



Viewpoint

"Natural Experiments" in Health Care Research

Dhruv Khullar, MD, MPP; Anupam B. Jena, MD, PhD

In "natural experiments," the treatment or intervention is determined by variation not under the control of the researcher. These designs, used in economics and epidemiology to support inferences about causal relationships between interventions and outcomes, are useful tools to help improve the rigor of observational studies in health policy and medicine. Perhaps the first natural experiment in medicine was that of the English physician John Snow in the mid-nineteenth century. In 1854, a cholera outbreak struck Broad Street in London, killing hundreds. Studying case clusters, Snow discovered that neighborhoods supplied with water downstream of where sewage was discharged into the Thames River experienced high levels of disease, while neighborhoods receiving upstream water had low disease levels.¹ Snow described the populations as similar in age, occupation, income, and social rank, divided into groups without choice, illustrating an essential component of natural experiments: similar but distinct populations that are exposed to a condition outside the researchers' control, allowing for reasonable conclusions about the potential causal link between exposure and outcome.

Randomized clinical trials (RCTs) have traditionally been viewed as the primary method for establishing causality in health care, but they have important limitations: they are expensive; it is not always possible to randomize patients; and their findings may not be generalizable to different patient populations or nonexperimental settings. When RCTs are not possible, medical and health policy researchers have turned to observational studies. In observational studies, however, individuals are not assigned to the intervention independently of potential confounding factors that could also influence outcomes, making it difficult to separate the treatment effect from other factors that may be associated with receiving the treatment.

By contrast, natural experiments rely on variation in treatment exposure that may be unrelated to other factors associated with the outcomes. Suppose researchers are interested in examining the likelihood of long-term use and adverse outcomes for patients after an initial opioid prescription. An observational analysis might be confounded if the factors that influence a clinician's decision to prescribe opioids (eg, cancer-related pain) also affect long-term outcomes (eg, opioid dependence). An RCT might resolve this issue but would be ethically and practically challenging. Instead, researchers could examine how long-term opioid use varies among opioid-naive individuals who, by chance, are exposed to physicians with a high propensity vs low propensity to prescribe opioids (eg, when assigned to the next available physician in an emergency department).² In this scenario, the long-term outcomes following an initial opioid prescription could be identified by variation in the drug's use associated with prescriber variation that is plausibly unrelated to variation in unobservable patient factors associated with both initial opioid use and long-term outcomes.

Natural experiments use *quasi-randomization*, a method of allocation to study groups that is not truly random and is not assigned by a researcher, such as a specific date, age, or event. These study designs have an important feature: the similarity of the groups can be measured. Treatment and control groups should be similar in sociodemographic characteristics, comorbidities, prior health care utilization, and any other factors that might be associated with outcomes, but often this is not the case and adjustments are needed based on these observed variables. Natural experiments attempt to control for unobserved variables. When well implemented, natural experiments may be more informative than traditional observational studies that do not control for unobservable confounders, but are less informative than RCTs in establishing true cause and effect. With natural experiments,

+ Editorial

Author affiliations and article information are listed at the end of this article.

 **Open Access.** This is an open access article distributed under the terms of the CC-BY License.

the more closely the study design resembles an RCT, the more confidence we may have in the validity of the findings.

Five types of natural experiments are particularly relevant for observational studies in health policy and medicine: regression discontinuity designs (RDD), instrumental variable designs, difference-in-differences (DID) analyses, event-study analyses, and interrupted time-series²⁻⁴ (Table). This is an overview of these types of studies with health policy examples and is not intended to provide a detailed assessment of these designs.

Regression discontinuity designs identify effect sizes associated with an intervention by studying individuals with treatment assignment that differs by position on either side of a specific, arbitrary cutoff (eg, a treatment threshold, policy implementation date, an age threshold, or a geographic discontinuity).⁵ In this design, the probability of being exposed to the intervention changes discontinuously at this cutoff. Studies using RDD rely on the assumption that individuals on either side of the cutoff are similar, so their treatment assignment is nearly independent of their characteristics, both observed and unobserved. For example, a 2018 study evaluated the phased introduction of Medicare's Value-Based Payment Modifier program.³ Researchers used the program's practice size thresholds (eg, 100 or more clinicians) to evaluate whether the program was associated with practice performance—with the assumption that practices just above and below the cutoff did not differ in important ways—and found that the program was not associated with improved practice performance and may have exacerbated health disparities.

Instrumental variable analyses using quasi-random variation in assignment to treatment or intervention have also been used to study clinical and health policy interventions.⁶ For example, health policy researchers have been interested in whether higher spending hospitals achieve better outcomes, a relationship that is confounded by the fact that higher spending hospitals may treat patients that are disproportionately sicker, which could spuriously suggest that higher hospital spending leads to worse outcomes. To address this issue, a study examined the association between hospital spending and mortality by using quasi-random variation in ambulance dispatching patterns as an instrumental variable.⁷ Ambulances may have preferences for which hospitals patients are taken to for reasons that are unrelated to patient clinical severity; this, in turn, may lead otherwise similar patients to be transported to (and treated at) higher vs lower spending hospitals.

Other natural experiments use different types of analyses to assess potential causal relationships. These include DID, event study, and interrupted time series analyses. In DID analysis, researchers compare outcomes in 2 groups that were similar before an intervention (natural or otherwise) that affected only 1 of the groups.⁸ The DID analysis postulates that if the treatment had no effect, the differences between the groups would be unchanged after the treatment. One such

Table. Examples of "Natural Experiments" in Health Care Research

Example	Natural experiment type	Testable assumption
Roberts et al, 2017 ³ Using the Medicare value modifier program's practice size participation threshold to estimate the association between program exposure and performance on quality and spending measures	Regression discontinuity design: Quasi-random differences in health insurance coverage around an age-based eligibility threshold	Individuals on either side of an arbitrary threshold are similar but for differences in the probability of the intervention
Barnett et al, 2017 ² Using quasi-random assignment of patients to physicians within an emergency department, and variation across physicians in their propensity to prescribe opioids, to estimate the effect size of the association between an initial opioid prescription and long-term opioid dependence	Instrumental variable analysis: Differences in treatment based on quasi-random variation in practice patterns across clinicians, hospitals and treatment centers, and third-party payers	Physicians and hospitals vary in the probability of offering a particular treatment, but patients are otherwise similar
Sabety et al, 2021 ⁴ Using the timing of primary care physician exit from the workforce (eg, retirement) to study the associations between longitudinal care continuity and health care use and outcomes	Difference-in-differences, event-study, and interrupted time-series analyses: Differences in treatment resulting from the timing of specific events	Individuals whose likelihood of treatment may be affected by the timing of an event are similar to those in surrounding periods whose treatment likelihood is unaffected

study found lower long-term mortality rates after Hurricane Katrina among people who had been living in New Orleans compared with those who had been living in other similar cities, which they concluded represented the effect of migration because New Orleans residents migrated to areas with better socioeconomic conditions and lower baseline mortality after the hurricane.⁹ A randomized experiment on the effects of resettling a population on that scale would have been infeasible.

In event-study analyses, researchers rely on exogenous and variable timing of interventions in exposed groups to study changes within groups over time (eg, estimating the effect size for the association between care continuity and outcomes by studying patients whose primary care physicians retired at different points in time). Although event-study analyses do not require control groups, control groups without any exposure are frequently incorporated into this approach.⁴ Interrupted time series analyses are similar, but typically focus on changes in outcomes before and after a single event that affects a population of interest (eg, a citywide soda excise tax).¹⁰

Each of these types of study designs and analyses have important limitations that should be considered, including not controlling for unobserved or unmeasured differences between the groups, risk of selection bias due to allocation that cannot be concealed from the researchers, non-parallel trends that could affect comparisons between the groups, and spillover influences from 1 group to the other. For example, in RDD studies, assumptions must be tested to ensure that observed variables are continuous at the point where the treatment and outcome discontinuities occur, such that there are no abrupt changes in the relationship between the observed variables and the treatment or outcome except at the discontinuity cutoff. Similarly, studies that use instrumental variable analysis must ensure that an appropriate instrumental variable is selected and should acknowledge the possible threats to validity from unmeasured confounding factors.

Natural experiments offer an important approach for examining potential causal links between interventions and outcomes. Studies that appropriately use these methods could help provide data to inform questions affecting the health of patients that otherwise may remain unanswered.

ARTICLE INFORMATION

Published: June 11, 2021. doi:[10.1001/jamahealthforum.2021.0290](https://doi.org/10.1001/jamahealthforum.2021.0290)

Open Access: This is an open access article distributed under the terms of the [CC-BY License](https://creativecommons.org/licenses/by/4.0/). © 2021 Khullar D et al. *JAMA Health Forum*.

Corresponding Author: Anupam B. Jena, MD, PhD, Department of Health Care Policy, Harvard Medical School, 180 Longwood Ave, Boston, MA 02115 (jena@hcp.med.harvard.edu).

Author Affiliations: Division of Health Policy and Economics, Department of Population Health Sciences, Weill Cornell Medical College, New York, New York (Khullar); Division of General Internal Medicine, Department of Medicine, Weill Cornell Medical College, New York, New York (Khullar); Department of Health Care Policy, Harvard Medical School, Boston, Massachusetts (Jena); Department of Medicine, Massachusetts General Hospital, Boston (Jena); National Bureau of Economic Research, Cambridge, Massachusetts (Jena).

Conflict of Interest Disclosures: Dr Jena reports consulting fees from Pfizer, Bioerativ, Bristol Myers Squibb, Merck Sharp & Dohme, Janssen, Edwards Life Sciences, Novartis, Amgen, Eli Lilly, Vertex Pharmaceuticals, AstraZeneca, Celgene, Tesaro, Sanofi, Aventis, Precision Health Economics, and Analysis Group, all outside of the submitted work.

REFERENCES

1. Bor J. Capitalizing on natural experiments to improve our understanding of population health. *Am J Public Health*. 2016;106(8):1388-1389. doi:[10.2105/AJPH.2016.303294](https://doi.org/10.2105/AJPH.2016.303294)
2. Barnett ML, Olenski AR, Jena AB. Opioid-prescribing patterns of emergency physicians and risk of long-term use. *N Engl J Med*. 2017;376(7):663-673. doi:[10.1056/NEJMsa1610524](https://doi.org/10.1056/NEJMsa1610524)
3. Roberts ET, Zaslavsky AM, McWilliams JM. The value-based payment modifier: program outcomes and implications for disparities. *Ann Intern Med*. 2018;168(4):255-265. doi:[10.7326/M17-1740](https://doi.org/10.7326/M17-1740)

4. Sabety AH, Jena AB, Barnett ML. Changes in health care use and outcomes after turnover in primary care. *JAMA Intern Med*. 2021;181(2):186-194. doi:[10.1001/jamainternmed.2020.6288](https://doi.org/10.1001/jamainternmed.2020.6288)
5. Maciejewski ML, Basu A. Regression Discontinuity Design. *JAMA*. 2020;324(4):381-382. doi:[10.1001/jama.2020.3822](https://doi.org/10.1001/jama.2020.3822)
6. Maciejewski ML, Brookhart MA. Using instrumental variables to address bias from unobserved confounders. *JAMA*. 2019;321(21):2124-2125. doi:[10.1001/jama.2019.5646](https://doi.org/10.1001/jama.2019.5646)
7. Doyle J, Graves J, Gruber J, Kleiner S. Measuring returns to hospital care: evidence from ambulance referral patterns. *J Polit Econ*. 2015;123(1):170-214. doi:[10.1086/677756](https://doi.org/10.1086/677756)
8. Dimick JB, Ryan AM. Methods for evaluating changes in health care policy: the difference-in-differences approach. *JAMA*. 2014;312(22):2401-2402. doi:[10.1001/jama.2014.16153](https://doi.org/10.1001/jama.2014.16153)
9. Deryugina T, Molitor D. Does when you die depend on where you live? Evidence from Hurricane Katrina. *Am Econ Rev*. 2020;110(11):3602-3633. doi:[10.1257/aer.20181026](https://doi.org/10.1257/aer.20181026)
10. Powell LM, Leider J. Evaluation of changes in beverage prices and volume sold following the implementation and repeal of a sweetened beverage tax in Cook County, Illinois. *JAMA Netw Open*. 2020;3(12):e2031083. doi:[10.1001/jamanetworkopen.2020.31083](https://doi.org/10.1001/jamanetworkopen.2020.31083)