



A Strategy for Improving U.S. Health Care Delivery — Conducting More Randomized, Controlled Trials

Amy Finkelstein, Ph.D.

In December, the 2019 Nobel Prize in Economic Sciences was awarded to Abhijit Banerjee, Esther Duflo, and Michael Kremer “for their experimental approach to alleviating global poverty.” Through

their research and other efforts, including Banerjee and Duflo’s founding of the Abdul Latif Jameel Poverty Action Lab (J-PAL), the laureates pushed their field of development economics to more regularly use rigorous research methods — specifically, randomized, controlled trials (RCTs).

Until recently, the experimental approach that transformed the fight against global poverty had yet to make its mark on efforts to improve health care delivery in the United States. Despite the prominence of RCTs in medical research, they have too rarely been used to evaluate health care delivery. Between 2009 and 2013, just 18% of studies of U.S. health care delivery interventions used

randomization, as compared with 86% of drug studies and 66% of studies of nondrug medical interventions.¹ Recently, however, researchers, practitioners, and policymakers have started to find ways to overcome barriers to the widespread use of RCTs to improve health care delivery.

Some important and commonly discussed barriers include ethical concerns about rationing services, the time and financial costs involved in conducting RCTs, and whether RCTs can help tackle important, systemwide issues such as health equity or payment reform. My experiences as a researcher and as a cochair of J-PAL North America’s U.S. Health Care Delivery Initiative have illu-

minated many circumstances in which these challenges have been overcome. Of course, RCTs cannot be used for every health care policy question, but recent cases in which they have been successfully used should inspire practitioners, insurers, nonprofit organizations, and government agencies to use RCTs more often (see table).

A primary concern is whether it is ethically appropriate to conduct RCTs in a given circumstance. In health care delivery, logistical and financial constraints often mean that more people are eligible for a program than can be served. In such cases, randomization can ethically — and equitably — be used to allocate scarce resources. In 2008, for example, Oregon decided — without input from researchers — that a random lottery was the fairest way to allocate a limited number of Medicaid slots. This

Examples of Randomized, Controlled Trials (RCTs) in Health Care Delivery in the United States.*

Health Care Issue	Research Question	Implementer	Location	Level of Randomization and Sample	Outcomes
Medicaid expansion	What is Medicaid's effect on health, health care utilization, and economic outcomes?	Oregon Health Plan standard Medicaid program	Oregon	Roughly 75,000 people who signed up for the state's Medicaid lottery	Medicaid coverage increased the use of health care services, reduced financial hardship, reduced rates of depression, and improved self-reported health. Medicaid had no effect on physical health, employment, or earnings.
Medicare payment reform	What are the effects of bundled payments for knee and hip replacement on Medicare spending, health care utilization, and quality?	Centers for Medicare and Medicaid Services	United States	196 metropolitan statistical areas	In several studies, the bundled-payment program slightly reduced health care use and had no effect on quality of care, patient volume, or patient composition.
Overprescribing	Can peer-comparison letters reduce inappropriate prescribing of controlled substances?	Centers for Medicare and Medicaid Services	United States	1518 Medicare Part D high prescribers of Schedule II drugs	Letters had no detectable effect on prescribing. The Centers for Medicare and Medicaid Services and researchers continued to collaborate to test alternative versions of letters.
Overprescribing	Can strongly worded, peer-comparison letters reduce overprescribing of Seroquel without harming patients?	Centers for Medicare and Medicaid Services	United States	5055 Medicare Part D high prescribers of Seroquel	Letters caused substantial, long-lasting reductions in Seroquel prescribing without evidence of harm among patients.
Quality improvement	What can be learned from 10+ low-cost, rapid-cycle RCTs of quality-improvement efforts?	NYU Langone Health	New York	Various	Researchers concluded that the program of experimentation paid for itself by increasing take-up of preventive care.
Racial disparities in health	What is the effect of race concordance among patients and physicians and incentives on the take-up of preventive health services by black men?	Clinic established by researchers to provide preventive services	Oakland, CA	1374 black men	For black men, seeing a black male doctor significantly boosted demand for preventive health services, especially for more invasive tests. Incentives increased demand but didn't fully substitute for meeting with a black doctor. The findings suggest that black doctors could reduce the gap in cardiovascular mortality between black and white men by 19%.
"Superutilizers"	Can a care-transition program that provides intensive, time-limited clinical and social assistance to "superutilizers" of the health care system who have complex needs reduce hospital readmissions?	Camden Coalition of Healthcare Providers	Camden, NJ	800 patients with complex medical and social needs	The intervention had no effect on the rate of hospital readmission within 6 months after hospital discharge or on mortality.
Workplace wellness	What is a large-scale workplace wellness program's effect on employee health behaviors, self-reported health status, productivity, and medical expenditures?	University of Illinois-Urbana-Champaign	Urbana, IL	12,459 employees	The program had no measurable impact on employee health behaviors, productivity, self-reported health status, or total medical expenditures. Of the 41 measured outcomes, there were only two significant effects of program participation, both based on follow-up survey responses.
Workplace wellness	What is a workplace wellness program's effect on self-reported and clinical health, health spending, utilization, and employment?	BJ's Wholesale Club and Wellness Workdays	Northeast United States	200 BJ's Wholesale Club sites	After 18 months, employees at worksites where the wellness programs were offered reported improved health behaviors in 2 of 29 areas measured, but there were no effects on any other self-reported, clinical, spending, utilization, or employment outcome.

* Information is from J-PAL² and Horwitz et al.⁵

decision provided an opportunity for me and my coauthors to rigorously evaluate the effects of expanding Medicaid to poor, uninsured adults. Contrary to the pessimistic view that Medicaid coverage was worthless because clinicians wouldn't accept its low reimbursement rates, the randomized allocation of Medicaid coverage showed that such coverage increased use of preventive and primary health care and reduced financial hardship for beneficiaries.² Contrary to the optimistic view that increasing Medicaid coverage would save money by getting previously uninsured people out of emergency departments and into cheaper primary care settings, the evidence indicated that Medicaid coverage increased the use of emergency departments and increased total health care spending by about 25%.² The rigorous evidence produced by this RCT permitted the public to have an informed discussion about the trade-offs involved in expanding Medicaid.

Conducting RCTs is also ethically appropriate when there is equipoise, or uncertainty about a program's benefits. When programs require large investments and have uncertain effects, it's important to estimate the return on investment in order to consider whether more resources should be allocated to a given program or whether limited resources could be better deployed elsewhere. For example, the Affordable Care Act encouraged employers to offer workplace wellness programs, which observational studies had suggested could produce substantial health benefits and yield savings. Yet two RCTs of comprehensive workplace wellness programs showed that they had limited effects on

employees' health habits and no effect on their health, employment, or health care costs during the initial years.² The findings suggest that the programs may have failed to achieve what policymakers intended.

Regarding potential financial and time costs, the increasing availability and use of administrative data have made implementing RCTs easier and less expensive than it once was. A recent RCT of the Camden Coalition of Healthcare Providers "hotspotting" program (for which I was a coauthor) is a case in point.³ Two decades ago, in an RCT of a similar care-transition program, researchers used telephone interviews with patients to obtain information on readmissions after discharge, the study's primary outcome.⁴ This time around, improved data systems allowed us to instead use existing hospital discharge data from the four Camden hospital systems and the Camden Coalition Health Information Exchange database. Because we no longer needed to collect survey data, we could conduct the evaluation at substantially lower cost and effort and with less risk of nonresponse bias.²

Administrative data also enable use of RCTs for low-cost, rapid testing of repeatedly fine-tuned interventions. For example, researchers partnered with the Centers for Medicare and Medicaid Services (CMS) on an RCT evaluating the effect of sending letters to physicians who prescribed disproportionately large amounts of Schedule II drugs. CMS sent the letters in September 2014; researchers then used Medicare administrative data to learn that the letters didn't reduce prescribing within 30 or 90 days and built on psychological

and other research to quickly and cheaply modify an alternative letter for a subsequent RCT by April 2015.² This second study found that strongly worded letters comparing physicians' prescribing levels with those of their peers caused substantial, long-lasting reductions in prescribing of quetiapine (Seroquel), a nonscheduled but often overprescribed drug, with no evidence of negative effects on patients.²

RCTs can also help address systemwide challenges that seem intractable — such as racial disparities in health — by breaking them down into answerable questions to help identify root causes. In one recent RCT, black male patients at a clinic in Oakland, California, were randomly assigned to either a black or nonblack (white or Asian) physician.² When patients first saw a photo of their assigned doctor and indicated what preventive services they wanted to receive, the physician's race had no effect on their preferences. Once patients met with the physician and could revise their decisions, however, those who saw a black male doctor were more likely to take up preventive services, especially invasive services (such as influenza vaccination). Designing the trial in this way allowed the researchers to infer that patients' encounters and communication with the physicians, rather than physicians' race alone, drove the findings. The researchers are now conducting another RCT to learn which aspects of communication are important for influencing health behaviors.²

J-PAL North America supported many of these studies and dozens more, but we are only a small part of a larger movement that is gaining momentum as

the barriers to conducting RCTs of U.S. health care delivery interventions are shrinking. For example, NYU Langone Health recently completed 10 randomized, rapid-cycle quality-improvement projects in 1 year. The researchers concluded that this program of experimentation paid for itself by increasing take-up of preventive care.⁵ The U.S. government has also started using RCTs to study systemwide payment reforms. In 2016, CMS launched a nationwide, 5-year RCT of Medicare bundled-payment reform for hip and knee replacements that is mandatory for hospitals — the first of its kind. The initial results indicate that this bundled-payment program produced small reductions in patients' health care use and had no apparent effect on quality of

 **An audio interview with Dr. Finkelstein is available at NEJM.org**

care, patient volume, or patient composition.² CMS plans to launch similar RCTs this year examining the effects of systemwide payment reform in Medicare's end-stage renal disease program and in radiation oncology.

For so long, RCTs have been rare in U.S. health care delivery, but — as we are increasingly seeing — they don't have to be. Many energetic people and organizations are at the forefront of creating innovative health care delivery models to meet the needs of patients more effectively and efficiently. We owe it to patients to rigorously evaluate these efforts. The awarding of the 2019 Nobel Prize in Economic Sciences is a validation of the value of science in policymaking. It should serve as a beacon for health policy researchers to generate more rigorous evidence.

Disclosure forms provided by the author are available at NEJM.org.

From the Department of Economics, Massachusetts Institute of Technology, Cambridge.

1. Finkelstein A, Taubman S. Health care policy: randomize evaluations to improve health care delivery. *Science* 2015;347:720-2.
2. Results from J-PAL North America Supported Randomized Controlled Trials to Improve US Health Care Delivery. Cambridge, MA: Abdul Latif Jameel Poverty Action Lab, March 2020 (<https://www.povertyactionlab.org/publication/hcdi-evidence-wrap>).
3. Finkelstein A, Zhou A, Taubman S, Doyle J. Health care hotspotting — a randomized, controlled trial. *N Engl J Med* 2020; 382:152-62.
4. Naylor MD, Brooten D, Campbell R, et al. Comprehensive discharge planning and home follow-up of hospitalized elders: a randomized clinical trial. *JAMA* 1999;281:613-20.
5. Horwitz LI, Kuznetsova M, Jones SA. Creating a learning health system through rapid-cycle, randomized testing. *N Engl J Med* 2019;381:1175-9.

DOI: 10.1056/NEJMp1915762

Copyright © 2020 Massachusetts Medical Society.

Evidence from Pragmatic Trials during Routine Care — Slouching toward a Learning Health System

Gregory E. Simon, M.D., M.P.H., Richard Platt, M.D., and Adrian F. Hernandez, M.D., M.H.S.

In 2010, the National Academy of Medicine (NAM) called for development of a learning health system, setting a goal that by 2020, “90 percent of clinical decisions will be supported by accurate, timely, and up-to-date clinical information, and will reflect the best available evidence.”¹ Yet in 2020, clinicians and patients still lack high-quality evidence to guide the majority of common and consequential decisions regarding alternative treatments.² Absent such evidence, the type of care provided is determined by the

haphazard influences of financial incentives, clinicians' anecdotal experiences, and patients' or clinicians' exposure to marketing messages.

Evidence from traditional clinical trials will not be timely or relevant enough to fill the many evidence gaps. Highly selected participants, tightly controlled treatment delivery, and research-specific data collection in traditional trials all slow the pace of clinical research and undermine its generalizability for real-world decisions. The current Covid-19

pandemic underscores the need to rapidly generate evidence regarding common clinical decisions.

Observational comparisons using data from health records can fill only some of the evidence gaps. Population-based data from health records can be used to rapidly and efficiently generate evidence directly relevant to real-world practice. Detailed historical records and increasingly sophisticated methods to account for pretreatment differences can often support valid inference.