Research Registries and the Credibility Crisis: An Empirical and Theoretical Investigation

Eliot Abrams* Jonathan Libgober† John A. List‡

October 22, 2025

Abstract: We analyze one prominent policy solution to the credibility crisis in experimental research—research registries—with primary focus on the AEA RCT Registry. We find that the AEA RCT Registry has had a limited impact to date. A majority of recent economics experiments are not registered, only about half of those registered did so prior to intervention, and most of these preregistrations lack sufficient detail. We find broad similarities when comparing these patterns to ClinicalTrials.gov. We then advance an economic model of registration to explore potential improvements to registries generally. The model shows banning late registration can significantly increase registry effectiveness.

Keywords: Research Registries; Randomised Controlled Trials; Publication Bias
JEL: B41; C9; C93; D83

*Stripe, eabrams@uchicago.edu
†University of Southern California, corresponding author, libgober@usc.edu
‡University of Chicago and Australian National University, jlist@uchicago.edu

This paper is a revision of "Research Registries: Myths, Facts, and Possible Improvements." Natasha Denisyuk, David Franks, Nicoló Generoso, Xinyi Hong, Jiwon Hwang, Jonathan Lambrinos, Ariel Listo and Sai Zhang, and especially Franco Daniel Albino, provided excellent research assistance. We are also grateful to the numerous other research assistants whose dedicated work in assembling the data made this project possible. Discussant Matthias Greiff gave useful feedback, and discussions with Vittorio Bassi, Eszter Czibor, Stefano DellaVigna, Rachel Glennerster, Michael Kremer, Min Sok Lee, Shengwu Li, Ulrike Malmendier, Torben Mideksa, Gautum Rao, Sutanuka Roy, and Uri Simonsohn significantly improved the paper. We also thank participants of the ASSA Virtual Winter Meetings, the Advances in Field Experiments Conference, and the UChicago Experimental Economics Seminar for their feedback and the Becker Friedman Institute for financial support.

1 Introduction

The last several decades have brought significant change to the empirical landscape in economics. New approaches to generating data in the lab and field have opened up several unique lines of research into the "whys" behind observed behaviors.[1] These experimental approaches have helped to clarify identification, inference, and interpretability. However, critics in the broader social sciences have recently called for the experimental movement to proceed more cautiously. An active debate has emerged over claims that experiments face a "credibility crisis."[2] This charge follows from the fact that data are ultimately finite, so that researchers must choose which hypotheses to test, report, and trumpet in a system where publication incentives imply that not all results are equally likely to get published. Economists, along with researchers in other empirical disciplines, have recognised that these limitations could lead to a departure from socially optimal experiment conduct.

This paper conducts an empirical and theoretical examination of one of the most significant policy prescriptions that advocates have proposed to improve the credibility crisis—the establishment of research registries for randomised controlled trials (RCTs), focusing primarily (but not exclusively) on economics. These registries provide a venue for researchers to document their experiment setup (notably, including sample size), execution, hypotheses and results in a site that is searchable by external audiences. In principle, if used appropriately, research registries can tackle key issues in the credibility crisis.[3] We examine the extent to which research registries address two concrete issues that have received particular attention and, to the best of our knowledge, form the primary motivation for the establishment of registries in the first place:[4]

[1] See Harrison and List (2004) for discussion of the potential insights provided by both field and lab experiments.

[2] See Jennions and Møller (2003), Ioannidis (2005), Nosek, Spies and Motyl (2012), Bettis (2012), Maniadis, Tufano and List (2014), and Dreber et al. (2015) for discussions of the extent of the credibility crisis.

[3] We acknowledge that the credibility crisis applies to empirical research more broadly and goes back to at least Edward Leamer, who famously advocated taking the "con" out of econometrics in Leamer (1983). However, discussions of the crisis and policy prescriptions (including research registries) tend to focus on RCTs. We believe that this is because RCTs are seen as low hanging fruit—each RCT is ostensibly designed to test a small set of interventions and has an explicit start and end date. One notable exception is the Open Science Framework (OSF) Registries Network. The OSF advocates for open collaboration in science research and their registries network permits the registration of observational studies. Other web services, such as AsPredicted, also facilitate recording any research hypothesis. However, unlike research registries, AsPredicted does not provide a way to search the recorded hypotheses. We study AsPredicted in Section 4.2. See Burlig (2018) for a discussion of registration of observational studies.

[4] We focus on these two issues due to their concreteness and since they are, as far as we are aware, the most

- *The file drawer problem*, namely that many studies are never made public, and so relegated to the proverbial "file drawer."
- *Scope for p-hacking*, namely that researchers often make adaptive data analysis decisions in the pursuit of results that are statistically significant at conventional levels.

A registry can address the file drawer problem for RCTs to the extent that researchers record all RCTs started and their outcomes. A registry can address p-hacking in RCTs to the extent that researchers document their initial experimental design and analysis plan along with changes to these over time in their registrations.

We focus our examination on the American Economic Association's registry for randomised controlled trials (the AEA RCT Registry) and utilise the ClinicalTrials.gov medical research registry as a benchmark. Launched in 2013, the AEA RCT Registry is the most commonly used registration database in economics (see Section 4.2). The AEA RCT Registry lists 9,923 studies across over 139 countries as of February 4, 2025. ClinicalTrials.gov is maintained by the National Institutes of Health and is the largest research registry overall. It contains 525,007 trial registrations from over 227 countries as of February 4, 2025. An existing literature (reviewed in Appendix D) has assessed the mixed effectiveness of ClinicalTrials.gov. To the best of our knowledge, we are the first to provide a systematic assessment of the AEA RCT Registry.

We make two specific contributions to the literature on policy prescriptions for the credibility crisis. First, we empirically evaluate the extent to which the AEA RCT Registry has been effective at solving the file drawer and p-hacking problems. Second, we advance a model of registration that suggests alternative registry designs that could improve registry effectiveness broadly. Our theoretical analysis focuses on one concrete design issue, namely that both the AEA RCT Registry and ClinicalTrials.gov accommodate *late registration*. While typical motivations for promoting registration rely upon the assumption that it is done prior to the start of the experiment intervention,[5] the AEA RCT Registry and ClinicalTrials.gov permit the registration of completed trials.[6]

salient issues registries are designed to address. While other issues are certainly interesting (such as transparency and hypothesis selection broadly defined), we leave empirical examinations of these to future work. See Christensen and Miguel (2018) for a notable discussion of transparency in economics research.

[5]For instance, because researchers may be more likely to "relegate an experimental finding to the file drawer" if the results are negative.

[6]The AEA RCT Registry chose to allow late registration primarily to facilitate the registration of RCTs that started

To understand the extent to which the AEA RCT Registry mitigates the file drawer problem, we perform a census of papers conducting RCTs published in leading economics journals. While the full universe of started experiments is unobserved, the extent to which the registry covers this known sample of prominent experiments serves as an informal upper bound on its coverage of all experiments. This exercise reveals that registration has become more widespread over time, but is far from universal. As we explain in Appendix C, the AEA RCT Registry is primarily targeted at the registration of field experiments, making it more difficult to interpret any success or failure of the registry in addressing the file drawer problem for lab experiments. Correspondingly, we focus the analysis of the file drawer problem on field experiments to prevent this distinction from interfering with the interpretation of our conclusions. Including data from lab experiments strengthens our insights. We find that 62% of the field experiments[7] published in top economics journals between 2017 and the end of 2023 are registered. We find that approximately 10% of lab experiments published in top economics journals between 2017 and the end of 2023 are registered, suggesting a significant gap in norms regarding registration across different fields.

We next examine the extent to which the AEA RCT Registry mitigates the p-hacking problem for economics experiments. Registrations can reduce p-hacking to the extent that they occur before the intervention begins (i.e., are preregistrations); sharply specify their primary outcomes; and match the resulting published or working papers. To allow time for researchers to learn about the registry's existence, we examine the subset of trials that report an intervention start date on or after January 1, 2014.[8] Of these trials, only 52% registered before their intervention began. We then randomly select 1,000 preregistrations from the AEA RCT Registry, and instruct a set of RAs to (i) assess the specificity of the primary outcomes reported by each preregistration; (ii)

prior to the registry's establishment. However, our understanding is that there is no plan to revisit this design choice now. ClinicalTrials.gov generally allows late registration although several categories of medical experiments are required to preregister by law. As far as we know, no existing laws require either the registration or preregistration of economics experiments.

[7]Throughout the paper, we refer to any study that uses randomisation to assign treatment as an RCT. Our definition of field experiments follows Harrison and List (2004), with the salient feature being that treatment and control units are observed in the setting of interest rather than a controlled environment. As we explain in Appendix C, the AEA RCT Registry is primarily targeted at the registration of field experiments, making it more difficult to interpret any success or failure of the registry in addressing the file drawer problem for lab experiments. Correspondingly, we focus the analysis of the file drawer problem on field experiments to prevent this distinction from interfering with the interpretation of our conclusions. Including data from lab experiments strengthens our insights.

[8]The registry became widely known after David McKenzie's October 14, 2013 World Bank Development Impact blog post.

identify the latest working or published paper associated with each preregistered RCT; and (iii) compare the outcomes reported in the paper to the preregistered primary outcomes. We find that the preregistrations leave significant latitude. Even the most detailed primary outcomes generally fail to provide a specific variable construction or measurement timeframe. As we discuss in more detail below, these primary outcomes are similar to "number of fruits each experimental subject consumes" rather than to "number of apples each experimental subject consumes in March, 2024." That said, we find that published and working papers do generally match their preregistration. In the average paper, 88% of the primary outcomes are consistent with their preregistered construction. As part of this analysis, we also examine whether preregistered studies generally maintain the sample size described in their registration. We find that such departures are frequent (in more than half of all registrations) and often large, approximately 31.4% of the time consisting of a deviation by more than 25% of the registered sample size.[9] These results are troubling to the extent that potentially endogenous choices by researchers about when to stop collecting data are not taken into account when testing statistical significance.

To statistically assess the impact of the registry on p-hacking, we randomly selected 200 published papers with differing registration statuses from the population of papers assessed in the file drawer exercise. We assigned RAs to collect the p-values for the primary outcomes for each paper. We then applied a battery of tests for p-hacking which have been previously proposed in the literature.[10] Overall, this analysis shows that the sample of published registered RCTs and the sample of published unregistered RCTs provide similar evidence for p-hacking. In both samples, a discontinuity test around the significance threshold strongly rejects the null hypothesis of no p-hacking and the remaining tests fail to reject the null hypothesis with comparable levels of confidence.

In sum, our empirical analyses suggest that the AEA RCT Registry does not yet sufficiently address either the file drawer or p-hacking problems. A theme that emerges from our analysis is that in economics, the social norm of registration is limited. Many trials fail to register and those that

[9]Overall, when a deviation occurs, we find that it is toward using less data rather than more. For deviations of at least 25%, roughly equal fractions are larger than registered compared to smaller.

[10]One comment is that the tests we employ only test for *marginal* p-hacking—in other words, p-hacking which only occurs nearby the significance threshold. Tests for non-marginal p-hacking (i.e., p-hacking that occurs well beyond this range) is generally more challenging and left to future work.

do register often do not provide the detail necessary for an appropriate examination of the integrity of their experimental design and data analysis plan. Insofar as formal registration requirements are fairly weak (which was arguably a deliberate choice in order to encourage participation and help *establish* a norm for registration), this unfortunately implies that the impact of the registry on credibility is fairly weak as well.

Assessments of ClinicalTrials.gov provide a useful benchmark for our results on the AEA RCT Registry. The former focuses on medical trials, in contrast to the latter's focus on economics. We extend the existing literature on ClinicalTrials.gov by examining the restrictiveness and fidelity of primary outcomes reported by 300 randomly chosen preregistrations from the first five years of ClinicalTrials.gov.[11] We find that the ClinicalTrials.gov preregistrations are only slightly more restrictive than the AEA RCT Registry preregistrations. We also find that papers associated with the ClinicalTrials.gov preregistrations and the AEA RCT Registry preregistrations have similar fidelity to the registered primary outcomes. This result, combined with the literature review in Appendix D, suggests that there is little reason to be optimistic that existing research registries will significantly impact the credibility crisis in economics on the current trajectory.

In an effort to improve research registry designs, we construct a simple model of registration. This model speaks to registration design generally, not only within economics, but our discussion focuses on its implications for the AEA RCT Registry. The model is a novel dynamic signalling game which may be of independent interest beyond our particular application. Specifically, we consider a researcher endowed with an experiment on an underlying hypothesis whose payoffs improve as an "outsider" becomes more optimistic that the researcher's hypothesis is true. The researcher first chooses whether to preregister and conduct the experiment. The researcher then chooses whether to register late. Finally, the researcher receives a payoff based on the outsider's updated belief about the underlying hypothesis after seeing the registration decision and the experiment outcome. Preregistration allows researchers to signal confidence in their hypotheses, for instance from strong intuition based on prior work or domain expertise. But late registration is tempting due to option value—there is a chance that registration is not worth it ex-post since there are

[11] We focus on the first five years of ClinicalTrials.gov to provide a reasonable comparison to the launch of the AEA RCT Registry.

costs associated with registration.[12] We acknowledge that this simple model abstracts away from several important factors—most notably, the potential for registration to shape the experimentation process itself—in order to focus on the endogenous factors that influence a researcher's registration decision. For instance, registration may act as a commitment device or "moral compass," helping researchers avoid actions that would ultimately reduce the informativeness of the resulting experiment. However, such considerations would imply registered experiments are exogenous more informative, an assumption we seek to avoid making a priori.

We use our model to explore policy counterfactuals focusing the discussion on the consideration of a late registration ban. First, we identify plausible conditions under which a ban on late registration increases the *total* number of registrations (directly improving registry effectiveness against the file drawer problem and potentially improving effectiveness against p-hacking via increasing preregistrations). One might find it natural to conjecture that allowing late registration would only increase the number of registered experiments by providing researchers more opportunities to register. However, we highlight that the option value associated with late registration gives researchers on the margin of registering an added incentive to delay their registration decision. Therefore, banning late registration always increases registration rates for the marginal experiment. And, since not all researchers who delay their registration decision will find it worthwhile to ultimately register, this effect can be sufficiently strong to overturn the natural conjecture.

We use a calibration exercise to argue that this insight is empirically relevant for the AEA RCT Registry. Generally, the comparison between registration rates with and without a late registration ban is ambiguous due to the competing effects identified in the previous paragraph. Our calibration exercise provides some suggestions about which way this may resolve in practice. Under parameter values that match historical registration rates, we show that banning late registration strictly increases total registrations for the AEA RCT Registry. Altering model parameters to match current registration rates (as explained in more detail in Section 6.3), we find even stronger support for this conclusion.

So where do we go from here? Our recommendation is to prohibit late registration while simultaneously providing incentives for researchers to preregister their work such as mandating preregistration as a condition for publication. We note that incentives (particularly mandates) are

[12] Many of the costs outlined by Olken (2015) regarding pre-analysis plans apply to registration as well.

costly to implement, and that greater enforcement and clarity related to existing mandates (e.g., what counts as a lab versus field experiment for the AEA publishing criteria) is one area of low-hanging fruit. That said, to the extent that the ultimate goal of a research registry is to mitigate the file drawer problem and p-hacking, this dual approach can move us in that direction.

The remainder of our paper is organised as follows. Section 2 summarises related literature. Section 3 presents our empirical assessment of the extent to which the AEA RCT Registry is currently solving the file drawer and p-hacking problems. Section 4 compares the AEA RCT Registry to ClinicalTrials.gov and discusses other registration venues as well. Section 5 presents our model of a researcher's registration decision. Section 6 reports our calibration exercise. Section 7 considers key model extensions. Section 8 concludes. We highlight that Appendix C contains background information on the AEA RCT Registry relevant for our analysis and that Appendix D surveys past work on ClinicalTrials.gov. All tables, figures, and proofs are in the respective appendices.

2 Literature Review

Our contributions are both empirical and theoretical in nature. On the former, we contribute to a growing literature that seeks to assess the credibility of the research process.[13] Imai et al. (2025) document the rapid growth of preregistration in economics, presenting survey evidence suggesting sharp disagreement on the proper scope of registration. They interpret this finding as suggesting the need for sharper guidelines from professional organisations, a point further underscored by our results on the latitude left open by many registrations. Imai et al. (2025) also find that researchers view preregistered tests as more credible; our model proposes one explanation for this perception and analyzes the implications of this mechanism. Focusing on economics research, Brodeur et al. (2016) provide evidence that published studies tend to inflate their p-values. In a similar spirit, Vivalt (2018) and Brodeur, Cook and Heyes (2020) show that certain identification strategies may be more susceptible to p-hacking. Chassang and Kapon (2023) discuss a number of strategies

[13] A related line of work explores a more decentralised approach to alleviate the crisis of confidence: using, and incentivising, a greater number of replications (see, e.g., Butera et al. (2020); Dreber et al. (2015) for examples).

which may facilitate external validity, with registration among them. Asri, Imai and Leight (2024) link registrations from the first few years of the AEA RCT Registry to research output, finding that 86% of their sample has a corresponding working paper or publication. Examining the content of the research outputs, they document a null-results penalty at top-five journals, but also suggests that a higher share of null results in a paper's abstract increases the overall likelihood of journal publication. Chopra et al. (2024) provide further evidence of a null results penalty in economics, based on an experiment with 500 researchers in economics departments in which study characteristics were varied exogenously.

In independent and contemporaneous work, Brodeur et al. (2024) consider the impact of preregistration and pre-analysis plans on p-hacking and publication bias, ultimately concluding that the latter enhances credibility but the former does not. Our results suggest, like theirs, that pre-registrations appear to have limited impact. While we do not consider pre-analysis plans, our work on p-hacking and publication bias complements theirs by modifying the analysis to restrict to primary outcomes. This modification is of interest in part because Brodeur et al. (2024) also find that pre-registered studies report more statistics, raising the question of whether more pronounced differences might emerge when filtering out for this difference. Beyond this analysis, our work also complements Brodeur et al. (2024) by including hand-coded data on outcome sharpness, comparing the AEA RCT Registry to ClinicalTrials.gov (to compare economics to other fields), and advancing a theoretical model to improve registry design. Our analysis of sample size in Section 3.2.3 complements the analysis of the impact of power analyses in Brodeur et al. (2024). They show that the distribution of test statistics among studies that include such a discussion exhibits noticeably less bunching around the 5% threshold. Our analysis, in turn, shows significant deviations in sample size can indeed be seen among registrations, suggesting a potential mechanism driving p-hacking.

The literature on registries as a distinct mechanism for research credibility has thus far focused primarily on ClinicalTrials.gov. Broadly, the literature shows that ClinicalTrials.gov fails to capture a census of all relevant trials (e.g., Manheimer and Anderson (2002) and Dickersin and Rennie (2003)); that many trials that do register do not provide sufficient information (e.g., Zarin et al.

(2011) and Zarin et al. (2017)); and that registered trials often fail to report their results (e.g., Anderson et al. (2015) and Nguyen et al. (2013)). We provide a more systematic discussion in Appendix D. However, ClinicalTrials.gov is a particularly unique registry. As the foremost medical research registry, significant aspects of the registration process are enforced by law and a large fraction of studies in ClinicalTrials.gov are funded by industry.[14] An open question is the performance of registries when these mechanisms are removed. Our assessment of the AEA RCT Registry, which is for the most part isolated from legal mandates or industry funding, suggests that the performance is similarly poor.

Finally, we also add to an important theory literature that utilises communication models to speak to researcher incentives. Examples include Di Tillio, Ottaviani and Sorenson (2021), Libgober (2022), Frankel and Kasy (2022), Al-Ubaydli, List and Suskind (2019), Tetenov (2016), and Anderson and Magruder (2017). The closest paper in this literature to our contribution is Williams (2021). While his model also provides a role for researcher signalling, it differs on a number of other key dimensions—most notably, it does not allow for a researcher to register early versus late. In contrast, motivated by our empirical findings on registration timing, this distinction plays a defining role in our theoretical exploration and results.

The signalling model that we develop is a communication game with ex-post verifiable information. That is, we imagine preregistration as a signalling action, taken prior to the observation of the study's (ex-post verifiable) results. A number of other papers have documented how ex-post verifiable information can dramatically influence the structure of equilibria (see, e.g., Feltovich, Harbaugh and To (2002), Chen, Ishida and Suen (2022), Daley and Green (2014), and Kremer and Skrzypacz (2007)). The particular signalling model we develop is distinguished by a binary signalling decision (registration) which can be delayed, whereas these related papers allow sender actions to belong to a larger set. Our setup makes full separation impossible by definition, which we view as realistic for our application.[15] More to the point, our goal is to provide tractable comparisons of registration across a variety of environments, leading us to instead seek conditions

[14] Of note, Oostrom (2024) documents that industry funded studies of psychiatric drug efficacy have inflated test statistics and that registration helps reduce this bias.

[15] For instance, while we view it as realistic to assume registration is costly, it seems less realistic to assume that the level of cost is a publicly observable choice variable. Such an assumption, however, would be necessary to allow each researcher type to send a unique message.

under which computable partitional and monotone equilibria exist.

3 Analysis of AEA RCT Registry

Academic journals tend to selectively publish studies that reject a null hypothesis to the exclusion of studies that confirm a null hypothesis or provide inconclusive results. Robert Rosenthal coined the term the “file drawer problem" in 1979 to describe the bias this selection introduces into the scientific literature.[16] This selection also directly gives researchers an incentive to repeatedly re-choose their data, outcome variables, and analysis method until they are able to reject the null hypothesis of interest at conventional levels of statistical significance. The process of repeatedly re-choosing data, outcome variables, and analysis method is commonly referred to as “p-hacking." Together, these two effects can undermine public trust in empirical research and cause inefficient resource allocations.

We start by empirically examining the extent to which the AEA RCT Registry is currently capturing the universe of economics RCTs (i.e., addressing the file drawer problem) and the extent to which it succeeds in pre-committing researchers to assessing a specific set of outcome variables (i.e., addressing p-hacking). We consider the AEA RCT Registry from its launch on May 15, 2013 through the end of 2023.[17] The AEA RCT Registry is primarily targeted at the registration of field experiments, making it more difficult to interpret any success or failure in the registration of lab experiments. As such, we focus our analysis of the file drawer problem on field experiments to prevent this distinction from interfering with the interpretation of our conclusions.[18]

[16]For example, consider 100 researchers who each conduct an experiment to test the null hypothesis that some parameter is less than or equal to 0 against the alternative that the parameter is greater than 0. At least 5 of the researchers are likely to find that the parameter is greater than 0 at a 5% significance level. If journals only publish significant results, then only these 5 studies will be published. Seeing 5 out of 5 studies rejecting the null, outside researchers might incorrectly conclude that there is strong evidence that the parameter is greater than 0.

[17]Our previous version of this paper ended its analysis in 2021; the current version has been updated to include additional data for this time period.

[18]See Appendix C for a detailed description of the AEA RCT Registry.

3.1 File Drawer

We first examine whether the AEA RCT Registry is effective at mitigating the file drawer problem for field experiments. Informally, a registry can address the file drawer problem to the extent that every field experiment that is started is added to the registry and experiment results are added to the registry at the conclusion of the experiment. Because the universe of started field experiments is unknown, we cannot determine the fraction that register with accuracy. As such, we assess the registration rate for field experiments published in leading economics journals, which we argue forms an informal upper bound for the overall registration rate.

Table I presents the registration rates for RCTs appearing in the following outlets over 2017-2023, based on our handcollected data, showing the percentage of field experiments and lab experiments that are registered as well as registration rates over time for field experiments:

- American Economic Journal: Applied Economics (AEJ-AE)
- American Economic Journal: Economic Policy (AEJ-EP)
- American Economic Journal: Microeconomics (AEJ-Mic)
- American Economic Review (AER)
- American Economic Review: Insights (AERI)
- Econometrica (ECTA)
- Experimental Economics (EE)
- Journal of Development Economics (JDE)
- Journal of Labor Economics (JLE)
- Journal of Political Economy (JPE)
- Journal of Political Economy: Microeconomics (JPE-Mic)
- Journal of the European Economic Association (JEEA)
- Quarterly Journal of Economics (QJE)
- Review of Economic Studies (ReStud)
- Review of Economics and Statistics (REStat)
- The Economic Journal (EJ)

The data show that only 61.5% of published field experiments registered across these journals (414 out of 673 papers). The median journal had a slightly lower registration rate of 54.7%. Overall, we do see some heterogeneity in registration rates, with *AER: Insights*, *QJE*, *AER*, and *AEJ-Applied* having among the highest registration rates. For lab experiments, we find the registration rates are much lower, with 10.3% of studies registered (51 out of 494), yielding an overall registration rate of 39.8%.

Part of the reason for this gap between registration rates may in part be due to journal requirements.

The AEA journals require that field experiments be registered as a condition for publication as of January 2018 (this requirement does not apply to lab experiments).[19] Table I investigates the effectiveness of this requirement by reporting the registration rates for field experiments by journal and year. Relative to the first circulation of this manuscript in 2021, we find that the registration rate among AEA journals has increased, but is still below 100%. Specifically, for 2018-2021 we find a registration rate of 81% (80 out of 99), while over the full range of 2018-2023 this rate has increased to 86% (136 out of 158), which corresponds to a rate of roughly 95% for 2022–2023. Ambiguity in the designation of a given RCT as a field or lab experiment could be responsible for the finding that registration rates are less than 100% even over the last few years (i.e., that some studies might count as lab or field, depending on the criterion used).

The second step to addressing the file drawer problem is reporting results. To this end, the AEA RCT Registry requests that researchers complete a series of post trial questions.[20] We note that the AEA RCT Registry does not collect data on whether a given RCT is a field experiment or lab experiment. As such, we examine the response rate to these questions across all registered trials. We find that the response rate is surprisingly low. Of the 7,531 registered trials with a planned end date by December 31, 2023, only 21.6% responded to *any* of the post trial questions in the AEA RCT Registry by December 31, 2024. Data on post trial results was even sparser. Only 10.9% of the 7,531 registered trials provided information on resulting papers, reports, and other material by December 31, 2024. This finding is not driven by the short horizon, although recent years have seen a slight improvement. Of the 3,176 trials that ended by December 31, 2019, 33.2% responded to one or more post trial questions and 20.0% provided information on resulting papers, reports, and other material. These results underscore that norms surrounding post-trial reporting are not established and that incentives to do so are weak, particularly as researchers may feel such reporting is unnecessary once a paper is published.

[19] See Appendix C for additional background.

[20] These questions cover intervention completion; data collection completion; data publication; program files; and resulting papers, reports, and other material.

3.2 P-Hacking

We next examine whether the AEA RCT Registry is effective at attenuating p-hacking. Informally, registrations can reduce p-hacking to the extent that they occur before the intervention begins (i.e., are preregistrations); sharply specify their primary outcomes; and match the resulting published or working papers.[21] We examine each of these points in turn, and find that most registrations in the AEA RCT Registry are late registrations that do not provide sharp informational content on their primary outcomes. We conclude by directly comparing the evidence for p-hacking from registered published RCTs to that from unregistered published RCTs using tests from the literature. Reflecting the above, these tests suggest an indistinguishable amount of p-hacking across the two samples.

3.2.1 Preregistration

We find that approximately half of registrations with the AEA RCT Registry are preregistrations. To allow time for researchers to learn about the registry's existence, we examine the subset trials that report an intervention start date on or after January 1, 2014.[22] Of these trials, only 52% (4,688 out of of 9,077 trials) registered before their intervention began.

We note the trend in registrations over time paints a somewhat more positive picture. Figure I plots the cumulative number of preregistrations and late registrations, and Figure II plots the number of preregistrations and late registrations each quarter. Preregistrations per quarter have outpaced late registrations per quarter since the start of 2021. Looking ahead to our policy counterfactual, one might think this trend would imply a late registration ban is unnecessary. However, our analysis in Section 6.3 shows the opposite: perhaps counterintuitively, given these changes, a late registration ban could be even more effective.

[21] When interpreting a paper's fidelity to the preregistration, it is important to remember that a preregistration is a statement of the initial plans for the experiment. A preregistration does not prohibit altering the experiment to navigate realised hurdles or explore unanticipated paths. Plans can and do change both before and during execution. Correspondingly, researchers can update the registration to reflect how and why the initial plans changed or explain any such changes in the paper itself.

[22] The registry became widely known after David McKenzie's October 14, 2013 World Bank Development Impact blog post.

3.2.2 Restrictiveness

Of course, preregistration alone is not sufficient for limiting p-hacking. The preregistration must also detail the primary outcomes with enough specificity to constrain variable constructions.[23] To examine the restrictiveness of preregistrations in the AEA RCT Registry, we randomly sample 1,000 registrations to assess the specificity of the primary outcomes reported by each preregistration.[24] We followed a simple protocol: we counted the number of primary outcomes listed and scored the outcome descriptions on a scale of 0 (not specific) to 5 (very specific). For example, we marked "health" as a 0, "nutritional intake" as a 1, "number of fruits consumed" as a 2, "number of fruits consumed at school per week" as a 3, "number of fruits consumed at school per week during Spring quarter" as a 4, and "number of bananas consumed at school per week during Spring quarter" as a 5." Online Appendix H.1 provides the full instructions.

Delecourt and Ng's preregistration of "Unpacking the Gender Profit Gap: Evidence from Micro-Businesses in India" provides a useful illustration. The authors plan to "test whether giving men and women the same business closes the gap in profitability. We set up our own market stalls, to which we randomly assign male and female vendors. We thus exogenously vary gender, holding the business constant." The authors' primary outcomes are (at the vendor level) "daily profit, daily revenue, number of "missed" clients, number of purchasing clients" and (at the product level) "quoted price, price paid." Note that profit, revenue, and number of purchasing clients are specific except for missing a time period; quoted price and price paid are missing both a specification of the products to be considered (likely the primary outcomes of interest will actually be price indexes) and a time period; and number of "missed" clients is missing both a specification of how missed will be measured and a time period. In this example, RAs would have been instructed to score the maximumly restrictive outcome as a 4 and the minimally restrictive outcome as a 2.

We find that existing preregistrations leave significant latitude. Table II reports the assessed restrictiveness. The average preregistration specified 3 primary outcomes. The average minimumly restrictive outcome, maximumly restrictive outcome, and median restrictive outcome are classified

[23] Researchers are notably not required to specify their secondary outcomes or submit a pre-analysis plan.

[24] This final sample reflects several updates of our data corresponding to various revisions of this paper. We describe the full process through which we arrived at our sample in Appendix E.

as weakly above a 2.0, i.e. roughly as precise as "number of fruits consumed." The preregistrations generally do not specify a precise measurement unit or measurement time period.

3.2.3 Fidelity

For the AEA RCT Registry to mitigate p-hacking, it is also essential that primary outcomes reported in the associated working and published papers match the preregistered primary outcomes. The p-hacking concern here is that researchers might change the construction of primary outcomes to achieve significant results, add additional outcomes that have a significant relationship, or not report outcomes that fail to have a significant relationship. To assess fidelity, we identified the latest working or published paper associated with each preregistered RCT and compared the outcomes reported in the paper to the preregistered primary outcomes.[25]

Table III reports the assessed fidelity of the papers associated with the preregistrations. In the average paper, 88% of the primary outcomes match their preregistered construction. This figure is encouraging, but may be somewhat misleading because the vast majority of preregistered primary outcomes are unspecific—to use Delecourt and Ng's example, there are many ways to construct a variable that reports the "price paid" for products sold by micro-businesses in India. More troubling, 8.8% of the papers report additional primary outcomes (i.e. highlight an unregistered variable in their abstract, introduction, or conclusion—see Online Appendix H.1). The average paper reports 0.57 additional primary outcomes. Similarly, 7.6% of the papers fail to report at least one primary outcome with the average paper under-reporting 0.42 primary outcomes.

Table IV focuses on the extent to which the *sample size* reported in a registration matches those in terms of final output. For *most* studies, we find that there are discrepancies between the reported sample size in the registration with those in the research output, with 38.3% of field experiments and 25.5% of lab experiments having matching sample sizes. We also find that, when there is a departure in terms of sample size, it is more often toward having less data rather than more data. These departures are not necessarily minor: For both field and lab experiments, when the sample size is larger than listed in the registration, more than half the time it is larger by 25%. Similarly,

[25] We found working or published papers for 289 of the 1000 preregistrations.

when the sample size is smaller than listed in the registration, approximately half the time it is smaller than 25%. These results are potentially troubling to the extent that these deviations are caused by endogenous choices, as it is well known that decisions about when to stop collecting data based on perceived significance can invalidate traditional hypothesis tests. While there are many innocent explanations for such departures—e.g., unforeseen funding constraints or other institutional roadblocks—we interpret these findings as underscoring our message the registrations in practice seem to be fairly weak in terms of tying researcher hands. Given that our work does not answer why such deviations occur or the extent of researcher transparency around them when they occur, we leave open how these findings may influence interpretations of statistical significance.

3.2.4 Impact of Registry on Publication Bias and P-Hacking[26]

To directly assess the impact of the AEA RCT Registry on p-hacking, we randomly selected 103 published papers with unregistered RCTs and 97 published papers with registered RCTs from the population of papers assessed in the file drawer exercise.[27][28] We assigned RAs to identify the primary outcomes reported by each paper along with the associated statistical significance. To limit confirmation bias, the RAs conducted this exercise blind to the registration status. See Online Appendix H.1 for the full instructions.[29]

We focus our analysis on the primary outcomes as these are the objects of interest for p-hacking.

[26] All analysis in this section was introduced after the first circulation of our manuscript and was not registered. It should therefore be viewed as exploratory. See Appendix E for details on the construction of all data samples.

[27] Despite the significant review effort here, we acknowledge that this sample size may still provide limited power. We first drop Experimental Economics, which published 180+ RCTs over the period of interest consisting primarily of unregistered lab experiments. This choice helps ensure a balanced coverage of journals in the final sample. After these deletions the population consists of 885 RCTs.

[28] We expanded this sample in response to reviewer feedback suggesting we include data through 2023. When we initially performed this exercise in Fall 2022, this sample consisted of 60 papers in each category; in fall 2024, when updating our data for the current version of this paper, we included an additional 40 papers in each. When initially selected papers we aimed to have an equal number of each, but in the course of checking our data we found 3 papers that had been incorrectly labeled as registered. The results from the empirical analysis with the original dataset we collected is included in Online Appendix J.

[29] In brief, we asked the RAs to identify the top two primary outcomes for the paper based on the abstract and verbally clarified that, if the paper only examined one primary outcome, then that is all the RAs should report. The RAs identified the primary outcomes, effect sizes, and either the standard errors, t-statistics, or p-values for 196 out of the 200 papers. The RAs reported just one primary outcome for seven papers (five unregistered papers and two registered papers) and reported two primary outcomes (per the instructions at Online Appendix H.1) for the remaining papers. Matching the results from Brodeur, Cook and Heyes (2020), the majority of the papers only provide standard errors. Following Brodeur, Cook and Heyes (2020), we convert t-statistics and standard errors to p-values associated with two-sided t-tests based on the standard normal distribution.

Several other papers conducting similar exercises do so examining a broader set of p-values from the results sections of published papers. However, a challenge with this approach is that papers report varying numbers of secondary outcomes, alternative specifications, and robustness checks.[30] These additional statistical tests and the inevitable variation in the types and quantities of tests expected across economic sub-disciplines complicates the interpretation of results.

Figure IV displays a histogram of the resulting p-values for the primary outcomes from the registered RCTs (in orange) and the unregistered RCTs (in blue). The distributions are visually similar. Figure V repeats these histograms for t-statistics. The distributions are again similar, though the graph suggests that the t-statistics for the unregistered RCTs may have a fatter right tail than the t-statistics for the registered RCTs. These results suggest that registration, as is, has a limited impact on p-hacking.

We conduct a battery of tests previously employed in the literature to formally test for evidence of p-hacking in each sample. Of interest is whether the sample of registered RCTs provides less evidence for p-hacking than the sample of unregistered RCTs. While these tests vary, one common theme is that all test for the presence of p-hacking near the significance threshold—that is, *marginal p-hacking*. Potential differences among these populations in terms of *non-marginal p-hacking*, which occurs significantly outside of this region, are beyond the scope of our exercise.

First, we apply the tests of Andrews and Kasy (2019) to determine the publication probability as a function of a study's findings. This test allows us to assess the extent to which publication bias differs across the two samples. We estimate their model using our full dataset and separately for registered and non-registered studies. Here, we find that both registered studies and non-registered studies appear to be selected, although with slightly less selection among registered studies. Specifically, we find that a registered study is approximately four times more likely to be published with a significant result, while a non-registered study is approximately 7.9 times more likely. This finding is intriguing in contrast to Brodeur et al. (2024), who detect no differences between registered and non-registered studies when including *all* test statistics in papers published

[30]For examples of papers following this approach, see Brodeur, Cook and Heyes (2020); Brodeur et al. (2024); Elliott, Kudrin and Wüthrich (2022). In particular, Brodeur, Cook and Heyes (2020) report over 30 p-values per paper, underscoring this point.

between 2018 and 2021.[31] Importantly, Brodeur et al. (2024) also find preregistered studies report *more* test statistics on average. In light of their insights, we interpret our findings as suggesting that p-hacking may be more significant for primary findings than secondary findings, with pre-registration influencing but nevertheless failing to significantly prevent publication bias among primary results.

Second, we apply the tests proposed by Elliott, Kudrin and Wüthrich (2022). Since the data do not only contain t-tests, we consider tests based on nonincreasingness and continuity of the p-curve. Namely, a binomial test on [0.01, 0.05], Fisher's test, a density discontinuity test at 0.05, a histogram-based test for non-increasingness (CS1), a histogram-based test for 2-monotonicity (CS2B), and the LCM test.[32] Table VI reports the results. These tests are less encouraging for pre-registration compared to those from Andrews and Kasy (2019).

- The binomial test, Fisher's Test, and LCM test fail to reject the null hypothesis of no p-hacking across both samples. The binomial test has a much lower p-value for the registered sample than the non-registered sample, but both are far from rejecting the null hypothesis. For the other tests, these tests' p-values for the registered RCT sample closely match those for the unregistered RCT sample. These tests' p-values also closely match the values that Elliott, Kudrin and Wüthrich (2022) report for the sample of published economics papers they consider (see their Figure 3).[33]

- The discontinuity test strongly rejects the null across both samples (consistent with Figure IV, which shows missing masses at 0.05).

- The CS1 test and the CS2B test detect p-hacking in the *registered* sample, but not the non-

[31] These results of this test when restricting to this timeframe are in Online Appendix J. We mention that while we detect less of a difference when restricting our data, the broad pattern remains, with registered studied being approximately 3.4 times more likely to be published and non-registered approximately 5.9 times more likely.

[32] Note that when there is also publication bias, these are joint tests for p-hacking and publication bias. We increase the range used for the Binomial test from Elliott, Kudrin and Wüthrich (2022)'s range of [0.04, 0.05] in order to increase power. There are 31 p-values in the range [0.01, 0.05] for the Not Registered sample, as well as for the Registered Sample.

[33] Elliott, Kudrin and Wüthrich (2022) apply the tests to Brodeur et al. (2016)'s sample of t-tests from 641 papers published in the AER, QJE, and JPE from 2005-2011. The findings of interest are the p-values from the tests on the full sample of de-rounded data reported in Figure 3. Elliott, Kudrin and Wüthrich (2022) report the following p-values: 0.679 for the Binomial test; 1.0 for Fisher's Test; 0.795 for the discontinuity test; 0.492 for the CS1 test; 0.428 for the CS2B test; and 1.0 for the LCM test.

registered sample.[34]

Across all tests, there is no indication that the registered sample has less evidence of p-hacking than the unregistered sample. This finding underscores the previous theme, that preregistration still leaves significant scope for p-hacking across published research in economics more generally.[35]

3.3 Proper Practices on Registration

We pause to highlight that our restrictiveness exercise posits guidelines for registration content. We acknowledge that this is somewhat subjective and that others may have different views. Nevertheless, we believe researchers interested in our views about ideal registrations should consider the following:

- First, all outcomes should be recorded.
- Second, outcomes should be as specific as possible. Each primary outcome should include the precise outcome variable, the measurement unit, and the measurement time period.
- Third, the number of observations anticipated should be recorded with the unit clearly stated. If the statistical analysis will be clustered at a higher level, the registration should also provide the relevant variable, anticipated number of clusters, and unit.
- Lastly, any changes to the primary outcomes or population should be documented.

We recognise that some researchers may be cautious about providing this level of detail given that reviewers might interpret unanticipated deviations negatively. We underscore that it is important for editors and reviewers to appreciate the inclusion of this detail with the understanding that some departures may reflect best practices rather than an unclear design. Lastly, we mention that these

[34] This result is not obtained when using data only through 2021 (see Online Appendix J).

[35] Brodeur, Cook and Heyes (2020) examine the evidence for p-hacking by identification method from a sample of causal inference papers published by 25 top journals in economics from 2015-2018. The authors report that papers which rely on difference-in-difference specifications or instrumental variables show more evidence for p-hacking than papers that rely on RCTs or regression discontinuity designs. However, the reported differences are minimal in their tighter specifications. For example, in reference to their randomisation tests, the authors note "all methods have a statistically significant discontinuity when the analysis window becomes small enough." Similarly, the authors' caliper tests show that RCT, difference-in-difference, and regression discontinuity designs provide similar evidence for p-hacking after either field or journal fixed effects are included in the caliper test specification. Given these results, we are not surprised that the evidence for p-hacking in our RCT sample is similar to the evidence for p-hacking across published research in economics more generally.

guidelines only apply to registration — for a discussion of the ideal scope of pre-analysis plans, see List (2025).

4 Registration in Other Venues

4.1 Analysis of ClinicalTrials.gov

We conduct a new survey of ClinicalTrials.gov to more precisely benchmark our results on the restrictiveness and fidelity of AEA RCT Registry preregistrations. We emphasise that our analysis of ClinicalTrials.gov builds on a large literature, which we survey in Appendix D. That said, we are not aware of any previous comparisons to the AEA RCT Registry, which is the main contribution of this section. Of note, Section 4.2 examines whether economists use other research registries in addition to or in place of the AEA RCT Registry. Verifying the consensus view, we find that the AEA RCT Registry is indeed the dominant registry for economists.

We proceed in the same manner as in Section 3.2 and focus on the launch of ClinicalTrials.gov to provide a reasonable comparison. We find that preregistrations from the first five years of ClinicalTrials.gov are somewhat more restrictive than the AEA RCT Registry preregistrations. We also find that published and working papers associated with the ClinicalTrials.gov preregistrations and with the AEA RCT Registry preregistrations have similar fidelity to the registered primary outcomes. This result, combined with the literature review in Appendix D, suggests that if ClinicalTrials.gov gives a sign of where the AEA RCT Registry is headed, then there is little reason to be optimistic that the current approach will significantly dent the credibility crisis in economics.

More precisely, we randomly sampled 300 trials that preregistered with ClinicalTrials.gov between March 1, 2000 and July 1, 2005. We choose this period since it runs from the start of the ClinicalTrials.gov website through the enforcement of the International Committee of Medical Journal Editors' (ICMJE) policy requiring investigators to preregister trials as a condition for publication. We then used the same rubric as for the AEA RCT Registry: we assessed (1) the extent to which the trial's preregistration specifies the primary outcomes in detail and (2) whether the primary outcomes reported in the latest published or working paper match those

registered.[36] Online Appendix I repeats this analysis for preregistrations with ClinicalTrials.gov after the implementation of the Final Rule for Clinical Trials Registration and Results Information Submission and reaches similar conclusions.

Table VIII reports the assessed restrictiveness of the 300 randomly selected ClinicalTrials.gov preregistrations. The average preregistration specified 2 primary outcomes—1 less than the average AEA RCT preregistration. The average minimumly restrictive outcome is classified as a 2.8, the average median restrictive outcome as 3, and the average maximumly restrictive outcome as 3.4—each roughly 1 unit more restrictive than the equivalent value for the AEA RCT preregistrations. Put another way, the median primary outcome from a ClinicalTrials.gov preregistration is roughly as specific as "number of fruits consumed at school per week." In contrast, the median primary outcome from an AEA RCT Registry preregistration is just "number of fruits consumed."[37]

We were able to associate published or working papers with 279 of the 300 ClinicalTrials.gov preregistrations. Table IX reports the assessed fidelity of the primary outcomes reported in these papers to those in the registration. In the average paper, 80% of the primary outcomes matched their registered construction—as compared to 88% for the AEA RCT Registry.[38] However, as with the AEA RCT Registry results, this figure may be misleading because the vast majority of registered primary outcomes are vague enough to match with multiple possible variable constructions. Perhaps more telling, the average paper reported 0.4 primary outcomes that were not registered and failed to report 0.4 registered primary outcomes. These values closely match those found for the AEA RCT Registry.

[36] We assessed the first available registration for each clinical trial. However, the ClinicalTrials.gov database was reset on June 23, 2005. As such, the first available registration for the majority of trials in the sample period is the version as of June 23, 2005. Because investigators may have updated their registration between the initial submission and June 23, 2005, the following analysis provides an upper bound on the restrictiveness of the preregistrations and on the fidelity of the reported primary outcomes.

[37] The last two rows in Table VIII report empirical results from comparing the latest version of the registration to the first available registration. We find 51% of the 300 assessed preregistrations later changed a primary outcome and 64% changed their sample specification. These results are an order of magnitude above those for the AEA RCT Registry. This difference could be due to the longer future horizon available for the ClinicalTrials.gov preregistrations.

[38] Of note, Ewart, Lausen and Millian (2009) find a similar 70% fidelity rate for primary outcomes registered with ClinicalTrials.gov.

4.2 Analysis of AsPredicted and Other Registries

Before turning to our theoretical model, we briefly consider AsPredicted which launched in December 2015, and also discuss our analysis of other registries. While both the AEA RCT Registry and ClinicalTrials.gov facilitate research registration and search over registrations, AsPredicted only provides permits registration — AsPredicted does not make its body of registration searchable. Put another way, AsPredicted is concerned only with p-hacking, rather than mitigating the file drawer problem.[39] For the same sample of papers in our census from Section 3.1, we search for whether an AsPredicted registration was linked. The results are presented in Table VII. The first paper we document with an AsPredicted registration is published in 2020. Overall, the number of publications describing an AsPredicted registration is quite low, although given the even more limited timeframe this is to be expected to some extent. While we conclude that it may be too soon to assess the success of AsPredicted, the fact that researchers do appear to be voluntarily using this venue over the AEA RCT Registry suggests that indeed this venue is meeting some demand among researchers that the AEA RCT Registry has not satisfied. Indeed, Imai et al. (2025) document a small decrease in the share of registrations in the AEA RCT Registry, with the remaining registrations largely going to AsPredicted together with the Open Science Framework (OSF). Their survey of researchers attributes its use to "simplicity, speed, and flexibility as a platform for concise registration." However, their findings also suggest that the use of AsPredicted, while growing, is still relatively limited compared to the AEA RCT Registry, making a more complete analysis of this venue difficult. We therefore leave a more in-depth analysis to future work.

We also sought to address whether other registries aside from those above have significant usage among economists. In particular, the Registry for International Development Impact Evaluations (RIDIE), OSF and the Evidence in Governance and Politics (EGAP) Registry all target audiences overlapping significant with the AEA RCT Registry (although, as of October 15, 2023, EGAP no longer accepts registrations). OSF in particular permits the registration of *both* RCTs and

[39] Indeed, the AsPredicted website writes that "[the file drawer benefits] are unlikely to materialize because to actually help combat the file-drawer problem authors need not only commit to telling us that a study was performed, they need to commit to reporting the result and describing the study in enough details that its quality can be assessed and the study can be easily found. The ClinicalTrials.gov experience suggests the first two requirements are unlikely to be met."

observational studies, whereas the AEA RCT Registry is targeted only at the former. Our preliminary, exploratory searches of these registries suggest fairly limited usage, although many registrations in the OSF registry do mention economics, potentially suggesting demand for registration beyond RCTs.[40] To proceed more systematically, our census also tracked whether published studies mentioned registrations in *any* venue other than the AEA RCT Registry. The results are presented in Table VII; we find only 23 registrations over all years, and are unable to discern any particular notable trend in these registrations. We conclude that, at least among papers involving RCTs, the AEA RCT Registry is the primary venue.

5 Theoretical Model of Registration

With the empirical estimates in hand, this section introduces a simple model that articulates the incentives behind registration and the implications of the registration timing decision. We view this model as relevant to empirical research generally — not just within economics — although we are motivated by our analysis in previous sections to address the question of whether potential changes to registry design could result in improvement. In our calibration exercise, we focus on the implications of the model with the AEA RCT Registry as a leading example.

We consider a researcher who faces a dynamic decision of when to register her experiment. We articulate a central tradeoff: in equilibrium, researchers more confident that their hypothesis is true register earlier, while less confident researchers may experiment without preregistering. Thus, registered results are *endogenously viewed as more credible*. While registration may influence experimentation through a number of channels, many of these would exogenously specify which researchers register versus do not. Our goal is to remain agnostic by presenting a minimal model capturing the above tradeoff, yielding endogenously determined registration rates pinned down by researcher incentives. We *then* discuss how the conclusions change under other considerations.

This section also presents our first main theoretical result: economically meaningful conditions

[40] A search conducted in June 2021 found that there was no single quarter with more than 25 economics registrations in either RIDIE or the EGAP Registry. In OSF, a general search for the keywords "Economics" or "economics" on June 13, 2021 returned only 945 registrations—many of which were, on inspection, observational studies conducted by psychologists or sociologists—although the same search in January 2025 found a marked increase in studies in the general search, returning 4227 registrations, a more than fourfold increase compared to our initial search.

under which banning late registration increases (total) registration rates. While one might have conjectured that eliminating registration opportunities would yield less registration, incentive considerations create nuances to this logic.[41] Section 6 discusses partitional equilibria, computation, and our calibration; Section 7 discusses considerations beyond the scope of our simple model.

5.1 Model

5.1.1 Players and Actions

Our model is a two-stage interaction involving a researcher (she) who takes actions and an outsider (he) who is passive but updates beliefs (e.g., a journal editor)—these beliefs, in turn, influence the researcher's payoffs. The researcher is endowed with an experiment related to state $\theta \in \{T, F\}$—for instance, reflecting whether an intervention causes a significant treatment effect. Initially, the researcher and outsider share a common prior over θ, denoting the probability they initially assign to $\theta = T$ by p_0. The researcher chooses actions in two stages:

- **Stage One**: The researcher *privately* observes a signal drawn according to $s_1 \sim f(\cdot \mid \theta)$. After observing s_1, she decides whether to (a) conduct the experiment, and (b) register the experiment if conducted. The game ends if the experiment is not conducted. We think of s_1 as reflecting expertise or the researcher's own prior work, but could reflect any information the researcher has prior to conducting the experiment which is not reflected in the prior.

  We assume $\frac{d}{ds_1} \log f(s_1 \mid T) > \frac{d}{ds_1} \log f(s_1 \mid F)$, $s_1 \in [\underline{s}_1, \overline{s}_1]$, and that $f(s \mid \theta)$ is continuously differentiable for both θ. The assumptions on f imply the strict monotone likelihood ratio property is satisfied (see Milgrom (1981)),[42] a standard assumption ensuring that higher signal realisations of s_1 should be interpreted more favorably.

- **Stage Two**: The researcher and outsider *both* observe signal $s_2 \sim g(\cdot \mid \theta)$ (e.g., the experimental results). The possible actions in Stage Two depend on the action taken in Stage

[41] While showing that *pre*-registration rates increase when late registration is banned is more straightforward (and intuitive), our goal is to be agnostic on the relative value of late registrations versus pre-registration.

[42] In our binary state context, the strict monotone likelihood ratio property states that $\frac{f(s_1|T)}{f(s_1|F)}$ is strictly increasing in s_1. To see this implication, note that our condition can be equivalently stated as $\frac{d}{ds_1} \log\left(\frac{f(s_1|T)}{f(s_1|F)}\right) > 0$; since log is a strictly monotone transformation, $\frac{f(s_1|T)}{f(s_1|F)}$ is strictly increasing if and only if its log is.

One, as well as whether late registration is banned. If late registration is *allowed* and the researcher did not register at Stage One, then she has the option to register in Stage Two. If the researcher registered in Stage One, then no further actions are taken. If late registration is *banned*, then the researcher cannot register in Stage Two, and so again no actions are taken. Whether late registration is allowed or banned is decided before the game begins.

We impose the same assumptions on g as f; so that $\frac{d}{ds_2}\log g(s_2 \mid T) > \frac{d}{ds_2}\log g(s_2 \mid F)$ for all s_2. We note that this also implies that the distribution over s_2 is larger in the FOSD order if $\theta = T$ than if $\theta = F$.[43] We assume $s_2 \in [\underline{s}_2, \overline{s}_2]$ and that $g(s \mid \theta)$ is continuously differentiable for both θ.

We thus take s_1 and s_2 to be independent conditional on θ. We refer to the triple of $f, g,$ and p_0 as the *informational environment*. And, we assume registration decisions are publicly observed (e.g., a journal editor would have access to the registration history), in addition to the signal s_2. Therefore, since s_1 is only privately observed by the researcher, the outsider's belief at the end of the interaction can be written as $\hat{p}(s_2, d)$, where $d = 1$ if registration is at time 1, $d = 2$ if registration is at time 2, and $d = \emptyset$ if registration does not occur. Note that if late registration is banned, $d = 2$ is impossible. We denote the posterior probability that $\theta = T$ following signals s_1 and s_2 as $q(s_1, s_2)$. Abusing notation slightly, we let $q(s_1, \emptyset) = \mathbb{P}[\theta = T \mid s_1]$ denote the probability that $\theta = T$ given signal only s_1 (i.e., interpreting $\emptyset$ as "no observation"). So while $\hat{p}$ refers to the outsider's beliefs over θ, q refers to the researcher's.

5.1.2 Payoffs

The researcher's final payoff depends on the outsider's belief and the actions she takes during the course of the interaction.[44] The researcher incurs a cost of $c_E \geq 0$ when conducting the experiment and incurs a cost of $c_R \geq 0$ whenever registering the experiment (whether early or late). For simplicity, we take both of these to be independent of s_1.

We think of the outsider's belief as influencing publication or citation prospects. The researcher

[43] This assumption allows us to argue that the expected second-period signal is uniformly more favorable for the researcher when $\theta = T$ is more likely.

[44] We postpone a discussion of welfare until our discussion of extensions.

obtains a benefit of $b_R(\hat{p}(s_2, d))$ if the experiment is registered and $b_N(\hat{p}(s_2, \emptyset))$ if it is not. Note that this makes the benefit of a pre-registered study equal to that of a late registered one, *fixing the outsider's belief*. This seems plausible since journals with registration requirements still publish studies registered late, so it is not clear where any such payoff difference would emerge for two studies inducing the same outsider beliefs.[45] Registering thus yields a final researcher payoff of:

$$b_R(\hat{p}(s_2, d)) - c_R - c_E,$$

where $d \in \{1, 2\}$. If the researcher does not register, her final payoff is:

$$b_N(\hat{p}(s_2, \emptyset)) - c_E.$$

If the researcher does not conduct the experiment, her final payoff is 0.

We impose the following assumption on payoffs for our analysis:

Assumption 1. *The payoff functions $b_N(p)$ and $b_R(p)$ are continuous and strictly increasing in p with $0 \leq b_N(p) \leq b_R(p)$ for all $p \in [0, 1]$. Furthermore, the difference in payoffs between registration decision, $b_R(p) - b_N(p)$, is strictly increasing in p.*

The fact that payoffs are increasing in beliefs reflects a preference for positive results; for instance, the payoff could reflect a benefit from publication where the probability of publication is increasing in the outsider's belief that $\theta = T$ (see Brodeur et al. (2016) and Andrews and Kasy (2019) for empirical evidence suggestive of this preference). The increasing difference assumption says that the gain to registration is higher when the outsider's belief is more optimistic. Equivalently, this assumption says that additional optimism benefits the researcher more following registration, suggesting complementarities between beliefs and registration. We emphasise that researcher payoffs as a function of beliefs may arise from a variety of sources (e.g., reputational considerations).

In Appendix G.3.2, we discuss a few simple microfoundations of payoffs that provide more

[45] In the context of our model, harder publication prospects for late-registered studies would emerge if, e.g., editors viewed late-registered experiments as less likely to be influential, holding fixed the updated beliefs about θ. Briefly, if late registered experiments are treated as not registered, there would be no incentive to register late in our model—and sufficient discounting of late registered studies would amount to a ban of late registration, as we consider extensively below.

context for why the increasing differences assumption is plausible in some cases. We note that strictly speaker this assumption is stronger than necessary, but imposing it simplifies the intuition dramatically and will suffice for the simple specifications we consider in our calibration exercise.

5.1.3 Equilibrium

We are interested in equilibria where:

- The researcher chooses whether or not to register at time 1 as a function of s_1 to maximise her payoffs, given $\hat{p}(s_2, d)$.
- The outsider updates his beliefs after observing the researcher's first-period action according to Bayes rule if possible (i.e., if the probability of that action is nonzero), and
- The outsider's final beliefs are further updated after observing s_2, with the researcher then deciding whether or not to register at time 2 to maximise her payoffs.
- We impose the additional requirement on equilibrium beliefs that $\hat{p}(s_2, 2) = \hat{p}(s_2, \emptyset)$, and also that $\hat{p}(s_2, d) \in [q(\underline{s}_1, s_2), q(\overline{s}_1, s_2)]$.

The last point imposes two restrictions on beliefs; first, independence from the second-period registration decision, and second, a restriction of the range of possible values given s_2. Note that the former is a requirement that beliefs not distinguish between whether a study is registered late or not registered at all. To understand motivation for this restriction, note that we would automatically have $\hat{p}(s_2, 2) = \hat{p}(s_2, \emptyset)$ if the $t = 2$ registration decision were unobservable. We could have made this assumption instead, but as discussed, in practice this decision can be observed. Still, we seek to capture the idea that the researcher would register late to satisfy a requirement, and not provide further information about the state (unlike the Stage One registration decision). Practically speaking it strikes us as unlikely that late registration conveys information about past special expertise or initial intuition, which we think of s_1 as capturing—even though theoretically this could emerge using mixed strategies.

As for the restriction that $\hat{p}(s_2, d) \in [q(\underline{s}_1, s_2), q(\overline{s}_1, s_2)]$, we note that this holds whenever beliefs are updated according to Bayes rule, and is a version of "no-signalling-what-you-don't-know" (Fudenberg and Tirole (1991)), since the researcher's belief will always be in this range. We seek to avoid issues in defining the outsider's beliefs if the researcher deviates to a registration decision that occurs with probability 0 in equilibrium. Such issues are well-known in dynamic games of incomplete information. Thus, even if the researcher were expected to never register but did, the outsider must still assign a probability to the event that $\theta = T$ that could emerge given some (probability distribution over) s_1 given s_2.

In Appendix G.1, we show equilibrium existence in this model, without needing to invoke Assumption 1. We also show that with Assumption 1, if late registration is allowed, then the researcher's second-period registration decision will be deterministic and characterised by a threshold above which the researcher registers late and below which the researcher does not register in either period. Our first main result is derived without further restrictions on equilibrium. However, to *compute* equilibria in Section 6, we will restrict to equilibria where the researcher's first-period decision is characterised by a threshold.[46]

5.1.4 Discussion and Other Comments

Before presenting our results, we briefly comment on our model's intuition and some technical considerations. We defer our discussion of interpretations and extensions until later.

First, the basic tradeoff between early and late registration which we seek to highlight is the following. On the one hand, delaying registration is tempting because it preserves *option value*. There is a chance that registration would not be worth it given the cost c_R, depending on the realisation of s_2. Alternatively, this benefit only arises if the researcher *expected* to need it, and this event is less likely if the initial signal s_1 is favorable. Therefore, registering early can be seen as the researcher declaring there is no need for this option value, thus signalling confidence that the hypothesis is true. The main benefit to *early* registration (i.e. choosing $d = 1$) rather than

[46] In particular, taking $b_R = b_N$ would imply that late registration never occurs, under the refinement we impose. However, it is still possible to have equilibria with *pre*-registration even in this case, since as we discuss in Section 6, it is possible to ensure increasing differences in s_1 holds using only assumptions on b_R, and not b_N.

late in this model is this signalling effect. We interpret this property as an *endogenous increase in credibility* caused by registration. One could imagine that there are other benefits to registering early; we discuss this further in an extension. However, from our conversations with colleagues, experiences, and personal introspection, a major driver of registration is the fear that results will be discounted if they are not preregistered. Signalling captures this incentive.

Second, we emphasise that in pursuit of minimality, we have avoided introducing additional model elements which, while plausibly significant, do not help elucidate the above tradeoff. Our goal in constructing the model has been simplicity, but this should *not* be interpreted as a view that these other elements are not important. For instance, one could imagine that registering late would be less costly than registering early. While in our view it is plausible that time and effort costs should be independent of registration timing, other aspects of registration costs might differ depending on registration timing. Still, the extent of any such "discount" from late registration seems less obvious. Note that without *any* costs of late registration, all studies would be registered late, inconsistent with our data. In any event, our analysis would be essentially unchanged with this extra parameter, except our calibration exercise would involve more degrees of freedom. We have also posited that the outsider can distinguish whether an experiment is registered or not. This reflects the idea that registration is done, for instance, as part of submission to a journal (at which point editors can see it was not pre-registered), and not as a means of "pooling." In reality this decision may indeed be imperfectly observed, but we do not see a simple way of determining how editors (for instance) make inferences about this parameter. Doing so would require us to specify the beliefs over registration, a strange object to include given that registration status and timing is observable, even though in practice this distinction may matter. We discuss additional model elements in Section 7, notably the possibility that registration increases informativeness per se.

Third, our model allows equilibria where the outsider interprets registration as coming from a researcher with $s_1 = \underline{s}_1$. Since registration is off-path in such an equilibrium, the outsider's interpretation is not restricted. We mention that we are suspicious of this equilibrium. For instance, it suggests that registration is a *negative* signal, something that, at least at first blush, seems out of line with practice; in reality, researchers in economics often advertise that their study was

preregistered, as if the inference should be that the results are *more* credible. We find it hard to justify beliefs that interpret registration as such a negative signal. Additionally, our proof of Proposition 1 also shows that such equilibria do not emerge for certain ranges of parameters.

Fourth, we have two main motivations for restricting to a setting where signals are continuously distributed—first, to be able to appeal to the simplicity of the intuition that emerges when considering the incentives of the indifferent researcher. Second, for the ability to compute equilibria in our calibration exercise by determining the s_1 signal which induces indifference. That said, strictly speaking, these benefits only emerge once we restrict to partitional equilibria—i.e., equilibria where researchers register (or experiment) if and only if initial signals are sufficiently high. We discuss the corresponding issues at length in Section 6.

Finally, despite our imposition of Assumption 1 ensuring that a more favorable (final) outsider belief leads to a larger gain to registration, by itself this is *insufficient* to ensure that higher s_1 signals more gain from registration. Intuitively, this is because the value of registration is endogenous in our model. For general equilibria, it need not be the case that the expectation of $\hat{p}(s_2, 1) - \hat{p}(s_2, \emptyset)$ conditional on s_1 is increasing in s_1.[47] Thus, even if $b_R(p) - b_N(p)$ is increasing, it might be that the gain to *early* registration is *decreasing* in s_1. This would not present an issue for our first theoretical result, Proposition 1, but would present difficulties in guaranteeing partitional equilibria (see Section 6.1). For this reason, Section 6 presents a sufficient condition on equilibrium beliefs which guarantees increasing differences in s_1, which we then use in some subsequent analysis.

5.2 Implications of a Late Ban on Registrations

Our key theoretical result is that the model features a range of parameters such that there are more registrations when late registration is banned, for general equilibria (as defined in Section 5.1.3). More precisely, we show that the probability the researcher registers increases, no matter which equilibrium is selected in either regime. On the one hand, this argument and its intuition is general and does not hinge sensitively on parametric assumptions (but instead economically interpretable

[47] For instance, if high realisations of s_1 ensured realisations of s_2 which revealed $\theta = T$, then this would be violated.

ones). On the other hand, this is only a possibility result, and indeed need not emerge for all possible parameterisations of the model. Thus, this result cannot speak to whether this prediction is reasonable, or to richer questions related to equilibrium registration patterns. To say more, Section 6 introduces partitional equilibria which we then compute using the parameterisation in Section 6.2, arguing that this prediction does not seem knife edge and is empirically relevant.

Our conditions are designed to be economically interpretable, and one of the key parameters toward that end relates to the strength of the signal s_1. The following definition quantifies this aspect of the problem. Recall that $q(s_1, s_2)$ denotes the belief that $\theta = T$ given s_1 and s_2.

Definition 1. *The **maximal influence of past expertise** is defined as the following quantity:*

$$\max_{s_2} q(\overline{s}_1, s_2) - q(\underline{s}_1, s_2).$$

This quantity bounds the impact the initial signal can have on the beliefs of the outsider. An initial signal of $\overline{s}_1$ induces as much optimism over the state as possible. A signal of $\underline{s}_1$ induces as much pessimism over the state as possible. Thus, for every s_2, $q(\overline{s}_1, s_2) - q(\underline{s}_1, s_2)$ is the difference between the belief of the most optimistic and most pessimistic researcher. Taking this maximal quantity, this parameter reflects how responsive to the first-period signal beliefs could possibly be.

The condition we need for our main proposition is that the maximal influence of past expertise is not too large. Informally, insofar as the prior reflects public knowledge about the hypothesis, this condition imposes the researcher, before conducting the experiment, not possessing too much additional knowledge relative to the public:

Proposition 1 (Implications of a Late Registration Ban on Total Number of Registrations)**.** *Fix b_R, b_N, and $g(s_2 \mid \theta)$. There exists δ such that whenever f induces $\max_{s_2} q(\overline{s}_1, s_2) - q(\underline{s}_1, s_2) < \delta$, then the following holds: The probability of registration is higher in any equilibrium when late registration is banned than in any equilibrium where late registration is allowed, for some ranges of registration costs, say $[\underline{c}_R, \overline{c}_R] \ni c_R$, and experiment costs, say $[\underline{c}_E, \overline{c}_E] \ni c_E$.*

That is, if the maximal influence of past expertise (as defined in Definition 1) is sufficiently small, then there exists a range of registration costs and experiment costs such that, in any equilibrium

(as defined in Section 5.1.3), the probability the researcher registers is higher when late registration is allowed than when it is banned.

In the proof of this proposition, we focus on the case where $\underline{c}_E, \overline{c}_E$ are such that $c_E \in [\underline{c}_E, \overline{c}_E]$ is small, to avoid cases where experimenting is too prohibitively costly to be undertaken; the assumptions that $b_N(p) \geq 0$ and $b_N(p)$ is increasing ensure that, when this is the case, researchers do not decide to stop *conducting* experiments when late registration is banned. In general, a late registration ban may discourage experiments from being conducted, which we discuss in Section 7.5. For this proposition, we do not need to consider this margin when evaluating the comparisons, though certainly policymakers might be concerned with this consequence.

Some intuition for the proposition is as follows. Consider a researcher who finds herself indifferent between registering at Stage One and delaying registration (when late registration is allowed). As a result, she finds the gain due to the "bump" in beliefs exactly offset by the option value delaying registration provides. Now suppose late registration is suddenly banned, for the moment assuming the outsider does not change his belief updating function $\hat{p}(s_2, d)$ despite the ban. The researcher finds her option value is removed—she either registers today or never register. Early registration thus becomes relatively more attractive, making her strictly prefer registering.

This argument suggests that, when late registration is banned, the probability that this marginal researcher registers increases. Indeed, by definition, if she registers early, she registers; but if she were to instead delay registration, she would not register whenever her s_2 realisation is sufficiently unfavorable, which occurs with positive probability. Thus, the probability this marginal researcher registers increases when late registration is banned. We caution that this need not immediately imply that *overall* registrations will increase in general, since we would have to worry about what a "non-marginal" researcher will do—such a researcher might never register in Stage One, whether or not late registration is allowed. For *this* (non-marginal) researcher, registration occurs with probability 0 under a late registration ban, but with positive probability when late registration is allowed (in case the s_2 realisation is sufficiently favorable).

Despite this contrary force, we show that when the initial signal is not too informative, the comparison is quite stark: all researchers are induced to pre-register when the option to delay is

removed, provided registration costs are chosen appropriately. Exhibiting an overall increase in the registration probability requires an intermediate value of c_R; whenever c_R is too low, then all researchers would register in either regime, and whenever c_R is too high then none would even under a ban. We determine such an intermediate range of registration costs where:

- Researchers would want to register early if late registration is banned, but
- Researchers would delay registration if allowed, and
- Some unfavorable subsequent realisations of s_2 would lead researchers to not register at all.

Taking the initial signal to be not too informative facilitates the argument since there is no reason to register early when late registration is allowed, since the signalling benefit is negligible. This makes the second point more immediate. Putting these together, we show a late registration ban will increase total registration rates for this range of costs given weak initial signals.

One difficulty in proving this result is the lack of any restriction on the equilibrium. As a result, we do not have enough structure on $\hat{p}(s_2, d)$ to allow us to study the incentives of such a "marginal researcher" in the first period, for general equilibria. It is also worth noting that the above intuition explicitly assumes that the outsider's updating rule does not change when late registration is banned, but this assumption will typically not be valid. A late registration ban will change the incentives of *all* researchers—possibly depressing the gains to registration for s_1 realisations which induced *more* eagerness to register. In principle, this could lead to such researchers no longer registering, recalling the discussion from Section 5.1.4 that Assumption 1 is insufficient to ensure higher s_1 implies a higher willingness to register. The proof instead looks at the limiting case where the maximal influence of past expertise is 0, in which case equilibrium requires the researcher and outsider to hold identical beliefs at time 1, and these complications are significantly diminished. We then invoke continuity properties to show that the same conclusion holds even when the signal s_1 conveys some information about the state.[48] See the Appendix for details.

[48] Aside from showing that the conclusion is not knife-edge, this step is necessary to enable $\frac{d\log(f(s_1|T))}{ds_1} > \frac{d\log(f(s_1|F))}{ds_1}$ for all s_1. It is straightforward to find distributions satisfying this condition given any $\delta > 0$; for instance, let $\tilde{f}(s_1 \mid T), \tilde{f}(s_1 \mid F)$ satisfy this condition and $\tilde{\tilde{f}}(s)$ be a distribution independent of θ; then given any $\delta > 0$ we can find $\alpha > 0$ sufficiently small such that $f(s_1 \mid \theta) = \alpha\tilde{f}(s_1 \mid \theta) + (1-\alpha)\tilde{\tilde{f}}(s_1)$ will satisfy $\frac{d\log(f(s_1|T))}{ds_1} > \frac{d\log(f(s_1|F))}{ds_1}$ with the maximal influence of past expertise less than δ.

We make one final comment on this result before turning to our calibration. It may seem that the previous proposition is only of interest insofar as the first signal is not too informative. On the one hand, our empirical analysis in Section 3.2 suggests this is likely the practically relevant case. If the initial signal is very strong, then this increased confidence should translate into a detectable difference in the distribution of test statistics due to preregistration. With the usual caveats applied to statistical tests, this is not the case, suggesting economics may indeed be close to the limiting case. On the other hand, our calibration exercise suggests, while the informativeness of s_1 does play a role in determining whether a late ban is effective, the comparisons are not particularly knife-edge. Part of our motivation for the calibration exercise is to compare the effect driving Proposition 1, on the removal of option value when late registration is banned, from the more basic (contrary) force that providing more opportunities to register may lead to more registrations.

6 Calibrating the Model and Exploring Registry Improvements

Our result on the impact of banning late registration holds for general parameterisations of the model and arbitrary equilibria. However, the conditions are derived in limiting cases of the model, and so are somewhat hard to assess. To say more requires the computation of explicit equilibria.

We begin this section with a slight detour to discuss a particular class of equilibria that are tractable to compute, namely *partitional equilibria*. We also discuss the components of this notion which are restrictive and those that are not. In Appendix F, we discuss their existence, in the process underscoring why we view them as reasonable descriptors for our application.

We then turn to a calibration exploration of the above model to explore the impact of banning late registration for the AEA RCT Registry, as suggested by Proposition 1.[49] This allows us to add empirical content to the theoretical predictions provided by the model. These calculations allow us to speak to the relative costs and benefits of a late registration ban, without needing to consider particular limiting cases on the informational environment as in that proposition.

[49] Statistics on the registrations and published papers unfortunately do not provide enough information to estimate a rich structural model.

6.1 Partitional Equilibria

Formally, we define the class of partitional equilibria for the model as follows:

Definition 2. *A **partitional equilibrium** is characterised by thresholds $s^*_{1,\emptyset}$, $s^*_{1,R}$, $s^*_{2,R}$ (where possibly $s^*_{1,\emptyset} = s^*_{1,R}$) such that:*

- *The researcher conducts the experiment whenever $s_1 > s^*_{1,\emptyset}$,*
- *The researcher preregisters the experiment whenever $s_1 > s^*_{1,R}$, and*
- *If the researcher does not preregister, then the researcher registers the experiment late whenever $s_2 > s^*_{2,R}$.*

Note that $s^*_{1,\emptyset} > s^*_{1,R}$ cannot hold in equilibrium, since not experimenting yields payoff 0, whereas registering an experiment and not conducting it yields a negative payoff.

There are two reasons we find partitional equilibria appealing. First, if an equilibrium is non-partitional, then this means either (a) a positive measure of s_1 realisations randomise over registration decisions or (b) an increase in s_1 would convince a researcher to not register. While one could debate whether these might reflect something economically substantive, at first blush neither seems particularly relevant to our application (though ruling them out in general would require complicating features).[50]

Second, there is a dramatic gain in tractability relative to other classes of equilibria. Partitional equilibria are convenient to work with because the threshold signal realisation makes the researcher indifferent between actions on each side of the threshold—that is, a researcher with signal $s^*_{1,R}$ should be indifferent between preregistration and not, and likewise for other signal realisations in this definition.[51] To arrive at comparative statics, it is often a lot easier to simply consider

[50] That said, Proposition 2 in Appendix F highlights that partitional equilibria exist in cases where returns from registration are sufficiently convex; the disproportionate returns to positive results which have been documented suggest that this assumption is reasonable in practice.

[51] Note that in a partitional equilibrium of our model, the distribution of the observed second-period signal will be truncated at $s^*_{2,R}$ for studies that are registered late; by contrast, pre-registered studies will display no such *second-period* truncation. There is empirical support for this contrast when using our preferred interpretation of s_2 as the experiment results; Adda, Decker and Ottaviani (2020) show empirically that experimental results on ClinicalTrials.gov do not display a clustering just above the significance threshold, even though this is frequently found in published studies across disciplines. Insofar as late registration may be a requirement for publication among studies not registered early, we view this result as supportive of our formulation of preregistration as well as this particular equilibrium.

the incentives of the marginal researcher, as opposed to considering the registration probability for every possible s_1 and s_2 signal realisation. In Section 7, we discuss other simple comparative statics that can be obtained by considering these marginal incentives. As alluded to in the discussion of Proposition 1, arriving at similar comparative statics is substantially less straightforward with richer equilibria classes, since a change in the decision of a marginal s_1 generally influences the incentives behind registration for *all* other values s_1.

6.2 Information Structure Specification for the Calibration Exercise

Our calibration exercises uses a simple, illustrative parameterisation of the researcher's information structure. Specifically, we assume that for $\underline{s}_1, \underline{s}_2 < 1/2$:

- $f(s_1 \mid T) = k_1 s_1$ with $s_1 \in [\underline{s}_1, 1 - \underline{s}_1]$; similarly, $g(s_2 \mid T) = k_2 s_2$ with $s_2 \in [\underline{s}_2, 1 - \underline{s}_2]$
- $f(s_1 \mid F) = k_1(1 - s_1)$ with $s_1 \in [\underline{s}_1, 1 - \underline{s}_1]$; similarly, $g(s_2 \mid F) = k_2(1 - s_2)$ with $s_2 \in [\underline{s}_2, 1 - \underline{s}_2]$

We take k_1 and k_2 to be such that densities integrate to 1. Note that given that the prior is p_0, the *researcher's* belief following a signal of s_1 is:

$$q(s_1, \emptyset) = \frac{s_1 p_0}{s_1 p_0 + (1 - s_1)(1 - p_0)}$$

After seeing s_2, the *researcher's* belief is given by:

$$q(s_1, s_2) = \frac{s_1 s_2 p_0}{s_1 s_2 p_0 + (1 - s_1)(1 - s_2)(1 - p_0)}$$

Letting $\sigma(1, s_1)$ denote the probability the researcher chooses $d = 1$ following signal s_1, then after observing $d = 1$, the outsider's belief that $\theta = T$ is:

$$\frac{p_0 \int_{\underline{s}_1}^{\bar{s}_1} s_1 \sigma(1, s_1) f(s_1 \mid T) ds_1}{p_0 \int_{\underline{s}_1}^{\bar{s}_1} s_1 \sigma(1, s_1) f(s_1 \mid T) ds_1 + (1 - p_0) \int_{\underline{s}_1}^{\bar{s}_1} (1 - s_1) \sigma(1, s_1) f(s_1 \mid F) ds_1}.$$

6.3 Numerical Calibration

We now present the details and results of our calibration exercise. Here, we make use of a number of results in Appendix F which ensure the existence of threshold equilibria for the specification of the model in Section 6.2.

Our first exercise chooses parameters to match registration rates over the entire history of the registry, during which approximately half of all registered studies are preregistered. Our second exercise chooses parameters to match rates since the first draft of this paper was circulated, during which approximately two-thirds of all registered studies are preregistered.

In our view, it is a priori unclear which specifications of the parameters are most compelling. We therefore seek to be permissive in the specifications we consider, while focusing on an information acquisition technology that allows us to tractably vary signal informativeness. Specifically, we let the first- and second-period signals have the distribution specified in Section 6.2, taking $\underline{s}_2 = 0$. Note that the informativeness of the first-period signal is decreasing in $\underline{s}_1$. For simplicity, we next assume that the payoff functions are linear—taking $b_R(\hat{p}) = \hat{p}$ and $b_N(\hat{p}) = \kappa\hat{p}$, for $\kappa < 1$.[52] We will in particular assume $\kappa = 0.80$ for our first calibration exercise, but will allow κ to vary in our second exercise.[53]

The remaining model parameters are the cost of experimentation c_E, first-period signal lower bound $\underline{s}_1$ (introduced above), the initial prior p_0, and the cost of registration c_R. We take as given that all researchers experiment, and so set $c_E = 0$.[54] For the first exercise, we focus our attention on values for $\underline{s}_1$, p_0, and c_R that produce equilibria wherein the percentage of RCTs that preregister closely matches the percentage of RCTs that register late. We further restrict attention to parameter ranges that by Proposition 3 (in the Appendix) ensure that we do not have to worry about existence issues, so that to determine an equilibrium, it suffices to find a first-period indifference condition.

[52] We only consider $\kappa < 1$ throughout to have non-zero late registration rates. As discussed above, strictly increasing differences are necessary to guarantee an interior $s^*_{1,R}$ threshold. We note that it is possible to have non-zero preregistration rates even with $\kappa = 1$; for instance, taking $p_0 = .2, c_R = .075, \underline{s}_1 = .35, c_E = 0$, we compute $s^*_{1,R} = 0.55584$. Assuming $b_R(p) - b_N(p)$ increasing seems the most immediate way of delivering non-zero late registration rates, particularly given the registration requirement for AEA journals.

[53] Recall that Proposition 3 holds independently of κ, in particular since increasing gains to early registration (Definition 1) does not depend on $b_N(p)$.

[54] This decision avoids considerations of how registration timing influences the external margin of experimentation. Note that this margin is unobserved because we are unable to determine how many potential experiments are not conducted.

Appendix G.3.1 provides the computation details.[55] Very briefly, we note that we can solve for $s^*_{2,R}$ as a function of $s^*_{1,R}$, and then perform a grid search over $s^*_{1,R}$ to determine the researcher's time 1 indifference condition. We first fix a particular choice of $\underline{s}_1$. We then vary c_R according to some small increment and, in tandem, determine the value of p_0 which yields approximately half of all registrations being early. We choose values of c_R such that the prior also lies within the range of values for which Proposition 3 holds. Using these parameters, we compute both the equilibrium when late registration is allowed, as well as when it is banned.

Table X presents the results. Columns 1 through 3 report the input $\underline{s}_1$, p_0, and c_R. Column 4 gives the percentage of RCTs that preregister in equilibrium. Column 5 confirms that this value matches the percentage of RCTs that register late. Column 6 displays the total registration rate. Note that the total registration rate is increasing in $\underline{s}_1$. That is, the registration rate is decreasing in the informativeness of the first-period signal.

Table X Column 7 reports the counterfactual of interest—How does banning late registration impact registration rates? In all cases, we find that banning late registration causes a sharp increase in preregistration. At the least, the percentage of experiments that preregister roughly doubles. In half the parameterisations, we also find that banning late registration causes an increase in overall registration with the increase being larger when the first-period signal is less informative. The critical value for $\underline{s}_1$ such that the conclusion in Proposition 1 holds is slightly above 0.36.

As discussed in our empirical analysis, since we first circulated this paper in 2019, the AEA RCT Registry has experienced a significant uptick in preregistrations. In some recent quarters, the share of new registrations that are preregistrations has been close to 2/3. Within the context of our model, what could possibly explain this change? In our view, the time horizon is too short for it to reflect anything related to the nature of experimental research. As such, changes in p_0 or $\underline{s}_t$ seem unlikely culprits. We are also unaware of any notable changes in registration or experimentation costs—ruling out significant movements in c_R or c_E. The rewards for publication or dissemination of research have likely stayed constant as well, making it difficult to see why b_R might have increased. By process of elimination, we conjecture that there has been a decrease in

[55] The computations of equilibria were performed using Mathematica. The notebook used is available here, as well as at https://www.jonlib.com/working-papers.

b_N. Our view is that referees have taken to treating RCTs that are not preregistered more harshly.

This change over time motivates our second calibration exercise. Fixing the parameters from the first calibration exercise (i.e., b_R, $\underline{s}_1$, p_0, and c_R) we ask what change in b_N rationalises the new preregistration rate of 2/3 versus the prior preregistration rate of 1/2. Using the new b_N, we then recompute all the quantities from the first calibration exercise. In particular, we re-examine the impact of banning late registration on overall registration rates.

Table XI presents the empirical results. Columns 1 through 3 again report the input $\underline{s}_1$, p_0, and c_R. Column 4 gives the imputed value for κ — note $b_N(p) = \kappa p$. Columns 5 through 7 present the preregistration rate, late registration rate, and overall registration rate under the status quo where late registration is allowed. Finally, Column 8 presents the registration rate under the counterfactual wherein late registration is banned.

Perhaps surprisingly, we find that the new b_N implies that a late registration ban is *even more* effective. The intuition is that the loss in option value is starker when b_N is lower. In the first calibration exercise, a late registration ban only increased the overall registration rate for $\underline{s}_1 \in \{0.38, 0.40\}$. In this second calibration exercise, a late registration ban causes roughly twice the increase in registration rates when $\underline{s}_1 \in \{0.38, 0.40\}$ and there is now also a marked increase in the registration rate for $\underline{s}_1 = 0.35$.

We conclude that the uptick in preregistration in recent quarters provides evidence in favor of imposing a late registration ban on the AEA RCT Registry. We acknowledge that our simple model omits other elements guiding registration decisions that may be significant. We also acknowledge that the nature of this exercise is such that parameters cannot be pinned down more precisely. Thus, we caution against the assertion that banning late registration *must* increase overall registration. We briefly mention that it is straightforward to show that a late registration ban unambiguously increases *pre*registrations (as this lowers the $s^*_{1,R}$ threshold). Insofar as society values preregistrations significantly more than late registrations, this observation tilts the scales even more strongly in favor of our proposal of a late registration ban.

7 Extensions and Further Discussion

We note that our model of registration is *not* driven by our focus on the AEA RCT Registry. The only component that may be specific to economics research is a microfoundation for the benefit of registration that we provide in Appendix G.3.2. In other settings, the benefit function may depend on the registration timing itself or may derive from other sources. That said, we now turn to select model extensions. The following discussion speaks to the general impact of registration on experiment informativeness, welfare, and the incentive to conduct experiments.

7.1 Discussion of Model Assumptions and Why Researchers Register

Our model reflects the idea that registration facilitates the dissemination of results and may facilitate publication in certain outlets. While this increases the benefits of registration (as some results may have greater ability to be published in certain venues, as is the case), it also may involve costs for negative results (e.g., by making public the existence of some failure to find results from researchers). We have also studied how this *exogenous* force leads to *endogenous* influences of outsider beliefs on registration. We interpret this as microfounding increased credibility of registered studies, beyond those implied by formal statistical analysis.

Our model does not provide scope for a registration itself to influence the outsider's beliefs, implicitly assuming that if the researcher registers the choice of how to do so is degenerate. In practice, researchers may find that a detailed pre-analysis plan conveys other information that may increase a work's publication process. Conversely, a poorly done registration or pre-analysis plan may be viewed as a negative signal. We do not suggest that the form of registration should not be relevant for how results are perceived, but rather avoid these complications in our main analysis.

At the same time, some researchers may simply have moral views that registration should always be done at a certain stage of research, or simply be unaware of the requirements that some journals have. Additionally, the fact that registration makes some aspect of an experiment public in a verifiable way may have both benefits to some researchers, as well as costs (as we discuss in Appendix C).

7.2 Private Registration

Our model assumes that the first-period registration decision is public, as is the case with the AEA RCT Registry. As discussed in Section 4.2, it is *not* the case for AsPredicted, where the decision of making the registration decision observable is at the discretion of the researcher. In principle, both the initial registration as well as the revelation of that registration may involve costs, but it seems highly implausible to us that this second step involves any significant costs. Assuming that (a) these latter costs are indeed 0, and that (b) c_R is the same for both private and public registration, this essentially amounts to a research's second-period payoff being:

$$\max\{b_N(p), b_R(p)\}.$$

Given any fixed belief distribution over $\hat{p}(s_2, 1)$, private registration weakly dominates public registration (and strictly so if $b_N(\hat{p}(s_2, 1)) > b_R(\hat{p}(s_2, 1))$ with positive probability. As a result, modifying the model to allow for private registration with payoffs as such, if we impose that (a) the outsider does not distinguish between registration based on whether it is public or private, then (b) no researchers register publicly. Note that this conclusion hinges sensitively on the assumption that the benefit of preregistration does not depend on if it is public or private—in practice, some benefits of registration accrue *only* if it occurs in the AEA RCT registry, suggesting that indeed these payoff functions should be different.

On the one hand, allowing for private registration increases the benefit of early registration relative to when this is not allowed, intuitively since researchers are not "stuck with" their registration. The implication is that the indifference threshold in the equilibrium with private registration will be *lower* than the indifference threshold in the equilibrium with only public registration. Thus, our model predicts that private registration will be more effective at spurring preregistration that public registration. The obvious catch, however, is that not all of these private preregistration will become observable registrations in the second stage.[56]

We also mention that there may be equilibria where private registration and public registration

[56] As stated in Footnote 39, AsPredicted's website explicitly maintains skepticism that registration is in itself valuable—something we have actively tried to remain agnostic on—implicitly arguing that this impact is insignificant.

coexist, where researchers with the highest values of s_1 register publicly, those with slightly lower values register privately, and those with even lower values do not preregister at all. Intuitively, this is possible because higher values of s_1 value option value less, which private registration maintains.

7.3 Informativeness

In pursuit of a model with minimal ancillary elements, we have assumed that registration does not have an impact on the nature and outcomes of the experiment itself. In practice, however, it is likely that registering an experiment induces a researcher to think through more contingencies. Alternatively, registration may provide a commitment mechanism preventing manipulation in Stage Two (i.e., when the results are obtained). Such a view may suggest that experiments that are registered are of higher quality. One could accommodate this possibility by assuming that instead, in the second stage, the researcher observes a signal $s_2 \sim g_\gamma(\cdot \mid \theta)$ where $\gamma \in \{0, 1\}$ reflects whether the experiment is registered $(\gamma = 1)$ or not $(\gamma = 0)$. And impose an additional assumption on g_γ so that the experiment is more informative when $\gamma = 1$ than when $\gamma = 0$. This modification could also reflect the case where the experiment is not p-hacked when $\gamma = 1$, but maximally p-hacked when $\gamma = 0$.[57]

Increasing the informativeness of experiments through say encouraging more detailed preregistrations could be one avenue to helpfully increase preregistration. Intuition for this claim follows from considering comparative statics in a partitional equilibrium. A researcher with strictly convex[58] $b_R(p)$ has strictly higher payoffs when the resulting experiment is more (Blackwell) informative.[59] So assuming this convexity, a researcher with signal $s^*_{1,R}$, initially indifferent between registering early versus not registering, would now have a strict incentive to register. This observation suggests

[57] We note that it is not a priori obvious how repeated sample selection may influence informativeness. Suppose each test is an independent observation of $s_i = h(\theta) + \varepsilon_i$ for $\varepsilon_i \sim H$, $h : \{T, F\} \to \mathbb{R}$, and $i \in \{1, \ldots, K\}$, with the researcher only reporting the largest of the s_i. Results from Di Tillio, Ottaviani and Sorenson (2021) show informativeness decreases in K only if $-\log H$ is logconvex. If this distribution is logconcave, then increasing K *increases* informativeness. Still, other formulations of p-hacking (e.g., dropping independence, or assuming p-hacked experiments are inherently uninformative) would, of course, change this result.

[58] Libgober (2022) shows that this convexity condition is naturally generated if follow-on work is proportional to beliefs and if the researcher prefers follow-on work when $\theta = T$.

[59] Several characterisations of the Blackwell order exist; one is that an experiment $\mathcal{I}_1$ is Blackwell-more informative than an experiment $\mathcal{I}_2$ iff $\mathcal{I}_2$ can be represented via some (potentially stochastic, but θ-independent) transformation of the outcome of $\mathcal{I}_1$ (See Blackwell (1953) and the literature following).

that if the potential damage caused by the lack of ex-ante guidance in experimentation increases, then researchers will have more incentives to preregister. In the extreme case where all researchers then pre-register, this would suggest no need to make any mandates regarding registration timing.

7.4 Welfare

Thus far we have had little to say about welfare. In our view, if the goal of a research registry is to be used, then the relevance of our results for welfare is immediate. In Appendix G.3.3 we discuss this view more thoroughly and articulate a precise welfare criterion. Under this welfare criterion, we formally equate "more registration" with "higher welfare." The idea is that registration has value via helping results escape the file drawer problem and yielding better experiment practices.

That said, we do not consider every possible trade-off. For instance, one point worth recognising is that the costs and benefits of registration may not be uniform across the discipline.[60] In particular, researchers with significant resources may be able to ensure more time and effort toward completing pre-registrations, in a way that may be difficult for researchers with tighter budget or time constraints. Thus, the costs to researchers of a late registration ban may fall disproportionately on less established researchers. Outside of the discussion below on the incentive to conduct experiments, we leave such questions to future work.

7.5 Incentive to Conduct Experiments

One potential downside to encouraging registration is that it may dampen the incentive to conduct experiments in the first place. This issue is articulated by Duflo et al. (2020). To our knowledge, our model provides a first formalisation of their observations. The concern is that an increase in the prevalence of registration in a partitional equilibrium amounts to a decrease in $s^*_{1,R}$. Since this means the best signals in the interval $[s^*_{1,\emptyset}, s^*_{1,R}]$ are now registering late, the belief following non-registration is decreasing as well. Consider the incentives of the researcher who is on the margin between experimenting and not. When $s^*_{1,R}$ decreases, the payoff from conducting an experiment decreases as well, since the results are then viewed less favorably. As a result, a researcher who was

[60] We are grateful to an anonymous referee for encouraging us to highlight this point.

indifferent between experimenting and not when $s^*_{1,R}$ is higher will strictly prefer not to experiment when $s^*_{1,R}$ is lower. Accordingly, when the threshold for registering decreases, the threshold for conducting the experiment increases, and hence fewer experiments are conducted in equilibrium. Ensuring that the effect of discounting non-registered RCTs is not *too* strong may be necessary to avoid welfare loss due to experiment conduct choice.

8 Conclusion

This paper provides a relatively sobering assessment of research registries — suggesting that so far they have not had a transformative impact on research credibility. In the case of the AEA RCT Registry, most field experiments do not register and many registrations are done for experiments that have already concluded. Perhaps most disconcerting is that even when preregistrations are completed, they often do not provide enough information to significantly attenuate p-hacking concerns. These findings are particularly surprising because preregistrations are first and foremost statements of intentions. They are not prohibitions against altering experiments to navigate realised hurdles or explore unanticipated paths.

By introducing an economic model that clarifies the costs and benefits inherent in this knowledge creation market, we are able to provide specific policy recommendations. Namely, we recommend prohibiting late registration and simultaneously providing incentives for more (and more detailed) preregistrations. This dual approach should both maximise preregistrations and increase the overall registration rate. Together these changes would significantly increase the ability of research registries to mitigate the file drawer problem and p-hacking.

Along these lines, our analysis highlights the importance of making it *as easy as possible* for registrations to be *as detailed as possible*. Indeed, requirements for longer but less detailed registrations are likely to impose additional costs on researchers while providing little benefit to consumers of research. There may also be lessons on how to best achieve this goal from other registries, such as AsPredicted. We further refer readers to our views on the proper scope of registrations in Section 3.3.

We acknowledge that much of the behavior regarding registration is undoubtedly guided by

norms. In our economic model, this takes the form of treating the costs and benefits as exogenous. Certain norms might make publishing without preregistration very difficult. If a norm change were to occur, then our analysis suggests that this feature alone could induce a higher bar for undertaking an experiment in the first place and a lower bar for registration. We suspect that this trade-off is something policymakers are cognisant of, but which our analysis formalises. Still, it is not clear what might lead to changes in norms, or whether some efforts to create such a shift may have unintended consequences. We do not doubt some of these changes are underway, and we encourage the profession to have extended discussions about the virtues and costs of registration. Awareness of the issues may influence which studies researchers view as more significant, creating shifts above and beyond publication incentives alone.

We also acknowledge that some researchers may remain critical of the enterprise of preregistration. Section 7.4 alluded to the possibility that the costs of adhering to registrations may not be uniform across the profession. Others may find that expectations of strict preregistration may inhibit exploratory analysis or that the lack of clarity on current norms creates scope for unfair standards and may stifle progress. Our broad message suggests there is merit to some of these concerns, given that norms around proper registration appear vague. At the same time, preregistration does address a genuine market failure in the form of the file-drawer problem. We believe that well-designed registration standards should address these concerns without losing sight of registration's true promise. In addition, registration may be only part of the solution — for instance, Registered Reports, allowing studies to be peer-reviewed before data collection, may be another. We leave the question of which specific endeavors may be most effective to future work.

How will research registries impact experimentation in the long run? While we have some hints from our discussion of ClinicalTrials.gov, new norms might lead to other changes in experimental conduct that would need to be considered. For instance, we do not observe researchers repeating an experiment multiple times with a new registration each time. But this behavior might emerge if the requirement to register early is sufficiently stringent. We should note that the impact of this behavior on the informativeness of experiments is generally ambiguous (see for instance Di Tillio, Ottaviani and Sorenson (2021) and Glaeser (2008)). We leave investigating this question to future

work, but we view it as important to take such concerns seriously when considering optimal policy in the knowledge creation market.

References

Adda, Jérôme, Christian Decker, and Marco Ottaviani. 2020. "P-hacking in clinical trials and how incentives shape the distribution of results across phases." *Proceedings of the National Academy of Sciences*, 117(24): 13386–13392.

Al-Ubaydli, Omar, John A List, and Dana Suskind. 2019. "The Science of Using Science: Towards an Understanding of the Threats to Scaling Experiments." National Bureau of Economic Research.

American Economic Association. n.d.. "AEA RCT Registry." https://www.socialscienceregistry.org/, Dataset.

Anderson, Michael L., and Jeremy Magruder. 2017. "Split-Sample Strategies for Avoiding False Discoveries." University of California, Berkeley, Department of Agricultural and Resource Economics.

Anderson, Monique L, Karen Chiswell, Eric D Peterson, Asba Tasneem, James Topping, and Robert M Califf. 2015. "Compliance with Results Reporting at ClinicalTrials.gov." *New England Journal of Medicine*, 372(11): 1031–1039.

Andrews, Isaiah, and Maximilian Kasy. 2019. "Identification of and Correction for Publication Bias." *American Economic Review*, 109(8): 2766–94.

Asri, Viola, Taisuke Imai, and Jessica Leight. 2024. "Pathways from Registration to Publication: Evidence from the AEA RCT Registry." Working Paper.

Bettis, Richard A. 2012. "The Search for Asterisks: Compromised Statistical Tests and Flawed Theories." *Strategic Management Journal*, 33(1): 108–113.

Blackwell, David. 1953. "Equivalent Comparison of Experiments." *Annals of Mathematical Statistics*, 24(2): 265–272.

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.

Brodeur, Abel, Nikolai Cook, and Anthony Heyes. 2020. "Methods matter: P-hacking and publication bias in causal analysis in economics." *American Economic Review*, 110(11): 3634–60.

Brodeur, Abel, Nikolai M. Cook, Jonathan S. Hartley, and Anthony Heyes. 2024. "Do Preregistration and Preanalysis Plans Reduce p-Hacking and Publication Bias? Evidence from 15,992 Test Statistics and Suggestions for Improvement." *Journal of Political Economy Microeconomics*, 2(3): 527–561.

Burlig, Fiona. 2018. "Improving transparency in observational social science research: A pre-analysis plan approach." *Economics Letters*, 168: 56–60.

Butera, Luigi, Philip J. Grossman, Dan Houser, John A List, and Marie-Claire Villeval. 2020. "A New Mechanism to Alleviate the Crisis of Confidence in Science–With an Application to the Public Goods Game." National Bureau of Economic Research.

Chassang, Sylvain, and Samuel Kapon. 2023. "Designing Randomized Controlled Trials with External Validity in Mind." *Working Paper*.

Chen, Chia-Hui, Junichiro Ishida, and Wing Suen. 2022. "Signaling Under Double-Crossing Preferences." *Econometrica*, 90(3): 1225–1260.

Chopra, Felix, Ingar Haaland, Christopher Roth, and Andreas Stegmann. 2024. "The Null Result Penalty." *The Economic Journal*, 134(657): 193–219.

Christensen, Garrett, and Edward Miguel. 2018. "Transparency, Reproducibility, and the Credibility of Economics Research." *Journal of Economic Literature*, 56(2): 920–980.

Daley, Brendan, and Brett Green. 2014. "Market signaling with grades." *Journal of Economic Theory*, 151: 114–145.

Dickersin, Kay, and Drummond Rennie. 2003. "Registering Clinical Trials." *JAMA*, 290(4): 516–523.

Di Tillio, Alfredo, Marco Ottaviani, and Peter N. Sorenson. 2021. "Strategic Sample Selection." *Econometrica*, 89(2): 911–953.

Dreber, Anna, Thomas Pfeiffer, Johan Almenberg, Siri Isaksson, Brad Wilson, Yiling Chen, Brian A Nosek, and Magnus Johannesson. 2015. "Using Prediction Markets to Estimate the Reproducibility of Scientific Research." *Proceedings of the National Academy of Sciences*, 112(50): 15343–15347.

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." National Bureau of Economic Research.

Elliott, Graham, Nikolay Kudrin, and Kaspar Wüthrich. 2022. "Detecting p-Hacking." *Econometrica*, 90(2): 887–906.

Ewart, Robert, Harald Lausen, and Norman Millian. 2009. "Undisclosed Changes in Outcomes in Randomized Controlled Trials: An Observational Study." *The Annals of Family Medicine*, 7(6): 542–546.

Feltovich, Nicholas J, R Harbaugh, and T To. 2002. "Too Cool for School? Signalling and Countersignalling." *The RAND Journal of Economics*, 33(4): 630–649.

Frankel, Alexander, and Maximilian Kasy. 2022. "Which Findings Should be Published?" *American Economic Journal: Microeconomics*, 14(1): 1–38.

Fudenberg, Drew, and Jean Tirole. 1991. "Perfect Bayesian equilibrium and sequential equilibrium." *Journal of Economic Theory*, 53(2): 236–260.

Glaeser, Edward. 2008. "Researcher Incentives and Empirical Methods." *The Foundations of Positive and Normative Economics*, 300–319.

Harrison, Glenn W, and John A List. 2004. "Field Experiments." *Journal of Economic literature*, 42(4): 1009–1055.

Imai, Taisuke, Séverine Toussaert, Aurélien Baillon, Anna Dreber, Seda Ertaç, Magnus Johannesson, Levent Neyse, and Marie Claire Villeval. 2025. "Pre-Registration and Pre-Analysis Plans in Experimental Economics." Institute for Replication (I4R) 220.

Ioannidis, John PA. 2005. "Why Most Published Research Findings are False." *PLos med*, 2(8): e124.

Jennions, Michael D, and Anders Pape Møller. 2003. "A Survey of the Statistical Power of Research in Behavioral Ecology and Animal Behavior." *Behavioral Ecology*, 14(3): 438–445.

Kremer, Ilan, and Andrzej Skrzypacz. 2007. "Dynamic Signaling and Market Breakdown." *Journal of Economic Theory*, 133(1): 58–82.

Leamer, Edward E. 1983. "Let's Take the Con Out of Econometrics." *American Economic Review*, 73(1): 31–43.

Libgober, Jonathan. 2022. "False Positives and Transparency." *American Economic Journal: Microeconomics*, 14(2): 478–505.

List, John A. 2025. *Experimental Economics: Theory and Practice.* University of Chicago Press.

Manheimer, Eric, and Diana Anderson. 2002. "Survey of Public Information About Ongoing Clinical Trials Funded by Industry: Evaluation of Completeness and Accessibility." *BMJ*, 325(7363): 528–531.

Maniadis, Zacharias, Fabio Tufano, and John A List. 2014. "One Swallow Doesn't Make a Summer: New Evidence on Anchoring Effects." *American Economic Review*, 104(1): 277–90.

Milgrom, Paul. 1981. "Good news and bad news: representation theorems and applications." *The Bell Journal of Economics*, 12(2): 380–391.

Nguyen, Thi-Anh-Hoa, Agnes Dechartres, Soraya Belgherbi, and Philippe Ravaud. 2013. "Public Availability of Results of Trials Assessing Cancer Drugs in the United States." *Journal of Clinical Oncology*, 31(24): 2998–3003.

Nosek, Brian A, Jeffrey R Spies, and Matt Motyl. 2012. "Scientific Utopia: II. Restructuring Incentives and Practices to Promote Truth Over Publishability." *Perspectives on Psychological Science*, 7(6): 615–631.

Olken, Benjamin A. 2015. "Promises and Perils of Pre-Analysis Plans." *Journal of Economic Perspectives*, 29(3): 61–80.

Oostrom, Tamar. 2024. "Funding of Clinical Trials and Reported Drug Efficacy." *Journal of Political Economy*, 132(10): 3298–3333.

Tetenov, Aleksey. 2016. "An Economic Theory of Statistical Testing." University of Geneva.

Vivalt, Eva. 2018. "Specification Searching and Significance Inflation Across Time, Methods and Disciplines." *Oxford Bulletin of Economics and Statistics*, 81(4): 797–816.

Williams, Cole. 2021. "Preregistration and Incentives." *Working Paper*.

Zarin, Deborah A, Tony Tse, Rebecca J Williams, and Thiyagu Rajakannan. 2017. "Update on Trial Registration 11 Years After the ICMJE Policy was Established." *New England Journal of Medicine*, 376(4): 383–391.

Zarin, Deborah A, Tony Tse, Rebecca J Williams, Robert M Califf, and Nicholas C Ide. 2011. "The ClinicalTrials.gov Results Database——Update and Key Issues." *New England Journal of Medicine*, 364(9): 852–860.

A Tables

Table I: Registration of published RCTs over 2017-2023 by journal, experiment type, and year

	Count		Fraction Registered		Count (Field)							Fraction Registered (Field)						
	Field	Lab	Field	Lab	2017	2018	2019	2020	2021	2022	2023	2017	2018	2019	2020	2021	2022	2023
Journal																		
AEJ-AE	64	5	0.80	0.20	9	12	14	8	8	5	8	0.67	0.67	0.71	0.88	1.00	0.80	1.00
AEJ-EP	29	4	0.76	0.00	1	4	5	2	3	10	4	0.00	0.50	0.60	0.50	1.00	0.90	1.00
AEJ-Mic	2	32	0.50	0.09	0	0	1	0	0	1	0	nan	nan	0.00	nan	nan	1.00	nan
AER	83	28	0.83	0.25	19	6	6	13	13	15	11	0.53	0.50	1.00	0.92	1.00	1.00	0.91
AERI	9	1	1.00	0.00	0	0	3	0	1	2	3	nan	nan	1.00	nan	1.00	1.00	1.00
ECTA	9	9	0.56	0.11	0	1	2	1	2	1	2	nan	1.00	0.50	1.00	0.50	0.00	0.50
EE	32	249	0.22	0.06	3	5	1	2	4	12	5	0.00	0.00	0.00	0.50	0.50	0.00	0.80
EJ	64	56	0.44	0.07	6	7	5	6	11	9	20	0.33	0.29	0.20	0.50	0.64	0.44	0.45
JDE	167	11	0.60	0.18	11	22	11	21	27	32	43	0.45	0.41	0.55	0.52	0.70	0.62	0.70
JEEA	30	39	0.40	0.13	3	4	5	3	7	5	3	0.00	0.50	0.40	0.33	0.29	0.40	1.00
JLE	20	2	0.45	0.00	2	3	2	2	4	3	4	0.00	0.33	0.50	0.50	0.50	0.67	0.50
JPE	32	12	0.56	0.25	3	4	4	8	4	4	5	0.00	1.00	0.50	0.50	0.50	1.00	0.40
JPE-Mic	1	2	0.00	0.50	0	0	0	0	0	0	1	nan	nan	nan	nan	nan	nan	0.00
QJE	45	8	0.84	0.38	6	7	7	4	6	7	8	0.50	0.71	1.00	1.00	1.00	0.71	1.00
ReStat	52	19	0.54	0.16	5	4	5	6	11	11	10	0.20	0.25	0.20	0.50	0.73	0.55	0.80
ReStud	34	17	0.50	0.18	1	2	10	4	10	7	0	1.00	1.00	0.50	0.25	0.30	0.71	nan

Table II: Assessment of the extent to which 1000 randomly chosen AEA RCT Registry experiment preregistrations precisely specify their primary outcomes

	Mean	Std	Min	10%	25%	50%	75%	90%	Max
Number of Outcomes	3.00	2.56	0.00	1.00	1.00	2.00	4.00	6.00	28.00
Minimumly Restrictive Outcome	2.18	1.31	0.00	1.00	1.00	2.00	3.00	4.00	5.00
Maximumly Restrictive Outcome	2.48	1.41	0.00	1.00	1.00	2.00	3.50	5.00	5.00
Median Restrictive Outcome	2.34	1.28	0.00	1.00	1.00	2.00	3.00	4.00	5.00
Outcome Changed (Yes/No)	0.05	0.21	0.00	0.00	0.00	0.00	0.00	0.00	1.00
Sample Changed (Yes/No)	0.08	0.26	0.00	0.00	0.00	0.00	0.00	0.00	1.00

Notes: Preregistrations were randomly sampled from the period May 15, 2013 to December 31, 2021. The RAs were instructed to mark unspecific outcomes as a 0 and very specific outcomes as a 5. The instructions (which include a scoring example) are presented in Online Appendix H.1. Percentiles are computed using linear interpolation.

Table III: Assessment of the extent to which working and published papers report the primary outcomes preregistered with the AEA RCT Registry

	Mean	Std	Min	10%	25%	50%	75%	90%	Max
Fraction of Matching Outcomes	0.88	0.26	0.00	0.50	1.00	1.00	1.00	1.00	1.00
Number of Additional Outcomes	0.57	1.26	0.00	0.00	0.00	0.00	1.00	2.00	9.00
Number of Missing Outcomes	0.42	0.95	0.00	0.00	0.00	0.00	0.50	1.00	7.00

Notes: Associated papers were found for 289 of the 1000 preregistrations. Percentiles are computed using linear interpolation.

Table IV: Assessment of the extent to which working and published papers match the sample sizes listed in the registrations

Lab/Field	Frac. Equal	Frac. Greater	Frac. Smaller	Frac. > 1.05	Frac. > 1.10	Frac. > 1.25	Frac. < 0.95	Frac. < 0.90	Frac. < 0.75
Total	36.92%	26.15	36.92	20.66	17.36	14.51	30.11	26.15	16.92
Field	38.33	25.80	35.87	20.15	16.71	14.25	29.73	25.80	15.97
Lab	25.53	27.66	46.81	23.40	21.28	14.89	34.04	29.79	25.53

Notes: Data was collected in the same stage as our restrictiveness exercise. See Online Appendix H.2 for details on the instructions given.

Table V: Evidence for p-hacking using the procedure of Andrews and Kasy (2019)

	μ (SE)	τ (SE)	df (SE)	$[0, 1.96]$ (SE)
Overall	0.009 (0.004)	0.012 (0.010)	1.094 (0.083)	0.163 (0.028)
Registered	0.020 (0.007)	0.030 (0.012)	1.266 (0.154)	0.244 (0.056)
Not Registered	0.005 (0.004)	0.006 (0.002)	1.030 (0.098)	0.126 (0.025)

Notes: We use the specification of the publication probability which is symmetric, whose errors follow a student-t distribution, allowing for a single step at 1.96. The stated parameters μ, τ and df represent parameters of the model. The last column gives the publication probability for a result insignificant at the 5 percent level relative to a significant result. A value of 1 in this column implies no selection, whereas 1 divided by this column gives how much more likely a study with a significant result is to be published relative to an insignificant one. Standard errors of all estimates are in parentheses.

Table VI: Evidence for p-hacking by registration status based on the tests from Elliott, Kudrin and Wüthrich (2022)

Test	Not Registered	Registered
Binomial	0.96	0.50
Discontinuity	0.00	0.00
CS1	0.42	0.02
CS2B	0.37	0.02
LCM	1.00	1.00

Notes: There are 190 p-values in the Registered sample and 195 p-values in the Not Registered sample. Per Elliott, Kudrin and Wüthrich (2022), since the data do not only contain t-tests, we consider tests based on nonincreasingness and continuity of the p-curve (Theorem 1). Namely, a binomial test on [0.01, 0.05], Fisher's test, a density discontinuity test at 0.05, a histogram-based test for non-increasingness (CS1), and the LCM test. The CS1 test uses 30 bins. We increase the range used for the Binomial test from Elliott, Kudrin and Wüthrich (2022)'s range of [0.04, 0.05] in order to increase power. There are 31 p-values in the range [0.01, 0.05] in the Not Registered sample and 31 p-values in this range in the Registered Sample. Fisher's Test returns a value of 1 for both the Registered and Not Registered sample, and is hence not included in this table.

Table VII: Registrations in AsPredicted and other registries by journal and year.

	Number in As Predicted							Number Registered elsewhere						
Journal/Year	17	18	19	20	21	22	23	17	18	19	20	21	22	23
AEJ-AE	0	0	0	0	0	0	0	0	0	1	0	0	0	0
AEJ-EP	0	0	0	0	0	0	0	0	0	0	0	0	0	0
AEJ-Mic	0	0	0	0	0	1	0	0	0	0	0	0	0	0
AER	0	0	0	0	1	0	1	1	0	0	0	0	1	0
AERI	NA	NA	0	0	0	0	0	NA	NA	0	0	0	0	0
ECTA	0	0	0	0	0	0	0	0	0	0	0	0	0	0
EE	0	0	0	0	0	1	2	0	0	0	1	0	2	0
EJ	0	0	0	0	0	0	2	0	1	0	0	0	0	1
JDE	0	0	0	0	0	0	0	0	1	0	1	1	2	1
JEEA	0	0	0	0	0	1	0	0	0	0	1	0	0	0
JLE	0	0	0	0	0	0	0	0	0	0	0	0	0	0
JPE	0	0	0	0	0	0	0	0	0	0	2	0	0	0
JPE-Mic	NA	NA	NA	NA	NA	NA	0	NA	NA	NA	NA	NA	NA	0
QJE	0	0	0	0	0	0	0	0	0	0	0	1	2	1
ReStat	0	0	0	0	0	0	1	0	0	0	0	0	0	1
ReStud	0	0	0	1	0	0	NA	0	0	0	0	0	1	NA

Notes: Table includes both lab and field experiments. Entries of NA are for years prior to when the journal had started publishing issues.

Table VIII: Assessment of the extent to which 300 randomly chosen ClinicalTrials.gov preregistrations precisely specify their primary outcomes

	Mean	Std	Min	10%	25%	50%	75%	90%	Max
Number of Outcomes	1.96	1.19	0.50	1.00	1.00	1.50	3.00	4.00	6.00
Minimumly Restrictive Outcome	2.76	0.99	1.00	1.50	2.00	3.00	3.50	4.00	5.00
Maximumly Restrictive Outcome	3.34	0.99	1.00	2.00	3.00	3.50	4.00	4.50	5.00
Median Restrictive Outcome	3.04	0.91	1.00	2.00	2.50	3.00	3.52	4.03	5.00
Outcome Changed (Yes/No)	0.51	0.45	0.00	0.00	0.00	0.50	1.00	1.00	1.00
Sample Changed (Yes/No)	0.64	0.44	0.00	0.00	0.00	1.00	1.00	1.00	1.00

Notes: Preregistrations were randomly sampled from the period March 1, 2000 to July 1, 2005. This period corresponds to the first five years of the ClinicalTrials.gov registry and predates the ICMJE policy requiring preregistration for publication in most medical journals. Percentiles are computed using linear interpolation.

Table IX: Assessment of the extent to which working and published papers report the primary outcomes preregistered with ClinicalTrials.gov

	Mean	Std	Min	10%	25%	50%	75%	90%	Max
Fraction of Matching Outcomes	0.80	0.30	0.00	0.33	0.60	1.00	1.00	1.00	1.0
Number of Additional Outcomes	0.37	0.93	0.00	0.00	0.00	0.00	0.50	1.00	10.5
Number of Missing Outcomes	0.37	0.84	0.00	0.00	0.00	0.00	0.50	1.50	7.0

Notes: Associated papers were found for 279 of the 300 preregistrations. Percentiles are computed using linear interpolation.

Table X: Equilibrium registration rates for various model specifications

$\underline{s}_1$	p_0	c_R	% Preregister	% Register Late	% Register	% Preregister (Late Ban)
0.33	0.124	0.100	3.73	3.85	7.58	6.35
0.33	0.159	0.120	4.05	3.99	8.04	6.30
0.33	0.198	0.140	3.91	3.95	7.86	5.66
0.35	0.149	0.100	5.59	5.29	10.9	10.1
0.35	0.169	0.110	5.52	5.46	11.0	9.73
0.35	0.190	0.120	5.30	5.58	10.9	9.20
0.38	0.148	0.080	7.98	7.84	15.8	19.0
0.38	0.172	0.090	8.28	8.26	16.5	19.1
0.38	0.198	0.100	9.09	8.61	17.7	19.6
0.40	0.102	0.050	8.43	8.93	17.4	28.3
0.40	0.167	0.075	10.1	10.6	20.6	30.3
0.40	0.242	0.100	12.3	12.0	24.3	32.0

Notes: Table computes registration rates for choices of $\underline{s}_1, p_0, c_R$ such that roughly half of all registrations are preregistrations. Each calculation uses $c_E = 0$, $\underline{s}_2 = 0$, $b_R(\hat{p}) = \hat{p}$, and $b_N(\hat{p}) = 0.8\hat{p}$. Columns 1 through 3 report the input $\underline{s}_1$, p_0, and c_R. Columns 4 and 5 present the percent of experiments that preregister or register late in equilibrium, respectively. Column 6 displays the total registration rate, the sum of these two numbers. Column 7 reports the registration rate (which is also the preregistration rate) under a ban on late registration.

Table XI: Equilibrium registration rates with b_N set to match recent registration rates

$\underline{s}_1$	p_0	c_R	Implied b_N	% Early	% Late	% Total	% Register (Late Ban)
0.33	0.124	0.100	0.730	11.5	5.77	17.3	17.7
0.33	0.159	0.120	0.745	11.6	5.81	17.4	16.9
0.33	0.198	0.140	0.755	11.7	5.81	17.5	16.1
0.35	0.149	0.100	0.750	13.8	6.83	20.6	22.2
0.35	0.169	0.110	0.755	13.7	6.97	20.7	21.6
0.35	0.190	0.120	0.755	14.4	7.23	21.6	22.2
0.38	0.148	0.080	0.76	17.6	8.82	26.4	34.3
0.38	0.172	0.090	0.76	18.9	9.24	28.1	35.8
0.38	0.198	0.100	0.765	19.5	9.48	28.9	35.7
0.40	0.102	0.050	0.760	19.2	9.57	28.8	47.1
0.40	0.167	0.075	0.765	22.0	10.9	33.0	50.4
0.40	0.242	0.100	0.775	23.6	12.0	35.6	50.1

Notes: Table recomputes equilibria from Table X, but assuming b_N such that 2/3 of all registrations are preregistrations. Columns 1 through 3 report the input $\underline{s}_1$, p_0, and c_R. Column 4 reports the closest b_N within a .005 increment such that approximately 2/3 of all registrations are early, given these parameters. Columns 5 and 6 present the percent of experiments that preregister or register late in equilibrium, respectively. Column 7 displays the total registration rate, the sum of these two numbers. Column 8 reports the registration rate (which is also the preregistration rate) under a ban on late registration.

B Figures

Figure I: Cumulative number of AEA RCT preregistrations and late registrations

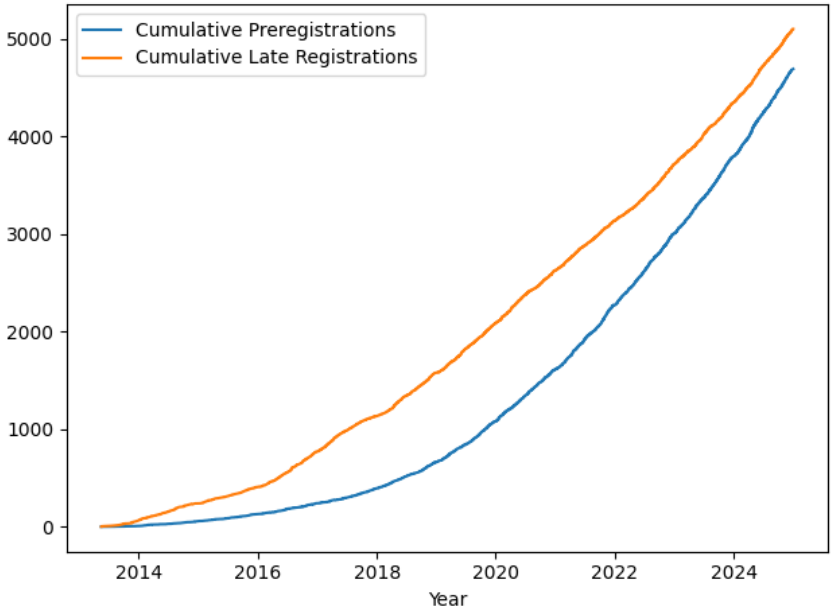


Figure II: Number of AEA RCT preregistrations and late registrations by quarter

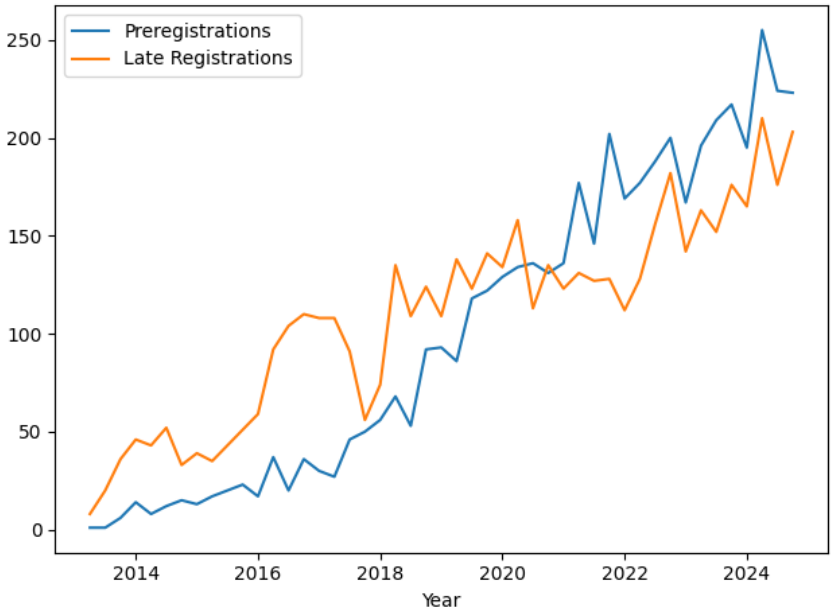


Figure III: Days between intervention start and AEA RCT registration for RCTs started after January 1, 2014. Positive values indicate that the intervention began after the registration.

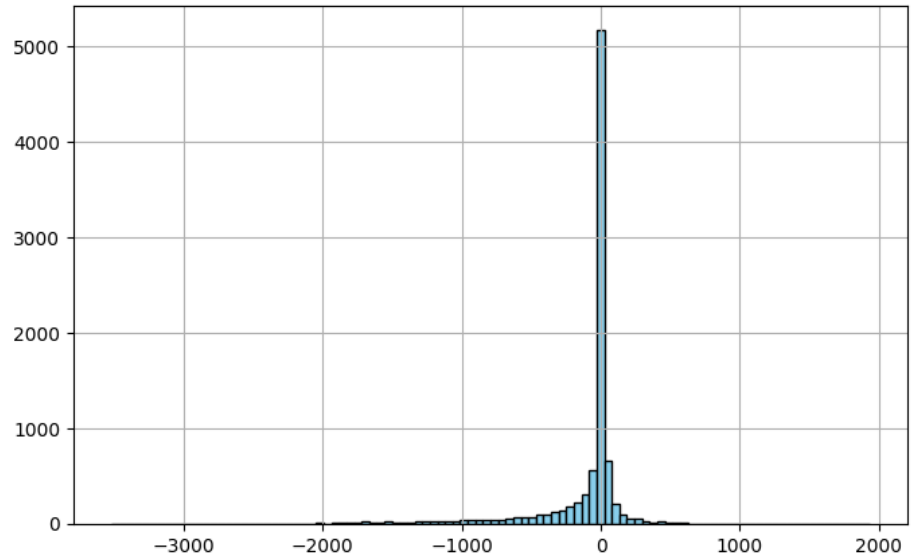


Figure IV: Histogram of p-values by registration status

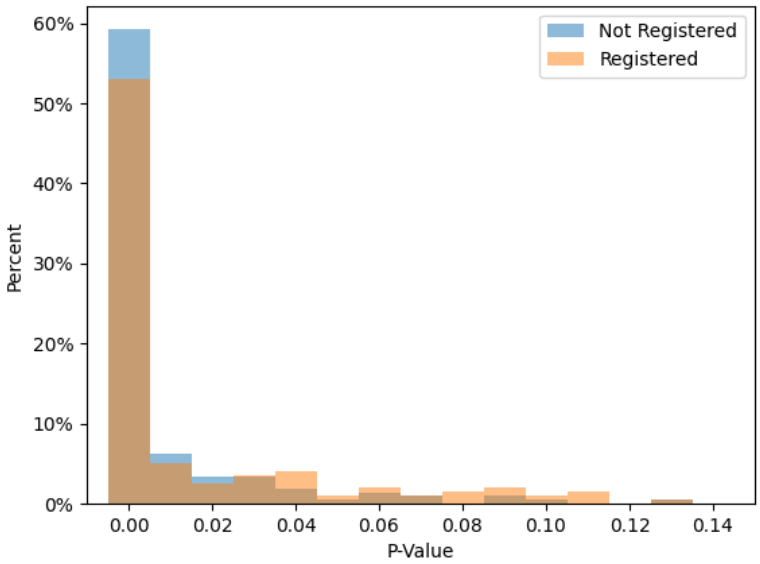
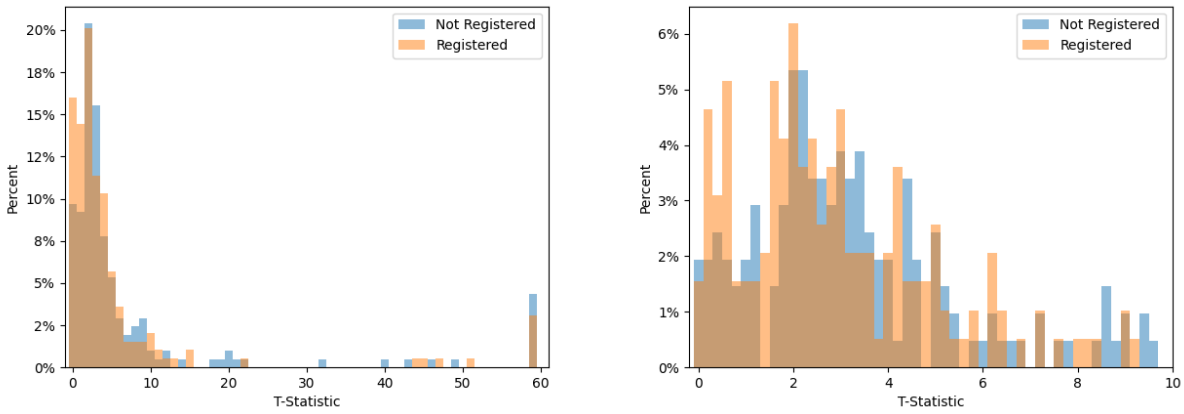


Figure V: Histogram of t-statistics by registration status



(a) All Values (Clipped at 60)

(b) Subset to T-Stat < 10

Figure VI: Timing of moves in the model

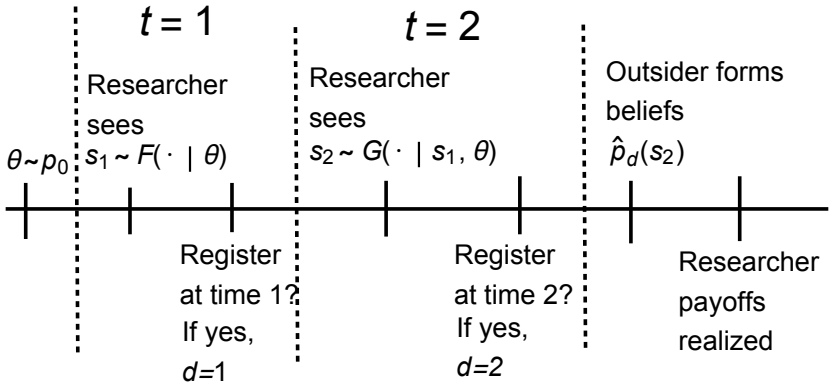
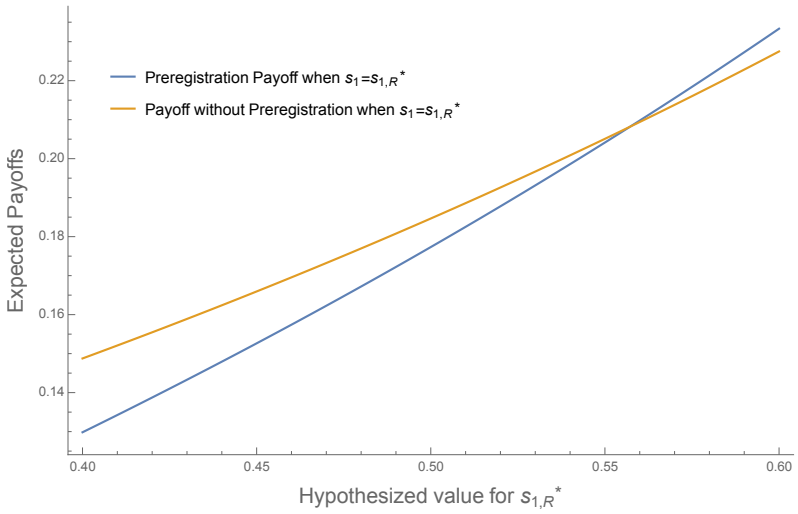
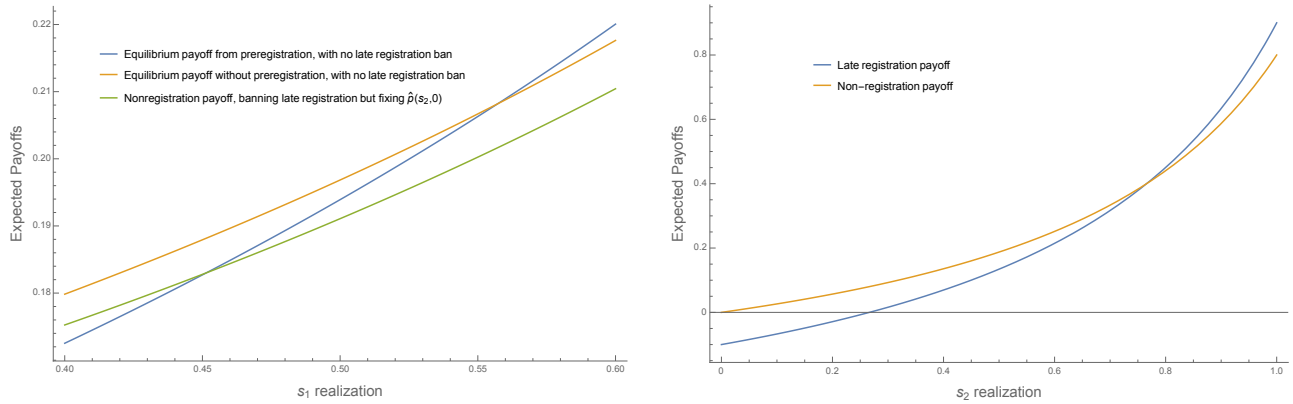


Figure VII: Researcher payoff upon receiving signal $s_1 = s^*_{1,R}$ assuming the equilibrium registration threshold is conjectured by the outsider to be $s^*_{1,R}$



Notes: Illustration of the Equilibrium Construction. The (conjectured) equilibrium threshold is the intersection point of these two lines. Payoffs and information structure are as in Table X, with $\underline{s}_1 = .4$, $c_R = .1$, and $p_0 = .25$.

Figure VIII: Researcher equilibrium expected payoffs when late registration is allowed, in the first period (left panel) and in the second period (right panel), given each possible action as a function of that period's signal



Notes: Equilibrium illustration with $\underline{s}_1 = .4$, $c_R = .1$, and $p_0 = .25$. In the left panel, the intersection of the blue and orange lines is $s^*_{1,R}$. The green line represents the intuition for Proposition 1 from the text: removing the option to register late removes option value, and thus lowers the (pre)registration threshold (the intersection between the green and blue lines) if the outsider does not adjust $\hat{p}(s_2, 0)$. Note that in equilibrium, the (pre)registration threshold will be higher than this intersection, since when beliefs adjust, the payoff from preregistration will be lower than the blue line. In the right panel, $s^*_{2,R}$ is the intersection of the two lines.

C Background on the AEA RCT Registry

The AEA launched the AEA RCT Registry in May 2013 to capture ongoing, completed, and terminated RCTs in economics and other social sciences (see About the Registry on the AEA registry webpage).[1] At the time, existing registries, such as ClinicalTrials.gov, focused on medical trials.[2] The AEA chose to implement a streamlined registration process to encourage participation—registration only requires answering a few questions and researchers are able to register at any time *even after the RCT is completed*. The required questions ask for a title, short abstract, start date, primary outcomes, treatment arms, and IRB approval number.[3] The AEA decided to focus on RCTs since experiments have a fairly distinct beginning and end. While one could imagine allowing non-RCT projects to register as well, it is difficult to implement a credible registration approach for research on pre-existing data.[4]

We use the term preregistration to denote a registration that occurs before the start date of the RCT's intervention. A related concept is a pre-analysis plan. We define a pre-analysis plan as a statistical analysis plan that is added to the registration before the start date of the RCT's intervention. Duflo et al. (2020) propose that a pre-analysis plan should answer two questions: "What are the key outcomes and analyses?" and "What is the planned regression framework or statistical test for those outcomes?" A more detailed pre-analysis plan may go further and specify *all* steps involved in analyzing the data. Of note, registries generally allow researchers to record a pre-analysis plan via text fields within the registration or via uploading a separate document. See Ofosu and Posner (2023) for an assessment of the pre-analysis plans that have been added as attached documents to the AEA RCT Registry.

Crucially, neither a preregistration nor a pre-analysis plan prevent the researcher from altering the experiment to navigate realised hurdles or explore unanticipated paths. Plans can and do change

[1] We are indebted to Rachel Glennerster for providing context about the registry's creation.

[2] The International Initiative for Impact Evaluation (3ie) Registry, which focuses on experiments in developing countries, and the Evidence in Governance and Politics (EGAP) Registry, which focuses on political science experiments, were launched contemporaneously.

[3] Many RCTs in economics require IRB approval, but the IRB approvals are not made publicly available. We note that a policy that either made external registration a condition for IRB approval or made IRB approvals public could help address the file drawer problem.

[4] See Burlig (2018) for a more thorough discussion of issues related to the registration of and pre-analysis plans for non-RCT empirical studies.

both before and during execution. Correspondingly, registries, including the AEA RCT Registry, generally allow (and even encourage) researchers to update the registration or analysis plan to reflect and explain any changes to the initial experiment design.

Though not explicitly stated on the website, the AEA RCT Registry is primarily focused on capturing economics field experiments. While many lab experiments have chosen to register, some of the registry questions are less natural for certain lab experiments. In the same spirit, as of January 2018, the AEA journals require that field experiments, but not necessarily lab experiments, be registered as a condition for publication.[5] In any case, no economics journal requires that any experiment preregister—instead allowing registration to be done at the time of submission.[6] In contrast, most medical journals require preregistration of clinical trials.

The timing of an AEA RCT Registry registration can be determined from its listing in the registry database. Preregistered trials are marked by a small orange clock. Trials that registered after their intervention start date are instead marked by a grey clock. That said, it is not clear to us whether this distinction is salient or appreciated by consumers of research (or referees and editors). Unfortunately, we are not able to precisely study the extent to which the time of registration is distinguishable to someone who searches the registry. Our own conjecture is that the distinction is minor,[7] though researchers may emphasise that a study was preregistered in the corresponding written paper.

Finally, two other aspects of the AEA RCT Registry prove important in practice. First, the registry sends automatic reminders to encourage researchers to complete fields that become relevant during and after the RCT. For example, after the trial has concluded, researchers are asked to link to any data, program files, or results that they have made public. Second, researchers are able to hide several fields in the registration from public view until later dates (specifically, the trial's location, intervention description, experimental design, names of any sponsors or partners, and supporting

[5]The specific policy is "As of January 2018, registration in the RCT registry is mandatory for all applicable submissions. This applies to field experiments. Laboratory experiments do not need to be registered at this time."

[6]The official policy states, emphasis added, "If the research in your paper involves an RCT, please register (registration is free), prior to submitting. We also kindly ask you to acknowledge compliance by including your RCT ID number in the introductory footnote of your manuscript. *Registration ideally happens before the project launches, but registering at the time of submission is also acceptable.*" (Emphasis added)

[7]Anecdotally, despite our own familiarity with the registry, we never realised these clock icons existed until starting this project. Likewise, in our informal discussions of this paper with colleagues—several of whom regularly referee field experiments—many were not aware of how to determine this distinction prior to our informing them.

documents). On this last point, we stress that, as is, allowing these fields to be temporarily hidden does not fully eliminate the possibility that registering could invalidate an RCT's experimental design. One oversight is that the registry does not permit hiding the researchers' names, experiment title, and start date. This oversight is problematic for any RCT where identification relies on the intervention's occurrence being undisclosed to participants in the control and/or treatment arms.[8]

D Survey of ClinicalTrials.gov Literature

Assessments of ClinicalTrials.gov provide a useful contrast between economics and medical disciplines. Since ClinicalTrials.gov (launched in February 2000) has a much longer history than the AEA RCT Registry, these assessments may also provide hints about how the AEA registry could evolve going forward. Unfortunately, previous studies show that ClinicalTrials.gov has foundational problems similar to the AEA registry.

First, ClinicalTrials.gov, by itself, does not capture a census of all relevant trials. In an early survey of industry-sponsored phase III drug trials, Manheimer and Anderson (2002) found that 25% of prostate cancer drug trials and 40% of colon cancer drug trials failed to register with ClinicalTrials.gov (or any other applicable registry). Dickersin and Rennie (2003) raised similar concerns for academic trials. In response to this issue, the International Committee of Medical Journal Editors (ICMJE) mandated that clinical trials register before the onset of patient enrollment as a condition of consideration for publication.[9] This policy change provides a rough upper bound on the voluntary registration rate. Zarin et al. (2007) document that ClinicalTrials.gov received an average of 30 new registrations per week prior to the full implementation of the ICMJE policy in September 2005 and 220 new registrations per week after. These values imply that fewer than 14% of all clinical trials voluntarily registered with ClinicalTrials.gov.[10]

[8] This issue was raised to us anonymously after we first circulated this paper. A specific concern is that the public might inadvertently discover the RCT through web searches for the researchers' names and historical paper titles.

[9] The policy required new trials to preregister from July 1, 2005 on and existing trials to register by September 13, 2005. The policy did not specify a required registry, but the announcement noted that only ClinicalTrials.gov currently fulfilled the ICMJE's specifications. See DeAngelis et al. (2005).

[10] 14% is likely a high upper bound because the ICMJE policy does not impact most industry-sponsored trials. Also, enforcement of the ICMJE policy increased over time. Mathieu et al. (2009) find a 73% registration rate for trials in three medical areas (cardiology, rheumatology, and gastroenterology) indexed in the ten general medical journals and specialty journals with the highest impact factors in 2008. Meanwhile, Huser and Cimino (2013) find a 96%

Second, many trials that do register do not provide sufficient information. Zarin et al. (2011) examine the primary outcome measures from 100 randomly selected non–phase I trials that registered with ClinicalTrials.gov in August 2010 and find that 61% lacked either a specific metric and/or time frame. Zarin et al. (2017) repeat this analysis for 80 articles published in the New England Journal of Medicine and the Journal of the American Medical Association over 2015-2016 and find that 42.6% of the primary outcomes listed in the associated ClinicalTrials.gov registrations lacked either a specific metric and/or time frame.[11] More surprisingly, even basic ClinicalTrials.gov information fields are often completed incorrectly. Chaturvedi et al. (2019) survey registrations over 2005-2015 and find that 17% of the listed primary investigator names are not those of real persons, but instead, to use their term, "junk information."

Third, most registered trials fail to report their results. ClinicalTrials.gov launched a results database in September 2008 to implement Section 801 of the Food and Drug Administration Amendments Act of 2007 (FDAAA), which requires the submission of "basic results" for most clinical trials of drugs and biologics within one year of their completion.[12] Despite this law, Law, Kawasumi and Morgan (2011) find that fewer than 13% of relevant registered trials completed between October 2008 and May 2010 reported results on time. Prayle, Hurley and Smyth (2012) and Anderson et al. (2015) show similarly poor reporting compliance rates for registered trials that completed in 2009 and over 2008-2012 respectively. Examining longer time frames, Nguyen et al. (2013) note that 50% of cancer drug trials failed to report results three years after completion. And Fain et al. (2018) find that 25% of industry-sponsored trials failed to report results even seven years after completion.[13] Adda, Decker and Ottaviani (2020) show an excess in the number of significant results in Phase III investigation relative to Phase II investigations for small industry sponsors; they argue this is consistent with the selective reporting of results.

Finally, when registered trials do report results these often differ from the published results. Hartung et al. (2014) explore these inconsistencies by taking a 10% random sample of Phase III

registration rate for trials published in five ICMJE founding journals over 2010-2011.

[11]The authors also find that 33% of the trials that registered over 2012-2014 registered more than three months after their start date.

[12]The FDAAA also mandates the registration of most non-phase I trials of FDA-regulated drug, biological, and device products.

[13]In a partial counterpoint, Oostrom (2024) finds that requirements to preregister psychiatric drug trials with ClinicalTrials.gov help limit the effect of financial sponsorship on reported drug efficacy via capturing negative results.

and IV trials that both proceeded to publication and reported results on ClinicalTrials.gov before January 1, 2009. The authors find that 80% were inconsistent in the number of secondary outcomes considered, 35% inconsistently stated the number of individuals with a serious adverse event, 20% had inconsistencies in a primary outcome value, and 15% described a primary outcome inconsistently. Becker et al. (2014) similarly find that nearly all trials published in high-impact journals that reported results on ClinicalTrials.gov had a least one significant discrepancy. Perhaps more ominously, Earley, Lau and Uhlig (2013) highlight differences between the number of deaths reported on ClinicalTrials.gov and in corresponding published papers.

E Background on Sample Generation

Here we describe the process we followed to generate the data used in Section 3.2.2 to assess the extent to which registrations with the AEA RCT Registry are sufficiently detailed to limit p-hacking.

- Initial Draft (April 2019): We took a 20% sample of registrations. Specifically, we selected 300 out of the 1,527 registrations through the end of 2017. We left a one plus year post period in order to ensure time post registration for all papers over which to observe follow-up and posting of interim and or final results.

- First Revision (January 2022): We added a 20% sample from the population of new registrations — selecting 600 out of 3,188 registrations from the start of 2018 to the end of 2021.

- Current Revision (June 2024): The years 2022 and 2023 added an additional 2,711 registrations. Due to limited time and resources, it was impractical to evaluate a 20% random sample of the additional 2,711 registrations; thus, we randomly selected 100 registrations — bringing the total of evaluated registrations to 1,000.

All sampling was done in Python using the Pandas sample function and specifying the number of selected papers and a random state (choosing random state = 1).

We also detail the process used to generate the data in Section 3.2.4 to statistically assess the extent to which registration limits p-hacking.

- First Revision (September 2022): We added 3.2.4 for the first revision of the paper. The analysis was conducted in September 2022 using a sample of 60 registered published papers and 60 unregistered published papers randomly selected from the universe considered in the file drawer exercise (RCTs published in top journals between 2017 and 2021).

- Current Revision (September 2024): We added an additional 40 registered papers and 40 unregistered papers from the universe considered in the file drawer exercise subset to papers published in 2022 and 2023.

When initially selected papers we aimed to have an equal number of registered and unregistered papers, but in the course of checking our data we found 3 papers that had been incorrectly labeled as registered. The results from the empirical analysis with the original dataset we collected for the first revision is included in Appendix J.

F Existence and Construction of Partitional Equilibria

This Appendix presents results on the existence and construction of partitional equilibria. Parititonal equilibria require three "threshold signal realisations" corresponding to the first-period registration decision, the first-period experimentation decision, and the second-period registration decision. We discuss these in reverse order.

We start with the second-period registration decision. If late registration is allowed, then as discussed in Section 5.1.3, *all* equilibria under Assumption 1 where outsider beliefs do not respond to registration decisions will involve a threshold $s^*_{2,R}$. If late registration is banned, then this is equivalent to forcing $s^*_{2,R} = \overline{s}$.

Next, we show that, under the assumptions we have imposed, the decision of whether or not to undertake the experiment will also be partitional. Since the decision to not experiment delivers a fixed payoff of 0, we can show that more optimistic researchers will be more eager to experiment no matter what their equilibrium registration decision would be:

Lemma F.1. *Under Assumption 1, the first-period experimentation decision takes a partitional form,*[14] *both when late registration is allowed and when it is banned.*

Thus, the existence of thresholds $s^*_{2,R}$ and $s_{1,\emptyset}$ holds generally.

What is more restrictive is the existence of a threshold $s^*_{1,R}$. We alluded to these difficulties in Section 5.1.4, where we pointed out that the gain to registration may be decreasing in s_1. Our intuition suggests this should be unusual. Specifically, if a researcher *always* expected to register with a very high s_1, then the option value associated with delayed registration would be lower than for a researcher with a lower s_1. The complication is that such a researcher may be so confident that s_2 will be very favorable that she sees no need to register early.

It need not be the case that partitional equilibria exist, even under Assumption 1, and we are not aware of existing conditions which would deliver it in our setting. One issue is the endogeneity of the outsider's beliefs. Another is the need to impose conditions on the informational environment

[14] By "takes a partitional form," we mean that there exists some signal realisation in $[\underline{s}_1, \overline{s}_1]$ such that one choice is made on one side of the threshold, and the other choice is made on the other side. Here, this means that there exists some $s^*_{1,\emptyset} \in [\underline{s}_1, \overline{s}_1]$ such that the researcher experiments whenever $s_1 > s^*_{1,\emptyset}$.

and not simply payoffs alone (as in Assumption 1), since the informational environment influences the researcher's expected payoffs.[15] We now present results ensuring the first-period registration decision is partitional. We first describe an increasing differences condition that we use in our analysis:

Definition 1. *Let* $\mathbb{E}[g(s_2 \mid \theta) \mid s_1]$ *denote the expected value of* $g(s_2 \mid \theta)$ *given signal* s_1 *(emphasising that this expectation is taken with respect to both* s_2 *and* θ*). We say that* $\hat{p}(s_2, d)$ *satisfies increasing gains to early registration if:*

$$\int_{-\infty}^{\infty} (b_R(\hat{p}(s_2, 1)) - b_R(\hat{p}(s_2, 2)))\mathbb{E}[g(s_2 \mid \theta) \mid s_1]ds_2 \tag{1}$$

is increasing in s_1*.*

Notably, this condition does not depend on b_N; or, for that matter, anything related to the researcher's decision at time 2 other than her beliefs (which are pinned down after time 1). We briefly mention that this property will turn out to be useful in our calibration exercise. We also note that whether this holds *in equilibrium* may depend on the researcher's strategy, through its influence on $\hat{p}(s_2, d)$. This condition says that if a researcher were to register, beliefs are such that it is even better to register early rather than late when the initial signal realisation is higher. This seems to be the practically relevant case since, as mentioned above, it appears researchers with more favorable results are generally more eager to register early.

Our interest in this condition can be seen from the following Lemma. First, it shows the existence of an $s^*_{1,R}$ threshold, whenever $\hat{p}(s_2, d)$ satisfies increasing gains to early registration. Second, indifference conditions pin down the relevant thresholds:

Lemma F.2. *Suppose* $\hat{p}(s_2, d)$ *satisfies increasing gains to early registration, with both early and delayed registration being on-path. Then under Assumption 1, the first-period registration decision must be of a partitional form (see Footnote 14). Furthermore, given partition thresholds* $\underline{s}_1 \leq s^*_{1,\emptyset} < s^*_{1,R} < \overline{s}_1$*, if the increasing gains to early registration condition is satisfied for the induced*

[15] While higher first-period signals may be associated with higher beliefs (and thus a larger gain in expectation), they are *also* associated with higher second-period signal realisations, and this may dampen the gain to signalling. For intermediate signals, signalling could be a powerful motivator, but not for higher signals. See Feltovich, Harbaugh and To (2002) for an exposition of a signalling model with "countersignalling" equilibria.

$\hat{p}(s_2, d)$ *and the researcher is indifferent between (i) "experiment" and "don't experiment" at* $s^*_{1,\emptyset}$ *and "register at time 1" and (ii) "don't register at time 1" at* $s^*_{1,R}$*, then the thresholds are part of a partitional equilibrium (together with the appropriate* $s^*_{2,R}$*).*

At a high level, this Lemma shows increasing gains to early registration is a "global condition" on the receiver's beliefs which is sufficient for the "local conditions" for optimality—namely, indifference at each threshold—to define an equilibrium. If it is violated, then it is possible that the thresholds make the researcher indifferent between actions, but that some "non-threshold" s_1 would prefer to deviate from their prescribed action. This is precisely what single-crossing conditions provide in signalling models. Increasing gains to early registration is thus the relevant single-crossing condition in our environment.

To summarise, there are two main reasons the notion of increasing gains to early registration is useful. First, it justifies our focus on partitional equilibria, given our intuition that researchers *should* feel an increased eagerness to register early when they are more optimistic about θ. Second, it is simpler to check than the overall increasing differences condition. One way that this can be seen is that it does not require us to compute s^*_2, whereas overall monotonicity does. Note that while it will typically be straightforward to find *beliefs* such that $b_R(\hat{p}(s_2, 2)) - c_R = b_N(p(s_2, \emptyset))$, these beliefs will depend on the registration threshold $s^*_{1,R}$, and so s^*_2 will depend on $s^*_{1,R}$ as well.

We put these Lemmata together to deliver the following pair of results related to the existence of partitional equilibria. The proofs reduce to checking increasing gains to early registration:

Proposition 2. *Fix an informational environment where* $f(s_1 \mid \theta), g(s_2 \mid \theta) > c > 0$*, for* $s_1 \in [\underline{s}_1, \overline{s}_1]$*,* $s_2 \in [\underline{s}_2, \overline{s}_2]$*,* $\theta \in \{T, F\}$ *and some* $c > 0$*. Consider any family of* b_R, b_N, c_R *where (a)* b_R *is three-times differentiable with* $b'''_R \geq 0$ *and, (b) for some fixed* ε*, letting* $q(\emptyset, s_2)$ *be the probability that* $\theta = T$ *following signal* s_2 *alone,*

$$\mathbb{E}_{s_2}[b_N(q(\emptyset, s_2))] > \mathbb{E}_{s_2}[b_R(q(\emptyset, s_2))] - c_R + \varepsilon.$$

There exists η *such that every equilibrium is partitional whenever* $\frac{b''_R(p)}{b'_R(p)} > \eta$*.*

Our second application of the increasing gains to early registration condition is an existence result

explicitly for our numerical calibration. The usefulness of this result comes from its implication that finding indifference thresholds suffices to determine equilibria:

Proposition 3. *Consider the informational environment in Section 6.2 with* $\underline{s}_2 = 0$*, and suppose* $b_R(\hat{p}) = \hat{p}$*. A partitional equilibrium strategy exists for* $p_0 \in [0, 0.3]$ *and* $\underline{s}_1 \in [0.2, 0.5]$.

We note that the conclusion Proposition 3 holds for a larger set of parameters than stated, but this region is sufficient to cover our calibration exercise. To conclude and summarise, we briefly describe how we compute the threshold $s^*_{1,R}$. The procedure is simple, and illustrated in Figure VII.[16] For each signal s_1, we conjecture that $s^*_{1,R} = s_1$, and then given this conjecture, compute the payoff from (a) preregistering and (b) experimenting without registration. We then compute $s^*_{1,R}$ to be the intersection of these lines. Thanks to Proposition 3, this procedure guarantees we have found an equilibrium. And indeed, the intersection point of the two lines in Figure VII is unique and therefore pins down a unique interior $s^*_{1,R}$ for this environment.

[16]This figure uses the same informational environment described in our numerical calibration.

G Proofs

This appendix is organised as follows. In Section G.1, we provide technical lemmas which are useful for our analysis. Interested readers are encouraged to skip this section and instead refer to it as needed. We then present the proofs from the main text in Section G.2. We conclude with some additional discussion referenced in the main text, while elaborating on some of the model subtleties, in Section G.3.

G.1 Preliminary Results

Lemma G.1. *$q(s_1, \emptyset)$ is strictly increasing in s_1, and $q(s_1, s_2)$ strictly increasing in s_2 for all s_1.*

Proof. Standard and thus omitted. □

Lemma G.2. *In any equilibrium, $\hat{p}(s_2, 1)$ is strictly increasing in s_2 if $d = 1$ is on-path.*

Proof of Lemma G.2. First consider the fictitious environment where s_1 were observable to the outsider, so that his beliefs are $q(s_1, s_2)$. Differentiating $q(s_1, s_2)$, we have that it is proportional to:

$$g'(s_2 \mid T)f(s_1 \mid T)\mathbb{P}[T]\cdot g(s_2 \mid F)f(s_1 \mid F)\mathbb{P}[F] - g'(s_2 \mid F)f(s_1 \mid F)\mathbb{P}[F]g(s_2 \mid T)f(s_1 \mid T)\mathbb{P}[T].$$

Note that we can rewrite this as

$$f(s_1 \mid T)\mathbb{P}[T]f(s_1 \mid F)\mathbb{P}[F])(g(s_2 \mid T)g(s_2 \mid F)) \cdot \left(\frac{g'(s_2 \mid T)}{g(s_2 \mid T)} - \frac{g'(s_2 \mid F)}{g(s_2 \mid F)}\right),$$

which must be strictly greater than 0 for all $s_1 \in (\underline{s}_1, \overline{s}_1)$, since $\frac{d}{ds_2} \log g(s_2 \mid T) > \frac{d}{ds_2} \log g(s_2 \mid F)$, and since all other densities and probabilities are positive as well. Thus, $q(s_1, s_2)$ is strictly increasing in s_2.

We now consider $\hat{p}(s_2, 1)$ instead of $q(s_1, s_2)$. Let $\sigma(d \mid s_1)$ denote the probability the researcher chooses d after observing signal s_1. If $d = 1$ is on-path, then:

$$\hat{p}(s_2, 1) = \frac{g(s_2 \mid T) \int_{\underline{s}_1}^{\overline{s}_1} \sigma(1 \mid s_1) f(s_1 \mid T) p_0 ds_1}{g(s_2 \mid T) \int_{\underline{s}_1}^{\overline{s}_1} \sigma(1 \mid s_1) f(s_1 \mid T) p_0 ds_1 + g(s_2 \mid F) \int_{\underline{s}_1}^{\overline{s}_1} \sigma(1 \mid s_1) f(s_1 \mid F)(1 - p_0) ds_1}.$$

From inspection, we see that the expression for $\hat{p}(s_2, 1)$ is the same as the expression for $q(s_1, s_2)$, replacing $f(s_1 \mid \theta)$ with $\int_{\underline{s}_1}^{\overline{s}_1} \sigma(1 \mid s_1) f(s_1 \mid \theta) ds_1$. In the calculation of the derivative of $q(s_1, s_2)$, this comes out as a constant, so an identical calculation proves the Lemma. □

Lemma G.3. *There exists an equilibrium of the game in Section 5.1; in particular, there exists a PBE where $\hat{p}(s_2, 2) = \hat{p}(s_2, \emptyset)$ (i.e., where the outsider's belief does not depend on the second-period registration decision), and where $\hat{p}(s_2, d) \in [q(\underline{s}_1, s_2), q(\overline{s}_1, s_2)]$.*

Proof. We present the proof for the case when late registration is not banned; the argument for when it is is identical. We consider a slightly different game than the one presented in the main text; the auxiliary game coincides with the one from the main text, except (a) we force the condition that $\hat{p}(s_2, 2) = \hat{p}(s_2, \emptyset)$, which restricts off-path beliefs, and (b) we explicitly make the game static so that the existence theorem of Milgrom and Weber (1985) can be applied. However, we then argue that the equilibrium of this auxiliary game also defines an equilibrium of the game in Section 5.1.

Specifically, assume that (a) the second-period registration decision is unobserved, and (b) the outsider chooses an action $a = (a_y, a_n) \in [q(\underline{s}_1, \emptyset), q(\overline{s}_1, \emptyset)] \times [q(\underline{s}_1, \emptyset), q(\overline{s}_1, \emptyset)]$ at the same time as the researcher. Specifically:

- There is a type of nature $(\theta, s_2) \in \{1, 0\} \times [\underline{s}_2, \overline{s}_2]$, where $\theta = 1$ corresponds to the event that $\theta = T$ in the main model.
- The outsider has no private information, and chooses an action to maximise:

$$u_O(a, d_R, \theta, s_1, s_2) = -(a_y - \theta)^2 \mathbf{1}[d_R = Y] - (a_n - \theta)^2 \mathbf{1}[d_R = N].$$

- The researcher chooses, as a function of s_1, $d_R \in \{0, Y, N\}$; payoffs are:

$$
\begin{aligned}
u_R(a, Y, \theta, s_1, s_2) &= b_R\left(\frac{a_y g(s_2 \mid T)}{a_y g(s_2 \mid T) + (1-a_y) g(s_2 \mid F)}\right) - c_R - c_E, \\
u_R(a, N, \theta, s_1, s_2) &= \max\left\{ b_R\left(\frac{a_n g(s_2 \mid T)}{a_n g(s_2 \mid T) + (1-a_n) g(s_2 \mid F)}\right) - c_R, \right. \\
&\qquad \left. b_N\left(\frac{a_n g(s_2 \mid T)}{a_n g(s_2 \mid T) + (1-a_n) g(s_2 \mid F)}\right)\right\} - c_E, \\
u_R(a, 0, \theta, s_1, s_2) &= 0.
\end{aligned}
$$

It is straightforward to apply the existence theorem of Milgrom and Weber (1985) to this game, for every ε; (a) since type spaces and action spaces are compact and payoff functions are continuous, utility functions are uniformly continuous, and (b) since the marginal distributions are supported on $\{0,1\}, [\underline{s}_1, \overline{s}_1]$ and $[\underline{s}_2, \overline{s}_2]$, and since the joint distribution is supported on $\{0,1\} \times [\underline{s}_1, \overline{s}_1] \times [\underline{s}_2, \overline{s}_2]$, the absolute continuity condition is satisfied as well.

Now, given the equilibrium of this game, we set:

$$
\hat{p}(s_2, 1) = \frac{a_y g(s_2 \mid T)}{a_y g(s_2 \mid T) + (1-a_y) g(s_2 \mid F)}, \;\; \hat{p}(s_2, 2) = \hat{p}(s_2, \emptyset) = \frac{a_n g(s_2 \mid T)}{a_n g(s_2 \mid T) + (1-a_n) g(s_2 \mid F)}.
$$

Note that under this definition, $\hat{p}(s_2, 1)$ is the posterior belief that $\theta = T$ given prior belief a_y and signal s_2; similarly, $\hat{p}(s_2, \emptyset)$ is the posterior belief that $\theta = T$ given prior belief a_n and signal s_2. The registration strategy of the researcher and belief profile of the outsider induced by this pair form an equilibrium in the model of Section 5.1. □

Lemma G.4. *Suppose late registration is allowed. In any equilibrium under Assumption 1 where the researcher registers early with probability less than 1, then there either exists a single threshold s_2^* such that a researcher who has not registered at time 1 will do so at time 2 if $s_2 > s_2^*$; in particular, the late-registration decision is deterministic (except possibly at s_2^*).*

Proof of Lemma G.4. Let $\hat{p}(s_2, d \neq 1)$ denote the outsider's belief after observing the researcher *did not* register at time 1, but before seeing whether late registration occurred or not. The same

argument from Lemma G.2 implies that $\hat{p}(s_2, d \neq 1)$ is (strictly) increasing in s_2, since we assume this is on-path. But since $\hat{p}(s_2, 2) = \hat{p}(s_2, \emptyset)$, and since we must have

$$\hat{p}(s_2, d \neq 1) = \mathbb{E}_d[\hat{p}(s_2, d)],$$

we conclude that $\hat{p}(s_2, d \neq 1) = \hat{p}(s_2, 2) = \hat{p}(s_2, \emptyset)$, so that the same holds if we instead considered $\hat{p}(s_2, 2)$ (or $\hat{p}(s_2, \emptyset)$).

Consider any signal s_2 where the researcher were to mix over the registration decision. At any such signal, we must have $b_R(\hat{p}(s_2, 2)) - c_R = b_N(\hat{p}(s_2, \emptyset))$, since otherwise there would be a strict incentive to deviate. Since $b_R(\hat{p}) - b_N(\hat{p})$ is strictly increasing by Assumption 1, and since $\hat{p}(s_2, 2)$ is strictly increasing in s_2 as well, there is either a unique signal where this indifference is satisfied. Then the researcher finds it strictly optimal to register if $s_2 > s_2^*$ and not if $s_2 < s_2^*$, so that the researcher's action is a deterministic function of s_2 for all s_2 except possibly s_2^*. In fact, recall that $\hat{p}(s_2, 2) = \hat{p}(s_2, \emptyset)$ is also continuous (in fact, differentiable) in s_2. Thus, if $b_R(\hat{p}(s_2, 2)) - c_R - b_N(\hat{p}(s_2, \emptyset))$ is neither always positive nor always negative, then by the intermediate value theorem, there is some belief in the range of possible second-period beliefs where this is equal to 0; since this belief must be unique, we can set the signal inducing it equal to s_2^*. If $b_R(\hat{p}(s_2, 2)) - c_R - b_N(\hat{p}(s_2, \emptyset))$ is always positive or always negative, then the equilibrium is trivially deterministic and characterised by a threshold outside of the support of s_2. □

Lemma G.5. *Consider the distribution of s_2 conditional on the researcher's belief after s_1; that is, where s_2 has density:*

$$q(s_1, \emptyset)g(s_2 \mid T) + (1 - q(s_1, \emptyset))g(s_2 \mid F)$$

This distribution is FOSD increasing in s_1.

Proof. Recall that $g(\cdot \mid T)$ first-order stochastically dominates $g(\cdot \mid F)$, so that given any increasing function $u(s_2)$, we have $\int_{s_2} u(s_2)g(s_2 \mid T)ds_2 \geq \int_{s_2} u(s_2)g(s_2 \mid F)ds_2$. Furthermore, $q(s_1, \emptyset)$ is increasing in s_1, meaning that if $s_1' > s_1''$, we also have

$$\int_{s_2} u(s_2)(q(s_1', \emptyset)g(s_2 \mid T) + (1 - q(s_1', \emptyset))g(s_2 \mid F))ds_2 \geq$$
$$\int_{s_2} u(s_2)(q(s_1'', \emptyset)g(s_2 \mid T) + (1 - q(s_1'', \emptyset))g(s_2 \mid F))ds_2$$

so that $s_1' > s_1''$ implies $q(s_1', \emptyset)g(s_2 \mid T) + (1 - q(s_1', \emptyset))g(s_2 \mid F)$ first order stochastically dominates $q(s_1'', \emptyset)g(s_2 \mid T) + (1 - q(s_1'', \emptyset))g(s_2 \mid F)$, as claimed. □

G.2 Proofs from the Main Text

G.2.1 Proof of Proposition 1

We recall that the proof allows us to select c_R and c_E freely. The proof proceeds in two steps:

1. We exhibit a range of $[\underline{c}_R, \overline{c}_R]$ with $c_E = 0$ satisfying the conclusion of the proposition in the special case that $f(s_1 \mid T) = f(s_1 \mid F)$—in which case, the maximal influence of past expertise is 0. Note that in this case, while we will violate the assumption that $\frac{d \log f(s_1|T)}{ds_1} > \frac{d \log f(s_1|F)}{ds_1}$, this assumption will not play a role in the proof.

2. We then argue that when the maximal influence of past expertise is sufficiently small, the same conclusions can be reached, potentially changing the values for $\overline{c}_R$ and $\underline{c}_R$ slightly, even allowing for a range of possible values for c_E.

Step one: Since $q(\underline{s}_1, s_2) = q(\overline{s}_1, s_2)$ if $f(s_1 \mid T) = f(s_1 \mid F)$ for all s_1, our equilibrium concept pins down $\hat{p}(s_2, d)$ uniquely. Letting $q^*(s_2) := q(\underline{s}_1, s_2)$, we note that equilibrium requires $q(\underline{s}_1, s_2) = \hat{p}(s_2, d)$. Also note that the researcher's belief that $\theta = T$ after observing s_1 coincides with the prior, p_0, and that $\hat{p}(s_2, d)$ must be strictly increasing in s_2 by Lemma G.2

If late registration is banned, then the researcher is willing to register whenever:

$$\int_{\underline{s}_2}^{\overline{s}_2} (b_R(q^*(s_2)) - c_R)(p_0 g(s_2 \mid T) + (1-p_0)g(s_2 \mid F))ds_2$$
$$\geq \int_{\underline{s}_2}^{\overline{s}_2} b_N(q^*(s_2))(p_0 g(s_2 \mid T) + (1-p_0)g(s_2 \mid F))ds_2,$$

or, rearranging:

$$\int_{\underline{s}_2}^{\overline{s}_2} (b_R(q^*(s_2)) - b_N(q^*(s_2)))(p_0 g(s_2 \mid T) + (1-p_0)g(s_2 \mid F))ds_2 \geq c_R. \tag{2}$$

Now suppose that late registration is allowed. Let $\hat{s}_2(c_R)$ denote a value of s_2 which is indifferent between registering late and not registering if some such value exists, i.e., the implicit solution to:

$$b_R(q^*(\hat{s}_2(c_R))) - c_R = b_N(q^*(\hat{s}_2(c_R)));$$

If $b_R(q^*(s_2)) - c_R < b_N(q^*(s_2))$ for all s_2, then set $\hat{s}_2(c_R) = \overline{s}_2$; if $b_R(q^*(s_2)) - c_R > b_N(q^*(s_2))$ for all s_2 then set $\hat{s}_2(c_R) = \underline{s}_2$. Note that in particular since $q^*(s_2)$ is strictly increasing in s_2 (since we assume $\frac{d\log(g(s_2|T))}{ds_2} > \frac{d\log(g(s_2|F))}{ds_2}$) and $b_R(p) - b_N(p)$ is strictly increasing in p there is exactly one value for $\hat{s}_2(c_R)$ thus defined. Also note that, by the assumed smoothness properties of g, $q^*(s_2)$ must be continuous; using the continuity of b_R and b_N, we have that $\hat{s}_2(c_R)$ is continuous as well, a fact that will be useful later.

The researcher will strictly prefer to not pre-register if:

$$\int_{\underline{s}_2}^{\overline{s}_2} (b_R(q^*(s_2)) - c_R)(p_0 g(s_2 \mid T) + (1-p_0)g(s_2 \mid F))ds_2$$
$$< \int_{\underline{s}_2}^{\hat{s}_2(c_R))} b_N(q^*(s_2))(p_0 g(s_2 \mid T) + (1-p_0)g(s_2 \mid F))ds_2$$
$$+ \int_{\hat{s}_2(c_R)}^{\overline{s}_2} (b_R(q^*(s_2)) - c_R)(p_0 g(s_2 \mid T) + (1-p_0)g(s_2 \mid F))ds_2,$$

or, rearranging:

$$\int_{\underline{s}_2}^{\hat{s}_2(c_R)} (b_R(q^*(s_2)) - c_R - b_N(q^*(s_2)))(p_0 g(s_2 \mid T) + (1-p_0)g(s_2 \mid F))ds_2 < 0. \tag{3}$$

Indeed, this condition is satisfied whenever $\hat{s}_2(c_R) > \underline{s}_2$, since by definition of $\hat{s}_2(c_R)$ and increasing differences, $b_R(q^*(s_2)) - c_R < b_N(q^*(s_2))$ whenever $s_2 < \hat{s}_2(c_R)$, which occurs with positive probability.

Define c_R^* to be the value of c_R such that (2) holds with equality. Let $\tilde{c}_R$ be the supremum over the set of values of c_R such that $\hat{s}_2(c_R) = \underline{s}_2$. We note that $\tilde{c}_R < c_R^*$; indeed, since the integrand in (2) is strictly increasing by the conditions of the proposition, if (2) holds with equality at $c_R = c_R^*$ then we must have $b_R(q^*(\overline{s}_2)) - c_R^* > b_N(q^*(\overline{s}_2))$ and $b_R(q^*(\underline{s}_2)) - c_R^* < b_N(q^*(\underline{s}_2))$. Thus $\hat{s}_2(c_R^*) > \underline{s}_2$. Since decreasing c_R decreases $\hat{s}_2(c_R)$, we have that we must lower c_R by some amount from c_R^* in order to reach $\tilde{c}_R$. We note that the precise amount will depend on b_R, b_N, p_0 and g.

Pick $\underline{c}_R$ and $\overline{c}_R$ such that $\tilde{c}_R < \underline{c}_R < \overline{c}_R < c_R^*$. In this case:

- If late registration is banned, all researchers will strictly prefer to preregister—this holds since (2) holds with strict inequality.
- If late registration is allowed, all researchers will strictly prefer to delay registration—this holds by the observation that (3) holds since $\hat{s}_2(c_R) \in (\underline{s}_2, \hat{s}_2(c_R^*))$.
- After delaying their registration decision when late registration is allowed, not all researchers will register, again since $\hat{s}_2(c_R) > \underline{s}_2$, implying a positive probability of a signal realisation such that the researcher would not register.

Putting this together, we see that under a late registration ban, registration occurs with probability 1, whereas when late registration is allowed, not all researchers register. This shows the proof of the claim.

Step Two: Now we allow for choices of f which convey non-zero information about θ. In this case, equilibrium still allows us to bound the beliefs of the outsider, which are at worst $q(\underline{s}_1, s_2)$

and at best $q(\underline{s}_1, s_2) + \delta$. For the stated range of c_R, the researcher strictly preferred to register under a ban. We thus have that, for any equilibrium inference, a researcher observing signal s_1 would strictly prefer to register under a ban, for any equilibrium outsider beliefs, if the following condition held:

$$\int_{\underline{s}_2}^{\overline{s}_2} (b_R(q(\underline{s}_1, s_2)) - c_R)(q(s_1, \emptyset)g(s_2 \mid T) + (1 - q(s_1, \emptyset))g(s_2 \mid F))ds_2 \\ > \int_{\underline{s}_2}^{\overline{s}_2} b_N(q(\underline{s}_1, s_2) + \delta)(q(s_1, \emptyset)g(s_2 \mid T) + (1 - q(s_1, \emptyset))g(s_2 \mid F))ds_2,$$

since $q(\overline{s}_1, s_2) < q(\underline{s}_1, s_2) + \delta$. We note that as $\delta \to 0$, $q(\underline{s}_1, s_2) \to q^*(s_2)$ and $q(s_1, \emptyset) \to p_0$. Thus, if δ is sufficiently small, continuity of b_R and b_N imply that all researchers will still strictly prefer to delay registration under a ban, given that they did in the $\delta = 0$ limit, for the stated range of c_R.

As for when late registration is allowed, define $\hat{\hat{s}}_2(c_R)$ as the implicit solution to:

$$b_R(q(\underline{s}_1, \hat{\hat{s}}_2(c_R))) - c_R = b_N(q^*(\underline{s}_1, \hat{\hat{s}}_2(c_R)));$$

Note in particular that as $\delta \to 0$, we also have $\hat{\hat{s}}_2(c_R) \to \hat{s}_2(c_R)$, since $q(\underline{s}_1, s_2) \to q^*(s_2)$, $b_R(p)$ and $b_N(p)$ are continuous in p and satisfy strictly increasing differences, and $q(\underline{s}_1, s_2)$ and $q^*(s_2)$ are continuous and strictly increasing in s_2; thus, by taking δ sufficiently small (where the precise maximum value will be a function of all other parameters), any f inducing $q(\overline{s}, s_2) - q(\underline{s}, s_2) < \delta$ for all s_2 will induce a threshold $\hat{\hat{s}}_2(c_R)$ that is interior (i.e., in the interval $(\underline{s}_2, \overline{s}_2)$). At this point, similar arguments from the late-ban case imply that when late registration is allowed, all researchers prefer to delay registration if δ is sufficiently small. In particular, since $\hat{\hat{s}}_2(c_R)$ is interior, we have the probability that the researcher registers late is less than 1.

Putting this together, we see that the same interval will yield the desired conclusion of the proposition, provided the maximal influence of past expertise is not too large, even if it is non-zero.

Finally, note that since $b_N(p) \geq 0$, and provided no signal reveals the state (which, in particular,

is the case whenever the maximal influence of past expertise is sufficiently small), we have that the payoff from non-registration is bounded away from 0 (since $b_N(p)$ is greater than 0 for all p and strictly increasing). Thus, the researcher will always strictly prefer experimenting and not registering to not experimenting at all; this implies we can take $\overline{c}_E > 0$, completing the proof.

G.2.2 Proof of Lemma F.1

Consider the researcher's payoffs from early registration:

$$-c_R + \int_{\underline{s}_2}^{\overline{s}_2} b_R(\hat{p}(s_2, 1)) \left(q(s_1, \emptyset) g(s_2 \mid T) + (1 - q(s_1, \emptyset)) g(s_2 \mid F)\right) ds_2. \tag{4}$$

We write the payoff from late registration (using that $\hat{p}(s_2, 2) = \hat{p}(s_2, \emptyset)$) as:

$$\int_{\underline{s}_2}^{\overline{s}_2} \max\{(b_R(\hat{p}(s_2, 2)) - c_R), b_N(\hat{p}(s_2, 2))\} \left(q(s_1, \emptyset) g(s_2 \mid T) + (1 - q(s_1, \emptyset)) g(s_2 \mid F)\right) ds_2. \tag{5}$$

Since the researcher's payoffs in both expressions are increasing in the relevant $\hat{p}(s_2, d)$, which in turn is increasing in s_2, both expressions are increasing in s_2. Since these expressions are both expectations over s_2 conditional on s_1, by Lemma G.5, both of these expressions are increasing in s_1. Hence if some type s_1 does not prefer to undertake the experiment, then neither do any lower types, since this implies both of the expressions are negative at s_1 and are therefore also negative at higher s_1. Likewise, if some type s_1 prefers to undertake the experiment, then it means at least one of these is positive, and hence is also positive at higher s_1, as desired.

G.2.3 Proof of Lemma F.2

Recalling that $q(s_1, \emptyset)$ is the researcher's belief that $\theta = T$ following a first-period signal s_1, recall that the payoff from early registration is given by (4); we write the payoff for late registration slightly differently: Starting with

$$\int_{s^*_{2,R}}^{\bar{s}_2} \left(b_R(\hat{p}(s_2,2)) - c_R\right)\left(q(s_1,\emptyset)g(s_2 \mid T) + (1-q(s_1,\emptyset))g(s_2 \mid F)\right) ds_2$$
$$+ \int_{\underline{s}_2}^{s^*_{2,R}} b_N(\hat{p}(s_2,2))(q(s_1,\emptyset)g(s_2 \mid T) + (1-q(s_1,\emptyset))g(s_2 \mid F))ds_2,$$

if we add and subtract $\int_{\underline{s}_2}^{s^*_{2,R}} \left(b_R(\hat{p}(s_2,2)) - c_R\right)\left(q(s_1,\emptyset)g(s_2 \mid T) + (1-q(s_1,\emptyset))g(s_2 \mid F)\right) ds_2$ to this expression, we can alternatively write this as:

$$\int_{\underline{s}_2}^{\bar{s}_2} \left(b_R(\hat{p}(s_2,2)) - c_R\right)\left(q(s_1,\emptyset)g(s_2 \mid T) + (1-q(s_1,\emptyset))g(s_2 \mid F)\right) ds_2$$
$$+\int_{\underline{s}_2}^{s^*_{2,R}} \max\{0, (b_N(\hat{p}(s_2,2)) - (b_R(\hat{p}(s_2,2)) - c_R))(q(s_1,\emptyset)\}g(s_2 \mid T) + (1-q(s_1,\emptyset))g(s_2 \mid F))ds_2. \tag{6}$$

We subtract (6) from (4) and obtain:

$$\overbrace{\int_{\underline{s}}^{\bar{s}} \left(b_R(\hat{p}(s_2,1)) - b_R(\hat{p}(s_2,2))\right)\left(q(s_1,\emptyset)g(s_2 \mid T) + (1-q(s_1,\emptyset))g(s_2 \mid F)\right) ds_2}^{\dagger}$$
$$+\overbrace{\int_{\underline{s}}^{s^*_{2,R}} \left(b_R(\hat{p}(s_2,2)) - c_R\right) - b_N(\hat{p}(s_2,2))(q(s_1,\emptyset)g(s_2 \mid T) + (1-q(s_1,\emptyset))g(s_2 \mid F))ds_2}^{\dagger\dagger} > 0.$$

We wish to show that

- (i) if this holds for some s_1, then this also holds at any larger s_1, and
- (ii) if this is violated at some s_1 then it is violated at smaller s_1.

Note these immediately imply the equilibrium registration decision is characterised by a threshold. Note that this expression considers the difference as the sum of two terms: The first term is the *belief increase* (†) due to *pre*registration, and the second is the loss due to *option value* (††).

Increasing gains to early registration being satisfied immediately gives that $(\dagger)$ is increasing in s_1. We show that $(\dagger\dagger)$ is strictly increasing in s_1. First suppose that late registration is allowed. We first note that first order stochastic dominance is maintained under monotone transformations,[17] and that $\hat{p}(s_2, 2)$ is a monotone transformation of s_2, as argued in the proof of Lemma G.4 (following the proof of Lemma G.2). Note that since $b_R(\hat{p}(s_2, 2)) - c_R - b_N(\hat{p}(s_2, 2)) < 0$ if and only if $s_2 < s^*_{2,R}$, we can rewrite $(\dagger\dagger)$ as:

$$\int_{\underline{s}}^{\overline{s}} \min\{(b_R(\hat{p}(s_2, 2)) - c_R) - b_N(\hat{p}(s_2, 2)), 0\}(q(s_1, \emptyset)g(s_2 \mid T) + (1 - q(s_1, \emptyset))g(s_2 \mid F))ds_2.$$

Therefore, since $\min\{(b_R(\hat{p}(s_2, 2)) - c_R) - b_N(\hat{p}(s_2, 2)), 0\}$ is strictly increasing, this expression is the expectation of an increasing function of s_2. As a result, it increases (strictly) with strict first order stochastic dominance shifts in the distribution of s_2, and therefore by Lemma G.5, is strictly increasing s_1.

As for when late registration is banned, note that this is equivalent to setting $s^*_{2,R} = \overline{s}$. Again the integrand is strictly increasing in s_2, meaning the same reasoning as the previous case applies, completing the proof.

G.2.4 Proof of Proposition 2

We check the increasing gains to early registration condition in Definition 1, using an arbitrary candidate belief profile. The result then follows by invoking Lemma F.2.

First, note that, in any equilibrium under the conditions of the Proposition, we must have $\hat{p}(s_2, 1) > \hat{p}(s_2, \emptyset) + k$ for all s_2, for some constant k. To see this, note that if we had some equilibrium with $\hat{p}(s_2, 1) \leq \hat{p}(s_2, \emptyset)$ for *some* s_2, then we would also have the inequality for *all* s_2; indeed, both of these are obtained by Bayesian updating of the outsider's belief over θ conditional on d alone; thus, $\hat{p}(s_2, 1) \leq \hat{p}(s_2, \emptyset)$ holds if and only if the outsider's belief following $d = 1$ is

[17] For a quick proof for reference (since this property is important for the proof), note that if $\mathbb{P}[A \leq x] \leq \mathbb{P}[B \leq x]$, for all $x \in \mathbb{R}$, then for any monotone f we have $\mathbb{P}[f(A) \leq f(x)] \leq \mathbb{P}[f(B) \leq f(x)]$ for all x. Then we also have $\mathbb{P}[f(A) \leq y] \leq \mathbb{P}[f(B) \leq y]$, for all $y \in \mathbb{R}$—either y is in the image of f in which case this is immediate, or it is not in which case either both probabilities are equal to 0 or both probabilities are equal to 1.

lower than when observing no registration, which in turn implies the inequality for all s_2. [18] Now, note that the registration decision is a binary signal about $\theta = T$ to the outsider; in particular, if observing $d = 1$ is a "negative signal" then the complementary event must be a "positive signal," so that we would further have $\hat{p}(s_2, \emptyset) \geq q(\emptyset, s_2) \geq \hat{p}(s_2, 1)$. Therefore, not registering would be a profitable deviation, since we would have:

$$\mathbb{E}_{s_2}[b_N(\hat{p}(s_2, \emptyset))] \geq \mathbb{E}_{s_2}[b_N(q(\emptyset, s_2))] > \mathbb{E}_{s_2}[b_R(q(\emptyset, s_2))] - c_R + \varepsilon > \mathbb{E}_{s_2}[b_R(\hat{p}(s_2, 1))] - c_R$$

where this string of inequalities uses the monotonicity of payoffs and condition (b) in the proposition. Compactness of $[\underline{s}_2, \overline{s}_2]$ and continuity of the density functions implies that $\hat{p}(s_2, 1) - \hat{p}(s_2, \emptyset) > k > 0$, for some k, by the extreme value theorem.

Now, the assumption that the densities are differentiable implies that $\hat{p}(s_2, 1)$ and $\hat{p}(s_2, 2)$ are differentiable. Now, note that since $b_R''' \geq 0$, $b_R'(p + k) \geq b_R'(p) + k b_R''(p)$ by properties of convex functions. Furthermore, $\frac{d\hat{p}(s_2,1)}{ds_2}$ and $\frac{d\hat{p}(s_2,2)}{ds_2}$ are bounded from above and below, under the assumption that densities are uniformly bounded from below and continuously differentiable. So consider the derivative of $b_R(\hat{p}(s_2, 1)) - b_R(\hat{p}(s_2, 2))$. We have that:

$$\begin{aligned}\frac{d}{ds_2}(b_R(\hat{p}(s_2, 1)) - b_R(\hat{p}(s_2, 2)) &= b_R'(\hat{p}(s_2, 1))\frac{d\hat{p}(s_2, 1)}{ds_2} - b_R'(\hat{p}(s_2, 2))\frac{d\hat{p}(s_2, 2)}{ds_2} \\ &> b_R'(\hat{p}(s_2, 2))\left(\frac{d\hat{p}(s_2, 1)}{ds_2} - \frac{d\hat{p}(s_2, 2)}{ds_2}\right) + k b_R''(\hat{p}(s_2, 2))\frac{d\hat{p}(s_2, 1)}{ds_2}.\end{aligned}$$

Thus, the derivative is positive if:

$$\left(\frac{d\hat{p}(s_2, 1)}{ds_2} - \frac{d\hat{p}(s_2, 2)}{ds_2}\right) + k\frac{b_R''(\hat{p}(s_2, 2))}{b_R'(\hat{p}(s_2, 2))}\frac{d\hat{p}(s_2, 1)}{ds_2} > 0.$$

Since $\frac{d\hat{p}(s_2,1)}{ds_2}$ can be bounded below, and $\frac{d\hat{p}(s_2,1)}{ds_2}$ can be bounded above, this implies we have that

[18] If it is possible that $g(s_2 \mid \theta) = 0$, then some s_2 would reveal the state irrespective of the equilibrium. Thus, noting that $\hat{p}(s_2, 1) = \hat{p}(s_2, \emptyset)$ for some s_2 would not suffice to show that the same holds for all s_2. However, if all distributions have strictly positive support, then this is not possible. In other words, $\frac{p_L g(s_2|T)}{p_L g(s_2|T) + (1-p_L) g(s_2|F)} < \frac{p_H g(s_2|T)}{p_H g(s_2|T) + (1-p_H) g(s_2|F)}$ if and only if $p_H > p_L$ whenever $g(s_2 \mid T), g(s_2 \mid F) > 0$.

the derivative is positive if b_R''/b_R' is sufficiently large.

Therefore, the Proposition follows from invoking Lemma G.5; since this Lemma implies that an increase in s_1 leads to a FOSD shift in the conditional distribution of s_2, and since $b_R(\hat{p}(s_2,1)) - b_R(\hat{p}(s_2,2))$ is increasing in s_2, an increase in s_1 will therefore cause (1) to increase, as desired.

G.2.5 Proof of Proposition 3

As stated, we assume the particular informational environment from Example 6.2, demonstrate that increasing gains to early registration holds for any $\hat{p}(s_2,d)$ which emerges from a partitional equilibrium, and then note that the result follows from Lemma F.2.

We note that the proof of this result follows from a numerical verification of the condition, but we first describe how to determine some useful parameters. First, we write out the outsider's beliefs, conditional on the first-period signal being in some interval $[s_*,s^*]$ and the second-period signal being s_2. This is:

$$\begin{aligned}
&\frac{p_0\int_{s_*}^{s^*}2s_1 2s_2 ds_1}{p_0\int_{s_*}^{s^*}2s_1 2s_2 ds_1+(1-p_0)\int_{s_*}^{s^*}2(1-s_1)2(1-s_2)ds_1}\\
=&\frac{p_0((s^*)^2-(s_*)^2)s_2}{p_0((s^*)^2-(s_*)^2)s_2+(1-p_0)((1-s_*)^2-(1-s^*)^2)(1-s_2)}\\
=&\frac{p_0(s^*+s_*)s_2}{p_0(s^*+s_*)s_2+(1-p_0)(2-s_*-s^*)(1-s_2)}.
\end{aligned}$$

In a partitional equilibrium, the outsider's belief $\hat{p}(s_2,2)$ (or $\hat{p}(s_2,\emptyset)$) will be given by this expression, setting $s_*=\underline{s}_1$ and s^* to be the signal at which the researcher is indifferent between early registration decisions. Similarly, the outsider's belief $\hat{p}(s_2,1)$ will be given by this expression, setting s_* to be the signal at which the researcher is indifferent between early registration decisions, and $s^*=\overline{s}_1$.

Regarding the first-period beliefs of the researcher, we note that the expression for the first period is the same as the previous expression in the special case where $s_2=1/2$. Thus, the highest possible belief corresponds to the case where $s^*=s_*=1-\overline{s}$, meaning that the first-period belief is less than:

$$\frac{p_0(1-\overline{s})}{p_0(1-\overline{s})+(1-p_0)\overline{s}},$$

which approaches p_0 as $\overline{s} \to 1/2$ and 1 as $\overline{s} \to 0$.

Using these computations, we numerically verify that the increasing gains to early registration condition, Definition 1, is satisfied for all p_0 and possible first-period belief thresholds. This is done using Mathematica.[19] This code verifies that the gain to early registration is increasing in the belief that $\theta = T$ (which, in turn, holds if and only if (1) is satisfied, since this belief changes monotonically in s_1).

G.3 Additional Results and Discussion

G.3.1 Computing Equilibria

We describe how equilibria to the baseline model can be computed in the case where s_1 and s_2 follow the distribution in the example. We assume that experiment costs are sufficiently low so that researchers always choose to experiment. In this case, given a signal of s_1, the expected distribution over s_2 is:

$$\mathbb{E}[g(s_2 \mid \theta) \mid s_1] = 2s_2 \frac{p_0 s_1}{p_0 s_1 + (1-p_0)(1-s_1)} + 2(1-s_2)\left(1 - \frac{p_0 s_1}{p_0 s_1 + (1-p_0)(1-s_1)}\right).$$

Suppose the outsider conjectures the first-period registration threshold is $s^*_{1,R}$. In this case, given second-period signal s_2, we have the belief is (see the Proof of Proposition 3):

$$\hat{p}(s_2, 1) = \frac{p_0(s^*_{1,R} + \overline{s}_1)s_2}{p_0(s^*_{1,R} + \overline{s}_1)s_2 + (1-p_0)(2 - s^*_{1,R} - \overline{s}_1)(1-s_2)},$$
$$\hat{p}(s_2, 0) = \frac{p_0(s^*_{1,R} + \underline{s}_1)s_2}{p_0(s^*_{1,R} + \underline{s}_1)s_2 + (1-p_0)(2 - s^*_{1,R} - \underline{s}_1)(1-s_2)}$$

We note that the second-period threshold, $s^*_{2,R}$, is determined by the condition:

$$\hat{p}(s^*_{2,R}, 0) - c_R = k\hat{p}(s^*_{2,R}, 0),$$

[19] The relevant Mathematica code is available here, as well as on https://www.jonlib.com/working-papers

which yields a closed-form solution for $s^*_{2,R}$, as a function of $s^*_{1,R}$ and other parameters; letting $\zeta = \frac{c_R}{1-k}$, we have:

$$s^*_{2,R}(s^*_{1,R}) = \frac{\zeta(1-p_0)(2-s^*_{1,R}-\underline{s}_1)}{(1-\zeta)p_0(s^*_{1,R}+\underline{s}_1)+\zeta(1-p_0)(2-s^*_{1,R}-\underline{s}_1)}.$$

Using these observations, $s^*_{1,R}$ is pinned down by the condition:

$$\int_{\underline{s}_2}^{\overline{s}_2} \hat{p}(s_2,1)\mathbb{E}[g(s_2 \mid \theta) \mid s^*_{1,R}]ds_2 - c_R = \int_{\underline{s}_2}^{s^*_{2,R}(s^*_{1,R})} k\hat{p}(s_2,0)\mathbb{E}[g(s_2 \mid \theta) \mid s^*_{1,R}]ds_2 + \int_{s^*_{2,R}(s^*_{1,R})}^{\overline{s}_2} \hat{p}(s_2,1)\mathbb{E}[g(s_2 \mid \theta) \mid s^*_{1,R}]ds_2 - c_R. \quad (7)$$

where we emphasise that $s^*_{1,R}$ also appears in the expressions for $\hat{p}(s_2,1)$. If this equation is satisfied, then we know that a researcher with $s_1 > s^*_{1,R}$ prefers to register early and a researcher with $s_1 < s^*_{1,R}$ prefers to delay registration.

To solve for the threshold when late registration is banned, we can simply consider the same condition except setting $s^*_{2,R}(s^*_{1,R}) = \overline{s}_2$ (thus not allowing the researcher to choose to register in the second period). In that case, (7) still defines the relevant threshold $s^*_{1,R}$.

G.3.2 Justifying Increasing Differences

We first present some examples explaining why the increasing differences condition may emerge naturally:

Example 1. *Suppose that whether publication ultimately occurs only depends on $\hat{p}(s_2,d)$, with this probability being denoted by $\pi(\hat{p}(s_2,d))$ for an increasing function $\pi(\cdot)$. However, the ultimate venue depends on registration; the expected value of a registered publication is β_R and the expected value of a non-registered publication is β_N. In this case, the increasing difference condition is satisfied, since the difference in payoffs is $(\beta_R - \beta_N)\pi(\hat{p}(s_2,d))$.*

In the previous example, registration does not impact whether publication occurs, but it does impact the expected tier of the ultimate venue, for instance, due to the AEA requirement that

experiments register to be published. We can also consider the opposite case, where the tier of the final outcome is irrelevant, but registration leads to additional independent possibilities for publication (again, since more possible journals are available).

Example 2. *Normalise the benefit of publication to 1, but suppose that the probability of publication is* $1-(1-\pi(\hat{p}(s_2,d)))^{\beta}$*, where* $\beta=\beta_R$ *when registered and* $\beta=\beta_N$ *when not registered, where* $\beta_R>\beta_N$*, for a differentiable and increasing* $\pi(\cdot)$*. Taking derivatives and simplifying, we have that the increasing difference condition is satisfied whenever:*

$$\beta_R(1-\pi(\hat{p}(s_2,d)))^{\beta_R-1}>\beta_N(1-\pi(\hat{p}(s_2,d)))^{\beta_N-1}$$

The expression $\beta(1-\pi)^{\beta-1}$ *is increasing in* β*, for* $\pi\in(0,1)$*, whenever* $1+\beta\cdot\log(1-\pi)>0$*, which can be rewritten as* $\pi<1-e^{-1/\beta}$*. Hence, this is satisfied whenever the probability of publication is low, relative to the number of venues. Considering a case where* $\beta_R=5$ *and* $\beta_N=4$ *(e.g., reflecting one less publication opportunity and a strong preference for top 5 publications), increasing differences reduces to the requirement that* $\pi(\hat{p}(s_2,d))<.2$*; while whether this restriction is reasonable may very well depend on context, we note that overall publication rates are less than 10 %. While we expect the probability of publication to perhaps be higher than this for the most favorable possible results, one still expects this to hold for a large range of possible results.*

To emphasise, these examples are simply meant as a way to assist the reader in calibrating the increasing differences assumption. This assumption is standard in the signaling literature, and the complementarity may come from other sources not explicitly considered in the above examples.

G.3.3 On Social Welfare

To translate our results into welfare statements, we briefly describe a particular social welfare function that organises our thoughts about how policymakers may seek to design the registry. We have in mind that registration not only interacts with the researcher's decision, but has an added social benefit of increasing the probability that the findings are disseminated—i.e., solving the file

drawer problem.

To model this, we imagine that society benefits when the experiment $\{g_{\tilde{\gamma}}(\cdot \mid \theta)\}_{\theta \in \{T,F\}}$ is more informative, but that society can only become aware of an experimental result if it is (1) published or (2) registered. Publication depends on $\hat{p}(s_2, d)$, and so we take $q(\hat{p}(s_2, d))$ to be the probability society observes the experiment when the outsider's belief is $\hat{p}(s_2, d)$. Registration increases the probability such awareness occurs; if the experiment is registered and the belief is $\hat{p}(s_2, d)$, then this probability increases by $r(\hat{p}(s_2, d))$.

Let $I_\gamma(s_2)$ be the social value from observing outcome s_2, and let $h(s_2 \mid s_1) = p_0 f(s_1 \mid T) g_{\tilde{\gamma}}(s_2 \mid T) + (1 - p_0) f(s_1 \mid F) g_{\tilde{\gamma}}(s_2 \mid F)$ be the distribution over s_2 given s_1. Social welfare is:

$$W = \int_{s_1 \text{:Experiment Conducted}} \int_{s_2} (q(\hat{p}(s_2, d)) + \mathbf{1}[d \neq \emptyset : s_1, s_2] r(\hat{p}(s_2, d))) I_{\tilde{\gamma}}(s_2) h(s_1, s_2) ds_1 ds_2.$$

With this social welfare function, how might registration improve welfare? The short answer is by addressing p-hacking (i.e., increasing informativeness) and the file drawer problem (i.e., wider dissemination of results independent of publication). By increasing the informativeness of the experiment, welfare may be increasing in $\tilde{\gamma}$, as society gains more from more informative experiments. On the other hand, it may be that the value of $I_{\tilde{\gamma}}(s_2)$ does not align with the probability of publication—for instance, if negative results are valuable in aggregate even though publication is more likely for positive results. In this case, $r(\hat{p}(s_2, d))$ might be larger when $\hat{p}(s_2, d)$ is smaller, correcting a loss when negative results are not seen. Thus, effectively addressing the file drawer problem could create a welfare improvement. In our view, this welfare function captures the first-order issues relevant to the evaluation of registry design; while we do not doubt other issues are important in maximising the advancement of science, it is harder to see how a registry is well-suited to address them.[20]

[20]However, we do emphasise that registries, in turn, are one of a *very* small number of *concrete* policy proposals which have been implemented *widely*.

References

Adda, Jérôme, Christian Decker, and Marco Ottaviani. 2020. "P-hacking in clinical trials and how incentives shape the distribution of results across phases." *Proceedings of the National Academy of Sciences*, 117(24): 13386–13392.

Anderson, Monique L, Karen Chiswell, Eric D Peterson, Asba Tasneem, James Topping, and Robert M Califf. 2015. "Compliance with Results Reporting at ClinicalTrials.gov." *New England Journal of Medicine*, 372(11): 1031–1039.

Andrews, Isaiah, and Maximilian Kasy. 2019. "Identification of and Correction for Publication Bias." *American Economic Review*, 109(8): 2766–94.

Becker, Jessica E, Harlan M Krumholz, Gal Ben-Josef, and Joseph S Ross. 2014. "Reporting of Results in ClinicalTrials.gov and High-Impact Journals." *JAMA*, 311(10): 1063–1065.

Burlig, Fiona. 2018. "Improving transparency in observational social science research: A pre-analysis plan approach." *Economics Letters*, 168: 56–60.

Chaturvedi, Neha, Bagish Mehrotra, Sangeeta Kumari, Saurabh Gupta, HS Subramanya, and Gayatri Saberwal. 2019. "Some Data Quality Issues at ClinicalTrials.gov." *Trials*, 20(1): 378.

DeAngelis, Catherine D, Jeffrey M Drazen, Frank A Frizelle, Charlotte Haug, John Hoey, Richard Horton, Sheldon Kotzin, Christine Laine, Ana Marusic, A John PM Overbeke, et al. 2005. "Clinical trial registration: a statement from the International Committee of Medical Journal Editors." *Archives of dermatology*, 141(1): 76–77.

Dickersin, Kay, and Drummond Rennie. 2003. "Registering Clinical Trials." *JAMA*, 290(4): 516–523.

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." National Bureau of Economic Research.

Earley, Amy, Joseph Lau, and Katrin Uhlig. 2013. "Haphazard Reporting of Deaths in Clinical Trials: A Review of Cases of ClinicalTrials.gov Records and Matched Publications—A Cross-Sectional Study." *BMJ open*, 3(1): e001963.

Elliott, Graham, Nikolay Kudrin, and Kaspar Wüthrich. 2022. "Detecting p-Hacking." *Econometrica*, 90(2): 887–906.

Fain, Kevin M, Thiyagu Rajakannan, Tony Tse, Rebecca J Williams, and Deborah A Zarin. 2018. "Results Reporting for Trials with the Same Sponsor, Drug, and Condition in Clinicaltrials.gov and Peer-Reviewed Publications." *JAMA Internal Medicine*, 178(7): 990–992.

Feltovich, Nicholas J, R Harbaugh, and T To. 2002. "Too Cool for School? Signalling and Countersignalling." *The RAND Journal of Economics*, 33(4): 630–649.

Hartung, Daniel, Deborah A Zarin, Jeanne-Marie Guise, Marian McDonagh, Robin Paynter, and Mark Helfand. 2014. "Reporting Discrepancies between the ClinicalTrials.gov Results Database and Peer Reviewed Publications." *Annals of Internal Medicine*, 160(7): 477.

Huser, Vojtech, and James J Cimino. 2013. "Evaluating Adherence to the International Committee of Medical Journal Editors' Policy of Mandatory, Timely Clinical Trial Registration." *Journal of the American Medical Informatics Association*, 20(e1): e169–e174.

Law, Michael R, Yuko Kawasumi, and Steven G Morgan. 2011. "Despite Law, Fewer than One in Eight Completed Studies of Drugs and Biologics are Reported on Time on ClinicalTrials.gov." *Health Affairs*, 30(12): 2338–2345.

Manheimer, Eric, and Diana Anderson. 2002. "Survey of Public Information About Ongoing Clinical Trials Funded by Industry: Evaluation of Completeness and Accessibility." *BMJ*, 325(7363): 528–531.

Mathieu, Sylvain, Isabelle Boutron, David Moher, Douglas G Altman, and Philippe Ravaud. 2009. "Comparison of Registered and Published Primary Outcomes in Randomized Controlled Trials." *JAMA*, 302(9): 977–984.

Milgrom, Paul R., and Robert J. Weber. 1985. "Distributional Strategies for Games with Incomplete Information." *Mathematics of Operations Research*, 10(4): 619–632.

Nguyen, Thi-Anh-Hoa, Agnes Dechartres, Soraya Belgherbi, and Philippe Ravaud. 2013. "Public Availability of Results of Trials Assessing Cancer Drugs in the United States." *Journal of Clinical Oncology*, 31(24): 2998–3003.

Ofosu, George K, and Daniel N Posner. 2023. "Pre-Analysis Plans: A Stocktaking." *Perspectives on Politics*, 21(1): 174–190.

Oostrom, Tamar. 2024. "Funding of Clinical Trials and Reported Drug Efficacy." *Journal of Political Economy*, 132(10): 3298–3333.

Prayle, Andrew P., Matthew N. Hurley, and Alan R. Smyth. 2012. "Compliance with Mandatory Reporting of Clinical Trial Results on ClinicalTrials.gov: Cross Sectional Study." *BMJ*, 344: d7373.

Zarin, Deborah A, Nicholas C Ide, Tony Tse, William R Harlan, Joyce C West, and Donald AB Lindberg. 2007. "Issues in the Registration of Clinical Trials." *Jama*, 297(19): 2112–2120.

Zarin, Deborah A, Tony Tse, Rebecca J Williams, and Thiyagu Rajakannan. 2017. "Update on Trial Registration 11 Years After the ICMJE Policy was Established." *New England Journal of Medicine*, 376(4): 383–391.

Zarin, Deborah A, Tony Tse, Rebecca J Williams, Robert M Califf, and Nicholas C Ide. 2011. "The ClinicalTrials.gov Results Database——Update and Key Issues." *New England Journal of Medicine*, 364(9): 852–860.

Online Appendix for
Research Registries and the Credibility Crisis:
An Empirical and Theoretical Investigation

H RA Instructions

H.1 RA Instructions for Evidence for P-Hacking Assessment

Background

- Research Registries and the Credibility Crisis: An Empirical and Theoretical Investigation
- https://www.socialscienceregistry.org/

Instructions

For each paper listed in the Excel, identify the *top 2 key outcome variables* that the paper is focused on based on the abstract. For each, record the following information:

- *Outcome Name*
- *How large an impact the main treatment had on the outcome, i.e., the "Effect Size"*
- *Standard Error of the Effect Size*
  - ***OR*** *P-Value of the Effect Size*
  - ***OR*** *T- Statistic for the Effect Size*

Save the Excel sheet with the date and email it to jlambrinos@uchicago.edu cc-ing eliot.a.abrams@gmail.com

Example

https://law.yale.edu/sites/default/files/area/workshop/leo/leo16_starr.pdf

Title	*Outcome*	*Outcome Name*	*Effect Size*	*Standard Error*	*P-Value*	*T-Statistic*
BAN THE BOX, CRIMINAL RECORDS, AND STATISTICAL DISCRIMINATION: A FIELD EXPERIMENT	*1*					
	2					

H.2 RA Instructions for Restrictiveness Assessment

Rubric for assessing pre-registration restrictiveness:

Use the Trial History button to get to the last pre-registry version before the Intervention Start Date with a +1 week buffer.

Primary Outcomes

- Number of outcomes listed ____

Note: Be mindful of indices. In some cases, PIs may list the variables which make up an index to be more specific. In these cases, the index itself should be counted as one primary outcome variable and the variables that make up the index should not be counted. Some of this information may appear in the "Primary Outcomes (explanation)" field.

- Specificity of outcomes listed

Score each outcome based on the example scale below and report the

  - Minimum ____
  - Maximum ____
  - Median ____

Example Scale: Mark "health" as a 0, "nutritional intake" as a 1, "number of fruits consumed" as a 2, "number of fruits consumed at school per week" as a 3, "number of fruits consumed at school per week during Spring quarter" as a 4, and "number of bananas consumed at school per week during Spring quarter" as a 5.

- Did the number of outcomes or their descriptions change after the Intervention Start Date?
  - Yes = 1
  - No = 0

Notes: Please click on View Changes and check that significant changes have been made. Minor semantic changes or typos do not count as changes.

Sample Information (found in Experiment Characteristics under Experimental Details):

- Estimate or prediction for final sample size____

Use field ***Sample size: planned number of observations.*** Put 0 if a specific number is not given

- Number of populations used ___

Add 1 for each population used.
For example, Put 3 if the analyses are run for all, then for men, then for women

- Did the sample size or sample splits change after the Intervention Start Date?
  - Yes = 1
  - No = 0

H.3 RA Instructions for Fidelity Assessment

Rubric for assessing fidelity of working/published paper to registration

Compare latest version of the paper available to the pre-registered version assessed above. You will likely need to search for the paper by title and then by authors. Titles will change.

Primary Outcomes

- Fraction of variables whose construction remains true to the pre-registry ___

  Example:

  - If 1 out of 5 variables changes, then report 0.80
  - The construction of a variable changes if the pre-registration lists "number of bananas consumed at school per week during Spring quarter" but the paper reports "number of bananas consumed at school per week during summer".

- Number of primary outcomes introduced in the paper but not previously registered ___

- Number of primary outcomes listed in the registry but not in the paper ___

Note: For this section, a primary outcome is a variable mentioned in the abstract, introduction, or conclusion.

Sample Information

- Number of observations reported in the paper ___

- Number of populations introduced in the paper, but not registered ___

For example, the paper may repeat analyses for rich household and for poor households. If these sub-populations are not mentioned in the preregistration, then put 2.

- Number of populations listed in the registry, but not mentioned in the paper ___

I ClinicalTrials.gov after the Final Rule

Section 801 of the Food and Drug Administration Amendments Act of 2007 (FDAAA 801) established requirements for clinical trials to register with and report results to ClinicalTrials.gov. These requirements were clarified and expanded by the the Final Rule for Clinical Trials Registration and Results Information Submission (Final Rule), which went into effect on January 18, 2017. See ClinicalTrials.gov for a comprehensive overview.

FDAAA 801 and the Final Rule may have increased the restrictiveness of trials' preregistrations with ClinicalTrials.gov along with the fidelity of trials' results to their preregistrations. To examine this hypothesis, we repeat our assessment of ClinicalTrials.gov for the year following the effective date of the Final Rule.[1] Specifically, we randomly sample 100 ClinicalTrials.gov preregistrations that occurred between January 18, 2017 and December 31, 2017. We then instruct two RAs to assess each preregistration and associated working or published paper using the same rubric as in Section 4.1.

Table I reports our assessment of the extent to which these preregistrations precisely specify their primary outcomes. We find that the preregistrations over January 18, 2017 to December 31, 2017 are slightly *less* specific than over the March 1, 2000 to July 1, 2005 period examined in Section 4.1. We also find that fewer trials changed their outcome or sample—likely due to the shorter follow-up period here.

Table II reports our assessment of the fidelity of working or published papers associated with the preregistrations. We identify working or published papers for 37 of the 100 preregistrations. These papers show slightly more fidelity to the preregistrations than those from the earlier March 1, 2000 to July 1, 2005 period. 90% of the primary outcomes reported by the average paper matched the construction specified in the preregistration, but the average paper still reported 0.36 additional primary outcomes and failed to report 0.15 primary outcomes.

[1] This choice provides at least two years of follow-up for all trials.

Table I: Assessment of the extent to which 100 randomly chosen ClinicalTrials.gov preregistrations precisely specify their primary outcomes

	Mean	Std	Min	10%	25%	50%	75%	90%	Max
Number of Outcomes	1.87	1.59	1.0	1.00	1.0	1.0	2.00	3.10	8.0
Minimumly Restrictive Outcome	2.60	1.10	0.5	1.50	1.5	2.5	3.12	4.05	5.0
Maximumly Restrictive Outcome	2.84	1.14	0.5	1.50	2.0	3.0	3.50	4.50	5.0
Median Restrictive Outcome	2.70	1.13	0.5	1.45	2.0	2.5	3.50	4.50	5.0
Outcome Changed (Yes/No)	0.12	0.28	0.0	0.00	0.0	0.0	0.00	0.50	1.0
Sample Changed (Yes/No)	0.34	0.46	0.0	0.00	0.0	0.0	1.00	1.00	1.0

Notes: Preregistrations were randomly sampled from the period January 18, 2017 and December 31, 2017. This period corresponds to the first year after the implementation of the Final Rule for Clinical Trials Registration and Results Information Submission (42 CFR Part 11). Each registration was assessed by two RAs. The values presented are based on the average of the two assessments.

Table II: Assessment of the extent to which working and published papers report the primary outcomes preregistered with ClinicalTrials.gov

	Mean	Std	Min	10%	25%	50%	75%	90%	Max
Fraction of Matching Outcomes	0.90	0.25	0.0	0.71	1.0	1.0	1.0	1.00	1.0
Number of Additional Outcomes	0.36	0.64	0.0	0.00	0.0	0.0	0.5	1.25	2.0
Number of Missing Outcomes	0.15	0.33	0.0	0.00	0.0	0.0	0.0	0.75	1.0

Notes: Working or published papers were found for 37 of the 100 preregistrations.

J P-Hacking Analysis, Using Data Prior to Fall 2024 Data Updates

Table III: Evidence for p-hacking by registration status based on the tests from Elliott, Kudrin and Wüthrich (2022)

Test	Not Registered	Registered
Binomial	0.87	0.73
CS2B	0.70	0.18
Discontinuity	0.00	0.00
CS1	0.53	0.57
LCM	1.00	0.96

Notes: There are 118 p-values in the Registered sample and 117 p-values in the Not Registered sample. Per Elliott, Kudrin and Wüthrich (2022), since the data do not only contain t-tests, we consider tests based on nonincreasingness and continuity of the p-curve (Theorem 1). Namely, a binomial test on [0.01, 0.05], Fisher's test, a density discontinuity test at 0.05, a histogram-based test for non-increasingness (CS1), and the LCM test. The CS1 test uses 15 bins. We increase the range used for the Binomial test from Elliott, Kudrin and Wüthrich (2022)'s range of [0.04, 0.05] in order to increase power. There are 13 p-values in the range [0.01, 0.05] in the Not Registered sample and 11 p-values in this range in the Registered Sample. Fisher's Test returns a value of 1 for both the Registered and Not Registered sample, and is hence not included in this table.

Table IV: Evidence for p-hacking using the procedure of Andrews and Kasy (2019)

	μ (SE)	τ (SE)	df (SE)	[0,1.96] (SE)
Overall	0.011 (0.006)	0.025 (0.018)	0.930 (0.092)	0.221 (0.049)
Registered	0.026 (0.012)	0.046 (0.017)	1.202 (0.198)	0.297 (0.085)
Not Registered	0.002 (0.004)	0.002 (0.004)	0.797 (0.087)	0.169 (0.041)

Notes: We use the specification of the publication probability which is symmetric, whose errors follow a student-t distribution, allowing for a single step at 1.96. The stated parameters μ, τ and df represent parameters of the model. The last column gives the publication probability for a result insignificant at the 5 percent level relative to a significant result. A value of 1 in this column implies no selection, whereas 1 divided by this column gives how much more likely a study with a significant result is to be published relative to an insignificant one. Standard errors of all estimates are in parentheses.

References

Andrews, Isaiah, and Maximilian Kasy. 2019. "Identification of and Correction for Publication Bias." *American Economic Review*, 109(8): 2766–94.

Elliott, Graham, Nikolay Kudrin, and Kaspar Wüthrich. 2022. "Detecting p-Hacking." *Econometrica*, 90(2): 887–906.