

Pre-analysis Plans: A Stocktaking *

George K. Ofofu[†]

Daniel N. Posner[‡]

December 12, 2019

Abstract

The evidence-based community has championed the public registration of pre-analysis plans (PAPs) as a solution to the problem of research credibility, but without any evidence that PAPs actually bolster the credibility of research. We analyze a representative sample of 195 PAPs from the American Economic Association (AEA) and Evidence in Governance and Politics (EGAP) registration platforms to assess whether PAPs are sufficiently clear, precise and comprehensive to be able to achieve their objectives of preventing “fishing” and reducing the scope for post-hoc adjustment of research hypotheses. We also analyze a subset of 93 PAPs from projects that have resulted in publicly available papers to ascertain how faithfully they adhere to their pre-registered specifications and hypotheses. We find significant variation in the extent to which PAPs are accomplishing the goals they were designed to achieve.

*The authors thank Maxim Ananyev and Merabi Chkhenkeli for excellent research assistance, and Graeme Blair, Alex Coppock, Steven Glazerman, Don Green, Macartan Humphreys, Matt Lisiiecki, and David McKenzie, as well as seminar participants at EGAP, Oxford and Berkeley, for valuable comments. The authors gratefully acknowledge funding from the Social Science Meta-Analysis and Research Transparency (SSMART) program of the Berkeley Initiative for Transparency in the Social Science (BITSS). The survey of researchers who register pre-analysis plans, which provides some of the data reported in the paper, was determined to be exempt from IRB review (UCLA IRB # 19-000063). We registered our study at the Open Science Framework (OSF) registry:<https://osf.io/xrtqm/>.

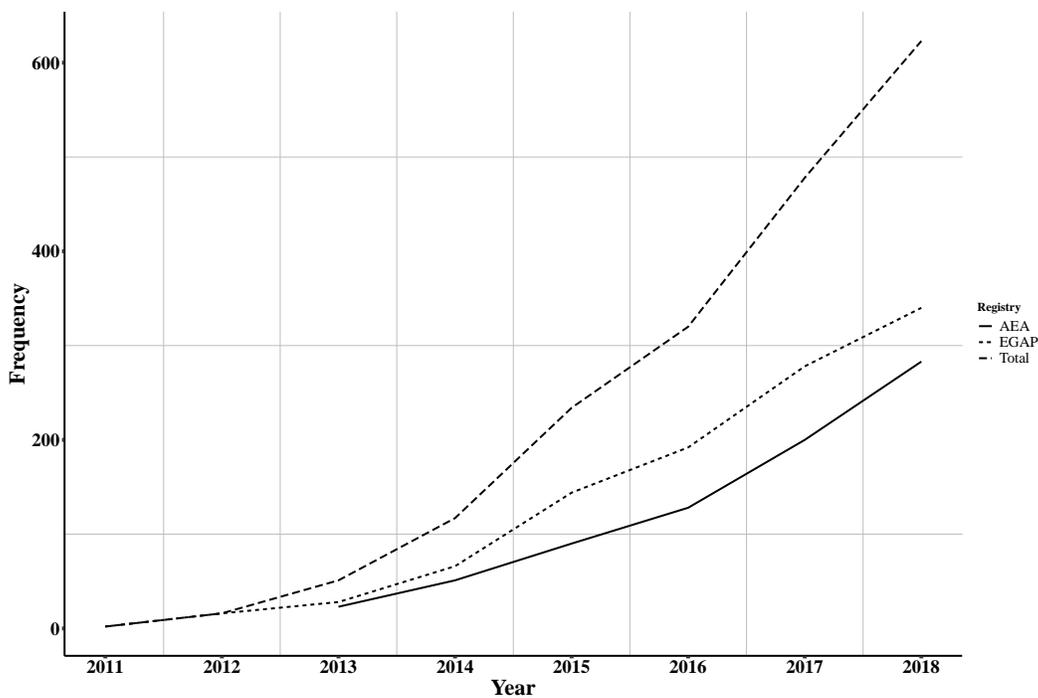
[†]Department of Government, London School of Economics and Political Science, Email: g.ofosu@lse.ac.uk.

[‡]Department of Political Science, University of California, Los Angeles. Email: dposner@polisci.ucla.edu

1 Introduction

Pre-analysis plans (PAPs), public documents that specify in advance the hypotheses a researcher will investigate and how the data will be collected and analyzed, have been championed as an important tool for addressing the problem of research credibility in the social sciences (Ioannidis, 2005; Franco, Malhotra and Simonovits, 2014; Simonsohn, Nelson and Simmons, 2014; Open Science Collaboration, 2015; Christensen, Freese and Miguel, 2019).¹ Their use has skyrocketed in recent years. As shown in Figure 1, which displays the number of PAPs registered on the American Economic Association (AEA) and Evidence in Governance and Politics (EGAP) registries since 2011, their numbers have grown exponentially.² Graduate students are now taught that registering a PAP is a *de rigeur* part of undertaking their research projects.

Figure 1: PAP registrations on the AEA and EGAP registries, 2011-2018



¹PAPs are a special case of pre-registration, which involves publicly declaring one's intention to undertake a study that investigates a particular hypothesis. PAPs go beyond pre-registration by also providing specific details about how the proposed analysis is to be undertaken.

²Our stocktaking focuses on patterns in Economics and Political Science, and thus on the two major registries in these disciplines. Other prominent social science registries, whose contents we do not review, include the Registry for International Development Impact Evaluations (RIDIE), the Open Science Framework (OSF) Registry, and the website AsPredicted. In 2020, the EGAP Registry will merge with the OSF Registry.

PAPs are advocated for two main reasons. First, they prevent “fishing” (also referred to as “*p*-hacking” or “data mining”). Fishing is the practice of selectively reporting, from among the many possible results that might be generated from a given set of data, the subset of findings that are statistically significant, novel, or allow the researcher to tell a cleaner or more compelling story.³ PAPs solve this problem by specifying in advance exactly which econometric specifications, outcome variables, coding rules, covariates, sub-samples, and inclusion rules will be used to generate the results that will be presented as the definitive test of the research question. Specifying the key details of the analysis in advance reduces the “researcher degrees of freedom” that provide latitude for consciously or unconsciously selecting particular specifications that make the results more striking.⁴

Second, PAPs prevent hypothesizing after results are known (sometimes abbreviated as “HARKing”). HARKing involves interpreting results *ex post* based on the results of the analysis rather than *ex ante* based on expectations derived from theory. PAPs solve this problem by specifying in advance which hypotheses a researcher is intending to test, thus preventing the researcher from succumbing to hindsight bias and emphasizing in the presentation of her findings the hypotheses that happened to find support in the data (Nosek et al., 2018). Registering research hypotheses in advance in a PAP need not prevent researchers from using their data to conduct exploratory research. Pre-registration simply clarifies which of the analyses the researcher ultimately presents in the paper are confirmatory (i.e., testing predictive hypotheses specified before the results were known) and which should be treated as exploratory (i.e., products of learning and new hypothesis generation based on the patterns that emerged in the data). Both confirmatory and exploratory findings can be sources of insight, but the evidentiary status of each is quite different.

These two benefits of PAPs are clear and, for those committed to improving the credibility of social science research, compelling. But whether PAPs are actually achieving these goals in practice is an empirical question—albeit an extremely challenging one to answer definitively.⁵ One cannot compare

³For an illuminating illustration of the scope for fishing within a real study, see Casey, Glennerster and Miguel (2012). For evidence of the prevalence of fishing in Economics, see Brodeur et al. (2016); for Political Science, see Gerber and Malhotra (2008).

⁴For a discussion of researcher degrees of freedom, see Simmons, Nelson and Simonsohn (2011) and Wicherts et al. (2016). For discussions of the incentives researchers have to generate more striking results, see Nosek et al. (2018) and Laitin and Reich (2017).

⁵For a notable attempt to estimate the causal effect of registration in the medical field, see Fang, Gordon and Humphreys (2015).

the degree of fishing and post-hoc hypothesis adjustment in studies implemented with and without PAPs because, absent a PAP, there is no record of the analyses or hypotheses that were pre-specified. And even if such a comparison were possible, the conclusions one could draw would be undermined by the fact that researchers self-select into whether or not they pre-register their analyses, and the researchers who file PAPs are quite likely different from those who do not. Moreover, even researchers who regularly file PAPs do not register them for all of their studies, so the lack of randomness in who pre-registers a PAP is compounded by within-researcher selection across projects.⁶

We therefore adopt a different approach. We draw a representative sample of PAPs and analyze their content to determine whether they are sufficiently clear, precise, and comprehensive as to meaningfully limit the scope for fishing and post-hoc hypothesis adjustment. We also assess whether PAPs do, in fact, tie researchers hands by comparing publicly available papers that report the findings of pre-registered studies to the PAPs that were registered when those studies were initiated. These are, of course, subjective judgements. But we have undertaken in our coding rules and our procedures to be both transparent and objective in the judgements we make. We also underscore that none of our findings should be taken as evidence of the causal effects of pre-registration. As noted, such an analysis would require knowing what researchers would have pre-specified had they done so, as well as a random assignment of projects to be pre-registered. Notwithstanding these unavoidable weaknesses, we believe that our analysis can provide an illuminating assessment of whether PAPs, as they are actually written and used, are able to accomplish the main objectives that have motivated their widespread promotion and adoption.⁷

The importance of such an assessment is rooted in the significant costs associated with writing and following a PAP (Olken, 2015; Coffman and Niederle, 2015; van't Veer and Giner-Sorolla, 2016). The modal researcher in our PAP users' survey (discussed below) reports spending 2-4 weeks preparing her pre-registration materials, and more than a quarter of researchers report spending more than a month. Beyond the time they take to write, the hand-tying that PAPs entail is claimed to limit the scope for

⁶In our PAP users' survey (discussed below) 78% of researchers said they had at least one ongoing research project for which they did not register a PAP.

⁷An analogous stocktaking, involving the comparison of published and unpublished papers with the proposals that secured their funding, is provided in Franco, Malhotra and Simonovits (2014). For a similar effort in Psychology, see Claesen et al. (2019).

breakthroughs that come from unexpected, surprise findings and to generate boring, mechanical papers that are not favored by reviewers or journal editors. PAPs are also said to force researchers to undertake analyses that they know to be inappropriate or sub-optimal once they have encountered their data. Other critics point out that, whatever the benefits of pre-registration may be in theory, PAPs are unlikely to enhance research credibility without vigorous policing—something the disciplines generally do not reward (Laitin, 2013; Laitin and Reich, 2017). Still others argue that publicly posting the details of one’s proposed analyses creates a risk of getting “scooped.” This is especially a concern for junior scholars and other researchers who may lack the resources to quickly implement a promising research design. While there are good responses to many of these objections (many of which we discuss below), they nonetheless underscore the importance of assessing how much weight should be put on the positive side of the pre-registration ledger. Doing so requires taking the stocktaking exercise we undertake here.

2 Methods

2.1 Sampling

To evaluate whether PAPs are written sufficiently clearly and comprehensively to achieve their intended objectives, we drew a representative sample of PAPs from the universe of studies registered on the AEA and EGAP registries between their initiation and 2016. Although the web forms that investigators complete when registering their studies on both of these sites provide opportunities for describing many details of the proposed research, including much of the information that ordinarily goes into a formal PAP, our analysis only includes studies for which a formal PAP was uploaded.⁸ Because we were interested not just in the PAPs’ contents but also in how those contents shaped the reporting of the research that was undertaken, we drew our sample so that roughly half of the PAPs would be from studies that had resulted in publicly available papers (journal articles or working papers).

Our procedures, described in detail in Appendix A, yielded a sample of 204 PAPs, equally distributed between those with and without publicly available papers. In nine instances, working papers that had been found on authors’ websites at the time we included the PAPs in our sample were no longer

⁸To the extent that the information provided in PAPs is more complete than the information provided on registry web forms alone, our findings are likely to represent an upper bound on the hand-tying provided by pre-registration more broadly.

publicly available by the time we began our coding. We therefore coded 93 PAPs with papers, bringing our final sample of coded PAPs to 195.

2.2 Coding

All 195 PAPs in our sample were coded according to a common rubric that recorded details of the hypotheses that were pre-specified, the dependent and independent variables that would be used in the analysis, the sampling strategy, inclusion and exclusion rules, and the statistical models to be run, among other features. For the sub-sample of 93 PAPs for which publicly available papers were available, we added further questions that addressed how faithfully the study authors adhered to the pre-specified details of the analysis in the resulting paper. The complete coding rubric for PAPs with papers is provided in Appendix B. All PAPs were coded by at least two different people—a research assistant and one of this paper’s authors—and any discrepancies between them were investigated and recoded.

Although much of the information collected in the coding rubric was straightforward and unambiguous—for example, whether the PAP was registered prior to data collection; whether the PAP included a power analysis, committed to a multiple testing adjustment, or was ever gated or private—a number of the key coding items involved subjective judgements. Chief among these was whether the main research hypotheses and the key causal and outcome variables were specified sufficiently clearly as to prevent post-hoc adjustments. For the latter, our coding rules asked the coder to consider, following Olken (2015), whether “if you gave the PAP to two different programmers and asked each to prepare the data for the primary dependent/independent variable(s), they [would] both be able to do so without asking any questions, and they [would] both be able to get the same answer.” As for the clarity of the research hypotheses, we defined a “clear hypothesis” as one that describes a relationship between an independent and dependent variable in which the direction of the effect is specified.

In the discussion that follows, we occasionally draw on examples from the PAPs we analyzed to illustrate our points. When we do so, we change the details to protect the anonymity of the PAP authors. This is in keeping with our goal of identifying broad patterns in how PAPs are written and used, not singling out individual authors for particularly weak (or strong) practices.

2.3 Survey of PAP users

We supplemented our coding of PAPs with an anonymous survey of PAP users to elicit their experiences with writing and using PAPs in their research. We were especially interested in collecting information about how the practice of composing and registering a PAP had changed the ways in which the investigators went about their work. The survey (reproduced in full in Appendix C) was sent to all affiliated researchers in the EGAP and Innovations for Poverty Action (IPA) research networks (N=664). We received 155 responses, of which 81% reported having registered a PAP for at least one project and 60% reported having registered multiple PAPs.

3 Are PAPs achieving their objectives?

Before turning to our findings, it will be useful to say something about the sample of PAPs on which our stocktaking is based. The overwhelming majority of the 195 PAPs we coded were from field (63%), survey (27%), or lab (4%) experiments; observational studies comprised just 4% of our sample.⁹ Eighty-one percent of PAPs were registered on the AEA or EGAP websites prior to data collection, and another 19% were registered after data collection but before the researchers had access to their data or began their analysis.¹⁰ Among the PAPs with papers, 66% were working papers and 33% were journal articles. In keeping with their share in the population of PAPs registered on the AEA and EGAP registries during the period we studied, 45% of the PAPs we coded were registered in 2016. The fact that such a large share of our sample comes from the final year of our analysis speaks to the exponential growth in the registration of PAPs over the period we studied. It also (somewhat) allays concerns that the findings we present come from the early period of PAP usage, when researchers were still just learning how to use PAPs as tools in their research.

⁹Tellingly, among the 77% of researchers in our PAP users' survey who reported having other projects for which they did not register a PAP, the most commonly cited reason was because they were observational studies.

¹⁰In the PAP users' survey, several researchers said they hesitated to register PAPs for studies drawing on data that was, in principle, available to them prior to drafting the PAP, as there was no way to prove that they had not looked at the data. This may account for the lower share of studies registered after data collection had commenced.

3.1 Do PAPs reduce the scope for fishing?

Fishing is made possible by imprecise variable definitions and by lack of clarity about the statistical models that will be run, the covariates that will be included, and the rules that will be applied for excluding cases, among other details of the analysis that will be undertaken. The failure to clearly specify these aspects of the research design in advance provides scope for researchers to run their analyses multiple ways and then present as their “test” of the hypothesis in question the specification that happens to generate the most appealing results.¹¹ This can happen either nefariously (by researchers searching for findings that they think are more likely to be published or bring them renown) or inadvertently via post-hoc rationalization (“Of course this was the right specification to run! Silly of me not to have seen this at the outset!”)—a skill at which human beings are dangerously accomplished (Nosek et al., 2018). Whatever the source, fishing undermines the credibility of the research findings.

One of the key features we coded in our sample of PAPs was whether the primary dependent and independent/treatment variables were operationalized sufficiently clearly as to prevent post-hoc adjustments. Examples of lack of clarity include defining the outcome of interest as “political participation” without specifying how political participation is to be measured or hypothesizing that the partisan alignment of a radio station will matter for the impact of the messages it broadcasts, with partisan alignment determined through “content analysis of programming.” Promising to “create an index” without specifying exactly how the index is to be constructed offers another illustration. None of these examples would pass the Olken test described above.

These violations are relatively rare, however. In our sample of PAPs, 77% of primary dependent variables and 93% of independent/treatment variables were judged to have been clearly specified.¹²

PAP authors were less good, however, at clearly specifying their control variables. Many PAPs indicated the researchers’ intentions to “include baseline controls to improve precision” or to control for imprecisely defined covariates such as “wealth,” “demographic characteristics,” “employment status,” or “cognitive ability.” While these variables may well be relevant to include, describing them in the PAP

¹¹See Humphreys, De la Sierra and Van der Windt (2013) for a simulation-based exercise demonstrating the scope for generating erroneously significant results due to poor pre-specification of different aspects of the research design.

¹²The high rate of clearly specified main independent variables stems from the fact that in most cases—90% in our sample—this variable was simply a treatment dummy whose details were unambiguous.

in such broad and non-specific terms leaves wide scope for fishing at the data analysis stage.¹³ Lack of clarity in variable definition is not the only issue. In 44% of PAPs, the number of pre-specified control variables was judged to be unclear, making it nearly impossible to compare what was pre-registered with what is ultimately presented in the resulting paper. Lenz and Sahn (2019) show that the flexibility stemming from such imprecision provides wide scope for generating results that might not otherwise have reached traditional levels of statistical significance.¹⁴

Further scope for fishing comes from imprecision in the empirical models that are pre-specified.¹⁵ Insofar as researchers can generate different results if they run their analyses using ordinary least squares, weighted least squares, multinomial logit, or other approaches—not to mention with or without particular adjustments for calculating standard errors—it is critical to commit in advance to a particular statistical model. Sixty-eight percent of PAPs were judged to have spelled out the precise statistical model to be tested; 37% specified how they would estimate their standard errors. In 19% of cases, the models presented in the resulting papers deviated from the models specified in the PAP—for example, two stage least squares was run when ordinary least squares was pre-specified; controls were added or omitted; covariate adjustment was specified in the PAP but not undertaken in the paper. Such deviations are not a problem if they are noted and a rationale is provided for the divergence from what was pre-registered. However, in the fourteen instances in our sample where deviations occurred, the change was noted in only one case.

Additional latitude for specification searching comes from lack of clarity about the rules that researchers will apply to include or exclude units from their analyses and, in experimental work, to deal with unanticipated imbalances across treatment and control groups. Such rules are important because unforeseen implementation challenges—attrition, noncompliance, project delays, problems with

¹³Even when attempts are made to clarify how such variables are to be measured, the clarifications are often equally problematic. For example, defining “wealth” as an index based on characteristics such as the condition of a respondent’s dwelling, a household’s ownership of animals, agricultural equipment, and other assets, or the number of days a household goes without food still leaves broad latitude for subjectivity (which dwelling conditions? which specific animals? what kinds of agricultural equipment? what if there is enough food for some family members but not others?) and fails the Olken test.

¹⁴Lenz and Sahn (2019) find that 30-40% of observational studies report findings that depend on covariates to increase their effect sizes to the point where they cross the threshold of statistical significance, and that the authors of these studies almost never disclose that their results depend on the particular constellation of covariates they have chosen to include.

¹⁵The simulations in Humphreys, De la Sierra and Van der Windt (2013) suggest that discretion over model selection is not a major source of fishing opportunities. However, the test presented in the paper is for discretion over using linear, logit, or probit models for binary variables, and may not apply to other aspects of model choice in other applications.

randomization—often force researchers to make fixes at the analysis stage that can bias the results, intentionally or unintentionally, toward a particular conclusion. Twenty-five percent of PAPs specified how they would deal with missing values and/or attrition; 13% specified how they would deal with non-compliance; 8% specified how they would deal with outliers; and 20% specified how they would deal with covariate imbalances. It would appear that study authors are less careful about pre-specifying what they will do if their implementation does not go according to plan than they are about pre-specifying other details of their proposed analysis. While all of the studies for which rules about missingness, non-compliance, and outliers were pre-specified followed them in the resulting papers, the fact that so many PAPs were silent on these issues underscores the incompleteness of most PAPs—and the opportunities that such omissions provide for researchers to tweak their analyses in ways that generate particular results.

The practical difficulties with pre-specifying responses to every possible implementation problem that might arise has led to the development of standard operating procedures (SOPs)—a set of default practices adopted by a lab or research group to which study authors can commit in advance to guide decisions that are not addressed specifically in the PAP (Lin and Green, 2016). However, notwithstanding the utility (and time savings) that might come from invoking SOPs, just 3% of the PAPs in our sample indicated that they would rely on SOPs to deal with unanticipated deviations from their pre-registered designs.

3.2 Do PAPs reduce the scope for post-hoc hypothesis adjustment?

The clearest strategy for eliminating the scope for post-hoc hypothesis adjustment is to specify the hypotheses in a way that leaves no ambiguity about the propositions that the research will test. In this respect, PAP authors do quite well. Ninety percent of the PAPs we coded were judged to have specified clear hypotheses.

However, even clearly specified hypotheses can leave scope for HARKing if authors pre-specify so many hypotheses that they can pick and choose which ones to report after they have seen their results. In this respect, PAP authors fared less well. While 34% of PAPs specified between one and five hypotheses—a number sufficiently small as to limit the leeway for selective presentation of results

downstream—18% specified between six and ten hypotheses; 18% specified between 11 and 20 hypotheses; 21% specified between 21 and 50 hypotheses; and 8% specified more than 50 hypotheses (see Panel A of Figure 2). PAPs that pre-specify so many hypotheses raise questions about the value of pre-registration.¹⁶

Figure 2: Number of pre-specified hypotheses

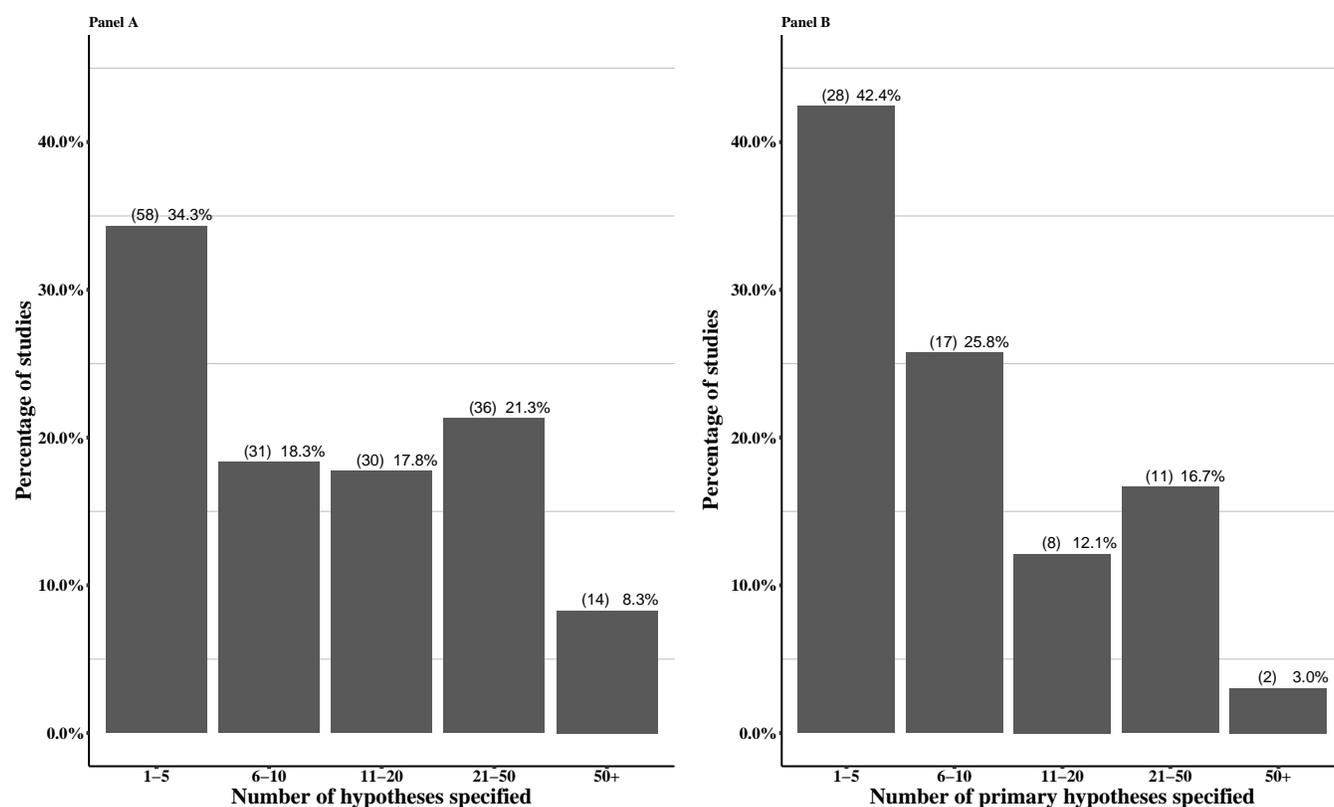


Figure 2 shows the distribution of the number of hypotheses pre-specified in the full sample of PAPs (Panel A) and in the subset of PAPs that pre-specified more than 5 hypotheses in all and that distinguished between primary and secondary hypotheses (Panel B).

One safeguard against this pitfall is to distinguish between primary and secondary hypotheses. Many PAPs adopt this protection: among authors who pre-specified more than five hypotheses, 60% make such a distinction. But they often do so in ways that do little to solve the underlying problem. As shown in Panel B of Figure 2, 42% of PAPs that distinguished between primary and secondary hypotheses limited the number of primary hypotheses they specified to five or fewer. Twenty-six percent

¹⁶Closely related to the number of hypotheses is the length of many PAPs. While the median PAP in our sample was 11 single-spaced pages, the longest 10 percent were more than 31 pages, and three were over 90 pages long.

pre-specified 6-10 primary hypotheses; 12% pre-specified 11-20; 17% pre-specified 21-50; and 3% pre-specified more than 50. From the standpoint of reducing the scope for selective presentation of research findings, distinguishing between primary and secondary hypotheses is only useful if the number of primary hypotheses is kept small.

Another safeguard is to pre-commit to a multiple testing adjustment. Multiple testing adjustments down-weight the statistical significance of any single result based on the number of hypotheses that are being tested, thus guarding against the cherry-picking results in instances where there are many possible findings to choose from and the chances of generating a false positive are high. Among the PAPs in our sample that pre-specified more than five hypotheses, 28% pre-committed to a multiple testing adjustment.

Taken together, these practices leave significant leeway for authors to omit results that are null or that complicate the story they wish to tell. But do authors take advantage of this latitude in practice? To find out, we examined the sub-sample of 93 PAPs we coded that had publicly available papers and compared the primary hypotheses pre-specified in the PAP with the hypotheses discussed in the paper and/or its appendices.¹⁷ We find that study authors faithfully presented the results of all their pre-registered primary hypotheses in their paper or its appendices in 61% of cases. More than a third of studies had at least one pre-registered hypotheses that was never reported. Taking primary and secondary hypotheses together, the median paper in our sample neglected to report 25% of the hypotheses that had been pre-specified in the PAP. To be sure, constraints on journal space, the desire to package a study's results in a more readable form, and sometimes the requests of editors and/or reviewers, rather than unscrupulous research practice, likely accounts for many of the omitted hypotheses.¹⁸ But the frequency of the mismatch between what is pre-registered and what is presented undermines research credibility.

Apart from pre-registering hypotheses that are not reported in the paper, authors may also deviate from the PAP by reporting the results of hypotheses that were not pre-registered at all. We found that

¹⁷Researchers will sometimes register a PAP for an entire project, intending that different parts of the project will be discussed in different papers. In such a situation, a single paper may only report a subset of the pre-registered hypotheses in the PAP. In undertaking our coding, we looked for language indicating that the paper was reporting only a subset of the pre-registered hypotheses, with the others to be discussed in future work. We note, however, that, absent the careful pre-specification of which hypotheses will be presented in which papers, such situations create opportunities for selective presentation of results. It is impossible to know whether an author has cherry-picked the hypotheses to report in the “first” paper, never intending to (or not putting significant value on) dealing with the other hypotheses in follow-on paper—a within-study version of the “file drawer” problem discussed below.

¹⁸Consistent with this hypothesis, the median share of pre-specified hypotheses that were left out of the resulting paper was higher for published articles (25%) than for working papers (18%).

18% of the papers in our sample presented tests of novel hypotheses that were not pre-registered. Such deviations need not be a problem for research credibility if authors are transparent about the fact that the hypotheses were generated after the PAP was filed. But authors that presented results based on hypotheses that were not pre-registered failed to mention this in 82% of cases.

3.3 Other issues

3.3.1 Solving the “file drawer problem”

Beyond reducing the scope for fishing and post-hoc hypothesis adjustment, PAPs, which can be thought of as a particularly rigorous form of pre-registration, can help address the “file drawer problem.”¹⁹ The file drawer problem refers to the bias in the published literature on a given topic resulting from the tendency for authors not to submit, reviewers not to support, and/or journals not to publish results that fail to reach conventional thresholds for statistical significance (Rosenthal, 1979). Absent pre-registration, consumers of research only have access to the subset of studies that have been published or made publicly available as working papers. With pre-registration, consumers of research gain access to a record of studies that were initiated but that were never published. Among the subset of these studies that were actually completed—studies commonly fail for reasons that are uncorrelated with the outcomes that they might have generated—we have reason to suspect that at least some were not published because they generated null results (Gerber and Malhotra, 2008; Franco, Malhotra and Simonovits, 2014). Hence, study registries can put consumers of research in a position to make an educated inference about how likely it is that the findings reported in the studies in the public domain are representative of the underlying distribution of results that have been generated. If social science registries contain dozens of pre-registered studies on a given topic but the literature contains only a handful of publications, then researchers would be right to put less emphasis on the published findings.

Whether pre-registration aids in addressing this problem, however, depends on whether researchers actually consult registries to learn whether investigations on a given topic that have not been

¹⁹Filing a PAP is not, strictly speaking, necessary to address the file drawer problem. Pre-registration, which involves simply publicly declaring one’s intention to undertake a study that investigates a particular hypothesis, should be sufficient: this is why the AEA registry encourages pre-registration even in the absence of a formal PAP. However, pre-registering a PAP does this and more, so it makes sense to include the contribution to solving the file drawer problem in a discussion of the benefits of PAPs.

published may have nonetheless been undertaken. We asked researchers about this in our PAP users' survey and 38% reported that they had ever consulted a registry for this purpose. Like a tree falling in a forest with nobody nearby to hear it, PAPs—and pre-registration more generally—will do little to reduce the file drawer problem if researchers do not take advantage of the public record they provide about what has been done.

An argument is also made that registering a PAP can put researchers in a stronger position to publish the results of their studies, even if they generate null effects. Indeed, several journals in Political Science and Economics have begun experimenting with soliciting “registered reports” in which authors submit PAPs in lieu of finished research papers. Editors and reviewers then evaluate these submissions based on the importance of the questions that motivate the research and the quality of the proposed designs, with strong submissions accepted in principle on the condition that the data is collected and analyzed as proposed.²⁰ One such experiment in Political Science, a 2016 special issue of *Comparative Political Studies*, generated mixed reviews. Study authors generally liked the results-free submission and review process (Bush et al., 2018), but the journal editors concluded that the costs outweighed the potential benefits and indicated that they would not be moving toward a registered reports model for the journal writ large (Ansell and Samuels, 2016; Findley et al., 2016). Another experiment, at the *Journal of Development Economics*, appears to have been more positive, although the pilot's organizers identified a number of challenges, including the difficulty in judging submissions without seeing the final research findings, the up-front costs of composing guidelines for authors and reviewers, and the considerable effort required to guide authors and reviewers through a novel process that was demanding and “out of their comfort zone” (Foster et al., 2019).

3.3.2 Protecting against research partners with rival interests

Another leading rationale for PAPs is that they can help protect researchers against partners with rival interests. Donors and governments often fund the research activities for which PAPs are written. Like

²⁰Journals in Psychology and the medical sciences have long run their submission processes in this manner. In Political Science and Economics, journals that have embraced results-free submissions include the *Journal of Experimental Political Science*, *Research and Politics*, the *Journal of Development Economics*, *Experimental Economics*, and the *Japanese Journal of Political Science*. A longer list of journals have experimented with special issues that solicited registered reports, even if they have not (yet) adopted the approach as a regular submission option. A full list is available at <https://osf.io/8mpji/wiki/home/>.

pharmaceutical companies that underwrite research in the medical sciences, these actors may have interests in having the research generate particular conclusions. By providing an opportunity to discuss and agree in advance on both the results that will be reported and the specifications that will be employed to generate them, PAPs can help protect against pressure from such partners to favor particular empirical approaches or findings once the data analysis has begun and the results are becoming clear. Although most researchers in our PAP users' survey indicated that they had not yet used a PAP to protect themselves against a research partner with rival interests, several indicated that they had, and others indicated that they imagined that a PAP could be useful for this purpose.

4 Objections to PAPs

Our assessment also puts us in a position to address some of the objections to PAPs that have been raised in the literature.²¹

4.1 Too time consuming

Foremost among the objections to PAPs is that they are too time-consuming to prepare. Eighty-eight percent of researchers in our PAP users' survey reported devoting a week or more to writing the PAP for a typical project, with 32% reporting spending an average of 2-4 weeks and 26% reporting spending more than a month. It is perhaps not surprising, then, that 34% of researchers said that writing a PAP delayed their project's implementation.

However, while the PAP users we surveyed nearly all agreed that writing a PAP was costly in terms of time, 64% agreed with the statement that "it takes a considerable amount of time, but it is worth it."²² An overwhelming majority (8 in 10) said that drafting a PAP caused them to discover things about their project that led to refinements in their research protocols and/or data analysis plans. Sixty-five percent said that it put them in a position to receive useful feedback on their project design that they otherwise would not have received. And 52% said that they experienced downstream time savings from

²¹Useful discussions of objections to PAPs that go beyond the ones discussed here—and that echo several of the challenges articulated by respondents in our PAP users' survey—are provided in Humphreys, De la Sierra and Van der Windt (2013), Coffman and Niederle (2015), Olken (2015), van't Veer and Giner-Sorolla (2016), and Nosek et al. (2018).

²²Six percent said: "it doesn't take much time, so the cost is low." Thirty percent said: "it takes a considerable amount of time, and I am not certain of the value in the end."

having written a PAP, with 64% (so, 33% overall) indicating that these savings were equal to or greater than the time spent to draft the PAP in the first place. PAPs thus appear to shift the timing of work on research projects—from the back end, when the analysis is done, the results are written up, and most of the careful thinking about the project has traditionally taken place, to the front end. But for at least some researchers it is not clear that, on net, PAPs generate significantly more work.

4.2 Limit scope for new discoveries

Another major critique of PAPs is that they limit the scope for new discoveries that come from unconstrained explorations of one's data. One researcher in our PAP users' survey faulted PAPs for forcing her/him to "think about the lowest risk research I can run with the least potential for surprising findings." Another described PAPs as "stifling creativity" and worried that they "are being used as ammunition against careful researchers with integrity who genuinely want to learn from data." Yet these appear to be outlier views. Eleven percent of researchers said they thought that the existence of a PAP restricted their ability to fully explore and analyze their data "quite a bit," whereas 43% reported feeling not at all constrained and 46% reported feeling somewhat constrained. Similarly, 15% said they thought that having registered a PAP prevented them "quite a bit" from stumbling on unexpected, surprise results, whereas 37% reported that the existence of a PAP had not at all prevented them from generating unanticipated findings and 48% reported being somewhat prevented.

One response to the hand-tying generated by pre-specification is to pre-commit to an iterative approach in which the results from one part of the study inform the analysis of subsequent parts in carefully pre-specified ways.²³ Such an approach can be particularly attractive in situations where prior information about the subject of study is limited, making it difficult for the researcher to be confident that they are pre-specifying the full set of relevant or interesting hypotheses. While theoretically attractive, such iterative PAPs are tricky to implement in practice. For example, without a neutral gatekeeper, it can be challenging for researchers to document that iterations were truly pre-specified (Bidwell, Casey and Glennerster, Forthcoming).

²³Examples include Bidwell, Casey and Glennerster (Forthcoming) and Blair et al. (2019).

The more common approach is to freely undertake exploratory investigations that go beyond the PAP, so long as the results of such investigations are clearly labeled in the paper as coming from analyses that were not pre-specified, with an explanation provided for why they were added. Such an approach allows study authors to investigate new hypotheses that occur to them after they have immersed themselves in the data, while offering high transparency about the research process that generated results they report. But pursuing such a strategy faithfully, with findings clearly marked as pre-registered or exploratory and explanations provided for each deviation from the PAP, may come at the expense of the tight narrative that reviewers and journal editors are thought to favor.²⁴

4.3 Policing

By providing a record of the hypotheses a researcher intends to investigate and the analyses she commits herself to employ to test them, a PAP makes it possible for deviations from these pre-specified plans to be identified—but *only* if reviewers, editors, and/or consumers of the published work invest the considerable time and energy to track down the PAP and compare it (and, sometimes, its several iterations) side-by-side with the working paper or published article.²⁵ Laitin (2013) makes the point strongly: “registration without a community of scholars interested and incentivized to challenge findings is worthless.”

Is there any evidence that such policing actually happens? We asked the researchers in our PAP users’ survey whether, when they had submitted a paper with pre-registered analyses for review at a journal, reviewers had ever mentioned their PAP. Thirty-nine percent reported that reviewers had. This relatively low share may reflect the fact that only 28% of PAP users said that had ever included their PAP when they submitted their paper to a journal (however, another 50% said that this was because the paper mentioned the PAP and they assumed that reviewers could easily find it).²⁶ A similar share (35%)

²⁴In a separate analysis (Ofosu and Posner, 2019), we compare the publication outcomes of experimental NBER working papers that report the results of studies that do and do not include PAPs. We find that studies with PAPs were less likely to be published but, if published, significantly more likely to be published in a top-five journal.

²⁵As we have learned in our coding work for this project, this is challenging, time-consuming work—especially, as Bidwell, Casey and Glennerster (Forthcoming) emphasize, in the case of complex, iterative pre-specified designs. The unfortunate fact is that innovation to solve one problem (overly rigid designs that make it impossible for researchers to update their approach as they learn from their data) creates problems on another dimension (the difficulty of policing deviations from complicated, iterative PAPs that attempt to provide study authors with such flexibility).

²⁶Among Political Scientists, the ability of reviewers to examine a publicly posted PAP is complicated by the double-blind review process employed in most Political Science journals. To maintain the double-blind standard, authors submitting their PAP for review with their paper would have to submit an anonymized version (which, we note, is in tension with the desirability of having PAPs be public documents).

said that other researchers had invoked their PAP when discussing their paper outside of the formal review process, or that they themselves had consulted the PAP of a paper they were reviewing (34%). While PAPs may make policing possible, the norms and practices among reviewers, journal editors, and seminar participants seem not to have yet evolved to generate the strong policing equilibrium that would be required for PAPs to play the hand-tying role that is often imagined.²⁷

Policing involves not just effort on the part of reviewers, seminar participants, and other consumers of research, but also cooperation from the researcher producers themselves. The willingness of study authors to respond to queries about their work—especially when replication data, survey instruments, or code have not been made publicly available, or when PAPs remain private or gated—are essential companions to pre-registration.²⁸ It is therefore noteworthy that only 68% of the authors whose private/gated PAPs were randomly selected into our sample, and who we contacted to request that they share their PAPs with us, even replied to our email, and only 58% were willing to share their PAP.²⁹ Given the emerging norms in both Economics and Political Science about the importance of adopting open science practices (Christensen et al., 2019), registering a PAP is taken as a signal of type. However, such signals become uninformative if researchers who embrace some open science practices (such as pre-registration) are unwilling to do the (admittedly hard) work of following through when other researchers request additional information.

There is a sentiment, in some parts of the PAP users' community, that PAPs offer the worst of both worlds, in the sense that they tie researchers' hands, preventing them from investigating interesting threads that emerge in their analysis, while still leaving them open to demands from reviewers for endless robustness tests. As one PAP user wrote: "I've gotten an absurd number of requests for sensitivity analyses for strictly pre-specified empirical work. The existing norm appears to keep me from looking for unexpected results while providing no protection from readers or reviewers who want to dig through the data trying to kill off empirical results they don't agree with." Although 46% of PAP users report having invoked their PAP to respond to the suggestions of reviewers or workshop participants regarding additional analyses to run, one lamented that pointing to the PAP does little good, since "referees and

²⁷ An insightful discussion of policing norms in Political Science, and the challenges of changing them, is provided in Laitin and Reich (2017).

²⁸ For recent evidence on the adoption of such open social science practices, see Christensen et al. (2019).

²⁹ Further details of our efforts to contact the authors of private/gated PAPs are provided in Appendix A.

editors ignore them/refuse to be bound by them.” Again, the absence of common norms about what PAPs obligate both producers and consumers of research to do leaves pre-registration well short of achieving its goals.

4.4 Getting scooped

We also asked researchers in our PAP users’ survey whether, in contemplating registering a PAP, they had any concern that others might scoop their ideas. Forty-six percent reported having no concern whatsoever, with another 39% saying they had slight concern. Eleven percent said that they were unconcerned because the PAP was gated or private. If we assume that preventing others from stealing their ideas was the only reason why these researchers gated their PAPs, then the total share of researchers expressing significant concern about getting scooped is below 15%.

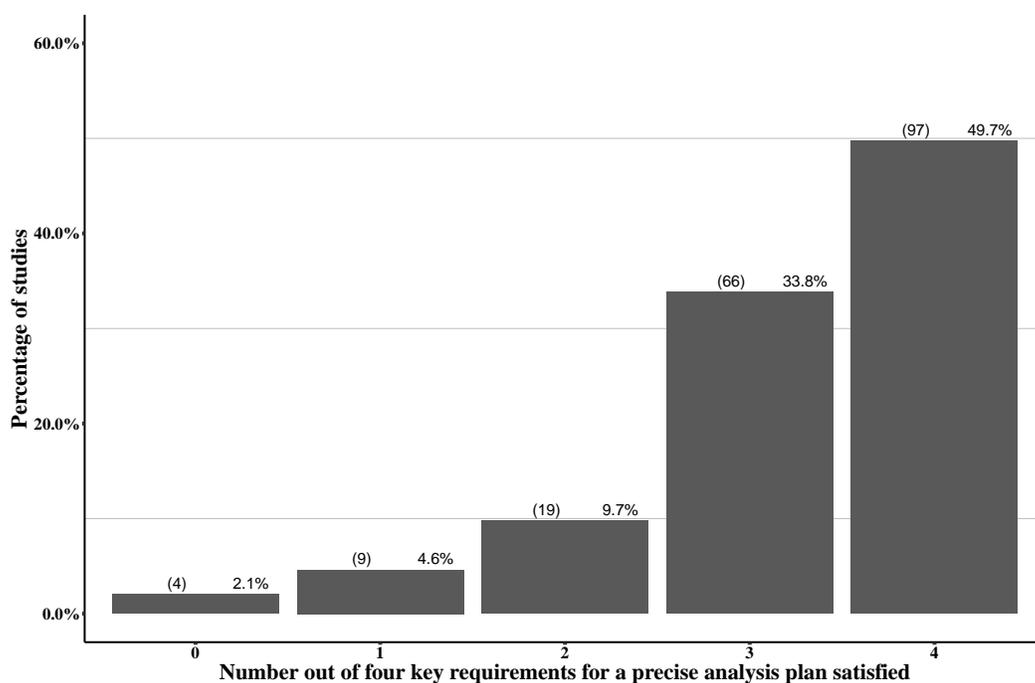
5 The balance sheet

Our stocktaking suggests that PAPs are often not written or used in a way that allows them to do everything that their proponents had hoped. Many PAP authors are insufficiently clear about the hypotheses they are testing to prevent them from moving the goal posts once they have seen the patterns in their data. The details of the analyses that PAPs pre-specify—how outcome and causal variables are to be operationalized; which controls will be included; what the statistical model will look like; how imbalances, outliers, and attrition will be dealt with—are not always adequate to reduce researcher degrees of freedom in a meaningful way. In addition, papers that result from pre-registered analyses do not always follow what was pre-registered. Some papers introduce entirely novel hypotheses; others present only a subset of the hypotheses that were pre-registered.

But showing that not all PAPs adequately addresses all of the problems they were designed to solve does not imply that the growing use of PAPs in Economics and Political Science is not generating more credible research. Figure 3 reports the share of PAPs in our sample that meet what we take to be the four key requirements for a complete, well-specified PAP: specifying a clear hypothesis; specifying the primary dependent and independent/treatment variable(s) sufficiently clearly so as to prevent post-hoc adjustments; and spelling out the precise statistical model to be tested. Just under half of the 195

PAPs we analyzed were judged to meet all four of these criteria, and another third were judged to satisfy three of the four. Although this is hardly a perfect record, it seems reasonable to view our stocktaking as suggesting that the glass is half full rather than half empty—especially when one recognizes that the counterfactual condition would be a world with no PAPs at all. Even if the scope for fishing and HARKing is not foreclosed by *every* PAP, such opportunities are limited to at least some degree in most. Even imperfect PAPs increase the credibility of (at least some aspects of) the research studies for which they are written.

Figure 3: Number and share of PAPs satisfying the four key requirements of a complete PAP



Note: Figure 3 shows the number and share of PAPs that satisfy the four key requirements of a complete PAP: 1) specifying a clear hypothesis; 2) specifying the primary dependent variable(s) sufficiently clearly so as to prevent post-hoc adjustments; 3) specifying the treatment or main explanatory variable sufficiently clearly so as to prevent post-hoc adjustments; and 4) spelling out the precise statistical model to be tested including functional forms and estimator.

As PAP skeptics point out, however, these improvements to research credibility come at a price. Writing a PAP occupies weeks of valuable research time, and adhering faithfully to what was pre-specified may limit creativity, reduce the scope for new discoveries, and result in research papers that more closely resemble lab reports than the sorts of exciting write-ups that reviewers and journal editors are thought to favor—or so critics claim. While the time costs of writing a PAP are real, the alleged constraints on creativity and exploration can be loosened by simply labeling one’s investigations as ex-

ploratory or confirmatory. The concern that adherence to PAPs results in boring, rote papers can be addressed by a combination of better writing and a re-weighting of priorities toward scientific rigor over compelling narrative. Equally important, the data from our PAP users' survey suggest that PAPs do not restrict researchers' investigations or gum up the research process nearly as much as their detractors claim. On balance, researchers report that the benefits of writing a PAP outweigh the costs. For every researcher who describes PAPs as "an additional hassle" or "toxic to the process of doing research" there is another who says that writing a PAP "makes me more thoughtful and deliberate" or "causes me to really think through design and analysis decisions that, honestly, were often done on the back end."

6 The importance of complementary norms and institutions

Our stocktaking exercise was motivated by a desire to assess the extent to which PAPs, as they are actually written and used, generate meaningful improvements in research credibility. Our strategy for answering this question is to scrutinize whether PAPs are sufficiently clear, precise and comprehensive to prevent fishing and HARKing. However, as we have hinted at several points in the discussion, the impact of PAPs on research credibility may depend less on the contents of the PAPs themselves than on the presence of a set of complementary norms and institutions that provide guidance on how PAPs should be used in the research process and that create incentives for researchers to invest the time and energy to produce and police them.

A first, crucial set of norms speak to what, exactly, a complete PAP should contain. Although several publications provide recommendations for what authors should include in their PAPs (McKenzie, 2012; Glennerster and Takavarasha, 2013; Christensen, Freese and Miguel, 2019; EGAP, N.d.), there are no universally agreed upon rules for what a comprehensive PAP should look like. This lack of common standards may account for some of the deficiencies we identified in our coding exercise. Recent innovations such as *Declare Design* (Blair et al., 2019), which provides software that allows researchers to formally describe (and troubleshoot) the details of their proposed analyses, are powerful tools that may help remedy this problem. But they are new and have yet to become widely adopted.

Alongside clarifying the standards for what PAPs should include, a major issue is the development of norms about how PAPs should be used by the research community. Laitin articulates the problem

well when he writes that “all the pre-analysis plans...we produce do not serve science if no one has a career interest in deciphering them or confirming the results that followed from them. We have increased the supply of transparency but have given insufficient attention to generating a demand for it” (Laitin, 2018). Scrutinizing PAPs and comparing their contents to what is reported in the resulting publications and working papers is tedious work, but it is necessary for the benefits of PAPs to be fully realized. Creating disciplinary incentives for such policing is a critical challenge.³⁰

The most logical venue for such scrutiny is the journal review process. But here too the disciplines lack clear norms. Should researchers be required to submit their PAPs along with their papers? Should reviewers be expected to go through the PAP and certify that the analyses presented in the paper match those that were pre-specified? What should reviewers or editors do if, as we found in many of the PAPs we analyzed, the pre-specification of hypotheses and/or procedures is too unclear or incomplete to remove the scope for fishing or HARKing? Or what if, as in Bidwell, Casey and Glennerster (Forthcoming), the PAP was periodically updated during the course of the project, making the task of identifying deviations maddeningly complex? Is it fair for reviewers to ask authors of papers with PAPs to present multiple robustness tests as a condition for acceptance? These and other questions will need to be debated and answered in order to better harness the formal review process to more fully leverage the transparency that PAPs offer.

While much of the enhanced research credibility generated through pre-registration accrues to the pre-registered studies themselves, some of the benefits depend on the adoption of pre-registration by the discipline as a whole. For example, the role that pre-registration plays in addressing the file drawer problem depends on researchers becoming habituated to consulting study registries for clues about the true distribution of findings in a given area. But such consultations will only be informative if the registries are complete and comprehensive. Bolstering the usefulness of registries as repositories of what has been done will thus require bolstering norms about the necessity of pre-registration.

Convincing researchers who do not currently pre-register their projects to begin doing so (much less convincing them to begin composing and filing formal PAPs) is no easy task, however. The re-

³⁰An increasingly common assignment in many graduate seminars in Political Science is to have students replicate the analyses presented in published studies. Similar assignments could be devised in which students are tasked with comparing published articles or working papers with the PAPs that were registered at the time the projects were initiated.

cently completed State of Social Science Survey (Christensen et al., 2019) finds that while the majority of researchers in Economics and Political Science are aware of and support the norm of pre-registration, behavior in adopting the practice is significantly lagging. One key obstacle is the hesitancy, revealed both in our data and in the evidence summarized in Christensen et al. (2019), of authors of observational studies to file PAPs. In part, this reluctance stems from the fact that observational data is often available to researchers prior to initiating their projects, which makes it difficult or impossible for them to demonstrate that they composed their PAPs prior to looking at the data. Institutions for embargoing data or involving independent third party actors, along the lines suggested in Bidwell, Casey and Glennerster (Forthcoming) and Fafchamps and Labonne (2017), might increase the perceived value of PAPs among researchers using historical or administrative data and lead to their adoption by a broader set of scholars.

Another strategy for increasing the value of PAPs is to invest in institutions and norms that allow the researchers who write them to receive helpful feedback on their study designs. Groups such as EGAP and the Working Group in African Political Economy regularly reserve slots at their meetings for the discussion of PAPs, alongside completed working papers. Such discussions provide opportunities for receiving comments and suggestions at a key, early stage in a project's development. The promotion of norms that make seminar presentations of PAPs equally acceptable as presentations of finished papers would lead to the proliferation of such opportunities. This, in turn, would provide tangible benefits to PAP authors that offset the cost of composing the PAP, and thus make them more willing to make such investments in the first place.

Although their use has risen steeply in recent years, PAPs are still in their infancy. Our analysis, which covers PAPs registered between 2011 and 2016, captures the early years of PAP usage when many authors were filing their first PAPs and norms about what authors should include in their PAPs and how they should deal with deviations from what they pre-registered were still emerging. Although nearly half of our sample comes from 2016, the final year in this period, we think it is likely that PAPs registered today may be, on average, more precise and complete than those whose contents we analyzed—and that the contribution of PAPs to research credibility today may be even greater than what is suggested by our stocktaking. The further development of norms and complementary institutions that can both augment

the power of PAs to improve research credibility and create incentives for researchers to invest the time and energy to produce and police them will only reinforce these positive trends.

References

- Ansell, Ben and David Samuels. 2016. "Journal editors and "results-free" research: A cautionary note." *Comparative Political Studies* 49(13):1809–1815.
- Bidwell, Kelly, Katherine Casey and Rachel Glennerster. Forthcoming. "Debates: Voting and expenditure responses to political communication." *Journal of Political Economy* .
- Blair, Graeme, Fotini Christia, Cyrus Samii and Jeremy Weinstein. 2019. "Meta-analysis pre-analysis plan: Community policing Metaketa project." Pre-analysis plan.
- Brodeur, Abel, Mathias Lé, Marc Sangnier and Yanos Zylberberg. 2016. "Star wars: The empirics strike back." *American Economic Journal: Applied Economics* 8(1):1–32.
- Bush, Sarah, Aaron Erlich, Lauren Prather and Yael Zeira. 2018. "Lessons from results-blind review." *The Political Economist* (Spring).
- Casey, Katherine, Rachel Glennerster and Edward Miguel. 2012. "Reshaping institutions: Evidence on aid impacts using a preanalysis plan." *Quarterly Journal of Economics* 127(4):1755–1812.
- Christensen, Garret, Jeremy Freese and Edward Miguel. 2019. *Transparent and reproducible social science research: How to do open science*. University of California Press.
- Christensen, Garret, Zenan Want, Elizabeth L. Paluck, Nicholas Swanson, David J. Birke, Edward Miguel and Rebecca Littman. 2019. "Open science practices are on the rise: The state of social science (3S) survey." Working paper.
- Claesen, Aline, Sara Gomes, Francis Tuerlinckx and Wolf Vanpaemel. 2019. "Preregistration: Comparing dream to reality." Working paper.
- Coffman, Lucas C. and Muriel Niederle. 2015. "Pre-analysis plans have limited upside, especially where replications are feasible." *Journal of Economic Perspectives* 29(3):81–98.
- EGAP. N.d. "10 things to know about pre-analysis plans."
URL: <http://egap.org/methods-guides/10-things-pre-analysis-plans>
- Fafchamps, Marcel and Julien Labonne. 2017. "Using split samples to improve inference on causal effects." *Political Analysis* 25(4):465–482.
- Fang, Albert, Grant Gordon and Macartan Humphreys. 2015. "Does registration reduce publication bias? Evidence from medical sciences."
- Findley, Michael G., Nathan M. Jensen, Edmund J. Malesky and Thomas B. Pepinsky. 2016. "Can results-free review reduce publication bias? The results and implications of a pilot study." *Comparative Political Studies* 49(13):1667–1703.
- Foster, Andrew, Dean Karlan, Edward Miguel and Aleksandar Bogdanoski. 2019. "Pre-results review at the Journal of Development Economics: Lessons learned so far." *Development Impact* .
URL: <https://blogs.worldbank.org/impactevaluations/pre-results-review-journal-development-economics-lessons-learned-so-far>

- Franco, Annie, Neil Malhotra and Gabor Simonovits. 2014. "Publication bias in the social sciences: Unlocking the file drawer." *Science* 345(6203):1502–1505.
- Gerber, Alan and Neil Malhotra. 2008. "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.
- Glennerster, Rachel and Kudzai Takavarasha. 2013. *Running randomized evaluations: A practical guide*. Princeton University Press.
- 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):1–20.
- Ioannidis, John PA. 2005. "Why most published research findings are false." *PLoS medicine* 2(8):e124.
- Laitin, David D. 2013. "Fisheries management." *Political Analysis* 21(1):42–47.
- Laitin, David D. 2018. "Is there transparency if no one is looking?" *The Political Economist* (Spring).
- Laitin, David D. and Rob Reich. 2017. "Trust, transparency, and replication in political science." *PS: Political Science & Politics* 50(1):172–175.
- Lenz, Gabriel and Alexander Sahn. 2019. "Achieving statistical significance with covariates and without transparency." Working paper.
- Lin, Winston and Donald P. Green. 2016. "Standard operating procedures: A safety net for pre-analysis plans." *PS: Political Science & Politics* 49(3):495–500.
- McKenzie, David. 2012. "A pre-analysis plan checklist." *Development Impact* .
URL: <https://blogs.worldbank.org/impacetevaluations/a-pre-analysis-plan-checklist>
- Nosek, Brian A., Charles R. Ebersole, Alexander C. DeHaven and David T. Mellor. 2018. "The preregistration revolution." *Proceedings of the National Academy of Sciences* 115(11):2600–2606.
- Ofofu, George K. and Daniel N. Posner. 2019. "Do pre-analysis plans hamper publication?" Working paper.
- Olken, Benjamin A. 2015. "Promises and perils of pre-analysis plans." *Journal of Economic Perspectives* 29(3):61–80.
- Open Science Collaboration. 2015. "Estimating the reproducibility of psychological science." *Science* 349(6251):aac4716.
- Rosenthal, Robert. 1979. "The 'file drawer problem' and the tolerance for null results." *Psychological Bulletin* 86(3):638–641.
- Simmons, Joseph P., Leif D. Nelson and Uri Simonsohn. 2011. "False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant." *Psychological Science* 22(11):1359–1366.
- Simonsohn, Uri, Leif D. Nelson and Joseph P. Simmons. 2014. "P-curve: a key to the file-drawer." *Journal of Experimental Psychology: General* 143(2):534.

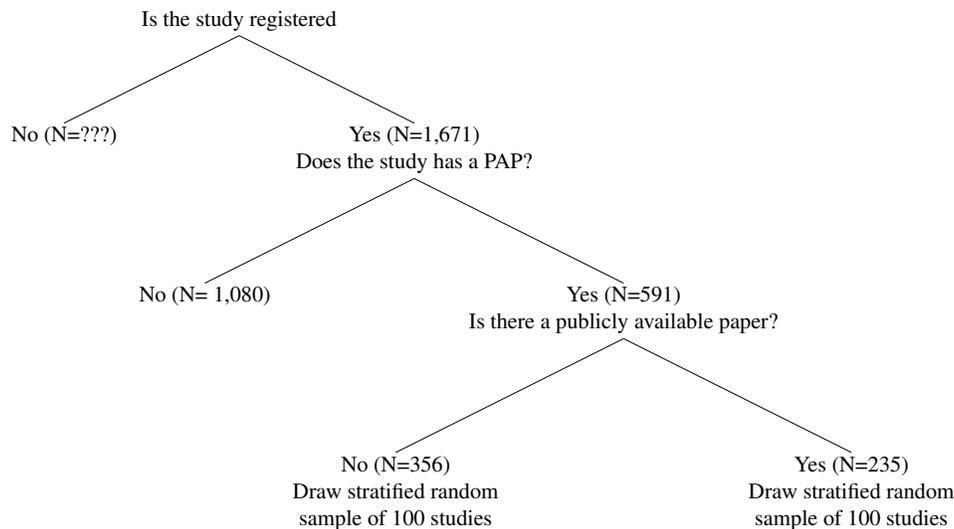
- van't Veer, Anna Elisabeth and Roger Giner-Sorolla. 2016. "Pre-registration in social psychology—A discussion and suggested template." *Journal of Experimental Social Psychology* 67:2–12.
- Wicherts, Jelte M., Coosje L.S. Veldkamp, Hilde E.M. Augusteijn, Marjan Bakker, Robbie Van Aert and Marcel A.L.M. Van Assen. 2016. "Degrees of freedom in planning, running, analyzing, and reporting psychological studies: A checklist to avoid p-hacking." *Frontiers in Psychology* 7:1832.

Appendix

A Sampling procedures

As shown in Figure A.1, a total of 1,671 studies were registered on the AEA (1,235) and EGAP (436) registries during the period we studied.³¹ Of these, 591 had PAPs (269 of the AEA-registered studies and 322 of the EGAP-registered studies).³² We then identified whether each study had resulted in a publicly available paper. To do this, we conducted web searches of each study author’s web page, as well as key-word searches based on the project’s title and abstract. Of the 591 studies with PAPs, we found 235 that had resulted in a publicly available paper by the time of our search.

Figure A.1: Sampling Procedure



Notes: Stratification is by year, initially gated status, and study registry (AEA and EGAP).

We then drew a random sample of 100 of these studies, alongside a random sample of 100 studies that had not yet resulted in a publicly available paper. In drawing these samples, we stratified by three

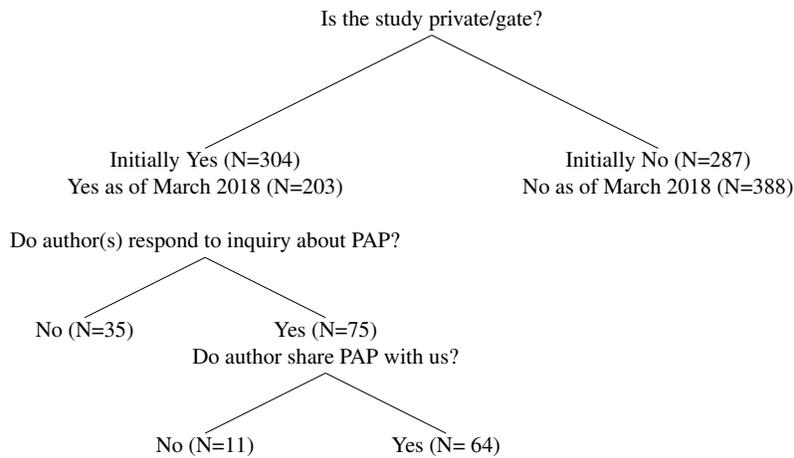
³¹In addition to the 1,671 registered studies there are, of course, also an unknown number of studies that are not registered at all, and that therefore fall outside the scope of our analysis. We underscore that the absence of these studies in our sample prevents us from making causal claims about the effects of PAPs.

³²The reason for the smaller share of studies with PAPs on the AEA registry is because many of the projects registered there were included to provide a record of the fact that they had been undertaken rather than to pre-register a set of research procedures or hypothesis. To avoid including PAPs written by graduate students as part of a class exercise, we limit our analysis to PAPs written by researchers holding an academic appointment or, if not at an academic institution (i.e., at the World Bank), then holding a PhD.

criteria: the year the study was registered (2011, 2012, 2013, 2014, 2015, 2016), whether the study was registered on the AEA or EGAP registries, and whether or not the PAP was initially gated/private.

The fact that not all PAPs are made public at the time a study is registered created a challenge for our coding exercise. The AEA registry affords study authors the opportunity to keep their PAPs private and the EGAP registry, while strongly discouraging researchers from doing so, permits study authors to gate their PAP for a period of time.³³ As shown in Figure A.2, of the 591 studies with PAPs, 304 were initially private/gate, although 101 of those had become public/ungated by the time we drew our sample in March 2018.

Figure A.2: Dealing with private/gated PAPs



Notes: We contacted the authors of 123 studies and can confirm 110 read our email.

To reach our goal of coding 100 PAPs from projects that had resulted in publicly available papers and another set of 100 that had not, and anticipating that some authors of private/gated PAPs might be unresponsive to our request that they share their PAPs with us, we oversampled 30 percent of private/gated plans in each category. The oversample contained 265 PAPs (132 with papers and 133 without), of which 123 were still private/gated as of March, 2018. We contacted the authors of these private/gated PAPs via email to ask them to confidentially share their PAPs with us.³⁴ Of the 120 authors who we can confirm

³³Our suspicion is that the kinds of authors who keep their PAPs private or who gate them for an initial period may be different from those who make them public from the start. Hence our decision to stratify our sample by this criterion.

³⁴A copy of the email, which was sent on April 10, 2018, is provided in Appendix A.1. We sent a reminder email nine days later to authors who had not yet responded.

received and read our email, we received replies from 75 (68 percent), of which 64 (58 percent) were willing to share their PAP.³⁵ Summary statistics are provided in Appendix A.2.

³⁵Of the 11 study authors who replied to our query but did not share their PAP, five reported that their study was still ongoing and one reported that the study was cancelled. Others reported that there was no PAP (even though the registration suggested there was one) or insisted that they had made the PAP public, even though we were not able to access it.

A.1 Email to authors of private/gated PAPs

Dear Author (Dr. XXX),

As part of an ongoing research project on pre-registration and research transparency in the social sciences, we are collecting pre-analysis plans of studies that have been pre-registered on the AEA (socialscienceregistry.org) and EGAP (egap.org) registries. We see that your study, “[study title]” (registration # “[registration number]”) is currently private (or gated), which means that we are unable to view your pre-analysis plan.

We are writing to ask if you would be willing to share your pre-analysis plan with us. We pledge not share it publicly and to use it only for our analysis.

We would very much appreciate receiving your response within the next two weeks. You can send your pre-analysis plan to us as an attachment to PAPstudy2018@gmail.com.

Thank for your help in advance.

Best regards,

George Ofosu, Washington University, St. Louis

Daniel Posner, UCLA

A.2 Summary statistics

Table A.1: Summary statistics of population and sample of PAPs

Features	Population N=591	Sample (coded) N=195
With publication	0.396	0.477
Private (gated)	0.514	0.503
Registry		
American Economic Association	0.455	0.467
Evidence in Governance and Politics	0.545	0.533
Year		
2011	0.003	0.005
2012	0.012	0.010
2013	0.071	0.077
2014	0.149	0.169
2015	0.299	0.287
2016	0.465	0.451

B Coding rubric for PAPs with papers

Coding Instrument for PAPs with Papers

Coding Instrument

Study identification

1. Enter PAP code:

2. What kind of study does the PAP describe?

Mark only one oval.

- Field experiment
- Lab experiment
- Lab-in-field experiment
- Natural experiment
- Observational study
- Other: _____

Hypotheses

3. Does the PAP specify a clear hypothesis (hypotheses) to be tested? A "clear hypothesis" is one that describes a relationship between a clearly identified independent and dependent variable in which the direction of the effect is specified.

Note: "providing citizens with information about service delivery will increase political participation" would be coded "yes"

Mark only one oval.

- Yes
- No

4. Does the PAP specify more than one hypothesis to be tested?

Mark only one oval.

- Yes
- No

5. If yes, how many in total?

6. If the PAP specifies more than one hypothesis, are some of the hypotheses designated as primary and some as secondary/exploratory?

Mark only one oval.

- Yes
- No
- Not applicable (the PAP specifies just one hypothesis)

7. If the PAP distinguishes between primary and secondary/exploratory hypotheses, is this distinction maintained in the paper? (That is, are the primary hypotheses still presented as primary and the secondary hypotheses as secondary, or are some hypotheses that were pre-specified as secondary presented in the paper as primary, and vice versa?)

Mark only one oval.

- Yes
- No
- Not applicable (the PAP specifies just one hypothesis)

8. If the PAP specifies more than one hypothesis, how many are classified as primary?

9. How many of the primary hypotheses that were pre-registered in the PAP are presented in the main body of the paper?

10. How many of the primary hypotheses that were pre-registered in the PAP are presented in the appendix to the paper?

11. How many of the primary hypotheses that were pre-registered in the PAP and are presented in the main body of the paper are supported by the research findings?

12. How many of the primary hypotheses that were pre-registered in the PAP and are presented in the appendix are supported by the research findings?

13. How many of the secondary/exploratory hypotheses that were pre-registered in the PAP and are presented in the main body of the paper?

14. How many of the secondary/exploratory hypotheses that were pre-registered in the PAP and are presented in the appendix of the paper?

15. Are there any results highlighted in the paper based on entirely new hypotheses that were not mentioned in the PAP?

Mark only one oval.

- Yes
- No

16. If yes, do the researchers point out that these analyses were not pre-specified?

Mark only one oval.

- Yes
- No
- Not applicable (no results highlighted in the paper that were not based on pre-specified hypotheses)

Clarity of variable definitions

17. Is (are) the primary dependent variable(s) in the study operationalized sufficiently clearly so as to prevent post-hoc adjustments?

Note: Following Olken, "a good rule of thumb is that if you gave the PAP to two different programmers, and asked each to prepare the data for the primary dependent variable(s), they should be both able to do so without asking any questions, and they should both be able to get the same answer." If multiple primary DVs are specified, code based on whether, in general, they are operationalized clearly. Where some dependent variables are designated as primary, focus only on those.

Mark only one oval.

- Yes
- No

18. Are there any results highlighted in the paper that are based on changes in the operationalization of the primary dependent variable(s) that were pre-specified in the PAP?

Mark only one oval.

- Yes, significant changes that might plausibly affect estimates
- Yes, minor changes that probably do not affect estimates
- No changes in variable operationalization

19. If there are changes in the operationalization of the primary dependent variable(s), do the authors note that they are different from what was pre-specified?

Mark only one oval.

- Yes
- No
- Not applicable (no changes in variable operationalization)

20. Briefly describe these deviations from the PAP with respect to the operationalization of the primary dependent variable(s)

21. If secondary/exploratory dependent variables are specified, are they operationalized sufficiently clearly so as to prevent post-hoc adjustments?

Note: Following Olken, "a good rule of thumb is that if you gave the PAP to two different programmers, and asked each to prepare the data for the secondary/exploratory dependent variable(s), they should be both able to do so without asking any questions, and they should both be able to get the same answer."

Mark only one oval.

- Yes
- No
- Not Applicable (secondary/exploratory DVs are not specified in the PAP)

22. Is the treatment or main explanatory variable specified sufficiently clear so as to prevent post-hoc adjustments?

Following Olken, "a good rule of thumb is that if you gave the PAP to two different programmers, and asked each to prepare the data for the treatment/main explanatory variable, they should be both able to do so without asking any questions, and they should both be able to get the same answer."

Mark only one oval.

- Yes
- No

23. Are there any results highlighted in the paper that are based on changes in the operationalization of the treatment or main explanatory variable that were pre-specified in the PAP?

Mark only one oval.

- Yes, significant changes that might plausibly affect estimates
- Yes, minor changes that probably do not affect estimates
- No changes in variable operationalization

24. If there are changes in the operationalization of the treatment or main explanatory variable, do the authors note that they are different from what was pre-specified?

Mark only one oval.

- Yes
- No
- Not applicable (no changes in variable operationalization)

25. Briefly describe these deviations from the PAP with respect to the operationalization of the treatment or main explanatory variable

26. How many control or other independent variables are pre-specified?

27. **How many of these control or independent variables are specified sufficiently clear so as to prevent post-hoc adjustments?**
-

Sampling and power analysis

28. **Does the PAP specify the population of interest and unit of analysis?**

Mark only one oval.

Yes

No

29. **Does the PAP specify the sampling frame?**

Mark only one oval.

Yes

No

30. **Does the PAP specify the sampling strategy?**

Mark only one oval.

Yes

No

31. **Does the PAP specify the conditions under which participants/units may be excluded from the sample?**

Mark only one oval.

Yes

No

32. **Does the PAP use a power analysis to justify the sample size required for the study?**

Mark only one oval.

Yes

No

Data collection

33. **Is the treatment/main independent variable under the control of (randomized by) the researchers?**

Mark only one oval.

Yes

No

34. **If yes, does the PAP specify how participants/units will be assigned to treatment (i.e., the randomization procedure)?**

Mark only one oval.

Yes

No

Not Applicable (treatment is not controlled by researcher)

35. Does the PAP specify a manipulation check (i.e. variables to report in a balance table)?

Mark only one oval.

- Yes
 No

36. If yes, does the PAP specify what to do in the event of an imbalance?

Mark only one oval.

- Yes
 No
 Not Applicable (PAP does not specify a manipulation check)

Inclusion and exclusion rules

37. Does the PAP specify rules on how to handle missing values and attrition?

Mark only one oval.

- Yes
 No

38. Does the paper follow the pre-specified protocols with respect to missing values and attrition?

Mark only one oval.

- Yes
 No
 Not applicable

39. Does the PAP specify rules for dealing with outliers?

Mark only one oval.

- Yes
 No

40. Does the paper follow the pre-specified protocols with respect to dealing with outliers?

Mark only one oval.

- Yes
 No
 Not applicable

41. Does the PAP specify rules for dealing with noncompliance?

Mark only one oval.

- Yes
 No
 Not applicable (noncompliance is not an issue in the study)

42. Does the paper follow the pre-specified protocols with respect to noncompliance?

Mark only one oval.

- Yes
 No
 Not applicable

Statistical model specification

43. Does the PAP spell out the precise statistical model to be tested including functional forms and estimator (ordinary least squares, probit, logit, Poisson, instrumental variables, and so on)?

Mark only one oval.

- Yes
 No

44. Are the results presented in the paper based on the exact model specifications that were pre-specified?

Mark only one oval.

- Yes
 No, minor deviations from pre-registered specifications that probably do not affect estimates
 No, minor deviations from pre-registered specifications that might plausibly affect estimates
 No, major deviations from pre-registered specifications

45. If there are deviations from the pre-registered model specifications, do the authors point out that the specifications they use are different from what was pre-registered?

Mark only one oval.

- Yes
 No

46. Briefly describe these deviations from the PAP with respect to model specifications.

47. Does the PAP state how standard errors will be treated (robust, clustering, bootstrapping, or other techniques)?

Mark only one oval.

- Yes
 No

48. Does the PAP specify a multiple testing adjustment?

Mark only one oval.

- Yes
- No
- Not Applicable (too few DVs for a multiple testing adjustment to be necessary)

49. Does the PAP commit the researchers to presenting a simple difference-in-means test?

Mark only one oval.

- Yes
- No

50. If yes, do the authors specify whether the test will be one-tailed or two-tailed?

Mark only one oval.

- Yes
- No
- Not applicable (no difference-in-means test is specified)

51. Does the PAP specify whether/how covariates (including fixed effects) will be included in the regression model?

Mark only one oval.

- Yes
- No

Other features

52. When was the PAP initially registered?

Mark only one oval.

- Prior to data collection
- After data collection but before authors had access to/analyzed data
- After analysis of data had begun
- Other: _____

53. Was the PAP ever gated/private?

Mark only one oval.

- Yes
- No

54. Was the PAP ever updated/refiled?

Mark only one oval.

- Yes
- No

55. If the PAP was updated/refiled, how many updates were registered?*Mark only one oval.*

- One
- Two
- Three or more
- Not applicable (PAP never updated)

56. If the PAP was updated/refiled, were the updates clearly identified as changes from the original PAP?*Mark only one oval.*

- Yes
- No
- Not Applicable (PAP never updated)

57. Does the PAP specify that the study has received IRB approval (or that such approval is pending)?*Mark only one oval.*

- Yes
- No
- Not Applicable (IRB approval is not necessary for a study of this type)

58. Do authors pre-specify what will guide their decisions when issues arise that were not anticipated in the PAP (i.e., the SOPs described in Lin and Green 2015)?*Mark only one oval.*

- Yes
- No

59. How many pages long is the PAP (in single-spaced pages)?*Note: Divide double-spaced documents by two.*

60. Do the authors provide a link to their replication data in the paper?*Note: Divide double-spaced documents by two.**Mark only one oval.*

- Yes, a link is provided and it works
- Yes, a link is provided but it does not work
- Authors indicate that the replication data is "available on request"
- No mention of replication data

61. Do the authors provide a link to their do files in the paper?*Note: Divide double-spaced documents by two.**Mark only one oval.*

- Yes, a link is provided and it works
- Yes, a link is provided but it does not work
- Authors indicate that the do files are "available on request"
- No mention of the do files

62. Do the authors make available their data collection instruments and/or lab protocols?

Note: Divide double-spaced documents by two.

Mark only one oval.

- Yes, they are provided in an appendix (including an online appendix)
 - Authors indicate that these materials are "available on request"
 - No mention is made of these materials
-

Powered by



C PAP users' survey

Pre-analysis plans survey

Survey Flow

Standard: Introduction (1 Question)

Block: Pre-registration practice (13 Questions)

Standard: PAPs in the paper writing process (4 Questions)

Standard: PAPs in the review process (6 Questions)

Standard: Open-ended questions (5 Questions)

Page Break

Start of Block: Introduction

We thank you for your participation. The survey will take only 5-10 minutes to complete. Please, feel free to skip any questions that you prefer not to answer.

End of Block: Introduction

Start of Block: Pre-registration practice

Q2.1 Have you ever registered a pre-analysis plan (PAP) for a research project?

- Yes (1)
- No (2)

Display This Question:

If Have you ever registered a pre-analysis plan (PAP) for a research project? = Yes

Q32 How many PAPs have you registered?

- 1-2 (1)
- 3-4 (2)
- 5 or more (3)

Display This Question:

If Have you ever registered a pre-analysis plan (PAP) for a research project? = Yes

Q2.2 Have you ever kept one or more of your PAPs gated/private for a period of time?

- Yes (1)
 - No (2)
-

Display This Question:

If Have you ever registered a pre-analysis plan (PAP) for a research project? = Yes

Q2.3 Do you have any ongoing research projects for which you did not register a PAP?

- Yes (1)
- No (2)

Display This Question:

If Do you have any ongoing research projects for which you did not register a PAP? = Yes

Q2.4 How many?

- 1-2 (1)
- 3-4 (2)
- 5 or more (3)

Q2.5 In contemplating registering a PAP, did you have any concern that others might scoop your ideas?

- No concern whatsoever (1)
- Slight concern (2)
- Significant concern (3)
- The PAP was gated (4)

Display This Question:

If Have you ever registered a pre-analysis plan (PAP) for a research project? = Yes

Q2.6 How long does it take you to draft a PAP for a typical project?

- A few days (1)
- About a week (2)
- 2-4 weeks (3)
- More than a month (4)

Display This Question:

If Have you ever registered a pre-analysis plan (PAP) for a research project? = Yes

Q2.7 Did writing the PAP cause you to discover anything about your project that led to refinements in your research protocols and/or data analysis plans

- Yes (1)
- No (2)

Display This Question:

If Have you ever registered a pre-analysis plan (PAP) for a research project? = Yes

Q2.8 Did writing the PAP put you in a position to receive useful feedback on your project design that you might not otherwise have received?

- Yes (1)
- No (2)

Display This Question:

If Have you ever registered a pre-analysis plan (PAP) for a research project? = Yes

Q2.9 Did writing the PAP delay the implementation of your project?

- Yes (1)
- No (2)

Display This Question:

If Have you ever registered a pre-analysis plan (PAP) for a research project? = Yes

Q2.10 Did you experience any downstream time savings from having written a PAP?

- Yes (1)
- No (2)

Display This Question:

If Did you experience any downstream time savings from having written a PAP? = Yes

Q2.11 Were these downstream time savings greater than the time spent to draft the PAP in the first place?

- Yes (1)
- No (2)
- About equal (3)

Display This Question:

If Have you ever registered a pre-analysis plan (PAP) for a research project? = Yes

Q2.12 Which of the following best characterizes your feelings about the time it takes to write and register a PAP?

- It doesn't take much time, so the cost is low. (1)
- It takes a considerable amount of time, but it is worth it. (2)
- It takes a considerable amount of time, and I am not certain of the value in the end. (3)

End of Block: Pre-registration practice

Start of Block: PAPs in the paper writing process

Display This Question:

If Have you ever registered a pre-analysis plan (PAP) for a research project? = Yes

Q3.1 To what extent do you think the existence of a PAP restricted your ability to fully explore and analyze your data?

- Not at all (1)
- Somewhat (2)
- Quite a bit (3)

Display This Question:

If Have you ever registered a pre-analysis plan (PAP) for a research project? = Yes

Q3.2 To what extent do you think the existence of a PAP made it more difficult to write a theoretically interesting paper?

- Not at all (1)
- Somewhat (2)
- Quite a bit (3)

Display This Question:

If Have you ever registered a pre-analysis plan (PAP) for a research project? = Yes

Q3.3 To what extent do you think the existence of a PAP prevented you from stumbling on unexpected, surprise results?

- Not at all (1)
 - Somewhat (2)
 - Quite a bit (3)
 - Don't know (4)
-

Q3.4 Have you ever consulted a registry (EGAP, AEA) to learn whether studies on a particular research topic have ever been initiated?

- Yes (1)
- No (2)

End of Block: PAPs in the paper writing process

Start of Block: PAPs in the review process

Display This Question:

If Have you ever registered a pre-analysis plan (PAP) for a research project? = Yes

Q4.1 Have you ever included a PAP with a paper you have submitted to a journal?

- Yes (1)
 - No (2)
 - No, but the paper mentions the PAP and I have assumed that reviewers could easily find it (3)
-

Display This Question:

If Have you ever registered a pre-analysis plan (PAP) for a research project? = Yes

Q4.2 When you have submitted a pre-registered paper for publication, have reviewers ever mentioned your PAP?

Yes (1)

No (2)

Display This Question:

If When you have submitted a pre-registered paper for publication, have reviewers ever mentioned you... = Yes

Q4.3 Did they (please choose all that apply):

compare pre-specified hypotheses with those presented in the main paper? (1)

compare pre-specified variable definitions with those presented in the main paper? (2)

compare pre-specified inclusion of control variables with those presented in the main paper? (3)

refer to the PAP for some other reason (Please specify) (4)

Display This Question:

If Have you ever registered a pre-analysis plan (PAP) for a research project? = Yes

Q4.4 Have you ever invoked your PAP to respond to the suggestions of reviewers or workshop participants regarding additional analyses to run?

Yes (1)

No (2)

Q4.5 As a reviewer, have you ever consulted the PAP of a paper you are reviewing for a journal?

Yes (1)

No (2)

Display This Question:

If Have you ever registered a pre-analysis plan (PAP) for a research project? = Yes

Q4.6 Outside of the formal review process at a journal, has another researcher ever invoked your PAP when discussing your paper?

Yes (1)

No (2)

End of Block: PAPs in the review process

Start of Block: Open-ended questions

Display This Question:

If Have you ever registered a pre-analysis plan (PAP) for a research project? = Yes

Q5.1 Thinking back to the part of your scholarly career before you began regularly pre-registering your studies, how, if at all, has preregistration changed the way you conduct research? *(Please write your response in the box below. You can write as much or as little as you want).*

Display This Question:

If Have you ever registered a pre-analysis plan (PAP) for a research project? = Yes

Q5.2 If you answered earlier that you have ongoing research studies for which you have not registered a PAP, what was it about those studies that made you decide that pre-registration was not necessary? *(Please write your response in the box below. You can write as much or as little as you want).*

Display This Question:

If Have you ever registered a pre-analysis plan (PAP) for a research project? = Yes

Q5.3 How have you dealt in your research papers with deviations from what you pre-registered in your PAP? *(Please write your response in the box below. You can write as much or as little as you want).*

Display This Question:

If Have you ever registered a pre-analysis plan (PAP) for a research project? = Yes

Q5.4 Has pre-registration ever been helpful for dealing with implementing partners or funders? If so, how? *(Please write your response in the box below. You can write as much or as little as you want).*

Display This Question:

If Have you ever registered a pre-analysis plan (PAP) for a research project? = Yes

Q5.5 Have your views on pre-registration changed over time? If so, how? *(Please write your response in the box below. You can write as much or as little as you want).*

End of Block: Open-ended questions
