently being chosen. Physicians, patients, and other decision makers take into account many other factors, such as convenience or the beliefs of friends, family members, and colleagues. By implicitly accounting for such information, sophisticated decision makers may be much better at predicting the value of research than formal VOI calculations. On the other hand, purely subjective assessments of the likelihood that a research study would change choices could also be subject to a variety of cognitive and other biases, so approaches that combine objective and subjective assessments may be most desirable, for example using a formal VOI analysis to inform a panel of experts who are ultimately charged with making a final decision.

For these and other reasons, formal VOI approaches will likely be most useful as an aid to decision makers, rather than as a replacement for the deliberations that they will undertake. Value-of-information analysis can be very helpful as a tool for prioritization of research, and it can focus discussions by highlighting critical information and areas of uncertainty. There is now a sizeable literature describing the use of value-of-information methods to prospectively identify the value of research,

5. Meltzer D. Addressing uncertainty in medical cost-effectiveness analysis: implications of expected utility maximization for methods to perform sensitivity analysis and the use of cost-effectiveness analysis to set priorities for medical research. Journal of Health Economics, 20(1):109-129, 2001.

and there is substantial experience applying these methods as part of a policy making process, most notably by the National Institute for Clinical Excellence (NICE) in the United Kingdom, which provides guidance to the National Health Service. The details of the NICE system are not the focus of the discussion here, nor do we envision a CER approach in the United States to be tightly bound to payers as it is in the U.K. But the NICE experience is instructive. Even when it performs or commissions a formal VOI analysis, the ultimate recommendations about a clinical decision are left in the hands of an expert panel. Thus the formal VOI analysis is an important input into the decision making process, not a replacement for committee deliberations. The next section of this paper focuses on who should make prioritization and processes that they might follow. The emphasis is more on structure than the content of decision making, but throughout, the concepts, measures and questions for decision makers that are listed in Table 1 can provide organizing principles, tools, and guides.

Experience of the National Institute of Health and Clinical Excellence (NICE) with Value-of-Research Calculations

The National Institute of Health and Clinical Excellence (NICE) has been a leader in the application of value-of-research calculations. One of the first examples of the use of these approaches to inform policy was a study, led by NICE-affiliated investigator Karl Claxton, that examined the use of acetylcholinesterase inhibitors in Alzheimer's Disease.

Although the use of acetylcholinesterase inhibitors is common in the treatment of Alzheimer's Disease, there has been a great deal of controversy about their effectiveness and, accordingly, their appropriate use and policies, including coverage policies, that could influence that use. In the United Kingdom, NICE, which is charged with providing guidance to the local health authorities that make most coverage decisions for the National Health Service, has assessed acetylcholinesterase inhibitors on several occasions. They performed very detailed formal value-of-research calculations to identify those aspects of uncertainty about the effects of acetylcholinesterase inhibitors that were most uncertain. Collaborating with Peter Neumann and other colleagues at Harvard to do a parallel study based on US data, they found tha information from further studies of the effectiveness of these medications would be valuable, worth more than $300 million in the US alone. Moreover, they identified that the single most important question about the effectiveness of these medications was the duration of benefit. Subsequently, Claxton and his colleagues examined the practicality of using value-of-research calculations as part of the policy process, examining its application to screening in age-related macular degeneration; alternative manual physiotherapy techniques in asthma and in chronic obstructive pulmonary disease; and use of alternative long-term, low-dose antibiotics in children with recurrent urinary tract infections.[1] They concluded that although data was insufficient in some cases to do an informative value of research calculation, the methods had potential to be applied in a timely and valuable fashion as part of the policy making process.

Who Will Make Prioritization Decisions?

The composition and authority of the group that makes decisions about the topics that will be studied can have as important an influence on the actual decisions as the principles that guide them. Determining who makes prioritization decisions is intimately related to the structure and governance of the entity that funds or oversees comparative effectiveness research. As the arrangement dictated by the ARRA shows, the entity responsible for conducting or sponsoring comparative effectiveness research need not have complete responsibility for prioritization decisions. The ARRA specified that a committee formed by the Institute of Medicine would compile a priority list of topics to be used to help guide the allocation of the $400 million of comparative effectiveness research funds that was designated for the use of the Secretary of Health and Human Services. Many other proposed frameworks for managing comparative effectiveness research assign complete responsibility for prioritization to the organization that will have responsibility for the conduct of comparative effectiveness research. Even if an external entity is advisory or makes prioritization decisions, federal agencies such as NIH that have responsibility for implementation of the priority list will need to transform such recommendations into specific actions. They will craft prioritized topics, into, for example, requests for applications. They will also need to ensure that the priority list is refined to reflect ongoing and planned research efforts, which may duplicate items on the priority list, and to reflect the costs of alternative topics and study designs. Their modifications may result in a list that differs considerably from the initial formulations of the topics. Thus the "who" of prioritization may consist of more than one group, with potentially overlapping responsibilities for determining which topics are selected for study.

Determining who makes prioritization decisions is intimately related to the structure and governance of the entity that funds or oversees comparative effectiveness research.

As is clear from the discussion of value-of-research principles, the selection of topics for study requires considerable knowledge about disease prevalence, disease consequences, the likely effects of all the interventions under study, and estimates of the likelihood of different outcomes of the research. It also requires assumptions about the changes in clinical management that would occur as a consequence of the research, and their effects on health outcomes. Prioritization requires, therefore, a blend of epidemiological, statistical, and clinical expertise.

In most respects, the expertise needed for prioritization is similar to the expertise needed to evaluate evidence of effectiveness in other contexts, such as the operations of the Medicare Evidence Development and Coverage Advisory Committee and similar bodies. In these settings, much of the relevant information is assembled and analyzed by outside experts and staff, then further scrutinized and discussed by a committee.

The characteristics of the committee matter greatly because implementation of a value-of-research approach cannot be simply formulaic. Many of the figures used in such an analysis are subjective estimates rather than readily measured values. Broad expertise is needed to ensure that all aspects of the problem are considered. This means the ability to determine whether the interventions are well-defined, whether they represent the full range of relevant alternatives, whether the patient populations are appropriately defined. Furthermore, the group must determine whether feasible study designs will

lead to answers to the questions that motivate the research.

The composition of the group making the prioritization decisions would therefore ideally be diverse enough to include people who, even if they lack expertise in specific areas, are capable of interpreting expert testimony and evidence presented to them. Furthermore, group members should be able to understand the needs of health care providers and especially patients and their families. Thus it is common to specify that committees with similar health care decision-making authority include clinicians from various specialties, methodological experts, consumers, and health plan and health product (drug and device) industry representatives. It is critical that government agencies responsible for producing, using, or disseminating the results of comparative effectiveness research also have input into prioritization.

With such broad representation, it is inevitable that conflicts of interest will arise. As is now widely recognized, financial interests are not the only sources of conflict that matter. An individual who has publicly stated a position on a topic under consideration may bring deep insight into the prioritization process, but by virtue of having reached a conclusion beforehand, can no longer be considered entirely impartial. It is not always obvious at what point a known point of view becomes an intellectual conflict of interest, nor does an individual's past statements always mean that he or she is unable to change views when new evidence is presented. Policies for addressing such conflicts are evolving, in part due to high-profile cases in which individuals were alleged to have failed to disclose, or to have disclosed incompletely, major financial interests that called into question the impartiality of their interpretation of research results. Many current policies toward conflict of interest in medical research share the following characteristics: 1) all financial and intellectual interests that might call into question the individual's ability to deal with the matters under study with impartiality must be fully disclosed before the members of the group are appointed; 2) the individual should be recused from any decision-making process that would directly affect his or her financial or intellectual interests. Thus, for example, a major shareholder in a medical device company should not have decision-making authority over any studies of that company's products or competitors; 3) general conflicts (e.g., a faculty member in a university that might conduct comparative effectiveness research or the director of a hospital whose revenues might be affected by decisions made on the basis of comparative effectiveness studies) do not necessarily disqualify the individual from participating in a prioritization process.

More complete and illuminating discussion is likely to occur when it includes individuals with competing conflicts of interest than when it consists solely of individuals free of financial or intellectual interests in the issues under discussion.

Resolving conflict of interest issues is a serious and important task, but the difficulties in doing so should not, in our opinion, lead to excessive restrictions on the composition of any entity that makes comparative effectiveness research prioritization decisions. More complete and illuminating discussion is likely to occur when it includes individuals with competing conflicts of interest than when it consists solely of individuals free of financial or intellectual interests in the issues under discussion. Complete disclosure of conflicts, however, is necessary to ensure that remarks can be interpreted in context and to lend greater credence to the deliberations.

Procedures for Setting Priorities

The group or groups setting priorities for comparative effectiveness research will be guided by general principles but should also follow specific procedures as they make their decisions. The procedures are necessary for the implementation of principles. Furthermore, procedures influence outcomes. Appropriate procedures can ensure transparency and fairness, and may enhance predictability. We have already addressed the importance of convening a group with a breadth of expertise and interests, and now discuss some of the chief features of the processes that will enable the group to carry out its tasks fairly, fully, and effectively.

Public Participation

Committee meetings should include ample opportunity for testimony from the public, including those with a direct interest in the interventions under consideration, providers who might administer the interventions, representatives of patients and consumers, and disciplinary and methodological experts. Effective incorporation of their input requires adequate notice of meetings and other opportunities to make either oral presentations or to submit written comments. In addition, public meetings can promote transparency by making available the discussion and collection of information used by the priority-setting body. Public participation can also help ensure the fairness of the proceedings, by offering opportunity for diverse, conflicting points of view to be expressed and incorporated into committee deliberations. It will be important to balance opportunity for public participation with the need for timeliness.

Implementation of Criteria for Prioritization

As noted above, the application of formal methods for estimating the value of research is challenging and unlikely to be adopted without modification by a deliberative body that must generate a list of priorities. However, the basic principles underlying formal value-of-research approaches can guide practical efforts.

Sometimes considerations that are not incorporated into a formal value of research approach are important for decisions. For example, a greater priority might be placed on research that addresses the needs of a population relatively underrepresented in previous research studies. There might be reason to believe, for example, that the responses of women and men to some cardiovascular treatments differ, yet the effects of those interventions in women have been studied little. Or treatments of conditions that disproportionately affect a racial or ethnic minority might have been under-studied. It is important to recognize these issues during prioritization deliberations.

A procedure for incorporating value-of-research principles into prioritization decisions is shown in Figures 1 and 2 on the next page.[6] The first of these figures displays a procedure for rating the benefit of the research, and the second a scoring scheme for characterizing the feasibility or cost of performing a study. When these two factors are combined, this framework can guide a group to make decisions under conditions of imperfect and inexact information.

The contrast between the quantitative emphasis of VOI techniques and more common approaches to prioritization of research are perhaps best illustrated by comparing the VOI to the typical approaches used by NIH study sections—groups of experts that review grant applications—in prioritizing research. These involve: 1) creation of grant opportunities either through more or less targeted requests for proposals, with a general emphasis historically on investigator-initiated proposals, 2) peer review of proposals, presented in a standard format, with an emphasis on shared criteria for review, and 3) prioritization of proposals based on significance, approach, innovation, investigators, and environment. The format of standard applications, with sections in the research plan including background and significance, preliminary studies (often including a discussion of the investigators prior work

6. Adapted from a proposed CER prioritization approach developed by a working group of industry and academic leaders, facilitated by McKinsey & Company.

FIGURE 1

Assessing the expected benefit of CE research

	Component	Question(s)	Example measures / considerations	Rating (1-5)*	Weight**
Magnitude assessment	Potential health impact from therapy/ intervention	Is there potential for large clinical impact due to use or nonuse of the intervention, based on _ The overall disease/condition and how prevalent and "severe" it is? _ Potential of the intervention to impact the patient's health state with disease/condition?	_ Incidence / prevalence _ Mortality / morbidity _ Quality of life / disability estimates _ Existing study outcomes / evidence		25%
	Potential cost impact from therapy/ intervention	Is there potential for large cost impact due to use or nonuse of the intervention, based on _ The overall economic burden of the clinical condition, to patient, payor, and society? _ Cost burden of therapy: in absolute terms, or relative to disease/condition or comparator(s)?	_ Economic cost (of disease to society) _ Expected spend on therapy (absolute, % of disease spend, % relative to alternatives)		25%
Evidence assessment	Variation/ lack of evidence	For the disease/condition, are there large potential consequences for a change in practice based on CE/HTA output, for example, either _ Significant variability in clinical use? _ Clear lack of evidence with respect to benefits and/or harms despite widespread use?	_ Variability in use or penetration _ Number and relative complexity of treatment options _ Level of evidence		40%
	New evidence/ debate trigger	For the therapy or disease/condition, is there either _ New evidence that has come to light? _ Escalation in the debate or discussion about one or more therapies' merits?	_ Activity in literature _ Expert opinion/ consensus		10%
			Estimated expected benefit (weighted average)**		

* 1 = Lowest, 2 = Lower, 3 = Medium, 4 = Higher, 5 = Highest

** Weighted average score calculated as the sum of component ratings multiplied by weights listed

FIGURE 2

Assessing the feasibility of conducting CE research

Feasibility assessment must answer the following questions

- Is there a specific, well-defined clinical question — including intervention, indication, population, comparators, etc.?
- Is the clinical question 'answerable'? (i.e., answer above a minimum threshold of confidence available in a relevant timeframe)
- Is the answer 'actionable'? (i.e., answer is sufficiently valid, reliable, relevant and timely to permit stakeholders to act)
- What research activities will be required to achieve this potential answer? How much will it cost, and how long will it take?

Feasibility can be rated by selecting the best-fit category from the table

Feasibility rating	0 (Infeasible)	1 (Lowest)	2 (Lower)	3 (Medium)	4 (Higher)	5 (Highest)
Description	Infeasible to obtain an assessment that yields a sufficient answer within a time horizon where the information will still be relevant _ E.g., a 4-year multiple arm study is required but set of approved comparator therapies is likely to have changed by that time	Feasible only with complex, long-duration, multicenter, or very resource-costly trial(s)	Feasible only with larger or more complex trial(s)	Feasible by generating new trials data that is relatively easy to obtain	Feasible with existing evidence base plus new analysis of existing data	Feasible without generating new 1º or 2º research data given existing evidence base
Likely timeline to results		>5 years	2-5 years	1-2 years	<1 year	<1 year
Order of magnitude cost		≥$100M	$50-100M	$5-25M	$0.5-2M	<$0.5M
Research type required (examples only)		Multi-year, multi-center, multiple-arm randomized controlled trial; multiple trials	Simple/ conventional RCT, with or without supporting observational studies	Short time-horizon trials (including pragmatic RCTs), new registry creation	Retrospective study using existing data, modeling, analysis of pooled clinical trials data	Systematic review, meta-analysis only
'Real world' example		ALLHAT**	Typical phase 3 trial +/- surveillance program	NCDR ICD registry	CABG vs. PCI for multivessel disease, Lancet 2009;373**	Standard AHRQ EPC project**

Expected shift in feasibility over time, with investments in supporting infrastructure for CE research

* A typical CER / technology assessment will always include a systematic review, and may well involve multiple studies at various levels of cost

** ALLHAT 8 years, ~$150M in 2008 dollars; Standard EPC project ~$250-400k; Lancet CABG vs. PCI study estimated value of work $600-700k

It is crucial that its members be selected for their expertise, understanding of the goals of comparative effectiveness research, and for their judgment.

and qualifications), and methods, and supporting documents such as biographical sketches letters of support, budgets, etc. are aligned with this basic approach.

Although the NIH approach seems to differ significantly from a strict value-of-research approach, there are several parallels. For example, the background and significance portion of a grant application often emphasizes that the research in question addresses a problem that is costly, imposes a large burden of illness, addresses a toxic intervention or an alternative to a toxic intervention, and so on. This part of the grant application frequently argues that the results of the proposed study will reduce uncertainty about the problem of interest and will change decisions related to that problem. The possible outcomes and the potential benefits of the research are key aspects of a value-of-research calculation. Indeed, when detailed information needed to calculate the expected value of research is missing, factors such as the overall burden of illness that are often cited in the grant applications are often the best information available to address this key consideration.

Priority-setting will ordinarily include considerations of study design, which can be informed by a value-of-information approach. Just as the probability of success would figure prominently in a value-of-research context, reviewers in a study section would assess the appropriateness of methods to the question being asked. Accordingly, specific study methods will necessarily vary with the specific intervention or clinical question under consideration. For example, the key uncertainty about a surgical treatment might be the complication rate in the "real-world" setting, since the surgeons and hospitals performing the operation differ from those that participated in the published studies. In this case, analysis of a registry or other observational database might be sufficient to address the key questions. In other circumstances, a randomized trial might be necessary. In still other circumstances, studies addressing more narrowly targeted questions – such as effects of an intervention on quality of life – might be needed to complement existing studies. In the end, the critical question to ask is whether the results of the study would be reasonably expected to change the decisions made about an option, recognizing that the relevant decision might simply be to pursue additional research. This implies that, in practice, judgment will be needed to determine the best study design, including the sample size and duration of follow-up, in each situation in order to set comparative effectiveness research priorities.

The priority given to a study will of course also be determined by how long a study might be expected to influence clinical decisions, which can vary greatly, depending on the rate at which new data is being generated in a field and the frequency with which new alternatives become available. The priority a study is given might also depend on the potential to improve the study by revision. The high rate at which NIH study sections ask for revision of proposed studies as part of the process of reviewing grant proposals is one indication of the importance of considering the potential for constructive revisions in prioritizing CER. A related issue is whether similar research is being performed by others, which directly affects what the NIH terms the degree of "innovation" of a study. Often a key question for comparative effectiveness research studies will be whether the same research might be done by private entities, such as a pharmaceutical company with a private interest in studying the question.

Recommendations for Setting Priorities for CER

Practical Consideration of Value of Information Principles

The principles of value of information approaches should be considered when approaching prioritization decisions. Their application can be informal and qualitative or more formal and quantitative when time, the availability of data, and other practical concerns allow. Measures such as the cost or burden of illness and the likelihood that a research study might change decision making should be considered, as should qualitative but structured approaches such as the approach illustrated in Figures 1 and 2. In some instances, it might even be possible to execute formal value of information studies, as NICE has done. When and how such studies should best be executed and what their value is in practice is an important area for study, but investing effort now in empirical applications of VOI that could be useful in informing future decisions about CER would provide valuable data on the potential of formal VOI methods to inform priorities for CER.

Composition of Priority-Setting Body and Management of Conflict of Interest

The priority-setting body, referred to here as the Advisory Committee on Comparative Effectiveness Priorities (ACCEP), will generally serve in an advisory capacity to any group that has overall responsibility for the comparative effectiveness research effort. Despite its advisory role, ACCEP will have substantial influence over the scope of the comparative effectiveness research effort. Thus it is crucial that its members be selected for their expertise, understanding of the goals of comparative effectiveness research, and for their judgment. Furthermore, they should represent broad interests. We believe it is crucial for most ACCEP members to have deep expertise and for all members to have basic familiarity with issues, such as study design and methods for measuring health outcomes.

Those without the relevant disciplinary skills could receive training through an orientation effort that would help expand their clinical and methodological capabilities. Orientation activities could focus on learning to interpret information used to determine health benefits that could be derived from the study of a particular topic and to reach judgments about the probable costs of such studies.

In deciding on the size of the advisory committee, there is a need to balance considerations of efficiency and representativeness. It is not possible to create a committee large enough to ensure direct representation of all relevant interests – such as every medical specialty or patient advocacy organization – nor does inclusion of a member of every group ensure that all members of that group feel well-represented. Consequently, committee

We suggest that ACCEP consist of between 12 and 20 members, with ex officio representation of stakeholders from key government agencies (AHRQ, NIH, CMS, VA, FDA, DOD, and the office of the Secretary of HHS), along with methodological and clinical experts, and representatives of consumers, device and drug manufacturers, private health insurance plans, and employers.

membership alone is not sufficient to ensure that input will always be obtained from every interested party. We suggest that ACCEP consist of between 12 and 20 members, with ex officio representation of stakeholders from key government agencies (AHRQ, NIH, CMS, VA, FDA, DOD, and the office of the Secretary of HHS), along with methodological and clinical experts, and representatives of consumers, device and drug manufacturers, private health insurance plans, and employers. The Medicare Evidence Development and Coverage Advisory Committee (MEDCAC) approach can serve as a model. "At-large" members of MEDCAC are chosen largely for methodological and clinical expertise. This committee also features nonvoting, designated representatives of industry and consumers. The nonvoting members participate fully in the deliberations and are able to question experts and other members of the public who testify at MEDCAC meetings.

Any attempt to limit ACCEP membership to individuals who do not have real or apparent conflicts is not only likely to fail, but to hinder efforts to include diverse points of view.

Like any committee that provides advice about health care decisions and health care research, ACCEP must have well-defined rules for dealing with conflicts of interest. As noted above, conflict of interest (COI) is an area of controversy today, and many groups—including agencies within HHS, academic medical centers, medical specialty societies, other provider groups, the Institute of Medicine, the Association of American Medical Colleges, journal editors, and Congress – are actively reviewing COI policies. We believe that ACCEP should draw upon the discussions and recommendations of other groups in developing a COI policy. We believe that the policy, however, should reflect the principles described below.

First, as many groups recognize, both real and apparent conflicts of interest can arise not only from financial interests, but also from intellectual perspectives, fiduciary responsibilities from serving nonprofit organizations, and a host of other involvements that may lead to appearances that members approach their committee work with preconceptions or points of view that call their impartiality into question. We agree that such conflicts are germane to work on ACCEP and that they are likely to be unavoidable.

A second principle is that broad and diverse points of view should be represented in the ACCEP. Any attempt to limit ACCEP membership to individuals who do not have real or apparent conflicts is not only likely to fail, but to hinder efforts to include diverse points of view. Many observers have commented on the tension between the need for expertise and the financial and intellectual conflicts that are common among members of FDA's advisory committees. Although there have been several calls to exclude all individuals with financial conflicts from such committees, any attempt to implement such a policy would have to address the exclusion of most experts in the field. Some critics of conflict-free policies have also noted that a committee free of financial conflicts would not necessarily be free of intellectual bias, whose consequences for decision making can be just as great.

These considerations lead us to recommend that the membership of ACCEP be diverse and not necessarily free of conflicts, but include balanced interests. We would expect much of the membership to have few if any financial conflicts and to be selected for broad expertise rather than for knowledge of a specific (clinical) area; this is likely to minimize the importance of conflicts even among individu-

als who have financial or intellectual conflicts. As is typical for such bodies, individuals with direct conflicts in a specific area (for example, a major equity stake in a company whose product is likely to be a comparator in a topic under consideration) would need to be recused from voting. Individuals whose conflicts are too extensive and far-reaching (e.g., major investors in a broad range of health care technologies) might not be able to serve, but such a policy would allow individuals with limited conflicts to serve. All relevant conflicts should be fully disclosed; the form of disclosure and the level of detail should be consistent with best practices, which are still evolving but should give some indication of whether financial conflicts are major (i.e., greater than a fixed value) and the nature of any intellectual conflicts (such as speeches or publications on issues affecting the priorities that are considered). Many conflicts will be evident simply by making complete biographical information publicly available.

Procedures for Priority Setting

Because no committee can possess all of the knowledge and expertise that is needed to consider the full range of possible comparative effectiveness research topics, ACCEP will need to follow procedures that ensure that there will be extensive public input into the priority-setting process. Fortunately, there is ample precedent for such a task. AHRQ has a well-developed approach for soliciting topics for study that can be easily modified to guide the operations of ACCEP. Like other such processes, AHRQ's approach goes to great effort to ensure that key individuals and groups with an interest in a general area of health care have an opportunity to offer suggestions and to provide comments about potential topics.[7] AHRQ solicits the general public and stakeholder groups for topic nominations, which are posted online and reviewed by program staff. Indeed, throughout the evidence evaluation process stakeholder input is solicited, which not only influences topic selection but can affect the study design, the outcome measures used, and the set of comparators.

The Institute of Medicine's Committee on Comparative Effectiveness Research Priorities engaged in a similar approach on a compressed time scale. The experience of the IOM committee and other bodies underscores the importance of staff review following topic solicitation. Staff may need to edit or reformulate questions to make them more specific, combine similar topic nominations where appropriate, and classify the topics into broader categories. AHRQ lists 14 priority areas, for example, into which many topics fall.

After this initial processing, we recommend following these steps:

1) Assembly of "dossiers" for each topic nomination by staff. Information in the dossiers includes all of the characteristics that would be used to determine the benefits to be derived from the information. This would include estimates of the prevalence of the condition studied in various population groups; a description of the potential comparators; an assessment of the current state of knowledge; measures of the health consequences of the condition; and identification of areas of uncertainty. In addition, the dossier would describe potential study designs and would include information that would be helpful in estimating the costs and likely outcomes of different study options.

2) Following a variant of the procedure used by the IOM Committee on Comparative Effectiveness Research Priorities, ACCEP members would engage in an initial prioritization ranking procedure from the universe of topics presented by staff. For this step of the process, the members of the committee would each be given points (e.g., twice as many points as the number of topics) which they would then allocate to the different topics. In the IOM procedure, the maximum number of points that a committee member could allocate to any topic was 10 percent of the total points available.

7. A description of the AHRQ process can be found at http://effectivehealthcare.ahrq.gov/aboutUs.cfm?abouttype=program#Topic.

3) Following this ranking procedure, the highest ranked topics (perhaps 30 or 40) would be discussed in a full committee meeting in which more detailed information about each topic could be presented and discussed. After the discussion, a second ranking procedure, similar to the one described above, would be performed.

4) After the second round of the ranking exercise, there would be a portfolio balancing step in which the top ranked priorities would be evaluated for coverage of key issues (such as suitability in under-represented populations) and for costs, and the committee could vote to reorder the choices to reflect these considerations.

There should be opportunities for public input during multiple steps of the process. We would suggest that written input be solicited before the initial ranking vote, and that there be opportunities for public participation in the meetings at subsequent stages of the process.

Infrastructure and Methodology

The priority setting process is highly dependent upon decisions made about infrastructure support and methodological developments. The principles guiding prioritization include assessment of both the benefits and the costs of each study. These costs and benefits, in turn, depend greatly upon the comparative effectiveness research infrastructure and the methodologies that are employed. For example, the costs of performing many kids of comparative effectiveness research studies will be greatly decreased if the federal government invests heavily in the dissemination of electronic health records and their linkage into a distributed data network that can be used to support comparative effectiveness research. Not only could such data be used to construct virtual registries of patients treated with particular surgical procedures or drugs, but the health records could be used to reduce the costs of carrying out traditional randomized trials or "cluster-randomized trials," in which hospitals or regions could serve as the units that are randomized. Similarly, methodological advances could make it possible to draw conclusions from observational studies that are more robust and can be more easily generalized than has been possible in the past.

We believe that discussion of investment in methodological research and in infrastructure should be part of the priority-setting process.

We believe that discussion of investment in methodological research and in infrastructure should be part of the priority-setting process. It may be better, for example, to defer a comparative effectiveness study of a specific technology until a linked database can be created that would dramatically reduce the cost of carrying out the study. Similarly, investments in methodology development could pay off quickly enough to lead to a shift in the priorities assigned to different topics, so the study of one topic that would otherwise rank highly would be deferred until key methodological questions are resolved.

Keeping Comparative Effectiveness Research Timely

Because medicine is characterized by rapid change and ongoing production of new technologies and new information, it is crucial to ensure that comparative effectiveness research not only delivers information in a timely manner but is updated frequently to incorporate new information promptly. We believe that prioritization, therefore, should be carried out at least semi-annually, and that high priority topics should be updated frequently. For example, if comparative effectiveness research funds are allocated to a large trial of alternative treatments for localized prostate cancer, there should be

ongoing evaluation of new information relevant to this topic coming from other sources. For comparative effectiveness research that consists primarily of evidence reviews, it will be particularly important to update the reviews on an ongoing basis, rapidly incorporating the results of new studies.

Conclusion

A value-of-research framework is often implicit in the ways that research prioritization decisions are made, even though the formal framework is rarely used in practical decision- making. The framework provides a set of questions that should always be answered as part of a priority-setting effort, even when reliable estimates of key parameters of value of research calculations, such as the degree to which the distribution of net effect sizes will change with the research, are unavailable.

A broad group representing diverse areas of expertise and interests is best positioned to make prioritization decisions; the group must be closely coordinated with the relevant federal agencies and with the potential users of the information. The procedures it uses to make prioritization decisions should be open, transparent, and incorporate comments and other information from the public.

We believe that prioritization, therefore, should be carried out at least semi-annually, and that high priority topics should be updated frequently.

Acknowledgements

The authors are grateful to Elizabeth Nabel, Harold Sox, Tim Taylor, Mark McClellan, and Joshua Benner for their constructive comments on a previous draft of this paper.

Strategies to Improve Comparative Effectiveness Research Methods and Data Infrastructure

Sean R. Tunis, MD, MSc

Introduction

Effective CER will require new research methods for reaching conclusions about the benefits, risks, and costs of actual medical practices, and a much better data infrastructure to provide the foundation for this evidence. Achieving these technical objectives requires addressing the following issues in the design and implementation of CER studies:

- Meaningfully involve patients, consumers, clinicians, payers, policy makers and other relevant decision makers in key phases of CER study development and implementation
- Develop methodological guidance or "best practices" for the design of CER studies that reflects decision maker needs and balance internal validity with feasibility and timeliness
- Improve research infrastructure to enhance the validity and efficiency of CER studies

The approach to addressing each of these issues is informed by the understanding that the primary purpose of CER is to help health care decision makers – patients, consumers, clinicians, payers and policy-makers – make informed clinical and health policy decisions. An important corollary of this observation is that, in order to be useful for decision-making, the evidence generated through CER must be valid, relevant, timely, feasible and actionable. In order to balance all of those considerations, it will be necessary to go beyond the current approaches to conducting clinical and health services research; this research should not be designed and implemented within the traditional research community alone. The primacy of the needs of health care decision makers at multiple levels has important implications for collaboration, research methods, and infrastructure needed for CER.

Comparative effectiveness research can be performed using a broad range of established and emerging methods, which generally fall into five major categories, as shown in table 1, next page:

TABLE 1

Categories of methods for CER

- Systematic reviews of existing research, including meta-analysis
- Decision modeling, with or without cost information
- Retrospective analysis of existing clinical or administrative data, including natural experiments
- Prospective non-experimental studies, including registries, which observe patterns of care and outcomes, but do not assign patients to specific study groups
- Experimental studies, including randomized clinical trials (RCTs), in which patients or groups of patients are assigned to alternative treatments, practices, or policies

Ultimately, the ability to produce CER information of sufficient rigor, quantity, and relevance will depend on using the right data and methods, for the right question, at the right time. The focus of this paper is on improving methods and infrastructure for primary comparative effectiveness research, or the lower three categories in Table 1. Experimental studies will continue to be a crucial source of CER information, and for those questions that are best addressed with these methods, it is critically important to develop study designs and infrastructure that will generate credible and relevant information, as quickly and inexpensively as possible. Non-experimental approaches are also a useful tool for CER, and will become increasingly important as such methods continue to be refined. There have been important advances in methods that improve the validity of analyses of non-experimental data, considerable progress in the design and use of clinical registries, and significant technical advances (and increased funding) that will exponentially increase the availability of encounter-generated data (claims and electronic medical records).[1,2,3]

Some methodological, infrastructure, and policy challenges associated with non-experimental studies would benefit from more extensive analysis than is provided here. This paper also does not address the synthesis of existing research, including systematic reviews and decision modeling. A recent report from the Institute of Medicine (IOM) offers many important insights and recommendations for improving the quality and utility of systematic reviews and clinical guidelines.[4]

II. Meaningful Involvement of Decision Makers

Perhaps the most significant and common failure of much of the clinical and health services research done in the past—and a major explanation for the extensive gaps in knowledge about health care interventions—is the lack of sustained, meaningful engagement of health care decision makers in the design and implementation phases of CER studies.[5] To fulfill the objectives of CER, new strategies will be required to support highly diverse, multi-disciplinary collaborative working groups for CER projects.

1. Schneeweiss S, Seeger JD, Landon J, Walker AM. Aprotinin during coronary-artery bypass grafting and risk of death. *NEJM*. 2008;358:771-83; Schneeweiss S, Solomon DH, Wang PS, Brookhart MA. Simultaneous assessment of short-term gastrointestinal benefits and cardiovascular risks of selective COX-2 inhibitors and non-selective NSAIDs: an instrumental variable analysis. *Arthritis Rheum*. 2006;54:3390-8.
2. Schneeweiss S, Seeger JD, Landon J, Walker AM. Aprotinin during coronary-artery bypass grafting and risk of death. *NEJM*. 2008;358:771-83; Schneeweiss S, Solomon DH, Wang PS, Brookhart MA. Simultaneous assessment of short-term gastrointestinal benefits and cardiovascular risks of selective COX-2 inhibitors and non-selective NSAIDs: an instrumental variable analysis. *Arthritis Rheum*. 2006;54:3390-8.
3. Institute of Medicine. The Learning Healthcare System: Workshop Summary. Washington, DC: The National Academies Press. April 2, 2007.
4. Institute of Medicine. Learning What Works Best: The Nation's Need for Evidence on Comparative Effectiveness in Health Care. National Academies. September 2007. http://www.iom.edu/Object.File/Master/57/393/Comparative%20Effectiveness%20White%20Paper%20(F).pdf. Accessed 1 July 2008.
5. Tunis SR. A Clinical Research Strategy for Shared Decision Making. Health Affairs. 2005; 24(1): 180-184.

Recommendation 1:
The Agency for Health Care Research and Quality (AHRQ) should conduct a systematic assessment of best practices for effective engagement of decision makers during various stages of clinical and health services research, including in priority-setting, protocol development, study implementation and dissemination.

Recommendation 2:
As a condition of receiving federal funding for any CER study, the investigators must form a stakeholder advisory committee comprised of at least five individuals who represent groups directly affected by the research, and whose specific functions should be determined based on the findings of the review in Recommendation 1.

Among the elements of the CER process that will require effective dialogue and consensus are:

- Selecting and prioritizing important research questions
- Refining research hypotheses and arriving at the specific questions to be addressed
- Feedback on specific elements of draft study protocols, including patient inclusion criteria, outcomes of interest, and methods. This will include advice on how to ensure that important patient subgroups are analyzed.
- Techniques to enhance enrollment of patients and clinical investigators in the trials
- Use and protection of patient-level information in administrative and clinical databases when used for research
- Strategies for effective dissemination of the results

Categories of participants to be included in these CER working groups include:

- Patients and consumers (representative of the general public)
- Practicing clinicians
- Medical professional organizations
- Evidence review groups (The Cochrane Collaboration, AHRQ's Evidence-based Practice Centers)
- Federal Agencies (AHRQ, NIH, FDA, Centers for Disease Control and Prevention, Veterans Affairs, Department of Defense)
- Public and private payers/purchasers (CMS, Medicaid, VA, Wellpoint, Blue Cross and Blue Shield health plans, employers)
- Life sciences industry (drugs, devices, other products and services)
- Representatives of a study's research team

It will be particularly important to fully understand and incorporate the perspectives of patients and consumers in the clinically and technically complex discussions that take place with respect to the design of health research. Therefore, it will be necessary to conduct careful analysis of existing models of collaborative clinical and health services research in order to identify best practices and lessons learned, resulting in a template for successful approaches that could be adopted for all CER projects. Some insights for effectively engaging patients and consumers can be gleaned from the work of the National Breast Cancer Coalition, the HIV/AIDS community, the Juvenile Diabetes Research Foundation, Consumers United for Evidence-based Medicine, and the Citizens Council of the National Institute of Health and Clinical Excellence.[6]

There does not appear to be any published systematic assessment of the best practices identified by these organizations with respect to the engagement of patients and consumers, nor is there much literature on techniques for effectively engaging

6. United States Cochrane Center. Consumers United for Evidence-based Healthcare (CUE). http://apps1.jhsph.edu/cochrane/uscccc.htm. Accessed May 25, 2009.

other types of decision makers in clinical or health services research, so further work on these issues would be extremely helpful. Even in the absence of fully developed functional models for collaborations, all CER projects should be able to clearly describe the mechanisms that are used to ensure that relevant expert, stakeholder, and decision maker perspectives are adequately considered in the design, implementation and dissemination of their work. CER stakeholder advisory committees are one mechanism to accomplish this objective.

Members of the advisory committee should be determined based on the specific health intervention under study and should include representatives of those groups most directly affected by the results. For a decision regarding initial treatment of prostate cancer, for example, the advisory committee might include a representative from a prostate cancer patient advocacy group, a prostate cancer survivor, a representative from a professional society of urologists, the medical director from a private health plan, and the author of a recent review on the clinical effectiveness of existing strategies. Members of the advisory committee are not members of the investigator team, but the committee's recommendations on each phase of the study outlined above should be documented by the investigators in all study reports and publications.

III. Developing Evidentiary Guidance for CER

Limitations of the traditional hierarchy of evidence

A crucial requirement of effective CER will be to employ the best possible analytic methods and data in studies of clinical and health policy questions. Historically, the best possible evidence has come from randomized clinical trials, which sit atop a "hierarchy of evidence." Expert opinion is the least desirable source of evidence in the hierarchy, and non-randomized – or observational – studies are in the middle. With the emergence of new questions, new data sources, and improvements on methodologies, this hierarchy is coming into question as a hard and fast rule for conducting research. Indeed, one of the major limitations of past efforts to generate evidence about "what works in health care" was the application of an inadequate range of analytic tools and insufficient attention to ensuring that the research process began with defining questions from the perspective of (and with the involvement of) the decision makers who would eventually be using the study results.

The decade-long struggle of public and private payers, both in the U.S. and abroad, to make evidence-based policy decisions on the use molecular imaging (primarily FDG-PET scanning) in oncology exemplifies the problems resulting from the absence of a well-defined and broadly accepted evidentiary framework for conducting CER. PET scans produce images that reflect the metabolic activity of internal structures, and because many cancers are highly active metabolically, there has been great enthusiasm for use of this technology in managing patients with cancer. Despite the publication of hundreds of clinical studies on various diagnostic uses of PET, systematic reviews continue to observe that the available evidence of clinical utility is limited or poor quality for many common clinical uses. Over the past several years, Medicare has provided coverage for tens of thousands of PET scans in the context of a national PET registry intended to fill this gap,[7] but some analysts argue that such studies are of limited value in assessing the diagnostic value of imaging technologies.[8] The government of Ontario has also implemented a conditional reimbursement program for PET scanning, devoting considerable analytic and political resources to conducting randomized trials on the clinical utility of PET in the management of oncology patients.[9] More recently, the National Oncologic PET Reg-

7. Hillner BE, Siegel BA, Liu D, et al. Impact of Positron Emission Tomography/Computed Tomography and Positron Emission Tomography (PET) Alone on Expected Management of Patients with Cancer: Initial Results from the National Oncologic PET Registry. *J Clin Oncol*. 2008; 26(13): 2155-2161.
8. Levine MN, Julian JA. Registries that show efficacy: good, but not good enough. *J Clin Oncol*. 2008; 26(33):5335-5343.

istry (NOPR) has been modified to include linkages to Medicare claims data with the intent of monitoring longitudinal outcomes, though questions remain about the validity and credibility of evidence generated through that approach.[10]

All of this work is taking place in the absence of any shared understanding among researchers, decision makers, and other stakeholders about what standard should be applied to evidence of the clinical utility of FDG-PET for management of oncology patients. Some continue to advocate for RCTs, and others for more sophisticated registries. Strongly divergent views exist about whether it is reasonable to expect compelling, direct evidence that diagnostic technology improves patients' health outcomes, while others note that the absence of such information makes it impossible to determine whether imaging is justifiable. If measuring health outcomes is necessary, can one reliably derive such information from Medicare or other claims data? Further, even with outcomes information, in the absence of large randomized studies, can the impact of this diagnostic technology on outcomes be elucidated? Until there is some clarity on what type of evidence will be acceptable to various decision makers, it seems unlikely that additional CER studies in this area would make an impact on physician practice patterns.

The lack of a well-defined evidentiary framework to guide the design of CER studies is equally problematic in other important clinical areas, including cardiac imaging, genetic testing, radiation therapy for cancer, treatments for chronic wounds, complementary and alternative medicine, and disease management programs. In each of these areas, intense debate exists about which methods will yield evidence that is credible and relevant for decision makers. For the most part, there has been limited systematic effort to reconcile or align the competing views on what constitutes acceptable CER evidence, and research activity therefore is guided to some degree by what is possible, rather than what is most desirable. Indeed, there is little hope of using CER to improve patient outcomes or the value of health care with CER until there is a clearly articulated consensus on what methodologies and data sources will yield evidence upon which decisions should be made.

For the most part, there has been limited systematic effort to reconcile or align the competing views on what constitutes acceptable CER evidence, and research activity therefore is guided to some degree by what is possible, rather than what is most desirable.

Principles of an evidentiary framework for CER

1. Traditional "hierarchies of evidence" are overly simplistic and should not necessarily guide the implementation of CER.

The need to think more broadly about the widely used hierarchy of evidence was described in detail in a recent paper by Sir Michael Rawlins, Chairman of the National Institute of Health and Clinical Excellence (NICE) in the United Kingdom.[11]

9. Ontario PET Steering Committee. PET Scan Primer: A Guide to the implementation of positron emission tomography in Ontario. Ministry of Health and Long-Term Care. August 31, 2008. http://www.health.gov.on.ca/english/providers/program/ohip/outofcountry/pdf/pet_scan_primer.pdf. Accessed May 25, 2009.
10. Center for Medicare and Medicaid Services. Decision Memo for Positron Emission Tomography (FDG) for Solid Tumors (CAG-00181R). April 3, 2009. http://www.cms.hhs.gov/mcd/viewdecisionmemo.asp?from2=viewdecisionmemo.asp&id=218&. Accessed 10 May 2009.
11. Rawlins M. Harveian Oration: on the evidence for decisions about the use of therapeutic interventions. *The Lancet*. December 2008; 372 (9656): 2152-2162.

> "Hierarchies of evidence should be replaced by accepting—indeed embracing—a diversity of approaches. This is not a plea to abandon RCTs and replace them with observational studies. Nor is it a claim that the Bayesian approaches to the design and analysis of experimental and non-experimental data should supplant all other statistical methods. Rather, it is a plea to investigators to continue to develop and improve their methods; to decision makers to avoid adopting entrenched positions about the nature of evidence; and for both to accept that the interpretation of evidence requires judgment."

Many CER questions will not be adequately addressed by analyzing routinely collected data from large administrative databases or electronic medical records, given the widely recognized challenges of making valid comparisons between alternative strategies for managing health conditions, without being able to confidently account for unmeasured differences between comparison groups.

Rawlins correctly points to the need for a more cognitive approach to evidence-based policy making, and his comments mark a broader shift in thinking that recognizes the importance of both experimental and non-experimental methods in CER. Because knowing that an intervention works under ideal circumstances (efficacy) is necessary but not sufficient for evaluating what is appropriate for patients in real-world practice settings, answering CER questions will require a more nuanced approach to the generation and appraisal of evidence than is reflected in the widely used linear evidence hierarchy.[12]

2. A range of methods are important in CER

All five categories of methods in Table 1 have an important role in comparative effectiveness research. Many questions in health care, particularly regarding strategies for delivering care, can be adequately addressed with methods other than RCTs, and perhaps the most critical challenge related to the methodology of CER will be to reframe the debate from arguments about the inherent superiority of one method over another (the standard linear evidence hierarchy) to a more productive dialogue about applying the most appropriate method with a high degree of skill to important questions, and within the right timeframe.

An increasingly important adjunct to RCTs in the context of CER will be data collected during the delivery of and payment for health care. There have been important advances in methods that improve the validity of analyses of non-experimental data,[13] considerable progress in the design and use of clinical registries,[14] and significant technical advances and funding that will exponentially increase the availability of encounter-generated data (claims and

12. Atkins, D. Creating and Synthesizing Evidence with Decision makers in Mind, Integrating evidence from clinical trials and other study designs, medical Care, Volume 45, Number 10 Suppl 2, October 2007.
13. Schneeweiss S, Seeger JD, Landon J, Walker AM. Aprotinin during coronary-artery bypass grafting and risk of death. NEJM. 2008;358:771-83; Schneeweiss S, Solomon DH, Wang PS, Brookhart MA. Simultaneous assessment of short-term gastrointestinal benefits and cardiovascular risks of selective COX-2 inhibitors and non-selective NSAIDs: an instrumental variable analysis. *Arthritis Rheum*. 2006;54:3390-8.
14. Gliklich RE, Dreyer NA, eds. Registries for Evaluating Patient Outcomes: A User's Guide. Prepared by Outcome DEcIDE Center. AHRQ Publication No. 07-EHC001-1. Rockville, MD: AHRQ. April 2007.

electronic medical records).[15] As methods and data repositories continue to develop, non-experimental CER is likely to increase in importance. However, many important CER questions seek to detect relatively small but clinically important differences in treatment effects, and such questions may be less reliably answered with existing non-experimental methods.

While there are legitimate concerns about the generalizability and cost of RCTs, particularly those performed for regulatory approval, this has been interpreted by some to suggest that non-experimental methods will be the primary approach used for comparative effectiveness research. This perspective fails to recognize the techniques available to improve the applicability, generalizability, and efficiency of experimental methods, a topic discussed in greater detail below. In fact, many CER questions will not be adequately addressed by analyzing routinely collected data from large administrative databases or electronic medical records, given the widely recognized challenges of making valid comparisons between alternative strategies for managing health conditions, without being able to confidently account for unmeasured differences between comparison groups.[16] Furthermore, existing databases used for non-experimental studies may not include important data elements, such as primary and secondary health outcomes. Finally, while many decisions that will apply CER information will benefit from information generated by non-experimental studies, some clinical and policy decisions will require experimental evidence because of the potential to do harm by widely applying potentially incorrect findings from non-experimental studies.

A well-known example of this is the widespread use of hormone replacement therapy based on numerous, large, and consistent epidemiologic studies showing dramatic reductions in heart disease in women taking these medications. Subsequent clinical trials did not demonstrate cardiac benefits, but many women were likely harmed by following the unequivocal clinical and policy recommendations resulting from overconfidence in the size and consistence of the non-experimental research on this topic.[17] A common misconception is that a high volume of data (available from large databases) will allow researchers to identify and adjust for problems in the quality of that data, thereby deriving accurate conclusions, yet this adjustment is only sometimes possible. Therefore, a learning health care system will need more randomized controlled trials—relying on more efficient, larger, simpler, and pragmatic designs—and also a greater variety of other experimental and non-experimental methods. For example, "cluster randomization" of comparable groups of patients to alternative policies (e.g. formularies with different preferred drugs) may be more feasible than patient-level randomization to drugs, and may also provide more direct evidence on the comparative effectiveness of different formulary designs.

A learning health care system will need more randomized controlled trials—relying on more efficient, larger, simpler, and pragmatic designs—and also a greater variety of other experimental and non-experimental methods.

15. IOM. The Learning Health Care System. 2007
16. Califf RM. Issues facing clinical trials of the future. Journal of Internal Medicine. 2003; 254:426-433.
17. Hormone Therapy Woes. The New York Times. July 11, 2002. http://www.nytimes.com/2002/07/11/opinion/hormone-therapy-woes.html?scp=3&sq=hormone%20replacement%20therapy%20saga&st=cse&pagewanted=print. Accessed May 25, 2009.

3. The right approach to a given CER study depends on the circumstances.

Without a definitive framework to guide selection from a number of potential methods and data sources, how should CER studies be designed?

A number of contextual considerations should be used to guide the choice of research methods for a given CER question. These include the specific decision to which the evidence will be applied (and consequences of making the wrong choice), the nature of the intervention under study, the status of current evidence on the topic, and the time and resources required for alternative approaches.

Studies that are faster, less costly, and more generalizable will often be acceptable as long as there is room for error in the decision-making process. Where there is less tolerance for error, experimental research will still be necessary, and new design techniques and implementation strategies will be required to make this research less resource-intensive and more broadly applicable. While the choice of CER methods should not be based in large part on the skills and preferences of a particular investigator, it is also important that investigators have expertise in the study methods they use.

IV. Selection and Improvement of CER Methods

Consistent with the principles above, two forms of guidance to CER investigators are needed, in addition to continued investment in methodological research to enhance CER methods. This methodological guidance and innovation will enable the CER community to design studies that are accurately targeted to produce the information needed by patients, consumers, clinicians, payers and policy makers. If done through effective multi-disciplinary consultation, the study design recommendations will reflect collective judgments about the inevitable trade-offs between internal validity, feasibility, timeliness, and generalizability.

Recommendation 3: At least 10 percent of funds allocated to CER in the next 10 years should be directed to the Secretary of HHS for use in the development of methodological guidance and innovation. Funded programs should address the needs for:

- Objective reviews of the strengths and limitations of alternative methods for CER, including examples of their implementation and identification of categories of CER topics for which each is potentially appropriate. The Institute of Medicine, AHRQ or another organization capable of convening a broad range of methodologists, clinicians and other stakeholders should develop a "translation table" linking CER methods to specific types of research questions, as illustrated in the preliminary table presented above. This should build on existing work on methodological and reporting standards for each of the major CER methods. Once this is developed, all proposals for federal funding of CER should refer to this framework when describing the rationale for the methods proposed for their work.

- Guidance documents for the consideration and selection of methods and data sources for specific priority CER topics. As lists of priorities for CER research are produced by the Institute of Medicine and others, a process should be developed to identify appropriate research methods for each high-priority research question. This process would include methodologists and content experts, and would carefully consider the pros and cons of different CER methods, then make specific recommendations for study designs necessary to answer these questions. These general design recommendations would be included in the government requests for proposals. The investigators who submit proposals would be free to propose any study design but would be expected to provide a rationale for deviating from the recommendations of the expert workgroup.

- Continued research and innovation to develop improved methods for experimental and non-experimental CER. AHRQ should develop and update annually a list of priority needs for research and development on the methods of CER, and award competitive grants for these studies.

Matching CER research questions to appropriate CER methods

One approach to matching CER methods to appropriate research questions involves a systematic assessment of the range of existing CER methods, accompanied by careful consideration of the types of research questions that these methods are particularly well-suited to address. This analysis should be applied to the entire portfolio of CER methods, including systematic reviews, decision modeling, retrospective analysis, non-experimental studies and experimental studies.

Each CER method offers a potentially useful approach to generating CER evidence for decision-making. Since most methods that will be commonly used for CER – including methods that involve randomization – will raise some concern that they are not as methodologically robust as traditional RCTs, special attention will be necessary to ensure that the methods used will stand up to the level of scrutiny that is inevitable if these studies are going to inform important clinical and health policy decisions. It will therefore be useful to have a framework that articulates how to select an appropriate study design for specific types of CER questions as well as best practices for those study methods to help facilitate credibility and timeliness in CER.

A number of organizations have developed consensus reporting guidelines for various types of clinical and health services research (CONSORT, MOOSE, GRACE, etc.), and considerable work has been done to describe best practices in the conduct of most of the major methods for CER.[18] These activities have been extremely valuable in improving the quality and consistency of such research. In addition to having standards for high-quality design, conduct and reporting of CER, there will also need to be a well-defined process for developing consensus about the appropriate use of each of those methods in addressing specified types of CER questions. This consensus must include both methodologists and decision makers, since it is not simply a function of the quality of the research but also the adequacy of that research to inform clinical and policy decisions.

Eventually, it will be useful to be able to complete a "translation table" similar to the basic draft below, which includes a few proposed uses for some categories of CER studies. This initial draft table was developed based on some preliminary work done by several members of an informal workgroup of methodologists known as the CER Innovation Collaborative (CER-IC) convened by the Institute of Medicine's Evidence-Based Medicine Roundtable.

The table represents some initial thoughts only, and a thorough structured process involving experts familiar with each type of method will be required to provide useful methodological guidance to the CER community. It is presented here in this preliminary form for illustrative purposes only, mainly to highlight the fact that CER methods must be carefully chosen for each CER question, since these methods have inherent properties that render them more or less well-suited to specific circumstances. Applying the wrong method to an important question is likely to produce results that are not sufficiently credible, relevant and timely, wasting resources.

18. EQUATOR Network. http://www.equator-network.org/. Accessed May 25, 2009.

TABLE 2

CER methods with examples matched to potential research questions

	Description	Example	Advantages
Pragmatic clinical trials	These RCTs are designed to demonstrate how a medical intervention works in a typical, real-world setting. Features of these trials can include all or a combination of the following: relaxed inclusion/exclusion criteria, relaxed protocol, longer term endpoints, active comparators, and outcome measures of relevance to patients, payers, and physicians.[19]	Are newer types of antihypertensive agents, which are currently more costly to purchase on average, as good or better than diuretics in reducing coronary heart disease incidence and progression?[20] The ALLHAT study used patient relevant outcomes, had minimal inclusion criteria, and there was some flexibility in the dosing of the therapies.	Because these trials are designed to meet the needs of decision makers, the results tend to be more generalizable, the outcomes are useful to patients and physicians making tough clinical choices, and the trial maintains all or much of the scientific rigor of traditional RCTs.
Cluster RCTs	Groups of people are randomized to an intervention instead of randomizing individuals. These groups can be, for example, communities, regional payers, purchasers, delivery systems, clinics, etc. Individuals within a cluster will tend to resemble each other, which needs to be taken into account in the statistical analysis.	What is the comparative effectiveness of the American Cancer Society smoking cessation program versus the American Lung Association smoking cessation program?[21] In this study, different clinics adopted different smoking cessation programs.	This approach is ideal for comparing alternative, established therapies with true equipoise, rather than new therapies, and common therapies rather than novel or high-profile therapies. Cluster RCTs can also provide a rigorous evaluation of the effectiveness of therapies in real-world settings, especially when consent by cluster is acceptable.
Bayesian / Adaptive trials	Unlike traditional RCTs, the Bayesian approach makes use of prior information on a medical intervention to estimate a prior distribution. This prior information is then combined with trial data to create a posterior distribution. Trial data can be analyzed frequently and compared to the prior information to inform the direction of the study.	What is the most effective treatment method for patients with a given biomarker profile?[22] The authors propose to adaptively randomize patients to one of four treatment groups with allocation based on prior accumulated data.	The Bayesian approach incorporates prior information into the data analysis, which can lead to mid-course modifications to the trial design, and potentially avoid the need to start new separate trials to reflect knowledge gained during the course of the trial.
N-of-1 trials	N-of-1 trials are single event case studies to look at the effect of an intervention in an individual. Generally, there are two or more periods, alternating when the participant receives the therapy and one where he does not. This allows physicians to look for clinically meaningful differences in outcomes.[23] Multiple N-of-1 trials can be combined to estimate population effects	What is the optimal dose of drug x in patient y that effectively balances drug's efficacy with its side effects? (A number of other study questions are described in a manuscript by Guyatt et al., 1990).[24] In general, this design is best for chronic and relatively stable conditions.	There is an emphasis on optimizing effectiveness for the individual rather than for a population of patients. This design allows researchers to obtain information on individual treatment response. This approach is useful for chronic conditions with readily assessable primary therapeutic effect.
Delayed-design or "advance coverage" trials	Many variations exist. The most common version for this design is that participants are randomized to either receive the intervention from the start of the trial, or to have the intervention withheld for a pre-specified amount of time. By the end of the trial, both study groups have received the study intervention.[25]	This trial design has been employed for several studies of neuroprotective treatments for Parkinson disease.[26] For example, what is the comparative effectiveness of early versus later initiation of rasagiline on progression of disability in patients with Parkinson disease?[27]	All participants are eventually given the potentially beneficial medical intervention, which overcomes some of the ethical concerns raised by traditional RCTs, while maintaining a control group. This approach may be particularly useful for CER when applied to cluster RCTs.

19. Thorpe KE, Zwarenstein M, Oxman AD et al. A pragmatic-explanatory continuum indicator summary (PRECIS): a tool to help trial designers. *J Clin Epidemio*l. 2009;62:464-75.
20. Davis BR, Cutler JA, Gordon DJ, Furberg CD, Wright JT, Jr., Cushman WC, Grimm RH, LaRosa J, Whelton PK, Perry HM, Alderman MH, Ford CE, Oparil S, Francis C, Proschan M, Pressel S, Black HR, Hawkins CM: Rationale and design for the Antihypertensive and Lipid Lowering Treatment to Prevent Heart Attack Trial (ALLHAT). ALLHAT Research Group. Am J Hypertens 1996;9:342-360.
21. Lando HA et al. Comparative Evaluation of American Cancer Society and American Lung Association Smoking Cessation Clinics. Am J Public Health 1990; 80:554-559.
22. Zhou X et al. Bayesian adaptive design for targeted therapy development in lung cancer-a step toward personalized medicine. Clinical Trials 2008; 5:181-193.
23. Guyatt GH and Jaeschke. N-of-1 Randomized Trials-Where do we stand? West J Med 2009; 152: 67-68.
24. Guyatt GH et al. The n-of-1 randomized controlled trial: clinical usefulness-Out three year experience. Ann Intern Med 1990; 112:293-299.
25. Maclure M, Carleton B, Schneeweiss S. Designed delays versus rigorous pragmatic trials: lower carat gold standards can produce relevant drug evaluations. Medical Care 2007;45:S44-49
26. Clarke CE. Are Delayed-Start Design Trials to Show Neuroprotection in Parkinson's Disease Fundamentally Flawed? Movement Disorders 2008; 23: 784-789.
27. Parkinson Study Group. A Controlled, Randomized, Delayed-Start Study of Rasagiline in Early Parkinson Disease. Arch Neurol 2004; 61: 561-566.

Although the table presented above focuses on issues associated with experimental methods for CER, it is equally important to apply a systematic analysis to determine the types of CER questions that are best addressed with non-experimental methods, including retrospective analysis of claims and electronic medical record data, decision modeling, cost-effectiveness analysis. The ultimate goal is to ensure that high-priority CER questions are consistently investigated with methods that can provide relevant and actionable information.

Given the differential response of patients to the same intervention and the imminent advances in personalized medicine, efforts to develop a methodological framework should focus on the utility of CER methods to provide evidence that is relevant to defined subgroups of patients. Significant investment will be required to develop and refine CER methods that are better suited to address questions at the subgroup level, and to achieve adequate statistical power to make valid inferences for such subgroups.[28]

Study design recommendations for specific categories of health care interventions

The discussion above focused on the conceptual work needed to determine the appropriate uses of existing and emerging CER methods. This will lead to an approach to the structured thinking necessary to match any given high-priority CER question with a subset of methods that are most likely to generate useful evidence on that question. The categories of available methods primarily address alternative mechanisms by which patients are assigned to study groups, not more specific elements of study design that will sometimes vary in predictable ways depending on the technology or service that is the subject of study.

The value of developing consensus regarding key features of CER study protocols can be illustrated with the example of treatments intended to improve healing of chronic wounds, such as pressure ulcers or diabetic ulcers. AHRQ has commissioned two systematic reviews of negative pressure wound therapy for chronic wounds, each of which identified significant limitations in the quality of the existing clinical studies. Among a number of recurring deficiencies were the decisions about which patients to include and exclude form the studies, the adequacy of care provided to the patients receiving standard care (i.e., choice of the comparator intervention), and the choice of primary outcomes. Furthermore, these and other reviews of wound care interventions generally exclude non-experimental studies because of concerns about the potential for important unmeasured baseline differences in treatment groups that might explain differences in reported outcomes. This is consistent with the FDA guidance document on treatments for chronic wounds, which also emphasizes the need for RCT designs in order to make reliable treatment comparisons.[29]

The manufacturers of treatments for chronic wounds have different perspectives about what constitutes adequate evidence to demonstrate clinical benefit for regulatory or reimbursement purposes. And many public and private payers provide coverage for and spend considerable sums on treatments for chronic wounds that do not meet the evidentiary standards applied in the systematic reviews of these treatments. All of this creates an environment of tremendous uncertainty for researchers planning to design future CER studies for treatment of chronic wounds. Without some effort to systematically gather and integrate different perspectives on the design of studies for treatment of chronic wounds, future studies will reflect a range of different designs, perpetuating the current situation in which the overall quality of evidence is poor and not aligned with what decision makers would most like to know.

28. Garber AM, Tunis SR. Does Comparative Effectiveness Research Threaten Personalized Medicine? *NEJM*. 2009; 360(19):1925-1927.
29. Food and Drug Administration. Center for Biologics Evaluation and Research (CBER).Guidance for Industry: Chronic Cutaneous Ulcer and Burn Wounds—Developing Products for Treatment. Updated June 1, 2006. http://www.fda.gov/cber/gdlns/ulcburn.htm. Accessed February 18, 2009.

In order for there to be rational investment in CER studies of chronic wounds, it would be valuable for experts, stakeholders and decision makers to reach agreement on best practices for designing these studies. Are there acceptable alternatives to traditional RCT methods for comparing interventions? Would adaptive designs be a potential option for greater efficiency and lower costs? Is it possible to better formalize the elements of standard care for purposes of the control-group intervention? Are there specific primary outcome measures that are preferable to others that might be used? For head-to-head comparative studies, are the alternatives to very large, blinded RCTs that would provide reasonable evidence on the comparative risks and benefits of competing forms of the same underlying intervention? Answers to these questions, developed through dialogue among clinical experts, researchers and decision makers, are essential to guiding investments in CER studies for chronic wounds. Without better defined, shared principles on critical elements for CER research protocols, there is a high risk that the investment in CER studies will not be meet the requirements of internal validity, feasibility, timeliness and relevance.

The same lack of agreement on best practices has hindered the collection of evidence on non-invasive imaging in the diagnosis of coronary artery disease (coronary CT angiography, or CCTA). Several recent systematic reviews of CCTA noted limitations in the existing evidence, with most studies being single center trials enrolling high risk patients and reporting only on diagnostic accuracy compared to invasive angiography. Medicare came to similar conclusions after reviewing the evidence in March 2008, and concluded that large, rigorous RCTs with measurement of hard clinical endpoints (like cardiac death and acute myocardial infarction) would be necessary to provide adequate evidence of the clinical utility of coronary CT angiography.[30] Many individuals from the cardiology community and imaging vendors firmly believe that the current evidence is sufficient to conclude that CCTA is an important advance in management of suspected coronary disease, and do not believe that additional studies are necessary.[31] The National Heart, Lung, and Blood Institute at the NIH is currently reviewing three proposals for large RCTs to study the clinical utility of CCTA, though results of these studies may not be available for five or more years. In the meantime, no progress has been made on addressing the underlying question regarding the appropriate design of studies to evaluate new and existing non-invasive cardiac imaging technologies. Once again, investments in CER studies by public or private entities will be challenging without greater clarity on the set of acceptable designs for these studies. We do not have a clear answer to the question of which studies are most informative for health care decision makers and can be completed at a time when the information is needed for clinical and policy decisions.

Progress on CER could be expedited by a process for developing technology-specific evidentiary guidance on major categories of health services and strategies. This guidance would need to be developed through a collaborative effort involving the full range of informed experts and stakeholders and would develop specific recommendations for the design of CER studies that would aim to balance of validity, relevance, feasibility, and timeliness. It may be challenging to align the interests of the different stakeholders to successfully reach consensus on CER methods, given the significant implications for patients, product developers, payers and other stakeholders. However, there are potential advantages of increased consistency, transparency and certainty for all stakeholders, which suggest that agreement on some common principles will be achievable.

30. Centers for Medicare and Medicaid Services. Decision Memo for Computed Tomographic Angiography (CAG-00385N). March 12, 2008. http://www.cms.hhs.gov/MCD/viewdecisionmemo.asp?id=206. Accessed April 14, 2009.
31. Douglas PS, Budoff M, Tunis S, et al. A new era for cardiovascular imaging? Implications of the revoked national coverage decision for CT angiography on future imaging reimbursement. *J Am Coll Cardiol Img*. 2008; 1:398-403

Intervention-specific evidentiary guidance would draw upon the methodological framework described above that matches various CER methods to the type of question for which those methods are and are not well-suited. These principles would be considered in the process of developing guidance for specific categories of technologies, since the type of research questions that will be important for those technologies can be matched to the various categories of available methods that may be appropriate. The guidance documents would provide greater specificity by seeking to create consensus not only on the mechanism used for patient assignment to treatment and comparison groups but also on all other elements of the study design.

Work in progress to develop methodological guidance for CER

The Center for Medical Technology Policy has been working on early prototypes of collaboratively developed Effectiveness Guidance Documents (EGDs) to provide specific recommendations to product developers and clinical researchers about the design of clinical studies that will produce the evidence desired by patients, clinicians and payers. Each EGD will focus on a specific category of health care technology. For example, draft documents are under development for treatment of chronic wounds, non-invasive cardiac imaging, and gene expression profiling for management of breast cancer. Additional topics to be addressed by the initiative include imaging in oncology and complementary and alternative medicine.

The goal is to describe clinical studies that would provide decision makers with a *reasonable level of confidence* that the technology improves health outcomes. In this respect, the guidance documents are intended to provide technology-specific methodological roadmaps for the design of prospective comparative effectiveness research. For therapeutic interventions, the primary focus will be on evidence of comparative clinical effectiveness, and for diagnostic interventions the primary focus will be on comparative clinical utility—how the diagnostic information affects clinical management and whether this leads to improved health outcomes.

EGDs are conceptualized to be analogous to FDA guidance documents, which are also targeted to product developers and clinical researchers, and provide guidance on the design of clinical studies that are intended to support regulatory decision making.

EGDs are conceptualized to be analogous to FDA guidance documents, which are also targeted to product developers and clinical researchers, and provide guidance on the design of clinical studies that are intended to support regulatory decision making. EGDs will serve a comparable function for product developers and clinical researchers, but are focused on the design of clinical studies to support "post-regulatory" decision-making. These post-regulatory decisions include individual clinical decisions made by patients and consumers, clinical recommendations made by clinicians, clinical policies generated by medical professional societies, and reimbursement decisions made by payers. Since there is no single organization that represents the universe of post-regulatory decision makers, neutral forums that bring all of the relevant perspectives into a sustained dialogue will be helpful for generating study design recommendations that align the information needs of decision makers, the diverse interests of stakeholders, and the research activities of the CER community.

By including the relevant FDA regulatory experts in the EGD development process, it is hoped that EGDs will reflect optimal alignment between study design elements intended for regulatory approval and those targeted to clinical and health policy decision-making. This may help to avoid the need for multiple studies to address these different evidentiary purposes.

These methodological guidelines would need to be developed through a collaborative effort involving the full range of informed experts and stakeholders and would aim to develop specific recommendations for the design of CER studies that would aim to balance of internal validity, generalizability, feasibility, and timeliness. These could be used by CER researchers conducting either publicly or privately funded research as one source of input during the process of protocol development. To the extent that they accurately reflect the information needs of patients, consumer, clinicians, payers, and policy makers, and to the extent that they successfully balance scientific and practical consideration, they should be a useful guide for designing CER studies.

A particularly important aspect of CER methods research will be the development of improved approaches to accounting for differences in treatment response in subgroups of patients enrolled in these studies.

Methodological research and innovation

Finally, efficient production of valid comparative effectiveness evidence over the long-term will depend on refinements in the scientific methods. For those CER questions in which primary research is necessary, methodological research will be needed to develop and refine methods that are efficient, valid, generalizable and relevant. Attention to methods research in these areas has been limited to date because the demand for these studies, and the funding available to support them, has only recently been identified as a public policy priority, mainly as part of the increased focus on CER.

Somewhat more funding and attention has been devoted over the past ten to twenty years to the use of non-experimental methods to address questions of comparative effectiveness, and substantial improvements in these methods have been achieved. There is optimism about the potential value of these methods, given the greater speed and lower costs associated with them, particularly those that use routinely collected data from claims and electronic medical records. However, there is also limited confidence in these methods, since there is considerable variation in the level of sophistication with which they are applied in the research community, and the resulting literature is of uneven quality. For these reasons, the impact of non-experimental evidence on health care decision-making will depend not only on having better methods but also on an expanded supply of skilled methodologists, principles with which to identify high-quality research, and a substantial change in the way that clinical policy-making organizations view and apply such evidence.

A particularly important aspect of CER methods research will be the development of improved approaches to accounting for differences in treatment response in subgroups of patients enrolled in these studies. With the rapid scientific discovery of genetic and molecular markers for the development of disease and responsiveness to treatment, the evidentiary framework for CER will need to be further

developed to produce information that is informative for homogeneous subgroups of patients as well as for individuals.[32]

A number of investigators are now working on methods research that can better address heterogeneity in clinical trials, during both trial design and analysis of trial results.[33] The FDA is also aware of the increasing interest in methods to reliably analyze clinical trial data to assess subgroup effects, with a focus on determining the potential importance of biomarkers in predicting different response to treatments, particularly in cancer.

The enthusiasm for determining subgroups in whom the benefits or risks of treatment are markedly different from the average must be tempered by the recognition that such findings may or may not prove to be reliable. Perhaps the most compelling illustration of the potential for error was provide by Richard Peto, who performed a retrospective subgroup analysis of data from a very large trial of interventions for the treatment of acute myocardial infarction, and found that the benefits of immediate aspirin therapy—which is consensus standard of care— were not measurable for patients belonging to 2 of the 12 astrological signs.[34] This highlights that large sample size and statistical significance are insufficient to guarantee that results are accurate or clinically meaningful and that there is a need to recognize the methodological challenges of generating reliable information about subgroups simply by increasing the diversity of patients in CER studies. Useful information about subgroups will likely require much larger or more informative studies to ensure that the subgroups can be analyzed with reasonable statistical power and further refinement of methods to more accurately adjust for baseline differences in non-experimental studies. The ability of CER to enable more personalized medical decision-making ultimately depends on adequate investments in larger studies and better methods.

V. Meeting the Infrastructure Needs for CER

New methods and standards for CER evidence must be complemented by an even larger investment in research infrastructure in order to produce CER more effectively and efficiently.

Infrastructure for more efficient CER trials

Recommendation 4: At least 10 percent of funds allocated to CER in the next 10 years should be directed by the Secretary of HHS for use in the development of infrastructure for enhancing the efficiency of CER trials. Funded programs should address the needs for:

- Ensuring that data standards created through the expansion of health information technology and deployed through electronic medical records are capable of supporting practice-based clinical research
- Development of informatics grids and other architecture to link practice-based research networks, creating a national network with sufficient scale for conducting priority CER trials
- Incentives for participation of investigators and patients in CER trials
- Standard contract language for CER trials that use network infrastructure
- Ethical guidance to institutional review boards that addresses human subjects protection issues commonly encountered in CER trials

32. Garber AM, Tunis SR. Does Comparative Effectiveness Research Threaten Personalized Medicine? *NEJM*. 2009; 360(19):1925-1927.
33. Greenfield S, Kravitz R, Duan N, Kaplan SH. Heterogeneity of treatment effects: implications for guidelines, payment, and quality assessment. *Am J Med*. 2007; 120(4): S3-9.
34. Collins R, Peto R, Beigent C, Sleight P. Aspirin, Heparin, and Fibrinolytic Therapy in Suspected Acute Myocardial Infarction. *NEJM*. 1997; 336(12):847-860.

Moving toward a national network of practice-based and medical center-based investigators with the tools to conduct CER using best practices – and doing so efficiently – will require significant investments targeted to specific barriers.

By virtue of its focus on the effectiveness of interventions in actual practice, rapid expansion of CER nationally will require the capacity to collect data efficiently from sites where care is delivered. Many observers believe that expanding the ability to conduct research in primary care and other practice environments will both improve the efficiency of research by adding to the numbers of investigators and potential human subjects, as well as pave the way for rapid application of results in the practice settings where research was conducted. In 1998, practice-based research networks (PBRNs) were noted by the IOM to be "the most promising infrastructure development that [the committee] could find to support better science in primary care."[35] In 2006, the NIH released an Inventory and Evaluation of Clinical Research Networks (IECRN) which counted nearly 250 research organizations nationwide and identified 29 among their "best practice" research networks. The IECRN recognized the stability of funding for PBRNs as a "pressing concern."[36]

One goal of the NIH Clinical and Translational Science Awards (CTSA) program is to improve the effectiveness and efficiency of clinical trials in medical research centers. A national consortium, the CTSAs now include 39 institutions in 23 states. By 2012 the program is expected to span approximately 60 CTSAs with an annual budget of $500 million. One area of focus for the program is to speed the initiation of clinical studies by improving processes while controlling costs and reducing the time taken to complete protocol approvals by ethics committees and contract negotiations. Many stakeholders have observed that ethics approvals and contract negotiations are key bottlenecks in the existing clinical research system.

While these initiatives represent promising directions, the new demand for large-scale and coordinated CER adds urgency to the need for these infrastructure improvements. Moving toward a national network of practice-based and medical center-based investigators with the tools to conduct CER using best practices – and doing so efficiently – will require significant investments targeted to specific barriers.

Infrastructure for learning from the delivery of health care

Recommendation 5: At least 10 percent of funds allocated to CER in the next 10 years should be directed to the Secretary of HHS for use in the development of infrastructure for learning from the delivery of health care. Funded programs should address the needs for:

- Distributed data networks for administrative and clinical databases – including Medicare and Medicaid data – and procedures for private-sec-

35. Donaldson MS, Yordy KD, Lohr KN, Vanselow NA (Committee on the Future of Primary Care, Institute of Medicine). Primary care: America's health in a new era. Washington DC: National Academy Press; 1996, p. 231.
36. Durako S. (Inventory and Evaluation of Clinical Research Networks, IECRN). Best practices study report. 2006 July; p. 2-65. Available from http://www.clinicalresearchnetworks.org/4f.asp#_Reports. Accessed June 1, 2009.

tor databases to be added to the network, procedures for investigator access to the network, and appropriate safeguards for ensuring the privacy and security of protected information
- Technical data standards and a common vocabulary to be used by all linked systems. The necessity of these standards to support CER should be a high-priority spending consideration for the funds allocated to expand health information technology
- Incentives for organizations with relevant data to adopt these standards and participate in research networks
- Ethical guidance to HHS, other data owners, study sponsors, and investigators that balances the need for evidence to inform decisions with the need to safeguard personal health information

For non-experimental studies, increasingly rich clinical and administrative data generated electronically through routine health care encounters can and will serve as a timely source of CER information. This data could be a valuable resource for evaluating treatments and outcomes across a broad range of patients, clinicians and practice settings and for identifying heterogeneity in treatment effects between defined subgroups of patients. The methodological challenges should not be underestimated, but the combination of data sources with the right level of clinical richness can allow valid inference of many important comparative effectiveness questions. These data sources may also be used to provide longitudinal data to supplement data collected in clinical registries and clinical trials. Data networks under development including the DeCIDE (Developing Evidence to Inform Decisions about Effectiveness) network supported by AHRQ, and the Sentinel network for monitoring post-market safety, supported by the FDA. Patient-level administrative data from Medicare and Medicaid, the two largest health insurance programs in America, are currently not readily available to researchers.

Distributed data networks have the advantage of allowing those organizations with data to keep it behind firewalls (thus avoiding pooling) while still using it for collaborative research. This is done using a set of informatics "pipes" that deliver standardized computer code that queries and analyzes each database the exact same way. The results from each database can then be combined for statistically robust estimates of comparative effectiveness. Linking together a variety of databases in such a network adds not only to the quantity of patients who can be studied but also to the quality of studies that can be performed if clinical and administrative data can be linked at the patient level. However, it is in the area of linkage that many persistent challenges and barriers remain. The informatics demands of data-sharing are perceived by most organizations as an additional burden, and thus far there has been only modest federal support for these efforts and no incentives provided to organizations to encourage participation. In addition to the technical challenges, proprietary concerns, as well as concerns about data privacy and security, are additional barriers to widespread use of these data sources.[37]

Distributed data networks have the advantage of allowing those organizations with data to keep it behind firewalls while still using it for collaborative research.

37. Institute of Medicine. Beyond the HIPAA Privacy Rule: Enhancing Privacy, Improving Health Through Research. Washington, DC: The National Academies Press. February 4, 2009.

Potentially valuable sources of linkable data include the large pools of claims from Medicare, Medicaid and private insurers. Although claims data typically lack detailed clinical information, they have been used extensively for health services research and have generated important information about drug safety and effectiveness.[38] Some examples of useful database combinations that would improve the validity of CER findings while preserving generalizability include 1) Medicare Part A, B, and D data with or without Medicaid and Minimum Data Set (MDS), which would allow the analysis of medical interventions and drugs in nursing home residents with detailed clinical information; and 2) Medicare Part A, B, and D data with the Medicare Current Beneficiary Survey (MCBS), which would allow studies in elderly outpatients. It will also be important to develop a pathway for incorporating private databases, which can provide important additional detail, such as inpatient hospital records. For example, commercial insurance claims data were recently linked with inpatient data to study the comparative frequency of major complications associated with an in-hospital intervention.[39]

Finally, related to the need for methodological standards discussed earlier is the need for quality assurance in such a distributed data network. Because non-experimental research can be easily biased by the use of inferior methods and data to control for confounding, there is a risk that the credibility of the data network—and of CER in general—may be undermined by improper use. Thus, appropriate credentialing systems should be developed to assure effective utilization of such a data network.

VI. Summary and Conclusion

One of the most important consequences of the recent funding for CER has been to finally focus serious attention on what CER is, how it is different from other clinical and health services research, and what will be required to ensure that it can be conducted successfully. This paper identifies a number of issues related to CER methods and infrastructure that will need to be addressed as the CER enterprise is expanded:

- Strategies and mechanisms that allow for the meaningful engagement of patients, consumers, clinicians, payers, and policy makers in key phases of CER must be developed and replicated.
- A clearly articulated framework will be necessary to ensure that high-priority CER questions are addressed with the methods that are most likely to provide meaningful, relevant evidence.
- Significant and sustained investment will be necessary to improve existing research methods for CER.
- Investment in the data collection infrastructure for experimental and non-experimental research is critical to building the capacity to conduct rigorous studies on a national scale.

While initial funding for CER should be devoted to addressing a broad range of high-priority clinical topics, a substantial portion of these initial funds should be directed to building the methodological framework and data collection infrastructure discussed in this paper.

38. Schneeweiss S, Solomon DH, Wang PS, Brookhart MA. Simultaneous assessment of short-term gastrointestinal benefits and cardiovascular risks of selective COX-2 inhibitors and non-selective NSAIDs: an instrumental variable analysis. Arthritis Rheum 2006;54:3390-8
39. Schneeweiss S, Seeger JD, Landon J, Walker AM. Aprotinin during coronary-artery bypass grafting and risk of death. N Engl J Med 2008;358:771-83

Acknowledgements:

This paper was substantially improved by the careful review and thoughtful feedback from a number of colleagues who reviewed an earlier draft of this paper, including Joe Selby, Rich Platt, Josh Ofman, Ellen Sigal, Rob Califf, Tim Taylor, and Sebastian Schneeweiss. I thank Mark McClellan for the clear vision involved in focusing on the critical topics addressed in the papers produced for this Brookings-Hamilton initiative on comparative effectiveness research. Josh Benner provided a substantial amount of insightful feedback on earlier drafts of this paper, and provided much useful material on the data infrastructure section. Finally, I very much appreciate the assistance of Danielle Whicher with writing, reviewing, and editing the manuscript.

From Better Evidence to Better Care: Using Comparative Effectiveness Research to Guide Practice and Policy

Steven Pearson, MD

Introduction

Comparative effectiveness research compares alternative approaches to prevent, diagnose, treat, and monitor clinical conditions. The goal of a national initiative in comparative effectiveness, however, is not simply to produce new research. Much more is expected. Comparative effectiveness is intended to support patient-centered care that will produce superior patient outcomes. It is also hoped that the uptake of important clinical innovations will become more rapid and uniform. At the same time, with health care costs a national priority, comparative effectiveness is expected to help reduce the rate of growth of costs and ensure that the highest value is obtained for every health care dollar. With this much expected of it, policy makers considering the future form and function of comparative effectiveness will require a clear strategy for how to make it "fit for purpose" – how this research can be produced, disseminated, and applied to help all stakeholders improve the quality and value of clinical practice.

Two examples can convey some of the different reasons a new national initiative on comparative effectiveness research is needed, and suggest the significant challenges that must be addressed in order for this research to achieve its goals:

Lag in adopting important innovations

> In 1993 the results of a large, well-conducted randomized controlled trial comparing two different treatments for a common condition, deep vein thrombosis, were published in the *New England Journal of Medicine*[1]. The results demonstrated that a short series of injections of a new drug, enoxaparin, which could be done in an outpatient setting or at home, was just as effective as the traditional treatment of intravenous heparin which required a 5-7 day hospital stay. In addition, patients treated with the injection required many fewer blood tests and reported higher satisfaction with their care.
>
> There being no authoritative body to evaluate the results of this study, synthesize it with other evidence, and present judgments on the comparative effectiveness of the two treatment options, enoxaparin took years to become widely used in practice. Results from studies on its use in the outpatient setting took four years to begin to appear in the published literature,[2,3] confirming equal or superior clinical outcomes, enhanced patient satisfaction, and substantial cost savings. But still, without a recognized entity to present an authoritative assessment, insurance coverage for outpatient treatment remained uneven, causing many patients in the mid-to-late 1990s to undergo unnecessary hospitalization. There was no coordinated system to manage rapid large-scale observational analyses of available clinical databases, and so smaller, uncoordinated studies confirming over and over again the superior clinical outcomes and cost-effectiveness of enoxaparin continued to be performed and published more than a decade after the original evidence had become available.

1. Hull R, Raskob G, Pineo G, Rosenbloom D, Evans W, Mallory T, Anquist K, Smith F, Hughes G, Green D, et al. A comparison of subcutaneous low-molecular-weight heparin with warfarin sodium for prophylaxis against deep-vein thrombosis after hip or knee implantation. N Engl J Med. 1993 Nov 4;329(19):1370-6.
2. Levine M, Jent M, Hirsh J, et al. A comparison of low-molecular-weight heparin administered primarily at home with unfractionated heparin administered in the hospital for proximal deep-vein thrombosis. N Engl J med. 1996;334:677-681.
3. Koopman MMW, Prandoni P, Piovella F, et al. Treatment of venous thrombosis with intravenous unfractionated heparin administered in the hospital as compared with subcutaneous low-molecular-weight heparin administered at home. N Engl J Med. 1996;334:682-687.

Failing to guide practice or policy

Beginning in the late 1990s, the use of a new form of external radiation therapy for prostate cancer, called IMRT, began to increase rapidly across the United States, driven largely by Medicare reimbursement rates that paid physicians more than four times as much to perform IMRT as other active treatment options, which included surgery and radioactive seed implantation. No randomized controlled trials or other forms of high quality studies were ever performed to evaluate IMRT against other options, and multiple published systematic reviews concluded that there was inadequate evidence to judge IMRT's comparative effectiveness. Nevertheless, its use continued to grow, in some parts of the country nearly displacing all other treatment options.[4]

In February 2008, the federal Agency for Healthcare Research and Quality (AHRQ) completed an extensive comparative effectiveness review on treatments for localized prostate cancer that, given evidence limitations, concluded that nothing could be inferred about possible differences in overall survival or cancer outcomes between surgery and all the various radiation treatments.[5] The review did not attempt to compare the side effects, long-term outcomes, or cost-effectiveness of external IMRT with radioactive seed implantation. When the review was completed, no system existed that would have linked the timing of its dissemination with updates to clinical society guidelines or policy statements. The review had no impact on Medicare or private insurer payment for IMRT. And, although the review called for further research, no major studies comparing IMRT to other radiation therapy options have been undertaken since the review appeared. By the time the AHRQ review summary materials were made available, patients, clinicians, and payers were considering whether to use and pay for several newer options being touted in the marketplace, including robotic prostatectomy, proton beam therapy, and "cyberknife" radiotherapy, none of which was discussed in the AHRQ materials. No update to the AHRQ review has been performed, and no information on when a review or update may be expected is publicly available.

These vignettes highlight the opportunities for comparative effectiveness research and some of the interlocking barriers that have limited its impact to date. Chief among these barriers are poor coordination of review and research efforts; impaired legitimacy of review organizations and processes; limited applicability of the framing and formats of research; inadequate development and deployment of medical policies that apply evidence in a flexible, targeted fashion; and misaligned payment incentive structures. The passage of the American Recovery and Reinvestment Act of 2009 (ARRA) offers an important opportunity for policy makers to consider how they can address these barriers moving forward. Moreover, the recent legislative proposals put forth by members of both houses of Congress to create a prominent new federal comparative effectiveness organization foretell that in the near future important choices will be made that will determine the role of comparative effectiveness as a component of health care reform and shape its efforts to help move the U.S. health care system toward higher quality, sustained innovation, and better value for every health care dollar. The purpose of this paper is to examine some of these key choices, and to propose a specific approach to the production, framing, and application of comparative effectiveness evidence to achieve all of the high goals set for it.

The Function of Comparative Effectiveness Evidence

Comparative effectiveness research combines two very different elements: 1) *synthesizing existing*

4. Saul, S. Profit and questions on prostate cancer therapy. New York Times, published December 1, 2006. Available at http://query.nytimes.com/gst/fullpage.html?sec=health&res=950CE6DA1E3EF932A35751C1A9609C8B63#. Accessed May 29, 2009.
5. AHRQ (Agency for Healthcare Research and Quality). Comparative Effectiveness of Therapies for Clinically Localized Prostate Cancer. February, 2008. Available at http://effectivehealthcare.ahrq.gov/healthInfo.cfm?infotype=rr&ProcessID=9%20&DocID=79. Accessed May 29, 2009.

evidence to inform decision-making; and 2) *generating new evidence* to address important evidence gaps. Synthesizing evidence can be accomplished through systematic review, which seeks to provide an objective survey of all existing published evidence; meta-analysis, which applies statistical approaches to merging evidence from different studies; and decision analysis, a modeling technique which makes use of existing evidence to simulate and compare the benefits, risks, and sometimes the costs of entire pathways of care. The generation of new evidence, on the other hand, requires the commissioning of prospective clinical trials or the conduct of new analyses of patient outcomes from data available in insurance claims systems, electronic medical records, or clinical registries.

When combined effectively, the elements of comparative effectiveness research represent a powerful tool for improving the evidence base and informing decisions by patients, clinicians, and policy makers. However, the uncertainty and political tensions surrounding the question of how comparative effectiveness information will be used have forced much of the public rhetoric into projecting only two options for the function of a comparative effectiveness program: one that focuses narrowly on providing information to patients and clinicians, or one that largely serves the interests of payers by making or recommending coverage decisions. But there is no need to assume that the function of comparative effectiveness evidence can only be designed in accord with one of these extremes. It should be possible to align the production, framing, and application of comparative effectiveness evidence to achieve a broader function: to give patients and clinicians more actionable information while also providing payers evidence framed specifically to support value-based coverage and payment policies (see Table 1). This third option, referred to as "providing guidance," would navigate a middle path between an overly passive "providing information" role and one too strictly directed toward "making decisions" for payers.

The function of a federal comparative effectiveness program should be to create and disseminate guidance to patients, clinicians, and payers in a format that will prove actionable for each.

A federal comparative effectiveness program should create and disseminate guidance to patients, clinicians, and payers in a format that will prove actionable for each. Providing guidance to payers means that evidence will be framed clearly, with evidentiary judgments made to support transparent consideration in coverage and payment policies. But the concept has a clear boundary: providing guidance does not mean making decisions. Explicit recommendations for coverage are not part of the scope of this approach to comparative effectiveness. Payers will remain responsible for assessing the evidence reviews emanating from a federal comparative effectiveness program, factoring in contextual considerations, and working with patients and clinicians to make publicly defensible coverage and payment decisions.

This construction of the function of comparative effectiveness evidence is consistent with the suggestions for "patient safeguards" included in the recent white paper by the Senate Finance Committee.[6] The Finance Committee proposes that a federal comparative effectiveness entity be prohibited from issuing medical practice recommendations or from making reimbursement or coverage decisions or recommendations. The option they describe for use of comparative effectiveness re-

6. Senate Finance Committee. Transforming the Health Care Delivery System: Proposals to Improve Patient Care and Reduce Health Care Costs. April 29, 2009. Available at http://finance.senate.gov/sitepages/leg/LEG%202009/042809%20Health%20Care%20Description%20of%20Policy%20Option.pdf. Accessed May 29, 2009.

In lieu of a dominant federal organization, comparative effectiveness research results have often emerged from a wide variety of federal, state, academic, and private sources without coordination and without any organized dissemination or implementation strategy.

sults by Medicare and other HHS agencies is that they be allowed to use the findings only through processes that are transparent, rely on all available evidence (not only comparative effectiveness results from a federal entity), consider the potential for effects on subpopulations of beneficiaries, and allow for public comment on any draft proposals that use the information. As the Finance Committee proposal notes, this approach would prohibit Medicare or others from creating a fast-track process to automatically link the research findings to coverage or reimbursement decisions in public programs.

In line with this view of a broad, balanced function for comparative effectiveness, this paper offers five recommendations and associated specific strategies to overcome the barriers to effective use of comparative effectiveness evidence as part of an overall strategy for maximizing the value of a national program in comparative effectiveness research.

Recommendation 1: Coordinate the production and dissemination of comparative effectiveness research results through a single high-profile organization at the federal level.

Specific Strategies:

- *Create an organizational structure at "arm's length" from government, preferably through a federally chartered corporation, that remains accountable to Congress but has stable funding and a governance structure insulated from overt political pressure*
- *Establish a clear process for stakeholder engagement across all functions of the organization while setting the governance structure and internal policies to minimize perceptions of bias*
- *Develop centrally a common methodology and format for comparative effectiveness reviews but commission their production from a national network of academic and private sector review organizations which can produce high-quality evidence reviews*
- *Develop reviews rapidly, keep them up to date, and ensure that they reflect the contributions of all stakeholders*

All of the other recommendations made in this paper will achieve their greatest influence if they are exercised or coordinated by a single high-profile, trusted national comparative effectiveness organization. Currently, the conduct of comparative effectiveness research is diffused across many sectors of the health care system. Until the passage of ARRA, AHRQ alone had been nominally charged with the conduct of comparative effectiveness research within the federal government, but it had received scant resources and operated within a restricted mandate. In lieu of a dominant federal organization, comparative effectiveness research results have often emerged from a wide variety of federal, state, academic, and private sources without coordination and without any organized dissemination or implementation strategy.

It can be argued that this relative diffusion of responsibility and activity has been a beneficial feature of comparative effectiveness research in the United States. According to this view, diversity provides a form of competitive market, constantly balancing products to the needs of consumers, and ensuring choice of methods and organizations without the risk of a single, dominant organization exercising too much control. Even varying methods, results, and conclusions across disparate comparative effectiveness evidence reviews can be seen as a useful means to gain a healthy, humble perspective on the nuances of comparative effectiveness methodology.

Diversity and competition have some salutary effects, but the broader impact of the lack of coordination has been to impair the ability of comparative effectiveness evidence to help patients, clinicians, and payers achieve their goals. Poor coordination has led to duplication of efforts and an inefficient use of limited assessment resources. Lack of coordination has reduced the opportunities to align comparative effectiveness review efforts with the prioritization and targeting of new comparative effectiveness studies to fill important evidence gaps. But perhaps most importantly, having multiple comparative effectiveness sources and approaches has hindered the development of consistent methods and review processes that can gain the trust of the public. It is important to distinguish here the two types of methods involved in comparative effectiveness review: scientific or statistical methods, and the methods used to frame and format results. It will be the role of a federal organization to foster a robust and diverse academic community of researchers engaged in the continuous development and evaluation of scientific and statistical methods. Ultimately, however, the organization should help bring researchers together in order to seek consensus and set consistent, high-quality standards for the methods used in comparative effectiveness reviews. The federal organization should exercise an even stronger role in developing a consistent process for the conduct of comparative effectiveness reviews, including rules for stakeholder engagement and transparency of internal procedures.

The framing and formatting of the results of comparative effectiveness reviews are a critical set of responsibilities that should be standardized by a single federal organization. As discussed in later recommendations, framing and formatting will be central to the ability of comparative effectiveness to be communicated to all stakeholders in a way that is viewed as reliable and actionable. The current multiplicity of frameworks and formats used to assess evidence and to communicate judgments on comparative effectiveness has produced confusion as a natural outcome. The lack of standardization has also offered an easy target for critics unhappy with the outcome of a review on their favored intervention.

The framing and formatting of the results of comparative effectiveness reviews are a critical set of responsibilities that should be standardized by a single federal organization.

A federal comparative effectiveness organization should therefore exercise a leading role in supporting methods development, establishing consensus for key methods and procedures within comparative effectiveness reviews, and developing consistent frameworks and formats for communicating results. But, just as AHRQ has done with its existing comparative effectiveness portfolio, much of the "work" of methods development, and all of the production of comparative effectiveness reviews, should be commissioned from a diverse network of academic and private research and review or-

ganizations. Some international programs, such as the National Institute for Health and Clinical Excellence (NICE) in England [see box A, pg. 77], have developed mature networks of independent academic groups whose experience allows them to meet tight timelines while maintaining a consistent rigor and transparency in their products. Similar efforts as part of a federal comparative effectiveness program in the U.S. would enable patients and clinicians to get early guidance on emerging interventions, while also supporting payers in their goals of creating initial coverage and payment policies that reflect reliable assessments of the comparative effectiveness of new interventions. A robust network of collaborating research and review organizations would also make the process of updating comparative effectiveness reviews more efficient and reliable.

The benefits of having a governing Board that to the public will appear independent and objective could prove decisive in judgments about the overall legitimacy of the entire comparative effectiveness program.

To exercise the broad function of coordinating this set of activities, to gain the trust of clinicians and the public, and to withstand the inevitable disappointments that are generated by judgments about evidence, a federal comparative effectiveness organization will need a structure and governance that can prove both responsive and durable. The Medicare Payment Advisory Commission (MedPAC) and other commentators have described the full spectrum of options for placement, including a federally funded research and development center (FFRDC), an independent federal agency within the executive branch, an independent federal agency within the legislative branch, and a congressionally chartered nonprofit organization.[7] Each option has potential advantages and disadvantages, but in many ways the chief issue is the relative balance between governmental accountability and stakeholder influence. For those who are most concerned about the potential influence and bias of special interests, having a federal comparative effectiveness organization inside government makes the most sense. In contrast, an independent structure outside of government is favored by those who worry most about direct political interference or who are concerned that placement within government would subvert broader goals to the needs of cost control of public health programs.

The ultimate decision on placement should be made with great attention to the perspectives of all stakeholders, since the future success of the organization will be rooted in the base of support and engagement it can build upon from its inception. The goal should be for the organization to be accountable to Congress yet structured and funded so that it has greater political insulation than existing structures inside government. I propose that this goal would best be met by a congressionally chartered nonprofit corporation. This is the structure proposed in legislation submitted by Senators Baucus and Conrad (S. 3408) and by Congressman Schrader (H.R. 2502). The sources of funding would be more stable than annual appropriations, coming instead from a mix of general revenues, contributions from the Medicare trust funds, and an assessment on private insurance in proportion to their share of total national health expenditures.

7. MedPAC. Report to the Congress: Reforming the Delivery System. Chapter 5: Producing comparative effectiveness information. June 2008. Available at http://www.medpac.gov/documents/Jun08_EntireReport.pdf. Accessed May 30, 2009.

This placement of the federal comparative effectiveness organization outside government requires close attention to the structure of its governance and the role of stakeholders. There are two general approaches to governance that could balance the goal of stakeholder involvement with the need to minimize the potential for bias due to conflicts of interest. The first approach is to have the stakeholders – patients, clinicians, manufacturers, payers, researchers – all serve together on the governing Board. Biases and conflicts of interest would largely be managed through disclosure and a conscious attempt to have potential conflicts neutralize each other by inclusion of all the different stakeholders. This approach has the advantage of bringing different perspectives into direct contact with each other at the highest level of the organization, possibly facilitating a learning process and a sharing of responsibility for the comparative effectiveness agenda that would lead to robust leadership and creative policies. On the other hand, it is possible that entrenched interests within a governing Board would bring a comparative effectiveness program into gridlock or make all decisions by the Board vulnerable to external critiques of the conflicts of interest of much of the leadership.

Another possible approach is to have a governing Board that is distinguished by its freedom from conflicts of interest in the results of comparative effectiveness. Such a Board could be composed of figures from science outside of medicine, former academic or political figures, patient and public representatives without specific advocacy positions, and methodological experts without active potential conflicts of interest. In this model, the involvement of stakeholders could be organized through an external Advisory Committee and through membership of stakeholders on the key operational evidence review and methods committees that will appraise evidence reviews, make methodological recommendations, and direct a large part of the program's output. Whether the risk of gridlock would be just as serious, or even worse, with stakeholders engaged primarily at this level, cannot be known. But the benefits of having a governing Board that to the public will appear independent and objective could prove decisive in judgments about the overall legitimacy of the entire comparative effectiveness program.

A practical approach would be to include stakeholders as a minority of a governing Board, and this is the approach taken in the most recent Senate and House bills. Governmental agencies, health plans, manufacturers, physician groups, patient advocacy groups are all represented but no one group or block dominates, and explicit rules for disclosure and recusal are given. The Comptroller General is responsible for making these appointments. A Chairperson and Vice-Chairperson are to be selected, and should be members with no representational or financial conflicts of interest. Importantly, the procedures for the operation of the Board should be based on simple majority vote. The Board will ultimately be responsible for approving priorities in research and fund allocations, methodological standards recommendations from a methodology committee, and the formatting and communication strategies for all products. If stakeholders are to be included on the Board, a simple majority vote procedure will be necessary to maintain the ability of the Board to make difficult decisions that may not suit the interests of a minority of members.

Recommendation 2: Convey judgments of the strength of evidence and of comparative effectiveness through explicit rating systems that meet the needs of patients, clinicians, and payers.

Specific Strategies:

- *In both evidence synthesis and evidence generation, seek to uncover clinical differences among patients that help predict which patients will benefit most from alternative interventions*
- *Use decision analytic modeling techniques on a regular basis to complement the findings of systematic review and meta-analysis of the published medical literature*

Better understanding of which types of patients are more likely to benefit from different care options is the most central of all the evidentiary goals of comparative effectiveness research, and formatting the information to capture the diversity of patient responses to different options will maximize the applicability of comparative effectiveness for all stakeholders.

- *Balance the framing of comparative effectiveness evidence reviews so that they neither provide too little summary judgment nor so much that they serve as a proxy for decisions*
- *Develop clear, reliable, and valid rating systems through which judgments of the strength of evidence and comparative clinical and cost-effectiveness can be communicated to all stakeholders*

Evidence produced by new studies or through the synthesis of existing sources will always reflect the outcomes of populations of patients. Clinicians and patients, however, may determine that population-based findings have little relevance when making test or treatment choices for individual patients, patients who often differ in many ways from the "average." For all decision-makers, therefore, comparative effectiveness evidence will best serve their needs if evidence syntheses are framed to highlight clinical or socio-demographic characteristics that are related to different relative risks and benefits of treatment. Better understanding of which types of patients are more likely to benefit from different care options is the most central of all the evidentiary goals of comparative effectiveness research, and formatting the information to capture the diversity of patient responses to different options will maximize the applicability of comparative effectiveness for all stakeholders.

One way to highlight the importance of patient subpopulations, even when ample published evidence is lacking, is decision analytic modeling. Decision analysis involves the creation of a decision tree to represent all the possible components of a pathway of care, including the course of treatments with all their possible benefits and risks. Probabilities based on best existing evidence, or expert opinion when evidence is lacking, are assigned to each fork in the decision tree, and values related to the effect on quality of life are assigned to each outcome. Decision analysis is thus especially useful when patients' perspectives and values are particularly important in judging the outcomes of care. Ultimately, the entire decision tree can be analyzed to compare the cumulative outcomes between two or more different pathways of care. Decision analysis is able to explicitly model the different relative risks and benefits of interventions for all known subpopulations of patients, and can therefore help highlight the importance of key clinical or socio-demographic characteristics in understanding the comparative effectiveness of alternative interventions for diverse patient groups.

Decision analysis is also useful when there is no direct head-to-head comparative evidence between two alternative treatments. In such situations, adherence to traditional evidentiary hierarchies might lead a systematic review to conclude that there is "inconclusive" or "inadequate" evidence to make a comparative judgment between two alternative therapies; but decision analytic modeling can take a broader range of available evidence, supplement

it where necessary with consensus or expert advice, and present a model through which the effectiveness of two treatments can be more usefully compared. Decision analysis has been used by the Medicare Coverage and Evidence Development Advisory Committee (MEDCAC) and the United States Preventive Services Task Force (USPSTF) in evaluating screening and diagnostic interventions, but its use should be expanded by a federal comparative effectiveness program as a component of the evaluation of therapeutic interventions as well. Modeling has its own set of acknowledged vulnerabilities, and usually adds complexity and time to any comparative effectiveness review, but inclusion of modeling as a standard approach within comparative effectiveness research would likely improve the overall usefulness of the results for patients, clinicians, and payers.

Another key step in making comparative effectiveness reviews more useful to all stakeholders is to provide enough judgment about the strength of evidence and the overall findings of the reviews to be actionable. A balance must be struck between trying too hard to appear "impartial" by providing only an undigested compendium of study findings, and going to the other extreme of making decisions that negate the role of clinicians and patients in sorting out how the evidence should be applied in the context of individual clinical decisions. In order for evidence syntheses to avoid these twin pitfalls, the best option would be to frame comparative effectiveness reviews to include important detail and overall judgments translated into an easy-to-understand set of grades or other rating system. The proposed procedure for the generation of ratings would follow the general pattern established by NICE and also used by Medicare for the MEDCAC. First, the academic work of comparative effectiveness *assessment* would be performed by one of the research and review groups commissioned by the federal organization. This assessment would follow an established process for rating the strength of evidence on the key questions related to comparative net health benefits between two or more alternative interventions. The assessment report would then flow back to the federal organization where an independent panel would be convened to *appraise* the evidence. It would be at this appraisal phase that revision or further work by the academic assessment group might be requested, but ultimately the appraisal committee would recommend overall grades or ratings on comparative effectiveness, including ratings for all key patient subpopulations. Patients will find these reporting formats easier to understand and incorporate in their discussions with clinicians. Clinicians will also find this framing of comparative effectiveness more actionable.

The best option would be to frame comparative effectiveness reviews to include important detail and overall judgments translated into an easy-to-understand set of grades or other rating system.

In addition, this kind of formatting is likely to make it easier for professional guidelines groups to incorporate comparative effectiveness into guidelines and appropriateness criteria, a critical feature of the effective dissemination of guidance to clinicians.

As with evidence on clinical effectiveness, evidence on cost-effectiveness will prove more understandable and helpful for decision-making if it is framed through some kind of grading or rating system as well. This is not to suggest that an intervention or pathway of care would be rated as "cost-effective" or "not cost-effective" according to some single cut-point, although it might be that among two or more different approaches to achieve a particular goal – e.g. identify a person at risk of early heart

Whether a rating system would use letters, or grades, or just verbal categories, the key is to move the complexity of comparative effectiveness review into a format that can be understood and can prove tangible enough to provide real guidance to stakeholders.

disease – one approach might be identified as more cost-effective than others. Costs and cost-effectiveness would be presented just like evidence on clinical effectiveness, as a means of providing guidance for decision-makers, not as a proxy for making the decision for them.

There are helpful examples of existing comparative effectiveness products whose formatting fits the model proposed above. For example, patient and clinician versions of comparative effectiveness reports created by AHRQ use specific templates with a three-dot grading system to portray judgments on the strength of evidence underlying the answers provided to key questions about the comparative effectiveness of alternative care options. Another notable example of formatting reviews to provide guidance to multiple audiences is the recommendation program of the USPSTF. The USPSTF, first convened by the U.S. Public Health Service in 1984, and since 1998 sponsored by AHRQ, is the leading independent panel of private-sector experts in prevention and primary care. Its members are independent primary care clinicians with methodological expertise in epidemiology, decision analysis, and health services research. The USPSTF conducts rigorous, impartial assessments of the scientific evidence for the effectiveness of a broad range of clinical preventive services, including screening, counseling, and preventive medications. Its recommendations are considered the gold standard for clinical preventive services.

The USPSTF makes recommendations not for coverage and payment but for whether specific preventive services should be integrated into routine clinical care. The methodological foundation of USPSTF recommendations are transparent judgments, communicated by a rating scheme, of the level of certainty provided by the evidence and a separate judgment indicating the magnitude of the net health benefit conferred by a preventive service. Ratings of the certainty and magnitude of net health benefit are ultimately combined to create an overall recommendation grade of A, B, C, or D; an "I" grade is assigned when evidence is judged to be insufficient to make any recommendation. Each grade is linked to standard recommendation language: an "A" grade, for example, indicates that the USPSTF "strongly recommends that clinicians routinely provide the service to eligible patients"; a "B" grade is a weaker but still positive recommendation, and so on.

The recommendations of the USPSTF are widely disseminated to professional audiences in relevant journals, such as *Annals of Internal Medicine*; on the AHRQ Web site (http://www.preventiveservices.ahrq.gov); in print through the annual *Guide to Clinical Preventive Services*; and in a Web-based Electronic Preventive Services Selector (http://epss.ahrq.gov/PDA/index.jsp), which is downloadable into personal digital assistant devices. Wide and innovative dissemination efforts, however, are not the primary reason why USPSTF recommendations have proven extremely influential with clinicians and payers. Their influence is largely due to the strong legitimacy that comes from an independent and rigorous evidence review process, and, importantly, from their adoption of this rating approach that places evidentiary judgments into a clearly defined scheme to facilitate comprehension

and use by all decision-makers. The USPSTF is not making or mandating coverage recommendations, but providing guidance that allows for rapid consideration by individual patients and clinicians as well as by outside decision-making bodies who are able to use this evidence in creating transparent, evidence-based guidelines and medical policies.

Several other comparative effectiveness review groups, including the Blue Cross Blue Shield Association Technology Evaluation Center (TEC), and the Institute for Clinical and Economic Review (ICER), have developed publicly available methods for rating or grading evidence and for providing judgments on net harms and benefits of alternative care options. The TEC convenes a Medical Advisory Panel (MAP) to appraise internally produced assessments along five evidentiary criteria, each of which is judged by the MAP to be "met" or "not met" for the particular technology under review (see Box B). In rendering these judgments the TEC is not making coverage decisions for Blue Cross or other health plans; the five criteria serve as a benchmark to help the TEC reviews provide decision-makers with a consistent framework across multiple types of interventions to guide further deliberation on coverage at the individual plan level.

ICER has developed a two-part rating system for separate judgments of comparative clinical effectiveness and comparative value, with the potential for integrating the ratings if desired by stakeholders (see Box C). ICER reviews with this rating system have been used to guide coverage decisions by private health plans and by one public decision-making body, the Washington state Health Care Authority, which is legislatively instructed to consider safety, clinical effectiveness, and cost-effectiveness, in making coverage decisions for the state's Medicaid and other public departments.[8]

A federal comparative effectiveness organization will need to examine the merits of these different rating systems and adopt one in close consultation with all stakeholders. Whether a rating system would use letters, or grades, or just verbal categories, the key is to move the complexity of comparative effectiveness review into a format that can be understood and can prove tangible enough to provide real guidance to stakeholders. It may be difficult to arrive at a single format that is suitable for patients, clinicians, and payers, but the legitimacy and the ultimate impact of comparative effectiveness will be greatly enhanced if a single basic approach can be devised and tailored slightly to meet individual stakeholders' needs. Rapid attention to this framing and formatting issue will be required at the onset of the activity of a federal program, for the earliest reviews that are produced will set important precedents for the communication of results. Further evolution of any rating system is likely as a federal program learns from early experience, but a rating system is likely to become nearly synonymous with the work of the entire federal program, and so a consistent, transparent, and well-known approach should become an early cornerstone of a federal comparative effectiveness program.

There is no more contentious element of comparative effectiveness policy than the question of what to do with cost-effectiveness.

8. Song, KM. Medical procedures rated by state's expert comparison shoppers. Seattle Times. July 7, 2008. Available at: http://seattletimes.nwsource.com/html/localnews/2008036631_evidence07m.html. Accessed May 31, 2009.

Recommendation 3: Make cost-effectiveness a core component of the scope of a federal comparative effectiveness program.

Specific Strategies:

- *Coordinate analyses of comparative cost-effectiveness in parallel with evaluations of comparative clinical effectiveness*
- *Base cost-effectiveness analyses on transparent decision analytic models that meet high standards for objectivity and technical quality*
- *Present comparative cost-effectiveness results in a fashion that allows all stakeholders to understand the components of cost and the perspective taken in the analysis*
- *Frame evidence on cost-effectiveness as one of many tools to highlight opportunities for greater value; it should not be promulgated as the single or dominant method of guiding clinical decisions or payer policies*
- *Recognize the different cycles of development for various types of medical interventions, and deploy cost-effectiveness analyses and all evidence synthesis efforts in a manner cognizant of the time needed for the maturation of younger technologies that show substantial promise for treating significant unmet needs*

I propose that the long-term interests of patients in the United States for high-quality, affordable health care makes the answer to this question clear; and that the challenge of incorporating cost-effectiveness should be met at the outset of the operation of a federal comparative effectiveness program.

There is no more contentious element of comparative effectiveness policy than the question of what to do with cost-effectiveness. With quite reasonable apprehension, some stakeholders worry that cost-effectiveness could be used as a single criterion by which to make recommendations for the insurance coverage of medical services. This is, after all, the general approach taken by NICE in England. There are also concerns that cost-effectiveness methodology is rooted in an overly utilitarian ethic, that it poorly captures the value of interventions for severely disabled patients, and that it can catch important innovations at an early, more expensive stage, and snuff them out before they have a chance to prove their worth.

However, not one of these concerns is new; not one is less than well-studied and well-known to methodologists and policy figures in the cost-effectiveness field. And, ultimately, the question boils down to whether it is better to have a federal comparative effectiveness research program ignore costs, or whether the program should include costs within its scope, consult with all stakeholders on the best way to frame and format cost information, help develop its methodology further, and make sure that cost is considered openly and always in the context of clinical effectiveness. I propose that the long-term interests of patients in the United States for high-quality, affordable health care makes the answer to this question clear; and that the challenge of incorporating cost-effectiveness should be met at the outset of the operation of a federal comparative effectiveness program.

How are cost-effectiveness analyses done? They are built upon the same decision tree model that underpins decision analysis, with costs for all doctor visits, hospitalizations, tests, and treatments included along with clinical outcomes. If desired, costs and benefits related to time lost from work and

the time and effort of caregivers can be included as well. Cost-effectiveness analysis then evaluates the health benefits and economic costs of alternative interventions when viewed as entire pathways of care. As with decision analysis, subpopulations can be easily considered separately. The result of a cost-effectiveness analysis can demonstrate that two alternative pathways are equally effective, in which case the least costly is more cost-effective. The result can also demonstrate that one intervention and its pathway of care are more effective, at least for some patient populations, and if the intervention and its pathway are more effective they can either be less costly overall (a great bargain), or the incremental clinical benefits may come at a higher cost. This ratio of the incremental cost divided by the incremental clinical benefit is the cost-effectiveness ratio, a ratio that can extend from small numbers, such as $100 per additional year of life gained, to very high numbers, such as $1 million for an additional year of life. The decision about whether to use a particular ratio as a "cost-effective" threshold is not inherent in the use of cost-effectiveness. In fact, no formal threshold need be established at all to use cost-effectiveness information to provide guidance to patients, clinicians, and payers.

At its most basic, cost-effectiveness analysis can provide information to help doctors consider which treatment pathways are most effective for different kinds of patients, and can help both patients and doctors select a cheaper option if all pathways are judged to be equally effective. For example, ICER performed a comparative effectiveness review of the use of coronary CT angiography (CCTA) to evaluate low-risk patients with acute chest pain in the emergency department.[9] The ICER review concluded that the clinical effectiveness of using CCTA to rapidly screen patients was comparable to that of standard screening protocols that require a minimum of eight hours in the emergency department. The cost-effectiveness analysis done as part of this comparative effectiveness review suggested that although CCTA was a more expensive test, its use would reduce overall emergency department charges and subsequent outpatient testing costs, results that suggested that CCTA would be more cost-effective overall than traditional screening approaches. CCTA received a rating of "high value" in the ICER review. Guided by this analysis, which would not have been possible without the inclusion of cost-effectiveness information, the Washington state Health Care Authority granted coverage to CCTA for the evaluation of low-risk patients in the emergency department.

There are additional reasons to favor the inclusion of cost-effectiveness in the scope of a federal comparative effectiveness program. Some professional societies have indicated that costs should be considered by clinicians developing appropriateness criteria or guideline recommendations.[10,11] But current comparative effectiveness reviews usually avoid costs as an explicit component, leaving guideline groups and payers with no framework within which costs can be explicitly considered. Looking at costs later, after an assessment of clinical effectiveness, tends to focus coverage decisions on a drug's price rather than on its overall value to patients and the health care system. Cost-effectiveness evidence produced as part of comparative effectiveness would help ensure that when payers consider cost issues it is done in a transparent, explicit manner, enhancing the legitimacy of coverage and payment policies. In addition, the inclusion of cost-effectiveness as a core component of comparative effectiveness would serve as a stimulus for medical innovators and manufacturers to focus their efforts on developing new products that can demonstrate evidence of significant improvements in patient outcomes at

9. Institute for Clinical and Economic Review (ICER). Final Appraisal Document: Coronary Computed Tomographic Angiography for Detection of Coronary Artery Disease. January, 2009. Available at: http://icer-review.org/index.php/ccta.html. Accessed June 1, 2009.
10. Patel MR, Spertus JA, Brindis RG, et al. ACCF proposed method for evaluating the appropriateness of cardiovascular imaging. J Am Coll Cardiol 2005;46:1606 –13.
11. American College of Physicians. Improved Availability of Comparative Effectiveness Information: An Essential Feature for a High-Quality and Efficient United States health Care System. Philadelphia: American College of Physicians; 2008: Position Paper. (Available from American College of Physicians, 190 N.

prices that will prove sustainable for the U.S. health care system in the long term.

There are two basic strategies for incorporating cost-effectiveness information that would make the most sense for a federal comparative effectiveness program. The first of these strategies is similar to a policy within Medicare called "least costly alternative." The overarching idea is to identify clinical alternatives for which the evidence demonstrates their effectiveness is "comparable," and only then apply cost-effectiveness to evaluate the alternative pathways of care. Ultimately, whichever pathway of care is less costly overall would be identified as the higher value option, providing guidance to allow patients, clinicians, and payers to favor comparable interventions with lower "total pathway" costs. It is important to note that, as was the case with the ICER review of CCTA, total pathway costs may be lower even when the initial acquisition cost of the primary intervention is higher. Thus it is not unexpected for some expensive new interventions to be shown to provide high value compared to existing alternatives.

This strategy for using cost-effectiveness to provide guidance has an intuitive appeal and is easily explained to clinicians and the public. Why shouldn't patients and clinicians, as well as payers, want to choose a clinical alternative that is just as good as any other but cheaper? Challenges would still remain, however. From a policy perspective, this approach puts great pressure on the judgment of whether two clinical alternatives should be labeled "comparable." If the acknowledgment of small, incremental advantages eliminates any possible consideration of cost-effectiveness, payers would have a significant interest in having a federal comparative effectiveness program adopt a relatively liberal conception of "comparable," whereas manufacturers and others might advocate strongly for minor differences to be all that is needed to make it impossible to call two alternatives comparable. In effect, this approach prevents the application of cost-effectiveness as a tool to do what it often does best: identify incremental clinical advantages and estimate the cost payers and patients must pay for that incremental advantage.

Moreover, ratings do not in themselves convey recommendations for clinical practice or coverage, but they can support the consideration of cost-effectiveness by clinicians in the design of clinical guidelines, and they can provide explicit, rigorous information to support more flexible and patient-centered medical policies seeking to encourage the use of high value options.

Therefore, a second possible strategy for using cost-effectiveness to provide guidance is to evaluate pathways of care and provide cost-effectiveness ratings not only when two alternatives are considered comparable but when one is marginally better. Ratings of cost-effectiveness would be a central component of this approach, since raw incremental cost-effectiveness ratios – in which the metric used is often the cost per additional life year gained or quality-adjusted life year gained – are unlikely to provide suitable guidance for patients, clinicians, and payers without additional context.

One of the areas of methods development that a federal comparative effectiveness program would

need to explore is how a rating system for cost-effectiveness might be developed to achieve the balanced goal of providing effective guidance to all stakeholders without stepping over the boundary of making specific clinical or policy recommendations. As mentioned earlier, ICER has developed a rating system for both clinical and cost-effectiveness information as part of its comparative effectiveness reviews. ICER rates the use of interventions for particular patient populations as having "high," "reasonable or comparable," or "low" comparative value. As is true for reviews of comparative clinical effectiveness, a rating system for cost-effectiveness conveys important judgments about evidence and presents information in a clear, transparent format. Moreover, ratings do not in themselves convey recommendations for clinical practice or coverage, but they can support the consideration of cost-effectiveness by clinicians in the design of clinical guidelines, and they can provide explicit, rigorous information to support more flexible and patient-centered medical policies seeking to encourage the use of high value options.

Despite the potential advantages of including cost-effectiveness in the scope of a federal program, this prospect remains highly controversial. Even some stakeholders who believe in the importance of cost-effectiveness feel that it would taint any evaluation of clinical effectiveness, undermining trust among patients and clinicians who might fear that the initiative is dominated by a goal of cutting costs. Alternatives to formal incorporation of cost-effectiveness into a federal program include funding the analyses separately within other branches of HHS or leaving cost-effectiveness entirely outside the scope by providing no funding and relying on payers to commission their own analyses when desired. While it may be possible to construct some kind of arm's length relationship between a federal comparative effectiveness program and an allied effort to conduct cost-effectiveness analyses, the logistics of bringing the work into alignment would likely create significant inefficiencies and inconsistencies in the research, and the important potential for the decision analyses underlying cost-effectiveness to inform judgments of clinical effectiveness would be lost. In addition, relegating cost-effectiveness to the payer community would perpetuate the existing problems with legitimacy and coordination of the research, and minimize the chances that cost-effectiveness information would ever be viewed as trustworthy by clinicians or patients. Overall, therefore, despite the challenges of aligning clinical effectiveness and cost-effectiveness research without creating the impression that comparative effectiveness is "just about the money," I propose that the best way to meet the goals for a federal comparative effectiveness program is to commission the two through parallel processes, and to format and disseminate the results as transparent ratings that make clear the methodological limitations of the findings, and the proper, circumscribed, role that cost-effectiveness information should have in resource allocation and medical policies.

> The production and wide dissemination of high-quality evidence cannot be the true goal of a federal comparative effectiveness program. Getting that evidence applied in practice and policy, and seeing it improve outcomes and value, has to be the measure of success.

Recommendation 4: Integrate comparative effectiveness reviews into clinical practice by giving clinician organizations a leadership role every step of the way.

Specific Strategies:

- *In collaboration with professional clinical societies, formal procedures should be established through which guidance can be rapidly integrated by the societies into professional guidelines, appropriateness criteria, and other forms of clinical guidance as a principal method of influencing clinician understanding of the link between comparative effectiveness and best practices*
- *Collaborate with patient and clinician societies in co-branding formats of evidence reviews that can be aligned with the development of health IT infrastructure to reach clinicians and patients at the point of care*

The production and wide dissemination of high-quality evidence cannot be the true goal of a federal comparative effectiveness program. Getting that evidence applied in practice and policy, and seeing it improve outcomes and value, has to be the measure of success. Unfortunately, there are numerous barriers to the uptake of comparative effectiveness in clinical practice. The limitations from the clinician's standpoint begin with questions about the provenance of the information – does it reflect the considered judgments of expert clinicians in the field? Or the mixed motives of either payers or manufacturers? Clinicians are bombarded with "evidence," and it will be very difficult to craft a new source of evidence that can pierce the protective wall many clinicians erect to protect themselves from information overload.

There are also significant difficulties in making comparative effectiveness information available when needed at the point of care. Bulky, paper manuals are likely to sit on a shelf and go unused. And too often the guidance provided to clinicians lacks sufficient flexibility and relevance to clinical practice. The format for clinicians can suffer from being either too simple or too complex. Guidelines, for example, can be framed as simplistic algorithms that may disregard many important individual patient characteristics. Clinicians have often complained about such "cookbook" approaches that fail to allow for critical nuances in individual patient situations. Conversely, technically complex and diffusely framed summaries of existing information shorn of any attempt to rate or judge key findings will be too difficult for individual, busy clinicians to sort through and apply in a busy daily practice.

How best, then, to address these barriers and get comparative effectiveness information into clinical practice? The primary proposal underlying this recommendation is to drive the process by letting leading clinician organizations take the reins. The relationship that NICE has with the Royal Colleges in England, in which all clinical guidelines are produced and promulgated in direct collaboration with the professional organizations, provides an example of how a clinician-driven approach to dissemination can increase both the legitimacy and applicability of comparative effectiveness information.

In practice, this approach would entail close collaboration at all stages of the assessment, appraisal, and dissemination of comparative effectiveness reviews. At the initiation of a review, professional organizations would be solicited for nominees of clinical and methodological experts to serve on the appraisal committee along with patient representatives, other stakeholders, and unaffiliated members. These clinical experts would inform the initial scope of the review by providing input to the academic review organization commissioned by the federal program to perform the assessment. At all stages of the assessment, clinical experts would provide feedback and advice. They would also have an important role in the appraisal committee deliberations that would ultimately approve the assessment and provide ratings of comparative clinical effectiveness and cost-effectiveness. When the review is being prepared for general dissemination, there would be a key step in which the clinical experts would work with their own organizations to incorporate the findings into

updated clinical guidelines, appropriateness criteria, or other evidence-based tools for clinicians. The goal would be to have the comparative effectiveness review emerge simultaneously from the federal program with authoritative guidance or commentary from the relevant clinical specialties.

It can be argued that a process for guideline development under direct control of the federal program itself would be better in that it would safeguard against parochial interests and potential conflicts among different medical specialties. But physicians' trust in professional society recommendations is very high, and as long as the collaborative relationship is maintained throughout the review process, and the results can be formatted to facilitate rapid consideration and integration by professional societies into their own work, clinicians will be more likely to view the final product as trustworthy, authoritative, and actionable.

Integration into clinical practice will also be augmented if consideration is given to aligning the dissemination of the linked evidence reviews-clinical guidelines with developments in the health IT infrastructure. The goal should be to consider formats of the reviews-guidelines that can be easily woven into the daily practice flow of clinicians as they adopt health IT systems. It will be particularly useful if these formats support a shared decision-making process between patients and their clinicians. One good example of comparative effectiveness information formatted in this way is the approach taken by the Foundation for Informed Decision Making in its clinician and patient materials for localized prostate cancer, low back pain, and other common conditions (see www.fimdm.org). These materials, available in hard copy, DVD, and web-based versions, not only convey information but also frame a specific approach for patients and clinicians to engage with each other as they consider the information to reach an efficient, informed decision.

A federal program in comparative effectiveness should seek the same broad- based approach to supporting shared decision-making through the materials it disseminates in conjunction with clinician specialty organizations.

The emergence of a trusted national source of comparative effectiveness information will make it possible to speed the adoption of innovative approaches that can enhance care for individuals while also helping to restrain the growth in health care costs.

Recommendation 5: Use comparative effectiveness results to support innovative coverage and payment policies that encourage the generation of better evidence to guide future care and align patient and clinician incentives to use costly interventions prudently.

Specific Strategies:

- *Public and private payers should be empowered to use improved evidence from comparative effectiveness to support:*
- *Value-based benefit designs to decrease patient out of pocket expenses for higher value care options*
- *Coverage with evidence development policies to provide access to promising technologies with requirements to gather further evidence on clinical and cost-effectiveness*
- *Payment strategies that reward innovation demonstrating significant patient benefits instead of following a model of "reimbursement" divorced from evi-*

dence of comparative clinical effectiveness

- *Medical management programs that are finely tuned to guide the use of interventions with high costs and/or high risks so that patients receive the right care at the right time from a clinician with the right training*
- *Quality performance measures based on the findings of comparative effectiveness reviews*

The way that purchasers and payers apply comparative effectiveness results to benefit design, coverage, and payment policies will have a profound influence on the impact of comparative effectiveness – and on how a federal program is ultimately judged by patients and clinicians. The most traditional application of evidence by payers to medical policy has been the dichotomous "cover/no cover" coverage decision. When this is the dominant coverage policy tool at payers' disposal, comparative effectiveness either becomes a relatively blunt tool to deny coverage to all patients, or, more often, comparative effectiveness information has little impact at all as coverage is granted even in the absence of good evidence, and payment for new interventions is determined by traditional formulas divorced from consideration of evidence on comparative clinical effectiveness.

But the emergence of a trusted national source of comparative effectiveness information will make it possible to speed the adoption of innovative approaches that can enhance care for individuals while also helping to restrain the growth in health care costs. As the Congressional Budget Office stated in its 2007 report on comparative effectiveness:[12]

> Although insurers could choose not to cover drugs, devices, or procedures that were found to be less effective or less cost-effective, they would have a number of additional options as well. They could adjust payments to doctors and hospitals to encourage the use of more-effective care. Alternatively, insurers could require enrollees to pay some or all of the additional costs of more expensive treatments that were shown to be less effective or less cost-effective (in which case enrollees would have to decide whether the added benefits were worth the added costs). Indeed, some recent proposals call for a "value-based" design of insurance, which encourages the use of services when the clinical benefits exceed the costs and likewise discourages the use of services when the benefits do not justify the costs. Although insurance plans generally vary cost sharing by the type of service provided, that new approach would be tailored to the patient and the treatment.

In the past, payers have routinely used considerations of clinical and cost-effectiveness to design tiered drug formularies, but have otherwise not had a practical model for tiering the benefits related to procedures, devices, and many specialty services. Instead, many benefit designs have relied on a "blunt" deductible that applies equally to all health care services, no matter what the effectiveness or value. Comparative effectiveness reviews, if formatted with ratings of comparative clinical and cost-effectiveness, could support a transparent, evidence-based tiered benefit design for these other services in which patients would pay less out of pocket for "high-value" diagnostic or treatment options. Value-based insurance designs dropping out-of-pocket payments for patients with certain chronic conditions have now been adopted by several large employers, with early results suggesting improved clinical outcomes and cost savings.[13]

Comparative effectiveness ratings could also be useful to support innovative coverage decisions that link coverage to requirements for clinicians and patients to participate in ongoing clinical research.[14] Called "coverage with evidence development" or "coverage under protocol," this approach to coverage has been used by Medicare for several high-profile services, including implantable cardio-defibrillators, and PET scans for cancer management. Coverage with evidence development seeks to balance the interest in making a promising intervention available to

12. Congressional Budget Office. 2007. Research on the comparative effectiveness of medical treatments. Washington, DC: CBO
13. Fendrick AM, Chernew ME. Value-based insurance design: a "clinically sensitive" approach to preserve quality and contain costs. *Am J Manag Care*. 2006;1:18-20.
14. Tunis SR, Pearson SD. Coverage Options For Promising Technologies: Medicare's 'Coverage With Evidence Development." *Health Affairs* 2006; 25: 1218-1230.

patients who might benefit with the need for better evidence on the intervention's true long-term risks and benefits. Public and private payers need a valid, objective way to decide which new technologies should receive this form of coverage, and a rating system for comparative effectiveness results could provide the foundation on which this type of approach could grow with greater legitimacy. Private health plans should continue to develop new insurance contract language and other ancillary policy tools so that they can be ready to capitalize on the existence of independent, authoritative comparative effectiveness information that can help identify good opportunities for coverage with evidence development. And a federal comparative effectiveness organization should work closely with Medicare to consider allocating research funds to support coverage with evidence development research efforts when there is no manufacturer or sponsor to pay for the research that would be required.

Authoritative comparative effectiveness information should also help support novel payment policies that will be more effective at rewarding evidence-based practice. Too often current payment arrangements primarily reward clinicians for doing more – more tests, more treatments, more visits. Traditional fee-for-service arrangements create financial incentives for service volume that do not reflect whether clinicians engage in evidence-based shared decision-making with patients, whether care is based on judicious application of evidence, or whether patients experience high quality clinical outcomes.

Payment for new services is where many observers, including MedPAC, have felt that Medicare has the best opportunities to incorporate comparative effectiveness results into its policies. For example, Medicare could set an initial payment for a new intervention at a level commensurate with existing alternatives if there is inadequate evidence to demonstrate that the newer option is superior; such a "reference pricing" approach might be linked with enhanced efforts by Medicare to grant coverage under coverage with evidence development to promising interventions earlier, before evidence is fully adequate to judge comparative effectiveness. Coverage with evidence development can be called "dynamic coverage"; comparative effectiveness evidence could also help support an important corollary policy: "dynamic pricing." The underlying shift would take the United States health care system from a traditional approach that has set payment largely on the basis of "reimbursement" for clinician time and materiel, and move toward payment tied to independently judged evidence of net benefit compared to existing alternatives. This shift would prove controversial with manufacturers and providers, and would require new legislative authorities, but could prove to be one of the best ways for comparative effectiveness to improve the value of health care services under Medicare.

Comparative effectiveness information will also be able to support many forms of medical management in private health plans that seek to make most prudent use of expensive interventions.

Comparative effectiveness information will also be able to support many forms of medical management in private health plans that seek to make most prudent use of expensive interventions. There are radiology authorization programs in which physicians requesting an expensive scan must apply through a pre-authorization process; step-edits in pharmacy management in which patients must try a less expensive alternative first before receiving coverage for a more expensive (and sometimes risky) alternative; and case management programs, in which patients' overall care is coordinated by nurse-physician teams using evidence-based protocols. Al-

though these programs are often resented by patients and physicians, they do represent an attempt to provide coverage of expensive interventions while attempting to keep their use within prudent bounds. A robust federal comparative effectiveness program will produce evidence that can support and fine-tune these kinds of efforts, providing greater distinction among patient subpopulations on which patients should have different pathways of care; and perhaps, in some cases, providing evidence that the newer, more expensive alternatives are in fact no better than existing options.

A federal comparative effectiveness organization should work closely with NQF, professional clinician organizations, and other affiliated groups to seek an alignment between what comparative effectiveness research suggests is "best practice" and the development of quality performance measures.

A final area in which the uptake and impact of comparative effectiveness can be enhanced by alignment with medical policy is in quality performance measurement. Clinicians around the United States are becoming accustomed to having the quality of their clinical care assessed, and often publicly reported, on the basis of quality performance measures. These measures usually assess whether the clinician performed a "process" of care believed to be indicated for the patient, such as ordering a mammogram for a woman over age 50, or prescribing beta blockers for a patient following a heart attack. In recent years a process for validation of quality performance measures has evolved and is now led at the national level by the National Quality Forum (NQF) (see www.nqf.org). The NQF is a not-for-profit membership organization created to develop and implement a national strategy for health care quality measurement and reporting. A federal comparative effectiveness organization should work closely with NQF, professional clinician organizations, and other affiliated groups to seek an alignment between what comparative effectiveness research suggests is "best practice" and the development of quality performance measures. Here again the support runs both ways: having NQF use comparative effectiveness reviews as the basis for designating quality performance measures will support the uptake and adoption of the products of a federal comparative effectiveness program. Similarly, these products will improve the underlying evidentiary base of quality performance measures, thus enhancing their acceptance and impact among clinicians. Alignment of quality measurement and comparative effectiveness will be one more piece of a strategy that will be necessary in order to make sure that comparative effectiveness is used and used well in improving the quality and value of the health care system.

Conclusion

The barriers to the use of comparative effectiveness information in the health care system are significant, and nothing in the recommendations or specific strategies above should be construed as implying that it will be easy to overcome poor coordination of review and research efforts, impaired legitimacy of review organizations and processes, limited applicability of the framing and formats of research, inadequate development and deployment of medical policies, and misaligned payment structures. Easy, no; but the opportunity for a federal comparative effectiveness program to make headway against

these barriers should also not be discounted. The central product of comparative effectiveness will be better evidence of what works best when, and for whom. If this information can be produced, disseminated, and implemented through an aligned set of approaches to facilitate decision-making, then the chances will be good for significant improvements in the quality and value of health care in the United States.

Some stakeholders would prefer a comparative effectiveness program with strict "safeguards" against use of the information by payers in coverage or pricing decisions. But the design of comparative effectiveness and its products can be framed to help payer policies contribute to a more personalized health care system. Knowing what works best, when, and for whom, will allow policies to reflect variation in patients and, indeed, increase the chances that patients will receive care that benefits them as individuals. At the same time, comparative effectiveness that embraces a rigorous, transparent approach to integrating reviews of cost-effectiveness with clinical effectiveness will provide the information for patients and clinicians to select higher value care options when clinical outcomes are equivalent. This information will also focus payer policies on cost-effectiveness and not simply cost, while providing rewards for medical innovators who develop new, significant approaches to meet unmet patient needs.

Comparative effectiveness that embraces a rigorous, transparent approach to integrating reviews of cost-effectiveness with clinical effectiveness will provide the information for patients and clinicians to select higher value care options when clinical outcomes are equivalent.

The federal technology assessment agencies of Australia, Canada, France, Germany, and England all present examples of comparative effectiveness programs whose primary function is to provide explicit recommendations to national coverage bodies responsible for coverage and payment decisions. Their experiences, both good and bad, provide many valuable lessons, but these organizations do not provide models for how the priorities, formats, and implementation of comparative effectiveness would best be designed for the pluralistic health care system of the United States. A uniquely American balance will need to be found in order to produce a unique approach to providing guidance to patients, clinicians, and payers. Balancing evidence synthesis and evidence generation, creating a process for greater stakeholder involvement while maintaining scientific integrity, developing methods for translating evidence into ratings rather than recommendations or decisions – all these challenges and more will reflect the unique dynamics and needs of patients and other stakeholders in the U.S. health care system.

A single, high-profile federal organization is the best structure to coordinate the broad functions of comparative effectiveness envisioned here: to set priorities, commission research, develop methods, and produce and disseminate reports. But its most important responsibility will be to communicate in a trustworthy fashion with the American public. There will be unavoidable tensions between eagerness for innovation and the limits of evidence; between the hopes for personalized care and the pragmatic boundaries of clinical systems and medical policies. A federal comparative effectiveness organization will need to provide independent, rigorous guidance without mandating practice or

policy decisions. To do so it will need to embody the highest standards of scientific excellence. It will need insulation from excessive influence from political or private concerns; but just as importantly, it will need to be guided by principles of integrity and transparency. It will need to embrace the patient's perspective as it wrestles with broader questions of limited health care resources. And, to overcome the many challenges it will face, it will need to exercise all of its functions in a spirit of reflection and humility as it seeks to fulfill a task of deep importance to the American people.

BOX A

The National Institute for Health and Clinical Excellence (NICE)

NICE was established in 1999 to provide health professionals in England and Wales with advice on securing the highest attainable standards of care for patients in the National Health Service (NHS). Since its inception NICE has gained prominence for the rigor and independence of its technology assessment process. In addition, NICE relies on cost-effectiveness as the fundamental basis of comparison between new technologies and their alternatives. For these reasons, and for its tested political durability, NICE has become the most influential technology assessment program in the world.

Structure, Function, and Methods

NICE has a broad mandate to set standards for the use of new technologies and procedures within the NHS and to produce guidelines for clinical, and now public, health. NICE methodology makes an important distinction between technology assessment and technology appraisal. Assessment puts the evidence together; appraisal judges it in the context of a decision to fund or not fund the technology across the NHS.

NICE commissions its technology assessments from a network of independent academic groups across England. Technology assessment follows a standard and well-described methodology that includes a full systematic review of the topic and a rigorous approach to economic modeling of cost-effectiveness. The appraisal phase sees the technology assessment delivered back from the academic unit to NICE, but only so that the evidence can be considered there by an independent advisory committee, chaired by NICE staff but whose members are drawn from clinicians, professional groups, researchers, and individuals with experience in patient advocacy. Based on its deliberations, the advisory committee makes its recommendation on funding to the NHS. Although health ministers in the government have reserve powers to advise the NHS to ignore NICE guidance, they have never done so.

Stakeholders (including relevant professional and patient organizations as well as manufacturers) are involved at all stages, from the preliminary scoping exercise that establishes the boundaries and comparator technologies for the appraisal, through the assessment phase when they have full access to the supporting systematic reviews; and on to the appraisal phase, when they are encouraged to comment on draft forms of guidance. There is also a formal appeals mechanism for stakeholders which, until recently, had proven robust enough to ward off formal legal challenges. Finally, the Institute attempts to ensure that its processes are as transparent as possible: NICE's work programs and timelines are publicized well in advance; and the data from which its conclusions are drawn are in the public domain with the exception of the details of studies that manufacturers insist remain "commercial-in-confidence." During its first seven years NICE held its independent appraisal committee meetings in private, but has recently opened up all meetings to the public.

Economic Evaluation

The key measure used by NICE to assess the comparative value of a technology is the additional cost per quality adjusted life year (QALY) gained. If appropriate data on quality of life are unavailable, cost-effectiveness is estimated using alternatives such as the cost per life year gained. NICE expects its advisory bodies to use estimates of

Continued on next page

BOX A-CONTINUED

cost effectiveness to inform, but not determine, their decisions. Nevertheless, NICE has arrived operationally at a band of approximately $30,600 – $45,900 per QALY (based on purchasing power parity of US$1 = £0.65) as the threshold above which it would be increasingly likely to reject a technology on grounds of cost-ineffectiveness. For example, the Institute has approved the use of eternacept and infliximab, both with incremental cost effectiveness ratios of $47,430 per QALY, in the treatment of rheumatoid arthritis; but it has rejected another treatment option, anakinra, which has an incremental ratio of $102,510 per QALY.

NICE does not take the budget impact of a new technology into account. For example, although a new drug might have a favorable cost-effectiveness ratio of $20,000/QALY, the overall impact might be quite significant for the NHS budget if large numbers of patients were to be eligible for treatment. The recent approval by NICE of Herceptin has created just such a bind for NHS budgets. Although the drug itself was found to be cost-effective, its relative high cost and the increasing number of patients eligible to take it have created a situation in which it has been estimated that 25 percent of the entire budget for cancer care in England could be spent on Herceptin. Nevertheless, NICE's methods continue to ignore issues of affordability; the government remains accountable for the overall NHS budget and therefore has the responsibility to judge a particular intervention unaffordable for the NHS even though NICE might have judged it cost-effective.

BOX B

The Blue Cross Blue Shield Association Technology Evaluation Center (TEC)

The TEC, established by the Blue Cross and Blue Shield Association in 1985, and now administered in partnership with Kaiser Permanente, performs assessments of clinical effectiveness and appropriateness within specific patient populations, but does not generally consider costs or cost-effectiveness. The framework that TEC uses to assess technologies has been adapted or fully adopted by many other technology assessment initiatives in the United States and so is worth exploring in some detail. The TEC uses five formal evaluation criteria for judging the effectiveness of a medical technology. These five criteria are:

- The technology must have the final approval from the appropriate government regulatory bodies, if applicable.
- The scientific evidence must permit conclusions concerning the effects of the technology on health outcomes.
- The technology must improve the net health outcome.
- The technology must be as beneficial as any established alternatives.
- The improvement must be attainable outside the investigational setting.

TEC uses a formal approach when reviewing the evidence and judging a technology against each of these five criteria. A technology assessment is first prepared by the core staff of the TEC and then presented to its Medical Advisory Panel, or MAP. The MAP is composed of nationally respected clinical and methodological experts, with membership diversified to ensure representation by a wide variety of viewpoints within the health care community. A majority of members hold academic appointments and are independent medical experts without affiliation to healthcare payers.

The MAP judges a technology separately upon each of the five TEC criteria. These five criteria, which have served TEC for over 20 years, were developed primarily to help private health plans judge whether a new technology was no longer investigational or experimental. This has been a key distinction and threshold for coverage decisions, since health plan contracts generally exclude coverage for all experimental and investigational services. With little support from court decisions for other evidence-based methods as a basis for judging medical necessity, disqualification as investigational or experimental has been the mainstay of the use of technology assessments by private health plans. Thus the TEC criteria were designed primarily to help provide the basis for a dichotomous "yes/no" coverage decision on new technologies and this approach continues to be the most prominent and respected approach in the United States to informing coverage decisions.

BOX C

The Institute for Clinical and Economic Review (ICER)

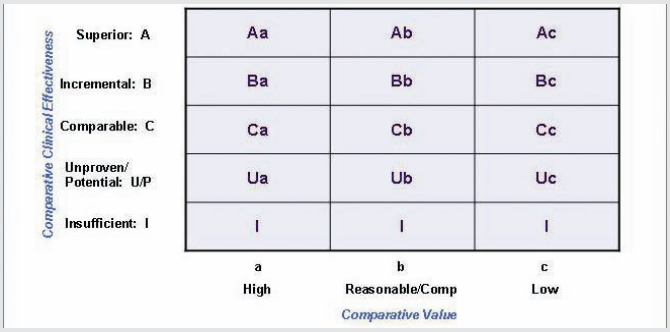
ICER provides independent evaluation of the clinical effectiveness and comparative value of new and emerging technologies. ICER is based at the Massachusetts General Hospital's Institute for Technology Assessment (ITA), an affiliate of Harvard Medical School. Purposely structured as a fully transparent organization, ICER is able to engage with all key stakeholders in its assessments while retaining complete independence in the formulation of its conclusions and the drafting of its reviews.

ICER develops its assessments in collaboration with faculty from the ITA and Harvard Medical School, as well as with clinical experts, patient groups, and policy leaders from around the country. To communicate its results in an actionable format, ICER assigns integrated ratings of clinical effectiveness and cost-effectiveness, ratings supported by rigorous systematic reviews and economic modeling. These integrated ratings answer not only the question, "Does it work?" but provide objective and trustworthy information with which to answer the question, "What is its relative value compared to other alternatives?"

The ICER rating system is specifically formatted to support innovative patient-clinician decision support tools and value-based coverage and reimbursement policy. But ICER aims to do more than produce actionable evidence reviews; its goal is to catalyze the application of evidence to achieve improved quality and value across the healthcare system. As an academic enterprise, ICER can serve as an "honest broker" to bring all stakeholders in health care together to assess evidence and then develop collaborative mechanisms for applying evidence to shift patterns of care towards higher value.

There are two components to the ICER Integrated Evidence Rating™, shown in the Figure below:

ICER Integrated Evidence Rating



"Comparative clinical effectiveness" is assessed through systematic review of the medical literature. Meta-analysis, decision modeling, and other quantitative methods are used where appropriate. A final comparative clinical effectiveness rating is assigned according to a structured consideration of the strength of evidence and the magnitude of the net health benefit (if any) provided by the new technology.

The second component of ICER's assessment is a rating of "comparative value." This is assessed on the basis of the results of an economic model that captures all clinical outcomes and costs. Aided by the input of an external Evidence Review Group, ICER assigns an ultimate rating of "high," "reasonable/comparable," or "low" comparative value based on economic outcomes such as the cost per case diagnosed, cost per hospitalization averted, and cost per quality-adjusted life year gained. The goal of ICER's value rating is to overcome the problem of elegant yet impenetrable cost-effectiveness information. ICER's cost-effectiveness work is of the highest scientific caliber, and it becomes a robust guide to dialogue and action through translation into tangible ratings.

TABLE 1

Options for the scope of CER, with aligned methods and mechanisms of implementation

Scope of CER	Research Approach	Topic Selection	Format of the products of CER	Mechanisms of Implementation
Exclusive Focus: Informing Patients and Clinicians Providing information for individual patients and clinicians to decide how to apply in care of the patient.	Evidence generation > evidence synthesis	1. General approaches to conditions > "drug vs. drug" 2. Clinical uncertainty > budget impact 3. Existing interventions > expensive emerging interventions	Evidence synthesis: • Emphasis on individual level differences in harms and benefits • Limited or no judgments of net health benefits • Only out-of-pocket costs are evaluated, if costs are considered at all Evidence generation: • Research results published with limited formal judgments of net health benefits or implications for use in clinical practice	1. Dissemination of information
Balanced Focus: Providing Guidance for Individual Decisions and Insurer Policies Translating information into guidance tailored for patients, clinicians, and insurers	Evidence generation = evidence synthesis	1. General approaches to conditions = "drug vs. drug" 2. Clinical uncertainty = budget impact 3. Existing interventions = expensive emerging interventions	Evidence synthesis: • Judgment of comparative clinical effectiveness for all clinically identifiable subgroups • Cost and cost-effectiveness evaluated and graded or rated separately from clinical effectiveness Evidence generation: • Research results rapidly reviewed and put in context as part of evidence synthesis	1. Dissemination of guidance 2. Rapid incorporation into professional guidelines 3. Transparent use by insurers in value-based policies
Dominant Focus: Making Payer Decisions Linking evidence directly to coverage decisions governing access to services at population levels	Evidence synthesis > evidence generation	1. "Drug vs. drug" > general approaches to conditions 2. Budget impact > clinical uncertainty 3. Expensive emerging interventions > existing interventions	Evidence synthesis: • Emphasis on judgment of net comparative clinical effectiveness at the population level • Cost-effectiveness from payer's perspective used with single threshold as a cut-off for coverage recommendations Evidence generation: • Research results rapidly reviewed and judged as part of evidence synthesis	1. Dissemination of guidance 2. Explicit recommendations for coverage or other direct links to medical policy decisions

Acknowledgements

The author is grateful to Marc Berger, Steven Findlay, Scott Gottlieb, Tim Taylor, Mark McClellan, and Joshua Benner for their constructive comments on a previous draft of this paper.

About the Authors

Comparative Effectiveness Research: Will It Bend the Health Care Cost Curve and Improve Quality?

Mark McClellan, MD, PhD

Director, Engelberg Center for Health Care Reform and Leonard D. Schaeffer Chair in Health Policy Studies, The Brookings Institution

As director of the Engelberg Center at Brookings, Mark McClellan's work focuses on promoting high-quality, innovative and affordable health care. A doctor and economist by training, he also has a highly distinguished record in public service and in academic research. Dr. McClellan is a former administrator of the Centers for Medicare and Medicaid Services (CMS) and former commissioner of the Food and Drug Administration (FDA), where he developed and implemented major reforms in health policy. These include the Medicare prescription drug benefit, the FDA's Critical Path Initiative, and public-private initiatives to develop better information on the quality and cost of care. Dr. McClellan previously served as a member of the President's Council of Economic Advisers and senior director for health care policy at the White House, as well as an associate professor of economics and medicine at Stanford University.

Joshua S. Benner, PharmD, ScD

Research Director, Engelberg Center for Health Care Reform, The Brookings Institution

Josh Benner is Research Director at the Engelberg Center for Health Care Reform at the Brookings Institution, and an adjunct scholar at the Center for Clinical Epidemiology & Biostatistics at the University of Pennsylvania School of Medicine. His work at Brookings focuses on the development of better evidence to inform health care decisions, including comparative effectiveness research, medical product safety surveillance, and quality measurement. It also includes initiatives to improve the effectiveness and efficiency of medical product development and regulatory approval. Prior to joining Brookings, he was Senior Principal in Health Economics and Outcomes Research at IMS Health. He was previously an AHRQ post-doctoral fellow in health services research at the Brigham and Women's Hospital Division of Pharmacoepidemiology & Pharmacoeconomics.

Setting Priorities for Comparative Effectiveness Research

Alan M. Garber, MD, PhD

Henry J. Kaiser Jr. Professor and Director of the Center for Health Policy and the Center for Primary Care and Outcomes Research, Stanford University School of Medicine and Staff Physician, Veterans Affairs Palo Alto Health Care System

Alan M. Garber is the Henry J. Kaiser Jr. Professor of Medicine at Stanford, where he directs the Center for Health Policy and the Center for Primary Care and Outcomes Research. He is a Staff Physician at the Palo Alto VA and directs the Health Care Program of the National Bureau of Economic Research (NBER). He is a member of the Panel of Health Advisers of the Congressional Budget Office, the American Society for Clinical Investigation, the Association of American Physicians, and the Institute of Medicine of the National Academy of Sciences. Dr. Garber served as the Chair of the Medicare Evidence Development and Coverage Advisory Committee (Centers for Medicare and Medicaid Services) and as a member of the National Advisory Council on Aging (National Institutes of Health), and as a member of many committees of the National Institutes of Health and of the National Academies. His work addresses methods for improving health care delivery and financing, particularly for the elderly. It encompasses technology evaluation, analysis of the causes of health expenditure growth, and health care productivity.

David O. Meltzer, MD, PhD
Chief, Section of Hospital Medicine
Director, Center for Health and the Social Sciences
The University of Chicago

David Meltzer is Director of the Center for Health and the Social Sciences, Chief of the Section of Hospital Medicine, and Director of the MD/PhD Program in the Social Sciences at the University of Chicago. His research examines the theoretical foundations of medical cost-effectiveness analysis and the determinants of the cost and quality of hospital care, with a focus on issues related to older adults. He is principal investigator of the Chicago Hospital Medicine and Economics Center for Education and Research in Therapeutics (CERT), funded by the Agency for Healthcare Research and Quality, and several research and training grants from the National Institute of Aging.

Strategies to Improve CER Methods and Data Infrastructure

Sean R. Tunis, MD, MSc
Founder and Director, Center for Medical Technology Policy

Sean Tunis is the Founder and Director of the Center for Medical Technology Policy, a non-profit organization that supports collaborative projects involving a broad range of experts and stakeholders, to improve the value of clinical research for health care decision making. He also advises numerous domestic and international health care organizations on issues of comparative effectiveness, evidence based medicine, clinical research and technology policy. Dr. Tunis was Chief Medical Officer at the Centers for Medicare and Medicaid Services (CMS) and also directed the Office of Clinical Standards and Quality. Prior to his work at CMS, Dr. Tunis served as the Director of the Health Program at the Congressional Office of Technology Assessment and as a health policy advisor to the U.S. Senate Committee on Labor and Human Resources, where he participated in policy development regarding pharmaceutical and device regulation.

From Better Evidence to Better Care: Using CER to Guide Practice and Policy

Steven Pearson, MD
Founder and President of the Institute for Clinical and Economic Review, Massachusetts General Hospital and Harvard Medical School

Steven Pearson is the Founder and President of the Institute for Clinical and Economic Review (ICER) at Massachusetts General Hospital and Harvard Medical School. He also serves as the Vice-Chair of the Medicare Evidence Development and Coverage Advisory Committee, and he is a Senior Scientist in the Department of Bioethics at the National Institutes of Health. He attended UCSF School of Medicine and completed his residency in internal medicine at Brigham and Women's Hospital in Boston. An internist, health services researcher, and ethicist, he was awarded an Atlantic Fellowship from the British government in 2004 to pursue policy studies at the National Institute for Health and Clinical Excellence (NICE) in London, England. He returned to the US to serve from 2005-2006 as Special Advisor, Technology and Coverage Policy, within the Coverage and Analysis Group at the Centers for Medicare and Medicaid Services. He also served as Senior Fellow at America's Health Insurance Plans from 2006-2008. Dr. Pearson co-chairs the Measurement Implementation Strategies Workgroup of the Quality Alliance Steering Committee.

BROOKINGS

The Brookings Institution
1775 Massachusetts Ave., NW, Washington, DC 20036
www.brookings.edu/healthreform ▪ www.hamiltonproject.org