Natural Experiments

Thad Dunning
Department of Political Science
Yale University

Draft entry for the *International Encyclopedia of Political Science*

Word count: 2, 523

The importance of *natural experiments* lies in their contribution to addressing confounding, a pervasive problem in the social sciences. Consider the obstacles to addressing the following hypothesis: extending property titles to poor land squatters boosts access to credit markets and promotes beliefs in individual political efficacy, thereby fostering socioeconomic development. To test this idea, researchers might compare poor squatters who possess land titles to those who do not. Yet, confounding may be a problem, because differences in individual attitudes and behaviors could in part be due to factors—such as family background—that also make certain poor squatters more likely to acquire titles to their property.

Investigators may therefore seek to control for potential confounders in observational (non-experimental) data. For instance, they may compare titled and untitled squatters, within strata defined by measures of family background. At the core of conventional quantitative methods is the hope that such confounders can be identified, measured, and controlled. Yet, this is not easy to do. Even within the strata defined by family background and intelligence, there may be other difficult-to-measure confounders—say, determination—that are associated with obtaining titles and that also influence economic and political behaviors.

Randomization is one way to eliminate confounding. In a randomized controlled experiment to estimate the effects of land titling, subjects could be randomly assigned to receive titles or not.

Family background, determination, and other possible confounders would be equivalent, on average and up to random error, across these two groups. Large post-titling differences would then be credible evidence for a causal effect of land titles. Yet, experimental research in such contexts may be expensive, impractical, or unethical.

Scholars therefore increasingly employ natural experiments—attempting to identify and analyze real-world situations in which some process of random or *as-if* random assignment places cases in alternative categories of the key independent variable. In the social sciences, this approach has been used to study the relationship between lottery winnings and political attitudes, the effect of voting costs on turnout, the impact of quotas for women village councilors on public goods provision in India, and many other topics. In the health sciences, a paradigmatic example comes from John Snow's nineteenth-century tests of the hypothesis that cholera is waterborne.

Natural experiments share one crucial attribute with true experiments and partially share a second attribute. First, outcomes are typically compared across subjects exposed to a treatment and those exposed to a control condition (or a different treatment). Second, in partial contrast with true experiments, subjects are often assigned to the treatment not at random, but rather *as-if* at random (though sometimes true randomization occurs, as in lottery studies). Given that the data come from naturally occurring phenomena that often entail social and political processes, the manipulation of the treatment is not under the control of the analyst; thus, the study is observational.

However, a researcher carrying out this type of study can often make a credible claim that the assignment of non-experimental subjects to treatment and control conditions is as good as random. This distinguishes natural experiments from "quasi-experiments," in which comparisons

are also made across treatment and control groups but *non-random* assignment to treatment is a key feature of the designs.

Yet, how can the claim of *as-if* random assignment in natural experiments be validated? And how much leverage do natural experiments in fact provide for causal inference? These questions are discussed below, after the initial example on land-titling is discussed at greater length.

How do Property Rights Affect the Poor?

An interesting social-scientific example comes from a study of how land titles influence the socio-economic development of poor communities. In 1981, urban squatters organized by the Catholic Church in Argentina occupied open land in the province of Buenos Aires, dividing the land into parcels that were allocated to individual families. A 1984 law, adopted after the return to democracy in 1983, expropriated this land with the intention of transferring titles to the squatters. However, some of the original landowners challenged the expropriation in court, leading to long delays in the transfer of titles to some of the squatters. By contrast, for other squatters, titles were granted immediately.

The legal action therefore created a (treatment) group of squatters to whom titles were granted promptly and a (control) group to whom titles were not granted. The authors of the study find subsequent differences across the two groups in standard social development indicators: average housing investment, household structure, and educational attainment of children. On the other hand, the authors do not find a difference in access to credit markets, which contradicts a well-known theory that the poor will use titled property to collateralize debt. They also find a positive effect of property rights on self-perceptions of individual efficacy. For instance,

squatters who were granted land titles—for reasons over which they apparently had no control—disproportionately agreed with statements that people get ahead in life due to hard work.

Is this a valid natural experiment? The key claim is that land titles were assigned to the squatters *as-if* at random, and the authors present various kinds of evidence to support this assertion. In 1981, for example, the eventual expropriation of land by the state and the transfer of titles to squatters could not have been predicted. Moreover, there was little basis for successful prediction by squatters or the Catholic Church organizers of which *particular* parcels would eventually have their titles transferred in 1984. Titled and untitled parcels sat side-by-side in the occupied area, and the parcels had similar characteristics, such as distance from polluted creeks. The authors also show that the squatters' characteristics, such as age and sex, were statistically unrelated to whether they received titles—as should be the case if titles were assigned at random. Finally, the government offered equivalent compensation—based on the size of the lot—to the original owners in both groups, suggesting that the value of the parcels does not explain which owners challenged expropriation and which did not. On the basis of extensive interviews and other qualitative fieldwork, the authors argue convincingly that idiosyncratic factors explain some owners' decisions to challenge expropriation, and that these factors were unrelated to the characteristics of squatters or their parcels.

The authors thus present compelling evidence for the equivalence of treated and untreated units. Along with qualitative evidence on the process by which the squatting took place, this evidence helps bolster the assertion that assignment is *as-if* random. Of course, assignment was not randomized, so the possibility of unobserved confounders cannot be entirely ruled out. Yet the argument for as-good-as-random assignment appears compelling. Note that the natural

experiment plays a crucial role. Without it, the intriguing findings about the self-reinforcing (not to mention self-deluding) beliefs of the squatters could have been explained as a result of unobserved characteristics of those squatters who did or did not successfully gain titles.

A Framework for Evaluating Natural Experiments

How much leverage for causal inference do natural experiments in fact provide? To address this question, it is helpful to discuss three dimensions along which natural experiments may vary: (1) plausibility of *as-if* random assignment; (2) credibility of the statistical models, which is closely connected with the simplicity and transparency of the data analysis; and (3) substantive relevance of the intervention—i.e., whether and in what ways the specific contrast between treatment and control provides insight into a wider range of important issues and contexts.

First, the key claim—and the definitional criterion—for a natural experiment is that treatment assignment is as good as random. Yet, one finds marked variation among studies that claim to use natural experiments in the plausibility of this claim.

How can the assertion of *as-if* random be at least partially validated? First, it should be supported by the available empirical evidence—for example, by showing equivalence on relevant pre-treatment variables (those whose values were determined before the intervention took place) across treatment and control groups, as would occur on average with true randomization. Qualitative knowledge about the process by which treatment assignment takes place can also play a key role in validating a natural experiment. The authors of the studies on

land-titling use both quantitative comparisons and qualitative evidence about the process of organizing squatters' settlements to validate the claim of *as-if* random assignment.

A second dimension along which natural experiments may vary is in the credibility of the statistical models used to analyze the data. As with true experiments, *as-if* random assignment implies that both known and unknown confounders are balanced (in expectation) across treatment and control groups, obviating the need to measure and control for confounding variables. This has the great advantage of permitting the use of simple analytic tools—for example, comparisons of means or percentages across the treatment and control groups—to make causal inferences. In principle, the simplicity and transparency of the statistical analysis provides natural experiments with an important advantage, relative to the conventional quantitative methods that have in recent years incurred substantial criticism from leading methodologists.

In practice, greater credibility of statistical models is not inherent in all studies that claim to use natural experiments. One recent survey of leading examples from political science and economics found that about half of the studies failed to present simple, unadjusted difference-of-means tests (in addition to any auxiliary analyses). Of course, in less-than-perfect natural experiments, in which the plausibility of *as-if* random is perhaps impeachable, researchers may feel compelled to control for potential confounders they can measure. Yet, any substantial changes after adjustment likely point to a lack of *as-if* random assignment—because randomization would ensure that control variables are independent of treatment assignment. Post-hoc statistical fixes can also lead to data mining, with only "significant" estimates of causal effects making their way into published reports. (Researchers also sometimes use multivariate

regression to reduce the variability of treatment effect estimators; yet, variance may be higher or lower after adjustment, and the standard errors calculated using the usual regression formulas do not apply). Thus, to bolster the credibility of the statistical models employed in natural-experimental designs, analysts should report unadjusted difference-of-means tests, in addition to any auxiliary analyses.

A third dimension along which natural experiments can be classified is the substantive relevance of the intervention. Here we ask: To what extent does *as-if* random assignment shed light on the wider social-scientific, substantive, and/or policy issues that motivate the study?

Answers to this question might be a cause for concern, for a number of reasons. For instance, the type of subjects or units exposed to the intervention might be more or less like the populations in which we are most interested. In lottery studies of electoral behavior, for example, levels of lottery winnings may be randomly assigned among lottery players, but we might doubt whether lottery players are like other populations (say, all voters). Next, the particular treatment might have idiosyncratic effects that are distinct from the effects of greatest interest. To continue the same example, levels of lottery winnings may or may not have similar effects on, say, political attitudes as income earned through work. Finally, natural-experimental interventions (like the interventions in some true experiments) may "bundle" many distinct treatments or components of treatments, which may limit the extent to which this approach isolates the effect of the explanatory variable about which we care most. Such ideas are often discussed under the rubric of "external validity," but the issue of substantive relevance involves a broader question: i.e., whether the intervention in fact yields causal inferences about the real causal hypothesis of concern.

To tie this discussion together, we can imagine a cube, in which the three axes are defined by these three dimensions: (1) plausibility of *as-if* random assignment; (2) credibility of statistical models; and (3) substantive relevance of intervention. In the front lower-left corner of the cube, we find those natural experiments that offer the least plausibility, credibility, and substantive relevance. In the back upper-left corner of the cube, we find those studies with the most plausibility, credibility, and relevance. The process of achieving a strong research design—in which natural experiments do provide substantial leverage for causal inference—may be understood as the process of moving from the front lower-left, "weak research design" corner of the cube, to the back upper-right, "strong research design" corner of the cube.

Two points are important to make in closing. First, there may be trade-offs in seeking to design a strong natural experiment, that is, in moving to the back upper-right corner of the cube. Different studies may manage the trade-off among these three dimensions in different ways, and which trade-offs are acceptable (or unavoidable) may depend on the question being asked.

Second, deep substantive knowledge, and a combination of quantitative and qualitative analysis, can help analysts better achieve success along all the three dimensions of the framework described above. Consider the studies of squatters in Argentina. There, substantive knowledge was necessary to recognize the potential to use a natural experiment to study the effect of land titling, and many field interviews were required to probe the plausibility of *as-if* randomness—that is, to validate the research design. Fieldwork can also enrich analysts' understanding and interpretation of the causal effects they estimate.

In sum, many modes of inquiry may contribute to successful causal inference using natural experiments; ultimately, the right mix of methods substantially depends on the research question

involved. Natural experiments, like regression analysis, do not provide a technical quick fix to the challenges of causal inference. In every study, analysts are challenged to think critically about the match between the assumptions of models and the empirical reality they are studying. Natural experiments are valuable to the extent they build on real substantive knowledge and appropriate methodological craftsmanship, and a full awareness of the trade-offs inherent in this style of research.

Further reading

Campbell, Donald T. and Julian C. Stanley. 1963. *Experimental and Quasi-Experimental Designs for Research.* Boston, MA: Houghton Mifflin Company.

Di Tella, Rafael, Sebastian Galiani, and Ernesto Schargrodsky. 2007. "The Formation of Beliefs: Evidence from the Allocation of Land Titles to Squatters." *Quarterly Journal of Economics* 122: 209–41.

Dunning, Thad. 2010. "Design-Based Inference: Beyond the Pitfalls of Regression Analysis?" In Henry E. Brady and David Collier, eds., *Rethinking Social Inquiry: Diverse Tools, Shared Standards.* Rowman & Littlefield, 2nd edition.

Galiani, Sebastian, and Ernesto Schargrodsky. 2004. "The Health Effects of Land Titling." *Economics and Human Biology* 2: 353–72.

Freedman, David A. 2010. "Statistical Models and Shoe Leather." Reprinted in David Collier, Jasjeet S. Sekhon, and Philip B. Stark, eds., *Statistical Models and Causal Inference*. New York: Cambridge University Press.

Snow, John. 1855. *On the Mode of Communication of Cholera.* London: John Churchill, New Burlington Street, England, 2nd edition. Reprinted in *Snow on Cholera*, London: Humphrey Milford: Oxford University Press, 1936.

Thistlethwaite, Donald L. and Donald T. Campbell. 1960. "Regression-discontinuity Analysis: An Alternative to the Ex-post Facto Experiment." *Journal of Educational Psychology* 51 no. 6: 309–17.