

# Health Policies Merit Rigorous Evaluation. But They Don't Get It.

By AUSTIN FRAKT

A few years ago, Oregon found itself in a position that you'd think would be more commonplace: It was able to evaluate the impact of a substantial, expensive health policy change.

In a collaboration by the state and researchers, Medicaid coverage was randomly extended to some low-income adults and not to others, and researchers have been tracking the consequences ever since.

Rigorous evaluations of health policy are exceedingly rare. The United States spends a tremendous amount on health care, but very little of it learning which health policies work and which don't. In fact, less than 0.1 percent of total spending on American health care is devoted to evaluating them.

As a result, there's a lot less solid evidence to inform decision making on programs like Medicaid or Medicare than you might think. There is a similar uncertainty over common medical treatments: Hundreds of thousands of clinical trials are conducted each year, yet half of treatments used in clinical practice lack sound evidence.

As bad as this sounds, the evidence base for health policy is even thinner.

A law signed this year, the Foundations for Evidence-Based Policymaking Act, could help. Intended to improve the collection

of data about government programs, and the ability to access it, the law also requires agencies to develop a way to evaluate these and other programs.

Evaluations of health policy have rarely been as rigorous as clinical trials. A small minority of policy evaluations have had randomized designs, which are widely regarded as the gold standard of evidence and commonplace in clinical science. Nearly 80 percent of studies of medical interventions are randomized trials, but only 18 percent of studies of U.S. health care policy are.

Because randomized health policy studies are so rare, those that do occur are influential. The RAND health insurance experiment is the classic example. This 1970s experiment randomly assigned families to different levels of health care cost sharing. It found that those responsible for more of the cost of care use far less of it — and with no short-term adverse health outcomes (except for the poorest families with relatively sicker members).

The results have influenced health care insurance design for decades. In large part, you can thank (or curse) this randomized study and its interpretation for your health care deductible and co-payments.

More recently, the study based on random access to Oregon's Medicaid program has been influential in the debate over Medicaid expansion. A state lottery —

which provided the opportunity for Medicaid coverage to low-income adults — offered rich material for researchers. The findings that Medicaid increases access to care, diminishes financial hardship and reduces rates of depression have provided justification for program expansion. But its lack of statistically significant findings of improvements in other health outcomes has been pointed to by some as evidence that Medicaid is ineffective.

Although there are other examples of randomized studies in health policy, the vast majority have far less rigorous designs.

Some of them are sponsored by the Center for Medicare and Medicaid Innovation, created by the Affordable Care Act. It has spent about \$1 billion a year on dozens of programs that pay for Medicare and Medicaid services in new ways intended to enhance quality and reduce spending.

Most of the innovation center's pilots lack randomized designs, for which it has been criticized. Also potentially problematic: Most of its programs rely on voluntary participation by health care organizations. There might be crucial differences between those that opt in and those that don't.

Mandatory participation poses its own set of challenges. "If you force a hospital to join a new program, but not its competitor down the street, you might put the hospital at an unfair financial

disadvantage," said Nicholas Bagley, a University of Michigan health law professor. Also, testing voluntary participation makes sense if the program is never intended to be mandatory in the first place.

In considering a mandatory program, you also have to be mindful of politics.

"There will always be winners and losers," said Darshak Sanghavi, a former senior official for the Center for Medicare and

## Very little is spent on figuring out what actually works.

Medicaid Innovation. "If losers are forced to remain in a program, that could cause a political backlash that might blow the whole thing up."

Randomization can also be challenging; it can be complex and hard to maintain. "A program with desirable features for evaluation, like randomization, that falls apart could be less valuable than one that was designed more realistically from the start," he said.

Problems can also plague roll-outs that are voluntary and not randomized. Programs showing promise suffer from diminishing participation as health care orga-

nizations drop out. The innovation center's pioneer accountable care organization program offered health care organizations the opportunity to earn bonuses in exchange for accepting some financial risk, provided they meet a set of quality targets. It started with 32 participants in 2012. Although studies showed it reduced spending and at least maintained, if not improved, quality, only nine remained by 2016 when the program ended.

Some of the largest innovation center programs — involving thousands of providers — bundle payments across services for some common treatments (like knee and hip replacements) instead of paying separately for each one. More efficient providers that can deliver the care for less than that price can keep some of the difference as profit. Those that can't lose money. Of six bundled payment programs, only one included random assignment.

Beginning in April 2016, Medicare randomly assigned 75 markets to be subject to bundled payments for knee and hip replacements, and 121 markets to business as usual. But the innovation center didn't maintain the design, announcing in November 2017 that hospitals could leave it. This will greatly limit what can be learned from the program.

Just as in clinical care, there are examples of incorrect thinking based on low-rigor studies that

more rigorous ones later overturn. For example, many low-quality studies suggest that wellness programs reduce employers' health care costs as they improve health outcomes. But when the programs have been subject to randomized controlled trials, none of these findings hold up.

Hospital cost shifting — the idea that shortfalls from Medicare or Medicaid cause hospitals to charge higher prices to private insurers — can also seem commonplace from studies without rigorous designs. But when subject to more careful evaluation, the phenomenon is almost never observed.

An apparent preference for ignorance is not unique to health care. Policies across governments at all levels are routinely put in without plans to find out if they work — or how to unwind them if they don't, or how to build on them if they do. A 2017 Government Accountability Office report found that the vast majority of managers of federal programs were not aware of any recent evaluation of the programs they oversaw. In most cases, none had been done. In others, none had been done in the past five years.

It's hard to rid ourselves of ideas that are little more than wishful thinking or to end policies that don't work. The first step would be to do more rigorous policy evaluations. The next would be to heed them.