We have the technology, and the vested interest, to build a better, stronger, fitter health system.



THE DOCTOR DOCTORS

PITT RESEARCHERS ARE FINDING OUT WHAT WORKS BEST IN MEDICINE. CAN THEY HELP SOLVE A $2.3 TRILLION QUESTION, TOO?

BY REID R. FRAZIER

ILLUSTRATIONS BY JESSE LENZ

One night in the 1980s, Mark Roberts phoned a transplant surgeon about a very sick patient in the ICU at what was then New England Deaconess Hospital in Boston, where Roberts was a resident. As a young MD, Mark Roberts had a habit of questioning his betters. A few years earlier, he'd earned a master's degree in public policy and health policy while in medical school. (As a third-year med student, he once drew up a schematic model on the options for a patient with an aortic aneurysm, after watching attending physicians and residents debate whether or not to operate, without basing their arguments on any scientific evidence.)

ILLUSTRATION: JESSE LENZ. PHOTO OF THE 6 MILLION DOLLAR MAN: CORBIS

The subject of the late-night phone call was a woman who'd had liver disease for 15 years and was awaiting a liver transplant. Roberts quizzed the transplant surgeon, Roger Jenkins, on why she hadn't been operated on sooner.

"I said, 'Roger, if you had a *Star Trek* replicator, and you could build an ABO-compatible liver of the right size for this woman, when should you have transplanted her?' And his answer was, 'I have no idea,'" Roberts says. Ideally, would it have been six months earlier? A year? There was no answer. No one had ever researched it, in part because there was no way to study the problem through traditional methods.

The question stayed with Roberts. With the help of Jenkins, he began working on a "virtual" trial to come up with an answer. Roberts produced a computer model of end-stage liver disease. The model took into account life expectancy with or without transplant, the likelihood of organ rejection, and the chaotic ways liver disease progresses in individuals—sometimes slowly, sometimes quickly. Roberts borrowed an algorithm designed for industrial plants. Models like Roberts' are now used routinely to determine which patients should get priority in liver allocation.

Questioning typical care decisions—like when a transplant should be done—might seem like common sense. But in many corners of the medical world, doctors employ remedies without much evidence that a chosen treatment is the best one.

For example, Roberts points out that there are 19 different drugs he can prescribe for high blood pressure. But they haven't been compared to one another across a spectrum of patients. So when he sees a hypertensive patient, Roberts is left without enough data to know which of the 19 drugs are best and for which patients.

Roberts—now a University of Pittsburgh professor of medicine and of industrial engineering, as well as professor and chair of health policy and management in Pitt's Graduate School of Public Health—has spent his career trying to find ways to answer these seemingly simple questions. He's designed models to study antiretroviral therapy for HIV patients, determine the effectiveness of alcoholism interventions, and figure out the optimal time to administer vaccine during an outbreak of flu.

He uses computer models because standard research techniques can't answer many of the questions he would like to answer. Randomized controlled trials, the gold standard for clinical medicine, are costly and time consuming. They work very well when asking how a new drug compares to placebo. But they're ill suited to find out how a drug might compare to 18 others in a diverse set of patients.

"Most of medical science is based on randomized controlled trials, which simply answer, *Is A better than B?* [with B being a placebo, typically]. But many of the questions we have in real clinical medicine aren't that dichotomous," says Roberts, a general internist. "They're, *What's the best time to intervene?* Or, *Under what conditions is this better than that?* It's much more continuous, much more calculus."

Real-life medicine is filled with questions that might appear simple but are, in fact, very hard to answer: What's the best way to avoid health care–associated infection? How do you best address obesity, hypertension, and diabetes in underserved populations? Should you treat back pain with surgery, physical therapy, or some combination thereof? These are examples of problems addressed by comparative effectiveness research (CER), a field that uses epidemiology, biostatistics, and health policy science to compare treatments and assess which work best, when, and for whom. (The field is new enough that it's been assigned several names, including "patient-centered outcomes research" and, in the UK, "clinical effectiveness.")

The many medical questions addressed by CER have taken on new urgency in the past few years as society has come to terms with budget-busting health care costs. Finding out which treatments work best should help us to stop paying for treatments that are inferior. That's why President Barack Obama is trumpeting CER's promise, and why Congress allocated $1.1 billion of stimulus funding to pay for research in CER.

COMMON SENSE GETS FUNDING

Since it was founded in 1990, the U.S. Agency for Healthcare Research and Quality (AHRQ, or "Ark," in the lingo) has been the major source of funding for comparative-effectiveness science. The agency is relatively small—its 2009 budget was about 1 percent of that of the National Institutes of Health (NIH), which funds most basic biomedical research. That will start to change. Congress' appropriation will essentially double the agency's budget.

Why the sudden attention? The biggest reason is economics. The United States spent $2.3 trillion on health care in 2008. With 32 million more people promised health coverage under the Patient Protection and Affordable Care Act, health care costs are expected to climb. Harvard University's Michael Chernew has called the state of health care costs "financial Armageddon."

In this gloomy picture, CER offers a ray of hope that we can eliminate unnecessary or ineffective treatments. For instance, new cancer drugs like cetuximab have been shown to work in colorectal cancer patients who have a normal expression of their KRAS gene, but not in those with a mutation in the gene. By simply giving the drug to only the appropriate patients, we could save $600 million in health care costs, research suggests.

A second, related issue is quality. Several well-known studies by the Dartmouth Institute for Health Policy and Clinical Practice found that health outcomes were no better in the highest-spending regions of the country than in those that spent the least. This phenomenon has been linked to the medical incentive system in America—more procedures, scans, surgeries, and prescriptions mean doctors and health care systems get paid more. Yet when doctors do more, that doesn't necessarily translate to better care. A well-known study of 6,700 patients found that doctors followed "recommended care" only

Roberts points out that there are 19 different drugs he can prescribe for high blood pressure. But they haven't been compared to one another across a spectrum of patients. So when he sees a hypertensive patient, Roberts is left without enough data to know which of the 19 drugs are best and for which patients.



Cleanup in the medicine aisle: A lack of convincing data may cause doctors to follow recommended-care guidelines only 55 percent of the time, according to one study. By revealing the best treatments, researchers think we could cut health care costs, which eat up nearly one-fifth of the U.S. economy.

55 percent of the time. This may be because guidelines are often too vague to be of much use in many care decisions. So doctors rely on their own experience, the advice of colleagues, and other unscientific information when making medical decisions.

The rapid advance of medical discovery is a third area where CER can be of use. New treatments often come online before adequate research establishes how they compare to other options. Drug-eluting stents for angioplasty, for example, were on the market for several years before the completion of any studies comparing them to older, bare-metal stents. Several studies found no difference in mortality. In some studies, drug-eluting stents were associated with slightly higher short-term mortality rates.

Add all these factors together and you get an expensive, inefficient health system. Most disconcerting, patients and doctors often make critical decisions without much help from the scientific data.

Hearing that an expensive test is unnecessary may be fine in the abstract, but when the patient is your wife or son, you're less likely to be receptive to the message.

This is where CER can help, says Wishwa Kapoor, chief of Pitt's Division of General Internal Medicine and vice chair of its Department of Medicine, as well as Falk Professor of Ambulatory Care and a professor of health policy and management. Kapoor has been comparing the effectiveness of treatments and strategies for the better part of 20 years. He has studied the best ways to treat pneumonia, administer antibiotics, and manage ICUs. It is in the gray areas of medical decision-making that advanced statistical modeling and measurement techniques can provide guidance, he notes.

"Sometimes there are no studies available to tell you what to do, but you can look at the evidence, add it all up, and come up with the best strategies for a particular condition," Kapoor says. "We deal with a lack of evidence all the time, but sometimes, you have to make a decision."

A MOVEMENT AFOOT

In the midst of this national CER effort, the University of Pittsburgh is moving quickly to establish itself in the field. Senior Vice Chancellor for the Health Sciences and Dean of the School of Medicine Arthur S. Levine chairs an NIH working group on the emerging field. The University has brought in several high-profile recruits to build up its CER operation. These include Sally Morton, former head of statistics for RAND and vice president for statistics and epidemiology at the research group RTI International. Morton is professor and chair of biostatistics in Pitt's Graduate School of Public Health. The University also recruited Everette James, who recently served as Pennsylvania's Secretary of Health and is an expert on health policy and cost containment. James has been tapped to lead Pitt and UPMC's compliance with health care reform, as well as to launch major research initiatives aimed at CER.

There are already hundreds of ongoing studies at Pitt that address comparative effectiveness, relating to everything from middle-ear infections to chemotherapy. With its access to UPMC's health network and electronic medical records, Pitt is poised to become a health services laboratory, says Steven Shapiro, Jack D. Myers Professor and Chair of Medicine and UPMC's chief medical and scientific officer.

One advantage Pittsburgh has over other academic medical centers, Shapiro says, is the presence of UPMC's integrated health insurance plan. This means that implementing less intensive—and thus less costly—treatments, say, watchful waiting instead of an MRI, won't hurt the system's overall budget.

"Being an integrated payer-provider allows us to focus exclusively on the best outcomes for the patient. That is, we only need to keep the patient healthy without concern for the 'health of our bottom line' in the current and unsustainable system of reimbursement."

CALLED TO SERVE

The daughter of physicians, Morton chose a career that allowed her to use math to solve real-world problems. The PhD is now a nationally recognized expert in using statistics to make health policy decisions. Morton has been asked to serve on several Institute of Medicine panels, including one with the (somewhat ambitious) title, "Knowing What Works in Health Care." Among its recommendations, the group called for the creation of an autonomous organization to evaluate the effectiveness of various interventions. A few months after the report came out, Congress created just such an entity, the public-private Patient-Centered Outcomes Research Institute (PCORI). Morton is quick to point out that "many others" had previously called for the creation of PCORI but admits that when the legislation came out, she felt a sense of pride. "It felt like, whoa—it really mattered what we said, that people were listening. As a statistician working in health policy, I'd never felt so close to decision-making."

Morton's next call from the IOM came during the beginning of the health care debate, in early 2009. The Institute was quickly convening a group of experts to create a list of the "Top 100" problems in health care for effectiveness researchers to address.

The call to serve felt profound to Morton. "Every single person who was asked to sit on that committee said, 'Yes,'" she says. In a few adrenaline-fueled months, the committee crafted a report, convened public hearings, and sent its draft for review. Typically, such a report takes 18 months to prepare. This one took five.

"We really thought the time was now—this was during health care reform," Morton says. "For a lot of us, we'd been working on health care for all our careers. To suddenly feel that things were changing, and we could contribute to that change in a positive way, was very uplifting."

Where is all of this headed? Morton has been watching with great interest the emergence of electronic medical records. She believes they have the power to transform how medical science is undertaken but need to be evaluated with caution. "There are people who have actually said that clinical trials should become obsolete," Morton says. "When you have a hundred million records, why do we need a trial?" Yet observational data like electronic medical records are more prone to bias than randomized experimental data, Morton notes. "The combination of EMRs and randomized data is where the strength lies," she says.

Kapoor and others have already embraced the computational potential of electronic records. He and his colleagues mined electronic records to study the use of spiral CT scans in the diagnosis of pulmonary embolism, a common cause of preventable mortality in

hospitals. The group wanted to know whether the proliferation of the spiral scan, which provides more detailed images than standard CT systems, actually improved outcomes. Combing through millions of medical records, the researchers isolated 38,000 hospital discharges for pulmonary embolism. Their conclusion: Spiral CT scans caught pulmonary embolism earlier, resulting in lower death rates for patients.

FACING UP TO THE COST QUESTION

In its landmark 2010 health care legislation, Congress explicitly forbade its new creation, PCORI, from funding studies that would compare the cost effectiveness of different therapies. The decision was made after extensive lobbying from the drug and device industries, which argued that the studies would reduce incentives for innovation. Some politicians warned that considering cost effectiveness would lead to care rationing.

Yet around the world, countries with national health insurance routinely consider cost effectiveness when deciding whether to pay for treatment. (This is one reason health care costs in most European countries are much lower than in the United States.) Among the most advanced of these systems is the United Kingdom's, whose National Health Service uses cost-effectiveness analyses (yes, there's an acronym: they're CEAs) to approve therapies for coverage. If a treatment cost falls above a threshold, around $50,000 per quality-adjusted life year (essentially an extra year of good health), the system might decide against paying for it. This approach has drawn criticism—the NHS has had to go back on some decisions to deny some forms of cancer treatment. Last year, the UK's new government decided to overhaul the health system and may yet scale back reliance on cost effectiveness.

A national health policy that ignores cost may simply be wishful thinking, says James, Pitt's recently appointed associate vice chancellor for health policy and planning. James ought to know. While leading the Pennsylvania Department of Health, James saw firsthand how hard it was to achieve meaningful cuts in health care costs.

On an overcast fall morning, James drops a photocopy handout on a desk in his sparsely decorated Scaife Hall office. "This is why we can't avoid cost," he says, putting his finger on a photocopy of a graph titled "How High Could It Go?" It shows the cost of health care in the United States from the 1960s until today, with a line graph portraying growth in private insurance, Medicare and Medicaid, and out-of-pocket expense costs. The graph looks like a staircase that gets steeper and steeper as it nears the top.

Under James, Pennsylvania set up a national model to reduce the prevalence of hospital-acquired infections (an effort led by many Western Pennsylvania hospital leaders, including those at UPMC). James tried, unsuccessfully, to enact measures proven to lower health costs in other places, like liability reform and apology legislation, which encourages open dialogue between doctors and patients in the event of medical error. Prior to taking the state post, James worked at a Washington, D.C., law firm, serving as counsel to both for-profit and nonprofit insurance companies. He has been on both sides of the public-private divide in health care and has a perfect vantage point from which to see the system's bottlenecks.

Part of James' job is to guide Pitt as it navigates around those bottlenecks. He will help create a data infrastructure to support CER research, giving investigators ways to access a warehouse of medical data. James acknowledges the difficulties of building the data architecture and legal agreements for CER while preserving patient privacy. But they are worth it, he says.

"It may be a headache, but it's going to be necessary to have a large enough sample size and enough utilization and outcome data to make the findings relevant," he says. First, James plans to reorient the hundreds of CER-related research projects across Pitt's schools of the health sciences under one umbrella for better access to funding, resources, and expertise. He has been logging 15-hour days on the job.

POLITICS AND PROSE

In late 2009, the U.S. Preventive Services Task Force announced new guidelines for mammograms for women ages 40 to 50. It suggested these women should consult with doctors about the need for mammograms, because the data showed the procedures were of limited benefit for women younger than 50.

The announcement came as the health care debate was coming to a boil, and reaction to the guidelines was swift and forceful. Patient advocacy groups opposed them. It seemed to them as though the government was trying to take a preventive tool away from women.

"The guidelines did not say that women 40 to 50 should not have a mammogram," says Morton. But that's not the way it was received after press reports, she notes.

The controversy was an important object-lesson for CER. The data do not always conform to the wants and expectations of patients, doctors, medical systems, or governments.

"I think patients think, *More is better.* And if they're faced with the question, *Should we do the test or not?* they think, *Why not?*" says Morton.

Herein lies a challenge. Hearing that an expensive test is unnecessary may be fine in the abstract, but when the patient is your wife or son, you're less likely to be receptive to the message.

"When people hear about CER, sometimes what they hear is, 'You're going to tell me what I can't get,'" Morton says. "They're not hearing, 'This is what works for you as an individual. This is what's best for you.'"

An area where comparative effectiveness has borne fruit is in cancer research. Nancy Davidson, director of the University of Pittsburgh Cancer Institute and UPMC Cancer Centers, is known for her work comparing treatments for certain kinds of breast cancers in premenopausal women. Davidson has also researched the importance of the HER-2 gene in breast cancer, and its role in treatment response. As a consequence of her studies and studies like them, gene testing is a common practice in breast cancer care.

"As we understand the molecular Achilles' heel for different types of cancers, we're better able to design treatments and make sure we're not exposing patients to treatments that won't work on them," Davidson says.

This kind of progress toward "personalized" medicine is exactly what Roberts would like to see in other areas of medicine. Clinical trials for new drugs are likely to exclude large numbers of patients, often because of age or disease. Yet most of the patients in Roberts' office are there because they are old or sick.

"If you ask the question, 'Would my patient have been included in the studies that were used to make those recommendations?' The answer is—a surprising number of times—'No,'" Roberts says.

"To me, the idea of comparative effectiveness is to try to understand, in real, live conditions, in real, live patients—who look like the patients I see—'How can I make better decisions about what's the best thing to do for them?'

"That's what we're trying to do." ■