

results. Such a perspective would undervalue contributions made by small-*N* methods. More generally, that kind of perspective ignores a crucial point: there are many ways to do good science.

Notes

¹ Fearon and Laitin show that close examination of typical cases (countries with no civil war and low probability of civil war according to the model) can be illuminating—a special and valuable feature of their research. Indeed, they use the cases to check the qualitative implications of their causal model.

² Seawright and Gerring (2008) give a clearer account of the matter, indicating that the relevant population must be large.

³ The setting for propensity-score matching is usually an observational study where subjects self-select into one of two conditions; call these “treatment” and “control.” The first step is usually to estimate the conditional probability that a subject winds up in treatment, given the covariates. Logit models are often used. This is not an activity to be undertaken with a small sample. For empirical evidence on the weaknesses of matching designs in large-*N* research, see for instance Arceneaux, Gerber, and Green (2006), Glazerman, Levy, and Myers (2003), Peikes, Moreno, and Orzol (2008), Wilde and Hollister (2007). For additional discussion pro and con, see *Review of Economics and Statistics* 86:1 (February 2004); *Journal of Econometrics* 125:1–2 (March–April 2005).

⁴ Suppose, for instance, that we have an experimental population of 1,000 subjects, with 400 chosen at random and assigned to treatment; the remaining 600 are the controls. Each subject has two potential responses: one is observed if the subject is assigned to treatment, and the other if assigned to control. The average response of the 400 is an unbiased estimate of what the average would be if all 1,000 subjects were assigned to treatment. Likewise, the average response of the 600 is an unbiased estimate of the average response if all 1,000 subjects were controls. The general principle is this: with a simple random sample, the sample average is an unbiased estimate of the population average. For additional details, see Freedman (2006).

⁵ For example, see Freedman, Pisani, and Purves (2007). Chapter 27 discusses experimental comparisons; technical detail is provided in A31–36. Chapter 29 explains what happens without randomization; also see Freedman (2008e).

References

- Arceneaux, Kevin, Alan S. Gerber, and Donald P. Green. 2006. “Comparing Experimental and Matching Methods Using a Large-Scale Voter Mobilization Experiment.” *Political Analysis* 14: 37–62.
- Berk, Richard A. 2004. *Regression Analysis: A Constructive Critique*. Sage Publications.
- Box-Steffensmeier, Janet M., Henry E. Brady, and David Collier. 2008. *The Oxford Handbook of Political Methodology*. New York: Oxford University Press.
- Brady, Henry E. and David Collier. 2004. *Rethinking Social Inquiry: Diverse Tools, Shared Standards*. Lanham, MD: Rowman and Littlefield.
- Fearon, James D. and David D. Laitin. 2008. “Integrating Qualitative and Quantitative Methods.” In *The Oxford Handbook of Political Methodology*. Janet Box-Steffensmeier, Henry E. Brady, and David Collier, eds. (New York: Oxford University Press), 756–76.
- Freedman, David A. 2005. *Statistical Models: Theory and Practice*. New York: Cambridge University Press.
- Freedman, David A. 2006. “Statistical Models for Causation: What Inferential Leverage Do They Provide?” *Evaluation Review* 30:

691–713.

- Freedman, David A. 2008a. “On Types of Scientific Enquiry: The Role of Qualitative Reasoning.” In *The Oxford Handbook of Political Methodology*. Janet Box-Steffensmeier, Henry E. Brady, and David Collier, eds. (New York: Oxford University Press), 300–18.
- Freedman, David A. 2008b. “On Regression Adjustments in Experiments with Several Treatments.” *Annals of Applied Statistics* 2: 176–96.
- Freedman, David A. 2008c. “Randomization Does Not Justify Logistic Regression.” *Statistical Science* 23: 237–49.
- Freedman, David A. 2008d. “Survival Analysis: A Primer.” *The American Statistician* 62: 110–19.
- Freedman, David A. 2008e. “Oasis or Mirage?” *Chance* 21: 59–61.
- Freedman, David A., Robert Pisani, and Roger A. Purves. 2007. *Statistics*. 4th edition. New York: W. W. Norton & Company, Inc.
- Gerring, John. 2008. “Case Selection for Case-Study Analysis: Qualitative and Quantitative Techniques.” In *The Oxford Handbook of Political Methodology*. Janet Box-Steffensmeier, Henry E. Brady, and David Collier, eds. (New York: Oxford University Press), 645–84.
- Glazerman, Steven, Dan M. Levy, and David Myers. 2003. “Non-experimental versus Experimental Estimates of Earnings Impacts.” *Annals of the American Academy of Political and Social Science* 589: 63–93.
- King, Gary, Robert O. Keohane, and Sidney Verba. 1994. *Designing Social Inquiry: Scientific Inference in Qualitative Research*. Princeton: Princeton University Press.
- Mahoney, James and Dietrich Rueschemeyer. 2003. *Comparative Historical Analysis in the Social Sciences*. New York: Cambridge University Press.
- Peikes, Deborah N., Lorenzo Moreno, and Sean Michael Orzol. 2008. “Propensity Score Matching: A Note of Caution for Evaluators of Social Programs.” *The American Statistician* 62: 222–31.
- Seawright, Jason and John Gerring. 2008. “Case Selection Techniques in Case Study Research: A Menu of Qualitative and Quantitative Options.” *Political Research Quarterly* 61: 294–308.
- Wilde, Elizabeth T. and Robinson Hollister. 2007. “How Close Is Close Enough? Evaluating Propensity Score Matching Using Data from a Class Size Reduction Experiment.” *Journal of Policy Analysis and Management* 26: 455–77.

Response to David Freedman

<https://doi.org/10.5281/zenodo.996340>

James D. Fearon

Stanford University

jfearon@stanford.edu

David D. Laitin

Stanford University

dlaitin@stanford.edu

As Jonathan Swift made mockingly clear, “modest proposals” that purport to solve previously unyielding problems can have horrible implications. Such proposals should be subjected to skeptical analysis. So we are pleased that our proposed random method of case selection for the qualitative component of multi-method research has attracted some skeptical commentary in the research community in political science.¹ And we are very grateful to David Freedman for providing a perspective on our approach. He is especially qualified to do

so, as he is a leading statistician who has long worried about inflated claims for statistical methods in the social sciences, and has been a champion of approaches that are sensitive to the particularities of each datapoint.

We completely agree with Freedman's claim that there are many ways to do good social science. Indeed, as Freedman quotes us, we argued that the random narratives approach is "a compelling complement" to large-*N* research. This is more modest than Freedman's implication that we believe we have discovered the one true path for multi-method research. In fact, if everyone did random narratives, there would be no expert narratives for the research community to consult!

Furthermore, as Freedman points out, the method has been applied only to our work on civil war onsets. Perhaps it will not be the best approach for other questions that scholars want to use multiple methods to address. We agree, although one goal of our article was to argue that there are good *a priori* (or theoretical) reasons to think that the approach could be valuable for research designs on topics other than civil war onset. Multi-method and other social science research inevitably involves a process of going back and forth between theory and data (despite the pristine hypothesis-testing scenario assumed in statistics textbooks). The random narratives approach is a way to discipline and make more productive this back-and-forth process in a fairly typical social science setting, where one has cross-sectional or panel data with which to document empirical patterns, and historical materials available to investigate causal mechanisms in particular cases.

In the case of our work on civil war, we constructed a country/year dataset with civil war onset as the dependent variable.² We estimated a statistical model that identified several correlates of civil war onsets for which we proposed possible causal interpretations. The interpretations were based on a reading of the statistical results *and* our previous knowledge of a set of cases well known to us. To look at those same cases for qualitative support for a causal interpretation would have been intellectual double-dipping. The method of randomly selecting cases for analysis of causal mechanisms behind peace or war onset helped us to avoid or at least reduce this bias. But we certainly do not maintain that random selection of cases for detailed analysis would always be the most effective and efficient approach in a multi-method research project, independent of the subject matter or the stage of the research (in terms of "back and forth").

Freedman writes that our "claimed superiority of large-*N* methods is obviously right if 'empirical regularities' are statistical measures of association, like regression coefficients. The thesis is less obvious if 'empirical regularities' are defined more broadly, so as to include (for example) causal relationships." We agree here as well, although we were trying in the cited sentences precisely to distinguish empirical regularities in the sense of mere associations from causal relationships. We would not claim a generalized superiority of large-*N* methods for identifying causal relationships. Indeed, the main idea of the multi-method approach we are endorsing is to use case-specific evidence systematically to assess whether causal in-

terpretations of the mere associations seen in a regression analysis are justified.

We do not therefore see how Freedman attributes to us the notions of the "superiority" of large-*N* methods or that "qualitative methods are useful only as checks on quantitative results." These claims may suggest incorrectly that we think causal relationships are easily read out of large-*N* statistical studies in social science. They also misread our view of the contributions made by small-*N* methods in the overall research process. In practice, as we noted above, there is a constant back-and-forth between data and theory in social science research, with case study evidence entering in more than one way. Knowledge of particular cases often helps to suggest causal mechanisms that may or may not be common and relevant in a larger sample of cases, and so may motivate and guide construction of a large-*N* study. A large-*N* study may in turn reveal new and different-from-expected patterns that stimulate new (or revised) theorizing about causal relationships, which may then be assessed by a return to case studies (chosen at random?). Those case studies may suggest new causal relationships that can subsequently be put to test with a newly constructed dataset. So it often happens in political science that case studies come into the scientific process at an early stage, motivating the research and the source of early conjectures, and then again at a later stage, after the regressions have been run.

Researchers in comparative politics invariably go back-and-forth between theory and data, and quite often they go back and forth between cases and broad patterns. Our modest proposal is an attempt to make progress on the question of by what principles to choose the cases in the context of the back and forth.

Notes

¹ See for example Evan Lieberman's critique of our proposed method, "Nested Analysis as a Mixed-Method Strategy for Comparative Research." *American Political Science Review* 99:3 (August 2005), 435–52.

² James D. Fearon and David D. Laitin, "Ethnicity, Insurgency, and Civil War." *American Political Science Review* 97:1 (February 2003), 75–90.

Techniques for Case Selection: A Response to David Freedman

John Gerring
Boston University
jgerring@bu.edu

Recognition of the problem posed by case selection in case study research stretches back, arguably, to the very beginnings of the genre, e.g., to early work by Frederic Le Play (1806–1882) and Florian Znaniecki (1882–1958). Harry Eckstein's (1975) classic study, a point of departure for political scientists today, appeared over three decades ago. Clearly, the field has been struggling with this issue for some time.¹