

Tie my hands loosely

Pre-analysis plans in political science

Daniel Rubenson , Ryerson University

Several years ago, I was at an American Political Science Association Annual Meeting at which there was a well-attended special session to discuss the newly proposed Data Access and Research Transparency (DART) initiative, which encouraged researchers to make their data available to facilitate evaluation of their findings. During a heated discussion, a very senior scholar who was skeptical of the application of DART in general and for qualitative research in particular, exclaimed in frustration, “It’s as if you think we’re trying to hide things!” A slightly more junior scholar who had been a proponent of greater transparency responded with a deadpan, “Yes, it’s as if you’re trying to hide things.” Much has happened in the 10 or so years since that meeting in terms of social and political science research practice and attitudes. One of the major developments has been the increasingly widespread adoption of the registration of pre-analysis plans (PAPs), in which researchers register their design and empirical specifications before accessing and analyzing (and often before collecting) their data. In many ways, this has been a natural extension of what economists Joshua Angrist and Jorn-Steffen Pischke (2010) call the “credibility revolution” in empirical economics and political economy—broadly speaking, the use of identification-driven research designs, most prominently randomized experiments.

Figure 1 shows the growth in the number of registered PAPs in two of the more prominent social science registries—the Evidence in Governance and Politics (EGAP) design registry¹ and the American Economic Association’s RCT Registry.² As can be seen, the number of registrations of PAPs has grown steadily each year since the introduction of these two registries in 2011 and 2013, respectively, with a slight dip (likely pandemic related) in the EGAP registry in 2020. As I alluded to earlier, this trend has been driven to a large extent by the

explosion of experimental research designs in political science and economics over the past two decades and by related organizations such as the EGAP research network, which has done much for the adoption of more transparent research practices. There have been several arguments presented for the adoption of PAPs and proposals for how this might work in social science. Humphreys et al. (2013), Nosek et al. (2015), and Munafo et al. (2017) are some of the more prominent recent ones.

In this short article, I want to provide some of my thoughts on these developments from the perspective of someone who writes PAPs and reads them as a reviewer, as well as from the perspective of a journal editor.

What’s the problem?

The problem for which PAPs are meant to be a (partial) solution is, in a nutshell, researcher degrees of freedom. That is, when I design and carry out an experiment (or any study), I have an immense amount of flexibility in the choices I can implement. I can, for instance, make decisions about how to measure key concepts, what data to collect, how to randomize, what model to estimate, and how to interpret the results. Researchers may face incentives to make choices that will be more likely to result in publishable work—for example, more likely to produce statistically significant results or results that confirm a hypothesis or are particularly novel. More specifically, three major issues arise: HARKing, fishing, and the file drawer problem.

HARKing (Hypothesizing After the Results are Known) refers to the practice of developing and presenting a hypothesis based on the results obtained from the analysis and passing it off as if it were an a priori hypothesis (i.e., developed before seeing the results) (Kerr, 1998). One obvious problem with this practice is that it leads to hypotheses that are always confirmed. It is difficult to know how much HARKing goes on, since, in the absence of a PAP, we do not know ex ante what the hypotheses were. However, Rubin (2017) provides some interesting evidence from surveys simply asking researchers whether they engage in the practice. Self-admission rates are high: on average, 43% of researchers

doi: 10.1017/pls.2021.23

Correspondence: Daniel Rubenson, Ryerson University, Toronto, M5B 2K3, Canada. Email: rubenson@ryerson.ca

¹See <https://egap.org/registry-0/> and <https://osf.io/registries/egap/discover>. The EGAP registry merged with OSF in 2019.

²See <https://www.socialscienceregistry.org>.

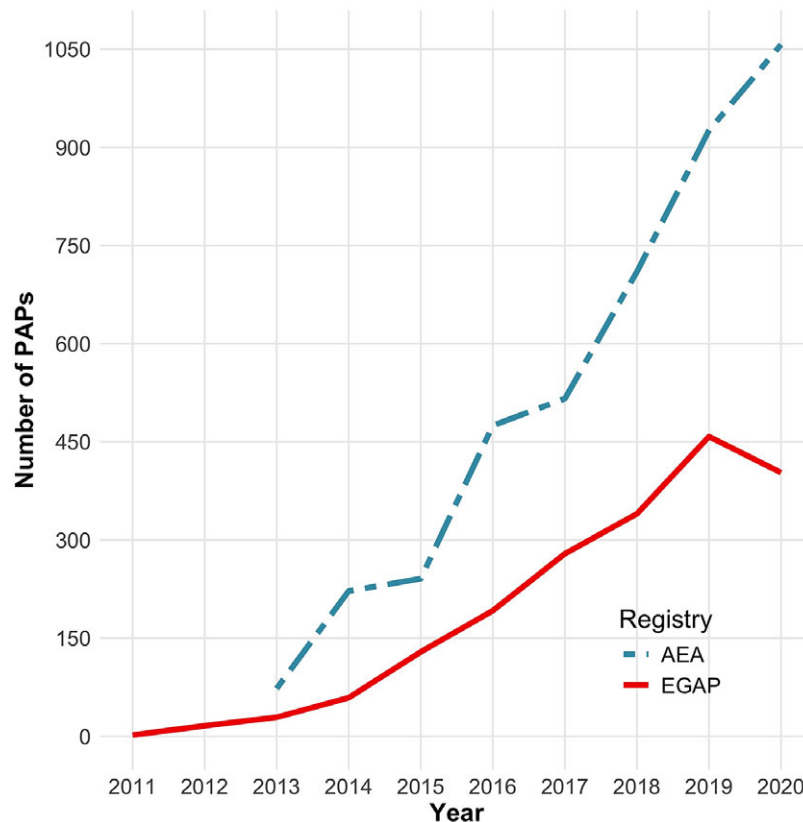


Figure 1. Number of PAPs registered 2011–2020 in the American Economic Association’s RCT Registry and the EGAP design registry. Note: The EGAP registry merged with OSF in 2020; therefore, data for 2020 have not been confirmed.

polled admitted to having HARKed (Rubin, 2017, p. 3). This is almost certainly an underestimate, since social desirability bias would mean people are loath to admit this behavior.

A second threat to the credibility of scientific research is what is known as “fishing” or “*p*-hacking” (Humphreys et al., 2013). Here, the issue is that a researcher may run regression after regression, fishing around until a statistically significant result (a small enough *p*-value) is found, and then report that result. As Humphreys et al. (2013) put it, “The problem with fishing is simple: selecting what results get reported can induce bias. If for example classical statistical tests are reported only when they yield ‘significant’ findings, then false positives will be overreported, true negatives will be underreported, and overall conclusions will be wrong” (p. 3).

Gerber and Malhotra (2008a), for example, find compelling evidence of fishing in the two leading political science journals—the *American Political Science Review* and the *American Journal of Political Science*.

Figure 2, reproduced from their article, shows the distribution of *z*-statistics in published work in these two journals. There is a clear spike of *z*-statistics just above 1.96, the critical value for statistical significance at the *p* = .05 level. Their evidence speaks to the problem of fishing and *p*-hacking. It is entirely possible—indeed, likely—that this spike in results at just the right side of statistical significance is the product of specification searches for *z*-statistics larger than the critical value.

The third problem, very much related to the first two, is that some research findings will not be submitted, let alone published—instead, they are consigned to the proverbial file drawer. There is a growing body of evidence suggesting that publication bias of this kind is quite widespread across social science disciplines (Christensen & Miguel, 2020). The evidence presented by Gerber and Malhotra also points to serious issues of publication bias. The kind of pattern seen in Figure 2 is, of course, more or less impossible just by chance, implying that studies with smaller *z*-statistics

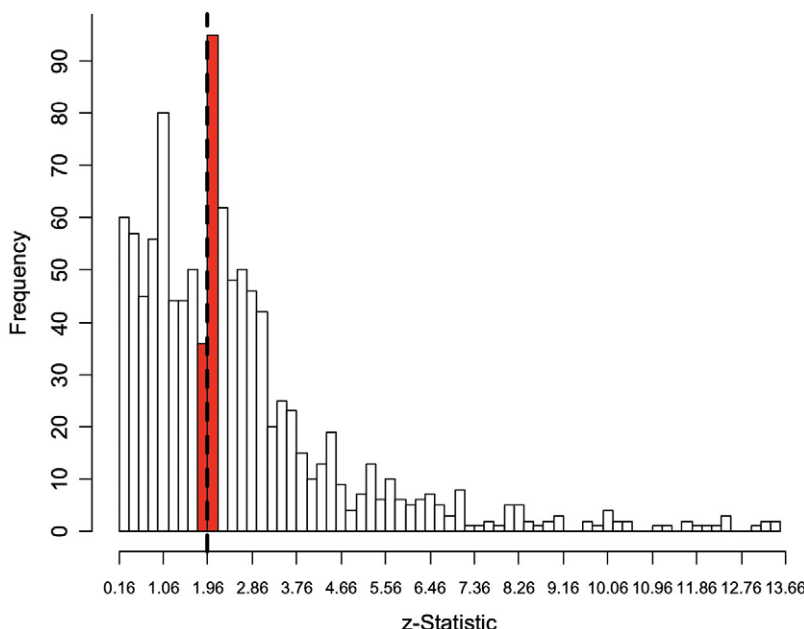


Figure 2. “Histogram of z -statistics, *APSR* & *AJPS* (Two-Tailed). Width of bars (0.20) approximately represents 10% caliper. Dotted line represents critical z -statistic (1.96) associated with $p = 0.05$ significance level for one-tailed tests” (reproduction of Figure 1a from Gerber & Malhotra, 2008a, p. 317).

(i.e., nonsignificant results) were less likely to be published. They find similar patterns in leading sociology journals (Gerber & Malhotra, 2008b) and in the political behavior literature (Gerber et al., 2010).

A great example of publication bias comes from a recent study reporting on National Heart, Lung, and Blood Institute (NHLBI) trials on the effectiveness of various treatments for cardiovascular disease. Since 2000, the NHLBI has required registration of primary outcomes in trials. As can be seen in Figure 3, the number of trials reporting benefits of the intervention went to almost zero after the requirement to register was introduced. This takes into account the relative proportions of trials types as well as industry sponsorship, neither of which was associated with the probability of reporting significant results.

Perhaps registration caused medicines to stop working! Or, probably more likely, prior to the registration of outcomes being required, researchers were more prone to only report and publish statistically significant results.

Issues such as HARKing, publication bias, and fishing badly undermine the credibility of social scientific research. But in fact, the problem is worse. These issues undermine our ability to *know* things and, at the risk of Rumsfeldian logic, to know whether we know things.

Put plainly, our ability as social scientists to advance knowledge is weakened. This is a problem we should take seriously, because if we want to push research and knowledge forward, we need to have confidence that the body of work we are building on rests on solid foundations. Moreover, insofar as social science research informs policy debates and policy design, those debates and policies will be of poorer quality if research is marred by these kinds of deficiencies.

Pre-analysis plans as part of the solution

One solution to the problems outlined here is the idea that researchers *ex ante* declare their design, what hypotheses they are testing, and what those tests look like and then deposit this pre-analysis plan with a publicly accessible registry. In this way, we can decrease the risks of issues such as HARKing, fishing, and eventual publication bias.

If you say what hypotheses you are testing before you even have any data, it becomes a lot harder to adjust those hypotheses *ex post* to fit whatever results you got. If you write down and then, in essence, publicly declare the exact tests you are going to perform, your freedom to *p*-hack your way to a significant result is severely

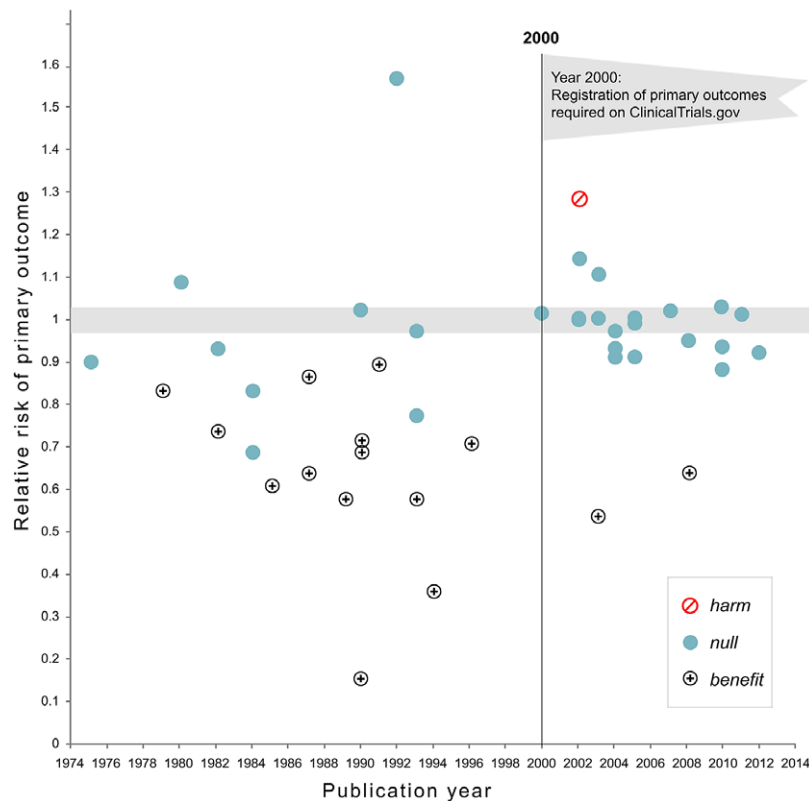


Figure 3. “Relative risk of showing benefit or harm of treatment by year of publication for large NHLBI trials on pharmaceutical and dietary supplement interventions” (reproduction of Figure 1 from Kaplan & Irvin, 2015, p. 5).

curtailed. And if we have a public record of all these designs, regardless of whether they led to novel or even interesting results, we have a much more complete picture of the research landscape than if we simply rely on what makes it through a publication process that massively rewards novel, statistically significant results over null findings. Moreover, if this practice becomes the standard, publishing norms and tastes for null findings will likely also change, making it easier to publish such results.

This all sounds great and quite simple. However, it turns out there is no standard format for PAPs or what ought to be included in them, nor is there much agreement on how stringently to adhere to them. In perhaps the most comprehensive treatment of PAPs in political science, Ofosu and Posner (2021) present evidence from an analysis of PAPs in political science along with a survey of PAP authors. They find there is a large amount of variation in what is included in PAPs. Moreover, their stock-taking of the ability of PAPs to achieve the goals their proponents claim reveals mixed results. As the authors point out, this is at least in part likely due to

the fact that their analysis covers only the earliest days of registration of PAPs in political science and economics and that researchers writing PAPs now are almost certainly doing so in a more sophisticated and detailed way (Ofosu & Posner, 2021, p. 11). They also make the case that at least as important as the specific contents of the PAP itself are the set of complementary institutions and norms that accompany their use. For example, we need more clarity on what should be included in PAPs. And we need to know what to do with them once they are written.

What’s in a PAP?

What should we include in a PAP? There is surprisingly little consensus on this. One minimalist approach is that of the Wharton School Credibility Lab’s As Predicted tool (aspredicted.org), which asks researchers to answer eight questions to create a PAP:

1. Have data been collected already?
2. What is the main question being asked or hypothesis being tested in this study?

3. Describe the key dependent variable(s) specifying how they will be measured.
4. How many and which conditions will participants be assigned to?
5. Specify exactly which analyses you will conduct to examine the main question/hypotheses.
6. Describe exactly how outliers will be defined and handled, and your precise rule(s) for excluding observations.
7. How many observations will be collected or what will determine sample size?
8. Anything else you would like to pre-register?

On the other end of the spectrum are tools such as DeclareDesign (Blair et al., 2019b) that encourage researchers to fully declare every aspect of their design, including how tables and figures will look in the eventual paper. It should be noted here and will be discussed more later that none of these models for PAPs seeks to tie researchers' hands so completely that no room for maneuvering exists. Whatever approach one takes, the main objectives of the PAP are to reduce the risk of the problems outlined earlier in this article and to improve the design.

What do we do with PAPs?

A second question without a particularly satisfactory answer is what, in fact, we use PAPs for in political science. It is all well and good to write a PAP every time we conduct an experiment and to teach our students that they should do the same, for the good of research transparency, open science, and credibility. But then what? I am one of the editors of the *Journal of Experimental Political Science* (JEPS). Sometimes submitted papers include a PAP. Sometimes they do not (JEPS does not require PAPs, unlike, for example, the new policy for experimental papers at the *Journal of Politics*). When a submission does include a PAP, the contents vary enormously. I read the PAP, and then I read the paper and try to make some determination of how well the latter reflects the former. I am sensitive to the fact that sometimes plans change, things happen, and exploratory analysis is done, but I make sure to instruct authors to clearly label these as such. If the paper is sent out for review, I give referees some instructions about reading the PAP. These are not extensive but simply remind referees there is a PAP, that I would like them to assess how well the authors have stuck to their plan, and to assess the authors' justification for any departure from

their PAP. I also remind referees that deviations are OK, but they need to be justified. What I want to avoid is that the PAP is either ignored (why would authors keep writing them if we do not pay attention?) or taken as an iron-clad promise that cannot be bent at all. But, based on referee reports, it is apparent there is massive divergence in terms of how referees treat PAPs. Some probably do not read them at all. Some take the other extreme and want to reject the paper if any deviation is made.

My point is that we need a clearer set of norms for how the peer-review process ought to deal with PAPs. Journal policies differ, editor preferences and behavior differ, and referee preferences and behavior differ. Some divergence is natural and even desirable. But I believe we need a certain level of consistency and clarity. This also applies to situations in which a PAP is not included with a submitted manuscript, and none exists. There are many reasons why a PAP might not exist. For example, it was not *that* long ago that we started with PAPs, and some experimental research projects predating that time are still producing papers. If they were done today, in all likelihood, they would have a PAP. But they do not. Should they be desk rejected? My view is definitely not. I believe the discipline is sensible enough not to throw the baby out with the bath water. I do not think many would argue it should not be published, as a rule. However, I also believe such work will and should be scrutinized differently than if it were accompanied by a PAP. To put it differently, preregistration of a PAP does not guarantee the design is good, nor does the lack of one suggest the design is bad, by definition. As referees, readers, and editors, we have to always evaluate based on design—not based on the presence of a PAP.

What then, do we do when a PAP exists? My own view is that authors ought to always include the PAP in an appendix when submitting the paper. There is little reason not to. When submitting the paper and the PAP, authors should also include a brief appendix outlining any deviations from the PAP. Again, I do not think anyone would argue that we should dismiss papers deviating from their PAP, but we do need to know what these deviations are and how they are reflected in the eventual analysis and write-up. For example, in a recent paper, my coauthors and I include in our appendix the full PAP but also a section detailing deviations from the PAP and a justification or interpretation of how those deviations are reflected in the paper (Slough et al., 2021b). The deviations and their justification range from the relatively trivial,

Deviation description: Re-orientation of the resource index outcome such that higher values indicate higher levels of use/extraction. **Justification/Interpretation:** This is simply an aesthetic change for presentation of results. No aspect of our analysis (other than negating the outcome) or interpretation is changed by this deviation. (Slough et al., 2021a, 39)

to the slightly more consequential,

Deviation description: Estimation of meta-analyses excluding ITT estimates from Brazil. **Justification/Interpretation:** In Section 12 #1, the PAP implies that implementation failures constitute grounds for exclusion. However we did not sufficiently specify a metric for implementation failures. As such, the estimates without Brazil represent a ‘partial’ deviation. Brazil achieved much lower levels of uptake of monitoring than the other sites (Figure 1). Pooling this site with the other estimates changes our definition of monitoring and may be misleading. We present all meta-analysis estimates on the full set of sites (as prespecified) as well as the set of sites excluding Brazil. (Slough et al., 2021a, p. 39)

The takeaway is that transparency is never a bad thing, and we are up front with readers about where we have drifted from the PAP and offer an explanation of how this affects the interpretation of that particular aspect of the paper. In other deviations, we flag how some results presented are exploratory. At the end of the day, it is up to readers whether or how much they want to discount such things. But it is up to authors to provide the information to make such an assessment possible.

When an editor receives a paper with a PAP, I think they ought to check whether the paper sticks to the PAP or, where deviations have occurred, whether they are adequately explained and documented. Serious deviations or undocumented ones should be grounds for desk rejection. When referees get papers to review with PAPs, they should expect a brief note from the editor explaining that this check has been done (but, of course, it is still up to referees to make their own check and assessment). Standards in political science (and other social sciences) are clearly evolving, and there is no set of agreed-upon procedures for assessing papers with PAPs, but at a minimum, referees should do a similar check as editors and recommend changes or rejection based on that. But they should also, as I have argued, not

take the overly stringent approach of only accepting anything that completely adheres to what was preregistered. Such practice will no doubt lead to some backlash against including PAPs, and it is simply not in the spirit of good science.

Critiques of PAPs

PAPs come with many advantages. both to the credibility of science in general, of course, but also to individual researchers. When it comes to researchers, PAPs can serve to clarify thinking around designs, they can get a lot of work that will end up in papers done up front, and, if presented or discussed with colleagues (as they should be), they can catch mistakes. In short, they can contribute to better research designs and thereby better papers. While the use of PAPs is increasing rapidly, journals are beginning to require them, and students are gradually seeing them as standard, there is still hesitation and some opposition. Some of this tentativeness is driven by lack of understanding or outright resistance to more transparency. But a lot of it is also the result of real worry that has to be taken seriously.

PAPs suppress discovery

Among the most common objections to PAPs is that they will suppress discovery. If we tie our own hands, we will miss out on the serendipitous discovery that sometimes results from exploratory work. While this sounds like a dangerous path to get on, I do not believe that many (or probably any) advocates of PAPs are arguing that exploratory work should be banned or that the only analysis allowed is that clearly specified in the PAP. On the contrary, exploratory work remains central to the research process. The difference in a world of PAPs is that we simply need to label it as such. Researchers are never prohibited from doing things not in the PAP, they just have to tell the reader that those analyses and results are exploratory and not part of the PAP and explain why they have now been performed.

One possibility for PAP writers is to signal where they want their hands tied really tightly and where they want to give themselves a bit more wiggle room. For example, measurement of outcomes and treatments, treatment assignment, main hypotheses, and other fundamental aspects of a design are areas in which one would think there should be little scope for deviation. Other things, such as model decisions that might depend on the structure of the data that one does know *ex ante*, could be

where authors remain a bit agnostic in the PAP, or where they give a few options of what they will do depending on how things look once the study is done. A key point here is that all of these options need to be justified in some principled way in the PAP. Again, being transparent about these things *ex ante* is probably never a bad idea.

PAPs take too much time

Another common critique of PAPs is that they take too much time. I do not think there is any doubt that in order to satisfy the main objectives of a PAP, it will take time. Ofosu and Posner (2021) surveyed social scientists who write PAPs and found that almost 90% reported spending at least a week on writing a PAP. About a third said they spent two to four weeks, and about a quarter said they spent more than a month on writing a PAP for a typical project (Ofosu & Posner, 2021, p. 7). In other words, for most researchers, the PAP is far from a trivial time commitment in the overall research process.

However, I think the criticism or worry about PAPs taking too much time is somewhat misplaced. First, although it takes time to write, the PAP is not wasted time. This is usually effort that needs to be exerted at some point in the research process anyway. And if done well, the PAP makes writing the eventual paper much easier. Some even go so far as to basically write the entire paper, including doing mock analysis on simulated data, producing the tables and figures that would go in an eventual paper, and so on (Blair et al., 2019a, 2021). So the PAP, if done well, would not necessarily add huge amounts of time and effort, and perhaps it could even save you time in the end by front-loading the effort.

Even if it is the case that some extra time and effort are required, this is not without benefit. One of the chief rewards of engaging in this effort is that it improves designs, leading to better research. The act of having to deliberately write out each important decision one will make in the research forces the researcher to think harder about these choices and trade-offs. Preregistering a design and analysis plan thus has more benefits than aiding in research transparency. It can also tell us something about design: whether our research questions are well posed, whether our strategies for collecting and analyzing data as well as our strategy for reporting the results to answer our question hang together in appropriate ways.

Along these lines, PAPs also provide an opportunity for feedback at the stage in the research process when it is most useful: at the design stage. While it is fun and can be rewarding to present fully complete papers and findings,

it is far more useful in the development of research to have the opportunity to get advice and help *before* the study is conducted. For example, one of the chief roles served by EGAP is to provide opportunities for the presentation and discussion of research designs. Many other small, mostly regional, meetings such as the Northeast Workshop in Empirical Political Science, the Working Group in African Political Economy, the Toronto Political Behaviour Workshop, and others are doing the same. This is particularly useful when there are partner organizations involved in the study, collaborating with researchers. Having partners involved at the design stage increases transparency between researchers and partner organizations, makes buy-in from partners easier to achieve, and likely leads to better design choices since partners often provide a useful perspective. Moreover, doing so provides more benefit to partners as they are a more involved party in the research (see Loewen & Rubenson, 2021, for a discussion of transparency and adding value when doing experiments with partners).

Other objections

Other objections raised to the increased adoption of PAPs include worries that researchers with fewer resources will be further disadvantaged; that the use of PAPs will lead to bias in the topics of study pursued by researchers; and that if we start requiring or expecting PAPs, it will put observational studies in an awkward position. While all of these are real concerns in research, it is not clear they are made worse (or better for that matter) by a more widespread use of PAPs. A PAP need not require any additional resources beyond the time to write it (and on that point, there may be efficiencies). And if there are biases in terms of the types of questions that are pursued (there may well be), they are likely not related to the prevalence (or dearth) of PAPs. It is certainly the case that PAPs raise interesting questions for observational and qualitative studies, but there are proposals on this front as well (see, e.g., Jacobs, 2020).

One important concern I would raise is the risk of applying a one-size-fits-all approach to PAPs. I do think there are some common elements that any PAP should include—we need to know the research question, the data collection strategy, randomization strategy, the measurement of outcomes, the tests to be performed—but the way in which these elements are included and discussed and the emphasis placed on them might differ depending on the discipline and on the type of study. Of course, observational versus experimental designs

represents a big difference. But there should not be many salient differences between, say, field, lab and survey experiments that all involve the random assignment of some treatment. Across these types of experiments, we still need to know about the same kinds of decisions. One area in which PAPs certainly differ is in the amount of other detail they contain. Some researchers include their entire theoretical discussion and the motivation for the study. Some include all tables and figures that will be in the final paper and analyze mock data. Others stick to the main elements of PAPs described above such as the research question, hypotheses, and tests. There are no rules about this—what is necessary is that we have enough information to avoid as best as possible the problems described earlier.

What now?

The rapid rise of experimental designs and, more recently, of the practice of preregistering analysis plans is a positive development in political science. It is still early days, and as a discipline, we are learning as we go. We need to develop clearer expectations of what is included in PAPs along with the norms and institutions that will support their use and further adoption in the discipline. This includes responsibilities for journal editors to better instruct both authors and referees in how to handle PAPs. One excellent development is the implementation of blind acceptance at several journals, including *JEPS* and *Politics and the Life Sciences (PLS)*.³ The idea is simple: Researchers would write up a paper exactly as normal with all the usual parts included, except they would not yet have conducted the study and collected the data. That manuscript is then assessed on the basis of the design, not the results (since there are no results yet). If the editor and referees think the question is important and the design is appropriate, it is agreed that the paper will be published regardless of the eventual results, assuming, of course, that the authors stick to their declared design. This path to publication ought to improve general open science and research transparency practices by providing incentives for researchers to carefully develop PAPs for their studies and incentives for editors and referees to evaluate designs rather than results.

³See *PLS's* Guidelines for Authors of Registered Reports (<https://www.cambridge.org/core/services/aop-file-manager/file/603683551cb1a12a6f125bf6/PLS-RegisteredReportsAuthorsGuidelines-210224.pdf>) as well as Arceneaux et al. (2018).

All of the articles in this special issue of *PLS* use this approach (Friesen et al., 2021; Heisbourg & Feitosa, 2021; Mansell et al., 2021; Ploger et al., 2021). Authors submitted full PAPs that were sent out for blind review, just as a regular study would be. Preregistered designs that passed the review process were then accepted in principle based on the design, regardless of the eventual results. Of course, the journal editors retained the right to reject papers that did not adhere to the PAP, while at the same time making allowances for exploratory analysis as long as it was clearly labeled as such. Once data collection, analysis, and write-up of results were complete, papers were sent for review again. Referees were instructed to assess how faithfully the authors executed their PAP and to make decisions about the justification for any deviations from the PAP.

This is an outstanding way to push research transparency forward. The way this issue of *PLS* has set up the preregistered report publication process ought to serve as a model both in creating incentives for researchers to engage in these practices but also in instructing referees on how to handle preregistered reports. One potential challenge with the preregistered report model of publishing is that referees might try to reject papers at the last stage by raising new arguments or objections. This is where editors need to exercise their prerogative much more strongly to push back against that sort of review. Not only is it unfair to authors, but it also undermines the basic idea of the preregistered report model.

There is still lots of work to be done in our discipline in terms of transparency in the research process. I am learning all the time as a researcher, a referee, a general consumer of research, and an editor. But we are certainly a long way down the right path compared with just a decade or so ago, when these conversations were hardly happening at all.

Acknowledgments

Many thanks to the organizers of this special issue and to the Association for Politics and the Life Sciences (APLS) for inviting me to write on the topic. I have had many conversations over the years with Graeme Blair, Alex Coppock, Chris Dawes, Miriam Golden, Don Green, Macartan Humphreys, Betsy Levy Paluck, and Tara Slough that have taught me much about research transparency and pre analysis plans. I thank them all for continuing to teach me and shape my thinking on the subject.

References

- Angrist, J. D., & Pischke, J.-S. (2010). The credibility revolution in empirical economics: How better research design is taking the con out of econometrics. *Journal of Economic Perspectives*, 24(2), 3–30.
- Arceneaux, K., Wilson, R., Boudreau, C., Bush, S., Jerit, J., Rubenson, D., Sinclair, B., Settle, J., Woon, J., & Zechmeister, E. (2018). The value of preregistration for open and credible science: An initiative from the journal of experimental political science. *Journal of Experimental Political Science*, 5(3), 165–166.
- Blair, G., Cooper, J., Coppock, A., & Humphreys, M. (2019a). Declaring and Diagnosing Research Designs. *American Political Science Review*, 113(3), 838–859.
- Blair, G., Cooper, J., Coppock, A., Humphreys, M., & Fultz, N. (2019b). *DeclareDesign*. [Software package for R]. Retrieved June 5, 2021, from <http://declaredesign.org>
- Blair, G., Coppock, A., & Humphreys, M. (2021). *Research design: Declaration, diagnosis, redesign*. Unpublished manuscript.
- Christensen, G., & Miguel, E. (2020). Transparency and reproducibility: Conceptualizing the problem. In C. Elman, J. Gerring, & J. Mahoney (Eds.), *The production of knowledge: Enhancing progress in social science* (pp. 129–164). Cambridge University Press.
- Friesen, A., Ksiazkiewicz, A., & Gothreau, C. (2021). Political taste: Exploring how perception of bitter substances may reveal risk tolerance and political preferences. *Politics and the Life Sciences*, 40(2), <https://doi.org/10.1017/pls.2021.20>.
- Gerber, A., & Malhotra, N. (2008a). Do statistical reporting standards affect what is published? publication bias in two leading political science journals. *Quarterly Journal of Political Science*, 3(3), 313–326.
- Gerber, A. S., & Malhotra, N. (2008b). Publication bias in empirical sociological research: Do arbitrary significance levels distort published results? *Sociological Methods & Research*, 37(1), 3–30.
- Gerber, A. S., Malhotra, N., Dowling, C. M., & Doherty, D. (2010). Publication bias in two political behavior literatures. *American Politics Research*, 38(4), 591–613.
- Heisbourg, E., & Feitosa, F. (2021). Does music affect citizens' evaluations of candidates? *Politics and the Life Sciences*, 40(2), <https://doi.org/10.1017/pls.2021.22>.
- Humphreys, M., De la Sierra, R. S., & Van der Windt, P. (2013). Fishing, commitment, and communication: A proposal for comprehensive nonbinding research registration. *Political Analysis*, 21(1), 1–20.
- Jacobs, A. M. (2020). Pre-registration and results-free review in observational and qualitative research. In C. Elman, J. Gerring, & J. Mahoney (Eds.), *The production of knowledge: Enhancing progress in social science* (pp. 221–264). Cambridge University Press.
- Kaplan, R. M., & Irvin, V. L. (2015). Likelihood of null effects of large NHLBI clinical trials has increased over time. *PLOS ONE*, 10(8), e0132382.
- Kerr, N. L. (1998). HARKing: Hypothesizing after the results are known. *Personality and Social Psychology Review*, 2(3), 196–217.
- Loewen, P. J., & Rubenson, D. (2021). *Value-added and transparent (VAT) experiments*. Working paper, <https://bit.ly/3iEmt1F>.
- Mansell, J., Mock, S., Rhea, C., Tecza, A., & Peirder, J. (2021). Measuring attitudes as a complex system: Structured thinking and support for the Canadian carbon tax. *Politics and the Life Sciences*, 40(2), <https://doi.org/10.1017/pls.2021.16>.
- Munafo, M. R., Nosek, B. A., Bishop, D. V., Button, K. S., Chambers, C. D., Du Sert, N. P., Simonsohn, U., Wagenmakers, E.-J., Ware, J. J., & Ioannidis, J. P. (2017). A manifesto for reproducible science. *Nature Human Behaviour*, 1(1), 1–9.
- Nosek, B. A., Alter, G., Banks, G. C., Borsboom, D., Bowman, S. D., Breckler, S. J., Buck, S., Chambers, C. D., Chin, G., Christensen, G., et al. (2015). Promoting an open research culture. *Science*, 348(6242), 1422–1425.
- Ofose, G. K., & Posner, D. N. (2021). Pre-analysis plans: An early stocktaking. *Perspectives on Politics*. Advance online publication. <https://doi.org/10.1017/S1537592721000931>
- Ploger, G., Dunaway, J., Fournier, P., & Soroka, S. (2021). The psychophysiological correlates of cognitive dissonance. *Politics and the Life Sciences*, 40(2), <https://doi.org/10.1017/pls.2021.15>.
- Rubin, M. (2017). When does HARKing hurt? Identifying when different types of undisclosed post hoc hypothesizing harm scientific progress. *Review of General Psychology*, 21, 308–320.
- Slough, T., Rubenson, D., Levy, R., Alpizar Rodriguez, F., Bernedo del Carpio, M., Buntaine, M. T., Christensen, D., Cooperman, A., Eisenbarth, S., Ferraro, P. J., Graham, L., Hartman, A., Kopas, J., McLarty, S., Rigerink, A. S., Samii, C., Seim, B., Urpelainen, J., & Zhang, B. (2021a). Supplementary information for “Adoption of Community Monitoring Improves Common Pool Resource Management Across Contexts.” *Proceedings of the National Academy of Sciences*, 118(29), e2015367118.

Tie my hands loosely

Slough, T., Rubenson, D., Levy, R., Alpizar Rodriguez, F., Bernedo del Carpio, M., Buntaine, M. T., Christensen, D., Cooperman, A., Eisenbarth, S., Ferraro, P. J., Graham, L., Hartman, A. C., Kopas, J., McLarty, S., Rigtterink, A. S.,

Samii, C., Seim, B., Urpelainen, J., & Zhang, B. (2021b). Adoption of community monitoring improves common pool resource management across contexts. *Proceedings of the National Academy of Sciences*, 118(29), e2015367118.