Research Registries: Facts, Myths, and Possible Improvements

Eliot Abrams* Jonathan Libgober† John A. List‡

May 30, 2020

Abstract: *The past few decades have ushered in an experimental revolution in economics whereby scholars are now much more likely to generate their own data. While there are virtues associated with this movement, there are concomitant difficulties. Several scientific disciplines, including economics, have launched research registries in an effort to attenuate key inferential issues. This study assesses registries both empirically and theoretically, with a special focus on the AEA registry. We find that over 90% of randomized control trials (RCTs) in economics do not register, only 50% of the RCTs that register do so before the intervention begins, and the majority of these preregistrations are not detailed enough to significantly aid inference. Our analysis further shows that using other scientific registries as aspirational examples is misguided, as their perceived success in tackling the main issues is largely a myth. In light of these facts, we advance a simple economic model to explore potential improvements. A key insight from the model is that removal of the (current) option to register completed RCTs could actually increase the fraction of trials that register. We also argue that linking IRB applications to registrations could further increase registry effectiveness.*

Keywords: Research Registries; Randomized Controlled Trials; Publication Bias
JEL: B41; C9; C91; C93

*Booth School of Business, University of Chicago, corresponding author, eabrams@uchicago.edu
†University of Southern California,libgober@usc.edu
‡University of Chicago jlist@uchicago.edu

We thank Ariel Listo and Xinyi Hong for excellent research assistance. Discussions with Vittorio Bassi, Eszter Czibor, Stefano DellaVigna, Michael Kremer, Min Sok Lee, Shengwu Li, Ulrike Malmendier, Torben Mideksa, Gautum Rao, and Sutanuka Roy significantly improved the paper. We also thank participants of the UChicago Experimental Economics Seminar for their feedback and the Becker Friedman Institute for financial support.

1 Introduction

"There is a property common to almost all the moral sciences, and by which they are distinguished from many of the physical... that it is seldom in our power to make experiments in them."
Mill (1836)

One immutable fact amongst economists is that there is rarely broad agreement about an issue of positive and normative import. Interestingly, one area where economists had seemingly agreed is how empiricism can be used to learn about the world. Whether it was John Stuart Mill in 1836, Milton Friedman in 1953, Joan Robinson in 1977, or William Nordhaus and Paul Samuelson in 1985, the general tenor was that, unlike chemists, physicists, and biologists, economists do not have the luxury of data generation via controlled experiments and therefore must rely on experiments that happen to occur (Friedman (1953), Robinson (1977), and Samuelson and Nordhaus (1985)). Of course, the general feelings of these icons were broadly shared amongst economists throughout the 19th and 20th centuries, which witnessed empirical advances primarily from extracting insights using naturally-occurring data.

Notwithstanding such ubiquitous advice, the last several decades have brought a significant change in the empirical landscape in economics. While use of historical evidence remains invaluable, new approaches to generate data in the lab and field have opened up several unique lines of research that go beyond measurement into the "whys" behind observed behaviors (see Harrison and List (2004)). While the experimental approach has helped to clarify identification, control, statistical inference, and interpretability, recently critics in the broader social sciences have called for the experimental movement to proceed more cautiously. An active debate has emerged that claims there is a "credibility crisis," whereby the foundation of the experimental approach and the credibility of the received results are called into question (see Nosek, Spies and Motyl (2012), Bettis (2012), Jennions and Møller (2003), Ioannidis (2005), Maniadis, Tufano and List (2014), and Dreber et al. (2015)).

The debate has evolved into several lines of inquiry, but the thread connecting them revolves around false positives, with lack of replication and external validity often carrying the water. This follows from the fact that data are ultimately finite, so that researchers must choose which hypothe-

ses 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 recognized these limitations could lead to a departure from socially optimal experimental conduct. This observation has led to a number of meta-analyses and policy prescriptions aimed at improving the social usefulness and credibility of empirical research (see Young (2018), Vivalt (2018), Vivalt (Forthcoming), Andrews and Kasy (2019), and Meager (2019) for meta-analyses, and Abadie (Forthcoming), McCloskey and Michaillat (2020), Christensen and Miguel (2018), Coffman, Niederle and Wilson (2017), and Glaeser (2008) for discussions of policy prescriptions).

In this paper, we study one such initiative in economics—the establishment of the AEA RCT Registry in May 2013. Briefly, the registry provides a venue for researchers to document their experiments in a manner that is searchable by external audiences. In principle, if used appropriately this innovation can tackle key issues in the credibility crisis. Our empirical work begins by evaluating the extent to which the registry has been successful in combating two particular areas that have received extensive attention:

- *The file drawer problem*, namely that not all experimental results are published and are therefore relegated to the "file drawer."
- *Scope for p-hacking and results manipulation*, namely that researchers often make ex-post decisions adaptively in a manner that is not accounted for in the empirical analyses.[1]

A registry can address the file drawer problem to the extent that it records all RCTs started and their outcomes. A registry can limit the scope for p-hacking by requiring researchers to specify their experimental design before commencement of the trial.[2]

Though still relatively new, the AEA RCT Registry appears to be the most commonly used reg-

[1]The extent to which results of empirical studies are manipulated in practice has been studied by Brodeur et al. (2016) and Vivalt (2018).

[2]The file drawer and p-hacking problems apply to empirical research generally, and not just RCTs. It is therefore puzzling that existing registries focus on RCTs. One explanation could be that RCTs are low hanging fruit——each RCT is ostensibly designed to test a small set of interventions and has an explicit start and end date. Other web services, such as AsPredicted, do facilitate recording *any* research hypothesis. However, unlike a registry, AsPredicted does not provide a way to search the recorded hypotheses. As such, AsPredicted does not attenuate the file drawer problem. AsPredicted might limit p-hacking in practice, but, in theory, researchers could abuse the service by (a) registering multiple hypotheses and (b) only revealing the confirmed hypothesis.

istration database in economics.[3] The largest research registry overall is ClinicalTrials.gov, which is maintained by the National Institutes of Health and contains 302,850 medical trial registrations from 208 countries as of April 1, 2019.[4] A growing body of research (that we review and extend in Section 3) has assessed the mixed effectiveness of ClinicalTrials.gov. To the best of our knowledge, we are the first to provide a corresponding systematic assessment of the AEA RCT Registry. We also believe our combined empirical and theoretical insights have the potential to serve as a starting point for registry improvement more generally.

Our goals in this paper are therefore twofold. First, to determine whether the AEA RCT Registry has been effective in solving the file drawer problem and limiting the scope for p-hacking. Second, to determine whether alternative registry designs might improve outcomes compared to the status quo. In this spirit, we focus on one concrete design issue, namely that the registry accommodates *late registration*. Specifically, while typical motivations for promoting registration rely upon the assumption that it is done prior to experimentation,[5] this need not be the case to be part of the AEA RCT Registry. It is noteworthy that ClinicalTrials.gov also allows late registration, although certain categories of experiments are required by law to be preregistered. To the best of our knowledge, no such laws exist for economics.

Unfortunately, we find little evidence that the AEA RCT Registry is sufficiently addressing either the file drawer problem or p-hacking. A theme that emerges from our analysis is that the social norm of registration appears rather limited. Many trials fail to register and those that do register often fail to provide a detailed description of their experimental designs beyond the mandatory registration requirements. Insofar as these requirements are fairly weak (which, we should emphasize, appears to have been a deliberate choice in order to encourage participation and help *establish* a norm for registration), this unfortunately implies that the impact of registration on credibility is fairly weak as well.

Regarding the file drawer problem, while the universe of started economics experiments is un-

[3] See Section 3.3 for evidence of this assertion. For discussion specifically around the role of the AEA registry in terms of promoting transparency, see Christensen and Miguel (2018).

[4] In contrast, the AEA RCT Registry lists 2,444 studies located in 133 countries as of April 1, 2019 (although the reader should bear in mind that the AEA RCT Registry is much newer than ClinicalTrials.gov).

[5] For instance, because researchers may be more likely to "relegate an experimental finding to the file drawer" if the results are negative. If researchers do not attempt to publish experiments with negative results, then they may not have sufficient incentives to distribute results ex-post.

observed, we provide an upper bound on the fraction of RCTs that register by performing a census of RCTs published in select economics journals and working paper series. Roughly half of the RCTs published in top journals (general interest or field) in economics—between 2017 and the end of 2019 Q2—are registered. While these journals represent a selected universe of top papers, perhaps more telling is that only 8% of the working papers in the NEP report on experimental economics are registered. Interestingly, *none* of the RCTs published in the premier field journal Experimental Economics are registered. These RCTs consist entirely of lab experiments, suggesting that the norm for registration of lab experiments is weak. This norm could partly explain the low registration rates for the NEP report on experimental economics. Moreover, we also find the AEA RCT Registry is not currently effective at capturing RCT outcomes. We find that only about one-third of registered studies follow-up with any outcome data as of April 1, 2019.[6]

Concerning p-hacking, we find that the AEA RCT Registry does not currently succeed in substantially limiting this credibility threat. On the one hand, the vast majority of registered trials (roughly 90%) do not provide a pre-analysis plan at registration. On the other hand, the information that is provided is often not specific enough to tie researchers to *one* experimental design. To highlight this fact, we assess the primary outcomes reported by 300 randomly chosen preregistrations. The average preregistration reports 3 primary outcomes, but even the most detailed of these outcomes fails to specify either a specific variable construction or measurement time frame. As we discuss more patiently below, even the most restrictive outcomes are similar to "number of fruits each experimental subject consumes" rather than to "number of apples each experimental subject consumes in March, 2019." We are able to find working papers associated with just over a third of the preregistrations.[7] Comparing the working papers to the preregistrations, we find that researchers change the construction of a primary outcome 10% of the time (e.g. report a number of vegetables consumed rather than a number of fruits consumed) and add a primary outcome 25% of the time (i.e. the working papers highlight an unregistered variable in their abstract, introduction, or conclusion).

Assessments of ClinicalTrials.gov provide a useful benchmark for our results on the AEA RCT

[6]This issue is not unique to economics. As we discuss in Section 3.1, ClinicalTrials.gov also faces problems with capturing outcome data.

[7]There are no associated published papers.

Registry. We first survey the existing literature on ClinicalTrials.gov and conclude that ClinicalTrials.gov has foundational problems similar to the AEA RCT Registry. We then examine the restrictiveness and fidelity of primary outcomes reported by 300 randomly chosen preregistrations from the first five years of ClinicalTrials.gov. We find that the ClinicalTrials.gov preregistrations are somewhat 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. Overall, these results 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.

In an effort to understand these data patterns and provide guidance moving forward, we construct a simple model of experimentation that allows for registration decisions and discusses the incentives underlying registration.[8] The model features a researcher endowed with an experiment on an underlying hypothesis and an outside "consumer" of research. The researcher first chooses whether to preregister (or not) and conduct the experiment (or not). The researcher then chooses to register late (or not) and receives a payoff based on the outsider's updated belief about the underlying hypothesis. Preregistration allows researchers to signal confidence in their hypotheses, for instance due to 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 given the experiment outcome.

The value of the model is twofold. First, we are able to formally scrutinize certain assertions regarding how registration influences research, particularly for researchers on the margin.[9] Second, we provide comparative statics that help determine how counterfactual policies influence registration decisions. In particular, we use the model to examine the implications of banning late registration. One might conjecture that allowing late registration could only increase the fraction

[8] A number of recent theoretical models seek to capture researcher incentives in order to speak to optimal design and conduct of experimentation. However, we are not aware of any models that speak to registration. For examples and further discussions of this growing literature, see Di Tillio, Ottaviani and Sorenson (2019), Libgober (2020), Al-Ubaydli, List and Suskind (2019), Tetenov (2016), and Anderson and Magruder (2017).

[9] For instance, we highlight that incentivizing registration may discourage experimentation by raising the bar for a study's results to appear significant, reflecting similar issues raised by Duflo et al. (2020).

of registered studies, since researchers have more opportunities to register if they can do so late. Our model highlights that this cannot be taken for granted. We show that late registration may be *at the expense of* preregistration, and as a result a ban may actually increase registrations *overall*. We use a calibration exercise to argue that this insight is empirically relevant for the AEA RCT Registry. Under several parameterizations of the model that match the current registration patterns, we find that banning late registration does strictly increase registrations. This suggests taking seriously the idea that a late registration ban could effectively address both the file drawer problem and p-hacking.

So, where do we go from here? Our model sheds insights into the potential for registries to attenuate false positives. Our first recommendation is to explore not allowing late registration, while simultaneously providing incentives for scholars to register their work (such as mandating that the work be registered before the experiment starts to be published). Insofar as the ultimate goal of a registry is to maximize preregistration, this dual approach can move us in that direction. Yet, this does not solve two other issues that our model highlights: 1) the lack of specificity in registrations, which is key to solving p-hacking and 2) fewer experiments will actually be conducted in equilibrium if registration costs are prohibitive. Our second recommendation tackles these issues: since RCTs require institutional review board (IRB) approval, we propose to have the researcher submit their IRB materials as a condition for preregistration. While admittedly the IRB materials are heterogeneous across schools, in our experience they contain enough uniformity and detail to provide a check on p-hacking. In addition, this approach avoids large additional costs since researchers can simply upload IRB forms that have already been completed.

The remainder of the paper is organized as follows. Section 2 provides a brief background on the AEA RCT Registry and then assesses whether the registry is currently solving the file drawer and p-hacking problems. Section 3 compares the AEA RCT Registry to ClinicalTrials.gov and discusses other registration venues. Section 4 provides and examines our model of a researcher's registration decision. Section 5 concludes. All tables, figures, and proofs are in the respective appendices.

2 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.[10] 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 cause empirical research to be undermined in the eyes of the policymaker, broader public, and the scientific community itself. Research registries are a prominent potential solution to both the file drawer problem and p-hacking. Here, we examine the extent to which the AEA RCT Registry is currently capturing the universe of started economics RCTs and the extent to which the registry succeeds in pre-committing researchers to assessing a specific set of outcome variables. We consider the AEA RCT Registry from its launch on May 15, 2013 up through April 5, 2019.

2.1 Background

The AEA RCT Registry is designed to capture a census of on-going, complete, or terminated RCTs in economics and other social sciences (see About the Registry on the AEA registry webpage). The registration process 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 details.[11]

[10] 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.

[11] Many RCTs in economics require IRB approval, but the IRB approvals are not made publicly available. A policy that either made external registration a condition for IRB approval or made IRB approvals public would directly help solve the file draw problem. Informed by our model, we also argue in the conclusion that requiring researchers to upload their IRB materials during registration could significantly improve the registry's ability to attenuate p-hacking at little cost.

A registration is distinct from a pre-analysis plan though in our experience they are often conflated. A registration is essentially metadata and a list of primary outcomes and treatment arms. In contrast, Duflo et al. (2020) propose that a pre-analysis plan should answer two detailed questions: "What are the key outcomes and analyses?" and "What is the planned regression framework or statistical test for those outcomes?" Put another way, a pre-analysis plan goes beyond a registration by specifying (1) a set of primary analyses and (2) the content of those analyses. A more detailed pre-analysis plan may go even further and specify *all* steps involved in analyzing the data. Of note, researchers have the ability to upload a pre-analysis plan as part of their registration (see Ofosu and Posner (2019) for an analysis of the pre-analysis plans that have been added to the AEA RCT Registry).

The AEA journals require that field experiments, but not necessarily lab experiments, be registered in order to be published.[12] However, no economics journal requires that any experiment *pre*-register, and in fact allow registration to be done at the time of submission.[13] In contrast, most medical journals require preregistration of experiments.

In principle, the timing of a registration can be determined from its listing in the AEA RCT Registry database. All RCTs are listed side-by-side with the preregistered trials marked by a small orange clock in the upper left corner of the trial entry. Trials that registered after data collection began are instead marked by a grey clock (see Figure I). 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,[14] though researchers may emphasize that a study was preregistered in the corresponding written paper.

Finally, a few other aspects of the AEA RCT Registry will prove important to our analysis. First, it is possible to update a registration after it is initially submitted although, as we document

[12]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."

[13]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.**"

[14]Anecdotally, despite our own familiarity with the registry, we never realized these clock icons existed until starting this project. Likewise, we informally discussed this paper with several colleagues and most were not aware of this distinction prior to our informing them.

below, this is rarely done. Second, it is also possible to hide certain fields in the registration from public view until later dates. This feature allows researchers to register without disclosing significant information. Third, the AEA RCT 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.

2.2 File Drawer Problem

We first assess whether the AEA RCT Registry is effective at solving the file drawer problem. Informally, a registry can help solve the file drawer problem to the extent that

1. Every RCT that is started is added to the register
2. RCT outcomes are added to the registry at the conclusion of the experiment

As the universe of started RCTs is unknown, we cannot determine the fraction of trials that register with accuracy.[15] That said, we can establish a rough upper bound by examining the registration rate for RCTs published in select economics journals and working paper series. Table I presents the registration rates for RCTs appearing in the following journals in 2017, 2018, and the first two quarters of 2019:

- American Economic Review (AER)
- American Economic Journal: Microeconomics (AEJ-M)
- American Economic Journal: Applied Economics (AEJ-AE)
- American Economic Journal: Economic Policy (AEJ-EP)
- Journal of Political Economy (JPE)
- Quarterly Journal of Economics (QJE)
- Review of Economic Studies (REStud)
- Journal of Development Economics (JDE)
- Experimental Economics (EE)
- Working papers in the area of experimental economics indexed by NEP-EXP

Columns 1-3 report the number of RCTs published in each journal in each year. The publication counts vary significantly by journal. On the high end, Experimental Economics publishes nearly 30 RCTs a year and the Journal of Development Economics publishes around 15 RCTs each year. On

[15] As mentioned above, while IRB approvals could conceivably be used to determine this, they are not publicly available or searchable.

the low end, AEJ-M and AEJ-EP only publish 2 RCTs each year. Columns 4-6 report the fraction of the published RCTs that registered with the AEA RCT Registry before August 2019. Registration rates across journals are heterogeneous and overall quite low: the AER, QJE, and AEJ-AE have the highest registration rates with only about 60% of the papers with an RCT registering in each year. The Journal of Development Economics and REStud have registration rates of around 33%. The remaining journals have registration rates under 10%. Of note, no RCTs published in Experimental Economics were registered. Experimental Economics primarily publishes lab experiments, i.e. RCTs that take place within a classroom or decision research lab within a university (see Harrison and List, 2004, for definitions of the various experiment types). This result suggests that it is not a norm within economics to register lab experiments.

As aforementioned, the AEA journals require registration prior to publication. Table II reports the registration rates for the AEA journals. Over the 2018-2019 period, the AEJ-EP, AEJ-AE, and AER published field experiments. However, the registration rates only hovered between 60% and 75%—far from full compliance. That said, this result could reflect ambiguity about what counts as a field experiment.[16] Over the 2018-2019 period, all four journals published lab experiments. None of the lab experiments registered.

The second step in solving the file draw problem is reporting outcomes. The registry data speaks immediately to whether outcomes are added to the registry at the conclusion of the RCT. Few registered trials add their outcomes. Of the 1,654 registered trials that ended before December 31, 2018, only 21% provided preliminary results or a link to a working paper by April 1, 2019. In fact, only 32% provided *any* follow-up information about the trial, e.g. intervention completion date, final number of observations, and whether there is public data available. This result is not driven by the short horizon. Of the 1,210 trials that ended before December 31, 2017, only 28% provide preliminary results or a link to a working paper and only 41% provide any follow-up information by April 1, 2019.

[16]For instance, an experiment that is conducted in a particular location and among a particular population may technically be counted as a lab experiment if the environment is directly administered. Such "lab in the field" experiments appear to not be bound to the AEA requirement.

2.2.1 Late Registration

As previously discussed, the AEA RCT Registry allows researchers to register RCTs even after the start of the intervention. Allowing late registration might help solve the file draw problem by facilitating more registrations. Here it is not per se important that the trial is registered immediately, just that it is registered. That said, late registration can also incentivize researchers to not register, insofar as they may attempt to delay registration and subsequently neglect to do so if not seeking to publish the study. This point is made more formally via our model, which highlights that allowing late registration may indeed come at the cost of diminishing the fraction of studies that register early. Late registration also enables a researcher to maximize her reputation by only registering projects after she is confident they will succeed.

In practice, it is not generally possible to tell if a given trial was registered late because the researcher did not know the registry existed or if the researcher purposely waited to register the trial.[17] Should the first case dominate, then allowing late registration helps to establish a census of trials. However, should the second case dominate, then allowing late registration may increase the file draw problem.

Fortunately, we are able to partially disentangle these two possibilities for the subset of researchers who register multiple trials over time. If a researcher registers her first trial late and then preregisters all of her future trials (i.e. those started after the first registration), then the late registration was likely due to not knowing about the registry. In contrast, if a researcher is repeatedly late in registering trials started at future dates, then the researcher is likely registering late on purpose.[18] Table VI displays all registrations made by three primary investigators. Each primary investigator here registered their first trial in 2014, proceeded to register multiple new trials (started after 2014) late, and registered their most recent trial over a year after the intervention began.

To investigate whether researchers purposely register late, we consider the subset of 1,209 distinct primary investigators who register a trial with a start date after January 1, 2014—note that

[17] For example, consider an unregistered project that a researcher is about to submit to the journal. There are at least two ways that this project enters the registry. First, the researcher may be unaware of the registry. On submission, the researcher learns of the registry from a referee and chooses to register. Second, the researcher may be aware of the registry. Before submission, the researcher decides to register and so is able to report that the paper was registered, which might be a signal of quality to the journal, or required by the journal (as it is for the AEA journals).

[18] At some point it becomes untenable to conclude that the researcher is just disorganized.

the registry itself opened in May 2013. To be further conservative, we only consider a registration as late if it occurred more than a week after the intervention began. 319 of the 1,209 primary investigators registered multiple trials. Of interest are the 231 primary investigators who registered at least one of their multiple trials late. 98 of these researchers registered multiple trials late at dates more than a quarter apart. This combination of observations suggests that many researchers register late on purpose. As such, we pay special attention to late registration as a model feature in Section 4.

2.3 P-Hacking

We now assess whether the AEA RCT Registry is effective at attenuating p-hacking. Informally, a registry can reduce p-hacking to the extent that

1. The RCT is registered before the intervention begins, i.e. it is preregistered
2. The registration specifies details of the experimental design (particularly the primary outcome variables)
3. Outside researchers routinely compare the published or working report on the RCT to its registration

We examine each of these issues in turn starting from the fraction of RCTs that preregister.

2.3.1 Preregistration

The registry data speaks directly to the fraction of RCTs that register before the intervention begins. As mentioned, the registry opened in May 2013. To allow time for researchers to learn about the registry's existence, we examine the subset of 1,792 trials whose start date is after January 1, 2014.[19] Of these trials, only 47% registered before the intervention began. Another 30% registered before the intervention ended.[20] Figure II presents the cumulative number of preregistrations and late registrations over time and Figure III presents the number of preregistrations and late registrations each quarter. While the fraction of RCTs that are preregistered has been weakly growing over time, the registry is still dominated by late registrations.

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

[20]The registry also allows trials to report a data collection completion date. Only 146 of these trials report this value. 76 of the 146 registered before the data collection completed.

Table V presents the preregistration rates for RCTs published in the journals considered above. Columns 1-3 report the number of papers with a registered RCT published in each journal by year. Columns 4-6 report the number of these papers whose RCT started post 2013 (note that the AEA RCT Registry opened in May 2013). Only a handful of papers published in each journal contained a registered RCT that started after 2013—in part, due to the lengthy process for some developmental economics experiments. Reflecting the above results, only one-third of the papers preregistered their trial.

2.3.2 Restrictiveness

When considering the extent to which the registration specifies the experimental design to be executed, we face a more challenging task. By design, a registration does not require the submission of a detailed pre-analysis plan. Correspondingly, only 11% of the 1,792 trials post a pre-analysis plan and the majority of these are not made publicly viewable until after the completion of the RCT. That said, registration does require the researcher to provide a basic description of:

- Primary outcomes[21]
- Randomization method
- Planned number of observations and treatment arms

As a first pass, we examine whether registrations specify the primary outcomes in enough detail to tie the researcher to specific variable constructions. Because the description of primary outcomes is sometimes open to interpretation, we had two research assistants (hereafter RAs) independently review each preregistration. The RAs were instructed to count the number of primary outcomes listed and score each outcome description based on its specificity on a scale of 0 to 5. The RAs were given the following 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." Appendix D provides the full RA instructions. The following statistics are based on the average of the two RAs' assessments.[22]

[21] Secondary outcomes are an option field. 25% of trials list a secondary outcome.

[22] The work was carried out by 10 RAs. Each RA was assigned two sets of 30 preregistrations. The average correlation of the restrictiveness scores across RA pairs was 70%.

Table III reports the assessed restrictiveness of 300 randomly selected RCT preregistrations. The average preregistration specified 3 primary outcomes. The average minimumly restrictive outcome and the median restrictive outcome are classified as a 2—these outcomes are only as precise as "number of fruits consumed." The preregistrations generally do not specify a precise measurement unit (say number of bananas) nor a measurement time frame. The average maximumly restrictive outcome is classified as a 2.5—so somewhere between "number of fruits consumed" and "number of fruits consumed at school per week." Only the 90th percentile maximumly restrictive outcome specified a precise measurement time frame. No outcome was as precise as "number of bananas consumed at school per week during Spring quarter."

Delecourt and Ng's preregistration of "Unpacking the Gender Profit Gap: Evidence from Micro-Businesses in India" provides a useful example. 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 expect for missing a time frame; quoted price and price paid are missing both a specification of the products to be considered (likely the primary outcome of interest will actually be a price index) and a time frame; and number of "missed" clients is missing both a specification of how missed will be measured and a time frame. The two RAs assessing this preregistration agreed that the maximumly restrictive outcome here is a 4 and the minimally restrictive outcome is a 2.

The RAs were also instructed to compare the most recent version of the registration to the preregistration to explore if any primary outcome or sample specification changed. The last two rows in Table III report the results. We find that 4% of the 300 assessed trials changed one of their primary outcomes after the preregistration. Similarly, 5% of the assessed trials changed some aspect of their sample specification after the preregistration.

2.3.3 Fidelity

We next assess the extent to which the primary outcomes reported in the associated working and published papers match the preregistered primary outcomes. The p-hacking concern here is that authors might change the construction of primary outcomes in order to achieve significant results, add additional outcomes that have a significant relationship, or not report outcomes that do not have a significant relationship. 281 of the 300 assessed preregistrations listed an intervention end date and 230 ended before June 2019. However, only 10 of the preregistrations provide a link to a working or published paper. As such, we instructed the RAs to use the reported link if present else to try to find an associated paper through Google Scholar via searching for the title and authors. The RAs conducted this search over August 2019 and found working papers for 119 of the preregistrations (there were no associated published papers). Given the above, we expect that this is close to the complete universe of working papers.

Table IV reports the assessed fidelity of working papers associated with the preregistrations. On average, 90% of the primary outcomes in a given working paper match their preregistered construction. However, this figure is somewhat misleading because the vast majority of preregistered primary outcomes were 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, roughly a quarter of the working papers report additional primary outcomes (i.e. the working papers highlight an unregistered variable in their abstract, introduction, or conclusion—see Appendix D). The average working paper reports 0.5 additional primary outcomes. Similarly, roughly a quarter of the working papers fail to report a primary outcome with the average working paper under-reporting 0.4 primary outcomes.

An important caveat to our analysis is that there are many valid reasons for researchers to deviate from their preregistered experimental design. For example, a sudden influx of monetary or technological support may enable a field experiment to record additional primary outcomes midway through an intervention. Rather than a binding constraint on what researchers can do and journals can publish, preregistration can be thought of as useful additional information for outside researchers. Preregistration provides value by distinguishing the initial hypotheses and testing

procedures from additional hypotheses and tests that became available or were developed during the course of the experiment.

3 ClinicalTrials.gov and Other Registries

3.1 ClinicalTrials.gov

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, we find that the success of ClinicalTrials.gov in solving the credibility crisis is largely mythical. 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.[23] 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.[24]

Second, many trials that do register do not provide sufficient information. Zarin et al. (2011)

[23] 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).

[24] 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% registration rate for trials published in five ICMJE founding journals over 2010-2011.

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.[25] 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.[26] 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.[27]

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 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%

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

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

[27]In a partial counterpoint, Oostrom (2020) 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.

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.

3.2 Restrictiveness and Fidelity of ClinicalTrials.gov

We now extend the existing literature by conducting a new survey of ClinicalTrials.gov. This new survey serves to more precisely benchmark our results on the restrictiveness of AEA RCT Registry preregistrations and on the fidelity of published or working papers to those preregistrations. 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. Overall, these results suggests that if ClinicalTrials.gov presents 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.

We proceed by randomly sampling 300 trials that preregistered with ClinicalTrials.gov between March 1, 2000 and July 1, 2005. This period runs from the start of the ClinicalTrials.gov website up through the enforcement of the International Committee of Medical Journal Editors' (ICMJE) policy requiring investigators to preregister trials as a condition for publication. We employ four RAs to independently review each trial. Using the same rubric as for the AEA RCT Registry, each RA 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.[28]

[28]The RAs 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.

Table VII 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 median restrictive outcome as 3, and the 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."[29] We were able to associate published or working papers with 278 of the 300 ClinicalTrials.gov preregistrations. Table VIII reports the assessed fidelity of the primary outcomes reported in these papers to those in the registration. On average, 80% of the primary outcomes in a given paper match their registered construction—as compared to 90% of the the AEA RCT Registry primary outcomes.[30] However, as with the AEA RCT Registry results, this figure is misleading because the vast majority of registered primary outcomes are vague enough to match with multiple possible variable constructions. 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.

3.3 Other Research Registries

A separate open question is whether economists use other research registeries in addition to or in place of the AEA RCT Registry. To answer this question, we first directly examine whether economists register in two specific alternative registries—the Registry for International Development Impact Evaluations (RIDIE) and the Evidence in Governance and Politics (EGAP) Registry. Figure V displays the number of economics registrations and the total number of registrations in RIDIE and EGAP by quarter over 2018 and 2019. We find that there is no single quarter with

[29]The last two rows in Table VII 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.

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

more than 25 economics registrations in either registry. This exercise provides some evidence that economists primarily use the AEA RCT Registry.

We next examine if and where each RCT published in Experimental Economics over 2016-2019 registered. This second exercise is motivated by the fact that no paper published in Experimental Economics registered with the AEA RCT Registry. Appendix D describes the exact search process. Surprisingly, we find no registrations. This result confirms our earlier conjecture that registration is not yet a norm for economists performing lab experiments. It is not the case that the AEA RCT Registry is somehow maladapted to economics lab experiments. Rather, economists running lab experiments generally do not register (or preregister) anywhere.

4 A Model of Registration

In this section, we introduce a simple model that articulates the incentives to register and the implications of registration timing. The model supports two sets of results regarding registration patterns. First, we find that removing the option to register an experiment after it has completed weakly increases the fraction of experiments that preregister and can even increase the total number of experiments that register overall. Second, we find that increasing the informativeness of a registration further increases the number of preregistrations. We end with a numerical calibration that provides some support for concluding that banning late registration will increase the number of economics experiments that register with the AEA RCT Registry.

4.1 Model Description and Assumptions

Our model is a simple two-stage experimentation problem. We consider a researcher who is endowed with an experiment related to state $\theta \in \{T, F\}$—for instance, reflecting whether an intervention causes a significant treatment effect. The researcher receives a signal on this underlying state in each stage, but an outsider (e.g. the public or journal editor) is only able to observe the signal from the second stage as well as the registration decision and time. The specific timing of

the researcher's actions in our model (illustrated in Figure VI) is:[31]

- First, the researcher observes an initial signal s_1 and then decides whether to conduct the experiment as well as whether to *preregister* (which we also refer to as *registering early*).

- After conducting the experiment, the researcher observes a second signal s_2 and, if the study was not preregistered, decides whether to *register late*.

We first describe our assumptions on the information environment and payoffs to the researcher. We then define a class of partitional equilibria wherein the researchers who register at each stage are those who have the most favorable signals.

4.1.1 Information Environment

We assume that the researcher and outsider initially share a common prior p_0 over $\theta \in \{T, F\}$. The researcher then receives two signals, each with a continuously differentiable density:

- The researcher's signal in the first stage is drawn according to $s_1 \sim f(\cdot \mid \theta)$ where we assume $\frac{d}{ds_1} \log f(s_1 \mid T) \geq \frac{d}{ds_1} \log f(s_1 \mid F)$, and take the distribution over s_1 to be larger according to first-order stochastic dominance when $\theta = T$ than when $\theta = F$.

- In the second stage, the researcher observes a second signal $s_2 \sim g_\gamma(\cdot \mid \theta)$ if she registered in the first stage where $\gamma \in \Gamma$ is exogenous and fixed. Else the researcher receives the second signal $s_2 \sim g_0(\cdot \mid \theta)$. We assume $\frac{d}{ds_2} \log g_{\tilde{\gamma}}(s_2 \mid T) \geq \frac{d}{ds_2} \log g_{\tilde{\gamma}}(s_2 \mid F)$, for all s_2 and $\tilde{\gamma} \in \Gamma \cup \{0\}$, and take the distribution over s_2 to be FOSD larger if $\theta = T$ than if $\theta = F$.

Importantly, $\gamma \in \Gamma$ parameterizes the informativeness of the second period signal following registration. For example, the process of registering may help experimenters think through additional contingencies that lead to an improved experimental design. We take the impact of registration on informativeness to be exogenous, although we study comparative statics in γ as well.

While only the researcher (directly) observes s_1, both the researcher and the outsider observe s_2 as well as the registration decision $d \in \{\emptyset, 1, 2\}$ (i.e. no registration, registration at $t = 1$, or

[31] A ban on late registration corresponds to an otherwise identical decision problem, except where the researcher is *only* able to preregister or not register at all.

registration at $t = 2$ respectively). We denote the belief of the outsider that $\theta = T$ by $\hat{p}_d(s_2)$. We think of s_1 as reflecting intuition or prior knowledge on the part of the researcher or information on the propensity of her sample to show treatment effects (for instance, as in the model of scaling results in Al-Ubaydli, List and Suskind (2019)). In contrast, s_2 reflects the experimental findings, which can be conveyed verifiably. The assumptions on the signals are standard technical assumptions that ensure that higher signals lead to positive updates on the truth of the hypothesis (which we verify in Appendix C). All signals are also assumed to have full support.

4.1.2 Researcher Payoffs

The researcher incurs a cost of $c_E \geq 0$ for conducting the experiment and also incurs a cost $c_R \geq 0$ whenever registering the experiment (whether registration is early or late). If the researcher does not conduct the experiment, then the researcher receives a payoff of 0. Else researcher receives a payoff which depends on the registration decision and the outsider's belief, $\hat{p}_d(s_2)$. We denote the payoff following registration as $b_R(\hat{p}_d(s_2))$ and the payoff following non-registration as $b_N(\hat{p}_\emptyset(s_2))$. We assume $b_N(p) \leq b_R(p)$ for all $p \in [0, 1]$. We also assume that $b_i(p)$ is continuous and increasing in p—reflecting a preference for positive results (see Brodeur et al. (2016) and Andrews and Kasy (2019) for empirical evidence suggestive of this preference). For some results below, we assume $b_R(p)$ is weakly convex in p, reflecting a (weak) preference for informative experiments.

4.1.3 Partitional Equilibrium

For some general results below and our calibration, we focus on the following class of equilibria:

Definition 1. *A **partitional equilibrium** is characterized by thresholds and $s^*_{1,\emptyset}$, $s^*_{1,R}$, $s^*_{2,R}$ 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 $s_2 > s^*_{2,R}$.*

Partitional equilibria are convenient to work with because the threshold signal is indifferent between actions on each side of the threshold.[32] We are able to derive the comparative statics results described below by studying these indifference conditions.[33]

The following assumption is necessary to ensure that the second period registration takes the partitional form, for all registration costs c_R:

Assumption 1. *The difference in payoffs between registration decision, $b_R(p) - b_N(p)$, is strictly increasing in p.*

This assumption says that the gain to registration is higher when the outsider's belief is more optimistic. Equivalently, this assumption says the additional optimism benefits the researcher more following registration, suggesting complementarities between beliefs and registration. Researcher payoffs as a function of beliefs may arise from a variety of sources (e.g., reputational considerations). In Appendix C, we discuss a few simple microfoundations of payoffs which provide more context for when this assumption is satisfied. However, we do not take a stand on microfoundations for this complementarity.[34]

A technical difficulty is that, while increasing differences is necessary for the second period signal to be a partition for all c_R, this is not enough to ensure the same holds for the first period registration decision.[35] While Propositions 2 and 3 take this as given, in Appendix C, we provide a sufficient condition which ensures that indeed, the local indifference condition ensures equilibrium holds globally. This condition states that as researchers grow more optimistic that $\theta = T$, their

[32] That is, a researcher with signal $s^*_{1,R}$ should be indifferent between preregistration and not, and likewise for other signals in this definition.

[33] Note that this model will always possess a pooling equilibrium whereby registration is seen as a *negative* signal. That is, consider a profile where researchers never preregisters, and any deviation is inferred as coming from the researchers with the worst possible signal. In this case, there is no incentive to preregister, since it is both seen negatively and sacrifices option value. Note that this equilibrium requires off-path beliefs susceptible to criticisms in the spirit of the intuitive criterion—assuming that earlier registration is interpreted more negatively may be unpalatable since it is the researchers with higher initial signals who have lower option value, and hence would have the least to lose by preregistering.

[34] In Appendix C, we show that, under Assumption 1, the second period registration decision does not convey information regarding the first period signal, which can fail more generally.

[35] The reason is the following: if the first period signals makes the researcher sufficiently optimistic that the second period signal will be favorable, *independently of the registration decision*, then the added benefit to registering earlier may decrease as well. The potential for non-monotonicity in signalling games is a well-known theoretical issue; see Feltovich, Harbaugh and To (2002) for a discussion of countersignalling equilibria, as well as Liu and Pei (2020) for a general treatment of potential non-monotonicities in signalling games and conditions under which they can be ruled out. Note that the latter paper shows single-crossing by itself does not ensure monotonicity.

preference for preregistration over late registration increases as well. This condition is useful for our numerical calibration, since checking that it holds implies the global conditions for equilibrium are satisfied given indifference at the threshold signals. We omit the technical details from the main text, in order to maintain focus on the implications.

For some of our comparative statics results in Section 4.2, it is important to rule out edge cases wherein all registration is early. To do so, we use the following assumption (which we emphasize is only used for our comparative statics results):

Assumption 2. *Let s_1^* denote the value of s_1 that causes the researcher to be indifferent between preregistration and not experimenting—assuming all higher signals preregister. Then type s_1^* would have a profitable deviation to experiment without registering if the observer interpreted such actions as implying $s_1 = s_1^*$.*

While this assumption may appear restrictive at first, it is necessary to rule out an edge cases where all researchers preregister in the absence of a late ban. In this case, a late registration ban would have no impact. Other comparative statics where this is violated can be treated separately, but follow similar logic.

Figure VII illustrates the content of Assumption 2, as well as how to find the equilibrium threshold $s_{1,R}^*$. This graph shows the expected payoffs from each (time 1) decision for a researcher with a signal equal to $s_{1,R}^*$, when the outsider conjectures $s_{1,R}^*$ as the registration threshold. Assumption 2 states that the payoff from delaying the registration decision is higher than the payoff from preregistration, when $s_{1,R}^*$ is the lowest possible researcher signal. We note that this assumption can be checked directly from model primitives, and is not an assumption on equilibrium. In contrast, the equilibrium condition requires these payoffs to be equal.

4.2 General Results

We present two sets of results. The first set considers the impact of banning late registration. We show that banning late registration can increase the total number of experiments that register—improving the registry's usefulness in solving the file draw problem. We also show that banning

late registration always weakly increases preregistrations—improving the registry's ability to attenuate p-hacking. The second set of results considers a change in the informativeness of the second period signal following registration. Here, we show that increasing the informativeness of registration further increases the number of preregistrations. We also articulate a subtle trade-off between incentivizing preregistration and incentivizing experimentation, reflecting similar concerns related to the social costs of pre-analysis plans by Duflo et al. (2020).

4.2.1 Implications of Banning Late Registration

We now turn to a discussion of a ban on late registration. Formally, we now assume that the researcher does not have the option of registering in the second stage. Ignoring researcher incentives, one could imagine that allowing late registration would lead to more trials registering. For instance, suppose the researcher simply decides to register in each period with some probability (independent of all other variables). In this case, a late registration ban would simply stop registrations that would have otherwise occurred. While this direct effect of banning late registrations is present in our model, the picture is more complicated since incentives cause researchers to substitute between early and late registration. The following proposition discusses conditions under which the substitution overwhelms the direct effect, resulting in a net increase in the fraction of studies that register under a late ban:

Proposition 1 (Overall Implications on a Late Registration Ban)**.** *Suppose $g_0 = g_\gamma$ and fix all other parameters besides c_R, c_E, and the distribution over s_1. There exists δ, such that banning late registration increases the overall number of registrations if $\max_{s_2} \hat{p}_{\bar{s}_1}(s_2) - \hat{p}_{\underline{s}_1}(s_2) < \delta$, for some set of c_R, c_E (in particular, c_E small and c_R sufficiently small but positive).*

The intuition is as follows. When deciding when to register the experiment, the researcher faces a tradeoff between the option value of delay and the potential to signal their confidence based on their initial information. When the initial signal is not too informative, the signalling benefit is low relative to the potentially significant option value. However, under a late ban, the researcher has no option value, and the tradeoff is instead between the expected benefit from registering or not. This larger difference induces them to register earlier, even when the initial information is

less favorable. The large increase in preregistration can overwhelm the lack of late registrations under a ban, leading to an overall increase in registrations. We note that this argument requires the registration cost to be intermediate—if it is too low, then there is no option value, but if it is too high, then the expected benefit may not be worth the cost.

The general comparative statics on the impacts of a late ban emerge by studying the incentives of the indifferent type:

Proposition 2 (Other Implications of a Late Registration Ban)**.** *Suppose Assumption 2 holds, and that the researcher's indifference conditions determine a partitional equilibrium. Then there exists an equilibrium under a late registration ban where:*

- *A weakly larger fraction of experiments preregister, and*
- *Weakly fewer experiments are started*

These increases are strict if the threshold signals $s^*_{1,\emptyset}$, $s^*_{1,R}$ and $s^*_{2,R}$ are all distinct and interior.

The proof and intuition are straightforward and come from considering the incentives of the marginal researcher indifferent between actions (i.e. experimentation and registration). As discussed above, banning late registration eliminates the researcher's option value from registering late. Thus the researcher that was marginal between registration decisions when the late registration is allowed will strictly prefer to register (early) under a ban. Similarly, the researcher that was marginal between experimenting or not when late registration is allowed now will strictly prefer to not experiment. We verify that the former change leads to more experiments preregistering, whereas the latter change leads to fewer experiments starting. The effect of diminishing the value of experimentation without registration highlights a potential trade-off between inducing preregistration and inducing experimentation.

4.2.2 Environmental Comparative Statics

Our second finding articulates conditions under which increasing the informativeness of registrations causes an increase in preregistrations. The logic behind this result closely follows our

previous comparative statics. Namely, this change increases the payoff to preregistration, and so encourages researchers who previously chose to delay to instead register early:

Proposition 3 (Informativeness Comparative Static)**.** *Suppose Assumption 2 holds and that the researcher's indifference conditions determine a partitional equilibrium. Consider a change in γ that makes preregistered experiments more (Blackwell) informative. If $b_R(\hat{p})$ is strictly convex, then there exists an equilibrium where the first period registration threshold weakly decreases (and strictly if the threshold is interior).*

Note that convexity is necessary in order to ensure that researchers gain from having more informative experiments.[36]

Insofar as the ultimate goal of registries is to maximize preregistration, we then recommend changes to the registration process that increase the informativeness of the subsequent experiment. These changes may include but are not limited to (1) requiring more detailed information about the experimental design at the time of registration, (2) requiring a pre-analysis plan, or (3) providing a mechanism for eligible subjects, be they individuals or communities, to join an experiment (as is possible with ClinicalTrials.gov). We emphasize that this is a normative statement. The issue is that increasing the informativeness of registration in this manner likely raises the cost of registration (see, for instance, Olken (2015) for discussion of the significance of planning costs in the context of pre-analysis plans) which may counteract the desired effect.

4.3 Numerical Calibration

We conclude by using a numerical calibration of the above model to explore the impact of banning late registration as suggested by Propositions 1 and 2. To do so, we need to specify an information acquisition technology and payoff functions. For the information acquisition technology, we let the first and second period signals have the distribution

- If $\theta = T$, then $s_t \sim f(s_t \mid T) \propto s_t, s_t \in [\underline{s}_t, 1 - \underline{s}_t]$
- If $\theta = F$, then $s_t \sim f(s_t \mid F) \propto (1 - s_t), s_t \in [\underline{s}_t, 1 - \underline{s}_t]$.

[36] Libgober (2020) 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 $\theta = T$.

for $t = 1$ and $t = 2$ respectively. We assume that $0 < \underline{s}_1 < 0.5$ in order to keep the first period belief bounded away from 1 and 0. And we take $\underline{s}_2 = 0$. Note that the second period signal is conditionally independent of the first period signal and that the informativeness of the first period signal is decreasing in $\underline{s}_1$. For simplicity, we then assume that the payoff functions are linear. Specifically, we take $b_R(\hat{p}) = \hat{p}$ and $b_N(\hat{p}) = 0.8\hat{p}$. This choice reflects that registration is required for publication in the AEA journals. This choice also ensures that the signal informativeness does not influence payoffs directly.

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. Without much loss of generality, we take $c_E = 0$. And guided by the observed timing of registrations with the AEA RCT Registry, we focus on values for $\underline{s}_1$, p_0, and c_R that produce equilibria wherein the percent of RCTs that preregister closely matches the percent of RCTs that register late.

Table IX presents the resulting equilibria. Columns 1 through 3 report the input $\underline{s}_1$, p_0, and c_R. Column 4 presents the percent of RCTs that preregister in equilibrium. Column 5 confirms that this value match the percent of RCTs that register late. Finally, 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.

Of interest, Table IX, Column 7 reports the registration rate (which is also the preregistration rate) under a ban on late registration. We find that, in all cases, banning late registration causes a sharp increase in preregistration. At the least, the percent of experiments that preregister nearly doubles. We also find that, in many cases, banning late registration causes an increase in overall registration—with the benefit being higher when the first period signal is less informative. When the first period signal is more informative ($\underline{s}_1 = 0.33$ and $\underline{s}_1 = 0.35$), banning late registration causes a small decline in overall registration. Whereas when the first period signal is less informative ($\underline{s}_1 = 0.38$ and $\underline{s}_1 = 0.4$), banning late registration causes a significant increase in overall registration. This trend is monotonic through other values that we used for $\underline{s}_1$.

In particular, these results demonstrate the empirical relevance of Proposition 1. The calibrations confirm that banning late registration increases overall registration under parameterizations

of the model that generate qualitatively similar patterns to the AEA RCT Registry data. Of course, our simple model omits other elements guiding registration decisions that may be significant, and we caution against the assertion that banning late registration *must* increase overall registration. Nevertheless, insofar as early registrations may be especially valuable, a late ban would likely be beneficial.

5 Conclusion

This paper provides a relatively sobering assessment of the AEA RCT Registry—suggesting that thus far it has not been transformative in tackling the major issues at hand. Most experimentalists do not register and many registrations are done for trials that are already at the submission phase. Perhaps most disconcerting is that even when registrations are completed, they often do not provide enough information to attenuate p-hacking concerns. Hence, even in the best case scenarios, we see limited progress towards solving the file drawer problem and p-hacking.

We then provide several policy suggestions informed by a simple theoretical model of registration based on neoclassical economic assumptions. Foremost, we recommend that the AEA RCT Registry prohibit the registration of RCTs after they have already begun. Further, we recommend changes to the registration procedure that would increase the information content of the subsequent RCT. These may include (1) requiring more detailed information about the experimental design at the time of registration, (2) requiring a pre-analysis plan, or (3) providing a mechanism for eligible subjects, be they individuals or communities, to join a trial.

Both proposals target maximizing preregistration. Yet, the model also highlights the trade-off between registration costs and the number of experiments that are started in equilibrium. This motivates examining low cost ways to increase the specificity of the registration itself in order to directly increase the registry's ability to attenuate p-hacking. In this spirit, we recommend requiring that the researcher submit their IRB materials as part of the registration. In our experience, IRB materials contain enough detail to help tie the researcher to a particular experimental design. This approach avoids large additional costs since researchers can simply upload IRB forms that have already been completed.

It is important to acknowledge that much of the behavior regarding registration is undoubtedly guided by norms. In our model, this takes the form of treating the benefits and costs as exogenous. Certain norms might make publishing without preregistration very difficult. If this were to occur, our analysis suggests this 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 cognizant of, but which our analysis formalizes.

Where will changes leave us 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 (2019) and Glaeser (2008)). We view it as important to take such concerns seriously when considering optimal policy in the knowledge creation market.

References

Abadie, Alberto. Forthcoming. "Statistical Non-Significance in Empirical Economics." *American Economic Review: Insights*.

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.

Anderson, Michael L., and Jeremy Magruder. 2017. "Split-Sample Strategies for Avoiding False Discoveries."

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.

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.

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.

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

Coffman, Lucas, Muriel Niederle, and Alistair Wilson. 2017. "A Proposal to Organize and Promote Replications." *American Economic Review: Papers and Proceedings*, 107(5): 41–45.

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.

Di Tillio, Alfredo, Marco Ottaviani, and Peter N. Sorenson. 2019. "Strategic Sample Selection."

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.

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.

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.

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.

Friedman, Milton. 1953. *Essays in Positive Economics.* University of Chicago press.

Gentzkow, Matthew, and Emir Kamenica. 2016. "A Rothschild-Stiglitz Approach to Bayesian Persuasion." *American Economic Review, Papers and Proceedings*, 106(5): 597–601.

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.

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.

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.

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.

Libgober, Jonathan. 2020. "False Positives and Transparency." *American Economic Journal: Microeconomics*, Forthcoming.

Liu, Shuo, and Harry Pei. 2020. "Monotone Equilibria in Signaling Games." *European Economic Review*, 124.

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.

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.

McCloskey, Adam, and Pascal Michaillat. 2020. "Incentive-Compatible Critical Values." *arXiv preprint arXiv:2005.04141*.

Meager, Rachael. 2019. "Understanding the Average Impact of Microcredit Expansions: A Bayesian Hierarchical Analysis of Seven Randomized Experiments." *American Economic Journal: Applied Economics*, 11(1): 57–91.

Mill, John Stuart. 1836. "On the Definition of Political Economy; and on the Method of Investigation Proper to it."

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.

Ofosu, George K, and Daniel N Posner. 2019. "Pre-Analysis Plans: A Stocktaking."

Olken, Benjamin A. 2015. "Promises and Perils of Pre-Analysis Plans." 3.

Oostrom, Tamar. 2020. "Funding of Clinical Trials and Reported Drug Efficacy."

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

Robinson, Joan. 1977. "What Are the Questions?" *Journal of Economic Literature*, 1318–1339.

Samuelson, Paul A, and William D Nordhaus. 1985. *Economics.* McGraw-Hill.

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

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

Vivalt, Eva. Forthcoming. "How Much Can We Generalize from Impact Evaluations?" *Journal of the European Economic Association*.

Young, Alwyn. 2018. "Channeling Fisher: Randomization Tests and the Statistical Insignificance of Seemingly Significant Experimental Results." *Quarterly Journal of Economics*, 134(2): 557–598.

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.

A Tables

Table I: Number of published papers with an RCT and fraction that registered with the AEA RCT Registry

	Number Published			Fraction Registered		
	2017	2018	2019 Q1 & Q2	2017	2018	2019 Q1 & Q2
Journal						
AEJ-M	2	1	4	0.00	0.00	0.00
AEJ-EP	2	2	2	0.00	0.50	1.00
AEJ-AE	9	13	8	0.67	0.62	0.50
AER	7	6	5	0.71	0.50	0.80
Develop Econ	12	19	10	0.33	0.42	0.50
Exp Econ	28	29	18	0.00	0.00	0.00
JPE	4	3	2	0.00	0.33	0.00
NEP-EXP	36	32	38	0.03	0.09	0.13
QJE	6	8	3	0.50	0.62	1.00
ReStud	3	7	9	0.33	0.43	0.22

Table II: Number of field and lab RCTs published over 2018-2019Q2 by AEA journals along with fraction of these published RCTs that register

	Number Published		Fraction Registered	
	Field	Lab	Field	Lab
Journal				
AEJ-AE	19	2	0.63	0.0
AEJ-EP	4	0	0.75	NaN
AEJ-M	0	5	NaN	0.0
AER	11	0	0.64	NaN

Table III: Assessment of the extent to which 300 randomly chosen AEA RCT Registry preregistrations precisely specify their primary outcomes

	Mean	Std	Min	10%	25%	50%	75%	90%	Max
Number of Outcomes	3.16	2.33	0.0	1.0	1.5	3.0	4.00	6.0	20.5
Minimumly Restrictive Outcome	1.90	1.07	0.0	0.5	1.0	2.0	2.50	3.5	4.5
Maximumly Restrictive Outcome	2.40	1.03	0.0	1.0	1.5	2.5	3.00	4.0	4.5
Median Restrictive Outcome	2.16	0.99	0.0	1.0	1.5	2.0	2.75	3.5	4.5
Outcome Changed (Yes/No)	0.04	0.16	0.0	0.0	0.0	0.0	0.00	0.0	1.0
Sample Changed (Yes/No)	0.05	0.18	0.0	0.0	0.0	0.0	0.00	0.0	1.0

Notes: Preregistrations were randomly sampled from the period May 15, 2013 to April 1, 2019. Each registration was assessed by two RAs. The values presented are based on the average of the two assessments. The RAs were instructed to market unspecific outcomes as a 0 and very specific outcomes as a 5. The instructions (which include a scoring example) are presented in Appendix D.

Table IV: 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	0.90	0.22	0.0	0.65	0.95	1.0	1.0	1.0	1.0
Number of Additional Outcomes	0.48	1.07	0.0	0.00	0.00	0.0	0.5	2.0	7.0
Number of Missing Outcomes	0.39	0.84	0.0	0.00	0.00	0.0	0.5	1.0	4.0

Notes: Working papers were found for 119 of the 300 preregistrations.

Table V: Number of published papers with an RCT that registered with the AEA RCT Registry, number whose RCT started after 2013, and fraction of papers whose RCT started after 2013 that preregistered

	Number Registered			Started Post 2013			Fraction Preregistered		
Journal	2017	2018	2019 Q1 & Q2	2017	2018	2019 Q1 & Q2	2017	2018	2019 Q1 & Q2
AEJ	0	0	0	0	0	0	NaN	NaN	NaN
AEJ EP	0	1	2	0	0	1	NaN	NaN	1.00
AEJ-AE	6	8	4	0	2	2	NaN	0.00	0.00
AER	5	3	4	1	3	3	1.0	0.67	0.33
Develop Econ	4	8	5	0	2	1	NaN	0.00	1.00
Exp Econ	0	0	0	0	0	0	NaN	NaN	NaN
JPE	0	1	0	0	0	0	NaN	NaN	NaN
NEP-EXP	1	3	5	1	2	4	1.0	0.00	0.50
QJE	3	5	3	1	3	2	0.0	0.33	0.00
ReStud	1	3	2	0	2	1	NaN	1.00	0.00

Table VI: Three primary investigators who knew about the AEA RCT Registry, but failed to pre-register multiple future RCTs

(a) Investigator A

	First Registered On	Start Date	Intervention Start Date
0	2014-10-07	2014-03-04	2014-04-29
1	2015-10-02	2015-08-14	2015-09-15
2	2016-04-06	2016-04-08	2016-04-29
3	2018-03-14	2016-02-09	2016-12-15
4	2018-06-26	2018-03-23	2018-08-01
5	2018-11-20	2019-02-01	2019-05-01
6	2019-03-16	2019-04-15	2019-04-22
7	2019-03-26	2016-02-09	2017-01-16

(b) Investigator B

	First Registered On	Start Date	Intervention Start Date
0	2014-03-31	2014-03-26	2014-03-26
1	2014-10-06	2014-08-01	2014-09-11
2	2015-10-16	2015-03-21	2015-10-27
3	2015-10-23	2015-09-28	2015-11-04
4	2016-12-14	2015-10-27	2016-12-15
5	2017-10-13	2016-04-01	2016-04-01

(c) Investigator C

	First Registered On	Start Date	Intervention Start Date
0	2014-05-07	2014-04-30	2014-04-30
1	2018-02-02	2018-05-01	2018-05-01
2	2018-09-10	2018-08-06	2018-08-06
3	2019-03-08	2017-01-01	2017-01-01

Notes: Each primary investigator here registered their first trial in 2014, proceeded to register multiple new trials (started after 2014) late, and registered their most recent trial over a year after the intervention began.

Table VII: 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	1.95	1.18	0.5	1.0	1.0	1.5	3.00	4.0	6.0
Minimumly Restrictive Outcome	2.77	0.99	1.0	1.5	2.0	3.0	3.50	4.0	5.0
Maximumly Restrictive Outcome	3.35	0.99	1.0	2.0	3.0	3.5	4.00	4.5	5.0
Median Restrictive Outcome	3.04	0.90	1.0	2.0	2.5	3.0	3.52	4.0	5.0
Outcome Changed (Yes/No)	0.51	0.45	0.0	0.0	0.0	0.5	1.00	1.0	1.0
Sample Changed (Yes/No)	0.64	0.45	0.0	0.0	0.0	1.0	1.00	1.0	1.0

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. Each registration was assessed by four RAs. The values presented are based on the median of the four assessments.

Table VIII: 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	0.41	0.0	0.32	0.58	1.0	1.0	1.0	5.0
Number of Additional Outcomes	0.39	0.94	0.0	0.00	0.00	0.0	0.5	1.0	10.5
Number of Missing Outcomes	0.38	0.84	0.0	0.00	0.00	0.0	0.5	1.5	7.0

Notes: Working or published papers were found for 278 of the 300 preregistrations.

Table IX: 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	3.87	3.85	7.72	6.49
0.33	0.159	0.120	3.99	3.99	7.98	6.24
0.33	0.198	0.140	3.96	3.95	7.91	5.71
0.35	0.149	0.100	5.3	5.3	10.6	9.75
0.35	0.169	0.110	5.46	5.46	10.92	9.66
0.35	0.190	0.120	5.55	5.58	11.13	9.45
0.38	0.148	0.080	7.88	7.85	15.73	18.86
0.38	0.172	0.090	8.28	8.26	16.54	19.07
0.38	0.198	0.100	8.68	8.64	17.32	19.18
0.40	0.102	0.050	8.98	8.88	17.87	28.88
0.40	0.167	0.075	10.06	10.57	20.63	30.31
0.40	0.242	0.100	12.1	12.02	24.12	31.84

Notes: Each simulation takes $c_E = 0$, $\underline{s}_2 = 0$, $b_R(\hat{p}) = \hat{p}$, and $b_N(\hat{p}) = 0.8\hat{p}$. Columns 1 through 3 report the input $\underline{s}_1$, p_0, and c_R. Column 4 presents the percent of experiments that preregister in equilibrium. Column 5 confirms that this value match the percent of experiments that register late. Column 6 displays the total registration rate. Column 7 reports the registration rate (which is also the preregistration rate) under a ban on late registration.

B Figures

Figure I: The AEA RCT Registry. Trials that register late are marked with a gray clock and trials that preregister are marked with an orange clock.

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

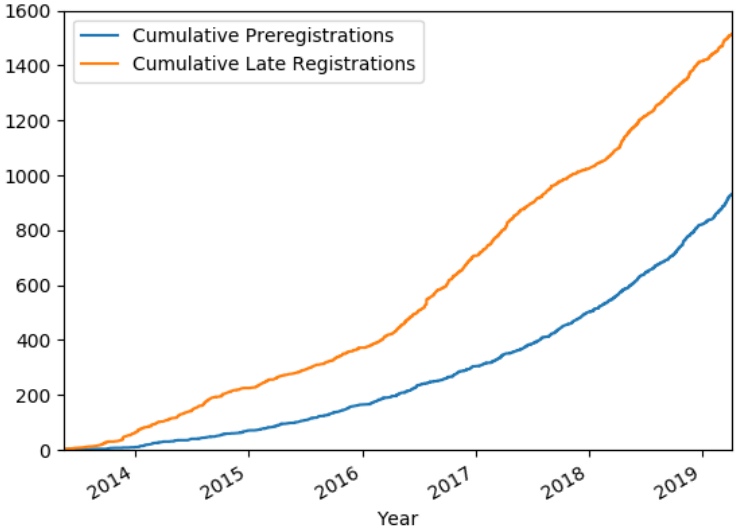


Figure III: Number of AEA RCT preregistrations and late registrations by quarter

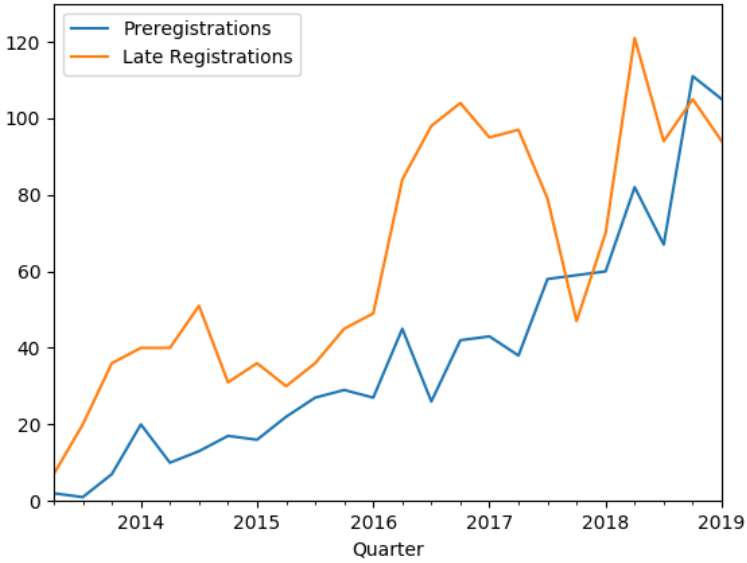


Figure IV: Days between intervention start and AEA RCT registration for RCTs started after January 1, 2014. Positive values indicate that the intervention began after the registration.

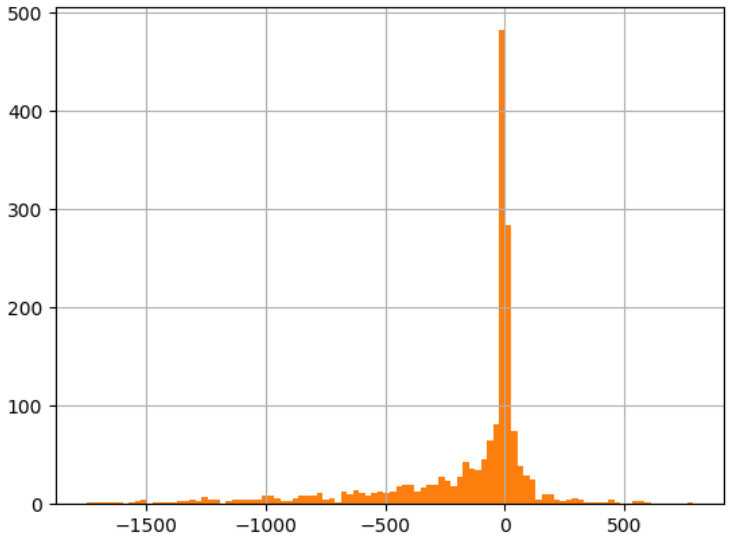


Figure V: Registration in the Registry for International Development Impact Evaluations (RIDIE) and in the Evidence in Governance and Politics (EGAP) registry over time. Solid lines display the total number of registrations while dashed lines present economics registrations

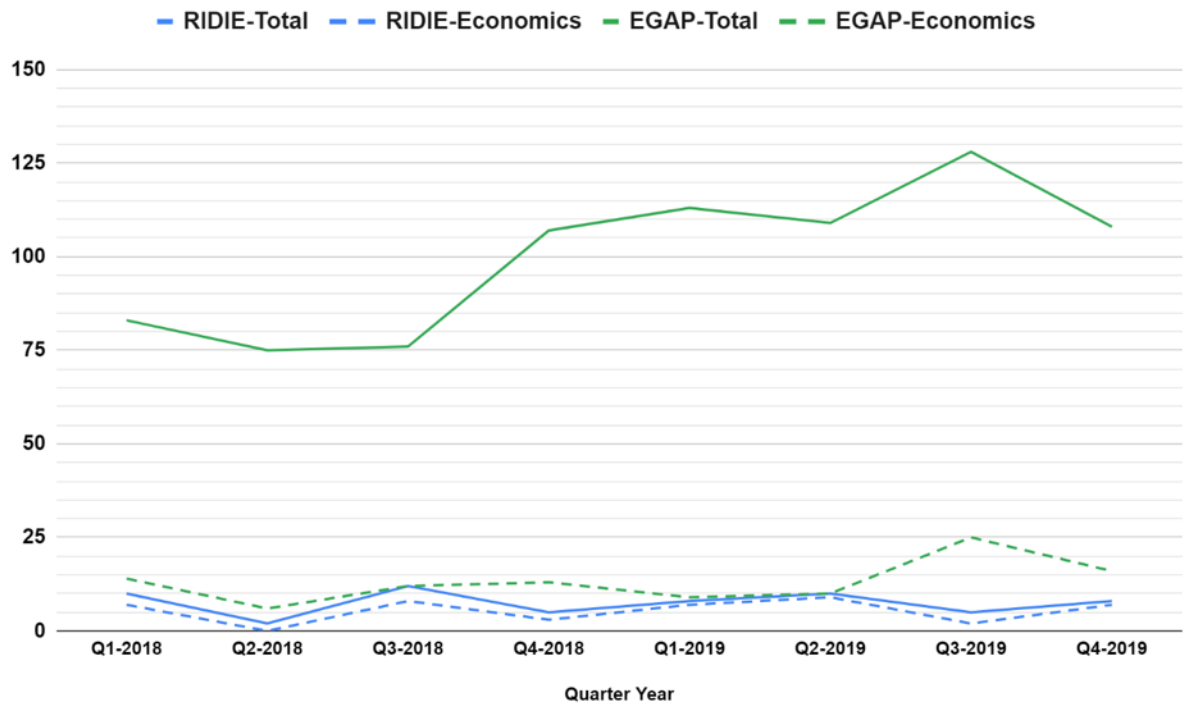


Figure VI: Timing of moves in the model

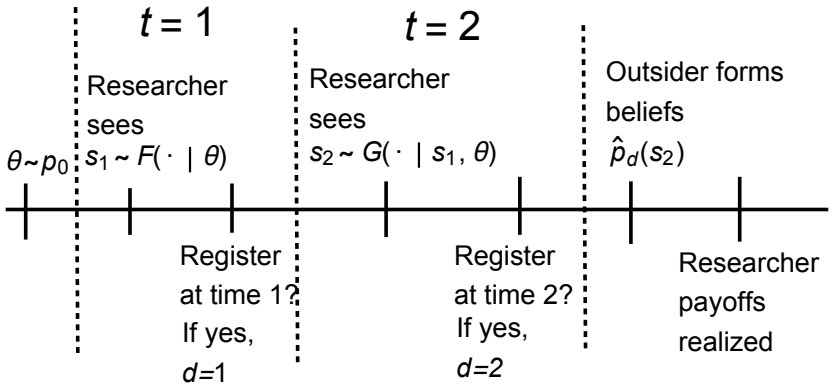
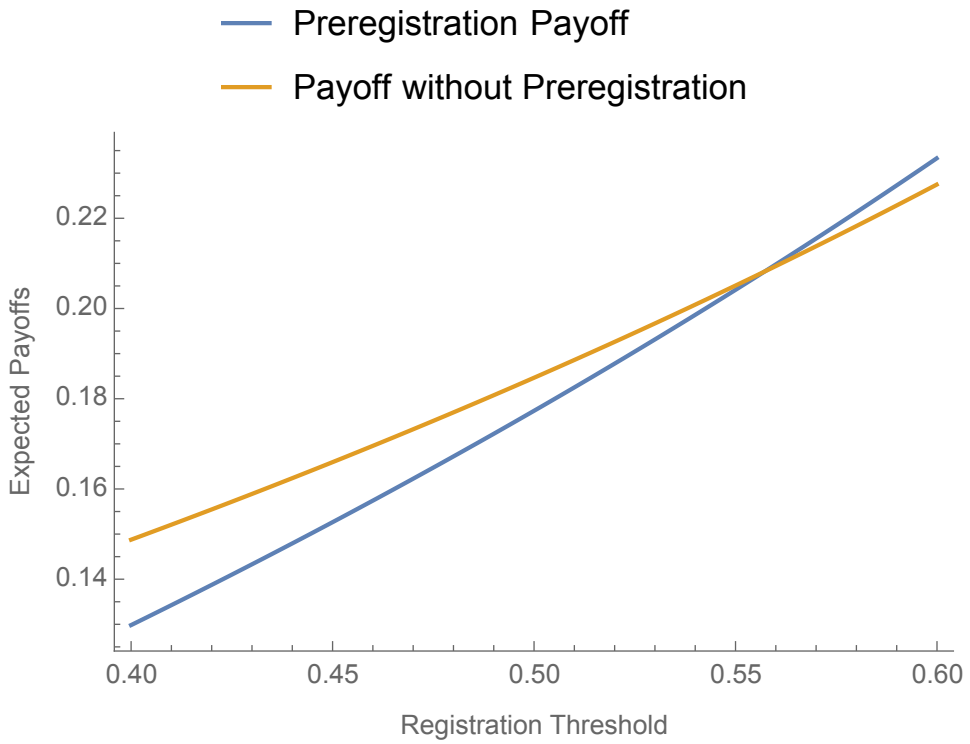


Figure VII: Researcher payoff upon receiving signal $s_1 = s^*_{1,R}$ assuming the equilibrium registration threshold is conjectured by the outsider to be $s^*_{1,R}$



Notes: Assumption 2 is seen to hold, since at the left endpoint, the orange line is higher than the blue line. The (conjectured) equilibrium threshold is the intersection point of these two lines. Payoffs and information structure are as in Table IX, with $\underline{s}_1 = .4$, $c_R = .1$, and $p_0 = .25$.

C Proofs

This appendix is organized as follows. First, we present proofs related to equilibrium beliefs. Then, we present proofs related to finding and checking partitional equilibria. With these results in hand, we present proofs of the comaprative statics from the main text. We subsequently verify the conditions for partitional equilibria which we use for our numerical calibration, and conclude with some additional discussion of the microfoundations of preferences which would lead to Assumption 1 being satisfied.

C.1 Properties of Beliefs

Proof that $\frac{d}{ds_1}\log f(s_1 \mid T) \geq \frac{d}{ds_1}\log f(s_1 \mid F) \Rightarrow \hat{p}(s_1)$ *is increasing.* While likely familiar, this argument is included for completeness. Note that

$$\hat{p}(s_1) = \frac{\mathbb{P}[\theta = T]f(s_1 \mid T)}{\mathbb{P}[\theta = T]f(s_1 \mid T) + \mathbb{P}[\theta = F]f(s_1 \mid F)}.$$

We take the derivative and obtain $\hat{p}'(s_1)$ has the same sign as:

$$\begin{aligned}
&\left(\mathbb{P}[\theta = T]f(s_1 \mid T) + \mathbb{P}[\theta = F]f(s_1 \mid F)\right)\mathbb{P}[\theta = T]f'(s_1 \mid T) \\
&\qquad - \left(\mathbb{P}[\theta = T]f'(s_1 \mid T) + \mathbb{P}[\theta = F]f'(s_1 \mid F)\right)\mathbb{P}[\theta = T]f(s_1 \mid T) \\
&\quad = \mathbb{P}[\theta = F]f(s_1 \mid F)\mathbb{P}[\theta = T]f'(s_1 \mid T) - \mathbb{P}[\theta = F]f'(s_1 \mid F)\mathbb{P}[\theta = T]f(s_1 \mid T),
\end{aligned}$$

which, since $\mathbb{P}[\theta = T] \in (0, 1)$, is greater than 0 if and only if:

$$\frac{f(s_1 \mid F)}{f'(s_1 \mid F)} \geq \frac{f(s_1 \mid T)}{f'(s_1 \mid T)} \Leftrightarrow \frac{d}{ds_1}\log f(s_1 \mid T) \geq \frac{d}{ds_1}\log f(s_1 \mid F),$$

as desired. □

Lemma C.1. *In any equilibrium,* $\hat{p}_1(s_2)$ *is increasing.*

Proof of Lemma C.1. As it will be useful for a later proof, we consider $\hat{p}_d(s_2)$ for any d. Let us first consider the fictitious environment where s_1 were observable to the outsider. Call this

$\tilde{p}_{s_1,d}(s_2)$. Differentiating $\tilde{p}_{s_1,d}(s_2)$ (which is equal to $\hat{p}_d(s_2)$ once integrating over the distribution of s_1 conditional on d), we have that it is proportional to:

$$g'_\gamma(s_2 \mid T)f(s_1 \mid T)\mathbb{P}[T] \cdot g_\gamma(s_2 \mid F)f(s_1 \mid F)\mathbb{P}[F] \\ - g'_\gamma(s_2 \mid F)f(\tilde{s}_1 \mid F)\mathbb{P}[F]g_\gamma(s_2 \mid T)f(s_1 \mid T)\mathbb{P}[T].$$

Following similar logic as the previous proof, we have:

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

which must be greater than 0 since $\frac{d}{ds_2}\log g_\gamma(s_2 \mid T) \geq \frac{d}{ds_2} g_\gamma(s_2 \mid F)$, and in addition since all other densities and probabilities are positive as well.

It remains to show that $\hat{p}_d(s_2)$ is increasing in s_2. Letting $\sigma(\cdot \mid s_2, d)$ denote the equilibrium measure over s_1 given s_2 and d, by the martingale property of beliefs:

$$\hat{p}_d(s_2) = \int_{s_1} \tilde{p}_{s_1,d}(s_2)\sigma(s_1 \mid s_2, d)ds_1,$$

as $\hat{p}_d(s_2)$ is simply the expectation over $\tilde{p}_{s_1,d}(s_2)$ after observing s_1, in addition to s_2 and d.

Now, if $d = 1$, then $\sigma(s_1 \mid s_2, d)$ is mechanically independent of s_2 on its support. So,

$$\hat{p}_1(s'_2) = \int_{s_1} \tilde{p}_{s_1,1}(s'_2)\sigma(s_1 \mid d = 1)ds_1 \geq \int_{s_1} \tilde{p}_{s_1,1}(s''_2)\sigma(s_1 \mid d = 1)ds_1 = \hat{p}_1(s''_2),$$

as claimed. □

Lemma C.2. *Under Assumption 1, the outsider's belief satisfies $\hat{p}_2(s_2) = \hat{p}_\emptyset(s_2)$ in equilibrium (that is, the belief following late registration is equal to the belief following no registration).*

Proof. Let $h(s_1, s_2 \mid \mathbf{1}[d = 1], \theta)$ denote the joint distribution of signals given the state θ and registration decision at time 1 and γ, and let $\sigma(d \mid s_1, s_2)$ denote the probability the registration

decision is d given signals s_1 and s_2. Note that $\hat{p}_d(s_2)$ is:

$$\frac{\int_{s_1} \sigma(d \mid s_1, s_2) h(s_1, s_2 \mid \mathbf{1}[d=1], T) \mathbb{P}[\theta = T] ds_1}{\int_{s_1} \sigma(d \mid s_1, s_2) h(s_1, s_2 \mid \mathbf{1}[d=1], T) \mathbb{P}[\theta = T] ds_1 + \int_{s_1} \sigma(d \mid s_1, s_2) h(s_1, s_2 \mid \mathbf{1}[d=1], F) \mathbb{P}[\theta = F] ds_1},$$

noting that σ also includes the event that the researcher undertakes the experiment.

Consider any signal s_2 where the researcher were to mix over the registration decision. At any such signal, we must have $b_R(\hat{p}_2(s_2)) - c_R = b_N(\hat{p}_\emptyset(s_2))$, since otherwise there would be a strict incentive to deviate. Since $b_R(\hat{p}) - b_N(\hat{p})$ is increasing, there can only be at most one belief where this indifference is satisfied, say p^*. Since there is only one belief that can be induced in order for the sender to be willing to mix over registration, we must therefore have that the registration decision is uninformative, i.e., that the same belief is induced for each registration decision.

However, note that for any strategy, $\mathbb{E}_{d \sim \sigma}[\hat{p}_d(s_2)]$ is equal to the probability that $\theta = T$ conditional on s_2 alone, by the martingale property of beliefs. Following the proof of Lemma C.1 to consider the event that $d \neq 1$, this belief is increasing in s_2, since it is increasing in s_2 for all s_1 and thus also increasing when we take an expectation over s_1 as well, for any measure over s_1. Since we require $\hat{p}_2(s_2) = \hat{p}_\emptyset(s_2) = p^*$ in order for the researcher to be willing to mix, it follows that there can only be a single signal s_2 where the registration decision is informative.

This shows that we must have a deterministic registration decision, for almost every s_2, and in particular one that is not informative. Thus, $\sigma(d \mid s_1, s_2)$ is independent of s_1 on the support of the signals that delay registration. Upon inspecting the expression of the outsider's beliefs, we observe that the second period registration decision does not influence the distribution over s_1. □

Lemma C.3. *Under Assumption 1, $\hat{p}_2(s_2) = \hat{p}_\emptyset(s_2)$ is increasing in s_2.*

Proof of Lemma C.3. Replicating the proof of Lemma C.1, Lemma C.2 implies that $\sigma(d \mid s_1, s_2)$ is also independent of s_2 when $d \in \{\emptyset, 2\}$, under Assumption 1. Thus, the same argument applies to this case as well. □

Remark 1. *The above arguments rely upon the increasing differences conditions in order to ensure that late registration does not convey information. Without this assumption, mixed strategy equilibria may emerge and cannot be ruled out immediately. To see why, suppose the increasing differ-*

ences condition is violated. Then, we can find p_1, p_2 such that $b_R(p_1) - b_N(p_1) = b_R(p_2) - b_N(p_2)$. We can then also find c_R such that $b_R(p_1) - c_R = b_N(p_1)$, which also implies $b_R(p_2) - c_R = b_N(p_2)$.

Now, to illustrate that this Lemma can fail, suppose for simplicity that the distribution of beliefs as a function of s_2, integrating over s_1, is a strict subset of (p_1, p_2). Then if the first period signal (among types that do not register) is sufficiently informative, there exists a strategy $\sigma : S_1 \times S_2 \to \{\emptyset, 2\}$ such that $\hat{p}_\emptyset(s_2) = p_1$ and $\hat{p}_2(s_2) = p_2$; that this can be done under the stated conditions follows immediatley from, for instance, Gentzkow and Kamenica (2016); their result implies that, given any "integrated out" second period belief $\hat{p}(s_2) \in (p_1, p_2)$, the martingale constraint alone dictates whether a distribution of beliefs can emerge under some information structure if s_1 fully reveals the state, and that this conclusion holds as long as the first period signal is sufficiently informative (how close to fully informative will depend on the parameters). By the stated conditions, given these beliefs, the researcher is indifferent between registration decisions. Note this actually describes two *equilibria; one could instead let $\hat{p}_\emptyset(s_2) = p_2$ and $\hat{p}_2(s_2) = p_1$.*

While many of the features of the above construction rely upon the assumption that the second period signal alone puts beliefs in (p_1, p_2), this is not strictly necessary and the argument would still work if there were registration strategies which put the support of the outsider's second period belief in $\{p_1, p_2\}$. We also do not see an easy way of ruling out "sufficiently informative" first period signals a priori. These equilibria are simpler to see in the extreme case where $b_N(p) = b_R(p) - c_R$ for all p. In this case, any *second period registration strategy forms an equilibrium for the researcher, including ones which reveal information about the first period signal. While theoretically interesting, our conjecture is that most researchers in practice would not vary their registration decision after experimenting based on information not conveyed in an experiment's results. As we view Assumption 1 as appealing, we leave an analysis of other conditions which would yield Lemma C.2 to future work.*

Lemma C.4. *Consider any partitional equilibrium, where decision d is taken by researchers with $s_1 \in [s_{min}, s_{max}]$. Then $\hat{p}_d(s_2)$ is uniformly increasing (i.e., increasing for all s_2) in s_{min}.*

Proof. We use the above characterization of $\hat{p}_d(s_2)$ which uses the martingale property of beliefs, i.e. that $\hat{p}_d(s_2) = \int_{s_{min}}^{s_{max}} \hat{p}_{d,s_1}(s_2) \mathbb{P}[s_2 \mid s_1 \in [s_{min}, s_{max}]]$. Consider $\mathbb{P}[s_2 \leq s' \mid s_1 \in [s_{min}, s_{max}]]$.

We claim that this is FOSD larger if s_{min} increases. The result is immediate given the claim, since $\hat{p}_{d,s1}(s_2)$ is increasing in s_2, which means integrating against an FOSD larger distribution leads to $\hat{p}_d(s_2)$ increasing, for all s_2. On the other hand, the claim is immediate as well; if we consider drawing s_1 and then using the resulting draw to determine s_2, increasing s_{min} simply increases the probability of using a more favorable s_1 draw. Hence the conclusion follows. □

C.2 Partitional Equilibria

In this Section, we walk through some additional details on partitional equilibria and conditions which ensure the indifference condition determines an equilibrium. While our sufficient condition for an equilibrium to be a partitional threshold equilibrium is restrictive, it is useful in that we are able to numerically verify it in many cases of interest, particularly the region where we calibrate our model to the data.

We first show that Assumption 1 implies the existence of the threshold equilibrium strategies in the second period:

Proposition 4. *In any equilibrium under Assumption 1, there exists a threshold s_2^* such that a researcher who has not registered at time 1 will do so at time 2 if $s_2 > s_2^*$.*

Proof. By Lemma C.2, the outsider's belief at time 2 depends only on s_2 and $\mathbf{1}[d = 1]$. So consider the range of $b_R(p) - c_R - b_N(p)$ over all s_2 given $d \neq 1$; note that this is either always positive, always negative, or positive for some values and negative for others. In the first two cases, the registration decision is degenerate and hence trivially of a threshold form (taking the thresold to be outside of the support of the signal distribution). In the latter case, since by assumption $b_R(p) - c_R - b_N(p)$ is increasing and continuous, by the intermediate value theorem we have there is some belief in the range of possible second period beliefs where this is equal to 0, say p^*, which corresponds to a signal s_2^*. Furthemore, we previously showed that the second period belief is increasing in s_2. Since it is also continuous in s_2, we thus have the researcher registers when $s_2 > s_2^*$ and not when $s_2 < s_2^*$, as desired. □

We now present our sufficient condition which ensures the existence of a partitional equilibrium

in the first period:

Definition 2. *We say that a registration strategy has increasing gains to early registration if:*

$$\int_{-\infty}^{\infty} b_R(\hat{p}_1(s_2))\mathbb{E}[g_\gamma(s_2 \mid \theta) \mid s_1] - b_R(\hat{p}_2(s_2))\mathbb{E}[g_0(s_2 \mid \theta) \mid s_1]ds_2 \tag{1}$$

is increasing in s_1.

Note that this condition does not depend on b_N; it says that given a researcher *will* register, it is even better to register early rather than late when the initial signal is higher.

Proposition 5. *Suppose a registration strategy satisfies increasing gains to early registration. Then under Assumption 1, then the first period registraiton decision must be of a partitional form.*

Proof of Proposition 5. Denote $\tilde{p}(s_1)$ as the *researcher's* belief that $\theta = T$ given a signal of s_1, and recall that $\tilde{p}(s_1)$ is a strictly increasing function of s_1. Furthermore, the researcher's benefit is independent of the *realized* s_1 (since this is not observed by the outsider). As a result, we can write the researcher's payoff without any reference to s_1 at all, and only the researcher's belief $\tilde{p}$. And to prove the theorem, it suffices to show that the payoff from registration increases more than the payoff from non-registration when $\tilde{p}$ increases.

Making this change of variables, we have the researcher's payoff is:

$$-c_R + \int_0^1 b_R(\hat{p}_1(s_2))\left(\tilde{p}f(s_2 \mid T) + (1-\tilde{p})f(s_2 \mid F)\right)ds_2. \tag{2}$$

The payoff from registration at time 2 is:

$$\int_{s_{2,R}}^1 \left(b_R(\hat{p}_2(s_2)) - c_R\right)\left(\tilde{p}f(s_2 \mid T) + (1-\tilde{p})f(s_2 \mid F)\right)ds_2 + \int_0^{s_{2,R}} b_N(\hat{p}_2(s_2))(\tilde{p}f(s_2 \mid T) + (1-\tilde{p})f(s_2 \mid F))ds_2.$$

Consider the difference between these two expressions, which can be written:

$$\int_{s_{2,R}}^{1} \left(b_R(\hat{p}_1(s_2)) - b_R(\hat{p}_2(s_2))\right)\left(\tilde{p}f(s_2 \mid T) + (1-\tilde{p})f(s_2 \mid F)\right) ds_2 \\ + \int_{0}^{s_{2,R}} (b_R(\hat{p}_1(s_2)) - c_R) - b_N(\hat{p}_2(s_2))(\tilde{p}f(s_2 \mid T) + (1-\tilde{p})f(s_2 \mid F))ds_2.$$

Rewriting this slightly, we wish to show that if:

$$\int_{0}^{1} \left(b_R(\hat{p}_1(s_2)) - b_R(\hat{p}_2(s_2))\right)\left(\tilde{p}f(s_2 \mid T) + (1-\tilde{p})f(s_2 \mid F)\right) ds_2 \\ + \int_{0}^{s_{2,R}} (b_R(\hat{p}_2(s_2)) - c_R) - b_N(\hat{p}_2(s_2))(\tilde{p}f(s_2 \mid T) + (1-\tilde{p})f(s_2 \mid F))ds_2 > 0.$$

then this also holds at any $\tilde{p}' > \tilde{p}$. Note that this expression considers the difference as the sum of two terms: The first term is the *belief increase* due to registration, and the second is the loss due to *option value*.

Now, if this is positive at some $\tilde{p}$ but not at $\tilde{p}' > \tilde{p}$, then it must be due to option value, since the first term is always positive, since the proposition considers equilibria where the increasing gains to early registration condition is satsfied. Thus, it suffices to show that the second integral is increasing in s_2.

To see this, first note that first order stochastic dominance is maintained under monotone transformations,[37] and that $\hat{p}_2(s_2)$ is a monotone transformation of s_2. As a result, consider the distribution over second period beliefs, say $g(p_2 \mid \theta)$. Since $f(s_2 \mid T)$ first order stochastically dominates $f(s_2 \mid F)$, we also have $g(p_2 \mid T)$ first order stochastically dominates $g(p_2 \mid F)$. we have this integral is:

$$\int_{0}^{s_{2,R}} (b_R(p_2) - c_R - b_N(p_2))(\tilde{p}g(p_2 \mid T) + (1-\tilde{p})g(p_2 \mid F))dp_2.$$

Now, recall $b_R(p_2) - b_N(p_2)$ is assumed to be increasing, and increases in $\tilde{p}$ yield increases in first order stochastic dominance shifts in the distribution over second period beliefs. Since the

[37] For a quick proof for reference, 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.

expectation of an increasing function of a random variable increases when the random variable distribution increases in first order stochastic dominance, we have that this integral increases as well.

We have thus showed that if some signal s_1 prefers to register, then so do all higher types as well. Likewise, if some signal s_1 prefers to not register, then so do all lower types. It follows that the registration decision partitions the support of the first period signal, as desired. □

Proposition 6. *Under Assumption 1, the first period experimentation decision takes a partitional form.*

Proof. Consider the researcher's payoffs from early registration, (2), as well as late registration, 2. As the proof of Proposition 5 states, 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. □

Note that this proposition does not rely upon the increasing gains to early registration condition. Indeed, it only relies upon the assumption that s_1 signals make $\theta = T$ more likely, and that b_R, b_N are increasing.

The following simple Corollary, which follows immediately from the above proofs, illustrates that to show that a particular partition is in fact an equilibrium, it suffices to check the increasing gains to early registration condition:

Corollary C.4.1. *Consider the strategy arrived at via the following algorithm:*

- *First, compute the second period beliefs that are indifferent between late registration and not registering, and*
- *Second, compute the first period signal which makes the researcher indifferent between registration decisions, given this signal.*

If the increasing gains to early registration condition is satisfied, then these thresholds define a partitional equilibrium.

Proof. Immediate from the above; given the indifference thresholds, higher first period signals imply higher payoffs to undertaking the experiment, and higher payoffs to registration. Hence the indifference conditions suffice to charactrize the equilibria. □

This corollary is used in our calibration.

C.3 Main Text Comparative Statics Proofs

Proof of Proposition 1. Consider a fictitious environment where early registration convinces outsiders that $s_1 = \overline{s}_1$, noting that early registration will give researchers lower payoff than this. Hence the payoff from early registration is at most:

$$-c_R + \int_{-\infty}^{\infty} b_R(\hat{p}_{\overline{s}_1}(s_2))\mathbb{E}_\theta[g(s_2 \mid \theta) \mid s_1]ds_2. \tag{3}$$

Since the worst-case from late registration is that the first period signal is $\underline{s}_1$, we have that the from late registration is at least:

$$\int_{-\infty}^{\infty} \max\{b_R(\hat{p}_{\underline{s}_1}(s_2)) - c_R, b_N(\hat{p}_{\underline{s}_1}(s_2)\}\mathbb{E}_\theta[g(s_2 \mid \theta) \mid s_1]ds_2. \tag{4}$$

Under the assumption that $\hat{p}_{\overline{s}_1}(s_2) - \hat{p}_{\underline{s}_1}(s_2) < \delta$, for some δ, continuity of b_i gives us that we can find some ε such that (4) is greater than:

$$-\varepsilon + \int_{-\infty}^{\infty} \max\{b_R(\hat{p}_{\overline{s}_1}(s_2)) - c_R, b_N(\hat{p}_{\overline{s}_1}(s_2)\}\mathbb{E}_\theta[g(s_2 \mid \theta) \mid s_1]ds_2.$$

If this equation is larger than (3), then we have that the payoff from late registration is higher than the payoff from early registration, which is our desired result. Subtracting this from (3) yields:

$$\overbrace{\varepsilon + \int_{-\infty}^{s_2^*} ((b_R(\hat{p}_{\overline{s}_1}(s_2)) - c_R) - b_N(\hat{p}_{\overline{s}_1}(s_2)))\mathbb{E}_\theta[g(s_2 \mid \theta) \mid s_1]ds_2}^{(*)}.$$

Now, in the limit we consider, we take the initial signal to be uninformative, but fix the informativeness of the second signal. While ε may be arbitrarily small given a sufficiently uninformative

first period signal, $(*)$ is bounded away from 0 and negative, as long as c_R is chosen so that some types would not register in the second period. Hence, it follows that this difference is negative. Therefore, the researcher's payoff from registering late is larger than registering early.

Now, recall $\hat{p}_{s_1}(s_2)$ is increasing in s_2, for all s_1. It follows that for all s_1, since $b_R(p) \geq b_N(p)$, with strict inequality for some beliefs, we further have, for all s_1:

$$\kappa(s_1) := \int_{-\infty}^{\infty} b_R(\hat{p}_{s_1}(s_2)) - b_N(\hat{p}_{s_1}(s_2))\mathbb{E}[g(s_2 \mid \theta) \mid s_1]ds_2 > 0.$$

Hence, as the initial signal becomes uninformative, we have that $\kappa(s_1) \to \kappa$, for all s_1. It follows that as long as $c_R < \kappa$, researchers prefer registration at time 1 to non-registration. Note that at $c_R = \kappa$, the researcher must have second period beliefs which would lead them to strictly prefer registering, meaning that $(*)$ is negative provided c_R is not too low.

We now make the main comparison of interest: if late registration is allowed, then by the above, the payoff from late registration is larger than the payoff from early registration, so all researchers register late. On the other hand, we also argued that there exists c_R such that researchers would still be willing to register early if forced to do so. Thus, we have that all researchers register early in this instance. We conclude that banning late registration leads to all researchers registering early, whereas some researchers never register, as claimed.

The above argument takes $c_E = 0$, and analogous reasoning shows this still applies when $c_E > 0$. □

Proof of Proposition 2. Note that a late ban is equivalent to adding the following to this researcher's payoff:

$$\int_{s_2^*}^{\infty} b_N(\hat{p}_2(s_2))\mathbb{E}_\theta[g_0(s_2 \mid \theta) \mid s_{1,R}^*] - (b_R(\hat{p}_2(s_2)) - c_R)\mathbb{E}_\theta[g_\gamma(s_2 \mid \theta) \mid s_{1,R}^*]ds_2.$$

By convexity, this term will increase if g_γ is replaced by g_0. On the other hand, by definition, when $s_2 > s_2^*$ we have $b_N(\hat{p}_2(s_2)) < b_R(\hat{p}_2(s_2)) - c_R$, so that the entire above expression is negative. Note that, by assumption, we cannot have $s_{1,R} = s_{1,\emptyset}$, since we are focused on the case where both counterfactuals lead to researchers registering late with positive probability.

Given this, we first show that weakly fewer experiments are conducted under a late ban, i.e. $s^*_{1,\emptyset}$ increases. Indeed, if this signal leads to a payoff of 0 from conducting an experiment when late registration is allowed, it therefore leads to negative payoff when late registration is banned. Hence the type that is indifferent between not conducting the experiment and conducting the experiment with delayed registration must increase.

Now consider $s^*_{1,R}$. Similarly, we have that now, these types strictly prefer to register. Note that $s^*_{1,R} = s^*_{1,\emptyset}$ coincides with the case where there is no delayed registration at all. By Assumption 2, the payoff of early registration in this instance is lower than delayed registration revealing the lowest type. Hence by the intermediate value theorem, there exists a new threshold value, say $\tilde{s}_{1,R}$ which makes the reseacher indifferent between delaying and not, with this threshold being between the original $s^*_{1,R}$ and $s^*_{1,\emptyset}$. □

Proof of Proposition 3. Follows the same reasoning as proposition 2; an increase in the informativeness of registration leads to a mean preserving spread in beliefs (see Blackwell (1953)). Hence because $b_{R,\gamma}$ is convex, the researcher who is indifferent between registering and not registering prior to an increase in the informativeness does stricly better by registering. The same reasoning allows us to conclude the threshold lowers. □

C.4 Additional Model Discussion

We first present some examples microfounding the increasing differences condition:

Example 1. *Suppose that whether publication ultimately occurs only depends on $\hat{p}_d(s_2)$, with this probability being denoted by $\pi(\hat{p}_d(s_2))$ 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 $(\beta_R - \beta_N)\pi(\hat{p}_d(s_2))$.*

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 in order 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, due to the fact that more possible journals are available).

Example 2. *Normalize the benefit of publication to 1, but suppose that the probability of publication is* $1-(1-\pi(\hat{p}_d(s_2)))^{\beta}$*, 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:*

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

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* $\beta_N=4$ *(an extreme view of the relative importance of top 5 publications), increasing differences reduces to the requirement that* $\pi(\hat{p}_d(s_2))<.2$ *(note that this condition implies that the maximum probability of publication is less than* 0.67*).*

To emphasize, 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.

C.5 Numerical Calibration

In this Section, we describe details in showing that our specification satisfies increasing gains to early registration. We first write out the outsider's beliefs, given that the first period signal is in some interval $[s_*,s^*]$, under the particular experimentation technology, and observation of the

signal s_2. This is:

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

Regarding the first period beliefs, we note that the expression for the first period is exactly 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 threshold 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 $\overline{s} \to 0$. We numerically verify that the condition in Proposition 5 is satisfied for all p_0 and possible first period belief thresholds. This calculation is done in Mathematica and is available from the authors' webpage or upon request.

D RA Instructions

D.1 Restrictiveness

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

D.2 Fidelity

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 ___

D.3 Experimental Economics Registrations

We instructed an RA to:

- Assemble a list of RCTs published in Experimental Economics between the years 2016 and 2019
- Find registrations corresponding to these RCTs via
    1. Searching Google for the paper title plus the word "register"
    2. Searching four registries (specifically https://aspredicted.org/, https://ridie.3ieimpact.org/, http://egap.org/content/registration, and https://cos.io/prereg/) for the paper title and for the authors