World Development 127 (2020) 104825

Contents lists available at ScienceDirect

World Development

journal homepage: www.elsevier.com/locate/worlddev

Commentary

Reasons for policy experimentation that have nothing to do with selection bias

Cyrus Samii

New York University Department of Politics and Evidence in Governance and Politics (EGAP), United States

ARTICLE INFO

Article history:
Accepted 1 December 2019

ABSTRACT

The conventional case in favor of policy experimentation focuses on how randomization controls selection bias. This is undoubtedly important. This essay focuses on additional benefits from experimentation that are completely distinct from controlling selection bias. These benefits derive directly from the fact that experimentation involves actively intervening to assign policy treatments. Experimentation "puts manipulability to the test" in ways that passive observation does not, and it allows for deeply engaged learning about policy formulation and implementation that ex post analyses miss.

The standard case for an experiment over observational methods that work off of naturally occurring treatment selection is that by controlling treatment assignment, experiments help to prevent selection bias. For example, it may be that under natural conditions, a treatment is much more likely to be selected by (or for) those with characteristics that will tend to make their outcomes more negative. Then if one used a simple comparison of treated and untreated individuals to estimate the effect of the treatment, the estimate would be biased in the negative direction. We can try to account for observable differences between those who choose the treatment and those who did not, but the observable differences may not be all that matters. An experiment intervenes in the natural process to assign treatments in ways that can remove such bias, for example by using randomization, in which case there would be no systematic relationship between the conditions that determine outcomes and treatment assignment. Of course, one can always get a bad draw, but claims about bias are about the general operating characteristics of a procedure, and moreover, methods such as blocking, restricted randomization, or even near-deterministic assignment methods are available to control the potential for bad draws (Banerjee, Chassang, & Snowberg, 2017). By such arguments, the case for experimentation depends primarily on how one values the control of such biases relative to the various costs that experiments introduce—e.g., implementation costs that arise from trying to achieve adequate scale.

In this essay I want to bring attention to two additional benefits of experimentation that have nothing to do with these questions of selection bias or statistical efficiency. I call these two benefits "putting manipulability to the test" and "deep engagement." The key here is to understand experimentation as *active manipulation of policy treatments with the intention to learn*. Box (1966) famously proposed that, "to find out what happens to a system when you interfere with it you have to interfere with it (not just passively observe it)." In an experiment, variation in policy treatments is due to such active manipulation with the intention to learn. This is different from observational studies in which variation in policy treatments are the product of naturally occurring and endogenous selection processes divorced from a learning motive. The sections that follow explain the implications of this difference.

1. Putting manipulability to the test

Whether one or another intervention is likely to be more effective depends both on the relevant mechanisms driving outcomes and, crucially, whether the mechanisms can be meaningfully affected through intervention. It is in addressing the second question that experimental studies are especially useful. Various approaches, including both qualitative and quantitative, are helpful in identifying important mechanisms that drive outcomes. But experiments can provide especially direct evidence on whether we can actually do anything to affect these mechanisms — that is, experiments put "manipulability" to the test.

The successful use of experiments in policy research typically requires drawing on insights from other research on relevant mechanisms. This other research defines debates about what policy makers should do and how they should do it. Experiments have a distinct role in such debates by clarifying what is materially possible through human intervention.

Related to this is replicability. From an experiment, in principle, you should have before you a recipe for creating an

E-mail address: cds2083@nyu.edu

https://doi.org/10.1016/j.worlddev.2019.104825
0305-750X/

effect. Context-dependence means that replicability may sometimes be elusive in practice. One could measure scientific progress in terms of the ability to fashion recipes (including the contextual conditions) for giving rise to predictable effects (Hacking, 1983). Observational studies are often deficient in this regard, because they do not control where and when we get the variation in treatments of interest. We are left to wonder whether we have really mastered what the observational data imply about causal effects. It's possible that we have not mastered it at all, overlooked important details or background conditions, and have merely tricked ourselves (or others!) into believing that our observational analysis gives us accurate insights about what would happen were we to apply a treatment recipe. Experiments test recipes directly.

2. Deep engagement

Experimental evaluations of policies or programs are prospective. They typically require deep engagement between researchers and implementers in processes of treatment formulation, beneficiary selection, and site selection. That is, they involve researchers in doing policy. This offers additional opportunities to learn about what might be realistic in terms of policy options. It also offers insights that can help in the interpretation of estimated effects.

We can compare this type of deep engagement in experimental research to ex post analyses, which typically lack access to such detail into policy formulation and implementation. It is for good reason then that you often hear from practitioners that ex post evaluators did not understand "what really went on" in the program. They weren't there from the beginning. In my experience, this is much less the case for experimental studies. Working prospectively, the researcher is there operating alongside implementation. The experimental method typically defines beneficiary selection. Constructing the experiment requires making programmatic goals concrete for defining interventions crisply and devising outcome measures. In my experience, implementing partners have found it useful to go through the process of making interventions and outcomes concrete in a manner that speaks to a scientific audience. It helps to make clear what is at stake.

Of course these two benefits should be taken alongside some of the limitations of experiments, which mix statistical and non-statistical considerations. Experiments face timescale challenges, since we can often only sustain experimental variation in treatment differentiation for so long, whether due to ethical or program cycle reasons. They also face spatial scale challenges: it can be impractical to develop well-powered experiments for macro level institutions or programs that cover large areas. For these reasons, in my opinion, the most important areas for methodological innovation for experiments are in finding ways to maximize the information that we can extract from relatively small experiments and in finding ways to reliably evaluate outcomes after only a short amount of time. There is also the logistical complexity of experiments, given all the up-front decisions that they require. I do not include external validity as a distinct challenge for experiments as external validity is an issue that all empirical research typically faces (Samii, 2016). And finally, the type of deep engagement discussed above is no substitute for more systematic ethnographic observation of policy processes. Nonetheless, it is useful to articulate these additional benefits of experimentation to clarify how they are about a lot more than addressing selection bias.

References

Banerjee, A., Chassang, S., & Snowberg, E. (2017). Decision Theoretic Approaches to Experiment Design and External Validity. In A. Banerjee & E. Duflo (Eds.). *Handbook of Field Experiments* (Volume 1). North Holland.

Box, G. E. P. (1966). Use and Abuse of Regression. *Technometrics, 8*(4), 625–629.

Hacking, I. (1983). *Representing and Intervening: Introductory Topics in the Philosophy of Natural Sciences*. Cambridge: Cambridge University Press.

Samii, C. (2016). Causal Empiricism in Quantitative Research. *Journal of Politics, 78* (3), 941–955.