THE VALUE OF PRE-ANALYSIS

DAVID YOKUM AND JAKE BOWERS

INTRODUCTION

We describe the idea of a pre-analysis plan (PAP) and explain why you should use one. We emphasize the potential *political* uses of PAPs and how the PAP is, in this respect, a uniquely powerful tool for advancing the next generation of evidence.[1] We give examples from our experiences with PAPs over the past decade.

WHAT IS A PRE-ANALYSIS PLAN?

A pre-analysis plan (PAP) is a document describing how a research project will be conducted, written *before* data is collected or analyzed. The document explains what questions will be asked and how data will be collected and analyzed to answer those questions. The "registration" of a PAP involves publishing the document, with a timestamp, into a public location where it cannot be further edited.[2] A registered PAP is, therefore, a transparent record of what a researcher believed before conducting a study and how the researcher intended to update their beliefs with data.

There is substantial variation in how PAPs are written.[3] A PAP may contain dozens of pages, or maybe only one page or even a few sentences. The description may (or may not) include literature reviews, hypothesis statements, equations, mock figures and tables, code, or data simulations.

People have offered templates, checklists, and guidelines in an attempt to standardize—or at least set minimal standards for—the content and level of detail within a PAP.[4,5,6,7,8] But, ultimately, the researcher must use their own judgment to decide how much detail to include in a PAP, given the context and aims of the study.

WHY USE A PRE-ANALYSIS PLAN?

Pre-analysis plans help individual research teams and evidence-based policy in general in three main ways:

- PAPs enhance research integrity.
- PAPs prompt project management best practices.
- PAPs can be leveraged to facilitate political decision making.

Depending on which uses researchers pursue and to what degree, more or less detail will be required in the PAP.

PAPs Enhance Research Integrity

The first and foremost benefit—and the most common reason PAPs are becoming a standard practice throughout the academic community—is that PAPs *enhance research integrity*. In particular, the publicly registered PAP is a strategy for hedging against risks of *p*-hacking, HARKing, and publication bias.

P-Hacking

In the course of a study, a researcher will make hundreds of decisions regarding the design of data collection and how those collected data will be analyzed and reported.[9] These decisions can substantially affect what results are uncovered and shared.[10] For example, in considering whether the U.S. economy is affected by whether Republicans or Democrats are in office, decisions need to be made about how to operationalize economic performance (for example, employment, inflation, GDP), which politicians to focus on (for example, presidents, governors, senators), which years to examine, whether to entertain exclusions (for example, ignore), whether models should be linear or nonlinear, and so forth. To *p*-hack would be to try combinations of those decisions until "statistically significant" results surface.[11] This could happen intentionally or, much more commonly, unintentionally.[12]

The website FiveThirtyEight provides an interactive tool to build your *p*-hacking intuitions. Visit the website at https://fivethirtyeight.com /features/science-isnt-broken/ (or search "Aschwanden, Science Isn't Broken"). Toggle values on the "Hack Your Way to Scientific Glory" applet (it is in the middle of the article) to experience firsthand how, depending on your choices, you can reach literally *any* conclusion about the impact of political party on the U.S. economy.

The PAP hedges against *p*-hacking by forcing researchers to make these methodological choices in advance, based on criteria such as theory or statistical best practice rather than being lured into jiggling choices until a desired result is achieved.[13]

HARKing

To HARK is to "*H*ypothesize *a*fter the *R*esults are *K*nown."[14] HARKing happens when a researcher presents post hoc hypotheses in a research report as if they were, in fact, a priori hypotheses. In other words, a result gets framed as predicted by theory when, in fact, the result was *not* expected given the beliefs held before the study was conducted; it is only upon seeing the results that the researcher updates their beliefs and develops a *new* theory-driven hypothesis that is consistent with the result.

The updating of beliefs is not the problem—quite to the contrary. If properly done, that is the very essence of scientific progress. The problem is how HARKing conceals and distorts the belief updating process.[15] HARKing is alchemy that presents exploratory results as if confirmatory. This sleight of hand is misleading for a variety of reasons.[16] For example, HARKing violates the principle of disconfirmability: if a hypothesis is handcrafted to match already observed data, then there is no opportunity for a hypothesis to be disconfirmed by the study. And it is disconfirmed hypotheses, not confirmed hypotheses, that most efficiently winnow the field of competing ideas and advance our understanding.[17] Consider also that HARKing disregards information: prior beliefs based on theory are ignored, and the hypothesis is, instead, constructed on the sand of currently observed data and cherry-picked rationales.

The PAP prevents HARKing by keeping clear which hypotheses were predicted in advance versus which hypotheses were generated based on new results.

Publication Bias

Researchers are more likely to write up—and journals are more likely to publish—results that are statistically significant, even holding constant the importance of the question and the quality of methods.[18] One study found that research with statistically significant results had a *40 percentage point* higher probability of being published than if results were nonsignificant.[19] Such selective reporting leads to bias in the academic literature. Positive findings become overrepresented. Null or inconclusive findings, in contrast, become underrepresented, condemned to the researcher's personal file drawer rather than shared with the community. When this happens, any review or meta-analysis of the literature is misleading. Zero or contradictory effect sizes are effectively censored, leaving only the positive and largest effect sizes in print—and, thus, false positives are more likely and effect sizes are overestimated. A job training program with two positive evaluations might seem effective, but less so when it is uncovered that ten other evaluations, never published, failed to find any benefits or perhaps even found negative side effects.

To correct publication bias, all results must be openly available, so researchers can potentially summarize the entire body of findings.

PAPs Prompt Project Management Best Practices

The second benefit is mundane but important all the same. It may be the most immediate benefit you feel by adopting PAP practices. The documentation inherent to a PAP *fosters project management best practices*. To properly write out a methodology, the team must plan for a wide variety of details. To explain how randomization will happen, for example, you must determine and map out a suite of implementation details—how exactly will the intervention be delivered and to whom and by whom and when and for how long? In mocking up a data visualization, you are forced to think clearly about what data is needed to create that figure. And so on. You are forced to conduct a sort of "pre-mortem," considering what implementation or interpretation challenges might derail the project. And that, in turn, empowers you to manage against those challenges from the outset. By documenting all these project management details, you also increase communication across the research team as well as build resiliency against staff turnover. Any new team member can be handed the PAP during onboarding to the project.

Note that the PAP process should not actually create any additional work. A PAP should, instead, alter *when* work happens, namely, sooner rather than later. The only way to avoid the PAP work is a naughty one: to plan (even if implicitly) *not* to write up details if you fail to uncover statistically significant results that advance your theorizing.

The registration of a PAP is uniquely helpful in an additional way. There is a tendency for people—especially when busy, which is essentially always the case for practitioners—to carefully review documents only when absolutely necessary. It is common for drafts of reports to be skimmed but not fully engaged. This can lead to the frustrating situation where a document is shared and everyone thinks they agree on its contents, only to later discover—when it is about to *really be* published publicly and so everyone finally really reads the thing—that disagreements or objections linger. In our experience, the fact that a PAP will be registered—it will be public and uneditable at that point—is an excellent catalyst for engaging a partner's full attention sooner rather than later.

Managing a partner's full attention may feel like an added burden. It can slow down the launch of a project because extra time may be needed to clarify questions or negotiate points of debate. But we submit that the advance time is well spent for two reasons. The time will eventually be spent anyway; if not in advance, then after the fact while clearing up confusions about what was done. Indeed, dealing with the consequences of the misunderstanding is usually *more* complicated that averting the misunderstanding in the first place. At the extreme, a partner may want you to redo the work entirely. The second, and most powerful, reason relates to the political uses of PAPs, so let's turn there now.

PAPs Can be Leveraged to Facilitate Political Decision Making

Despite slogans to "follow the science," facts alone cannot determine *any* decision. The reason is that science inevitably involves value judgments, which are created by processes other than measuring and counting.[20] There are necessary value judgments, for example, in deciding what constitutes a meaningful effect size and how much uncertainty should be tolerated in the estimate of that effect size. Resolving these decisions cannot be done on technical grounds. There is technical skill involved in the calculations—there are correct and incorrect ways to calculate a confidence interval or a p value, for instance—but subjective opinions always enter when considering

whether an impact is big enough, how to balance the risks of a false positive versus a false negative, whether to focus on mean or distributional effects, how to consider the opportunity costs of spending scarce resources on X rather than Y, and so on.

Scientists often make these value judgments entirely by themselves, either deliberately or by default in following a convention, such as setting $p < 0.05$ as the threshold for "statistical significance." In our experience, this is frequently the source of frustration on the part of stakeholders and the lay public. For example, empirical data can be marshalled to estimate how much mask-wearing reduces the transmission of COVID-19. But to step further into a decision about whether people *should* wear masks is to enter a realm of value trade-offs: the estimated benefits of reducing the risk of transmission must be weighed against the downsides of requiring people to purchase and cover their faces with masks, with added considerations for how to manage the risk of misestimating either side of the ledger.

The PAP is *a vehicle to clearly distinguish technical judgments from value judgments, and then to facilitate discussions on both fronts from the appropriate parties.*[21] For the technical components—for example, peer review of whether the randomization scheme was robust or double-checking statistical code—feedback from other experts is usually most fitting. But for the value components, it is usually the case that feedback is needed from the community affected by the research, either directly or via representatives who are making decisions on their behalf.

Consider the PAP used in an evaluation of the Washington, DC, police department's body-worn camera program.[22] Police officers were randomly assigned to wear a body camera or not (this was a randomized controlled trial), allowing the estimation of how much (if at all) body cameras reduced uses of force by way of comparing the group of officers with cameras against the group of officers without cameras. A key question was how long to run the study. From a technical standpoint, the more months of a treatment and a control group, the more precise the estimate will become. But how many months is enough? That is a political judgment. It requires assessments such as: How big of a reduction in use of force would be meaningful in policy terms? How certain do we need to be about that effect size estimate? How much are you willing to pay (in added research costs) to achieve a given precision of estimate? How much downside is there to a false positive or a false negative? And so on. The research team held over ten public events—at schools, in libraries, and beyond—taking pains to explain concepts such

as randomization, effect size coefficients, and confidence intervals, so the community could then have a robust discussion about how big of an effect size would be meaningful to them. The PAP was key to facilitating these discussions.

CONCLUSION

The PAP is a uniquely fit tool for advancing the "next generation of evidence," for it empowers all three components identified by Project Evident:

1. **Practitioner Centric**: The PAP, when properly fleshed out and created collaboratively, is geared toward practical decision making and realistic project management. Drafting the PAP requires a clear articulation of: the question(s); the parameters for what constitutes an acceptable answer(s); and how the data for that answering process can be obtained in the field.

2. **Embraces a Research and Development (R&D) Approach**: Despite being a static document, the registered PAP really is geared toward *changing beliefs*, the key nuance being that PAPs facilitate proper belief-updating by way of fostering transparency in when and why beliefs have changed.

3. **Elevates the Voices of the Community**: The PAP is a concrete document that the community can read, comment on, and, potentially, even help draft. The best PAPs are documents, plus associated events or tutorials, that explain the technical components in plain language so relevant stakeholders can engage, regardless of background.

OTHER FAQS ABOUT PAPS

Q1: Do PAPs restrict exploratory research?

A: No, absolutely not. Although PAPs are commonly applied for null hypothesis testing (where problems of p-hacking fester), there is nothing about the underlying concept—making transparent your beliefs and intentions before data collection—that is inconsistent with exploratory research. A 100 percent exploratory PAP could literally just say, "This study is exploratory; there are no predictions and every permutation of data analytics will be attempted and reported." Notice

how this simple PAP hedges against HARKing (no hypothesis at all!); alerts the reader of the many attempted statistical tests (and, thus, vigilance is needed to calibrate uncertainty estimates based on family-wise error rates, to mitigate false positives from *p*-hacking); and alleviates publication bias by creating a public record.[23]

Q2: Can I deviate from the PAP?

A: Yes, of course. Just be transparent. Insights surfaced during unanticipated, exploratory analyses are the source of many scientific breakthroughs. Not to mention, deviations are often practically necessary if the intervention was implemented differently than planned. The key is that PAPs empower everyone to keep clear on what was predicted versus what was learned through exploration. Register a new version of the PAP if you update before beginning analyses. If after, simply note in your write-up what was planned versus what was not planned.

Q3: Is the PAP process different from community engagement?

A: Yes. Any PAP that leans into political uses must entail community engagement; but community engagement (broadly defined) need not and usually does not entail a PAP. Even when researchers publicly discuss their work with stakeholders, it is relatively rare to facilitate a discussion of value judgments and then to publicly register those agreements.

Q4: Do PAPs have to be made public while a study is ongoing?

A: No. PAPs can be embargoed to have their contents hidden for a specified amount of time. What matters is that the date of their registration be trustworthy to readers.

NOTES

1. Project Evident, www.projectevident.org, describes the "Next Generation of Evidence" as: 1) practitioner centric; 2) embracing a research and development (R&D) approach; and 3) elevating the voices of the community. This is in contrast to the status quo, which more typically involves a point-in-time evaluation geared toward informing academia and external funders. See, generally, www.projectevident.org/nextgenevidence-campaign.

2. Common repositories of PAPs include https://clinicaltrials.gov/, the American Economic Association's RCT Registry (www.socialscienceregistry .org/), the Evidence in Governance and Politics Registry (https://egap.org /registry/), and the Open Science Framework (https://osf.io). The Center for Open Science. The Center for Open Science is dedicated to promoting open

science best practices, and to that end, their website contains a host of additional readings, events, and resources; www.cos.io/.

3.   George K. Ofosu and Daniel N. Posner, "Pre-Analysis Plans: An Early Stocktaking," *Perspectives on Politics* (2021), pp. 1–17, https://doi.org/10.1017/S1537592721000931.

4.   Joseph P. Simmons, Leif D. Nelson, and Uri Simonsohn, "False-Positive Psychology Undisclosed Flexibility in Data Collection and Analysis Allows Presenting Anything as Significant," *Psychological Science* 22, no. 11 (2011), pp. 1359–366.

5.   Jelte M. Wicherts and others, "Degrees of Freedom in Planning, Running, Analyzing, and Reporting Psychological Studies: A Checklist to Avoid p-Hacking," *Frontiers in Psychology* 7 (2016), https://doi.org/10.3389/fpsyg.2016.01832.

6.   Anna Elisabeth van't Veer and Roger Giner-Sorolla, "Pre-Registration in Social Psychology—A Discussion and Suggested Template," *Journal of Experimental Social Psychology*, Special Issue: Confirmatory, 67 (November 1, 2016), pp. 2–12, https://doi.org/10.1016/j.jesp.2016.03.004.

7.   Abhijit Banerjee and others, "In Praise of Moderation: Suggestions for the Scope and Use of Pre-Analysis Plans for RCTs in Economics," National Bureau of Economic Research, April 20, 2020, https://doi.org/10.3386/w26993.

8.   Nuole Chen and Chris Grady, "10 Things to Know about Pre-Analysis Plans," EGAP Methods Guide (2019), https://egap.org/resource/10-things-to-know-about-pre-analysis-plans/.

9.   J. M. Wicherts, C. L. Veldkamp, H. E Augusteijn, M. Bakker, R. Van Aert, and M. A. Van Assen, (2016), "Degrees of Freedom in Planning, Running, Analyzing, and Reporting Psychological Studies: A Checklist to Avoid P-Hacking, *Frontiers in Psychology*, November 2016, https://doi.org/10.3389/fpsyg.2016.01832.

10. Simmons, Nelson, and Simonsohn, "False-Positive Psychology Undisclosed Flexibility in Data Collection and Analysis Allows Presenting Anything as Significant."

11. The *p* value is a statistical measure of the probability of observing results as big or larger than the result observed in a sample, even if, in reality, there is no true effect in the population. So, for example, how likely would it be to get forty tails when flipping a fair coin fifty times? By convention, many scientists consider $p < 0.05$ to be "statistically significant." *See also* Aschwander, "Not Even Scientists Can Easily Explain P-Values."

12. A. Gelman and E. Loken, "The Garden of Forking Paths: Why Multiple Comparisons Can be a Problem, Even When There Is No "Fishing Expedition" or "P-Hacking" and the Research Hypothesis Was Posited ahead of Time," *Department of Statistics, Columbia University* 348 (2013), pp. 1–17.

13. Simmons and others suggest six simple requirements for authors to avoid *p*-packing: 1) decide the rule for terminating data collection before data collection begins; 2) collect at least twenty observations per cell or else provide a compelling cost-of-data-collection justification; 3) list all collected variables; 4) report all experimental conditions, including failed manipulations; 5) if observations are eliminated, also report the statistical results if those observations are included; and 6) if an analysis includes a covariate, also report the statistical results without the covariate.

14. Norbert L. Kerr, "HARKing: Hypothesizing after the Results Are Known," *Personality and Social Psychology Review* 2, no. 3 (August 1, 1998), pp. 196–217, https://doi.org/10.1207/s15327957pspr0203_4.

15. Brian A. Nosek and others, "The Preregistration Revolution," *Proceedings of the National Academy of Sciences* 115, no. 11 (March 13, 2018), pp. 2600–606, https://doi.org/10.1073/pnas.1708274114.

16. Kerr, "HARKing."

17. Karl Popper, *The Logic of Scientific Discovery*, 2nd edition (London: Routledge, 2002).

18. Annie Franco, Neil Malhotra, and Gabor Simonovits, "Publication Bias in the Social Sciences: Unlocking the File Drawer," *Science* 345, no. 6203 (September 19, 2014), pp. 1502–505, https://doi.org/10.1126/science.1255484.

19. Ibid.

20. Richard Rudner, "The Scientist Qua Scientist Makes Value Judgments," *Philosophy of Science* 20, no. 1 (1953), pp. 1–6.

21. David Yokum, "Psychology, Open Science, and Government: The Opportunity," *APS Observer* 29, no. 4 (March 31, 2016), www.psychologicalscience.org/observer/psychology-open-science-and-government-the-opportunity.

22. You can view this at https://osf.io/hpmrt/.

23. Some authors have expressed a concern that *journals* might miss this nuance and drift into only publishing papers that have pre-specified null hypothesis tests, which in effect would chill exploratory research. Indeed, journals should not do that. But note this is a concern about the potential *misuse* of PAPs, not a critique on the value of PAPs rightly used.

RECOMMENDED RESOURCES

Aschwanden, Christie. "Science Isn't Broken." *FiveThirtyEight* (blog), August 19, 2015. https://fivethirtyeight.com/features/science-isnt-broken/. A journalist explains in plain language—and with interactive visualizations—the problems of *p*-hacking and publication bias.

Franco, Annie, Neil Malhotra, and Gabor Simonovits. "Publication Bias in the Social Sciences: Unlocking the File Drawer." *Science* 345, no. 6203 (September 19, 2014), pp. 1502–505. https://doi.org/10.1126/science.1255484.

An empirical investigation of how severe is the problem of publication bias.

Kerr, Norbert L. "HARKing: Hypothesizing after the Results Are Known." *Personality and Social Psychology Review* 2, no. 3 (August 1, 1998), pp. 196–217. https://doi.org/10.1207/s15327957pspr0203_4. Coined the term "HARKing" and explores how HARKing undermines scientific progress.

Nelson, Leif D., Joseph Simmons, and Uri Simonsohn. "Psychology's Renaissance." *Annual Review of Psychology* 69, no. 1 (January 4, 2018), pp. 511–34. https://doi.org/10.1146/annurev-psych-122216-011836. A review of risks to research integrity and how registered PAPs hedge those risks.