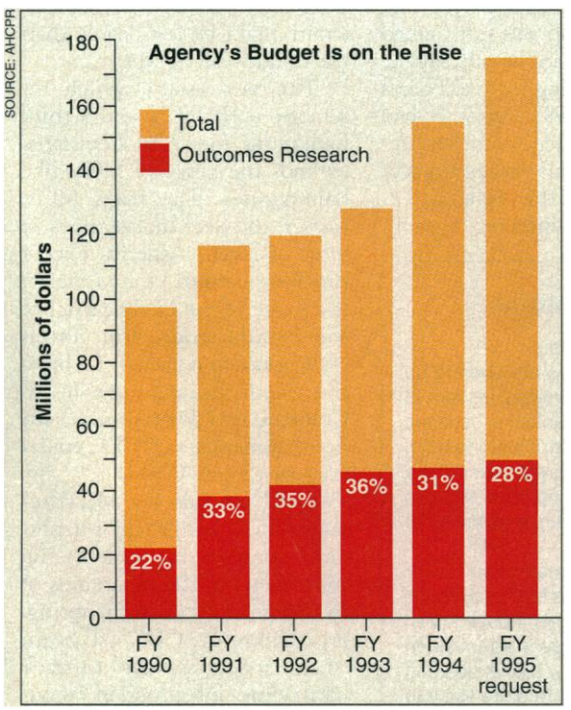
Measuring What Works in Health Care

Using health care records to measure outcomes was supposed to provide a quick and cheap alternative to clinical trials. But 5 years and $200 million later, critics are asking: Where's the beef?

Health care reformers who want to encourage physicians to use the most cost-effective treatments have a big problem: Physicians themselves, let alone government bureaucrats, often don't know which medical interventions work best. And even a relatively straightforward comparison—between a 10-cent aspirin and a $1,000 shot of a genetically engineered anticlotting drug, for example—requires a clinical trial costing tens of millions of dollars and lasting upward of a decade.

For more than 20 years, however, a group of epidemiologists has been arguing that some clues might be found relatively cheaply and easily in the records that have been accumulated by hospitals, insurers, and government health programs. Comb through these vast databases, they argued, and you might get a good idea of which medical interventions produced the best outcomes. In the late 1980s, with a combination of rising concern over health care costs and the advent of centralized electronic databases including millions of patients, the idea took off. In late 1989, Congress created the Agency for Health Care Policy and Research (AHCPR) largely to conduct such studies, known collectively as "outcomes research." In just 6 years its budget has climbed from $97 million to a planned $173 million, and President Clinton has proposed in his health care reform package to nearly quadruple the agency's current funding by 1998.



Mixed success. Spending on outcomes research is going up, but not as fast as AHCPR's overall budget.

But so far, outcomes research has yielded little that clinicians and policy makers can use to make rational decisions about health care. Instead, it has sparked a fierce debate between scientists over whether the approach itself is a cost-effective way of evaluating clinical techniques. Advocates say outcomes research is potentially cheaper and faster than clinical trials and can provide data on treatments that would otherwise never be evaluated. But critics argue that any research based on retrospective analysis of clinical records, whether they be Medicare claims databases, patient or hospital records, or tumor registries, is fatally flawed by hidden biases in the data.

After spending nearly $200 million on outcomes research (about one-third of the agency's budget goes for outcomes research; the rest is for health care policy research and clinical guidelines), AHCPR cannot point to a single case in which its database studies have changed general clinical practice. Just this month, its most definitive result—that "watchful waiting" is often a better option than surgery for benign prostate disease—was issued as a guideline to physicians. Even then, the researchers recommended a clinical trial to confirm their findings.

"A lot of money has been spent on nonrandomized outcomes research because the claim was made that it was going to give us reliable comparisons between the main effects of different treatments," says Oxford University epidemiologist Richard Peto. "It has utterly, totally, and predictably failed to do so." Peto argues that researchers cannot correct for the subtle reasons doctors choose one treatment over another for a particular patient. That bias, in turn, can undermine the entire premise of outcomes research.

Despite their differences, outcomes advocates and critics agree on some useful aspects of the research: It is clearly important to learn what doctors are actually doing in clinical practice, and wide variation in outcomes provides rich ground for further study, including generating hypotheses for clinical trials. Nevertheless, last summer AHCPR conceded that it had overemphasized the role of database analysis in its rules for the Patent Outcomes Reasearch Teams that conduct most of the agency's outcomes research. In the new rules, the teams no longer have to include database analysis, but can use a range of studies, including clinical trials.

The promise

Outcomes research would seem to have a lot going for it. Because it relies on nonrandomized "trials," either retrospective analysis of records or concurrent tracking of clinical practice, the data can be analyzed relatively cheaply and quickly. Unlike clinical trials, outcomes research doesn't require informed consent or permission from an institutional review board, and it usually doesn't interfere with the doctor-patient relationship.

Because outcomes research uses records from routine clinical care, its results are, in theory, relevant to every medical center in the country. It includes groups—the elderly, the poor, children, and minorities—that might not be widely represented in clinical trials. It compares medical techniques and technologies that are already in clinical use rather than experimental therapies. And outcomes research, as befits its name, focuses on patient outcomes rather than the type of intermediate physiological measures used in many clinical trials.

The premise of outcomes research, says Duke University epidemiologist David Eddy, is based on what he calls "natural experiments." For some conditions, the argument goes, treatments are so varied and the doctors' choices so unpredictable that the records approximate those derived from an arbitrary-assignment clinical trial. The idea, he said, was "just observe the experiment—mine the data—and you can get pretty good information" about what works and what doesn't.

That's what epidemiologist John Wennberg argued during the campaign to create AHCPR in the late 1980s. The rallying cry of "finding what works in medicine" was particularly appealing at that time because Medicare was tottering on the brink of bankruptcy. Wennberg, director of the Center for Evaluative Clinical Sciences at Dart-

mouth Medical School, had used database analysis to show wide—and presumably arbitrary—geographic variations in medical practice, and he became the most prominent advocate for a more scientific basis for health care decisions. He got the ear of Majority Leader George Mitchell (D–ME), and through him, the U.S. Congress.

In 1989, Congress passed a bill that created AHCPR as a stand-alone agency within the Department of Health and Human Services, replacing (and absorbing) the existing National Center for Health Services Research. The agency was given clear marching orders: AHCPR "shall conduct and support research...on the use of claims data and data on clinical and functional status of patients in determining the outcomes, effectiveness and appropriateness of...treatments," as well as set database standards to ensure that data will be useful.

"Congress was very impatient with the rate at which results were getting into practice," explains Richard Greene, director of the agency's Center for Medical Effectiveness Research. "They created AHCPR to do research on outcomes that they thought would be fast and cheap; and, because they weren't even willing to wait for that, they set up work on guidelines based on existing research." Driving this, he says, "were some researchers who had a dream that everybody's claim would be in this big computer and the answers will be all there."

The problems

Unfortunately, Greene says, that hasn't been the case. The limitations of the approach became clear as researchers applied their analytical skills to real head-to-head comparisons of treatments.

Perhaps the best-known example is a 1989 study of benign prostate disease by Wennberg himself, along with a team of U.S., British, Canadian, and Danish researchers. The researchers used claims data from the national health care systems of Canada, Denmark, and the United Kingdom to compare mortality rates for men treated in one of two ways for prostate disease—either traditional invasive surgery or a relatively new technique, called transurethral resection of the prostate (TURP), that did not involve open surgery. Although most urologists considered TURP to be the less traumatic operation, the researchers found that men who had the TURP treatment were significantly more likely to die or undergo another prostate operation within 8 years than those who had the more invasive surgery.

Wennberg and his colleagues were puzzled. They worried that statistical artifacts had distorted their analysis, perhaps because doctors tended to use the TURP procedure on patients who were sicker and less able to handle open surgery. But when attempts to compensate for such artifacts didn't significantly change their results, they published their original conclusion.

Two years late another team, headed by Yale Medical School epidemiologists John Concato and Alvan Feinstein, revealed the underlying flaws. Using patients' medical records from their own doctors, which required time-consuming individual review, the researchers found no difference in long-term mortality. Wennberg's results had apparently been skewed by inadequate classification in the Canadian data of the severity of the patients' other conditions. Wennberg says the true meaning of the data is still being debated.

Critics of outcomes research say such problems are inherent in the approach. Outcomes analyses, says Concato, are based on data "collected for a different purpose—for payments—and that has the potential for distortion." Peto concurs: "The patient characteristics that lead doctors to choose a particular treatment make most nonrandomized outcomes analysis untrustworthy." Asking doctors why they chose one treatment over another will not help, he says, because "you still don't know why this patient was referred to this doctor."

PROPOSED TOPICS FOR OUTCOMES RESEARCH

Condition	Treatment Options
Nancancerous uterus	Surgery vs. hormones vs. drugs vs. watchful waiting
Angina pectoris	Bypass surgery vs. angioplasty vs. drugs
Peripheral vascular diseases	Bypass vs. angioplasty vs. medical management
Cataracts	Lens extraction vs. watchful waiting
Arthritis of hip, knee	Surgery vs. medical management
Benign prostatic hyperplasia	Surgery vs. balloon dilation vs. drugs vs. microwave diathermy vs. waiting
Herniated disc	Surgery vs. various management strategies
Atherosclerosis of carotid artery	Carotid endarterectomy vs. aspirin

Source: Wennberg/*Annals of New York Academy of Sciences* (in press)

The middle ground

For Peto, the bottom line is clear: Large-scale randomized evidence, whether from clinical trials or meta-analyses of previous trials, has proved useful to clinical practice, and outcomes research has not. Investing in a lot of outcomes research "is worse than just destroying the money," he argues, "because it gives the illusion of information."

Congress has also begun to wonder what it's getting for its investment. In July, the congressional Office of Technology Assessment (OTA) is scheduled to release the results of an 18-month study that provides the first critical analysis of the AHCPR's outcomes research experiment. As one congressional aide says, "I think Congress was led to believe that we would be able to use databases in lieu of clinical trials. In that, it was sold a bill of goods." Next month the New York Academy of Sciences expects to release the proceedings of a conference held last March that analyzed the relative merits of outcomes research and clinical trials.

Outcomes researchers argue that, whatever the limitations of their methodology, it is better than nothing. And nothing, they say, is what they would get in many cases if they relied on clinical trials. Even Peto agrees that retrospective database analysis is often the only way to get data on treatments with rare complications that carry an extreme relative risk, such as acute leukemia as a side effect of some cancer therapies. And outcomes advocates point out that some treatments are best evaluated in terms of "soft outcomes," such as quality of life, whereas clinical trials work best with more precise endpoints, such as mortality.

The debate over database analysis has already affected how AHCPR spends its money. "What's clear is that it cannot alone be used to say if treatment A is better than treatment B," says Greene. "It's no substitute for clinical trials." Increasingly, he says, database analysis is used to supplement such tools as case-control studies, analysis of hospital records and tumor registries, meta-analysis of previous studies, and even clinical trials. The problem, however, is that such traditional techniques often are not much cheaper or faster than a clinical trial. And hidden biases can invalidate even these techniques.

But as the debate continues, supporters of outcomes research believe they hold a trump card—the nation's need to identify effective health care strategies. And although proponents have toned down their claims for the technique, they remain confident of its value to clinical practice. Today, says Harvard outcomes researcher Barbara McNeil, "the real aim is to ask: Can we identify variations of practice for further study? Can we develop better measures [of patient outcome] than mortality?" That approach is a long way from answering the question of "what works," but it may offer the best chance of success.

–Christopher Anderson