Journal of Economic Perspectives—Volume 25, Number 2—Spring 2011—Pages 27–46

The Pragmatist's Guide to Comparative Effectiveness Research

Amitabh Chandra, Anupam B. Jena, and Jonathan S. Skinner

All developed countries have been struggling with a trend toward health care absorbing an ever-larger fraction of government and private budgets. One potential solution is to rely more heavily on studies of the costs and effectiveness of new technologies in an effort to ensure that new spending is justified by a commensurate gain in consumer benefits. For most nonhealth commodities, markets function sufficiently well to perform this function unassisted. But in a market such as health care, effectiveness studies can (in theory) shed light on what patients would have demanded in the absence of moral hazard and adverse selection.

As one example, an Associated Press article described patient reactions to the price of a $93,000 drug (Provenge) that extends life for incurable prostate cancer by an average of four months (Marchione, 2010). One respondent, Bob Svensson, 80, a former corporate finance officer whose insurance was paying for the treatment, declared: "'I would not spend that money,' because the benefit doesn't seem worth it . . ." Perhaps reassuringly, this particular treatment would fail most cost-effectiveness guidelines.

■ *Amitabh Chandra is a Professor at the Harvard Kennedy School of Government, Harvard University, Cambridge, Massachusetts. Anupam Jena is a resident in the Department of Medicine at the Massachusetts General Hospital, Harvard Medical School, Boston, Massachusetts. Jonathan Skinner is John Sloan Dickey Third Century Chair in Economics and Professor of Family and Community Medicine, Dartmouth Institute of Health Care Policy and Clinical Practice, both at Dartmouth College, Hanover, New Hampshire. Chandra and Skinner are also Research Associates at the National Bureau of Economic Research, Cambridge, Massachusetts. Their e-mail addresses are ⟨amitabh_chandra@harvard.edu⟩, ⟨jena.anupam@mgh.harvard.edu⟩, and ⟨jonathan.skinner@dartmouth.edu⟩.*

doi=10.1257/jep.25.2.27

In many high-income countries, government agencies are responsible for making nationwide coverage decisions on medical therapies that are expensive and of uncertain benefit compared to cheaper alternatives. In the United Kingdom, for example, the National Institute for Health and Clinical Excellence (NICE) determines which treatments are reimbursed under the National Health Service, and tends to look unkindly on those that require spending more than about $50,000 to gain an extra (statistical) quality-adjusted year of life. Alternatively, payers can set "reference prices" or upper limits on payments for branded pharmaceuticals, as is done in Germany. A related approach to implementing cost effectiveness would be to pay more for new innovations only if they offered some clear advantage over existing treatments (Pearson and Bach, 2010).

In the United States, the original Medicare and Medicaid statutes prohibited the government from reimbursing expenses incurred for "items and services that are not *reasonable and necessary* for the diagnosis or treatment of illness or injury or to improve the functioning of a malformed body member." Whether "reasonable" implied cost effectiveness was unspecified; in practice, individual physicians were entrusted to make this determination. But in a world of fee-for-service reimbursement, this latitude encouraged the overuse of technologies of dubious value. Successive program administrators wanted to interpret "reasonable" as encompassing information about costs, and in the early 1990s, the government proposed regulation that would do precisely this. Unsurprisingly, there was massive opposition from patient advocacy groups (the American Association of Retired Persons), physician lobby groups (the American Medical Association), and pharmaceutical and device lobby groups (the Pharmaceutical Research and Manufacturers Association of America) who raised concerns about "rationing," leading to the withdrawal of the proposal. Consequently, the Medicare program continues to reimburse for any medical therapy regardless of the incremental value of its benefit. In light of charges about "death panels" in the debate surrounding the healthcare reform bill of 2010, Congress explicitly forbade the use of cost-effectiveness analysis in government programs (Sections 1182(b)(2), 1182(c)(1), 1182(e) of the Patient Protection and Affordable Care Act).

In this context, comparative effectiveness research emerged as an alternative strategy to understand better what works in health care. Put simply, comparative effectiveness research compares the efficacy of two or more diagnostic tests, treatments, or health care delivery methods *without any explicit consideration of costs.* To economists, the omission of costs from comparative effectiveness research might seem nonsensical, especially when healthcare reform was motivated in part to restrain runaway cost growth (Garber and Sox, 2010).

We argue that comparative effectiveness research still holds promise. First, it sidesteps one problem facing cost-effectiveness analysis—the widespread political resistance to the idea of using prices in health care. Such resistance is not just from political interest groups, but also from voters, who even in lab settings often dislike rationing based on cost effectiveness (Nord, Richardson, Street, Kuhse, and Singer, 1995). Second, there is little or no evidence on comparative effectiveness for a vast

array of treatments: for example, we don't know whether proton-beam therapy, a very expensive treatment for prostate cancer (which requires building a cyclotron and a facility the size of a football field) offers *any* advantage over conventional approaches. Most drug studies compare new drugs to placebos, rather than "head-to-head" with other drugs on the market, leaving a vacuum as to which drug works best (Nathan, 2010). Simply knowing what works and what doesn't will improve *productive* efficiency by shedding medical practices that are unsafe at any price.

But not everyone is a fan of comparative effectiveness. Critics have focused on heterogeneity of treatment effects across patients and physicians. A randomized trial may find no benefit on average, but this tells us much less about whether a specific subset of patients (or patients of particularly skilled physicians) might still gain from the treatment (Groopman, 2010). These critics suggest that "cookie-cutter" comparative effectiveness coverage decisions can introduce rationing and ultimately worsen patient outcomes. And while comparative effectiveness research can lead to cost savings (Perlroth, Goldman, and Garber, 2010), adopting any treatment that improves health outcomes, no matter what the cost, can worsen *allocative* inefficiency by paying dearly for small health gains. Of course, cost-effectiveness studies that explicitly account for both costs and benefits of healthcare choices would avoid this type of allocative inefficiency, but could introduce other problems, such as provider inertia or drug or device suppliers increasing prices so that they fall just short of cost-effective hurdle rate set by third-party payers.

The real question, though, is whether comparative effectiveness or cost-effectiveness research can help to break the inexorable growth in healthcare costs threatening the solvency of state governments and the U.S. federal government. Some moderating effects might be expected if such research can be used to nudge patients away from less-effective therapies, whether through improved decision making or by encouraging beefed-up copayments for cost-ineffective procedures. More promising still for reducing growth is the use of a comparative or cost-effectiveness framework to better understand where the real savings lie—and the real savings may well lie in figuring out the complex interaction and fragmentation of healthcare systems.

A Primer on Effectiveness Research

The Institute of Medicine, the independent nonprofit organization that is a part of the National Academy of Sciences, defines "comparative effectiveness research" as "the generation and synthesis of evidence that compares the benefits and harms of alternative methods to prevent, diagnose, treat, and monitor a clinical condition or improve the delivery of care." Benefits can be measured in a number of ways. When comparing treatments for hypertension, for example, efficacy may be measured in terms of life-years saved, strokes prevented, or heart attacks avoided. When comparing diagnostic tests such as CT colonography and colonoscopy to screen for colon cancer, efficacy may be measured by additional cases of disease diagnosed.

One common measure of effectiveness of a healthcare treatment is a "quality-adjusted life year" or QALY. This calculation is done by looking at the additional years of life a treatment provides, weighting those years by the quality of health in each year. The quality rankings range from a value of one for perfect health to a value of zero for death, and even allow negative values for especially unpleasant states of being alive. The quality rankings are determined by looking at the ability of individuals to function along five dimensions: mobility, pain/discomfort, self-care, anxiety/depression, and carrying out normal activities like work, study, and leisure. A healthcare treatment thus could add to years of live, or improve the quality of years of life, or some mixture of the two. The coarser the measure of benefits (for example, measuring survival, but not pain and nausea), the less useful the results of an effectiveness study.[1]

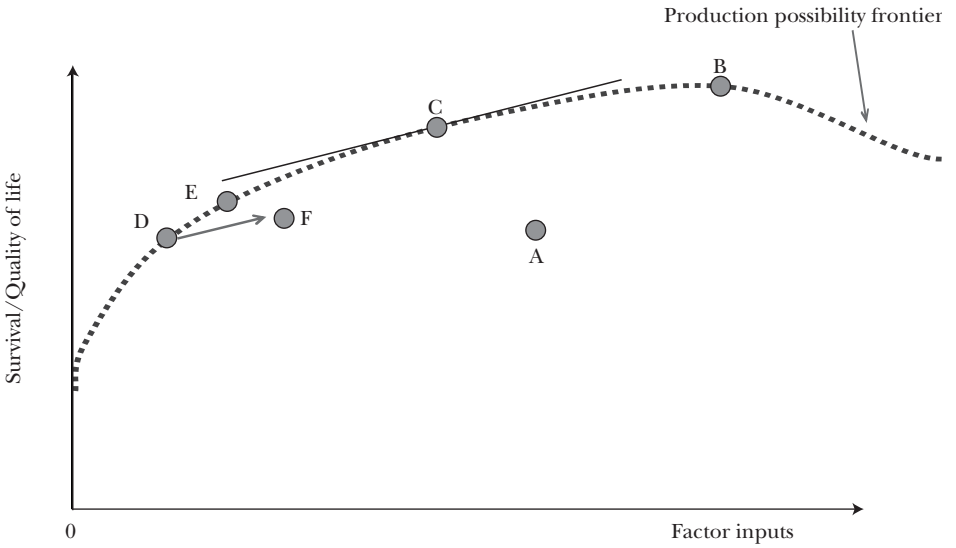
In comparative effectiveness research, the goal is to choose the option with the best health outcome. In an ideal world where all approaches to treating patients—given an existing body of scientific knowledge—are tested against one another, we could improve along each step of the way the overall health of the population. In this hill-climbing exercise, we would end up at a point where health is maximized, regardless of costs.

Figure 1 displays the association between factor inputs on the horizontal axis and survival/quality of life on the vertical axis. A concave production possibility frontier illustrates the maximum aggregate health for a given level of inputs. The U.S. healthcare system, represented by point *A* in Figure 1, falls far short of the production possibility frontier, whether because of wasteful costs (as discussed by Cutler and Ly in this issue), or because of shortfalls in health outcomes which would include both reductions in health and lives lost owing to sins of omission (lack of effective care such as prophylactic antibiotics prior to surgery) or sins of commission (Brenner and Hall, 2007, estimated that the overuse of CT and MRIs cause 1.5–2.0 percent of total cancers). The application of comparative effectiveness research to every possible treatment option would move the country to point *B*, at the peak of the production function. This point would almost certainly be more costly than our current status quo, but would represent a point where all possible health-related gains have been exhausted.

While point *B* is productively efficient, it is allocatively inefficient, given that the foregone consumption of attaining that last QALY is so high at point *B* (Garber and Skinner, 2008). By contrast, the objective of cost effectiveness is to adopt only those treatments that yield QALYs at a reasonable cost—where "reasonable" is of course open to interpretation. If we adopt for convenience a guideline of $100,000 per QALY or unadjusted life-year (a parameter we discuss later in the paper), then an exhaustive set of studies would again lead us into productive efficiency but at a different point on the production possibility frontier (specifically, at point *C*). At this point, the slope of the frontier is equal to the inverse of this cost-effectiveness

[1] An alternative to the QALY is disease-adjusted life-years (DALYs), which differ primarily in measuring disease (rather than health) and allows for age-based weights. See for example Robberstad (2005).

Figure 1
Cost Effectiveness and Comparative Efficiency in a Healthcare Production Function



Note: Figure 1 displays the association between factor inputs on the horizontal axis and survival/quality of life on the vertical axis. Point A falls far short of the production possibility frontier. Comparative effectiveness analysis can help the movement towards productive efficiency (point B), while cost-effectiveness analysis would identify the point at which productive and allocative efficiency is achieved (point C).

"hurdle" rate, in this case 1/$100,000. Economists would prefer point *C* to *B* since the forgone (nonhealth) consumption involved in getting from *C* to *B* exceeds the value of improved health. Intuitively, with the cost-effectiveness approach, all potential treatments are considered, but only those options that improve health for less than $100,000 per QALY, or that scale back on treatments costing more than $100,000 per QALY (such as that anticancer drug that cost an average of $93,000 for an average gain in life of four months) are chosen. In the aggregate, health outcomes would improve and costs would likely decline, but some subsets of the population, such as incurable prostate cancer patients, could end up being worse off (Weinstein and Skinner, 2010).

The Promise of Comparative Effectiveness Studies

In many cases, comparative effectiveness studies can lead to cost savings. One recent randomized trial compared patients with terminal lung cancer; half were randomized into early palliative care and the other half received regular chemotherapy treatments (Temel et al., 2010). Those in early palliative care experienced better quality of life, lower costs, and *longer* survival.

Another example of cost-saving comparative effectiveness research comes from arthroscopic surgery for osteoarthritis of the knee, in which surgeons enter the knee in a way that is minimally invasive and clean out particles from the joint using a sophisticated camera to guide their movements. Prior to 2002, over 650,000 such surgeries had been performed each year. In that same year, a landmark study (Moseley et al., 2002) demonstrated that compared to a control group of patients receiving "placebo surgery"—skin incisions and simulated surgery—there was no benefit from arthroscopic surgery, leading to a subsequent decline in its use (Hawker, Guan, Judge, and Dieppe, 2008). Perlroth, Goldman, and Garber (2010) suggest that comparative effectiveness research could save up to $3 billion annually by establishing that for prostate cancer patients, prostatectomy ($7,300 cost) yields results as good as brachytherapy or radiation seeds ($19,000) and radiation therapy ($46,900).

Similarly, with cost effectiveness and comparative effectiveness it is easy to make the case for the use of costly percutaneous coronary intervention (PCI), a technique in which narrowed or blocked blood vessels of the heart are opened by inserting an inflatable balloon and often kept open by introducing a coronary stent. (Percutaneous means that the intervention is done through the skin; coronary means that it is done for a blood vessel in the heart.) This technique has been shown to improve survival dramatically compared to drug therapy alone following a heart attack if performed within the first 12 or 24 hours following its onset, and thus it is highly effective (and cost-effective) for this use (Hartwell et al., 2005).

However, in the evaluation of PCI for patients with stable angina (chest pain and associated symptoms caused by strenuous activity), comparative effectiveness and cost effectiveness part ways. For this group, accounting for about one-third of all PCI procedures, clinical trials have found no mortality benefit and little (and transitory) symptom benefit of PCI relative to drug therapy alone (Boden et al., 2007; Weintraub et al., 2008). The positive benefit means that it passes the comparative effectiveness test; the small magnitude of the benefit and its high cost means that it fails the cost-effectiveness hurdle. But all is not lost for those who worry about allocative efficiency: armed with this new information, patients nervous about invasive procedures are now able to make better decisions, and evidence suggests that well-informed patients tend to want less cardiac surgery, not more (Morgan et al., 2000).[2]

Avoiding all mention of costs makes comparative effectiveness less appealing for economists but possibly more appealing for voters. An intriguing strand of the literature argues that voters, at least Australian ones, simply do not agree with the principle of cost-effectiveness analysis. In a survey conducted Down Under, respondents were asked about hypothetical choices between treating people with Disease *X*, which is treated cheaply, versus Disease *Y* requiring more expensive treatments (Nord et al., 1995). Respondents understood the trade-off and that spending a fixed budget to save people with Disease *Y* would lead to fewer overall lives saved.

[2] All this said, overall rates of PCI have continued to rise, suggesting that physicians who believe in the procedure, or those whose economic interests would be devastated by recommending against it, are having a greater impact than physicians who are nonbelievers.

Table 1

Five Different Ways to Allocate $1 Million Dollars, with Lives Saved of People with Diseases *X* and *Y*, and Most Preferred Options as Chosen by Survey Respondents

	I	*II*	*III*	*IV*	*V*
Number of people with Disease *X* saved	10	20	30	40	50
Number of people with Disease *Y* saved	8	6	4	2	0
Total saved	18	26	34	42	50
Percentage of survey respondents choosing each option	5%	27%	48%	14%	6%

Source: Nord, Richardson, Street, Kuhse, and Singer (1995, table 4).

Five options (I through V) are shown in Table 1, with total lives saved in the third row. Just 6 percent of the population chose the cost-effective solution (V), and about as many choose the least cost-effective approach (5 percent). Nearly half chose III, leading to just 34 lives saved instead of the maximum of 50. This result could reflect to some extent "central tendency" of respondents to choose what appears to be the median option, although it is also intriguing that option II substantially outpolled option IV. However, it also appears that the respondents viewed the cost-effective approach as unfair because it failed to insure against the risk of contracting a disease that was more costly to treat.

The Oregon experiment in cost-effective rationing can be viewed as a real-world example of the disconnect between the principles of cost effectiveness and voter preferences. Starting in 1989, Oregon embarked on a state-level effort to expand Medicaid health insurance coverage to more of its citizens and to finance this broader increase by providing a more limited package of healthcare services. Oregon ranked more than 700 healthcare services according to the desirability of coverage using a panel comprising patients and providers, and the Oregon legislature chose a level below which services would not be covered by state Medicaid. Controversy around this list arose when it was published: for example, life-saving surgical treatments for ectopic pregnancy and appendicitis were ranked below less-important procedures like dental caps for pulp exposure and splints for temporomandibular joint disorder (Hadorn, 1991). Though cost-effectiveness analysis suggested that the net value to society of treating 100 patients with painful temporomandibular joint disorder was of greater net value than saving a single life, the experiment failed. At a minimum, the episode suggests that even when there is some general level of acceptance for the cost-effectiveness argument, implementation is controversial and difficult.

More generally, there appears to be a disconnect between how people think of the whom-to-cover trade-off versus the what-to-cover trade-off (Baicker and Chandra, 2010). People seem to prefer that health care for the insured not be rationed. But providing generous coverage—drugs, hospitalizations, outpatient

services, proton beam therapy, long-term care—to some and nothing to others, is also a form of rationing.[3] It costs the same to insure 30 million people with a policy that has an annual premium of $6,000 per year as it does to insure 50 million people with a policy whose premium is $3,600. To date, the debate about the problem of healthcare costs has been mostly about excluding people from health insurance (Sack, 2011) rather than cutting the generosity of public benefit packages.

Finally, the comparative effectiveness research can prove a useful first step even in the absence of cost information if it provides key estimates of treatment effects, as Garber and Sox (2010) have noted. After all, such effects are typically expensive to determine and require years or even decades of data. Costs are much easier to measure, and can be appended at a later date as financial Armageddon draws closer.

Challenges to Using Comparative Effectiveness Studies

Critics of comparative effectiveness focus on the possibility for heterogeneous patient benefits, which reduce the benefits of what can be learnt from such studies. The effectiveness of a treatment for a given individual can be broken down further according to idiosyncratic patient attributes and according to the process or delivery system by which treatments are delivered. These two forms of heterogeneity may result in some patients benefiting from a treatment, while others are unaffected by it (or even harmed). Consider again Figure 1, where initially one begins on the production possibility frontier at point *D*. Now consider two approaches to expanding a new, potentially cost-effective treatment. In the case where only those appropriate for care get it, outcomes and costs improve, to point *E*, still on the "best practice" production possibility frontier. But in the case where treatment is extended across all patients, corresponding to point *F*, outcomes are worse, and costs are higher because the procedure is now done for a wider swath of patients. With variations in healthcare systems with regard to appropriate use of new technologies, extending treatment could even lead to a negative correlation between spending and outcomes, as illustrated by points *E* and *F*.

Heterogeneity in Patient Benefit

Comparative effectiveness research may demonstrate the superiority of one treatment over another when evaluated on average, even though the optimal treatment may vary across patients. This problem would most naturally arise if benefits are imprecisely measured—pain, nausea, or incontinence can be difficult to capture—and collapsed into a single outcome index.

[3] Some readers will note that the Emergency Medical Treatment and Active Labor Act (EMTALA) forces hospitals to provide emergency care in the emergency rooms without regard to citizenship or ability to pay. This is true, but EMTLA only requires emergency department physicians to stabilize the patients, not to treat them; a cancer patient would not receive any care for their cancer, nor a diabetic a prescription for insulin.

A more complicated situation arises when patient benefits are correctly measured, but some patients benefit more than others from a treatment—a phenomenon known as "treatment effect heterogeneity." To illustrate, consider the biologic drug panitumumab (brand name Vectibix) produced by Amgen. In 2007, the drug was evaluated in Europe for treatment of metastatic colorectal cancer. The drug was rejected on the basis of similar efficacy to pre-existing, less-expensive chemotherapy. After reviewing the initial submission data, it turned out that those patients with a specific normal gene type were far more likely to benefit from the drug than patients with a mutated gene. By the next year, the drug was approved for patients with a normal gene. A comparative effectiveness study that focused only on the average benefit of patients receiving panitumumab would miss the substantial benefit to a particular subset of patients more likely to benefit. Treatment effect heterogeneity is likely to increase in the future, as drug and biologic manufacturers develop therapies that are tailored to people with certain genes (Garber and Tunis, 2009).

The solution appears straightforward: conduct more studies for the relevant groups. However, this approach can be very expensive, particularly if one doesn't know which groups might benefit. If the results of average effects in a trial have just been announced, and subgroup analyses are precluded by poor statistical power, then what? Binary coverage decisions—cover/not cover—would raise concerns about potentially rationing valuable care in subpopulations, particularly where a physician believes on the basis of experience that a specific patient might benefit (Groopman, 2010). A less-stringent use of this new information would be to help "nudge" patients away from the treatment (Sunstein and Thaler, 2008). If comparative effectiveness studies are used to determine patient cost-sharing, or to design shared decision-making videos and to inform (but not determine) provider behavior, then the scope for claiming that valuable care is being withheld is substantially diminished.

Heterogeneity in Provider Skill

The effectiveness of a given technology may also depend on the skills of healthcare providers. Providers who use a certain technology repeatedly may find that there are economies of scale, learning by doing, or spillovers to other therapies. In heart disease, for example, patients receiving coronary stents in low-volume medical centers have higher 30-day mortality than patients treated in high-volume centers (McGrath et al., 2000). Chandra and Staiger (2007) find that regions that specialize in treating heart attack patients with intensive management (such as early percutaneous coronary interventions) obtain better results with the therapy than regions relying mainly on medications alone (like aspirin, beta-blockers, and statins).

Another example comes from carotid endarterectomy, a surgical procedure which removes plaque from the inside of the carotid artery that supplies the head with blood, thereby reducing the chance of stroke. In looking at hospital performance in the Asymptomatic Carotid Atherosclerosis Study (ACAS), Wennberg,

Lucas, Birkmeyer, Brendenberg, and Fisher (1998) note dramatic differences in the mortality that occurs within two weeks of the endarterectomy (known as "perioperative mortality") depending on whether the procedure was performed in one of the original clinical trial hospitals (1.4 percent mortality), in nontrial hospitals with high volumes of endarterectomies (1.7 percent), or in hospitals with low volumes (2.5 percent). In other words, procedures worth doing in academic medical centers may not be worth doing in community hospitals. This raises the bar even further for studies, requiring randomization across types of providers as well as patients.

How Much Will Comparative Effectiveness Research Cost?

Recall from Figure 1 that using a procedure only among appropriate patients leads to better outcomes at lower costs (point *E* rather than point *F*). But this ignores the costs of determining which subgroup is most appropriate for treatment. Thus, it is necessary to think about value-of-information studies, which assess the value of obtaining additional information on the clinical effectiveness of particular treatments (Dorsey and Meltzer, 2010). Broadly speaking, which treatments should be evaluated sooner rather than later will depend on the degree of uncertainty about clinical effectiveness (perhaps determined by expert panels and systematic review of the medical literature) and the potential total cost savings associated with recommending various treatments—which in turn will depend on the costs of various treatments and the number of people eligible for the treatment.

Observational studies and randomized control trials are two approaches to learning about effectiveness. The former is substantially cheaper than the latter yet carries many caveats. The simplest form of an observational study uses the standard "as treated" approach at the individual patient level with either propensity-score matching or regression analysis with covariates; there is no randomization and the researcher interprets the "treat/nontreat" coefficient as the treatment effect.

Observational studies based on regression adjustments are cheap and relatively easy to conduct, and an optimist might believe that observational studies with high-quality data identify treatment effects just as well as randomized control trials (Concato, Shah, and Horwitz, 2000). But "just as well" isn't always known until a trial is conducted. One prominent example in recent years is hormone replacement therapy, which was given to millions of women in the belief based on observational studies that it reduced menopausal symptoms and decreased heart attacks. However, randomized controlled trials in the late 1990s and early 2000s found that long-term use increased risks of heart attack and stroke (Taubes, 2007).

The discrepancy between randomized trials and observational studies is most salient in situations where the success of a treatment depends on patient factors that go beyond patient severity. It becomes very hard for observational studies to control effectively for confounding variables such as patient adherence, social and family support, and health literacy. All of these factors affect

outcomes, but each is notoriously difficult to measure and thus to control for in a regression. Moreover, the reason that some patients stick with a drug for a long time while others do not is that some patients experience side effects like pain and nausea, while others do not. In such a world, compliance is correlated with benefit, and simply "controlling" for patient factors is unlikely to yield a causal effect.

A more sophisticated class of observational studies tries to construct "natural experiments" to estimate treatment effects, and thus owes more to the econometrics literature. For example, distance to a catheterization lab has been used as an instrument for healthcare intensity (McClellan, McNeil, and Newhouse, 1994), while discontinuity designs (possible, for example, when birthweight cutoffs determine admission to a intensive care unit as in Almond, Doyle, Kowalski, and Williams, 2010) have been used to sidestep the otherwise daunting biases inherent in individual-level as-treated models. But the power of this methodologically superior approach is still limited because not every treatment displays a discontinuity or instrumental variable to mimic randomization. Furthermore, the estimated treatment effect is known to be valid only in the vicinity of the discontinuity.

For these reasons, the randomized controlled trial is viewed as the gold standard for evidence. Unfortunately, randomized trials are also expensive. For example, preapproval clinical testing done by pharmaceutical companies as part of getting approval from the Food and Drug Administration—so-called Phase III testing—for a single drug costs roughly $86 million (in 2000 dollars), according to a study by DiMasi, Hansen, and Grabowski (2003). Randomized control trials designed to generate comparative effectiveness research on drugs already known to be efficacious could be less expensive or more expensive, particularly if performed in expanded patient populations to study subgroup efficacy.

Who will cover the cost of these randomized trials? The answer is not always clear. In July 2005, clinical trials established the effectiveness of the biologic drug ranibizumab (brand name Lucentis) in the treatment of macular degeneration, in which older adults suffer retinal damage and severe vision loss (Martin, Maquire, and Fine, 2010). While awaiting approval from the Food and Drug Administration for the new drug, ophthalmologists used an alternative drug (bevacizumab or Avastin, which already had approval), that was essentially identical but significantly cheaper—only $50 per dose versus $2,000. Comparing the two treatments is an obvious application of comparative effectiveness. Yet no institution was initially willing to step up to fund such a study, and the Centers for Medicare and Medicaid Services for a variety of reasons could not pay for the full cost of the drugs; the study was saved only by the National Institutes of Health stepping in at the last minute to provide $25 million in funding on an ad hoc basis. This example underscores the importance of having a mechanism for paying for the treatment while it is being evaluated—a manufacturer may underwrite the costs for a trial in a group where the treatment is expected to work, but will be unwilling to do so for head-to-head comparisons where effectiveness is less clear. In addition to highlighting the issue of the costs of paying for treatments in trials,

the example of Lucentis versus Avastin also underscores the tremendous benefit from conducting trials of efficacy for similar drugs with dissimilar pricing. Both drugs are made by Genentech, they are in the same class, and have fundamentally similar mechanisms of action. But Lucentis was developed to capture the higher surplus associated with treatments for macular degeneration; its development was a mechanism to price discriminate.

Given that the United States now spends close to 18 percent of GDP on healthcare (a level of spending close to $8,000 per capita), it seems reasonable to pay a small fraction of this cost towards figuring out what works and what does not. The current National Institutes of Health budget is about $31 billion per year ($100 per person), and even tripling the NIH budget to do more effectiveness research would mean that approximately 2 percent of total healthcare spending would then be spent on how to make care more effective. Relative to the cost of developing a new drug this is a small amount of spending; one study estimates that recent drugs have cost $868 million per drug to develop, with a range of $500 million to $2 billion (Adams and Brantner, 2006). Further, spending more on learning what works should be viewed as an investment if it bends the cost curve trajectory. It is also plausible that this kind of knowledge has a strong public good aspect, which would imply that society has underinvested in such research. No individual insurer—whether Medicare or a commercial provider—will fully internalize the benefits of learning the appropriate frequency of office visits or whether Avastin increases survival among patients with metastatic colon cancer. The presence of such knowledge externalities would suggest a powerful role for federal funding of these trials, perhaps funded through taxes imposed on the healthcare industry.

Towards the Gold Standard: Adding Costs to Effectiveness Analysis

Of course, comparative effectiveness isn't enough for cost-effectiveness, which depends on the societal value of the additional life gained *and* the relative cost of achieving that gain. The costs and benefits in cost-effectiveness analysis should reflect the lifetime costs and benefits associated with each intervention (Meltzer, 1997). To compare value to costs, economists have proposed converting quality-adjusted life-years (QALYs) to dollars. The conversion factor was initially suggested as $50,000 per QALY, based on a 1984 Canadian study of annual costs of care for patients with end stage kidney disease on dialysis (Winkelmayer, Weinstein, Mittleman, Glynn, and Pliskin, 2002). Since that time, it has been updated for inflation to $100,000 per QALY (Lee, Chertow, and Zenios, 2009). An alternative way of rationalizing this figure is to use annual salaries in industrialized nations to value an additional year of life; for example, an annual salary of $30,000 for a 40-hour work week would lead to a value of a life-year of approximately $100,000 if leisure time were valued at the same rate as the market wage (also see Garber and Phelps, 1997). Clearly, there is considerable uncertainty about what is

"the" value of a life, with some estimates topping $300,000 per year (Murphy and Topel, 2006).

When treatments are comparatively effective but cost more than $100,000 per additional QALY in the United States, they are generally viewed as not being cost effective, even if Medicare or private insurance companies continue to pay the bills (Cutler, 2004). In countries such as the United Kingdom and Australia, treatments whose incremental life extensions cost more than $50,000 per QALY are routinely denied coverage. For example, in a highly publicized coverage decision regarding the biologic medication bevacuzimab (Avastin), the United Kingdom refused national coverage of the drug for patients with metastatic colorectal cancer on the basis that the drug improved life expectancy by six weeks over the preexisting standard of care but cost an additional $110,000 dollars per QALY to do so.

Aren't Prices Charged for Treatments Endogenous?

A subtler point is that most cost-effectiveness studies use the *price* charged to the national health plan or insurer as the measure of cost rather than the actual cost of production. The difference between price and cost is a particularly important issue in health care, where new technologies or patented inventions often have prices that far exceed the costs. A patented drug may have a high mark-up, while another drug that is equally costly to produce may be priced much cheaper if that drug's market is more competitive. If prices are used instead of costs in a cost-effectiveness analysis, the analysis may not lead to the socially efficient outcome (Jena and Philipson, 2010; Basu and Philipson, 2010). Indeed, in the case of multiple drug-resistant tuberculosis treatments in developing countries, global health leaders were able to negotiate the price of drugs down by as much as 90 percent, suggesting that many cost-effectiveness ratios using prices should be viewed as opening bids in a process of price negotiation (Kim et al., 2005).

Cost-Effectiveness Analysis of Healthcare Delivery Systems

Cost-effectiveness research may ultimately deliver its largest productivity improvements through the analyses of healthcare delivery systems, which vary greatly in their use of office visits, specialist consultations, outpatient services, and imaging technologies. Evaluating these interventions separately is tricky given the complicated production function that maps these inputs into health. But evaluating the overall productivity of different delivery systems offers great potential for substantial cost saving. To illustrate, leading medical care centers have nearly two-fold range in risk-adjusted costs in their care of patients with heart attacks, largely due to how frequently patients are seen and how often they are referred to specialists, cared for in the hospital, and subject to diagnostic testing and imaging (Fisher, Gottlieb, and Wennberg, 2004). These differences are unlikely to be the consequence of one hospital and not another having access to new technology, because every hospital in the sample is a teaching hospital; these differences primarily reflect "how" care is provided.

Table 2

Hospital-specific Measures of Mortality Outcomes and Medicare Expenditures for Five Large Hospitals, and Averages across 25 Hospitals, 1992–2004

	Average[a]	*Hosp. A*	*Hosp. B*	*Hosp. C*	*Hosp. D*	*Hosp. E*
Adj. 1 year mortality, 1992	0.346	0.366	0.415	0.326	0.361	0.291
Adj. 1 year mortality, 2004	0.297	0.250	0.305	0.289	0.356	0.294
Mortality diff.	–0.049	–0.116	–0.110	–0.037	–0.005	0.003
Adj. 1 year expenditures, 1992	19,991	14,785	16,492	22,961	18,799	15,425
Adj. 1 year expenditures, 2004	27,388	21,904	23,494	41,002	28,717	23,326
Expenditure diff.	7,397	7,119	7,001	18,041	9,918	7,901
PCI rate, 1992	0.27	0.33	0.17	0.23	0.23	0.43
PCI rate, 2004	0.47	0.59	0.43	0.42	0.35	0.53
Beta blocker, 1994/95	0.67	0.64	0.65	0.76	0.55	0.35
Aspirin (%), 1994/95	0.88	0.82	0.91	0.95	0.85	0.85
Effectiveness ratio	$12,455	$5,064	$5,251	$40,231	$163,633	Not defined

Source: Authors' calculations.
Notes: All prices in 2004 dollars. "Adj." means risj-adjusted as in Skinner, Staiger, and Fisher (2006). PCI is percuaneous coronary intervention.
[a]Averaged over all 25 hospitals with at least 250 AMI (acute myocardial infarction) patients in each year 1992–2004.

Comparative effectiveness analysis of the delivery system can identify efficiency at the system level. We illustrate the promise of this approach in Table 2. The first column provides summary measures of cost changes and outcomes changes for the 25 largest hospitals treating Medicare heart attack patients.[4] Since 1992, one-year mortality after heart attack has fallen by 4.9 per 100 heart attack patients (in medical terms, those who suffered an acute myocardial infarction). Most of this decline occurred in the early to mid-1990s; more recently mortality gains have slowed. Risk-adjusted inpatient Medicare expenditures for those with a heart attack rose by $7,397 during this period, implying a cost effectiveness of overall inpatient treatment of $12,455 per life-year (Cutler and McClellan, 2001).

This same calculation was then carried out for each of five large hospitals separately, shown in the remaining columns of Table 2. The hospitals are ranked by their own cost effectiveness, again defined as the change in expenditures divided by the change in risk-adjusted life expectancy. For the five hospitals chosen, individual cost-effectiveness ratios ranged from one that was highly favorable (A), just $5,064 per

[4] The results presented here are similar to those for the entire sample. We began with a 100 percent sample of Medicare Part A claims data from 1992–2004 to create a longitudinal cohort of fee-for-service enrollees, age 65 or over, coded with a new acute myocardial infarction, and risk-adjusted as in Skinner, Staiger, and Fisher (2006), limiting the sample to larger hospitals with at least 250 heart attack patients in any given year.

life-year, to a ratio of $163,633 for Hospital D, and to an undefined ratio for the least effective hospital (E), because expenditures rose while mortality did not change. While percutaneous coronary intervention (PCI) rates grew in all five hospitals (from 27 to 47 percent of patients on average), there was not a strong correlation between either levels or rates of this growth, whether among the five hospitals, or more generally among all hospitals. Clearly, the "cost effectiveness" of each hospital is determined by factors that have little to do with rationing care, and more to do with efficient organization of inpatient services and avoiding fragmented post-acute care once the patient has left the hospital.

At least one physician would seem to agree with our optimism for greater cost effectiveness of the delivery system. The surgeon Atul Gawande (2007) writes: "[T]he scientific effort to improve performance in medicine—an effort that at present gets only a miniscule of scientific budgets—can arguably save more lives in the next decade than bench science, more lives than research on the genome, stem cell therapy, cancer vaccines, and all the laboratory work we hear about in the news." In other words, simply learning how to better use what we already have may prove more valuable for patient health than new scientific discovery.

Conclusion

Comparative effectiveness analysis may appear inadequate to the task of taming healthcare cost growth in the U.S. The Patient-Centered Outcomes Research Institute, the nonprofit private entity created by the 2010 healthcare reform legislation, cannot even consider costs in its findings, as Congress prohibited its use of "a dollars-per-quality adjusted life year (or similar measure . . .) as a threshold to establish what type of health care is cost effective or recommended" (Garber and Sox, 2010).

But this view is too narrow. Comparative effectiveness research adds to the public knowledge-base about what works in healthcare and what doesn't, as Garber and Sox (2010) have emphasized. The costs of such research may appear large relative to the current size of the budget of the National Institutes of Health, but these investments in scientific knowledge, like traditional investments in biomedical research, can yield substantial benefits by adding to the long-term stock of scientific knowledge (Murphy and Topel, 2006). Still, opponents of comparative effectiveness research raise two concerns: first, that in a world where each patient responds differently to a treatment, a move towards greater effectiveness studies would reduce welfare by ignoring the heterogeneity in benefits; and second, that such efforts "ration" care. Both views are overly simplistic.

While recognizing the inherent downside of "cookie-cutter" rules for treating patients, one can still find value in comparative effectiveness research. It's certainly true that a randomized study may report an average-effect of the treatment, averaged over subgroups of patients, and thus not tell us treatment effects about a specific patient being seen by a specific doctor (Groopman, 2010). But knowing

the average effect is better than the *status quo* of having no published knowledge—a lacuna. Nor in the absence of comparative effectiveness studies can one rely on Bayesian physician learning to converge towards the universal optimum; different physicians converge to very different decision rules, and they are unlikely to all be correct (Sirovich, Gallagher, Wennberg, and Fisher, 2008). History has repeatedly shown that decision making based solely on physician experience can be wrong, and sometimes with devastating consequences. Examples in breast cancer include radical mastectomy (which offered higher morbidity along with no survival benefit) and high-dose chemotherapy followed by bone-marrow transplants to rebuild the immune system. In both cases, some physicians loudly proclaimed that trials were unethical because it was so clear that the more aggressive treatment was superior, an argument that was only silenced when trials came along.

Fears about how comparative effectiveness research will ration care also appear shortsighted. Some patients will get *more* valuable care with effectiveness studies. But offering treatments without regard to value—whether chemotherapy, angioplasty, proton beam therapy, or others—simply means greater financial pressures in the public and private sector to ration care to other patients by cutting insurance coverage (Sack, 2011).

Can comparative effectiveness and cost-effectiveness research really help to moderate healthcare cost growth? Our answer is a guarded yes: the research provides necessary but not sufficient information to change the behavior of patients and providers. Comparative effectiveness research generates useful information to assist patient decision making. But it's not enough just to publish the research; the information must also reach those patients who have overly optimistic perceptions of treatment benefits (Rothberg et al., 2010). Nor is it enough to assume that comparative effectiveness research will change physician behavior. One recent study found no benefit from vertebroplasty, a surgical procedure that injects cement into the spine for stabilization; still, one radiologist declared that, despite the study, he "will continue to recommend the surgery because he has seen its benefits" (Lazar, 2009). Even "black box" warnings about elevated heart attack risk from using the drug rosiglitazone for diabetes—that is, warnings on the package surrounded by a black box that is intended to emphasize the concern—led to only modest reductions in its use for some regions of the United States (Shah, Montori, Krumholz, Tu, Alexander, and Jackevicius, 2010).

Clearly, lack of research is not the only obstacle standing in the way of using comparative effectiveness research to reduce healthcare costs. The inability or unwillingness of providers and policymakers to use the information gleaned from comparative effectiveness research to make actual changes in reimbursement or patient cost-sharing may be just as important. Despite genuine efforts by Medicare officials to use cost-effectiveness analysis to determine reimbursement and coverage decisions in the Medicare program, Congress has been unwilling to do so. In court, insurance contracts are often interpreted in favor of the insured, and courts are reluctant to use published scientific literature to make rulings about what should be covered and what should not (Ferguson, Dubinsky, and Kirsch, 1993). Given

that Medicare reimburses without regard to the underlying value of health gains, or health gains relative to alternative treatments, it becomes very difficult for a private insurer, especially a single insurer, to take the lead on applying comparative effectiveness research (Chandra and Skinner, forthcoming).

Still, both private and public insurers might make more widespread use of comparative effectiveness research to determine patient cost-sharing based on the efficacy of a drug, therapy, or device. Private insurers are not forbidden from using "value-based" insurance design that lowers copayments and coinsurance for proven treatments and raises prices to patients (and perhaps lowers prices to providers) for procedures that are of marginal value in comparative effectiveness research (for example, Chernew, Rosen, and Fendrick, 2007). The key point, though, is that comparative effective research be used to nudge patients rather than to disallow coverage entirely, minimizing concerns about "rationing."

A more ambitious approach would use "dynamic pricing"—that is, Medicare would pay providers more for treatments with demonstrated superiority, and the same for two treatments with identical outcomes (Pearson and Bach, 2010). This switch would move away from the binary cover/not cover decision, but would also require substantial changes in law and political processes that could (unfairly) invoke cries of rationing.

One area where cost-effectiveness analysis may prove to be particularly potent is in evaluating the relative efficiency of different delivery systems—here, effectiveness analysis isn't being used to evaluate narrow scientific discoveries (drugs or procedures), but to direct how care is delivered. The efforts of the 2010 U.S. healthcare legislation to encourage "accountable care organizations" could in theory help to encourage greater attention to the cost effectiveness of healthcare *systems*; shared-saving "bonuses" would be provided to healthcare organizations that are able to provide high-quality care at lower costs (as discussed in more detail in the paper by McClellan in this symposium).

This in turn would presumably increase demand for learning about efficient institutional organization—such as weekend drop-in clinics rather than emergency room care—as well as cost-efficient procedures. Estimates from the literature on geographic variation in health spending suggest that, at a minimum, 20 percent of the $2.5 trillion spent by the United States on health care could be saved if cost-effectiveness research guided the redesign of inefficient healthcare systems (Skinner, Fisher, and Wennberg, 2005; Buntin and Cutler, 2009).

Over the medium- and long-term, as healthcare spending continues to rise, the financial pressure to consider such system-level cost effectiveness will become colossal. The implausibility of the marginal tax rates needed to finance government-provided health insurance—reaching 70 percent or more by 2060 (as discussed in Newhouse, 2010; Baicker and Skinner, 2011)—leads one to question not whether a fundamental shift in cost-growth will occur, but when. Comparative effectiveness research and its half-sibling cost-effectiveness research will provide a solid foundation for reform, once politicians and voters understand how dismal is the alternative.

■ *Chandra and Skinner acknowledge support from the National Institute of Aging (P01 AG19783). We thank without implicating Elliott Fisher, Victor Fuchs, Dana Goldman, and Tomas Philipson for conversations that have greatly influenced our thinking on this topic, and to David Autor, Alan Garber, Chad Jones, John List, and Timothy Taylor for insightful comments.*

References

Adams, Christopher P., and Van V. Brantner. 2006. "Estimating the Cost of New Drug Development: Is It Really $802 Million?" *Health Affairs,* 25(2): 420–28.

Almond, Douglas, Joseph Doyle, Amanda Kowalski, and Heidi Williams. 2010. "Estimating Marginal Returns to Medical Care: Evidence from At-Risk Newborns." *Quarterly Journal of Economics,* 125(2): 591–634.

Baicker, Katherine, and Amitabh Chandra. 2010. "Uncomfortable Arithmetic—Whom to Cover versus What to Cover." *New England Journal of Medicine,* 362 (2): 95–97.

Baicker, Katherine, and Jonathan S. Skinner. 2011. "Health Care Spending Growth and the Future of U.S. Tax Rates." NBER Working Paper 16772.

Basu, Anirban, and Tomas Philipson. 2010. "The Impact of Comparative Effectiveness Research on Health and Health Care Spending." NBER Working Paper 15633.

Boden, William E., Robert A. O'Rourke, Koon K. Teo et al., the COURAGE Trial Research Group. 2007. "Optimal Medical Therapy with or without PCI for Stable Coronary Disease." *New England Journal of Medicine,* 356(15): 1503–16.

Brenner, David J., and Eric J. Hall. 2007. "Computerized Tomography—An Increasing Source of Radiation Exposure." *New England Journal of Medicine,* 357(22): 2277–84.

Buntin, Melinda Beeuwkes, and David M. Cutler. 2009. "The Two Trillion Dollar Solution: Saving Money by Modernizing the Health Care System." *Center for American Progress.* Available at: http://www.americanprogress.org/issues/2009/06/2trillion_solution.html.

Chandra, Amitabh, and Jonathan Skinner. Forthcoming. "Technology Growth and Expenditure Growth in Health Care." *Journal of Economic Literature.*

Chandra, Amitabh, and Douglas O. Staiger. 2007. "Productivity Spillovers in Healthcare: Evidence from the Treatment of Heart Attacks." *Journal of Political Economy,* 115(February): 103–140.

Chernew, Michael E., Allison B. Rosen, and A. Mark Fendrick. 2007. "Value Based Insurance Design." *Health Affairs,* 26(2): 195–203.

Concato, John, Nirav Shah, and Ralph I. Horwitz. 2000. "Randomized, Controlled Trials, Observational Studies, and the Hierarchy of Research Designs." *New England Journal of Medicine,* 342(25): 1887–92.

Congressional Budget Office. 2007. "Research on the Comparative Effectiveness of Medical Treatments: Issues and Options for an Expanded Federal Role." http://www.cbo.gov/ftpdocs/88xx/doc8891/Frontmatter.1.2.shtml.

Cutler, David M. 2004. "Your Money or Your Life: Strong Medicine for America's Health Care System." New York: Oxford University Press.

Cutler, David M., and Mark McClellan. 2001. "Is Technological Change in Medicine Worth It?" *Health Affairs,* 20(5): 11–29.

DiMasi, Joseph A., Ronald W. Hansen, and Henry G. Grabowski. 2003. "The Price of Innovation: New Estimates of Drug Development Costs." *Journal of Health Economics,* 22(2): 151–85.

Dorsey, E. Ray, and David O. Meltzer. 2010. "The Economics of Comparative Effectiveness Research." *Neurology,* 75(6): 492–493.

Ferguson, John H., Michael Dubinsky, and Peter J. Kirsch. 1993. *Journal of the American Medical Association,* 269(16): 2116–21.

Fisher, Elliott, Dan Gottlieb, and David Wennberg. 2004. "Implications of Regional Differences in Spending." *Annals of Internal Medicine,* 140(2): 148–49.

Garber, Alan M., and Charles E. Phelps. 1997. "Economic Foundations of Cost-Effectiveness Analysis." *Journal of Health Economics,* 16(1): 1–31.

Garber, Alan M., and Jonathan Skinner. 2008. "Is American Health Care Uniquely Inefficient?" *Journal of Economic Perspectives,* 22(4): 27–50.

Garber, Alan M., and Harold C. Sox. 2010. "The Role of Costs in Comparative Effectiveness Research." *Health Affairs,* 29(10): 1806–07.

Garber, Alan M., and Sean R. Tunis. 2009. "Does Comparative Effectiveness Research Threaten Personalized Medicine?" *New England Journal of Medicine,* 360(19): 1925–27.

Gawande, Atul. 2007. *Better: A Surgeon's Notes on Performance.* Metropolitan Books.

Groopman, Jerome. 2010. "Health Care: Who Knows 'Best'?" *New York Review of Books,* February 11.

Hadorn, David. 1991. "Setting Health Care Priorities in Oregon: Cost-Effectiveness Meets the Rule of Rescue." *JAMA,* 265(17): 2218–25.

Hartwell, D., J. Colquitt, E. Loveman, A. J. Clegg, H. Brodin, N. Waugh, P. Royle, P. Davidson, L. Vale, and L. MacKenzie. 2005. "Clinical Effectiveness and Cost-Effectiveness of Immediate Angioplasty for Acute Myocardial Infarction: Systematic Review and Economic Evaluation." *Health Technology Assessment,* 9(17): 1–4.

Hawker, Gillian, Jun Guan, Andy Judge, and Paul Dieppe. 2008. "Knee Arthroscopy in England and Ontario: Patterns of Use, Changes over Time, and Relationship to Total Knee Replacement." *Journal of Bone and Joint Surgery,* 90A(11): 2337–45.

Jena, Anupam, and Tomas Philipson. 2010. "Endogenous Cost-Effectiveness Analysis in Health Care Technology Adoption." NBER Working Paper 15032.

Kim, J. Yong, Aaron Shakow, Kedar Mate, Chris Vanderwarker, Rajesh Gupta, and Paul Farmer. 2005. "Limited Good and Limited Vision: Multidrug-Resistant Tuberculosis and Global Health Policy." *Social Science and Medicine,* 61(4): 847–59.

Lazar, Kay. 2009. "The Long Run: Cutting Question." *The Boston Globe,* August 24, G.6, Living Arts.

Lee, Chris P., Clenn M. Chertow, and Stefanos A. Zenios. 2009. "An Empiric Estimate of the Value of Life: Updating the Renal Dialysis Cost-Effectiveness Standard." *Value in Health,* 12(1): 80–87.

Marchione, Marulynn. 2010, "$93,000 Cancer Drug: How Much is a Life Worth?" Associated Press, September 26.

Martin, Daniel F., Maureen G. Maguire, and Stuart L. Fine. 2010. "Identifying and Eliminating the Roadblocks to Comparative-Effectiveness Research." *New England Journal of Medicine,* 363(2): 105–107.

McClellan, Mark, Barbara McNeil, and Joseph Newhouse. 1994. "Does More Intensive Treatment of Acute Myocardial Infarction Reduce Mortality?" *JAMA,* 272(11): 859–66.

McGrath Paul D., David E. Wennberg, John D. Dickens Jr., Andrea E. Siewers F. Lee Lucas, David J. Malenka, Merle A. Kellett Jr., and Thomas J. Ryan Jr. 2000. "Relation between Operator and Hospital Volume and Outcomes following Percutaneous Coronary Interventions in the Era of the Coronary Stent." *JAMA,* 284(24): 3139–44.

Meltzer, David. 1997. "Accounting for Future Costs in Medical Cost-Effectiveness Analysis." *Journal of Health Economics,* 16(1):33–64.

Morgan, Matthew W., Raisa B. Deber, Hilary A. Llewellyn-Thomas, Peter Gladstone, R. J. Cusimano, Keith O'Rourke, George Tomlinson, and Alan S. Detsky. 2000. "Randomized, Controlled Trial of an Interactive Videodisc Decision Aid for Patients with Ischemic Heart Disease." *Journal of General Internal Medicine,* 15(10): 685–93.

Mosley, J. Bruce, et al. 2002. "A Controlled Trial of Arthroscopic Surgery for Osteoarthritis of the Knee." *New England Journal of Medicine,* 347(2): 81–8.Murphy, Kevin M., and Robert H. Topel. 2006. "The Value of Health and Longevity." *Journal of Political Economy,* 114(5): 871–904.

Murphy, Kevin M., and Robert H. Topel. 2006. "The Value of Health and Longevity." *Journal of Political Economy,* 114(5): 871–904.

Nathan, David M. 2010. "Time for Clinically Relevant Comparative Effectiveness Studies in Type 2 Diabetes." *Annals of Internal Medicine,* December. Available at: http://www.annals.org/content/early/2010/12/06/0003-4819-154-2-201101180-00303?col3.

Newhouse, Joseph P. 2010. "Assessing Health Reform's Impact on Four Key Groups of Americans." *Health Affairs,* 29(9): 1714–24.

Nord, Erik, Jeff Richardson, Andrew Street, Helga Kuhse, and Peter Singer. 1995. "Who Cares about Cost? Does Economic Analysis Impose or Reflect Social Values?" *Health Policy,* 34(2): 79–94.

Pearson, Steven D., and Peter B. Bach. 2010. "How Medicare Could Use Comparative Effectiveness Research in Deciding on New Coverage and Reimbursement." *Health Affairs,* 29(10): 1796–1804.

Perlroth, Daniella J., Dana P. Goldman, and Alan Garber. 2010. "The Potential Impact of Comparative Effectiveness Research on U.S. Health Care Expenditures. *Demography,* 47(Suppl.): S173–S190.

Polinski, Jennifer M., Philip S. Wang, and Michael A. Fisher. 2007. "Medicaid's Prior Authorization Program and Access to Atypical Antipsychotic Medications." *Health Affairs,* 26(3): 750–60.

Robberstad, Bjarne. 2005. "QALYs vs DALYs vs

LYs Gained: What are the Differences, and What Difference Do They Make for Health Care Priority Setting?" *Norsk Epidemiologi,* 15(2): 183–91.

Rothberg, Michael B., Senthil K. Sivalingam, Javed Ashraf, Paul Visintainer, John Joelson, Reval Klippel, Neelima Vallurupalli, and Marc J. Schweiger. 2010. "Patients' and Cardiologists' Perceptions of the Benefits of Percutaneous Coronary Intervention for Stable Coronary Disease." *Annals of Internal Medicine,* 153(5): 307–13.

Sack, Kevin. 2011. "Feeling Budget Pinch, States Cut Insurance." *The New York Times,* March 2.

Shah, Nilay D., Victor M. Montori, Harlan M. Krumholz, Karen Tu, Caleb Alexander, and Cynthia Jackevicius. 2010. "Responding to an FDA Warning—Geographic Variation in the Use of Rosiglitazone." *New England Journal of Medicine,* 363(22): 2081–84.

Sirovich, Brenda, Patricia M. Gallagher, David E. Wennberg, and Elliott S. Fisher. 2008. "Discretionary Decision Making by Primary Care Physicians and the Cost of U.S. Health Care." *Health Affairs,* 27(3): 813–23.

Skinner, Jonathan S., Elliott S. Fisher, and John Wennberg. 2005. "The Efficiency of Medicare." In *Analyses in the Economics of Aging,* ed. David A. Wise, 129–60. Chicago: University of Chicago Press.

Skinner, Jonathan, Douglas Staiger, and Elliott Fisher. 2006. "Is Technological Change in Medicine Always Worth It? The Case of Acute Myocardial Infarction." *Health Affairs,* 25(2): w34–247.

Sunstein, Cass, and Richard Thaler. 2008. *Nudge: Improving Decisions about Health, Wealth, and Happiness.* Yale University Press.

Taubes, Gary. 2007. "Do We Really Know What Makes Us Healthy?" *New York Times Magazine,* September 16.

Temel, Jennifer S., et al. 2010. "Early Palliative Care for Patients with Metastatic Non–Small-Cell Lung Cancer." *New England Journal of Medicine,* 393(8): 733–42.

U.S. Congress. 2010. *Patient Protection and Affordable Care Act.* H.R. 3590, 111th Congress of the United States, Second Session.

Weinstein, Milton C., and Jonathan A. Skinner. 2010. "Comparative Effectiveness and Health Care Spending—Implications for Reform." *New England Journal of Medicine,* 362(5): 460–65.

Weintraub, William S. et al. 2008. "Effect of PCI on Quality of Life in Patients with Stable Coronary Disease." *New England Journal of Medicine,* 359(7): 677–87.

Wennberg, David E., F. L. Lucas, John D. Birkmeyer, Carl E. Bredenberg, and Elliott S. Fisher. 1998. "Variations in Carotid Endarterectomy Mortality in the Medicare Population: Trial Hospitals, Volume, and Patient Characteristics." *JAMA,* 279(16): 1278–81.

Winkelmayer, Wolfgang C., Milton C. Weinstein, Murray A. Mittleman, Robert J. Glynn, and Joseph S. Pliskin. 2002. "Health Economic Evaluations: The Special Case of End-Stage Renal Disease Treatment." *Medical Decision Making,* 22(5): 417–30.