Political Analysis (2013) 21:40–41
doi:10.1093/pan/mps032

Preregistration of Studies and Mock Reports

Andrew Gelman
Departments of Statistics and Political Science, Columbia University, Amsterdam Ave. at 122 St., New York, NY 10027
e-mail: gelman@stat.columbia.edu

Edited by R. Michael Alvarez

The traditional system of scientific and scholarly publishing is breaking down in two different directions.

On one hand, we are moving away from relying on a small set of journals as gatekeepers: the number of papers and research projects is increasing, the number of publication outlets is increasing, and important manuscripts are being posted on SSRN, Arxiv, and other nonrefereed sites.

At the same time, many researchers are worried about the profusion of published claims that turn out to not replicate or in plain language, to be false. This concern is not new—some prominent discussions include Rosenthal (1979), Ioannidis (2005), and Vul et al. (2009)—but there is a growing sense that the scientific signal is being swamped by noise.

The papers at hand address the second concern in a helpful way, in both cases going beyond mere advice and criticism to actually implement their proposals in the context of their own ongoing research.

Monogan registered his study and chose his analysis ahead of time. He also carefully considers the practical concerns of registration and the advantages to the research community from specifying data collection protocols explicitly so they can be shared.

Another virtue of preregistration is that exposing an idea to outside eyes can reveal flaws in the plan. For example, Monogan writes, "A researcher interested in the effect of a treatment such as a state's Hispanic population on the probability of a senator's yea vote could create a research design ahead of time, post it publicly, and then finish the analysis once the roll call was reported." It does not make sense to me to consider a state's Hispanic population as a "treatment." If such a study were publicly registered, there is a chance that someone would raise this concern ahead of time, motivating the researchers to more clearly formulate their hypothesis.

My main concern with Monogan's proposal is that it may encourage a sort of robotic data analysis. His flagship example, an observational study of congressional positions and election outcomes, includes three tables with numbers presented to four decimal places and a nonparametric estimate of a treatment effect—but no plots of the raw data. Elsewhere, he seeks to "reduce the number of Type II errors we commit as a discipline," but I am bothered by thinking about "Type II errors" (or, for that matter, Type I errors) at all. A "Type II error" occurs when a true effect exists but it is claimed to be zero. But I see the problem as with making that claim; I would prefer to accept the uncertainty (Gelman and Stern 2006). It is fine to reject (or not reject) a null hypothesis, but in social science our hypotheses are complex and evolving. Nothing was stopping Monogan from plotting his data, of course, but perhaps the singleminded emphasis on testing a particular theory pushed him to limit his explorations.

Humphreys, Sanchez de la Sierra, and van der Windt go a step further, carrying out a plan that I and others have often recommended but have never actually done: before carrying out their analysis, they simulated fake data and used it to prepare a "mock report" which they publicly released. One advantage they found with this strategy was that it forced them to make various microdecisions about their analyses, choices which would ultimately need to be taken in any case

© The Author 2013. Published by Oxford University Press on behalf of the Society for Political Methodology.
All rights reserved. For Permissions, please email: journals.permissions@oup.com

https://doi.org/10.1093/pan/mps032 Published online by Cambridge University Press

but for which it is cleaner to decide ahead of time. Often it is only when focusing on the write-up that we fully engage with our research questions.

Given the high cost of collecting data compared with the relatively low cost of writing a mock report, we recommend the "mock report" strategy be done more often, especially for researchers planning a new and expensive study. The mock report is a form of pilot study and has similar virtues.

As with the other article under discussion, I am dissatisfied with the focus of Humphreys et al. on formal hypothesis testing. I am not particularly worried about "Type I errors" and "false positives" because I do not think these researchers are studying zero effects. As discussed in Gelman and Tuerlinckx (2000), I would be more concerned with Type S errors (getting the sign of the effect wrong) or Type M errors (when the magnitude of an effect is poorly estimated).

In the long term, I believe we as social scientists need to move beyond the paradigm in which a single study can establish a definitive result. In addition to the procedural innovations suggested in the papers at hand, I think we have to more seriously consider the integration of new studies with the existing literature, going beyond the simple (and wrong) dichotomy in which statistically significant findings are considered as true and nonsignificant results are taken to be zero. But registration of studies seems like a useful step in any case.

References

Gelman, A., and H. S. Stern. 2006. The difference between "significant" and "not significant" is not itself statistically significant. *American Statistician* 60:328–31.

Gelman, A., and F. Tuerlinckx. 2000. Type S error rates for classical and Bayesian single and multiple comparison procedures. *Computational Statistics* 15:373–90.

Humphreys, Macartan, Raul Sanchez de la Sierra, and Peter van der Windt. 2013. Fishing, commitment, and communication: A proposal for comprehensive nonbinding research registration. *Political Analysis* 21:1–20.

Ioannidis, J. P. A. 2005. Why most published research findings are false. Public Library of Science. *Medicine* 2:e124.

Monogan, James E., III. 2013. A case for registering studies of political outcomes: An application in the 2010 House elections. *Political Analysis* 21:21–37.

Rosenthal, R. 1979. The "file drawer problem" and tolerance for null results. *Psychological Bulletin* 86:638–41.

Vul, E., C. Harris, P. Winkielman, and H. Pashler. 2009. Puzzlingly high correlations in fMRI studies of emotion, personality, and social cognition. *Perspectives on Psychological Science* 4:274–90.

https://doi.org/10.1093/pan/mps032 Published online by Cambridge University Press