

Pre-Analysis Plans: An Early Stocktaking

George K. Ofosu and Daniel N. Posner

Pre-analysis plans (PAPs) have been championed 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 registered on the Evidence in Governance and Politics (EGAP) and American Economic Association (AEA) registration platforms to assess whether PAPs registered in the early days of pre-registration (2011–2016) were sufficiently clear, precise, and comprehensive to achieve their objective of preventing “fishing” and reducing the scope for post-hoc adjustment of research hypotheses. We also analyze a subset of ninety-three PAPs from projects that 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 registered during this period accomplished the goals they were designed to achieve. We discuss these findings in light of both the costs and benefits of pre-registration, showing how our results speak to the various arguments that have been made in support of and against PAPs. We also highlight the norms and institutions that will need to be strengthened to augment the power of PAPs to improve research credibility and to create incentives for researchers to invest in both producing and policing them.


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 As shown in figure 1, which displays the number of PAPs registered on the Evidence in Governance and Politics (EGAP) and American Economic Association (AEA) registries since 2011, their numbers have skyrocketed in recent years.² Graduate students are now taught that registering a PAP is a *de rigeur* part of undertaking their research projects.


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” (Simmons, Nelson, and Simonsohn 2011; Wicherts et al. 2016) that provide latitude for consciously or unconsciously selecting particular specifications that make the results more striking.

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 address 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 presented in the paper are confirmatory (i.e., testing hypotheses specified before the

A list of permanent links to Supplemental Materials provided by the authors precedes the References section.

Data replication sets are available in Harvard Dataverse at: <https://doi.org/10.7910/DVN/DOELUB>

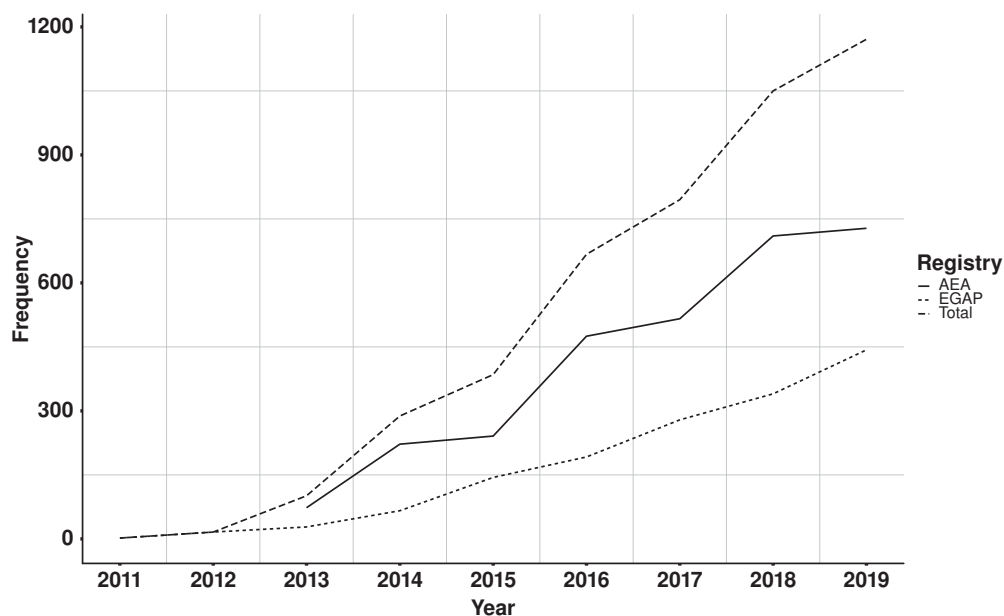
George K. Ofosu  is Assistant Professor in the Department of Government, London School of Economics and Political Science (g.ofosu@lse.ac.uk).

Daniel N. Posner  is James Coleman Professor of International Development in the Department of Political Science, University of California, Los Angeles (dposner@polisci.ucla.edu).

doi:10.1017/S1537592721000931

© The Author(s), 2021. Published by Cambridge University Press on behalf of the American Political Science Association. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted re-use, distribution, and reproduction in any medium, provided the original work is properly cited. 1

Figure 1
PAP registrations on the EGAP and AEA Registries, 2011–2019



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.⁴ One cannot compare 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 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.⁵

We therefore adopt a different approach. Rather than attempt to test whether PAPs *cause* research to be more credible, we ask whether PAPs are written and employed in a way that makes such an improvement in research credibility possible. To do this, we draw a representative sample of PAPs and analyze their contents 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 a subset of the PAPs we examine to the publicly available papers that report the findings of the investigations they pre-specified. These are, of course, subjective evaluations. But we have undertaken in our coding rules and our procedures to be both transparent and objective in the judgements we make.⁶ Our analysis provides 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 Our findings suggest that, in many cases, they are not.

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; Dufflo et al. 2020). The modal researcher in our 2018 potential PAP users' survey (discussed later) reports spending two to four weeks preparing her pre-registration materials, and more than one-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 breakthroughs that come from unexpected findings, restrict flexibility to adapt to changing circumstances or new opportunities, and generate boring, mechanical papers that are disfavored by reviewers and 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. In addition, 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 provide little guidance for undertaking and 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 promising research designs.

While there are good responses to many of these objections (many of which we discuss later), 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 undertaking the stocktaking exercise we present here. We have summarized our findings to be accessible to members of the discipline who have heard about PAPs but are not familiar with the rationale behind them or the debates surrounding their usage. In this respect, the paper serves as both an introduction to this important and relatively new research practice, and as an empirical evaluation of whether it is achieving its desired ends. Because this stocktaking covers only the first six years of PAP adoption, it provides clearer answers about the ability of the first generation of PAPs to reduce the scope for fishing and HARKing than about the clarity, precision, or completeness of PAPs registered in the last year or two. However, the discussion of the costs and benefits of pre-registration, along with the discussion of the complementary norms and institutions that might encourage and reinforce the positive impacts of registering a PAPs, remain highly relevant today.

Are PAPs Achieving Their Objectives?

Empirical Approach

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 EGAP and AEA registries between their initiation and 2016.⁸ 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 journal articles or working papers. Our procedures, which we describe in detail in the Appendix, yielded a sample of 195 PAPs, 93 of which had resulted in publicly available papers.

We coded all 195 PAPs according to a common rubric that recorded details of the pre-specified hypotheses; 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 ninety-three 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 online appendix C. 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 or whether it included a power analysis, committed to a multiple testing adjustment, or was ever private/gated—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 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.

We supplemented our coding of PAPs with an anonymous survey of potential PAP users to elicit their experiences with writing and using PAPs in their research. We were especially interested in collecting information about investigators' decisions surrounding whether or not to pre-register a study, and how the practice of composing and registering a PAP had changed the ways in which they went about their work, as well as how the rise of pre-registration affected their professional behavior more generally. The survey was conducted in 2018, so it captures a set of attitudes and behaviors closer to the present day than the patterns reflected in the PAPs we coded. The survey (reproduced in full in online appendix D) 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.⁹

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. Eighty-one percent of PAPs were

registered on the EGAP or AEA 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 EGAP and AEA registries during the period we studied, and reflecting the rapid uptake of pre-registration during this time frame, 45% of the PAPs we coded were registered in 2016, the final year of our analysis. This imbalance (somewhat) allays concerns that the findings we present come from the very early period of PAP usage, when researchers were still just learning how to use PAPs as tools in their research. However, it is impossible to rule out that the patterns we find are different from those we would have discovered had we focused on the present day rather than the first six years of pre-registration.¹¹

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.¹² 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 by ignoring or downplaying null/disconfirming results that, if reported, might provide a more accurate reflection of the true relationships in the area of study.

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 outcomes of interest in overly broad and unspecific terms—for example, “political participation,” “democratic consolidation,” or “educational attainment”—without specifying how these concepts are to be measured. Promising to “create an index” or do a “content analysis of programming” without specifying exactly how the index is to be constructed or the content analysis is to be undertaken offer other illustrations. None of these examples would pass the Olken test described earlier. 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.¹³

PAP authors were not as 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 vaguely 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 in such broad and non-specific terms leaves wide scope for fishing at the data analysis stage. Even when attempts are made to clarify how such variables are to be measured, the clarifications themselves are sometimes also problematic. For example, defining “wealth” as an index based on characteristics such as the condition of a respondent’s dwelling, asset ownership, or the number of days household members go without food still leaves broad latitude for subjectivity (which dwelling conditions? which specific assets? what if there is enough food for some family members but not others?) and fails the Olken test.

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. 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, and 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 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 or attrition; 13% specified how they would deal with noncompliance; 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, noncompliance, 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 of pre-specifying responses to every possible implementation problem that might arise are severe. As Duflo et al. (2020) underscore, “trying to write a detailed PAP that covers all contingencies, especially the ones that are *ex ante* unlikely, becomes an extraordinarily costly enterprise.” One response to this problem is the adoption 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 committing to 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.

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 research hypotheses in a way that leaves no ambiguity about the propositions that the analysis will test. In this respect, PAP authors in the sample we studied did 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 favorably. 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 figure 2, panel A). PAPs that pre-specify so many hypotheses raise questions about the value of pre-registration.¹⁶

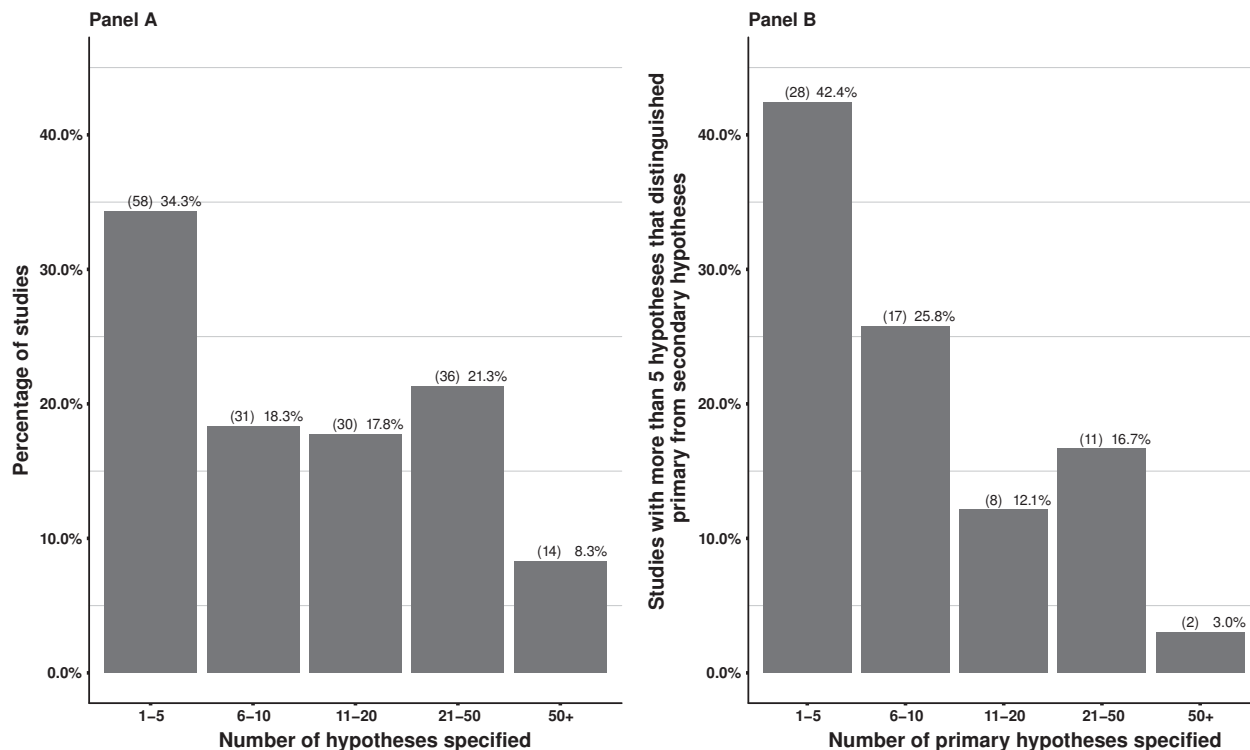
One safeguard against this pitfall is to distinguish between primary and secondary hypotheses. Many PAPs adopt this protection: among authors who pre-specified more than five hypotheses, 60% make such a distinction. But they often do so in ways that do little to solve the underlying problem. As shown 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 pre-specified six to ten primary hypotheses; 12% pre-specified eleven to twenty; 17% pre-specified twenty-one to fifty; and 3% pre-specified more than fifty. 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, 29% 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 subsample of ninety-three PAPs that had publicly available papers and compared the primary hypotheses pre-specified in the PAP with the hypotheses discussed in the paper 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 one-third of studies had at least one pre-registered hypothesis 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 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, sometimes in response to requests by reviewers, by

Figure 2
Number of pre-specified hypotheses



Notes: Panel A shows the distribution of the number of hypotheses pre-specified in the full sample of PAPs. Panel B limits the sample to the subset of PAPs that pre-specified more than five hypotheses and that distinguished between primary and secondary hypotheses.

reporting the results of hypotheses that were not pre-registered at all. We found that 18% of the papers in our sample presented tests of novel hypotheses that were not pre-registered.¹⁹ 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.

Other Issues

Addressing the “file drawer problem.” Beyond reducing the scope for fishing and post-hoc hypothesis adjustment, PAPs can help address the “file drawer problem” (Rosenthal 1979).²⁰ 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, or journals not to publish results that fail to reach conventional thresholds of statistical significance.²¹ Although the root of the file drawer problem lies in disciplinary norms that disfavor null results, pre-registration and PAPs can aid in addressing the dilemma.

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. Although studies commonly fail to result in publications or working papers for reasons that are uncorrelated with the outcomes that they generated, much evidence suggests that some fail to enter the public realm because they generate null results (Gerber and Malhotra 2008; Franco, Malhotra, and Simonovits 2014; Andrews and Kasy 2019). With pre-registration, consumers of research gain access to a record of studies that were initiated but never made public, thus enabling consumers of research to make an educated inference about how likely it is that the findings 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 be skeptical of 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 have been undertaken. We asked researchers about

this in our potential PAP users' survey, and 38% reported that they had ever consulted a registry for this purpose.²² 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 that pre-registration provides about what has been done.

Several journals in political science and economics have responded to the file drawer problem by experimenting with “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.²³ Registered reports enhance the probability of publishing null results on questions of theoretical importance and align the incentives of paper authors and reviewers to present the very best articulation of the theory and the most appropriate empirical tests.

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).

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 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 potential 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.

Objections to PAPs

In addition to allowing us to evaluate whether PAPs are delivering on their promise, our data also puts us in a position to address some of the objections to PAPs that have been raised in the literature.²⁴

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 potential PAP users' survey reported devoting a week or more to writing the PAP for a typical project, with 32% reporting spending an average of two to four 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. In some situations—for example, when there is a limited window of opportunity to initiate an experiment before an election takes place or a new policy comes into force—such delays can make it impossible to undertake the project at all, and the opportunity can be lost. The time cost of registering and adhering to a PAP may also exclude researchers from less well-resourced institutions who do not have the time, resources, or training to carry out research in the ways that PAPs require.

However, while the potential 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.”²⁵ An overwhelming majority (eight in ten) said that drafting a PAP caused them to discover things about their project that led to refinements in their research protocols 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 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. To the extent that they do, this cost must be weighed against the benefits to research credibility that result from a study whose analyses and hypotheses were pre-registered.

Limit Flexibility and Scope for New Discoveries

Another major critique of PAPs is that they constrain flexibility to adapt to unanticipated circumstances and limit the scope for new discoveries that come from unrestricted explorations of one's data.²⁶ One researcher in our

potential 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." Others worry more generally that a mode of inquiry focusing exclusively on the investigation of a narrow set of pre-specified relationships will remove opportunities for understanding relationships between variables, the sensitivity of different empirical tools to different types of data, and other investigations that provide seasoned researchers with the intuitions that set them apart from novices. These are important critiques, but they were outlier views in our users' survey. 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.²⁷ Such an approach can be particularly attractive in situations where prior information about the subject of study is limited, making it difficult for researchers 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 2020).

The more common approach—and the approach we advocate—is to freely undertake exploratory investigations that go beyond the PAP, clearly labeling the results of such investigations 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 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. It also allows researchers to avoid the selective attention trap highlighted by Yanai and Lercher (2020). Pursuing such a strategy faithfully, with findings clearly marked as pre-registered or exploratory and explanations provided for each deviation from the PAP—along with the mandatory reporting, in the body of the paper or the appendices, of every analysis that was

pre-specified—might appear to come at the expense of the tight narrative that reviewers and journal editors are thought to favor. However, in an analysis of publication outcomes of experimental NBER working papers that do and do not include PAPs, Ofosu and Posner (2020) find that while papers with PAPs are, in fact, slightly less likely to be published, they are more likely to land in a top-five journal, conditional on being published.

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, 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.²⁸ 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 potential 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 they 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).²⁹ A similar share said that other researchers had invoked their PAP when discussing their paper outside of the formal review process (35%), 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.³⁰

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.³¹ 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 e-mail, and only 58% were willing to share their PAP.³² 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.” Another expressed frustration with the different expectations of different participants in the review process: “Some reviewers didn’t like when we distinguish between hypotheses that were included in the PAP and those that were not. But other reviewers thought we were trying to hide something when we presented all the results (PAP and non-PAP) together.” 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 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.

Getting Scooped

We also asked researchers in our potential 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%.

The Balance Sheet

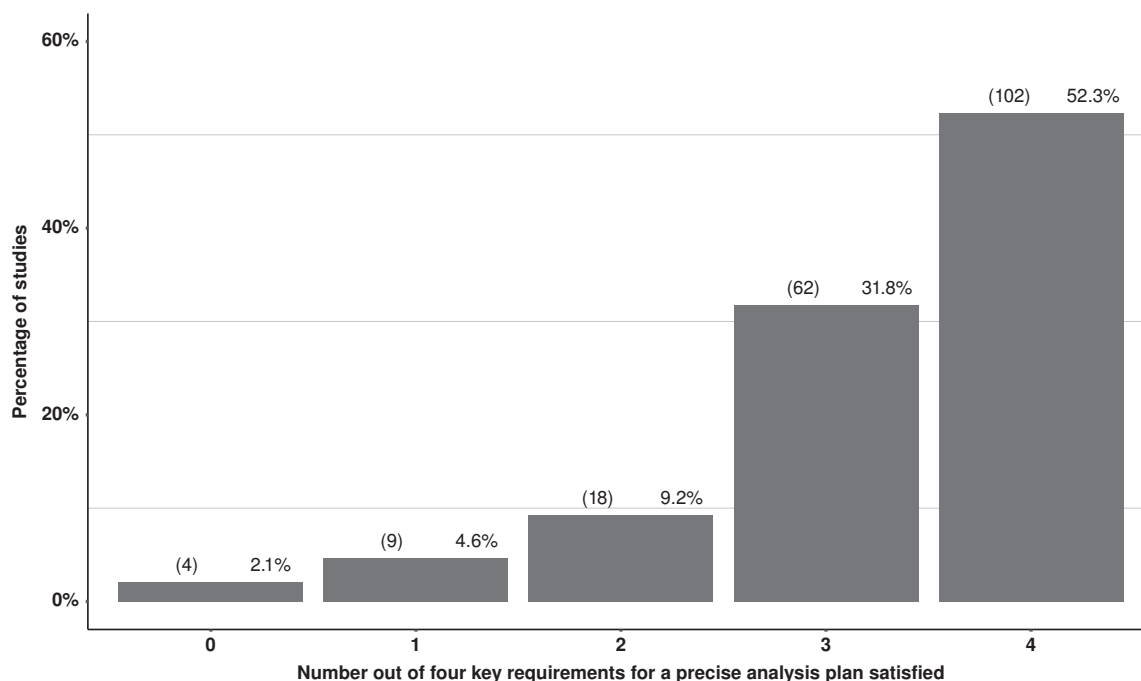
Our stocktaking suggests that PAPs registered during the first six years of the “pre-registration revolution” (Nosek et al. 2018) were often not written or used in a way that allowed them to do everything that their proponents had hoped. Many PAP authors were insufficiently clear about

the hypotheses they were testing to prevent them from moving the goal posts once they had seen the patterns in their data. The details of the analyses that PAPs pre-specified—how outcome and causal variables were to be operationalized; which controls would be included; what the statistical model would look like; how imbalances, outliers, and attrition would be dealt with—were not always adequate to reduce researcher degrees of freedom in a meaningful way. In addition, papers that resulted from pre-registered analyses did not always follow what was pre-registered. Some papers introduced entirely novel hypotheses; others presented only a subset of the hypotheses that were pre-registered.

But documenting that not all PAPs adequately addressed all of the problems they were designed to solve does not imply that the growing use of PAPs in political science and economics during this period did not generate 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 over half of the 195 PAPs we analyzed were judged to meet all four of these criteria, and about another third were judged to satisfy three of the four.³³ 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 was not foreclosed by *every* PAP, such opportunities were 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.

As PAP skeptics point out, however, these improvements to research credibility came at a price. Writing a PAP occupies weeks of valuable research time, and adhering faithfully to what was pre-specified may limit flexibility and 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 flexibility, creativity, and exploration can be loosened by simply labeling one’s investigations as exploratory or confirmatory or by explaining the exigent circumstances that necessitated the departure from what was pre-specified. 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 potential PAP users’ survey suggest that PAPs do not restrict researchers’

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



Notes: 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.

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.” The cost of writing and adhering to a complete and comprehensive PAP may simply be the price researchers need to pay for making their research more credible.

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 was to scrutinize whether PAPs were 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 and how PAPs should be adapted for observational studies, which comprise the majority of research projects undertaken in political science and economics (Burlig 2018; Jacobs 2020). Although several publications provide recommendations for what authors should (and need not) include in their PAPs (McKenzie 2012; Glennerster and Takavarasha 2013; Piñeiro and Rosenblatt 2016; Kern and Gleditsch 2017; Christensen, Freese, and Miguel 2019; Chen and Grady 2019; Dufflo et al. 2020), there are no universally agreed upon rules in either discipline 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 *DeclareDesign* (Blair et al. 2019), which provides software that allows researchers to formally describe (and troubleshoot) the

details of their proposed analyses, provide clear templates 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 credibility-enhancing benefits of PAPs to be fully realized. Creating disciplinary incentives for such policing is a critical challenge.

The most logical venue for such scrutiny is the journal review process.³⁴ 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 or procedures is too unclear or incomplete to remove the scope for fishing or HARK-ing? Or what if, as in Bidwell, Casey, and Glennerster (2020), 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 the enhanced research credibility generated through pre-registration accrues to the pre-registered studies themselves, some of the benefits of pre-registration depend on the adoption of the practice 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—especially if standards for the precision and comprehensiveness of PAPs are tightened in the ways we are suggesting they need to be.³⁵ The recently

completed State of Social Science Survey (Christensen et al. 2019) finds that while the majority of researchers in political science and economics are aware of and support the norm of pre-registration, behavior in adopting the practice is significantly lagging. One key obstacle, revealed both in our data and in the evidence summarized in Christensen et al. (2019), is the hesitancy of authors of observational studies to register 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 (2020) 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.³⁶

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, the Working Group in African Political Economy, and the Northeast Workshop in Empirical Political Science 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—including within professional associations like APSA and AEA—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 help to offset the cost of composing the PAP, and thus increase willingness to make such investments in the first place.

Although their use has risen steeply in recent years, PAPs are still in their relative infancy. Our analysis, which covers PAPs registered between 2011 and 2016, captures the early years of PAP usage. This was a time when many authors were registering their first PAPs, and when norms about both 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 PAPs 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.

Supplementary Materials

- A. E-mail to Authors of Private/Gated PAPs
- B. Summary Statistics
- C. Coding Rubric for PAPs with Papers
- D. Potential Pap Users' Survey

To view supplementary material for this article, please visit <http://dx.doi.org/10.1017/S1537592721000931>.

Acknowledgements

The authors thank Maxim Ananyev and Merabi Chkhenkeli for excellent research assistance, and Graeme Blair, Ashley Blum, Alex Coppock, Steven Glazerman, Donald Green, Macartan Humphreys, Matt Lisiecki, and David McKenzie, as well as seminar participants at EGAP; Oxford; and University of California, Berkeley; and four anonymous reviewers 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 Sciences (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/>.

Notes

- 1 PAPs 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.
- 2 Our stocktaking focuses on patterns in political science and economics, 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 merged with the OSF Registry.
- 3 An illuminating illustration of the scope for fishing within a real study is provided in Casey, Glennerster, and Miguel 2012. For evidence of the prevalence of fishing in political science, see Gerber and Malhotra 2008; for economics, see Brodeur et al. 2016. For discussions of the incentives for researchers to present more striking results, see Elman, Kapiszewski, and Lupia 2018; Noesk et al. 2018; and Laitin and Reich 2017.
- 4 For a notable attempt to estimate the causal effect of registration in the medical field, see Fang, Gordon, and Humphreys 2015.
- 5 In our potential PAP users' survey (discussed later) 78% of researchers said they had at least one ongoing research project for which they did not register a PAP.
- 6 We did not, however, pre-register our analysis or any specific hypotheses, as we view this research as a purely descriptive exercise.
- 7 Ocen and Woods 2019 provide a similar stocktaking, coding PAPs in terms of several key design features. However, their study makes no effort to evaluate whether each of these features is presented sufficiently clearly or comprehensively to reduce the scope for fishing or post-hoc theorizing. Another analogous effort, involving the comparison of published and unpublished papers with the proposals that secured their funding, is provided in Franco, Malhotra, and Simonovits 2014. For an analysis similar to our own in psychology, see Claesen et al. 2019.
- 8 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 PAP, our analysis only includes studies for which a PAP was uploaded. To 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 generally.
- 9 Because the survey was sent to a population of researchers likely to have registered a PAP, our results are biased toward the views and experiences of PAP users. This is not a problem—indeed it is a requirement—for questions about researchers' experiences with pre-registration. But it may bias responses to questions about other issues, such as whether or not the researcher has consulted a registry and, possibly, his/her views on the costs and benefits of writing and adhering to a PAP (although it was clear from our survey results that many respondents who reported registering PAPs did so because they thought the profession demanded it of them rather than because they were sold on their benefits). The results we discuss later should be read with this caveat in mind.
- 10 In the potential 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.
- 11 In other work, we are coding a random sample of more recent PAPs with the objective of comparing patterns across the early and contemporary periods.
- 12 See 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.
- 13 The 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.
 - 14 Lenz and Sahn (forthcoming) 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.
 - 15 The 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 they report 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.
 - 16 Closely related to the number of hypotheses is the length of many PAPs. While the median PAP in our sample was eleven single-spaced pages, the longest 10% were more than thirty-one pages, and three were over ninety pages long. As an insightful reviewer points out, one reason why PAPs are so long and unwieldy is because, just as with academic papers, tightening and sharpening them is hard intellectual work. Under the current set of disciplinary incentives, many researchers feel they will get little payoff for investing in this effort.
 - 17 Researchers 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 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 later.
 - 18 Consistent with this explanation, 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%), although this difference is not statistically significant.
 - 19 Consistent with the suspicion that the addition of novel hypotheses might be due to reviewers’ requests, published papers were twelve percentage points (80%) more likely to report hypotheses that were not pre-registered. However, this result is not statistically significant due to our small sample size.
 - 20 Filing 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.
 - 21 Although the file-drawer problem is commonly assumed only to affect confirmatory or quantitative research, Jacobs 2020 shows that it generates strong publication bias in qualitative studies as well.
 - 22 This figure is likely an overestimate of the frequency of registry consultation in the profession more broadly, as the PAP users’ survey captured the views of researchers more likely to be aware of registries and to recognize their utility for this purpose.
 - 23 Journals 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://cos.io/tr>.
 - 24 Useful discussions of objections to PAPs that go beyond the ones discussed here—and that echo several of the challenges articulated by respondents in our potential 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; Nosek et al. 2018; and Duffo et al. 2020.
 - 25 Six 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.”
 - 26 Yanai and Lercher (2020) demonstrate this point via an experiment in which participants were asked to analyze a fictitious dataset that, if plotted, clearly reveals the image of a gorilla. Half of the participants were given specific hypotheses to test in the data, and the other half were not. The latter, hypothesis-free,

participants were five times as likely to discover the gorilla pattern as the participants who were given hypotheses to investigate in advance. The authors explain this result as stemming from blindness due to selective attention “to the hypotheses that were given in advance,” which they characterize as a “hidden cost” of pre-specifying a hypothesis.

- 27 Examples include Bidwell, Casey, and Glennerster 2020 and Blair et al. 2019.
- 28 As we have learned in our coding work for this project, this is challenging, time-consuming work—especially, as Bidwell, Casey, and Glennerster 2020 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).
- 29 Among political scientists, the ability of reviewers to examine a publicly posted PAP is complicated by the double-blind review process employed in most disciplinary 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).
- 30 An insightful discussion of policing norms in political science and the challenges of changing them is provided in Laitin and Reich 2017.
- 31 For recent evidence on the adoption of such open social science practices, see Christensen et al. 2019.
- 32 Further details of our efforts to contact the authors of private/gated PAPs are provided in online appendix A.
- 33 We investigated whether these results differed across the roughly half of PAPs in our sample from 2016 versus PAPs from earlier years and find no statistically significant differences, suggesting the absence of a trend in improving or declining quality—at least across the six-year period we study.
- 34 An 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. Such efforts could complement the scrutiny provided by formal journal reviews.
- 35 Indeed, some have argued that design registries should be more lenient in terms of standards so as to encourage people to start using them, with standards tightened once the research community buys into the norm of pre-registration more fully.

- 36 For useful discussions of the challenges of pre-registering observational and qualitative research, see Burlig 2018; Elman, Kapiszewski, and Lupia 2018; Christensen, Freese, and Miguel 2019; and Jacobs 2020.

References

- Andrews, Isaiah, and Maximilian Kasy. 2019. “Identification of and Correction for Publication Bias.” *American Economic Review* 109(8): 2766–94.
- Ansell, Ben, and David Samuels. 2016. “Journal Editors and ‘Results-Free’ Research: A Cautionary Note.” *Comparative Political Studies* 49(13): 1809–15.
- Bidwell, Kelly, Katherine Casey, and Rachel Glennerster. 2020. “Debates: Voting and Expenditure Responses to Political Communication.” *Journal of Political Economy* 128(8): 2880–924.
- Blair, Graeme, Fotini Christia, Cyrus Samii, and Jeremy Weinstein. 2019. “Meta-analysis Pre-analysis Plan: Community Policing Metaketa Project.” Pre-analysis plan. (<https://osf.io/phjmd/>).
- 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.
- Burlig, Fiona. 2018. “Improving Transparency in Observational Social Science Research: A Pre-Analysis Plan Approach.” *Economics Letters* 168: 56–60.
- Bush, Sarah, Aaron Erlich, Lauren Prather, and Yael Zeira. 2018. “Lessons from Results-Blind Review.” *Political Economist* (Spring). (<https://docs.google.com/viewer?&v&pid=sites&srcid=dW1pY2guZWV1fHBvbGl0aWNhbC1lY29ub21pc3QtbmV3c2xldHRlcXneDoxZjU3NWl4ZjgyMjNmN2M4>).
- Casey, Katherine, Rachel Glennerster, and Edward Miguel. 2012. “Reshaping Institutions: Evidence on Aid Impacts Using a Preanalysis Plan.” *Quarterly Journal of Economics* 127(4): 1755–812.
- Chen, Lula, and Chris Grady. 2019. “10 Things to Know about Pre-analysis Plans.” *Evidence in Governance and Politics (EGAP)*, Institute of Government Studies, University of California, Berkeley. (<https://egap.org/resource/10-things-to-know-about-pre-analysis-plans/>).
- Christensen, Garret, Jeremy Freese, and Edward Miguel. 2019. *Transparent and Reproducible Social Science Research: How to Do Open Science*. Berkeley: University of California Press.
- Christensen, Garret, Zenan Wang, 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. (<https://osf.io/preprints/metaarxiv/5rksu/>).

- Claesen, Aline, Sara Gomes, Francis Tuerlinckx, and Wolf Vanpaemel. 2019. "Preregistration: Comparing Dream to Reality." Working paper. (<https://psyarxiv.com/d8wex/>).
- 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.
- Duflo, Esther, Abhijit Banerjee, Amy Finkelstein, Lawrence F. Katz, Benjamin A. Olken, and Anja Sautmann. 2020. "In Praise of Moderation: Suggestions for the Scope and Use of Pre-analysis Plans for RCTs in Economics." NBER Working Paper 26993. DOI 10.3386/w26993
- Elman, Colan, Diana Kapiszewski, and Arthur Lupia. 2018. "Transparent Social Inquiry: Implications for Political Science." *Annual Review of Political Science* 21:29–47.
- Fafchamps, Marcel, and Julien Labonne. 2017. "Using Split Samples to Improve Inference on Causal Effects." *Political Analysis* 25(4): 465–82.
- Fang, Albert, Grant Gordon, and Macartan Humphreys. 2015. "Does Registration Reduce Publication Bias? Evidence from Medical Sciences." (https://albertfang.com/research/ps_paper.pdf).
- 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): 1667–703.
- 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*, World Bank Blogs, July 15. (<https://blogs.worldbank.org/impacetevaluations/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): 1502–505.
- 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): 313–26.
- Glennerster, Rachel, and Kudzai Takavarasha. 2013. *Running Randomized Evaluations: A Practical Guide*. Princeton, NJ: 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 P.A. 2005. "Why Most Published Research Findings Are False." *PLoS Medicine* 2(8): e124.
- Jacobs, Alan M. 2020. "Pre-registration and Results-free Review in Observational and Qualitative Research." In *The Production of Knowledge: Enhancing Progress in Social Science*, ed. Colin Elman, John Gerring, and James Mahoney, 221–264. New York: Cambridge University Press.
- Kern, Florian G., and Kristian Skrede Gleditsch. 2017. "Exploring Pre-registration and Pre-analysis Plans for Qualitative Inference." Working paper. (https://www.researchgate.net/publication/319141144_Exploring_Pre-registration_and_Pre-analysis_Plans_for_Qualitative_Inference/link/599455d60f7e9b98953af045/download).
- Laitin, David D. 2013. "Fisheries Management." *Political Analysis* 21(1): 42–47.
- . 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): 172–75.
- Lenz, Gabriel, and Alexander Sahn. Forthcoming. "Achieving Statistical Significance with Covariates and without Transparency." *Political Analysis*. DOI 10.1017/pan.2020.31
- 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*, World Bank Blogs, October 28. (<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): 2600–606.
- Oceno, Marizia, and Logan T. Woods. 2019. "Preregistered Research Designs: Trends in Quality and Publication." Working paper. (https://drive.google.com/file/d/1jFxsL8w1WH5kdnzVAoIT-6AX4z5_3IXO/view).
- Ofosu, George K., and Daniel N. Posner. 2020. "Do Pre-analysis Plans Hamper Publication?" *AEA Papers & Proceedings* 110: 70–74.
- 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.
- Piñeiro, Rafael, and Fernando Rosenblatt. 2016. "Pre-analysis Plans for Qualitative Research." *Revista de Ciencia Política* 36(3): 785–96.
- Rosenthal, Robert. 1979. "The 'File Drawer Problem' and the Tolerance for Null Results." *Psychological Bulletin* 86(3): 638–41.

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): 1359–66.

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): 534–47.

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. (<https://dx.doi.org/10.3389%2Ffpsyg.2016.01832>).

Yanai, Itai, and Martin Lercher. 2020. “Selective Attention in Hypothesis-driven Data Analysis.” bioRxiv preprint. DOI <https://doi.org/10.1101/2020.07.30.228916>

Appendix

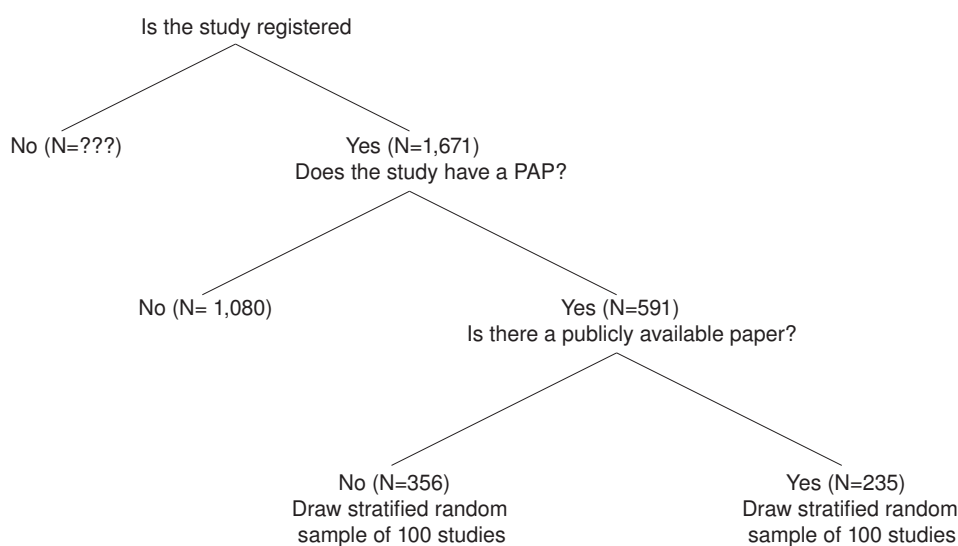
A total of 1,671 studies were registered on the EGAP (436) and AEA (1,235) registries during the period we studied (see figure A1).¹ Of these, 591 had PAPs (322 of the EGAP-registered studies and 269 of the AEA-registered studies).² We then identified whether each study had resulted in a publicly available 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.

We then drew a random sample of one hundred of these studies, alongside a random sample of one hundred studies

that had not yet resulted in a publicly available paper. In drawing these samples, we stratified by three criteria: the year the study was registered (2011, 2012, 2013, 2014, 2015, 2016), whether the study was registered on the EGAP or AEA registries, and whether the PAP was initially gated/private.

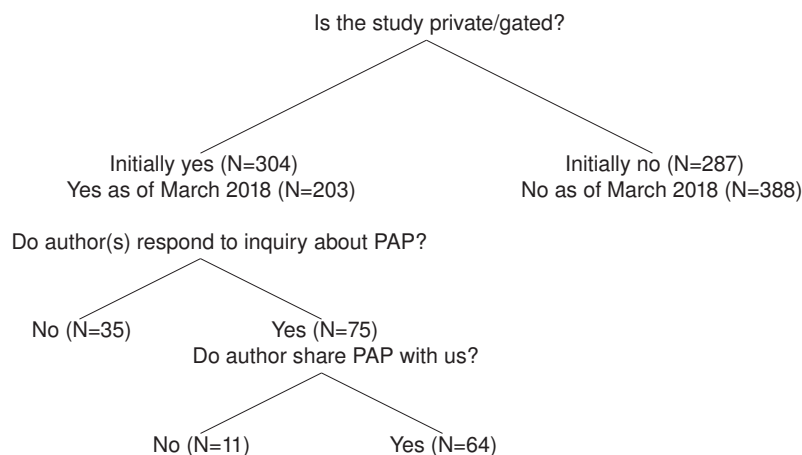
The fact that not all PAPs are made public at the time a study is registered created a challenge for our coding exercise. The AEA registry affords study authors the opportunity to keep their PAPs private and the EGAP registry, while strongly discouraging researchers from doing so, permits study authors to gate their PAP for a period of time.³ As shown in figure A2, of the 591 studies with PAPs, 304 were initially private/gated, although

Figure A1
Sampling procedures



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

Figure A2
Dealing with private/gated PAPs



Note: We contacted the authors of 123 studies and can confirm that 110 read our e-mail.

101 of those had become public/ungated by the time we drew our sample in March 2018.

To reach our goal of coding one hundred PAPs from projects that had resulted in publicly available papers and another set of one hundred 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% 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 e-mail to ask them to confidentially share their PAPs with us.⁴ Of the 120 authors who we can confirm received and read our e-mail, we received replies from 75 (68%), of which 64 (58%) were willing to share their PAP.⁵

Our procedures 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 drew our sample were no longer 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. Summary statistics are provided in online appendix B.

Appendix Notes

1 In 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.

- 2 The 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.
- 3 Our 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.
- 4 A copy of the e-mail, which was sent on April 10, 2018, is provided in online appendix A. We sent a reminder e-mail nine days later to authors who had not yet responded.
- 5 Of the eleven 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.