

Leveraging natural experiments to evaluate interventions in learning health systems

Sunita Desai ,¹ Eric Roberts²

¹Department of Population Health, School of Medicine, NYU, New York, New York, USA
²Health Policy and Management, University of Pittsburgh, Pittsburgh, Pennsylvania, USA

Correspondence to

Dr Sunita Desai, School of Medicine, Department of Population Health, NYU, New York, NY 10012-1126, USA; sunita.desai@nyulangone.org

Accepted 27 January 2020
 Published Online First
 6 February 2020

Health systems are increasingly testing interventions to reduce costs and improve patient care.¹ By leveraging their ability to deploy interventions across a network of providers and their data infrastructure, health systems have capacity to test new models of care delivery and to assess their effects.

What differentiates health systems from *learning* health systems, however, is how systems combine this infrastructure with rigorous research methods to generate evidence that can inform decision-making.^{1 2} Critical to this is isolating the causal effects of interventions from potential sources of bias (internal validity) and providing evidence that generalises to broader, real-world settings (external validity). As Walkey *et al* demonstrate in this issue of *BMJ Quality and Safety*, natural experiments that arise in the design and implementation of health system interventions can provide valuable opportunities to learn about the effectiveness of interventions, yielding findings that may offer both high internal and external validity.³

Natural experiments arise when the design or implementation of a programme generates variation in exposure to an intervention that is ‘as good as random’, conditional on observed confounders. For example, patients may be allocated to different treatments based on an arbitrary eligibility threshold (Walkey *et al* exploit this variation); health systems may stagger the roll-out of interventions across sites⁴; or patients may be entered into a lottery to allocate a number of intervention slots.⁵ When well conceived and thoroughly vetted, natural experiments retain many of the appealing statistical features of randomised controlled trials (RCTs) while offering the practical advantages of observational study

designs. While RCTs are the gold standard for clinical evaluation, they are often costly; raise ethical concerns; and because they are conducted in ‘idealised’ settings, can suffer from limited external generalisability. Moreover, the prospective nature of most RCTs—patients are enrolled in one or more treatment or control arms and are followed longitudinally—usually precludes timely findings that leaders of health systems seek to guide decision-making. Observational studies, which use real-world data without an explicitly experimental context, can overcome some of these practical limitations of RCTs, but usually at the expense of internal validity. For example, they may rely on strong and unrealistic assumptions (eg, a pre–post study design assumes the outcome would have remained unchanged in the absence of the intervention) and usually cannot control for important confounders, leading to biased inferences.

Walkey *et al* identify and exploit a natural experiment that arose in the targeting of care co-ordination interventions to prevent readmissions among patients in a large, academic health system. Exposure to these interventions was based on risk-score cut-offs: patients just above certain risk score thresholds received additional care co-ordination services compared with patients just below these risk score thresholds. Because patients above and below these thresholds are likely to be similar in all respects except for the intervention received (‘exchangeability’), the study setting provides variation in exposure to interventions that mimics a randomised trial within the vicinity of each threshold.⁶ Therefore, comparing individuals above and below these thresholds provides a plausible means of isolating causal effects of these interventions on readmission rates. To



© Author(s) (or their employer(s)) 2021. No commercial re-use. See rights and permissions. Published by BMJ.

To cite: Desai S, Roberts E. *BMJ Qual Saf* 2021;**30**:183–185.

address concerns about model mis-specification given the discrete distribution of the risk score variable, the authors apply a combined differences-in-differences/regression discontinuity approach to evaluate the impact of a readmission reduction initiative implemented among patients in a large academic medical centre.

However, natural experiments are not infallible and require careful investigation of their assumptions. Comprehensively reporting tests examining the plausibility of the assumptions in published studies using quasi-experimental methods should be standard practice. After all, study estimates reflect causal effects of the intervention only if the identifying assumptions are supported. Moreover, as opposed to RCTs where necessary assumptions are weak and likely to hold given appropriate implementation of randomisation, settings that appear to be natural experiments on the surface often turn out to be confounded on closer inspection.

In Walkey *et al*'s study, several assumptions warrant attention. First, the assumption of exchangeability—that patients above and below thresholds are similar in all respects except for the intervention received—is not fully testable in quasi-experimental studies, although careful descriptive analyses can lend support to this assumption. For example, the authors could have examined whether observed characteristics of the patients above and below the risk scores trended continuously through the thresholds or changed before versus after the intervention was implemented.

Second, the plausibility of quasi-random assignment in this context hinges on the assumption that neither patients nor providers could have manipulated the variable that determines eligibility for an intervention. To address this concern, Walkey *et al* examined the distribution of risk scores in the population before and after the intervention's start date and show no evidence of systematic changes or 'heaping' at the thresholds.⁷ Additional information about how the risk scores were constructed (eg, the extent to which scores were constructed prospectively versus modified after the introduction of the intervention) could help to substantiate the plausible randomness of the intervention with respect to potential outcomes. This underscores the important role institutional knowledge plays when leveraging a natural experiment to provide credible, causal estimates.

Moreover, the generalisability of these studies' findings may be limited beyond the population that was quasi-randomised to the intervention. For example, the treatment effects in Walkey *et al*'s study are specific to individuals whose risk scores lay within the vicinity of treatment eligibility thresholds but may not generalise across the full distribution of risk scores. As such, further detail on risk score construction is necessary to understand if and for whom an intervention would

be effective if it was applied in other settings or health systems.

Experimentation in health systems is sometimes accomplished by design, but as the term 'natural experiment' suggests, it often arises incidentally—an artefact of institutional decision-making. For example, to implement a pay-for-performance initiative for large provider groups, a large payer may draw an arbitrary boundary for eligibility based on practice size⁸; an exogenous, policy-determined threshold based on the proportion of a hospital's admitted patients who are low income may establish whether a hospital is eligible for discounts on drug purchases⁹; to allocate limited enrolment slots into a safety-net insurance programme, a state may enrol applicants via a lottery⁵; and to allocate scarce provider resources, a system may target high-risk patients based on an eligibility cut-off.³

Traditionally, harnessing these natural experiments has been the domain of social scientists—particularly economists—who often work outside the health system setting. Yet leaders and researchers embedded within health systems are often better informed about institutional details needed to identify potential natural experiments and determine whether a policy or intervention constitutes a credible natural experiment for evaluation. Thus, effective learning within health systems benefits from partnerships between health system leaders, social scientists and statisticians. 'Embedding' these trained researchers in health systems and expanding collaboration between health systems and their academic affiliates offers promise in enabling health systems to rigorously learn from and produce generalisable knowledge about the interventions they implement. Natural experiments generated from interventions must be implemented judiciously but can provide a creative and practical means for learning health systems to generate rigorous evidence in support of best practices.

Funding ER received salary support from a research scientist development award from the Agency for Healthcare Research and Quality (K01HS026727).

Competing interests None declared.

Patient consent for publication Not required.

Provenance and peer review Commissioned; internally peer reviewed.

ORCID iD

Sunita Desai <http://orcid.org/0000-0002-7962-5091>

REFERENCES

- 1 Horwitz LI, Kuznetsova M, Jones SA. Creating a learning health system through rapid-cycle, randomized testing. *N Engl J Med* 2019;381:1175–9.
- 2 Agency for Healthcare Research and Quality. *About learning health systems*. Rockville, MD: Agency for Healthcare Research and Quality, 2019. <https://www.ahrq.gov/learning-health-systems/about.html>

- 3 Walkey AJ, Bor J, Cordella NJ. Novel tools for a learning health system: a combined difference-in-difference/regression discontinuity approach to evaluate effectiveness of a readmission reduction initiative. *BMJ Qual Saf* 2020;29:161–7.
- 4 Garrido MM, Allman RM, Pizer SD, *et al*. Innovation in a learning health care system: veteran-directed home- and community-based services. *J Am Geriatr Soc* 2017;65:2446–51.
- 5 Finkelstein A, Taubman S, Wright B, *et al*. The Oregon health insurance experiment: evidence from the first year. *Q J Econ* 2012;127:1057–106.
- 6 Venkataramani AS, Bor J, Jena AB. Regression discontinuity designs in healthcare research. *BMJ* 2016;i:i1216.
- 7 Barreca AI, Lindo JM, Waddell GR. Heaping-induced bias in regression-discontinuity designs (NBER working paper 17408). *Econ Inq* 2011;54.
- 8 Roberts ET, Zaslavsky AM, McWilliams JM. The value-based payment modifier: program outcomes and implications for disparities. *Ann Intern Med* 2018;168:255–65.
- 9 Desai S, McWilliams JM. Consequences of the 340B drug pricing program. *N Engl J Med* 2018;378:539–48.