

The Case for Preregistering Quasi-Experimental Program and Policy Evaluations

Evaluation Review
2025, Vol. 0(0) 1–15
© The Author(s) 2025
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/0193841X251326738
journals.sagepub.com/home/erx



Thomas S. Dee¹ 

Abstract

The recognition that researcher discretion coupled with unconscious biases and motivated reasoning sometimes leads to false findings (“p-hacking”) led to the broad embrace of study preregistration and other open-science practices in experimental research. Paradoxically, the preregistration of quasi-experimental studies remains uncommon although such studies involve far more discretionary decisions and are the most prevalent approach to making causal claims in the social sciences. I discuss several forms of recent empirical evidence indicating that questionable research practices contribute to the comparative unreliability of quasi-experimental research and advocate for adopting the preregistration of such studies. The implementation of this recommendation would benefit from further consideration of key design details (e.g., how to balance data cleaning with credible preregistration) and a shift in research norms to allow for appropriately nuanced sensemaking across prespecified, confirmatory results and other exploratory findings.

Keywords

design and evaluation of programs and policies, quasi-experimental design < methodology (if appropriate)

¹Graduate School of Education, Stanford University, Stanford, CA, USA

Corresponding Author:

Thomas S. Dee, Graduate School of Education, Stanford University, 520 Galvez Mall, Stanford, CA 94305, USA.

Email: tdee@stanford.edu

Introduction

The precipitous declines in the confidence and trust placed in scientists and higher education, the growing distrust in expertise, and the growing climate of misinformation are arguably the most striking social developments in recent years. For example, surveys from the Pew Research Center ([Kennedy & Tyson, 2023](#)) indicate that the share of Americans who do not think that scientists act in the best interest of the public roughly doubled between 2019 and 2023 (i.e., from 13 to 27%). Similarly, Gallup surveys indicate that, between 2015 and 2024, the share of U.S. adults who have a “great deal” or “quite a lot” of confidence in higher education declined from 57% to 36% ([Jones, 2024](#)). While these recent trends appear closely related to growing partisanship in U.S. society, they also coincide with several high-profile incidents of scientific misconduct as well as a growing public awareness of “questionable research practices” in academic research.

However, concerns about the fundamental reliability of empirical studies in the social sciences are quite far from new. For example, [Spellman \(2015\)](#) organizes over a half century of critical commentary in psychology on issues such as questionable research practices, the poor replicability of social-science research, the biases against disseminating or publishing statistically insignificant findings, criticisms of null-hypothesis statistical testing (NHST), and the need for open data and study preregistration. Within economics, a widely discussed article by [Leamer \(1983\)](#) also promoted interest in research transparency and reproducibility “that later lost momentum and mostly died down” ([Christensen & Miguel, 2018](#)). That ebbing interest occurred despite the recognition that the use of higher-quality research practices has first-order implications for the reliability of the resulting findings. In particular, nearly four decades ago, Peter Rossi’s “stainless steel law of program evaluation” ([Rossi, 1987](#)) stated that better designed evaluations are more likely to find zero impact.

Much of the renewed and robust interest in research transparency or “open science” practices over the last decade can be traced to the “replication crisis” in psychology. [Romero \(2019\)](#) argues that this attention reflected both a growing awareness that several core findings in psychology often failed to replicate consistently and new evidence that the multiple forms of discretion readily available to researchers (e.g., searching among dependent variables and covariates and shaping the sample construction) can easily generate statistically significant findings for hypotheses that are not true ([John et al., 2012](#); [Simmons et al., 2011](#)). The impact of this awareness on research practices has arguably been substantial across multiple domains of social-scientific inquiry. In particular, the public preregistration of pre-analysis plans (henceforth, preregistration) has relatively quickly become the expected norm with respect to experimental studies in the social sciences ([Nosek et al., 2018](#)).

The rapid growth in the submissions to several online study registries attest to the substantial growth in preregistration among experimental studies. Other defining features of “open science” practices (e.g., requiring open data, study replication) have also expanded to some—but far from all—academic journals over this period (e.g., [Christensen & Miguel, 2018](#)).

While the evolving movement towards open science and research transparency has arguably had prominent and encouraging early success, I argue that its impact to date has actually been too narrow. In particular, the core thesis of this brief essay is that “open science” research practices have had far too little impact on the standards of practice among the many program and policy evaluations that seek to draw credible causal inferences from non-experimental or observational data (i.e., “quasi-experimental” studies). Others have made closely related observations. For example, [Hoces, Grant, and Miguel \(2021\)](#) note that “policy analysis has yet to systematically embrace transparency and reproducibility” and advocate for enhancing open access to relevant “outputs, analyses, and materials.” However, some open-science advocates have also suggested that preregistration would not be a convincing improvement to quasi-experimental research practices (e.g., [Christensen & Miguel, 2018](#)). I discuss such concerns and instead advocate for adapted preregistration procedures as a highly appealing and feasible way to bring research transparency to quasi-experimental studies.

Addressing the implied threat to the fundamental reliability of quasi-experimental studies through preregistration (and other open-science practices) is an imperative for several reasons. First, quasi-experimental studies are clearly a substantially more prevalent form of scientific inquiry than experiments. For example, [Brodeur et al. \(2020\)](#) collected data on the 684 quantitative studies that appeared in 25 leading economics journals published between 2015 and 2018. Within this recent sample, only 21% of the studies used experimental designs. Second, quasi-experimental studies allow us to learn about important policies and programs that either cannot or, in all likelihood, will not ever be evaluated in an experimental design. Quasi-experimental studies also focus on large-scale, real-world policy and program innovations that are not as easily amenable to replication as some experimentally manipulated contrasts. Third, quasi-experimental studies also have unique relevance in the production of useful knowledge because their external validity often compares favorably to those based on experimentally manipulated treatment contrasts implemented in contexts or with a fidelity that seldom exists in real-world settings. Similarly, the results from experiments where participants know the treatment has a finite term can be a poor guide to the impact of related policies that are credibly enduring. Finally, the increased distrust in evidence and expertise in contemporary society enhances the imperative to improve the credibility of quasi-experimental evaluations.

This essay proceeds in two broad parts. First, I discuss conceptual and empirical evidence that questionable research practices are uniquely common in quasi-experimental studies alongside evidence that quasi-experimental analyses have been comparatively unresponsive to the open-science movement. Second, I discuss potential solutions to this problem with a particular emphasis on an adapted form of preregistration as a new research standard for quasi-experimental studies. I also discuss potential criticisms of this recommendation. I conclude with thoughts on the practical challenges to adopting and implementing such changes in quasi-experimental research practices.

Evidence of Questionable Research Practices in Quasi-Experimental Studies

The most fundamental argument for the unique prevalence of questionable research practices in quasi-experimental studies is a purely conceptual one. Developing this point begins with noting that the core concern, which motivated the recent embrace of open-science practices, was a growing awareness that researchers make an extraordinary number of discretionary decisions (i.e., “researcher degrees of freedom”) that can easily lead to misleading results (Simmons et al., 2011). The multiplicity of important but discretionary design choices made by researchers includes the construction of the sample, the outcomes studied, the covariates used, the measurement of those key variables, and the exact approach to estimation and inference.

Why is this discretion problematic in the hands of expert researchers whose work is then subjected to rigorous rounds peer review? The concern is that researchers face high-powered incentives to engage tacitly in “fishing” or “p-hacking” in order to produce statistically significant results under NHST (i.e., $p < 0.05$). A recent study of “job market papers” from doctoral students in economics finds correlational evidence consistent with these strong incentives. Brodeur et al. (2024) examined 150 job-market papers and found that, conditional on other determinants, marginally significant quantitative findings were strongly related to subsequent academic placements.

Relatedly, Gelman and Loken (2013) underscore how questionable research practices can also be a stable equilibrium, stating “it’s easy to find a $p < .05$ comparison even if nothing is going on, if you look hard enough—and good scientists are skilled at looking hard enough and subsequently coming up with good stories (plausible even to themselves, as well as to their colleagues and peer reviewers) to back up any statistically significant comparisons they happen to come up with.” Simmons et al. (2011) stress that “this exploratory behavior is not the by-product of malicious intent” but cite evidence people are “self-serving in their interpretation of ambiguous information and

remarkably adept at reaching justifiable conclusions that mesh with their desires.”

In short, the argument is that critical researcher decisions can be unintentionally biased by motivated reasoning. Furthermore, hindsight bias (i.e., seeing outcomes as more predictable after they occurred) can reinforce researchers’ confidence in the self-serving results of these discretionary choices. Similarly, the practice of hypothesizing after the results are known or “HARKing” (Kerr, 1998) can also enhance an unjustified assurance in selected results.

Nosek et al. (2018) frame these research practices with respect to the epistemic distinction between prediction and postdiction. When conducting exploratory analyses based on a multiplicity of discretionary choices, researchers are effectively engaging in postdiction, an activity that can usefully inform hypothesis generation. However, in the absence of open science, such postdiction is often presented as hypothesis-testing prediction that is amenable to NHST. Notably, preregistration breaks this conflation by requiring researchers to make distinctions between “confirmatory” outcomes that are predicted *ex ante* and “exploratory” outcomes that do not reflect core predictions or are uncovered through subsequent postdiction.

The key insight in this context is that quasi-experimental studies have far more “researcher degrees of freedom” than experimental studies and, thus, more scope for p-hacked findings. The dire implications of enhanced researcher discretion have long been recognized. For example, Ioannidis (2005) states that “the greater the flexibility in designs, definitions, outcomes, and analytical modes in a scientific field, the less likely the research findings are to be true.” The comparative discretion available to quasi-experimental researchers was arguably true before the “preregistration revolution” (Nosek et al., 2018) simply because experimental designs place more restrictions on the treatment contrast under study, the estimation framework, and the availability of multiple measured outcomes. However, it is more emphatically true in the current climate where experimental studies are typically preregistered whereas the vast majority quasi-experimental analyses largely rely on what can be characterized as exploratory postdiction.

Ironically, several prominent study registries (e.g., the Open Science Framework, ClinicalTrials.gov) clearly accommodate quasi-experimental studies, with the notable exception of the one sponsored by the American Economic Association. However, the take-up of this option among researchers has been quite limited. To illustrate the comparative absence of preregistration for quasi-experimental studies, I organized the current data from a prominent education-focused study registry (Andersen et al., 2019) which accommodates quasi-experiments as well as experiments: the Registry of Efficacy and Effectiveness Studies (REES). At the time of my data collection (and after excluding retrospective registrations and a few studies with

unidentified designs), the REES repository included 606 research designs, some within the same study. Over 77% of the preregistered research designs relied on random assignment (i.e., 436 randomized control trials and 32 single-case designs). Only 21% of preregistered designs ($n = 129$) were quasi-experimental and only 9 preregistrations were for regression-discontinuity designs (RDD).

In other words, in REES, the preregistration of experimental designs outnumbers the preregistration of quasi-experimental designs by roughly 4 to 1. Notably, this likely overstates the broader preregistration of quasi-experimental designs because the U.S. Department of Education's Institute of Education Sciences (IES) funded the creation of the REES registry and encourages preregistration for the quasi-experimental studies it funds. Consistent with this bias, an examination of pre-analysis plans in economics and political science found that only 4% involved observational data (Ofosu & Posner, 2023). The stark underrepresentation of QED designs in study registries is particularly striking given that quasi-experimental studies are far more common than experimental ones. For example, the data collected by Brodeur et al. (2020) indicate that, in major economic journals, QEDs outnumber experimental designs by roughly 4 to 1. Similarly, an examination of working papers published by the National Bureau of Economic Research (NBER) found that non-experimental studies outnumbered experimental studies by nearly 9 to 1 (Ofosu & Posner, 2020).

Though the preregistration of quasi-experimental studies is quite uncommon, a possible rejoinder is that this static observation masks the ongoing, comparative growth of open-science practices among such studies. For example, Miguel (2021) discusses how the field of economics "is in a period of rapid transition toward new transparency norms in the areas of open data, preregistration and pre-analysis plans, and journal policies." However, this inference is based on a small sample of researchers. Furthermore, this documented growth in open-science practices (Miguel, 2021, Figure 1) largely reflects making data, code, and study instruments publicly available while the growth in study preregistration is much weaker (and presumably concentrated in experimental studies). To provide further evidence on this question, I organized counts of experimental and quasi-experimental preregistrations in REES by year since it began operations in 2018 (Figure 1). These data indicate that the preregistration of experimental designs grew rapidly relative to QED preregistrations and do not suggest any ongoing or more recent convergence.

In short, though the conceptual concerns that motivate preregistration (e.g., researcher discretion) are uniquely salient in quasi-experimental studies, the preregistration of such studies remains comparatively rare. Three other direct forms of empirical evidence indicate that, in the absence of preregistration, the current standards of practice in quasi-experimental research leads to potentially unreliable knowledge.

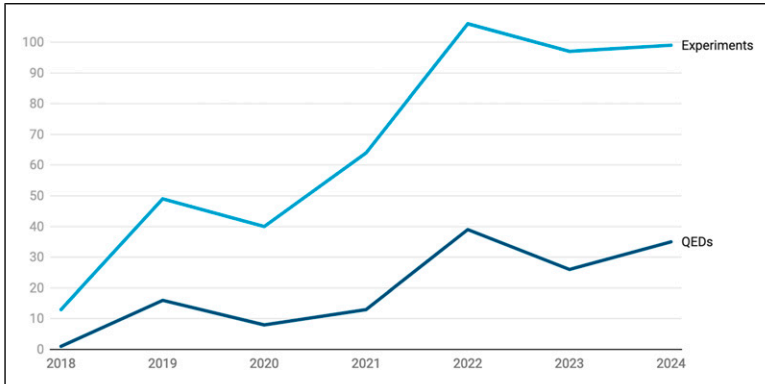


Figure 1. Preregistrations by research design. Source: Author Calculations Based on the Registry of Effectiveness and Efficacy Studies (REES).

First, a growing number of “many analyst” studies provide striking evidence on the empirical relevance of researcher discretion. The basic structure of these studies is to provide multiple research teams with a common data set and a shared research question and then to observe their subsequent analytical choices and conclusions. For example, in a seminal study by [Silberzahn and Uhlmann \(2015\)](#), 29 teams of researchers tested for referee bias based on skin tone in the receipt of red cards in soccer matches. The research teams in this study made diverse research-design choices (e.g., Bayesian clustering, logistic regression, linear modeling), which led to a variegated set of core findings. Specifically, roughly two-thirds of the research teams (i.e., 20 out of 29) reported statistically significant evidence of skin-tone bias but substantial variation in effect sizes. Notably, post-hoc discussion across the research teams did not build a consensus on a single, best approach.

In a larger and more recent study, [Brezna, Rinke, and Wuttke et al. \(2022\)](#) organized 73 research teams in using a shared data set to assess the same prominent hypothesis (i.e., that greater immigration reduces support for social policies among the public). Roughly 17% of the resulting point estimates indicated a positive and statistically significant effect while 25% indicated a statistically significant effect in the opposite (i.e., negative) direction. The remaining point estimates (i.e., about 58%) were not statistically significant. The authors argue that this strikingly diverse set of findings reflect, even for close observers, a “hidden universe of uncertainty.” Specifically, the authors qualitatively coded the identifiable decisions in each research team’s workflow. They found that these observed choices left 95% of the total variation in the point estimates unexplained.

Recent evidence on the distribution of test statistics across different study designs provide a second, critical form of evidence on the biasing effects of the

broad discretion that is uniquely available to quasi-experimental researchers. In particular, [Brodeur et al. \(2020\)](#) collected nearly 22,000 test statistics from quantitative studies that appeared in the top 25 economic journals between 2015 and 2018. They examined the distributions of these test statistics across four different research designs: difference-in-differences (DID), instrumental variables (IV), regression discontinuity (RD), and randomized control trials (RCT). They found that the test-statistic distributions for two widely used quasi-experimental designs (i.e., DID and IV) uniquely indicated missingness at values just before the conventional significance threshold (i.e., $z = 1.65$) and sharp surplus of test statistics at just higher values. They find no evidence that the prevalence of these misallocated test statistics is lower in more selective journals, that they decline through revisions, or that they are improving over time.

A third, related source of evidence focuses on the statistical power across experimental and quasi-experimental studies and corresponding estimates for the excess of statistically significant findings. The focus on statistical power reflects the concern that, because underpowered studies are less likely to generate statistically significant findings, they can increase the incentives for questionable research practices that select for spuriously significant (and publishable) findings. Recent evidence indicates this incentive for questionable research practices is highly concentrated in studies based on observational data. For example, [Askarov et al. \(2023\)](#) find that, in 31 leading economics journals, the median power of experimental studies (i.e., 78%, $n = 699$) is close to the 80% standard. However, the median power of quasi-experimental studies in the same journals (i.e., 7%, $n = 23,238$) is, strikingly, less than a 10th of this value. They also estimate that the excess of statistically significant findings in these journals is twice as large for observational studies (i.e., 19.1%) as for experimental studies (i.e., 9.7%).

The metascientific empirical evidence based on published results clearly suggests research-transparency problems unique to quasi-experimental studies. These patterns could reflect publication biases as well as p-hacking. Publication biases provide another important motivation for preregistration (i.e., addressing the “file drawer” problem). However, other recent evidence suggests the unique salience of p-hacking relative to publication biases. Specifically, [Brodeur et al. \(2024\)](#) find that the initial submissions to a journal display a suspicious pattern of missing and then bunching at levels of statistical significance. They conclude that the “peer review process has little effect on the distribution of test statistics.”

Assessing Open-Science Practices for Quasi-Experimental Research

Both conceptual reasoning and a growing body of empirical evidence indicate that questionable research practices are uniquely prevalent in quasi-

experimental studies relative to experimental studies. This fundamental threat to the epistemic validity of quasi-experimental research exists despite editorial oversight, expert peer review, and the recent growth in some open-science practices other than study preregistration (e.g., sharing data and code).

However, the encouraging evidence from some other open-science practices merits note and suggest the promise of further adoption and evaluation (Nosek et al., 2015). For example, editorial statements from eight health-economic journals that underscored the problem of p-hacking and the potential importance of statistically insignificant findings appear to have reduced the prevalence of tests rejecting the null hypothesis by 18 percentage points (Blanco-Perez & Brodeur, 2020). Similarly, a difference-in-differences study of 24 journals found that the adoption of a data-sharing policy reduced the reporting of significant results and the magnitude of the corresponding test statistics (Askarov et al., 2023).

However, there is also evidence that suggests the limited relevance of data-sharing and replication studies. For example, Schafmeister (2021) examined a large-scale replication effort in psychology and found no evidence that supportive or non-supportive replications influenced the citation patterns of the original article. Serra-Garcia and Gneezy (2021) found that social-science publications that fail to replicate are actually cited more than those that do. Moreover, this difference in citations is unresponsive to the publication of a replication failure. Indeed, only 12% of the post-replication citations of the original article even acknowledge the replication failure. The authors posit that, when a result is “interesting” readers apply lower standards regarding the relevance of replications.

Given this context, why hasn’t study preregistration—the most popular open-science practice—been as broadly adopted in quasi-experimental studies as they have been in experimental research? The most common stated objection is what can be called the verifiability critique. The argument is that, under the preregistration of quasi-experimental studies, malevolent researchers could secretly examine existing observational data and the findings based on different decisions before they file—and falsely attest to—an analysis plan with their preferred results. For example, Christensen and Miguel (2018) argue that with accessible, observational data “there is often no credible way to verify that preregistration took place before analysis was completed” and conclude “proponents of the preregistration of observational work have not formulated a convincing response to this obvious concern.” Similarly, Gelman and Loken (2013) claim “it would be close to meaningless to consider preregistration for data with which we are already so familiar.” Burlig (2018) argues that, due to this concern, the preregistration of quasi-experimental studies should be limited to three narrow cases where such verifiability is feasible. These occur in applications where the real-world event of interest has not yet occurred, where the relevant data has not yet been

collected, and, third, where the data can only be accessed through documented restricted-use procedures.

I believe the verifiability critique, while important, does not justify a wholesale dismissal of preregistration for quasi-experimental studies. First and foremost, it is based on tacit assumptions about the underlying character of questionable research practices that are not easily tenable. As noted above, metascientific researchers (e.g., [Simmons et al., 2011](#)) argue that questionable research practices are only rarely due to outright fraud and instead reflect the interactions of researcher discretion, unconscious biases, and motivated reasoning in the face of professional incentives to report interesting and statistically significant findings. Preregistration is well suited to addressing the pernicious effects this implies for the vast majority of quasi-experimental researchers and reported research. And precluding this change in practice because of a few high-profile “edge” cases effectively lets perfection become the enemy of meaningful improvement.

There are also logical inconsistencies implied by the verifiability critique’s argument that the possibility of intentional researcher fraud diminishes the appeal of preregistering studies based on pre-existing data. For example, this broad dismissal of quasi-experimental preregistration holds this important form of inquiry to a standard that experimental studies do not clearly meet. That is, if the underlying behavioral issue is outright researcher fraud, some preregistered experimental studies should also be viewed with increased suspicion given the possibility of misrepresentations (e.g., postdating an analysis to a date after a “preregistration”). More generally, if researcher fraud argues against adopting quasi-experimental preregistration, it can hardly be viewed as a compelling endorsement for the current status quo where such research is rarely preregistered. Instead, the logical implication of this reasoning is an extreme one: that quasi-experimental research—the overwhelming majority of social-science research making causal claims—is not and cannot be reliable.

I instead view the verifiability critique as having important implications for the sensible design and implementation of quasi-experimental preregistration. That is, a reasonable concern is that the researcher discretion, biases, and incentives that motivate preregistration can unintentionally seep into and corrupt the preregistration process. This is particularly so in quasi-experimental settings where concerns about data availability, quality, and missingness imply that researchers need to access and examine the relevant data before they can confidently prespecify key measures in much detail.

There are multiple ways in which possible adaptations to preregistration norms could address this concern in quasi-experimental applications. One approach would be to ask quasi-experimental researchers to preregister prior to accessing any data whatsoever. Then, understanding that the initial preregistration would be prospective, to encourage preregistration amendments

that clarified their key design decisions (e.g., identifying subsequent changes to key measures based on new information about the quality and character of the available data).

Alternatively, quasi-experimental preregistration procedures could ask researchers to create an intentional “firewall” between the data cleaning that precedes preregistration and their subsequent impact analysis. An explicit guideline (and attestation) in preregistration forms could operationalize this norm. Similarly, it would also be possible to task one subset of a research team, prior to preregistration, to clean and organize study data but to do so without clear access to data on treatment status. A research team could also prespecify an analysis plan based on the interrogation of simulated data prior to accessing the actual data and conducting that prespecified analysis.

A potentially important and practical concern about quasi-experimental preregistration sometimes voiced by researchers is that its strictures would seriously inhibit useful discoveries. Interestingly, this criticism also occurred over a decade ago in the context of early discussions about adopting preregistration for experimental studies (e.g., [Spellman, 2015](#)). There is little to indicate that those concerns have been borne out in the context of experimental research. Nor is there clear reason to believe that extending preregistration norms to quasi-experimental research would be any different. Preregistration does not appear to have limited discovery but instead has promoted clarity about those findings that are based on true predictions and those based on exploratory postdiction.

These clear distinctions between confirmatory and exploratory findings can do more than support the epistemic validity of a given study. They can also complement the ways in which researchers construct knowledge from a larger body of research. More specifically, researchers’ summative sense-making across multiple studies often turns on a kind of Bayesian updating based on multiple traits of those studies. That is, the extent to which a given study has compelling internal validity, a stronger claim to generalizability, or better measurement informs its contribution to the overall construction of knowledge. Preregistration only enhances these existing processes for constructing research consensus by providing useful transparency about whether a finding was predicted or based on exploratory postdiction. Nothing in this process precludes identifying and reporting findings that are exploratory in nature.

Conclusion

Quasi-experimental research is centrally relevant to empirical social science. It is by a large margin the most common form of research that makes causal claims. It allows us to make such inferences in real-world settings that can offer a generalizability often unavailable in experimental studies. It also allows us to examine the effects of programs, policies, and behaviors that simply

cannot or will not be subject to designed variation. However, questionable research practices constitute a substantial and currently unaddressed threat to reliability of this important form of intellectual inquiry.

In experimental research, it is now widely recognized that researcher discretion coupled with unconscious biases and powerful professional incentives to produce interesting and statistically significant results can lead to false findings (i.e., p-hacking). Over the last decade, that recognition has led experimentalists to broadly embrace open-science practices, especially pre-registration. However, the preregistration of quasi-experimental studies is, to date, surprisingly uncommon.

In this essay, I have argued that preregistration should be adapted and broadly adopted in quasi-experimental research. A core motivation is that the researcher discretion that motivated the preregistration of experimental studies is far more extensive in quasi-experimental applications. Furthermore, multiple forms of empirical evidence based on quasi-experimental studies—the evidence from “many analyst” studies, the distribution of test statistics, the comparatively low power and excess statistical significance—affirm the practical relevance of this straightforward conceptual insight.

However, the distinctive character of the workflow in quasi-experimental research implies that the optimal design details of such preregistrations will differ from experimental versions in ways that merit careful consideration. A particularly relevant issue involves how to navigate the tradeoffs between preregistrations that occur before or after data cleaning. Preregistrations filed prior to data cleaning establish a relatively clear demarcation between a researcher’s original scientific predictions and their subsequent analysis and findings. However, because the data used in quasi-experimental studies are sometimes quite messy (e.g., due to missingness, uncertain quality, etc.), this approach to preregistration would need either to be comparatively vague or to countenance subsequent revisions to the original analysis plan. Preregistrations filed after data cleaning can be more specific about key design details (i.e., the choice of confirmatory outcomes and their construction) and are less likely to require subsequent amendments. However, absent other changes in research procedures (i.e., “firewalls” that separate data cleaning from impact analyses), this approach could still allow for unintended researcher discretion. An additional but related issue is whether quasi-experimental researchers should preregister a single research design or a decision tree of research-design choices and robustness checks.

Despite the need to establish new standards of practice on these and possibly other issues, the transition to preregistering quasi-experimental studies is not only feasible but likely to be quite tractable. This is especially so if funders, journals, and academic societies provide more leadership on this recommended transition. Furthermore, as a practical matter, several prominent study registries already accommodate quasi-experimental studies

explicitly. The recent and successful transition to preregistration in experimental research also supports a considerable degree of optimism. The leading examples from the relatively few quasi-experimental studies that have preregistered (e.g., Bonilla et al., 2021; Dee & Pyne, 2022) also provide encouraging proof points.

This recommendation may also improve quasi-experimental studies by amplifying the considered role of theory and measurement in such research. That is, because preregistration compels a clearer distinction between prediction and exploratory postdiction, it strongly encourages researchers to think more deeply about their behavioral settings, their theoretical predictions, and their corresponding measures before they start generating results. More generally, quasi-experimental preregistrations can bring true transparency and credibility, which are currently lacking in this important form of scientific inquiry.

Acknowledgments

This essay is based on a talk given as a 2024 Recipient of the Peter H. Rossi Award for Contributions to the Theory or Practice of Program Evaluation at the 2024 meeting of the Association for Public Policy Analysis and Management (APPAM). I would like to thank Jeff Smith, Rebecca Maynard, Douglas Besharov, and Eugene Bardach, and participants at the APPAM conference for useful comments.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

ORCID iD

Thomas S. Dee  <https://orcid.org/0000-0001-7524-768X>

References

- Anderson, D., Spybrook, J., & Maynard, R. (2019). REES: A registry of efficacy and effectiveness studies in education. *Educational Researcher*, 48(1), 45–50. <https://doi.org/10.3102/0013189X18810513>
- Askarov, Z., Doucouliagos, A., Doucouliagos, H., & Stanley, T. D. (2023). The significance of data-sharing policy. *Journal of the European Economic Association*, 21(3), 1191–1226.

- Blanco-Perez, C., & Brodeur, A. (2020). Publication bias and editorial statement on negative findings. *The Economic Journal*, 130(629), 1226–1247.
- Bonilla, S., Dee, T. S., & Penner, E. K. (2021). Ethnic studies increases longer-run academic engagement and attainment. *Proceedings of the National Academy of Sciences*, 118(37), 1.
- Brezna, N., Rinke, E. M., Wuttke, A., Nguyen, H. H., Adem, M., Adriaans, J., & Van Assche, J. (2022). Observing many researchers using the same data and hypothesis reveals a hidden universe of uncertainty. *Proceedings of the National Academy of Sciences*, 119(44), e2203150119.
- Brodeur, A., Cook, N., & Heyes, A. (2020). Methods matter: P-Hacking and publication bias in causal analysis in economics. *The American Economic Review*, 110(11), 3634–3660.
- Brodeur, A., Kattan, L., & Musumeci, M. (2024). *Job market stars (No. 1514)*. GLO discussion paper.
- Burlig, F. (2018). Improving transparency in observational social science research: A pre-analysis plan approach. *Economics Letters*, 168(1), 56–60.
- Christensen, G., & Miguel, E. (2018). Transparency, reproducibility, and the credibility of economics research. *Journal of Economic Literature*, 56(3), 920–980.
- Dee, T., & Pyne, J. (2022). A community response approach to mental health and substance abuse crises reduced crime. *Science Advances*, 8(23), 1.
- Gelman, A., & Loken, E. (2013). 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(1-17), 3.
- Hoces de la Guardia, F., Grant, S., & Miguel, E. (2021). A framework for open policy analysis. *Science and Public Policy*, 48(2), 154–163.
- Ioannidis, J. P. (2005). Why most published research findings are false. *PLoS Medicine*, 2(8), Article e124.
- John, L. K., Loewenstein, G., & Prelec, D. (2012). Measuring the prevalence of questionable research practices with incentives for truth telling. *Psychological Science*, 23(5), 524–532.
- Jones, J. (2024). *US confidence in higher education now closely divided*. Gallup. Available at: <https://news.gallup.com/poll/646880/confidence-higher-education-closely-divided.aspx>
- Kennedy, B., & Tyson, A. (2023). *Americans’ trust in science, positive views of science continue to decline*. Pew Research Center. Available at: <https://www.pewresearch.org/science/2023/11/14/confidence-in-scientists-medical-scientists-and-other-groups-and-institutions-in-society/>
- Kerr, N. L. (1998). HARKing: Hypothesizing after the results are known. *Personality and Social Psychology Review*, 2(3), 196–217.
- Leamer, E. E. (1983). Let’s take the con out of econometrics. *The American Economic Review*, 73(1), 31–43.

- Miguel, E. (2021). Evidence on research transparency in economics. *The Journal of Economic Perspectives*, 35(3), 193–214.
- Nosek, B. A., Alter, G., Banks, G. C., Borsboom, D., Bowman, S. D., Breckler, S. J., & Yarkoni, T. (2015). Promoting an open research culture. *Science*, 348(6242), 1422–1425.
- Nosek, B. A., Ebersole, C. R., DeHaven, A. C., & Mellor, D. T. (2018). The pre-registration revolution. *Proceedings of the National Academy of Sciences*, 115(11), 2600–2606.
- Ofori, G. K., & Posner, D. N. (2020). Do pre-analysis plans hamper publication? In *AEA Papers and Proceedings*. (110, pp. 70–74). American Economic Association. 2014 Broadway, Suite 305.
- Ofori, G. K., & Posner, D. N. (2023). Pre-analysis plans: An early stocktaking. *Perspectives on Politics*, 21(1), 174–190.
- Romero, F. (2019). Philosophy of science and the replicability crisis. *Philosophy Compass*, 14(1), e12633. <https://doi.org/10.1111/phc3.12633>
- Rossi, P. (1987). The iron law of evaluation and other metallic rules. *Research in Social Problems and Public Policy*, 4(1), 3–20.
- Schafmeister, F. (2021). The effect of replications on citation patterns: Evidence from a large-scale reproducibility project. *Psychological Science*, 32(10), 1537–1548. <https://doi.org/10.1177/09567976211005767>
- Serra-Garcia, M., & Gneezy, U. (2021). Nonreplicable publications are cited more than replicable ones. *Science Advances*, 7(21), eabd1705. <https://doi.org/10.1126/sciadv.abd1705>
- Silberzahn, R., & Uhlmann, E. L. (2015). Crowdsourced research: Many hands make tight work. *Nature*, 526(7572), 189–191. <https://doi.org/10.1038/526189a>
- Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological Science*, 22(11), 1359–1366. <https://doi.org/10.1177/0956797611417632>
- Spellman, B. A. (2015). A short (personal) future history of revolution 2.0. *Perspectives on Psychological Science*, 10(6), 886–899. <https://doi.org/10.1177/1745691615609918>