desired pathways (3). The benefit of such an approach, which is typical of enzyme catalysts, is that it should lead to greater overall catalytic activity while retaining selectivity. Moreover, recent advances in our understanding of noncovalent interactions have provided a way to harness the power of noncovalent interactions to improve the design of better catalysts. Unfortunately, interactions are the relatively weak, nondirectional nature of these interactions, necessitating the introduction of numerous mathematical models through multivariate regression. These models can be used to design small catalysts. ■

**REFERENCES**


**HEALTH CARE POLICY**

**Randomize evaluations to improve health care delivery**

Administrative data and experimental designs lead the way

*By Amy Finkelstein*<sup>1,2,*</sup> and Sarah Taubman<sup>3</sup>

The medical profession has long recognized the importance of randomized evaluations; such designs are commonly used to evaluate the safety and efficacy of medical innovations such as drugs and devices. Unfortunately, innovations in how health care is delivered (e.g., health insurance structures, interventions to encourage the use of appropriate care, and care coordination approaches) are rarely evaluated using randomization. We consider barriers to conducting randomized trials in this setting and suggest ways for overcoming them. Randomized evaluations of fundamental issues in health care policy and delivery should be—and can be—closer to the norm than the exception.

There is particular interest in improving delivery of health care in the United States, where the health care sector accounts for almost one-fifth of the economy. The newly created Patient-Centered Outcomes Research Institute is providing an estimated $3.5 billion in research grants, and the latest round of Center for Medicare and Medicaid Innovation Health Care Innovation Awards provides about $1 billion in research grants—much of it aimed at improving the delivery of U.S. health care.

Studies of U.S. health care delivery typically rely on a range of observational and quasi-experimental methods. These can be extremely valuable for learning as much as possible from existing historical data and for studying questions that are not amenable to randomized designs. For prospective evaluation of new interventions, however, it is often possible to use a randomized design without adding substantially to the cost or difficulty...
of the study. When feasible, randomized designs have an unparalleled ability to provide credible evidence on an intervention's impact. This can be seen in the outsized and enduring influence of the 1970s RAND Health Insurance Experiment, a randomized evaluation of the impact of health insurance in the United States (1, 2). More recently, the attention paid to the 2008 Oregon Health Insurance Experiment (OHIE), a randomized evaluation of the impact of Medicaid (3–6), underscores the continued power and influence of such randomized evaluations in both the academy and public discourse.

To explore how commonly randomization is used in health care delivery studies, we examined papers published in a limited set of top journals in medicine, economics, and health services between 2009 and 2013 [see (7) for details on data and methods]. We included papers designed to study causal effects of an intervention (using either randomized or other methods). We focused on a handful of top journals to capture an illustrative set of high-profile studies; the picture may be different across all published (and unpublished) studies. We did, however, observe similar patterns in reviews of trials registered with clinicaltrials.gov and of reports from major contract research organizations (7).

On average, 18% of studies of U.S. health care delivery interventions used randomization (see the table). By comparison, 79% of studies of U.S. medical interventions were randomized (P-value for comparison < 0.001). Medical studies involving drugs were very likely to be randomized (86%), but randomization was also common in nondrug medical studies (66%).

Of course, regulatory and funding environments in medicine are quite different from those in the social sciences. However, we found several areas of social science where randomization is used far more than in health care delivery. In U.S. education studies in top economics journals, 36% were randomized (P-value for comparison = 0.028). More notably, 46% of international development studies in top economics journals were randomized (P < 0.001). Even within health care delivery, there appears to be more use of randomization internationally than within the United States. Looking across the same journals in medicine, economics, and health services as above, 41% of health care delivery studies conducted outside of the United States were randomized, compared with 18% in the United States (P < 0.001).

DATA AND DESIGN. To understand why randomized trials in U.S. health care delivery have been rare, we turn to some of the challenges in conducting such studies. We then propose practical approaches to managing these challenges.

We begin with potential ethical considerations. For medical innovations, randomized trials are considered essential in determining both safety and efficacy. In health care delivery, safety concerns tend to be less strong. However, there is often equipoise regarding effectiveness. Moreover, it is common in health care delivery for promising programs to reach only a small fraction of the individuals who might benefit. Where there are capacity constraints, random assignment can be the most equitable way to allocate limited slots. Indeed, the random selection used in the OHIE was designed by the state in conjunction with stakeholders specifically to address fairness concerns (8).

Another common concern is that randomized evaluations are prohibitively costly, but this does not have to be the case. It is true that the typical model for randomized controlled trials in medicine is expensive and time-consuming. Screening, recruiting, and obtaining informed consent from individual patients before randomization and then collecting follow-up data for the purposes of the study is labor-intensive and difficult. Historically, most randomized trials of health care delivery innovations have followed this model. Our review of randomized studies of U.S. health care delivery published between 2009 and 2013 in top medical journals (7) found that 80% recruited and requested consent from individuals and 85% collected primary data.

A corollary of this labor-intensive approach is that randomized evaluations frequently focus on very specific patient populations. Of the 31 randomized health care delivery studies from top medical journals included in the table, 77% were convenience samples (for example, patients at a single hospital). This raises important concerns about their generalizability.

This expensive, time-consuming, and convenience-sample approach may be necessary in most medical trials, where there are often real risks to participants. However, in most health care delivery interventions, there is usually only minimal risk of harm to participants. As a result, an alternative approach to randomization can produce valid causal estimates at substantially reduced cost. Randomization is done, with a waiver of informed consent, on a set of potentially eligible individuals, and those who are randomized into the treatment group are offered the intervention. All individuals included in the random assignment—including those who do not accept the offer of the intervention—are followed. Low take-up of the program (adherence to the assigned protocol) does not interfere with obtaining consistent estimates of the program's causal effects (9). This type of randomization design was used in the OHIE; those randomly selected were offered Medicaid applications, which allowed us to study the impact of Medicaid coverage.

Although this approach reduces the statistical power of the study, it is compatible with running large trials and trials with more representative samples, because it does not require individual recruitment, and individuals can be followed passively in administrative data. Data collected, used, and stored for
reasons other than the study—such as from insurance claims, hospital discharges, electronic medical records, employment records, and mortality records—often include a virtual census of the relevant individuals. These data allow researchers to examine a wide range of impacts at substantially lower cost than primary data collection.

Compared with surveys, such administrative data offer several additional advantages besides lower cost. They are less likely to suffer from bias due to differential nonresponse or attrition. They can provide close-to-real-time results on the impact of an intervention. They can also be used for following up on administrative data, randomized evaluations can be made no more costly than the prospective observational evaluations that are commonly done in U.S. health care delivery.

A final set of challenges revolves around the ways individuals and systems interact. Some of the most promising ideas for U.S. health care delivery interventions involve reforms to entire systems of care. Cluster-randomized designs can be a useful tool here. Some system-level or comprehensive interventions may even be amenable to patient-level randomization. For example, innovations such as including bundling payments for episodes of care and creating shared saving contracts—major themes in current health policy discussions—are often held up as examples of something hard to study through randomized evaluation. Yet as these payment mechanisms expand to take on new groups of patients, randomizing which individuals are included may be possible.

Of course, some interventions—including individual-level interventions—can have system-wide effects if implemented on a large scale. Consider the expansion of insurance coverage. A randomized study like the OHIE allows us to detect effects of covering a given individual with insurance, while holding the general health care environment constant. However, capacity constraints in the health care system may limit effects of market-wide expansions, particularly in the short run; alternatively, as suggested by quasi-experimental work (11), provider responses to a market-wide insurance expansion—such as adoption of new medical technology and changes in practice style—may amplify effects of market-wide expansions. In some cases, it is possible to design studies that look at these broader effects, by randomizing the proportion of individuals within the relevant unit who are assigned the treatment, as well as randomizing which individuals within the unit are assigned the treatment (12). In other cases, however, such approaches are not practical or feasible, and we need to draw on other methods.

More generally, this discussion highlights the value of experiments that are actively designed by researchers to shed light on specific mechanisms. The OHIE was not prospectively designed by researchers (8) and, as a result, leaves much to be debated, such as whether an alternatively designed Medicaid program could achieve most of the benefits but at lower cost (see, e.g., (13)). By contrast, the RAND Health Insurance Experiment was prospectively designed by researchers to shed light on tradeoffs involved in cost-sharing. It used multiple arms to randomly vary cost-sharing features of health insurance that individuals received. It has been widely used in policy and academic discussions of optimal cost-sharing designs.

Governments, insurers, employers, and health care providers are experimenting with a wide variety of innovations intended to improve health and reduce costs. Increased use of randomized evaluations offers a feasible way to more rigorously measure their efficacy and accelerate the pace at which we improve the health care delivery system.
Randomize evaluations to improve health care delivery
Amy Finkelstein and Sarah Taubman

Science 347 (6223), 720-722.
DOI: 10.1126/science.aaa2362