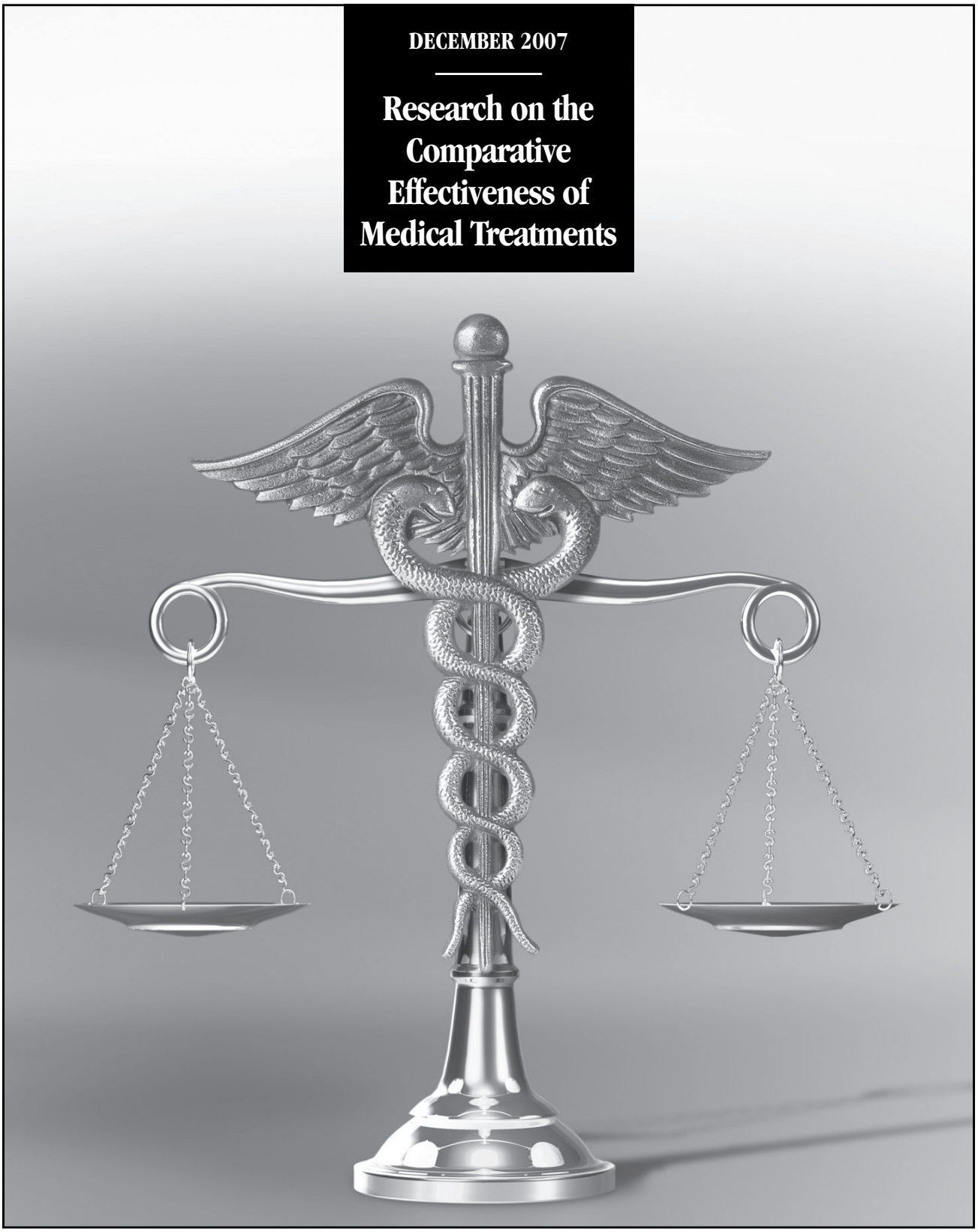
CONGRESS OF THE UNITED STATES
CONGRESSIONAL BUDGET OFFICE

A
CBO
PAPER

DECEMBER 2007

Research on the Comparative Effectiveness of Medical Treatments



© Comstock Images

Pub. No. 2975

A

CBO

PAPER

Research on the Comparative Effectiveness of Medical Treatments: Issues and Options for an Expanded Federal Role

December 2007

The Congress of the United States ■ Congressional Budget Office

Preface

Rising costs for health care represent a central challenge both for the federal government and the private sector, but opportunities may exist to constrain costs in both sectors without adverse health consequences. Perhaps the most compelling evidence of those opportunities involves the significant geographic differences in spending on health care within the United States, which do not, on average, translate into higher life expectancy or substantial improvements in other health statistics in the higher-spending regions. At the same time, only a limited amount of evidence is available about which treatments work best for which patients and whether the added benefits of more-effective but more-expensive services are sufficient to warrant their added costs. Together, those findings suggest that generating better information about the costs and benefits of different treatment options—through research on the comparative effectiveness of those options—could help reduce health care spending without adversely affecting health overall.

This Congressional Budget Office (CBO) paper—prepared at the request of the Chairmen of the Senate Budget and Finance Committees—examines options for expanding federal support for research on comparative effectiveness. It reviews the current state of such research in both the public and private sectors and discusses several mechanisms for organizing and funding additional research efforts. It also discusses the different types of research that could be pursued and their likely benefits and costs. Finally, it considers the potential effects that such research could have on health care spending and the difficult steps that public and private insurers would probably have to take to achieve substantial savings on the basis of that research—in particular, changing the financial incentives for doctors and patients to reflect that information. In accordance with CBO's mandate to provide objective, impartial analysis, this paper contains no recommendations.

Philip Ellis of CBO's Health and Human Resources Division prepared the paper, with valuable contributions from Colin Baker and Morgan Hanger. The analysis benefited from comments by Dr. Alan Garber, Henry J. Kaiser Professor of Medicine at Stanford University, and Dr. Sean Tunis of the Center for Medical Technology Policy. (The assistance of external reviewers implies no responsibility for the final product, which rests solely with CBO.)

John Skeen edited the paper, and Maureen Costantino prepared it for publication and designed the cover. Lenny Skutnik printed the initial copies, Linda Schimmel handled the print distribution, and Simone Thomas prepared the electronic version for CBO's Web site (www.cbo.gov).

Peter R. Orszag
Director

December 2007

Contents

Summary and Introduction 1

The Current State of Comparative Effectiveness Research 3

What Is Comparative Effectiveness? 3

Research in the Private Sector 7

Past and Current Federal Efforts 9

The Consequences of Limited Information 11

Options for Organizing and Funding New Federal Research Efforts 15

Options for Comparing the Effectiveness of Treatments 19

Methods of Research 19

The Scope and Focus of Analyses and the Dissemination of Results 25

Implications for Health Care Spending 28

The Potential for Savings on Health Care 29

Possible Responses by Private and Public Insurance Plans 30

Estimated Effects of a Recent Proposal 34

Tables

1. Requested, Proposed, and Actual Funding for the Agency for Health Care Research and Quality 10

2. Cost-Effectiveness of Different Screening Methods for Colorectal Cancer 21

Figures

1. Federal Spending for Medicare and Medicaid as a Percentage of Gross Domestic Product Under Different Assumptions About Excess Cost Growth 2

2. Medicare Spending per Capita in the United States, by Hospital Referral Region, 2003 13

3. Rates of Four Orthopedic Procedures Among Medicare Enrollees, 2002 and 2003 14

Boxes

1. Research on Comparative Effectiveness in Other Countries 6

2. The U.S. Preventive Services Task Force 18

Research on the Comparative Effectiveness of Medical Treatments

Summary and Introduction

Over the past 30 years, federal spending on Medicare and Medicaid has roughly tripled as a share of gross domestic product (GDP), rising from about 1.3 percent in 1975 to about 4 percent in 2007. According to the Congressional Budget Office's (CBO's) projections, under current policies such spending will reach about 12 percent of GDP by 2050—but substantial uncertainty surrounds that estimate.[1] If costs per enrollee continued growing over the next four decades as quickly as they have grown over the past four—about 2.5 percentage points faster than per capita GDP—then federal spending on those programs would reach about 17 percent of the economy. If, instead, costs per enrollee did not exceed the growth of GDP, those federal costs would reach about 6 percent of GDP in 2050 solely because of demographic changes (see Figure 1). As those figures indicate, the rate at which health care costs grow relative to income is the most important determinant of the country's long-term fiscal balance; it exerts a significantly larger influence on the budget over the long term than other commonly cited factors, such as the aging of the population or the coming retirement of the baby-boom generation.[2]

Rising health care costs represent a challenge not only for the federal government but also for private payers. Indeed, trends in both sectors reflect many of the same underlying forces—including the development and spread of new and more-expensive medical technologies—so controlling those federal costs over the long term will be difficult without addressing the forces that are also causing private costs for health care to rise. Total health care spending, which consumed about 8 percent of the U.S. economy in 1975, currently accounts for about 16 percent of GDP, and that share is projected to reach nearly 20 percent by 2016. About half of overall health spending in the United States is now publicly financed, and half is privately financed.

A variety of evidence suggests that opportunities exist to constrain health care costs both in the public programs and in the rest of the health system without adverse health consequences. Perhaps the most compelling evidence of those opportunities involves the substantial geographic differences in spending on health care—both among countries and within the United States—which do not translate into higher life expectancy or measured improvements in other health statistics in the higher-spending regions. For example, Medicare's costs per beneficiary vary significantly among different regions of the country, but much of the variation cannot be explained by differences in the population, and the higher-spending regions perform no better on available measures of average health outcomes than the lower-spending regions do.

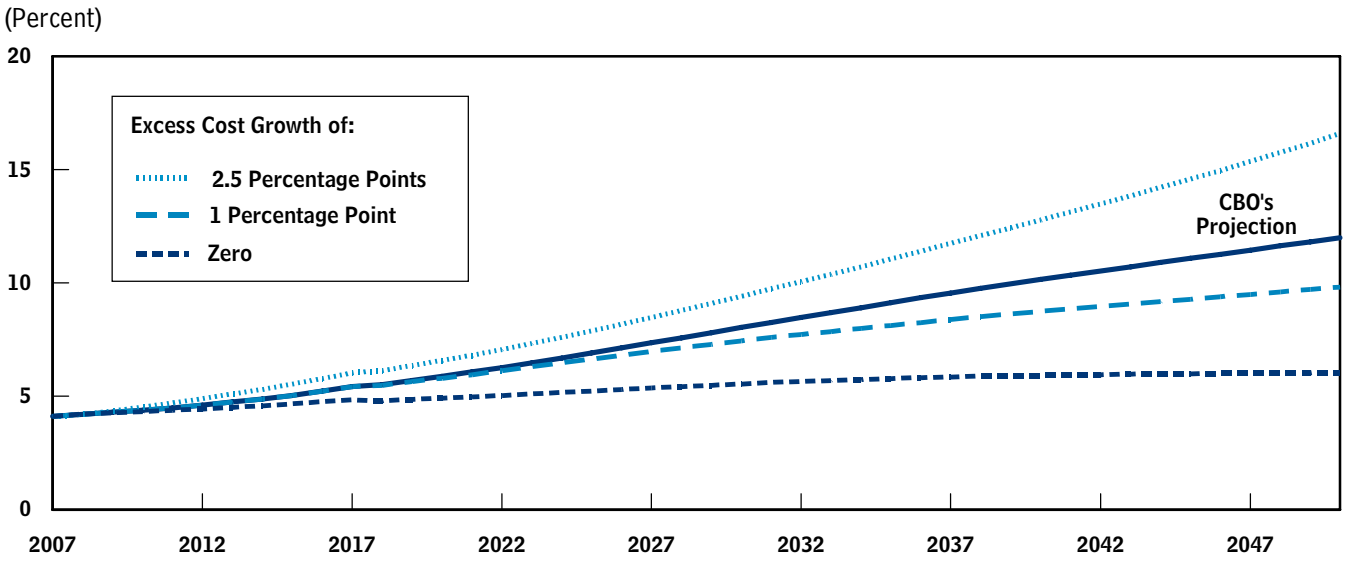
Furthermore, hard evidence is often unavailable about which treatments work best for which patients and whether the added benefits of more-effective but more-expensive services are sufficient to warrant their added costs—yet the current health system tends to adopt more-expensive treatments even in the absence of rigorous assessments of their impact. Indeed, the extent of the variation in treatments may be greatest when evidence about their relative effectiveness is lacking. Together, those findings suggest that better information about the costs, risks, and benefits of different treatment options,

1. Congressional Budget Office, *The Long-Term Outlook for Health Care Spending* (November 2007). The estimates of federal spending reflect Medicare's costs net of the premiums that enrollees pay and other offsetting receipts; the program's gross costs are about 15 percent higher than its net costs.

2. For additional discussion, see Congressional Budget Office, *The Long-Term Budget Outlook* (December 2007).

Figure 1.

Federal Spending for Medicare and Medicaid as a Percentage of Gross Domestic Product Under Different Assumptions About Excess Cost Growth



Source: Congressional Budget Office.

Note: Excess cost growth refers to the number of percentage points by which the growth of annual health care spending per beneficiary is assumed to exceed the growth of nominal gross domestic product per capita.

combined with new incentives reflecting the information, could eventually alter the way in which medicine is practiced and yield lower health care spending without having adverse effects on health. Over the long term, the potential reduction in spending below projected levels could be substantial.

Generating evidence that compares treatments is what research on "comparative effectiveness" does. This Congressional Budget Office (CBO) paper makes the following main points about the options that are available for an expanded federal role in supporting and organizing such research and about the impact that research could have on spending for health care:

- Because any private-sector entity (such as a health plan) has only a limited incentive to produce or pay for information that could benefit many entities—including its competitors—an argument can be made for a larger federal role in coordinating and funding research on comparative effectiveness. In addition, because federal health insurance programs play such a large role in financing medical care and account for such a large share of the budget, the federal government itself has an interest in generating evaluations of the effectiveness of different approaches to health care.

- If policymakers wanted to expand federal efforts to study comparative effectiveness, the endeavor could be organized in different ways—for instance, by augmenting an existing agency, by establishing a new agency, by supporting an existing quasi-governmental organization, or by creating a new public-private partnership. In choosing an organizational arrangement and a mechanism to provide federal funds to it, tradeoffs could arise between the entity's independence from political pressure and its accountability to policymakers and other interested parties. Efforts to bolster comparative effectiveness research would be more likely to change medical practice if the organization coordinating the research was respected and trusted by doctors and other professionals in the health sector.

- The level of funding required for a new or augmented entity would depend largely on what its additional activities involved. Synthesizing existing studies or analyzing available data on medical claims would be less expensive than conducting new head-to-head clinical trials to compare treatments but could also yield

less definitive results—and therefore might have a smaller impact on medical practice. Clinical trials could be more persuasive but also more time-consuming, and there is probably a limit to how many comparative trials could be undertaken effectively at any given time. If privacy concerns could be addressed, having more health records available in electronic form would facilitate the use of such data for research.

- Studies might need to compare not only broadly different treatment options—such as surgery versus drug therapy—but also different approaches to the same basic treatment—such as different levels of follow-up care after surgery. Studies that included an analysis of cost-effectiveness would probably have a larger impact than ones that compared only clinical effectiveness, because they would highlight cases where more-expensive treatments or approaches provided added benefits that were modest compared with their added costs (at least for some types of patients).

- To affect medical treatment and reduce health care spending in a meaningful way, the results of comparative effectiveness analyses would not only have to be persuasive but also would have to be used in ways that changed the behavior of doctors, other health professionals, and patients. For example, the higher-value care identified by comparative effectiveness research could be promoted in the health system through financial incentives—the payments doctors receive or the cost sharing that patients face. Making substantial changes in payment policies or coverage rules under the Medicare program to reflect information on comparative effectiveness would almost certainly require legislation.

- Making such substantial changes in the delivery of health care could prove difficult and controversial for a number of reasons. To inform new systems of incentives—designed to discourage the use of more costly treatments that provided little or no added benefits—the results of effectiveness studies would have to be sufficiently robust to minimize the risk of overlooking subgroups of patients who could benefit greatly from a treatment. Even with an expanded evidence base, some patients and providers might object to the use of such incentives, and keeping pace with new treatments and procedures would be an ongoing challenge.

- Generating additional information about comparative effectiveness and making corresponding changes in incentives would seem likely to reduce health care spending over time—potentially to a significant degree. The precise impact, however, depends on several factors and is difficult to predict. Given the time necessary to conduct the research, to alter incentives in a manner reflecting the results, and to affect behavior through those changes, any potential for substantial cost savings from new research would probably take a decade or more to materialize. Even so, generating additional information comparing treatments would tend to reduce federal health spending somewhat in the near term—but that effect may not be large enough to offset the full costs of conducting the research over that same time period.

The Current State of Comparative Effectiveness Research

In weighing options to expand and reorganize research efforts, it is useful to define what comparative effectiveness research means and to consider the arguments for and expanded federal role in conducting such research. Related issues include the reasons why the current stock of research on comparative effectiveness is limited and why treatments and procedures can gain wide use even when evidence about their relative effectiveness is lacking. Reviewing past and current research efforts—by private and public organizations in the United States and by other countries—also sheds light on several issues and challenges likely to arise in any future U.S. efforts. To the extent that past and current efforts are seen as inadequate, careful consideration of those shortcomings would inform the choice of an organizational approach and funding mechanism for new federal activities.

What Is Comparative Effectiveness?

As applied in the health care sector, an analysis of comparative effectiveness is simply a rigorous evaluation of the impact of different options that are available for treating a given medical condition for a particular set of patients. Such a study may compare similar treatments, such as competing drugs, or it may analyze very different approaches, such as surgery and drug therapy. The analysis may focus only on the relative medical benefits and risks of each option, or it may also weigh both the costs and the benefits of those options. In some cases, a given treatment may prove to be more effective clinically or more cost-effective for a broad range of patients, but fre-

quently a key issue is determining which specific types of patients would benefit most from it. Related terms include cost–benefit analysis, technology assessment, and evidence-based medicine, although the latter concepts do not ordinarily take costs into account.

While some information about the effectiveness of new drugs, medical devices, and procedures is usually available, rigorous comparisons of different treatment options are less common. Drugs and devices must be certified as safe and effective by the Food and Drug Administration (FDA) before they can be marketed in the United States, but with certain exceptions the regulatory process for approving those products does not evaluate them relative to alternatives.[3] Furthermore, physicians commonly prescribe drugs for "off-label" uses—that is, for treatments that have not been certified by the FDA. For drug manufacturers, the costs of conducting additional trials to demonstrate safety and efficacy for a broader set of patients or conditions may outweigh the benefits from the increased sales that would result; in particular, the potential gains from finding a favorable result for a different population would have to be weighed against the risk that safety and efficacy could not be demonstrated conclusively.

Medical procedures, which account for a much larger share of total spending on health care than drugs and devices do, can achieve widespread use without extensive clinical evaluation. In many cases, it may be reasonable to assume that the benefits of a treatment will be similar for related conditions or a broader group of patients. Without hard evidence, however, decisions about what treatments to recommend often depend on the individual experience and judgment of physicians. Various reasons have been cited to explain why the use of new medical technologies can spread even in the absence of proof about their effectiveness and why health costs tend to increase as a result; those reasons include fee-for-service payment of physicians (common in the private sector and prevalent in Medicare, that payment method typically gives doctors a financial incentive to provide more-expensive care) as well as enthusiasm for the newest technology on the part of both doctors and patients.[4] Furthermore, patients with insurance typically pay only a small share of the costs of their treatments, so their incentives to weigh the costs against the benefits are limited—a trade-off inherent in having insurance protection.

A recent example of a comparative effectiveness study indicates that careful analysis can sometimes disprove widely held assumptions about the relative merits of different treatments. The study, which involved patients who had stable coronary artery disease, compared the effects of two treatments: an angioplasty with a metal stent combined with a drug regimen versus the drug regimen alone.[5] Patients were randomly assigned to receive the two treatments, and although the study found that patients treated with angioplasty and a stent had better blood flow and fewer symptoms of heart problems initially, the differences declined over time.[6] More importantly, it found no differences between the two groups in survival rates or the occurrence of heart attacks over a five-year period.

Other examples of studies comparing the clinical effectiveness of different treatment options illustrate the types of findings that they can generate:

3. Clinical trials of new drugs must compare them to alternative medications only when the manufacturer wants to make a claim of superiority in its FDA-approved marketing materials or when giving trial participants a placebo would be unethical (for example, in the case of a study of AIDS drugs).

4. See Mark R. Chassin, "Is Health Care Ready for Six Sigma Quality?" *The Milbank Quarterly,* vol. 76, no. 4 (November 1998), pp. 565–591.

5. Coronary artery disease, or a buildup of plaque in the heart's arteries, is considered stable if a patient experiences some chest pain (angina) but does not have worsening pain over time and has not had a heart attack. In an angioplasty, a small balloon is surgically inserted into a clogged artery and then inflated to expand the opening; a stent—a small wire mesh tube—is commonly added in an effort to keep the artery open.

6. William E. Boden and others, "Optimal Medical Therapy With or Without PCI for Stable Coronary Disease," *The New England Journal of Medicine,* vol. 356, no. 15 (April 12, 2007), pp.1503–1516. Other studies have found that angioplasty with a stent has clear medical benefits for patients who are undergoing a heart attack, illustrating the point that results for a given treatment may differ significantly among different types of patients.

- One recent trial found that older, relatively inexpensive drugs for treating high blood pressure (known as diuretics) were more effective in preventing cardiovascular disease in patients age 55 or older than commonly used newer drugs known as angiotensin-converting enzyme inhibitors and calcium channel blockers.[7]

- Another trial compared the effects of surgery to reduce lung volume for patients suffering from emphysema—a treatment that had anecdotal support but lacked hard evidence about its effectiveness—with standard medical therapy for that disease. For many patients, lung surgery increased their risk of death slightly and did not improve their functional status, but for patients with certain types of lung problems and a limited capacity for exercise, the surgery yielded small net improvements in their quality of life (though not in their survival rates).[8]

- A trial of two statin drugs, which was sponsored by the maker of one of those drugs, found that its competitor's product was more effective both at lowering cholesterol levels and at reducing the risk of mortality—illustrating the point that comparative trials can be risky for manufacturers to conduct.[9]

- Recent studies have found that magnetic resonance imaging combined with mammography is more effective than mammography alone in detecting breast cancer for women with certain genetic markers that indicate a substantial increased risk of contracting that disease; the impact of that difference on survival rates, however, could not be measured.[10]

The range of findings that those studies yielded highlights several characteristics of research on comparative effectiveness. First, studies can examine not only treatments for health problems but also different procedures to screen for the presence of a disease. Second, the findings may have broad applicability or may pertain only to a very specific subset of patients and may also vary in the outcomes considered—such as effects on mortality or other measures of health gains.

Third, studies are often based on clinical trials, in which eligible patients are randomly assigned to the treatments under review—but there are several other methods available to compare treatments, each with its own strengths and weaknesses. Clinical trials can yield persuasive findings but can also be relatively costly and time-consuming to conduct. In particular, a trial designed to determine whether two treatments differ in their effectiveness may require a large number of enrollees to be followed for an extended period in order to generate results that are statistically significant. Less expensive approaches include systematic reviews of the evidence about treatment options, which are essentially meta-analyses of all available studies, and studies that use medical claims data, which can be used to follow large groups of patients who have already received different treatments. The impact of systematic reviews can be limited, however, by the fact that they simply reflect existing evidence, and studies using claims data can be subject to bias because the treatments are not randomly assigned to comparable patients.

The studies cited above focus on relative clinical effects, and not cost-effectiveness. For reasons discussed below, gauging cost-effectiveness as well as clinical effectiveness is sometimes controversial, and some observers believe that the two considerations are in separate fields. But cost-effectiveness analysis appears to be well within the scope of research on comparative effectiveness—and

7. Officers and Coordinators for the ALLHAT (Antihypertensive and Lipid-Lowering Treatment to Prevent Heart Attack Trial) Collaborative Research Group, "Major Outcomes in High-Risk Hypertensive Patients Randomized to Angiotensin-Converting Enzyme Inhibitor or Calcium Channel Blocker vs. Diuretic," *Journal of the American Medical Association*, vol. 288, no. 23 (December 18, 2002), pp. 2981–2997.

8. National Emphysema Treatment Trial Research Group, "A Randomized Trial Comparing Lung-Volume-Reduction Surgery with Medical Therapy for Severe Emphysema," *The New England Journal of Medicine*, vol. 348, no. 21 (May 22, 2003), pp. 2059–2073.

9. Christopher P. Cannon and others, "Intensive Versus Moderate Lipid Lowering with Statins After Acute Coronary Syndromes," *The New England Journal of Medicine*. vol. 350, no. 15 (April 8, 2004), pp. 1495–1504. Note that this study was undertaken in response to a similar one financed by the manufacturer of the other drug, which also showed that drug to be superior at lowering cholesterol levels but did not address mortality risks.

10. Ellen Warner and others, "Surveillance of BRCA1 and BRCA2 Mutation Carriers with Magnetic Resonance Imaging, Ultrasound, Mammography, and Clinical Breast Examination." *Journal of the American Medical Association,* vol. 292, no. 11 (September 15, 2004), pp. 1317–1325; and Mieke Kriege and others, "Efficacy of MRI and Mammography for Breast-Cancer Screening in Women with a Familial History or Genetic Predisposition," *The New England Journal of Medicine,* vol. 351, no. 5 (July 29, 2004), pp. 427–437.

Box 1.

Research on Comparative Effectiveness in Other Countries

Other developed countries also face challenges financing health care costs and have taken various steps to assess the comparative effectiveness of treatments. Unlike the United States, many of those countries establish overall budgets for their national health systems and regularly use the data on comparative effectiveness that are available to help determine the treatments and procedures to be covered and, in some cases, the payment rates. Despite differences in other countries' health insurance systems, the approaches that they have taken to organizing and funding those research and review activities could have lessons for any increased U.S. efforts.

Perhaps the best known example of an agency that assesses comparative effectiveness is the National Institute for Health and Clinical Excellence (NICE), which was established in 1999 as part of the United Kingdom's National Health Service (NHS). It analyzes both the clinical effectiveness and cost-effectiveness of new and existing medicines, procedures, and other technologies and provides guidance on appropriate treatments for specific diseases or types of patients. To date, NICE has published appraisals of over 100 specific technologies, guidance on the use of about 250 medical procedures, and about 60 sets of treatment guidelines—a substantial but not exhaustive list. If NICE approves a drug, device, or procedure, it must be covered by the NHS, but local health authorities make coverage decisions about treatments that NICE has not yet evaluated. With a staff of about 200 and an annual budget of about 30 million pounds (roughly $60 million), NICE does not fund new clinical trials or other forms of primary data collection. Instead, it commissions systematic reviews of existing research on clinical effectiveness and combines those findings with models of cost-effectiveness. Clinical trials are funded by the British Ministry of Health but (as in this country) data on total spending in the United Kingdom for research on comparative effectiveness are hard to come by.

Other countries such as Australia, Canada, France, and Germany have similar review processes, though the organizational and financing arrangements vary—and in several cases, the structures have recently been changed.[1] For example, France established a new agency in 2004 to bring together a number of related activities, including the evaluation of drugs, devices, and procedures, publication of clinical guidelines, accreditation of providers, and dissemination of medical information. Germany established a new agency in 2000 that conducts technology assessments and a new Institute for Quality and Efficiency

1. For additional information, see Institute of Medicine, *Learning What Works Best: The Nation's Need for Evidence on Comparative Effectiveness in Health Care* (September 2007), Appendix 2, available at www.iom.edu/ebm-effectiveness.

has been applied to many of to the treatments discussed above. For example, an additional analysis of lung-volume-reduction surgery, which focused on the patients likely to benefit from the surgery, found that it would be cost-effective if its benefits persisted for 10 years but might not be so if those benefits dissipated after three years.[11] (That study did not follow patients for a decade and therefore had to estimate the future benefits.) Similarly, another study examined the cost-effectiveness of more-expensive screening mechanisms for breast cancer and found that it varied substantially with the age of the patient.[12]

11. National Emphysema Treatment Trial Research Group, "Cost Effectiveness of Lung-Volume–Reduction Surgery for Patients with Severe Emphysema," *The New England Journal of Medicine*, vol. 348, no. 21 (May 22, 2003), pp. 2092–2102.

12. Sylvia K. Plevritis and others, "Cost-Effectiveness of Screening BRCA1/2 Mutation Carriers with Breast Magnetic Resonance Imaging," *Journal of the American Medical Association,* vol. 295, no. 20 (May 24/31, 2006), pp. 2374–2384.

Box 1.

Continued

in 2004 that evaluates health care services. Discussions about the use of comparative effectiveness in those countries sometimes focuses on their review processes for prescription drugs, but their efforts generally encompass all forms of acute medical care. (For all the attention they receive, drug costs represent less than 15 percent of health care spending in the United States—so research that focused only on medications would miss the vast majority of services and would not be able to compare drug therapy with surgical procedures or other interventions.)

Although those countries all have government-run health care systems, they have taken different approaches regarding the placement of and funding for their assessment bodies. In the United Kingdom and Australia, the agencies are part of the government's health departments; France and Canada have established independent not-for-profit organizations; and Germany has taken a mixed approach (the Institute for Quality and Efficiency is independent, but the technology assessment agency is an arm of the health ministry). Financing arrangements vary correspondingly: Funding in the United Kingdom and Australia comes from their health departments, whereas Germany's independent institute is funded by a levy on inpatient and outpatient health care services (which are mainly reimbursed by the country's regional health insurance funds), and the French agency gets its funding from a combination of taxes on promotional spending by drug companies, government subsidies, and accreditation fees. Health ministries in Australia, Canada, France, and Germany also help fund clinical trials and other forms of primary research, but total spending related to comparative effectiveness in those countries is also difficult to estimate.

Given the interest that has developed in many countries, it is not surprising that several international organizations have become involved in comparative effectiveness research. The best known may be the Cochrane Collaboration—a nonprofit organization that has a network of volunteers who conduct systematic reviews of treatments. Many of its activities are organized through centers located around the world, including one in the United States. Founded in 1993, the Cochrane Collaboration maintains an accessible database that now contains more than 4,500 reviews; its limited funding comes primarily from subscription fees for its quarterly journal. Any new or expanded U.S. entity that would organize and fund research on comparative effectiveness would probably draw upon Cochrane's findings and the results of research conducted in other countries (to the extent such research was applicable to U.S. patients).

More generally, the relative cost-effectiveness of treatment options is clear when a less expensive treatment yields comparable or superior health gains. In other cases, however, determining whether the additional medical benefits of a more expensive treatment warrant their added costs is complex. Typically, the benefits of different treatments are summarized as an increase in life expectancy or, more commonly, as an increase in quality-adjusted life years (QALYs) to account for effects on morbidity as well as mortality. That calculation reflects estimates of how much people value improving their health or avoiding various side effects, which are combined to create a single metric. By convention, cost-effectiveness analyses report results as the cost per QALY gained, so a lower dollar amount indicates a more cost-effective service. If that metric is used to determine whether specific health procedures are covered by an insurance program, choosing a cost-effectiveness threshold can be a controversial endeavor—but that need not be the manner in which such research is applied.

Research in the Private Sector

In the United States, most of the formal research that is done to examine the effects of drugs or medical devices is

conducted by the manufacturers of those products in the course of their development; as noted, however, it is the exception rather than the rule that those studies directly compare treatments or products.[13] Nevertheless, various other private organizations have also produced assessments and comparisons of some treatments. (Analyses conducted in other countries represent another source of information about treatments; see Box 1 on page 6.)

Several private-sector organizations exist primarily or exclusively to assess medical treatments and technologies. One prominent example is the Technology Evaluation Center that is part of the Blue Cross Blue Shield Association. Its analyses are based on systematic reviews of the available literature and therefore rely on clinical trials or other studies that have already been conducted. (In such reviews, more weight is given to studies that are judged to be of higher methodological quality.) The center produces about 20 to 25 new assessments of drugs, devices, and other technologies each year; the analyses consider clinical effectiveness but generally do not assess cost-effectiveness.

For-profit private-sector firms that specialize in technology assessments represent another source of analysis. Hayes, Inc., is one of the larger firms in the field. Such firms also conduct systematic reviews and evaluate medical and surgical procedures, drugs, and devices in return for a fee or on a subscription basis. Organizations that are similar but operate as nonprofit entities—sometimes affiliated with academic or medical centers—include the ECRI Institute and the Tufts-New England Medical Center's Cost-Effectiveness Analysis Registry (which provides an extensive list of the cost-effectiveness ratios that are available from published studies).

In addition, private health plans—most commonly, larger or more integrated ones—conduct their own reviews of evidence and sometimes undertake new analyses of comparative effectiveness using claims data for their enrollees.[14] Health plans may choose to publicize the results, or they may decide to keep their findings confidential and use them to shape their policies regarding coverage of and payment for the treatments in question. For example, health plans usually have an entity known as a pharmacy and therapeutic committee that considers the evidence regarding the relative effectiveness of different prescription drugs and makes recommendations about which ones should be covered (that is, included on formularies) or given preferred status. An example of a more public and collaborative effort is the HMO Research Network, a consortium of more than a dozen health maintenance organizations from different parts of the country; started in the mid-1990s, it brings together researchers to share findings and, in some cases, uses data from several plans as the basis for analysis.[15]

Notwithstanding those current efforts, the private sector generally will not produce as much research on comparative effectiveness as society would value. The knowledge created by such studies is costly to produce—but once it is produced, it can be disseminated at essentially no additional cost, and charging all users for access to that information is not always feasible. As a result, private insurers and other entities conducting research on comparative effectiveness often stand to capture only a portion of the resulting benefits and therefore do not invest as much in such research as they would if they took into account the benefits to all parties. In health plans that do not have exclusive provider networks, some of the benefits probably "spill over" to other health plans using the same doctors, because physicians tend to use a similar approach to care for all of their patients. Even if organizations could keep their findings confidential, so that they captured all of the benefits, some duplication of effort would probably occur. In such a situation, research constitutes a "public good," and economists have long recognized a role for government to increase the supply of such research toward the socially optimal level.

Another reason for the limited availability of information on comparative effectiveness is that public-sector health

13. In the limited number of instances in which manufacturers sponsor head-to-head trials, the comparisons tend to focus on the relative merits of products used to provide the same basic treatment. For example, a number of industry-sponsored trials have been conducted comparing different brands of coronary stents that are used during an angioplasty.

14. Although the Technology Evaluation Center discussed above is affiliated with the Blue Cross Blue Shield Association (which is an umbrella group that represents the Blue Cross and Blue Shield insurers from each state), that center does not work directly or exclusively for those insurers.

15. Medical specialty societies, such as the American Heart Association, represent another source of analysis of different treatment options—which typically take the form of treatment guidelines for various types of conditions and patients.

insurance programs—which collectively account for about 40 percent of all health care spending—have not sought to make extensive use of it. In particular, the Medicare program has made only limited use of comparative effectiveness data in making decisions about which treatments to cover and how much to pay for them. It stands to reason that the limited demand for such research from such a prominent payer has constrained the supply correspondingly. Conversely, increasing the amount of credible and objective research that was available could facilitate moving Medicare toward what former program administrator Mark McClellan has called a "fee-for-value" system rather than a fee-for-service one. (Options to incorporate research findings into Medicare's coverage and payment policies, along with the issues they raise, are discussed in the final section.)

Past and Current Federal Efforts

In the United States, the federal government has a rather long but somewhat checkered history of involvement in comparative effectiveness research and related efforts. Federal efforts date at least to the late 1970s and the short-lived National Center for Health Care Technology. Established in 1978 as part of the Department of Health, Education, and Welfare, it was given a broad mandate to conduct and promote research on health care technology, and it included an advisory board appointed by the Secretary to assist in setting research priorities. The center sponsored or cosponsored major evaluations of coronary artery bypass graft surgery, dental radiology, and cesarean delivery and made about 75 recommendations to the Medicare program about coverage. The center ceased operations at the end of 1981, however, reflecting changes in priorities for the new Administration and the Congress as well as opposition from some provider and industry groups.[16]

In that same period, the Office of Technology Assessment (OTA) was created as an advisory agency to the Congress, covering a broad set of issues, including health care. Given the agency's focus on evaluating technologies, much of its work would now be called research on comparative effectiveness; over the years, it studied a variety of health care topics, including the costs and benefits of screening tests for several diseases. OTA also produced an extensive review and analysis of the issues involved in and options for improving evidence about the clinical effectiveness and cost-effectiveness of medical treatments.[17] For a variety of reasons, however—having little to do with its health care studies specifically but instead reflecting broader questions about the agency's role—OTA was eliminated in 1995.

More recently, the Agency for Health Care Research and Quality (AHRQ) has been the most prominent federal agency supporting various types of research on the comparative effectiveness of medical treatments. Established in 1989 as the Agency for Health Care Policy and Research, AHRQ is an arm of the Department of Health and Human Services (HHS).[18] It currently has a staff of about 300 and an annual budget of over $300 million, which primarily funds research grants to and contracts with universities and other research organizations covering a wide range of topics in health services.

AHRQ has undertaken a number of initiatives related to comparative effectiveness. One such step—initially taken in collaboration with the American Medical Association and America's Health Insurance Plans, a coalition of insurance companies—has been the creation of a national clearinghouse for treatment guidelines, which are designed to summarize the available medical evidence on the appropriate treatments for various conditions. AHRQ has also endorsed about a dozen evidence-based practice centers around the country. Generally affiliated with a university, those centers analyze and synthesize existing evidence about treatments and technologies. Although many studies sponsored by AHRQ have examined only the relative clinical benefits of different treatments, some have also analyzed their cost-effectiveness. Research on comparative effectiveness has accounted for only a modest portion of AHRQ's budget, though.

As with other agencies examining the effectiveness of medical treatments or evaluating medical technologies, support for AHRQ has varied over time. In the mid-1990s, controversies arose after an agency-sponsored research team concluded that there was insufficient

16. See Seymour Perry, "The Brief Life of the National Center for Health Care Technology," *The New England Journal of Medicine*, vol. 307, no. 17 (October 21, 1982), pp. 1095–1100.

17. See Office of Technology Assessment, *Identifying Health Technologies that Work: Searching for Evidence,* OTA-H-608 (September 1994).

18. Prior to AHRQ's establishment as a separate agency, some of its functions were carried out by the National Center for Health Services Research within HHS.

Table 1.

Requested, Proposed, and Actual Funding for the Agency for Health Care Research and Quality

(Millions of dollars)

	Agency's Request	House Proposal	Senate Proposal	Appropriation
1991	109	88	138	115
1992	122	115	127	120
1993	125	118	130	128
1994	158	148	158	154
1995	171	154	166	162
1996	194	66	127	125
1997	144	125	144	143
1998	149	149	143	147
1999	171	171	171	171
2000	206	175	211	204
2001	250	224	270	270
2002	306	306	291	299
2003	250	0	314	304
2004	279	304	304	304
2005	304	304	319	319
2006	319	319	324	319

Source: Congressional Budget Office based on data from the Department of Health and Human Services, Agency for Health Care Research and Quality.

evidence to support certain spinal surgeries and, on the basis of that work, the agency issued practice guidelines for the treatment of back pain.[19] Strong opposition from back surgeons, along with broader questions about the value of the research that the agency had funded and other factors, led to proposals to eliminate the agency. Ultimately, the agency was retained, but its funding for fiscal year 1996 was reduced from prior levels (see Table 1). Since then, its overall budget has generally been maintained, at least in nominal terms, or increased. Again in 2002, however, the House of Representatives voted to cut off all funding for AHRQ, though in the end the agency received a small increase in its fiscal year 2003 appropriation.

Most recently, section 1013 of the Medicare Modernization Act of 2003 authorized AHRQ to spend up to $50 million in 2004 and additional amounts in future years to conduct and support research with a focus on "outcomes, comparative clinical effectiveness, and appropriateness of health care items and services (including prescription drugs)" for Medicare and Medicaid enrollees. The actual funding appropriated for that initiative has been $15 million per year. Using that funding, AHRQ has established an "Effective Health Care" program consisting of three main functions: reviewing and synthesizing existing evidence (using its evidence-based practice centers); generating new information using a set of approved research centers (such as the HMO Research Network) that have access to data from medical claims and electronic medical records; and publishing findings in formats that are geared to the differing needs of clinicians, patients, and policymakers.

Other federal agencies also engage in various activities related to comparative effectiveness research—efforts that receive less attention than AHRQ's activities but that are probably larger in dollar terms. The Department of Veterans Affairs (VA) has a very substantial research program that reviews evidence from the medical records of its patients, focusing particularly on the clinical effectiveness of treatments. The department also sponsors evidence reviews through a technology assessment program and helps fund clinical trials—including the study comparing

19. For a discussion, see Bradford H. Gray, Michael K. Gusmano, and Sara R. Collins, "AHCPR and the Changing Politics of Health Services Research," *Health Affairs,* Web Exclusive (June 25, 2003), pp. W3-283–W3-307.

stents to drug therapy mentioned above. Indeed, over the past 30 years, some of the most influential clinical trials have been supported by and conducted in the VA health system, including the first major trials that demonstrated the value of bypass surgery over medical therapy for some forms of coronary artery disease as well as head-to-head studies of drugs that treat prostate enlargement. Another source is the National Institutes of Health (NIH), part of HHS, which is the leading federal sponsor of medical research—primarily in the form of clinical trials. Although comparative effectiveness is not a focus of that research, over the years NIH has sponsored a number of trials that compare treatments directly.

The Centers for Medicare and Medicaid Services (CMS) has helped to sponsor a limited amount of research on comparative effectiveness (for example, it covered the medical costs of the study of lung-volume-reduction surgery). When making decisions about what services are covered, however, CMS generally considers only whether devices and procedures are clinically effective. It has sponsored some studies comparing the effectiveness of different treatments but has done so largely to determine whether to establish separate payment rates for similar treatments. For example, CMS is currently cosponsoring a trial with NIH that may eventually compare the effects of daily dialysis for kidney patients with the conventional treatment of dialysis three times per week.[20] If daily dialysis proves more effective for certain patients, CMS could modify its payment policy to cover the additional costs of more frequent treatment for those patients.

Estimating the total amount that is spent in the United States each year on research that compares the effectiveness of medical treatments is difficult. According to one recent analysis, the federal government spent about $1.5 billion in 2005 on all health services research, a broader category that includes some of the work on comparative effectiveness but also encompasses many other types of studies.[21] For example, that total included AHRQ's entire budget of roughly $300 million, whereas the funding devoted to the agency's effective health care program has been $15 million per year. At the same time, that aggregate figure may not include all federal funding for comparative trials or other efforts that are outside the traditional scope of health services research.

Estimating private expenditures is even more challenging. Although drug and device manufacturers spend billions of dollars each year on clinical trials aimed at demonstrating the safety and efficacy of new products, the vast majority of those efforts contribute to comparisons of treatments only indirectly. Data are simply not available on how much is spent by private organizations such a health plans, medical specialty societies, and technology assessment centers to compare medical treatments and procedures. Nevertheless, one recent study estimated that less that $2 billion is spent annually on comparative effectiveness research in this country—and even that rough estimate is subject to uncertainty.[22]

The Consequences of Limited Information

Whether the cause is limited supply or limited demand, the relative scarcity of rigorous data about comparative effectiveness has several effects. First and foremost, it means that decisions about what treatments to use often depend on anecdotal evidence, conjecture, and the experience and judgment of the individual physicians involved. In many cases, that basis may be sufficient; as some observers have noted, it is not necessary to conduct a randomized trial to determine whether to use a parachute when jumping out of an airplane. But if the benefits of a treatment—or risks of not providing it—are less obvious, the lack of hard data makes determining the appropriate choice of treatment difficult. Although estimates vary, some experts believe that less than half of all medical care is based on or supported by adequate evidence about its effectiveness.[23]

Evidence about treatments' effectiveness remains limited even though the number of rigorous studies has grown substantially in recent decades. To illustrate that point, one study simply examined the number of articles that were published each year in peer-reviewed medical journals that reported results from randomized trials.[24]

20. Initially, the study sought to test the feasibility of randomly assigning conventional or daily dialysis to a representative sample of patients.

21. AcademyHealth, *Placement, Coordination, and Funding of Health Services Research within the Federal Government* (September 2005), available at www.academyhealth.org/publications/placementreport.pdf.

22. See Institute of Medicine, *Learning What Works Best: The Nation's Need for Evidence on Comparative Effectiveness in Health Care* (September 2007), p. 8, available at www.iom.edu/ebm-effectiveness.

23. Institute of Medicine, *Learning What Works Best*, p. 2.

24. Mark R. Chassin, "Is Health Care Ready for Six Sigma Quality?"

Between 1966 and 1995, that number increased dramatically, from about 100 to nearly 10,000—with about half of the cumulative total over that period having been produced between 1990 and 1995. But even if the proportion of treatments based on hard evidence has increased as a result, the share remains relatively low. Furthermore, having the evidence base keep pace with the rapid development of new medical treatments and technologies will remain an ongoing challenge.

Another important effect of limited evidence—indeed, an indicator of that scarcity—is that the use of certain treatments and the types of care provided vary widely from one area of the country to another. For example, even after adjusting for differences in the age, sex, and race of Medicare enrollees, researchers at Dartmouth found about a fourfold variation in the share receiving a coronary artery bypass graft; and those differences were not correlated with rates of heart attacks in each region.[25] At the same time, those researchers found that overall surgery rates did not vary systematically; areas with above-average rates for certain procedures had below-average rates for others. Those differences in the use of treatments reflect at least in part the local practice norms that have arisen in each area, and the apparent variation in those norms indicates that there is not sufficient evidence to determine which approach is most appropriate.

Geographic differences in the types of care provided can remain substantial even among patients who turn out to be in their last six months of life. (Examining that period is an analytic approach that can be used in an effort to control for differences in the prevalence and severity of diseases patients have, on the grounds that large groups of patients who are nearing death are likely to have comparable health problems regardless of where they live.) For example, such patients spend nearly 20 days in the hospital over those last six months, on average, in the highest-use areas, compared with an average of about six hospital days in the lowest-use areas. Similarly, the average number of visits to physicians in that period is as high as 50 in some of the highest-use regions and as low as 16 in some of the lowest-use regions.[26]

The observed variations in the use of services correspond to substantial differences in Medicare spending per enrollee in different parts of the country (see Figure 2). In 2003, average costs ranged from about $4,500 in the areas with the lowest spending to nearly $12,000 in the areas with the highest spending (those averages were adjusted to account for differences in the age, sex, and race of Medicare beneficiaries in the various areas). Some of those differences in spending reflect varying rates of illness as well as 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. But according to the Dartmouth researchers, differences in illness rates account for less than 30 percent of the variation in spending among areas, and differences in prices can explain another 10 percent—indicating that more than 60 percent of the variation is due to other factors.[27] Other studies have found that a larger share of the variation in spending can be accounted for by differences in health status and demographic factors, but even so, the remaining differences are substantial in dollar terms.[28]

Of particular relevance to the issue of comparative effectiveness, there is some evidence that the degree of geographic variation in treatment patterns is greater when less consensus exists within the medical community about the best treatment to use. For example, patients who have fractured their hip need to be hospitalized, and there is relatively little variation in admission rates for Medicare beneficiaries with that diagnosis—but for hip replacements and for knee replacements, more discretion is involved and the surgery rates vary more widely

25. See 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. The analysis divided the country into about 300 "hospital referral regions," which reflect where Medicare beneficiaries typically receive hospital care. In 2003, bypass surgery rates ranged from about 2 to 3 per 1,000 Medicare beneficiaries in the lowest-use regions to about 9 to 10 per 1,000 in the highest-use regions. Although higher rates of bypass surgery could reflect higher rates of heart attacks, higher surgery rates could also prevent some heart attacks—a factor that could help explain the lack of correlation between those two measures.

26. Based on data from 2000 to 2003, available from www.dartmouthatlas.org.

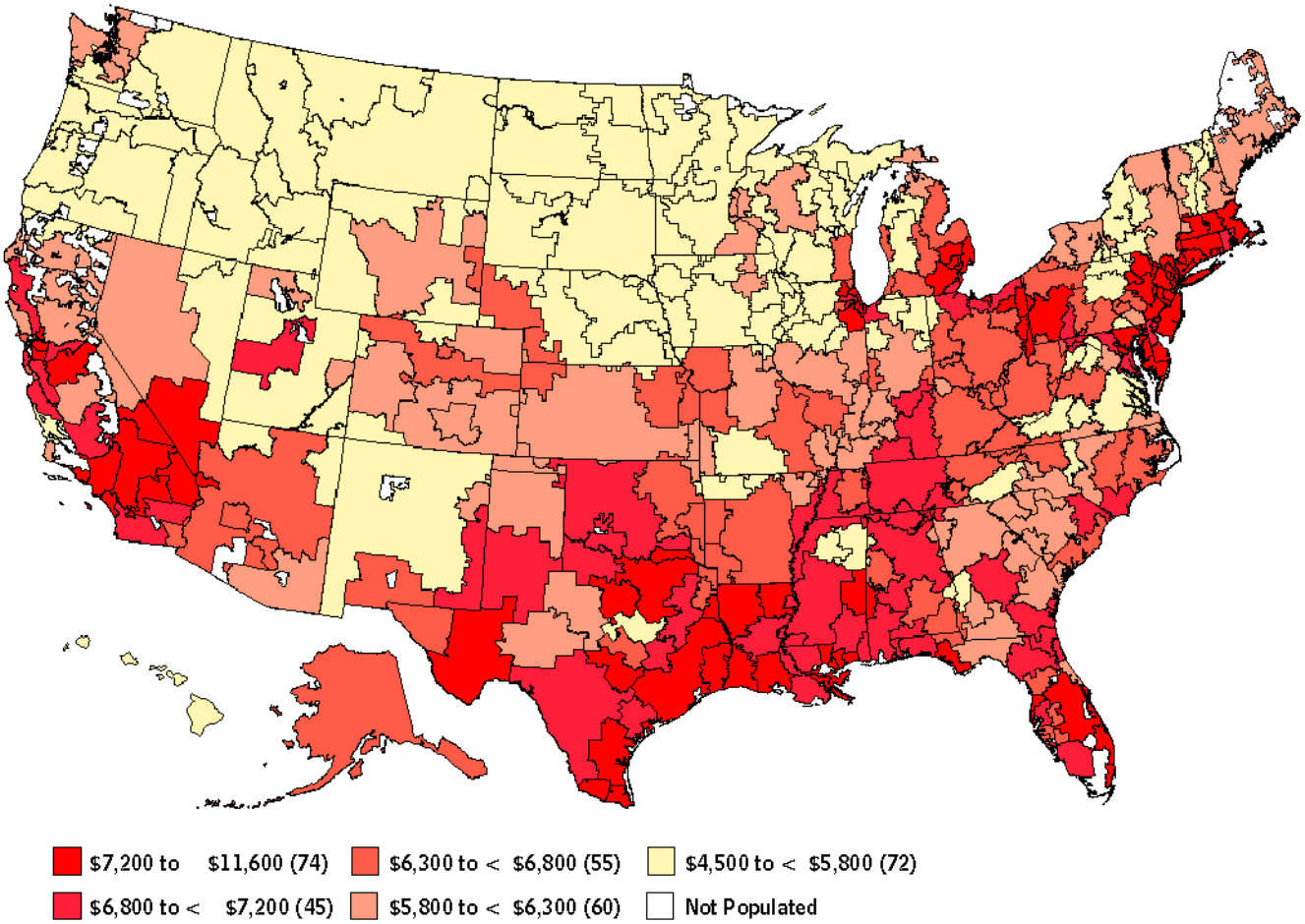
27. John E. Wennberg, Elliot S. Fisher, and Jonathan S. Skinner, "Geography and the Debate Over Medicare Reform"; and The Center for the Evaluative Clinical Sciences, Dartmouth Medical School, *The Dartmouth Atlas of Health Care 1999* (Dartmouth, N.H.: Health Forum, Inc., 1999), pp. 22–23.

28. David Cutler and Louise Sheiner, "The Geography of Medicare," *American Economic Review,* vol. 89, no. 2 (May 1999), pp. 228–233.

Figure 2.

Medicare Spending per Capita in the United States, by Hospital Referral Region, 2003

(Percent)



Source: *The Dartmouth Atlas of Health Care.*

Note: Numbers in parentheses refer to the number of hospital referral regions with per capita spending in each interval.

(see Figure 3). And there appears to be even more variation in the rates of back surgery—a treatment whose benefits have been the subject of substantial questions. Determining what share of any geographic variation in the use of procedures is due to differences in the treatments that doctors recommend and what share is due to differences in underlying illness rates is challenging, however, so the comparison of procedures may be sensitive to the manner in which the differences in illness rates are estimated.[29]

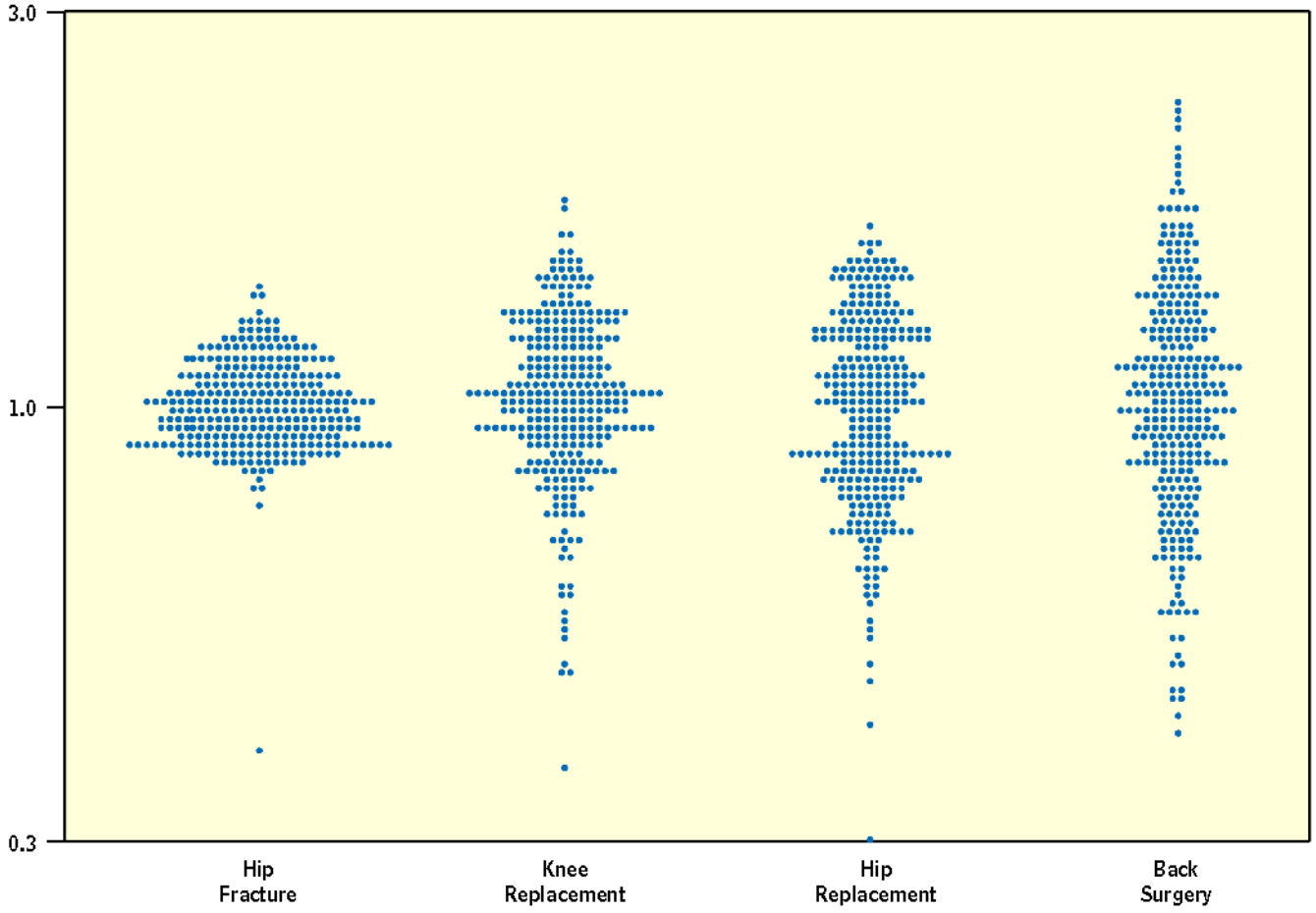
The implications of the observed variations in treatments and spending depend importantly on their relationship to health outcomes. If life expectancy and other measures were better in the areas with higher spending, that result would imply that increased spending in the low-cost areas would yield health benefits. One recent and well-

29. The data used in Figure 3 were adjusted to account for differences in illness rates among areas using data on five conditions, one of which was hip fracture. In the unadjusted data, the variation in knee and hip replacements is somewhat larger than the variation in hip fracture surgery—and variation in back surgery rates is larger still—but the differences are not as substantial. Whether the adjusted results were affected by including hip fracture rates both as an adjustment factor and in the comparison of procedures is not clear. Whether the prevalence of other diseases is correlated with the prevalence of those five conditions is also uncertain.

Figure 3.

Rates of Four Orthopedic Procedures Among Medicare Enrollees, 2002 and 2003

(Standardized discharge ratio, log scale)



Source: Dartmouth Atlas Project, *The Dartmouth Atlas of Health Care*.

Notes: In the figure, each point represents a hospital referral region; the country was divided into about 300 such regions on the basis of where Medicare enrollees typically receive their hospital care.

The points indicate how the rate at which the procedure is performed (per 1,000 Medicare enrollees) in each referral region compares with the national average rate (which has been normalized to 1.0). Differences in procedure rates were adjusted to account for differences among regions in the age, sex, and race of enrollees and for measures of illness rates.

designed study examined differences in hospital spending in Florida and found that areas with higher spending had lower mortality rates among Medicare patients who were treated in the emergency room for a heart attack.[30] Using data on Medicare enrollees nationwide, however, another study found that higher-spending regions did not, on average, have lower mortality rates than the lower-spending regions, even after adjustments to control for differing illness rates among patients and regions.[31] That study also found that higher spending did not slow the rate at which the elderly developed functional limitations (reflecting their ability to take care of themselves). Although more research is needed about the impact that differences in spending have on patients' morbidity and quality of life, perhaps using more-extensive measures of health outcomes, those findings suggest that spending in the high-cost areas could be reduced without adverse effects on the overall health of residents in those areas.

How much could spending be reduced? Some estimates of the potential savings from reducing the variations in treatments are quite large, although questions remain about what mechanisms could achieve those savings and what the effects on health would be. The Dartmouth researchers have suggested that Medicare spending—and perhaps all health spending in the country—could be cut by about 30 percent if the more conservative practice styles used in the lowest-spending one-fifth of the country could be adopted nationwide.[32] While they note the need for more research about the specific steps needed to reduce spending levels without harming health, their analysis indicates that the added spending is not contributing to better health outcomes. Other studies suggest that overall health might not suffer in the process of changing practice patterns but that patients who would benefit most from more-expensive treatments might be made worse off as a result, while patients who would do better with treatments that were less expensive would gain.[33]

Other studies of geographic variation indicate that there may be room to reduce spending without harming health in both high-use and low-use areas of the country. One older study, for example, had independent panels of doctors conduct after-the-fact reviews of the medical charts of Medicare enrollees who had had certain surgeries.[34] In areas with high use of the procedures, the study found that the share of surgeries that was clinically appropriate ranged from about 35 percent to about 70 percent; the remainder were either clinically inappropriate or of equivocal value. In low-use areas, the share considered appropriate ranged from about 40 percent to about 80 percent. In other words, the share of procedures deemed appropriate was slightly higher in the low-use areas, but that share was well below 100 percent in both high-use and low-use areas.

Options for Organizing and Funding New Federal Research Efforts

The approach that is taken for organizing and funding any increased federal efforts to support research on comparative effectiveness could play an important role in determining their impact. Some approaches would seek to insulate those efforts from political pressure by setting up an organization at "arm's length" from the government and by providing a dedicated source of financing. Many of the options that have been proposed seek to coordinate and centralize existing activities through one entity—which would tend to give any conclusions it reached more weight—but developing several competing sources of information about comparative effectiveness could also have value.

30. Joseph J. Doyle, Jr., "Returns to Local-Area Health Care Spending: Using Health Shocks to Patients Far From Home," NBER Working Paper 13301 (National Bureau of Economic Research, August 2007).

31. Elliott S. Fisher and others, "The Implications of Regional Variations in Medicare Spending, Part 2: Health Outcomes and Satisfaction with Care," *Annals of Internal Medicine*, vol. 138, no. 4 (February 18, 2003), pp. 288–298.

32. Elliott Fisher, "More Care is Not Better Care," *Expert Voices*, Issue 7 (National Institute for Health Care Management, January 2005), available at www.nihcm.org/publications/expert_voices.

33. Amitabh Chandra and Douglas O. Staiger, "Productivity Spillovers in Health Care: Evidence from the Treatment of Heart Attacks," *Journal of Political Economy,* vol. 115, no. 1 (February 2007), pp. 103–140.

34. Mark R. Chassin and others, "Does Inappropriate Use Explain Geographic Variations in the Use of Health Care Services? A Study of Three Procedures," *Journal of the American Medical Association,* vol. 258, no. 18 (November 13, 1987), pp. 2533–2537. The procedures studied were coronary angiography (which generally involves inserting a tube and special dyes into the heart to see how well blood flows through it), carotid endarterectomy (in which plaque is removed from the main artery that goes to the brain), and gastrointestinal endoscopy (in which a flexible tube with a small camera mounted on it is inserted into the intestines).

Specific options that have been put forward for organizing federal research on comparative effectiveness include the following (each of which could have many variants):[35]

- Expanding the role of an existing agency that already conducts or oversees research on health services generally—and comparative effectiveness specifically—such as AHRQ or NIH.
- Creating or "spinning off" a new agency, either within the Department of Health and Human Services or as an independent body that is part of either the executive or the legislative branch. The Federal Trade Commission and the Medicare Payment Advisory Commission (MedPAC) are potential models for such an option.
- Augmenting an existing quasi-governmental organization, such as the Institute of Medicine or the National Research Council. Such entities are often Congressionally chartered, but they are not subject to regular governmental oversight.[36] Even so, the Institute of Medicine receives most of its funding from government agencies, which is provided to finance specific studies that have been requested.
- Establishing a new public–private partnership to oversee and direct research. That option could be structured in various ways, but one such approach would be to set up a federally funded research and development center (FFRDC). FFRDCs are not-for-profit organizations that can accept some private payments but that get most of their funding from a federal agency that provides oversight and monitoring.

Regardless of the type of organization, several potential mechanisms (either individually or in combination) could be used to fund research on comparative effectiveness. Federal spending could be authorized and appropriated annually, as with other discretionary programs. Alternatively, funding could be drawn from Medicare's Hospital Insurance trust fund (which is financed primarily by payroll taxes) or specified as a percentage of mandatory federal outlays on health insurance programs.[37] Instead of or in addition to using existing sources of revenues, another option would be to require direct contributions from the health sector. For example, a new tax on health insurance premiums or other payments within the health sector could be established, with the resulting revenues dedicated to research on comparative effectiveness.

Trade-offs might arise between an entity's independence, credibility with the medical profession, and ability to reach controversial conclusions, on the one hand, and its accountability and responsiveness to policymakers and to other interested parties, on the other. For example, funding through appropriations would allow lawmakers to assess the new entity's contributions and accomplishments and to balance spending on those efforts against other federal priorities on an annual basis. But some observers have raised concerns that relying on annual appropriations would leave a new entity vulnerable to outside pressure and thus reluctant to undertake controversial studies or to reach conclusions that might generate opposition from affected groups. Indeed, the elimination of agencies engaged in such research that were funded by annual appropriations—or in the case of AHRQ, the occasional threat of elimination or substantial cuts in funding—may suggest the need for a different arrangement.

Alternatively, housing the new activities in an organization that was separate from the federal government and establishing automatic or dedicated funding mechanisms would give a new entity greater autonomy and potentially more influence on doctors and other health professionals. To be sure, lawmakers could change any funding formula that had been established—as is done frequently in Medicare—mitigating the degree to which the entity would lack oversight. Even with automatic funding, policymakers would want to periodically review the activities they

35. For a discussion of this issue, see Gail R. Wilnesky, "Developing a Center for Comparative Effectiveness Information," *Health Affairs,* Web Exclusive (November 7, 2006), pp. w572–w585.

36. Congressional Research Service, *The Quasi Government: Hybrid Organizations with Both Government and Private Sector Legal Characteristics,* RL30533 (updated February 13, 2007).

37. Current funding for AHRQ resembles a dedicated financing source in that it comes entirely from funds that are designated under the Public Health Service Act as available for evaluation activities. The total amount of funds available for such activities had been limited to 1 percent of certain expenditures (primarily those for research by NIH), but in recent years, that limit has been set at about 2 percent. As a practical matter, however, the agency's funding is like other discretionary appropriations. In previous years, some funding for AHRQ (and its predecessor agencies) came from regular appropriations, and a few million dollars was transferred from Medicare's trust funds.

were funding either to consider changes in the levels of spending or to adjust any funding formula to keep dedicated resources in line with spending trends—which could also provide a vehicle for pressure from interest groups. Nevertheless, automatic or dedicated funding mechanisms would tend to limit the influence of political pressure to some extent. But such mechanisms also would raise questions about how the entity set its priorities and allocated resources—and how it would be held accountable for those decisions. A nongovernmental organization might be able to act more quickly than a federal agency, but that speed could come at the expense of transparency.

Under any option, an advisory board (or governing council) could be established to serve several functions: providing guidance to the entity and establishing priorities for its research projects; creating an independent process for reviewing and possibly approving the findings that resulted from that research; and serving as a channel for interested parties to participate. For example, the board could include representatives of major federal health programs, private insurers, health care providers, advocacy groups for patients, and drug and device makers—as well as members of the general public and disinterested policy experts. Alternatively or in addition to including various stakeholders, a regular process could be established for getting input from interested parties. An example of that type of structure is the U.S. Preventive Services Task Force (see Box 2).

In designing such an oversight group, a number of issues would arise. The types of participants on any board and the manner in which members were chosen and replaced would have to be determined carefully to avoid giving one perspective undue influence. Similarly, conflict-of-interest rules governing the entity's staff would probably be needed. Trade-offs could exist between the extent to which many views and interests were represented and the ability of the council or board to make timely decisions or to reach consensus on contentious issues. Whether any oversight group was involved in reviewing or approving the results of research projects or focused instead on which projects to initiate and what those reviews entailed would also affect the entity's staffing requirements and the types of expertise that board members needed.

Another organizational issue is whether to establish a single or highly centralized entity or, instead, to design a more loosely coordinated system encompassing several distinct centers to produce independent analyses. Many of the options that have been proposed seek to centralize research activities through one entity—partly to address concerns about the lack of coordination among current U.S. efforts. An advantage of that centralized approach is that it would tend to give more weight to any conclusions reached. At the same time, that potential for having a greater impact could also lead the organization to adopt findings that were watered down to reach consensus; even if the entity did not have a formal approval process and instead simply released any results of approved projects, a single agency might be more reluctant to pursue research into more contentious questions. A decentralized approach could give individual research centers more latitude and encourage more competing perspectives to emerge. However, a more pluralistic approach could also involve some redundant efforts and, if it yielded any conflicting findings, would leave users with the task of reconciling the results.

An additional consideration—particularly if a new entity was created—would involve start-up costs and other implementation challenges. If funds were directed through an existing federal agency, some ongoing costs for additional staffing would be incurred, but the basic support infrastructure would largely exist already. By contrast, establishing a new agency or public–private partnership could require a greater effort before research could begin. At the same time, a quasi-governmental organization or public–private partnership could have more flexibility to develop and maintain its staff than a new or existing federal agency would have. Creating a new source of revenues (such as a tax on health insurance premiums) to help fund research on comparative effectiveness would also involve time and administrative costs.

Among existing organizations, their relative strengths and weaknesses could affect which one was best suited for new research efforts. NIH has extensive experience overseeing clinical trials but may not see research on comparative effectiveness as central to its mission of expanding the frontiers of biological and medical knowledge. AHRQ has substantial expertise in many areas of comparative effectiveness but has limited experience managing trials, and some observers have raised concerns about the impact that significantly expanded research about comparative effectiveness might have on that agency's other research endeavors. For its part, the Institute of Medicine is widely respected but does not have an extensive organizational capacity to conduct or oversee primary research,

Box 2.

The U.S. Preventive Services Task Force

The U.S. Preventive Services Task Force was established in 1984 by the Department of Health and Human Services to produce recommendations about which preventive health care services should be routinely provided to individuals who do not have any symptoms of a given disease. Such services include immunizations, tests to screen for the presence of diseases, and behavioral counseling (such as programs that encourage smokers to quit).[1]

The size and composition of the task force has varied over time, ranging from 10 to 20 members; the members are not federal employees but have generally been practicing clinicians. The task force's work is currently supported by the Agency for Health Care Research and Quality (AHRQ), with an annual budget of about $3 million. As a rule, the task force does not fund studies that evaluate preventive services but instead relies on existing evidence. Two research centers that AHRQ has designated generate summaries of that evidence—which are similar to but perhaps not as rigorous as systematic evidence reviews. Given the available time and resources, the task force has not sought to review all preventive services but instead has assigned priority to services that address significant health problems, that are likely to have new evidence available, or that have generated controversy about their use.

In developing its recommendations, the task force considers both the strength of the evidence and the magnitude of the expected benefits and risks. Risks can include adverse reactions to vaccines, false-positive test results that lead to unnecessary or even harmful follow-up care, and complications from invasive test procedures—which can have substantial aggregate effects even if their probabilities are low, because preventive services may be provided to very large numbers of people. The task force's recommendations cover which types of asymptomatic individuals should receive the services, taking into account how the risk of contracting a condition or disease varies by age, sex, and other factors.

1. For a general discussion, see Eileen Salinsky, *Clinical Preventive Services: When Is the Juice Worth the Squeeze?* Issue Brief No. 806 (Washington, D.C.: National Health Policy Forum, August 24, 2005).

and some observers believe its consensus-building process could make timely action difficult.

Among the options for a new entity, establishing an FFRDC has generated some interest, partly on the grounds that it would be somewhat insulated from political pressure. But most of an FFRDC's funding would have to come from a federal agency, so it is not clear why its activities (most of which, presumably, would also be contracted out to private researchers) would be subject to less pressure than the activities of an agency receiving direct funding. The argument is sometimes made that private contributions would make private payers more likely to accept and use the results of the research. If such contributions were voluntary, however, the incentives to make them would be modest because the benefits of the research would accrue to many parties. If such contributions were instead required, then the arrangement would be essentially equivalent to having the government collect the money and appropriate the funds via a federal agency.

More generally, competing perspectives exist about how the relative roles of public and private payers in funding research on comparative effectiveness would affect perceptions about the results of that research. In some quarters, the findings of research funded by the government are seen as reflecting political pressure, perhaps to accommodate the views of interest groups or to support budgetary objectives. Those concerns could be attenuated to some degree if the agency conducting the research was not also a payer for health care, such as CMS. At the same time, other observers have raised concerns about privately sponsored research, which is also seen as advancing cost-cutting objectives (if sponsored by insurers) or as promot-

Box 2.

Continued

The task force has presented its recommendations in a periodic series of reports, the most recent of which covers about 60 specific services. Those services are now given a letter grade, as follows:

- A, for services that are strongly recommended on the basis of solid evidence that the benefits of improved outcomes outweigh the risks of harm;
- B, for services that are recommended on the basis of reasonable evidence of net benefits;
- C, for services with no recommendation because the balance of benefits and risks is too close;
- D, for services that should not be routinely provided because the evidence indicates the services are ineffective or that the risks outweigh the benefits; and
- I, for services that do not have sufficient evidence on which to base a recommendation.

Initially, when formulating recommendations, the task force did not take into account the costs of providing preventive services or their cost-effectiveness.[2] According to one recent summary, however, the task force now "considers the total economic costs that result from providing a preventive service, both to individuals and to society, in making recommendations, but costs are not the first priority."[3] Although some immunizations against a disease have been shown to reduce total spending on health care, many other preventive services appear to increase spending on net—either because of the costs of providing those services to large segments of the population (only some of whom will be found to have the disease) or because the overall effects on treatment costs are modest. Analyses of cost-effectiveness would shed light on how the health benefits of preventive services compared with those increases in spending.

2. See Somnath Saha and others, "The Art and Science of Incorporating Cost-Effectiveness in Evidence-Based Recommendations for Clinical Preventive Services," *American Journal of Preventive Medicine*, vol. 20, no. 3 (April 2001), pp. 36–43.

3. Russell P. Harris and others, "Current Methods of the U.S. Preventive Services Task Force," *American Journal of Preventive Medicine*, vol. 20, no. 3 (April 2001), pp. 21–35.

ing the interests of drug and device manufacturers and of providers of health services.

Options for Comparing the Effectiveness of Treatments

The appropriate organizational form for any new or expanded federal entity, along with the mechanism and level of funding, may depend in large part on what activities it would carry out. For example, analyzing existing data would require a different set of skills, and would cost less, than overseeing new clinical trials that compared different treatments. In addition to setting priorities among the various methods of research, a new or expanded entity would have to define the scope of its analyses—both the types of comparisons it would commission and the questions that analyses would address. In particular, would the organization focus only on trying to determine which treatments conferred the greatest medical benefits, or would it also assess which treatments were most cost-effective? Whatever approach was taken, the manner in which the results were communicated to doctors, patients, and health insurers could play an important role in determining the impact on medical practice.

Methods of Research

Federal efforts to assess different treatment options could be pursued in a variety of ways. Options range from synthesizing existing research—a process known as a systematic review—to conducting new studies using data that are already available to funding new head-to-head clinical trials. Although those options are not mutually exclu-

sive—indeed, they could all be pursued at the same time—each one presents certain challenges, with potential trade-offs arising between the costs of the activities and the value of the information they provide.

Systematic Reviews of Existing Research. The approach that would probably be easiest to implement would be to review and summarize the results of existing studies in a systematic and rigorous way. For example, even though existing studies may only compare a single treatment to a placebo, the results of several studies of individual therapies could in some cases be combined to measure those treatments against one another. That effort could also critically assess the strengths and weaknesses of the available evidence and seek to reconcile conflicting findings or determine what the preponderance of the evidence indicated. Such reviews would be comparable to some of the work that AHRQ is already undertaking and to some current efforts based at universities or other public and private research centers such as ECRI and Hayes, Inc. One advantage of this approach is its relatively low expense; a single systematic review might cost a few hundred thousand dollars.

Because the evidence base for comparing treatment regimens is itself limited, however, how much additional insight can be gleaned from systematic reviews of existing research is not clear. Data from clinical trials that had already been conducted would naturally be the focus of any systematic review, because trials can provide the clearest evidence about a treatment's effects, but such studies also have limitations. Some analyses have indicated that clinical trials sponsored by interested parties—which is often the only source of such data—are more likely than independent studies to find favorable results.[38]

Another potential limitation is that existing information may not be sufficient to reach definitive conclusions. Studies may be difficult to compare or reconcile, either because they use different methodologies or analyze different populations of patients, or simply because they yield conflicting findings. For example, a number of independent studies have examined different screening techniques for colorectal cancer, each of which provides an estimate of the cost per enrollee for each increase in QALYs. But according to a recent review of those studies, the results varied to such an extent that reaching a definitive conclusion about which technique was most effective or most cost-effective was difficult (see Table 2).[39]

Available studies of treatments may have even more limitations than studies of screening tests, because trials of treatments for particular diseases frequently exclude patients with other health problems, elderly enrollees, or other populations that may be of considerable interest in gauging comparative effectiveness; as a result, determining how broadly the results apply or whether they will hold for other groups of patients is hard to do. The fundamental issue is that, no matter how rigorously a systematic review is conducted, its contribution is by definition constrained by the extent and quality of the underlying evidence.

A recent systematic review of drug treatments for one form of diabetes that was sponsored by AHRQ illustrates both the strengths and weaknesses of such research.[40] The review covered a large body of literature, consisting of over 200 reports, and it was able to reach a relatively clear conclusion: Older drugs were found to be at least as effective as newer drugs in controlling patients' blood sugar and cholesterol levels. Most of the studies that were reviewed had relatively short durations, however—two years or less—so they were not able to address the impact on mortality or other effects of diabetes on morbidity (which can take a long time to materialize). The studies also tended to focus on nonelderly white patients, so they could not address the effectiveness for other populations; indeed, the review recommended that several clinical trials be conducted to fill in those gaps. Moreover, study subjects typically had no other significant health problems, whereas most patients with diabetes also have other

38. See Justin E. Bekelman, Yan Li, and Cary P. Gross, "Scope and Impact of Financial Conflicts of Interest in Biomedical Research: A Systematic Review," *Journal of the American Medical Association,* vol. 289, no. 4 (January 22/29, 2003), pp. 454–465; Stephan Heres and others, "Why Olanzapine Beats Risperidone, Risperidone Beats Quetiapine, and Quetiapine Beats Olanzapine: An Exploratory Analysis of Head-to-Head Comparison Studies of Second-Generation Antipsychotics," *American Journal of Psychiatry,* vol. 163, no. 2 (February 2006) pp. 185–194; and Jeffrey Peppercorn and others, "Association Between Pharmaceutical Involvement and Outcomes in Breast Cancer Clinical Trials," *Cancer,* vol. 109, no. 7 (April 2007), pp. 1239–1246.

39. Medicare Payment Advisory Commission, *Report to the Congress: Increasing the Value of Medicare* (June 2006), pp. 232–233.

40. See Shari Bolen and others, "Systematic Review: Comparative-Effectiveness and Safety of Oral Medications for Type 2 Diabetes Mellitus," *Annals of Internal Medicine*, vol. 146, no. 6. (September 18, 2007), pp. 386–399.

Table 2.

Cost-Effectiveness of Different Screening Methods for Colorectal Cancer

(Dollars)

Screening Method	Lowest	Highest
Colonoscopy		
Every 5 Years	17,316	36,612
Every 10 Years	10,633	26,693
Fecal Occult Blood Testing		
Annually	4,643	25,860
Every 3 Years	2,942	10,861
Sigmoidoscopy		
Annually	1,391 [a]	1,391 [a]
Every 3 Years	16,318	20,727
Every 5 Years	14,384 [b]	42,310

Source: Congressional Budget Office based on Medicare Payment Advisory Commission, *Review and Analysis of Cost-Effectiveness Analyses for Two Medicare-Covered Services* (prepared by the Institute for Clinical Research and Health Policy Studies, New England Medical Center, June 2006), available at www.medpac.org.

Note: The cost-effectiveness ratio is the estimated cost per one-year increase in quality-adjusted years of life expectancy, in comparison with the result of no screening.

a. Only one study was available for analysis.

b. One study found that screening every five years yielded lower costs and better health outcomes than no screening.

diseases, limiting the potential usefulness of the findings. In addition, the implication of the review—that older drugs for diabetes should be tried first—was already the protocol recommended by the American Diabetes Association. Thus, although the review was relatively inexpensive to conduct and may well have been worth its costs, its contribution was also limited.

In some cases, the existing evidence may permit more clear-cut determinations, but many systematic reviews are inconclusive—so views differ about their overall contribution. Britain's National Institute for Clinical Excellence (NICE) relies solely on systematic reviews of available studies. It has nonetheless been able to analyze many different treatments on the basis of their cost-effectiveness and to develop an extensive set of clinical guidelines and recommendations about using medical technologies.[41] Whether that record indicates the greater strength of the evidence on the reviewed treatments or a greater willingness on NICE's part to draw conclusions from that evidence is not clear. Typically, though, systematic reviews find that the available evidence is not adequate to address many important questions, so the primary value of such reviews may lie in clearly identifying the gaps in knowledge that should be the subject of future research.

Analyses of Claims Records. A somewhat more challenging approach than reviewing existing studies would be to fund new analyses comparing medical treatments using existing sources of data, such as health insurance claims records. An advantage of that approach is that it could provide new information to help resolve uncertainties about treatments at relatively low cost—using data on patients that had already been treated.

A central difficulty in such studies, however, is accounting for the differences in patients' health status that play a role in determining which treatment they get—which can make simple comparisons misleading. Insurance claims typically do not include any information about health status. Yet patients with more severe heart disease, for example, are more likely to receive invasive and expensive surgical procedures such as an angioplasty or a bypass operation. The greater severity of their condition may also make them more likely to have a subsequent heart attack and more likely to die. As a result, a comparison with patients receiving less aggressive treatments—who are probably not as sick, on average, to begin with—could understate the benefits of more aggressive treatments. In other settings, patients receiving more aggressive treatments may be healthier, so even well-designed observational studies can generate misleading findings regarding the benefits of those treatments. Studies of

41. To estimate cost-effectiveness, NICE generally combines the results of such reviews with its own models of the impact of different treatment options on the use of health services and health care spending.

cancer care, for instance, may be particularly susceptible to such confounding effects.[42]

To address such problems, researchers might be able to analyze geographic differences in treatment patterns to compare the impact of different treatments on comparable types of patients. For example, one study using claims data for Medicare enrollees sought to exploit the fact that patients living farther away from hospitals that treat a high volume of heart attacks were less likely to receive an intensive treatment for that condition (such as an angioplasty or a bypass operation).[43] At the same time, patients living farther from such hospitals appeared to be about as healthy as patients living nearby—so grouping patients by distance could address the analytic problem noted above whereby sicker patients get more intensive treatments. The study found that patients living closer to high-volume cardiac hospitals had slightly lower mortality rates, but the difference arose on the first day of admission and thus did not seem related to which procedure (if any) they ultimately received.

That example illustrates some of the challenges involved in using observational studies that are based on claims data. A key one is finding a factor that can be used in the analysis that is correlated with the treatment that patients receive but is not correlated with their underlying health status. (Economists refer to such factors as "instrumental" variables.) Even if such factors are available, proving that other confounding effects did not influence the results can be difficult. Whether that study's findings were persuasive enough to reduce the use of intensive medical treatments for heart attack patients, for example, is not clear. And the case of heart attacks—where urgent hospital care can mean the difference between life and death but where a range of potentially effective treatments are available—may be easier to analyze than other conditions.

Other issues surround the claims data themselves. First, maintaining the privacy of the patients whose records were being examined would be an important matter but could also present a barrier to conducting such studies. For statistical reasons, extracting meaningful results could require a large volume of claims data (as was the case in the study of heart attack treatments). Second, the quality of the study that could be conducted would depend on the level of detail that the data provided. Comparisons of the effects of treatments on mortality rates would be easier to generate because that information is relatively easy to obtain. Effects on morbidity or on the extent to which symptoms are relieved, however, might be more difficult to ascertain—depending on whether the relevant data were readily available. In addition, private health plans might have difficulty in conducting longer-term comparative effectiveness studies using claims data on their enrollees given the turnover in insurance coverage; if patients who changed plans were different from those who remained, statistical obstacles might undermine the comparison.

The expanded use of electronic health records could facilitate more-sophisticated analyses, if the issues of access and privacy could be addressed. In particular, those records could provide more comprehensive information both about the health histories of different patients and about their health outcomes. That additional information would make controlling for differences among patients receiving different treatments easier and would allow studies to address a broader set of outcomes than mortality. Some work of that nature is currently being conducted through the HMO Research Network and through a broader network of centers that have access to electronic databases that AHRQ established in 2005.[44] One challenge, however, is that the electronic records of different health plans are not always compatible, making aggregating data difficult.

42. See Alan M. Garber, "Cost-Effectiveness and Evidence Evaluation as Criteria for Coverage Policy," *Health Affairs,* Web Exclusive (May 19, 2004), pp. W4-284–W4-296. Some analyses have found similar results for observational studies and randomized controlled trials of the same treatment, but others have found important differences in the magnitude of the treatments' effects, particularly when the nonrandomized studies were done retrospectively. See Kjell Benson and Arthur J. Hartz, "A Comparison of Observational Studies and Randomized, Controlled Trials," *The New England Journal of Medicine,* vol. 342, no. 25 (June 22, 2000), pp. 1878–1886; and John P. A. Ioannidis and others, "Comparison of Evidence of Treatment Effects in Randomized and Nonrandomized Studies," *Journal of the American Medical Association,* vol. 286, no. 7 (August 15, 2001), pp. 821–830.

43. Mark McClellan, Barbara J. McNeil, and Joseph P. Newhouse, "Does More Intensive Treatment of Acute Myocardial Infarction in the Elderly Reduce Mortality? Analysis Using Instrumental Variables," *Journal of the American Medical Association*, vol. 272, no. 11 (September 21, 1994), pp. 859–866.

44. For a discussion of those efforts, see Lynn M. Etheredge, "A Rapid-Learning Health System," *Health Affairs,* Web Exclusive (January 26, 2007), pp. w107–w118; and related articles contained in that supplemental issue.

Medical Registries. Another option that could supplement or help improve analyses of claims data would be to establish medical registries, which generally track patients who have a particular disease or who have received a specific treatment. Registries collect additional information that is typically not contained in claims records, such as measures of health status or test results. In the United States, a number of registries—established or managed by various entities, including medical specialty societies and product manufacturers—have been used to help determine the clinical effectiveness or cost-effectiveness of various products and services.[45] Some health plans establish registries of their enrollees, although a centrally managed registry would have the advantage of being able to track patients if they moved or changed health plans.

Data from medical registries could help improve claims-based analyses both by allowing a broader set of outcomes to be measured and by providing information to control for differences among patients getting different treatments, including the severity of their illness. But a number of challenges and trade-offs would exist. One issue would be how to recruit patients and their providers to participate in and provide information to the registries and to retain them over time. Voluntary participation might be easy to implement but could introduce bias into analyses if patients choosing to participate differed in important ways from patients who had opted out. Some form of mandatory participation could avoid that problem but might raise objections from participants. Registries focused on specific treatments could also be subject to bias if those patients differed systematically from patients who did not receive those treatments—a problem that could be addressed by including a comparison group in the registries. Another trade-off concerns the data elements to collect; a more extensive list would permit richer analyses but would raise the burden of participation. More-extensive registries and registries involving more patients would also be more expensive to operate, although the annual costs of maintaining a typical registry are probably on the order of several million dollars.

The establishment of registries could affect medical practice in various ways. For example, CMS recently instituted a policy of "coverage with evidence development" for Medicare, to address treatments with potentially promising but uncertain medical benefits. Under that policy, Medicare now covers the costs of implantable cardioverter-defibrillators for a broader set of heart conditions than had previously been eligible—but only if those new patients are included in a registry that is supposed to track their progress.[46] If CMS would otherwise have decided not to cover that treatment for those patients, then the new policy means an increase in spending in the near term, but it also allows broader access to that technology in order to help generate the kind of evidence needed to reach a conclusion about its value. The registry may also help ensure, through its documentation requirements, that all patients meet the medical criteria required for Medicare coverage. Another example comes from Sweden, where a registry of patients undergoing hip replacement surgery has been used to provide periodic feedback to doctors about their surgical techniques and to track which specific models of artificial hip have the lowest rates of complications. That effort is credited with reducing health costs by avoiding repeat operations to fix faulty or poorly installed hips.[47]

Randomized Controlled Trials. The method of research that would probably yield the most-definitive results involves randomized controlled trials to compare treatments head to head, but that approach would also be the most expensive and would take the longest to conduct. The main advantage of random assignment is that it usually ensures that any differences in outcomes reflect true differences among treatments and not confounding differences among patients (such as their health status or other factors that are more difficult to observe). But detecting differences that are statistically significant—that is, unlikely to have arisen simply by chance—can require a substantial number of patients to participate, and in some cases, they must be followed for several years. Total costs for conducting an extensive trial can exceed

45. For more information, see Richard E. Gliklich and Nancy A. Dreyer, eds., *Registries for Evaluating Patient Outcomes: A User's Guide*, AHRQ Publication No. 07-EHC001-1 (Rockville, Md.: Agency for Healthcare Research and Quality, April 2007).

46. See Sean R. Tunis and Steven D. Pearson, "Coverage for Promising Technologies: Medicare's 'Coverage with Evidence Development,'" *Health Affairs,* vol. 25, no. 5 (September/October 2006), pp. 1218–1230. An implantable cardioverter-defibrillator (ICD) is a device designed to quickly detect a life-threatening rapid heartbeat and to deliver an electric shock that converts the rhythm back to normal. Apparently, CMS has not yet implemented the longitudinal registry for ICD patients.

47. See Henrik Malchau and others, "The Swedish Total Hip Replacement Register," *The Journal of Bone and Joint Surgery,* vol. 84, no. 11 (November 2002), pp. S2–S20.

$100 million over the course of the study, although many trials are less expensive, and some may cost only a few million dollars. (One factor affecting the costs of funding a trial is whether the health care services that the participants receive will be paid for by a third party, such as Medicare.)

Although the number of studies reporting results from randomized controlled trials has increased sharply, a number of questions have also been raised about the findings that can be derived from the existing stock of trial results. Many trials are aimed at demonstrating efficacy rather than effectiveness—the distinction being that efficacy reflects optimal conditions, whereas effectiveness requires a demonstration in real-world medical settings.[48] Partly as a result, patients with other health problems or groups such as the elderly are often excluded from trials. Further, many trials focus on demonstrating efficacy for a narrowly defined set of patients, so the results may not be generalizable; and combining studies in order to compare multiple treatment options may offer its own difficulties because of the differences among the patients studied. Finally, questions about the objectivity of industry-sponsored trials have also been raised.

To address those problems, some observers have recommended a greater emphasis on "practical" clinical trials.[49] The two key features of such trials are that they compare treatment choices that clinicians face and include a wide variety of study participants drawn from a range of practice settings. Traditionally structured trials, such as those typically sponsored by NIH, can involve a relatively large number of participants and relatively long periods of follow-up observation and analysis. As a result, they may be relatively costly to implement. Trials that are simpler and less expensive and that take less time to carry out could provide a greater "bang for the buck," but at some risk of reduced accuracy.[50]

Because their results can be persuasive, well-structured trials can have a noticeable effect on the use of treatments. For example, according to one report, the findings of the trial (discussed above) comparing the use of angioplasty and a metal stent with nonsurgical management of patients with stable coronary artery disease—which found minimal advantages of stenting—may have reduced the use of that procedure.[51] Determining the precise effect of the trial is difficult, however, in part because the downward trend in stenting procedures began about eight months before the trial's results were publicized. Another example comes from the trial that CMS sponsored assessing lung-volume-reduction surgery. Although that study identified some types of patients who would benefit from the procedure, and Medicare decided to cover it nationwide in those cases, the number of Medicare enrollees undergoing that surgery actually declined after the study was published (apparently reflecting the risks of undergoing the procedure that were discovered).[52] Such effects on medical practice may not be typical, however, and in any event, it took seven or eight years to complete those trials and release the results.

In addition to trials' relatively high costs and long durations, other constraints limit the number of trials that can

48. In other words, a finding of efficacy shows that a treatment can work for some patients in some circumstances, whereas a test of effectiveness determines whether the treatment usually works for a broader set of patients.

49. See Sean R. Tunis, Daniel B. Stryer, and Carolyn M. Clancy, "Practical Clinical Trials: Increasing the Value of Clinical Research for Decision Making in Clinical and Health Policy," *Journal of the American Medical Association,* vol. 290, no. 12 (September 24, 2003), pp. 1624–1632.

50. A recent example may illustrate the risks of drawing conclusions from trial results too quickly. In 2002, a trial of hormones used to treat menopause was halted abruptly when the initial findings indicated widespread increases in the risk of heart attack for participants. Subsequent analysis, however, found that the effects varied substantially depending on the ages of the patients and that some groups would benefit from hormone replacement therapy. See Tara Parker-Pope, "How NIH Misread Hormone Study in 2002," *The Wall Street Journal,* July 9, 2007. Even so, the trial's results indicated that observational comparisons had generally overstated the benefits of hormone replacement therapy because they did not adequately account for differences between the patients who received that treatment and the ones who did not.

51. See Keith J. Winstein, "Stent Implants Declined in April; Doctors Attribute Drop to Study Showing Drugs May Have Similar Benefits," *Wall Street Journal,* May 17, 2007. According to that report, total spending in the United States on angioplasties with stents was about $14 billion in 2006, but the number of stenting procedures began to decline in mid-2006.

52. Prior to the initiation of the trial, Medicare did not have a national policy regarding coverage of lung-volume-reduction surgery, but many of the local organizations that process Medicare claims had been approving it and paying for it under existing billing codes. See Tunis, "Coverage Options for Promising Technologies"; and Gina Kolata, "Medicare Says It Will Pay, but Patients Say 'No Thanks,'" *New York Times,* March 3, 2006.

feasibly be conducted at any given time. One is getting a sufficient number of patients to participate to allow valid statistical comparisons of treatment outcomes. For medical conditions that are common, that may not be a substantial challenge, but the difficulty increases the more narrowly the target population is defined—just because fewer patients meet the criteria for participation in the trial. Ethical issues can also arise if one set of participants is assigned a treatment that is generally considered less effective, although such concerns may be less likely to arise when significant uncertainty exists in the medical community about the relative benefits of different treatments. In light of those constraints, significantly expanding comparative effectiveness research is likely to require a combination of randomized trials and other research methods.

Modeling. Another approach that has been suggested—as an alternative or supplement to clinical trials—is the use of computer models to simulate the effects of treatments on different populations of patients. While many well-designed models exist, perhaps the most prominent one is known as Archimedes; its development has been led by Dr. David Eddy with the support of the Kaiser Permanente health plan.[53] One benefit of that approach is that, once such a model is developed, it can be used to answer questions about effectiveness at relatively low cost. Indeed, that approach can even have advantages over analyses of claims data, electronic health records, or medical registries: If the model can accurately predict the effects of a new treatment, waiting for those treatments to be used and then tracking their effects on actual patients over time can be avoided in some cases.

Achieving that objective may be quite difficult, however, and a particular obstacle is that models rich enough to simulate real-world medical care may not be transparent enough to generate confidence in or acceptance of their results. Archimedes, for example, is a highly complex model that seeks to capture not only the behavior of doctors and patients but also many of the biological processes of the human body. Tests of the model have shown that under certain conditions, it is able to predict the results of trials with high accuracy. In those tests, a set of trials is examined—and usually, about half of them are used to calibrate the model, while the rest are used to test its predictions. It is not clear, however, how well the model would do when starting with a less extensive evidence base, so its primary contribution might be to fill in some gaps between existing trial results and to permit modest extensions of completed trials at relatively low cost. For more ambitious efforts, it would not be possible to tell whether the model's predictions proved correct or incorrect until after the treatment in question had been used and analyzed via the other methods described above.

53. See David M. Eddy, "Linking Electronic Medical Records to Large-Scale Simulation Models: Can We Put Rapid Learning on Turbo?" *Health Affairs,* Web Exclusive (January 26, 2007), pp. w125–w136.

The Scope and Focus of Analyses and the Dissemination of Results

In addition to determining what types of research to conduct, any organization sponsoring research on comparative effectiveness would have to make a number of decisions about the scope and focus of that research—or policymakers might decide to set parameters for those decisions. One important question is whether federally sponsored research would seek to assess both the relative clinical benefits and the cost-effectiveness of treatments. A second is what balance to strike between evaluating treatments already being used widely and examining new treatments that seemed likely to become common—and more generally, how to keep up with the rapid pace of technological development in health care. Another issue is whether and to what extent the research would compare the performance of different providers or types of providers (such as high-volume and low-volume hospitals). Last but not least is the issue of how to communicate results to doctors, patients, and other interested parties.

Clinical Effectiveness or Cost-Effectiveness? There are arguments both for and against having federally sponsored research on comparative effectiveness consider cost-effectiveness as well as clinical effectiveness. Those arguments involve the practical steps needed to do the analysis and the ultimate effects of the research.

One practical reason a federal entity might not seek to assess which treatment was most cost-effective for a given type of patient is that the answer to that question might vary by health plan. Health insurance plans have different cost structures and may pay different prices for the same services, so there is an argument for giving insurers (and other interested parties) more information about the relative benefits of different treatments and letting those parties calculate which one was most cost-effective. Indeed, the prices of the inputs involved are often subject to

negotiation. But those negotiations could be influenced by objective comparisons of medical benefits. Australia's health agency, for example, calculates a price at which a new drug is cost-effective, given its clinical benefits relative to existing therapies. (That agency conducts the reviews of clinical effectiveness as well—but because it also administers that national health insurance program, the example may not shed light on this country's debate about whether and where to conduct cost-effectiveness analyses.)

More fundamentally, objections to considering cost-effectiveness reflect concerns that such efforts would at least taint the analysis of clinical effectiveness—and might ultimately be used to restrict access to valuable treatments. To the extent that federally funded analyses of clinical effectiveness facilitated calculations of cost-effectiveness by other parties, however, the same concerns about their ultimate impact would seem to arise. Moreover, well-designed studies would primarily affect treatments whose added benefits did not appear to justify their added costs, and access to treatments would depend largely on how the results were applied by public and private insurers. Thus, a more substantial concern about the research itself is that having the same organization fund analyses of both clinical effectiveness and cost-effectiveness could reduce the impact of any findings about the former—because those findings might be perceived as reflecting cost-control objectives.

An alternative view, however, holds that federal sponsorship of research addressing cost-effectiveness would give that research more credibility. Such sponsorship could help address concerns about the consistency of the methodologies used to calculate cost-effectiveness and about the transparency of the process by which those calculations were made.[54] In addition, some observers believe that federally sponsored analyses would be viewed with less suspicion than are studies conducted by private insurers. As a practical matter, having the federal entity develop or support an initial cost-effectiveness analysis, along with a template that insurers or others could use to modify the calculation using different prices, could also avoid some duplication of effort.

A more basic argument in favor of including cost-effectiveness is that achieving the greatest possible gains in the efficiency of the health sector ultimately would require assessing both the benefits and costs of different treatment options to see whether the added benefits of more-expensive options were worth their added costs. On balance, research that included an analysis of cost-effectiveness would probably have a larger effect on medical practice than research that analyzed only the comparative clinical effectiveness of different treatments—primarily because the results would sometimes highlight that benefits were small relative to the incremental costs.

Even so, extending the scope of research to include cost-effectiveness would raise a number of additional challenges. For example, the methods of calculating quality-adjusted life years could be a source of controversy. Although there may be substantial agreement within the scientific community about the relative benefits of avoiding different adverse outcomes—such as degrees of disability and risks and side-effects of surgery—converting those differences into the common metric of QALYs might nevertheless raise concerns among patients and other interested parties. Similarly, deciding how broadly or narrowly any findings applied would be a very important consideration, because some treatments might be more effective for certain subgroups of patients than for an average patient. That consideration would affect the design of studies and the comparisons that would be undertaken; that is, the studies would need to be sufficiently robust to examine the potential variation in benefits among subgroups of patients—in order to limit the risk of overlooking patients who could benefit greatly from a treatment.

Finally, the very practice of placing a dollar value (or range of values) on an additional year of life has generated controversy; many people find the notion uncomfortable if not objectionable, and the sentiment that no expense should be spared to extend a patient's life is often expressed. Nevertheless, researchers have developed estimates of that value reflecting choices that individuals are observed to make in other settings (for example, when they purchase life insurance or accept the risks of driving). Estimates of about $100,000 per year are commonly cited, though higher and lower figures are often used. An agency charged with analyzing cost-effectiveness would not, however, have to determine what the appropriate threshold or range was—that decision could be left to purchasers and other decisionmakers. Instead, the agency

54. Standards for conducting analyses of cost-effectiveness have already been developed; for a discussion, see M. C. Weinstein and others, "Recommendations of the Panel on Cost-Effectiveness in Health and Medicine," *Journal of the American Medical Association,* vol. 276, no. 15 (October 16, 1996), pp. 1253–1258.

could estimate cost-effectiveness ratios and rank treatment options on that basis.

Other Questions of Scope and Focus. In addition to choosing which methods of research to pursue and whether to consider cost-effectiveness, a new or expanded agency would need to consider several other questions of scope and focus as well. Would it make recommendations about coverage of treatments as well? On which treatments would it focus attention, and how would it set those priorities? Would it compare different ways of providing a given treatment or concentrate on assessing broadly different options? Would it also try to assess doctors and other providers in terms of their effectiveness? And should it take explicit steps to expand the capacity for comparative research or anticipate that supply would grow to meet demand?

The question is whether the new or expanded federal entity would make recommendations about which treatments should be covered by insurance—either generally or for public programs—is related to but separate from the issue of whether to assess cost-effectiveness. Some observers have suggested that a U.S. entity focusing on comparative effectiveness should steer clear of making such recommendations because they would be controversial in themselves and because they might be seen as tainting findings about relative medical benefits. As a practical matter, furthermore, the entity would not have to make formal recommendations in order for its research to affect the use of medical care, as long as its findings on clinical effectiveness or cost-effectiveness were considered credible by doctors and other health professionals and could be easily used by insurers and other parties.

A more pressing issue is how a new or expanded entity would choose the specific treatments on which to focus its attention. Selecting broad areas of treatment (such as cardiovascular disease) might be relatively easy, but tradeoffs could arise between focusing on specific treatments that were widespread, expensive, and had uncertain benefits or, instead, on emerging treatments and technologies that promised to be expensive and might be adopted widely but had not yet become common practice. In the former case, data might be more readily available, but changing ingrained practice patterns might be difficult (short of producing evidence of actual harm). In the latter case, analyses might be more difficult to conduct given the limited claims data that would be available, while generating new data via clinical trials would take several years and thus might not be timely. A related question is how frequently to reassess treatments or variations on them; according to one study, systematic reviews typically require revision after about five years.[55]

An additional issue is whether to expand the scope and structure of comparisons so that they analyzed degrees of service use within a given treatment approach, not just broadly different approaches. As noted above, the literature on geographic variations in health care indicates that overall surgery rates do not vary systematically or in a manner that is strongly correlated with the variation in total Medicare spending. Rather, spending differences reflect more intensive use of hospital and physician services (as well as more use of ancillary services like tests). Therefore, future studies might need to examine different approaches to providing the same basic treatment, such as the extent of follow-up care provided or the frequency of using tests and imaging services—in addition to the "either/or" question of whether a given type of imaging or test was informative. Such analysis could also be applied to structured programs of care coordination or disease management, in order to assess their impact on health and their cost-effectiveness.

Another question is whether assessments would be limited to procedures and treatments or would also seek to evaluate the performance of individual doctors. In particular, the data from medical records that were used to compare the effectiveness of different treatments for a given type of patient could also be used to analyze the quality with which doctors provided each treatment. The potential gains from such analysis would include identifying doctors who delivered high-quality care and encouraging doctors who were not performing as well to improve—and doing both on the basis of objective evidence. At the same time, concerns could arise that evaluating doctors would detract from the focus on identifying effective procedures. Further, controlling for differences among patients that could affect the ratings of numerous individual doctors could be even more challenging than controlling for differences in patients when comparing a small set of treatments. Although such an approach could

55. See Karen J. Shojania and others, "How Quickly Do Systematic Reviews Go Out of Date? A Survival Analysis," *Annals of Internal Medicine*, vol. 147, no. 4 (August 21, 2007), pp. 224–233.

have a larger impact on health care than examining treatments alone, it could also be highly controversial.[56]

A final question regarding the scope of the activities funded revolves around the capacity to conduct research on comparative effectiveness. Several observers have indicated that the capacity is currently rather limited, which could serve as an important constraint on any expansion of federally funded efforts. In particular, some recent discussions have suggested that federal spending on that research should increase by billions of dollars per year, but it is not at all clear that such sums could be spent in an effective way in the near term.[57] At a minimum, additional research efforts would probably reach a point of diminishing returns. The entity that oversaw those efforts might therefore want to consider the option of using some funds to expand the pool of skilled researchers and to encourage steps that would make it easier to incorporate comparisons of effectiveness into the routine practice of medical care.[58]

The Dissemination of Results. Whatever types of results were produced, any new or augmented entity focused on comparative effectiveness would want to consider carefully how those results were communicated to doctors, patients, and other interested parties. (Such communication efforts represent an important element of AHRQ's program on effective health care.) Providing information to both technical and general audiences that was both useful and accurate would be challenging, though; a particular difficulty might be conveying the degree of uncertainty surrounding conclusions. A useful first step might be to conduct a critical assessment of past dissemination efforts in order to identify their strengths and weaknesses.

The mechanisms by which the results of comparative studies were disseminated could have important implications for their impact on medical practice. In particular, one such pathway could be the incorporation of any findings into computerized decision-support tools that some physicians and health plans now employ. Rather than having to recall any relevant evidence from memory, physicians could call up the results of comparative effectiveness research for a given patient's symptoms—or be presented with those findings (or their practical implications) automatically. The limited infrastructure for information technology that currently characterizes the health system, however, presents an obstacle to capturing the full potential of this approach. Alternatively, a few studies have found that presenting patients with comparative information about the benefits and risks of treatment alternatives—particularly in cases when elective surgery is one of the options being considered and when patients may vary in their valuation of the benefits and risks—leads them to choose less intensive treatments for certain conditions.[59] What process is most effective for presenting such information to patients, however, and how broadly those findings apply are less clear.

Implications for Health Care Spending

To affect medical treatment and reduce health care spending, the results of comparative effectiveness analyses would ultimately have to change the behavior of doctors and patients. For any large-scale changes to occur, the new or expanded entity would have to generate new findings for a substantial number of medical conditions—which would take many years. To have the maximum effect on behavior, those findings would then have to be incorporated into the incentives for providers and patients, a process of adjustment that might also take time. Although some patients and providers might object to such changes, over the long term the combination of

56. CMS has taken some initial steps toward assessing the quality of care that individual doctors provide. The Tax Relief and Health Care Act of 2006 provided for modest bonus payments under Medicare to doctors who elect to report information on certain measures of the care they provide in 2007. Although CMS will be able to provide feedback to doctors on how their performance compares to their peers', the payments do not depend on that performance. Furthermore, the measures that have been chosen cover areas of substantial consensus in the medical community about appropriate treatment protocols (for example, prescribing drugs known as beta blockers to patients who have had a heart attack). For a broader discussion, see Government Accountability Office, *Medicare: Focus on Physician Practice Patterns Can Lead to Greater Program Efficiency,* GAO-07-307 (April 2007).

57. The Health Industry Forum, *Comparative Effectiveness Forum: Executive Summary* (summary of a conference, Washington, D.C., November 30, 2006), available at http://healthforum.brandeis.edu/meetings/materials/2006-30-Nov./ExecBrief.pdf.

58. For a discussion, see Sean A. Tunis, "A Clinical Research Strategy to Support Shared Decision Making," *Health Affairs*, vol. 24, no. 1 (January/February 2005), pp. 180–184.

59. See Annette M. O'Connor, Hilary A. Llewellyn-Thomas, and Ann Barry Flood, "Modifying Unwarranted Variations in Health Care: Shared Decision Making Using Patient Decision Aids," *Health Affairs,* Web Exclusive (October 7, 2004), pp. VAR-63–VAR72.

additional information and revised incentives would tend to reduce spending for health care below currently projected levels, potentially to a substantial degree.

Currently, Medicare is effectively precluded from taking costs into account when making decisions about coverage and would probably need new legal authority to adjust payments to providers or cost-sharing requirements for enrollees to encourage the use of more cost-effective care. For their part, private insurers might not face legal barriers to limiting coverage of or altering payments for treatments that were shown to be less effective but still might be reluctant to do so if Medicare did not alter its own policies regarding coverage and payment. Thus, beyond conducting the analyses themselves, many difficult steps would probably need to be taken before spending on comparative effectiveness research translated into substantial savings for federal programs and the health care system. Even so, additional information comparing treatments would tend to reduce federal health spending in the near term—but probably not by enough to offset the full costs of conducting that research over the same period.

The Potential for Savings on Health Care

Predicting the impact that research on comparative effectiveness could have on health care spending is difficult because it is hard to know what that research will show. In some cases, the research could provide clearer evidence than exists today that the benefits of an expensive treatment outweighed the costs—in which case spending on such treatments could increase. Some observers have therefore suggested that comparative effectiveness research could also cause spending to increase on treatments already considered effective but not used as extensively as recommended protocols indicate.[60] By itself, however, new research on comparative effectiveness seems unlikely to increase the use of services that are already deemed effective, for two reasons. First, that research is unlikely to focus on such cases—instead, it would presumably target treatments of uncertain value. Second, even if it did address those types of care, an additional finding of effectiveness would be unlikely to have much impact on the use of already-recommended services without corresponding changes in the incentives to use them.

Although spending increases in some areas would be possible, current incentives already favor the adoption and spread of more-expensive treatments, so new research that found those treatments to be more effective or more cost-effective would probably increase their use only modestly. As a general rule, the fee-for-service reimbursement system by which health care is primarily financed in the United States—especially but not exclusively in Medicare—typically provides financial incentives for doctors and hospitals to adopt new treatments and procedures broadly even if hard evidence about their effectiveness is not available. For their part, insured individuals generally face only a portion of the costs of their care and, consequently, have only limited financial incentives to seek a lower-cost treatment. Although private health insurers have incentives to limit the use of ineffective care, they are currently constrained both by a lack of information and by public concerns about overly aggressive management (as was evident in a recent "backlash" against managed care plans).

Conversely, credible and well-designed studies that found that more-expensive treatments and approaches to care yielded little or no additional health benefits would have a greater potential to affect health care spending. Moreover, the evidence that additional spending and use of services in some parts of the country is not producing better health suggests that additional comparative research would be more likely to question than to support the value of more-expensive services. Research that affected the demand for treatments would also affect their supply; in particular, if the developers of new medical products and procedures had to demonstrate their value more clearly, those parties would not only have incentives to produce more evidence but also would be encouraged to focus their developmental efforts on approaches that were more clinically effective or more cost-effective. Over the long term, therefore, generating additional objective information about the relative costs and benefits of treatments seems much more likely to reduce total health care spending than to raise it—particularly if public and private insurers incorporated the findings into their coverage and payment policies.

Getting to the point at which additional research on comparative effectiveness could have a noticeable impact on health care spending would take several years. In addi-

60. One recent study found that patients typically received about half of recommended services, whether for preventive care, treatment of acute conditions, or treatment of chronic conditions. See Elizabeth A. McGlynn and others, "The Quality of Health Care Delivered to Adults in the United States," *The New England Journal of Medicine*, vol. 348, no. 26 (June 26, 2003), pp. 2635–2645.

tion to the time required to get the new activities under way, a lag would exist before results were generated, particularly if they depended upon the completion of new clinical trials. Initially, the available results would probably address a relatively small number of medical treatments and procedures; additional time would elapse before a substantial body of results was amassed. And in areas of medicine with significant levels of spending, many studies could be needed before a consensus emerged about the appropriate conclusions to be drawn—even if those studies did not generate conflicting results. For all of those reasons, it would probably be a decade or more before new research on comparative effectiveness had the potential to reduce health care spending in a substantial way.

The magnitude of that impact in the long term would depend primarily on how private and public insurers used that information and whether and how the results were incorporated into the incentives facing providers and patients. But additional information could have a modest effect on health care spending in the near term even if those incentive systems remained largely unchanged. The information would primarily affect spending in the private sector, where the scope for using comparative information is currently greater, but some "spillover" effects for enrollees in public programs would also be likely because doctors are inclined to provide similar care to all of their patients.

Possible Responses by Private and Public Insurance Plans

To affect medical treatment and reduce health care spending, the results of comparative effectiveness analyses would ultimately have to change the behavior of doctors and patients—that is, to get them to use fewer services or less intensive and less expensive services than are currently projected. Bringing about those changes would probably require action by public and private insurers to incorporate comparative effectiveness information into some combination of their coverage and payment policies. Those steps, however, could be difficult and controversial.

Private Insurers. One option for private insurers would be to not cover drugs, devices, or procedures that were found to be less effective or less cost-effective. That approach might prove to be particularly controversial, however, and the insurers would have a number of additional options as well. They could simply provide more information to providers and patients, which could improve compliance with treatment guidelines. For example, the use of medicines known as beta blockers, which is recommended following a heart attack to prevent a recurrence, has grown substantially in recent years—apparently as a result of reporting on the share of patients who receive prescriptions for them.[61] The availability of that information may have encouraged individuals to seek health plans whose doctors were more likely to prescribe beta blockers and may have encouraged doctors to prescribe them.

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); that approach is sometimes called value-based insurance design.[62] Or insurers could adjust payments to doctors and hospitals to encourage the use of more-effective care. According to one recent study, private insurers will currently cover a more effective treatment in nearly all cases—even if it is more costly—but it is also common for them to require that more costly treatments receive prior authorization before they are used or that patients try a less costly intervention first.[63] Research on comparative effectiveness could be used to determine when to apply those requirements.

Making substantial changes to insurance design and reimbursement would not be easy. Some patients, providers, and other interested parties would probably object to such arrangements or to the manner in which insurers established them. A particular concern would be that the average effects found by studies might not apply

61. Since 1996, the National Committee for Quality Assurance (NCQA), a not-for-profit organization that provides information about health care quality, has required private health care plans to report that information in order to receive accreditation. The average share increased from 63 percent in 1996 to 95 percent in 2005, and as a result, NCQA has now adopted a more stringent measure (which tracks actual use of those drugs). See Thomas H. Lee, "Eulogy for a Quality Measure," *The New England Journal of Medicine,* vol. 357, no. 12 (September 20, 2007), pp. 1175–1177.

62. See Michael E. Chernew, Allison B. Rosen, and A. Mark Fendrick, "Value-Based Insurance Design," *Health Affairs*, Web Exclusive (January 30, 2007), pp. w195–w203.

63. Garber, "Cost-Effectiveness and Evidence Evaluation as Criteria for Coverage Policy."

to all types of patients that were considered—so that subgroups of patients who could benefit greatly from a treatment might be overlooked. And as discussed above, having research studies keep pace with the development of new technologies would be an ongoing challenge. Consequently, any new incentive systems would probably be applied only in areas of care where the evidence was convincing.

Making such changes would also generate some new costs for insurers. Some administrative costs would be incurred to monitor whether patients met the medical criteria for which a given treatment had been proved effective or cost-effective. An exception or appeals process might also be needed to permit case-by-case reviews, and negotiating more complex reimbursement arrangements with providers would entail some costs. Those costs would probably be small in comparison to the change in health spending, given that insurers already monitor the use of treatments to ensure that they are medically necessary and generally have appeals processes in place. In addition, providing stronger incentives for patients and providers to use effective care would probably increase the use of services that are already deemed effective. The types of effective care that studies find are underprovided, however, tend to be relatively inexpensive screening and monitoring services for chronic health problems.

The steps that private insurers took could both affect public spending and be affected by public programs' responses to additional information about comparative effectiveness. To the extent that changes instituted by private insurers affected doctors' methods, there would probably be some "spillover" benefits for public programs. However, private insurers might be more reluctant to pursue such approaches aggressively, at least in the short term, if public insurance programs were not adopting similar methods.

Medicare. To reduce spending substantially under Medicare on the basis of comparative effectiveness research would probably require additional legislative authority to allow the program to consider relative benefits and costs in a more extensive way and to modify the financial incentives facing doctors and enrollees accordingly. Under current law, Medicare does not have clear authority to take costs into account when making decisions about what treatments are covered and has made only limited use of information about relative clinical effectiveness. Federal law does not explicitly prohibit Medicare from considering costs, but the Medicare statute provides that the program will pay for items or services if they are deemed "reasonable and necessary for the diagnosis or treatment of illness or injury or to improve the functioning of a malformed body member."[64] A regulation was proposed in 1989 that would have included cost-effectiveness as a factor in determining whether a treatment was reasonable and necessary, but that proposed regulation generated considerable opposition and was eventually withdrawn.[65]

Most recently, Medicare officials sought to clarify the definition of "reasonable and necessary" for the purpose of determining whether a new treatment or procedure would be covered. In 2000, they issued a "notice of intent" to publish a proposed rule on that topic.[66] Under the concept outlined in that notice, Medicare would generally require new treatments to provide "added value," which was defined in the following way:

- A "breakthrough" technology (one conferring substantially more benefits than existing treatments) would be covered without regard to its cost.
- A new item or service that had some medical benefits would be covered regardless of its cost if no other medically beneficial alternative was available or if the alternative treatment used a different "clinical modality." (That term was not defined precisely, but drug therapy and surgery would clearly be treated as different modalities.)
- An item or service equivalent in its benefits to a similar currently covered service (using the same modality) would be covered only if its costs were comparable to or lower than the cost of the currently covered service.

64. See section 1862(a)(1)(A) of the Social Security Act.

65. See Peter J. Neumann and others, "Medicare and Cost-Effectiveness Analysis," *The New England Journal of Medicine*, vol. 353, no. 14 (October 6, 2005), pp. 1516–1522.

66. Health Care Financing Administration, "Criteria for Making Coverage Decisions," *Federal Register,* vol. 65, no. 95 (May 16, 2000), pp. 31124–31129.

That nascent proposal also generated opposition, however, and to date, CMS has not taken the next step of issuing a proposed rule.[67]

As a practical matter, therefore, under current policy and law, Medicare generally covers any treatment or procedure that has net medical benefits—that is, benefits that outweigh the risks of the procedure—regardless of its cost or its effectiveness relative to alternative therapies. As noted earlier, Medicare officials recently developed an initiative that provides provisional coverage for new treatments that have uncertain medical benefits—but also requires the resulting evidence about their effects to be analyzed so that a more informed final decision on coverage can be made using those data. That initiative, however, may not involve comparing different treatments to see which is more effective and does not appear to take the costs of treatments into account. At the same time, CMS officials have given some indications that they will consider whether a new treatment is as good or better (on purely medical grounds) than currently covered alternatives when making coverage decisions, and a recent decision not to cover artificial spinal discs took into account a comparison of that option with other spinal surgeries that are covered.

Medicare currently has somewhat more flexibility regarding the payments it makes for covered services, which can take comparative medical benefits—and, in some cases, costs—into account on a limited basis. For example, in order for a hospital to receive an additional payment for using a new device during a covered procedure (known as a "pass-through" payment), the device must be shown to provide a substantial clinical improvement for Medicare beneficiaries compared with the current technology. Over time, however, Medicare's payments to hospitals are adjusted to account for the costs of new technologies (on an aggregate basis) without requiring an explicit analysis of their effectiveness. Similarly, CMS requires evidence that a new procedure or device offers improved medical benefits compared with similar items or services in order to qualify for a new procedure code (which is then assigned a payment rate). As noted above, CMS has been supporting research to determine whether more frequent dialysis for certain kidney patients has clinical advantages; if so, the agency could establish a new procedure code and payment amount for that service.

Although Medicare has not generally used information about effectiveness to set payment levels, a recent exception is its policy that bases payment rates on the "least costly alternative" for certain types of items. Under that policy, Medicare will not cover the additional cost of a more expensive product if a clinically comparable one is available that costs less; in other words, the program's payment rate for both products is set at the level of the least expensive one. That policy has been applied to payments for durable medical equipment and to certain comparable drugs, but wider application to products that are not very close substitutes would probably require additional statutory authority.[68]

Even those limited steps toward using information about comparative effectiveness have proven controversial, however. Medicare's decision to apply the least costly alternative policy to set the payment rate for certain drugs that treat prostate cancer, for example, has raised concerns about whether the policy has been administered consistently and questions about whether that approach has been superseded by a new system that sets the reimbursement rate for each drug as a function of its market price.[69] Similarly, CMS set the payment rate for a new antianemia drug equal to the rate for two existing drugs on the grounds that the products were "functionally equivalent"—but then the Medicare Modernization Act of 2003 prohibited CMS from applying a standard of functional equivalence in any future case involving Medicare's payments to hospital outpatient departments. (CMS's decision regarding antianemia drugs was not overturned.) Similarly, the provisions of that act governing AHRQ's research on comparative clinical effectiveness also specified that the CMS administrator could not use the results to withhold coverage of a prescription drug—although the private drug plans administering that benefit could presumably use relevant findings when designing their formularies.

67. For a further discussion, see Susan Bartlett Foote, "Why Medicare Cannot Promulgate a National Coverage Rule: A Case of *Regula Mortis*," *Journal of Health Politics, Policy and Law*, vol. 27, no. 5 (October 2002), pp. 707–730.

68. For further discussion about Medicare's current use of information on comparative effectiveness, see Medicare Payment Advisory Commission, *Report to the Congress: Issues in a Modernized Medicare Program* (June 2005), pp. 180–182.

69. For a discussion of the least costly alternative policy, see Medicare Payment Advisory Commission, *Report to the Congress: Impact of Changes in Medicare Payments for Part B Drugs* (January 2007), pp. 10–11.

If changes in law were made, Medicare could use information about comparative effectiveness to promote the use of more-effective care. It could, for example, choose not to cover treatments that were less effective or less cost-effective or it could exclude extremely inefficient providers from participating in the program—just as private insurers may do today. Alternatively, Medicare could tie its payments to providers to the cost of the most effective or most efficient treatment. If that payment was less than the cost of providing a more expensive service, then doctors and hospitals would probably elect not to provide it—so the change in Medicare's payment policy could have the same practical effect as a coverage decision. Even so, patients and providers might object more strongly to a decision not to cover a treatment than they would to a change in Medicare's payment for it. Alternatively, enrollees could be required to pay for the additional costs of less effective procedures (although the impact on patients' incentives and their use of care would depend on whether and to what extent they had supplemental insurance coverage that paid some or all of Medicare's cost-sharing requirements).

More modest steps that Medicare could take would include smaller-scale financial inducements to doctors and patients to encourage the use of cost-effective care. Doctors and hospitals could receive bonuses for practicing effective care or reductions in their payments for using less effective treatments (although the evidence to date about the effect of such pay-for-performance initiatives on health care spending is somewhat mixed).[70] Likewise, enrollees could be asked to pay only a portion of the additional costs of less efficient procedures. Or Medicare could simply provide information to doctors and their patients about their patterns of practice, which would create some pressure for doctors to use more-efficient approaches and could encourage patients to select more-efficient doctors. Adopting more modest measures to incorporate the findings of comparative effectiveness research, however, would probably yield smaller savings for the program.

Medicaid. As for Medicaid, state officials generally determine what specific services are covered—subject to broad federal requirements—and are reimbursed by the federal government for a portion of the reported costs using formulas specified in law. Because enrollees have low income, options for adjusting cost-sharing requirements to encourage the use of cost-effective care may be limited. Furthermore, a substantial portion of Medicaid spending pays for long-term care services such as nursing home care for elderly and disabled enrollees, which may be less amenable to comparative effectiveness research. At the same time, most of the poor mothers and children enrolled in the program receive their care through a private health insurance plan under contract to Medicaid, so spending for them would be directly affected by any changes that private insurers made. Another portion of Medicaid spending goes to cover cost-sharing requirements and payments of premiums for enrollees who are also on Medicare, so the impact on that spending would depend largely on what the Medicare program did.

An additional issue in applying the results of comparative effectiveness studies in Medicaid relates to the sharing of program costs between the federal and state governments. Federal matching rates under Medicaid currently range across states from 50 percent up to about 75 percent, and, by CBO's estimates, the federal government now covers 57 percent of the reported costs of health services provided by that program across the nation as a whole. At least in principle, those financing arrangements reduce the incentives for state Medicaid officials to limit coverage of less effective services—because, on net, states would face only a portion of those costs currently and would see only a portion of the savings that resulted from a programmatic change. Some coordination between state and federal officials might therefore be required to incorporate the results of comparative effectiveness research.

At the same time, many states recognize the growing fiscal burden posed by Medicaid costs, and several of them have already expressed interest in comparative effectiveness research. For example, more than a dozen state Medicaid programs are involved in a project (affiliated with the Oregon Health and Sciences University) assessing evidence about the relative safety and effectiveness of competing drugs in the same class. Similarly, the state of Washington has recently initiated a program to provide independent assessments of health technologies; a committee of physicians and other providers will review that evidence and make decisions about what treatments will be covered under the state's Medicaid program and other state-run health care programs. Oregon tried a broadly similar approach in its Medicaid program the 1990s (although controversies about the ranking of medical ser-

70. See Congressional Research Service, *Pay-for-Performance in Health Care*, CRS Report RL33713 (December 12, 2006).

vices according to their overall cost-effectiveness ultimately limited the impact of that experiment).[71]

Potential Effects on Research and Development. If public and private insurers incorporated the results of comparative effectiveness research into their coverage and payment policies, the types of new medical technologies that were supplied could be affected. Developers of new drugs, devices, and procedures would have clearer incentives than currently exist to generate products and services that yielded substantial health gains relative to current treatments or that could replicate the benefits of current care at a lower cost. Either outcome would tend to improve the efficiency of the health sector.

One potential concern is that the results of comparative research might enable the manufacturers of products with patent protection (such as prescription drugs) to charge higher prices, if their product was shown to be superior. But those same results would put downward pressure on the prices of competing products, which in turn could dampen the incentive to increase the price of the "winner." What is more, research that could establish the superiority of a specific product—rather than comparing broadly different treatment options—is more likely to be undertaken by the private sector, so the added impact of any federally sponsored studies in such circumstances might be modest. Overall, during the period of patent protection, prices of products found to be clinically more effective might rise, and those found to be less effective might fall relative to the current situation in which less is known about the comparative effectiveness of different products. The result would be relative prices that more appropriately reflected the relative values of products in terms of patients' health outcomes. Such a situation would signal producers to place a greater emphasis on developing products with greater clinical effectiveness.

Perhaps a greater concern is that extensive use of information about comparative effectiveness would discourage medical innovation and thus reduce the flow of new products and treatments—but the types most likely to be forgone are those that would have modest expected benefits or poor prospects for demonstrating cost-effectiveness. A particular concern may involve poorly constructed studies, which could provide inaccurate information about the relative merits of treatments and thus, in turn, skew research incentives; that possibility reinforces the importance of having new studies use rigorous methodologies. Overall, greater emphasis on using rigorous data about comparative effectiveness would seem likely to alter incentives for product development in ways that improved the efficiency of the health sector both at a point in time and over time.

Estimated Effects of a Recent Proposal

The near-term effects on health care spending that expanded federal research on comparative effectiveness could have are illustrated by CBO's estimate regarding a provision in legislation that was recently passed by the House of Representatives. Section 904 of H.R. 3162, the Children's Health and Medicare Protection Act of 2007, would do the following:

- Establish within AHRQ a Center for Comparative Effectiveness Research, which would fund research comparing the clinical effectiveness of treatments—using clinical trials, systematic reviews, observational studies, medical registries, and other methods. The center would develop methodological standards for conducting studies of comparative clinical "value" but would not fund studies that analyzed treatments' cost-effectiveness.

- Create a commission to oversee the center's activities, consisting of the Director of AHRQ, the Chief Medical Officer of CMS, and up to 15 additional members who would have relevant expertise and would represent clinicians, patients, researchers, insurers, and employers. The Comptroller General of the United States, in consultation with certain Members of Congress, would appoint those 15 members to multiyear terms.

- Provide authority to spend $300 million from 2008 to 2010 and $375 million per year thereafter, funded primarily by an annual per capita fee of about $2 imposed on private health insurance premiums that would start in 2011.

The provision would not change any of Medicare's or Medicaid's rules about which procedures and treatments were covered or how much was paid for them. Thus, any

71. See Thomas Bodenheimer, "The Oregon Health Plan—Lessons for the Nation," *The New England Journal of Medicine*, vol. 337. no. 9 (August 28, 1997), pp. 651–655; and Jonathan Oberlander, "Health Reform Interrupted: The Unraveling of the Oregon Health Plan," *Health Affairs,* Web Exclusive (December 19, 2006), pp. w96–w105.

impact that the resulting research would have on federal spending for health care would have to come primarily from changes such research induced in doctors' patterns of practice or patients' choices of treatments. Those changes—encouraged in some cases by private health insurers—would primarily affect private health spending, but some changes in treatment patterns would also be likely for enrollees in public programs because doctors tend to treat their patients in a similar manner regardless of their source of insurance.

To a lesser extent, some federal savings might also occur through changes in coverage that could be implemented under current law (although CBO did not make explicit assumptions about what those changes would be). For example, if research on comparative effectiveness determined that a service covered by Medicare did not confer any health benefits for certain types of patients or involved risks that outweighed the expected benefits, under its current coverage policies CMS would have clear authority to decide not to cover that service for those patients.

As discussed, evaluating the precise effect of new research is difficult because it is hard to know which studies will be undertaken and what they will find, but CBO estimates that such research would probably reduce spending for health care somewhat. Any impact of a given research study is likely to be felt over many years, so the change in spending in any given year would reflect the cumulative effects of past studies. Little evidence is available with which to estimate the precise magnitudes of the annual effects, although one comprehensive review of the issue indicated that additional information about the effectiveness of treatment options could "succeed in improving health care while paying for its own research-related costs through targeted health system cost reductions."[72] In estimating the effects of section 904, CBO assumed that the annual federal savings on health care would eventually reach a point at which they roughly equaled the annual outlays for research on comparative effectiveness—a process that would take about a decade.

Under H.R. 3162, budget authority for the Center for Comparative Effectiveness Research would be $1.1 billion over the 2008–2012 period and $2.9 billion over the 2008–2017 period. Because spending those funds would take some time, CBO estimates that outlays would amount to about $600 million over five years and $2.4 billion over 10 years. Direct spending by the federal government—mostly for Medicare and Medicaid—would be reduced by $0.1 billion over the 2008–2012 period and $1.3 billion over the 2008–2017 period. (Those amounts would constitute a very small fraction of cumulative federal outlays for those programs—less than one one-hundredth of 1 percent.) Thus, the net effect of enacting section 904 would be to increase federal direct spending by $0.5 billion over five years and $1.1 billion over 10 years, CBO estimates.

The impact on total spending on health care in the United States would be about five times as large as the effect on federal outlays, CBO estimates. Some of that effect would be seen in lower costs for providing health insurance to workers—costs that are excluded from income and payroll taxes. In turn, some of those savings on private insurance premiums would go to increase the taxable compensation of workers; by itself, that change would lead to a small increase in expected federal revenues. At the same time, the new fees on health insurance that would be used to finance the research would generate corresponding increases in health care costs for workers, which would tend to reduce taxable compensation modestly. Overall, those indirect effects of section 904 on revenues would be small.

72. Office of Technology Assessment, *Identifying Health Technologies that Work: Searching for Evidence*, p. 6.