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

Eliot Abrams* Jonathan Libgober† John A. List‡

September 17, 2021

Abstract: We analyze one prominent policy solution to the credibility crisis in experimental research—research registries—with a focus on the AEA registry. We find that a majority of economics field experiments do not register, only half of registrations are prior to intervention, and most of these preregistrations lack sufficient detail. We compare these facts to ClinicalTrials.gov and find broad similarities. As such, we conclude by advancing an economic model to explore potential improvements to registries generally. The model shows that banning late registration and linking Institutional Review Board applications to registrations can significantly increase registry effectiveness.

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

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

This paper is a revision of "Research Registries: Myths, Facts, and Possible Improvements," which included an analysis of economics lab experiments that registered with the AER RCT Registry. The current paper's discussion of the AER RCT Registry instead focuses on economics field experiments. We thank Ariel Listo and Xinyi Hong for excellent research assistance, as well as our discussant Matthias Greiff for useful feedback. Discussions with Vittorio Bassi, Eszter Czibor, Stefano DellaVigna, Rachel Glennerster, 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 ASSA Virtual Winter Meetings, the Advances in Field Experiments Conference, and the UChicago Experimental Economics Seminar for their feedback and the Becker Friedman Institute for financial support.

1 Introduction

The last several decades have brought significant change to the empirical landscape in economics. New approaches to generating data in the lab and field have opened up several unique lines of research into the "whys" behind observed behaviors.[1] These experimental approaches have helped to clarify identification, control, statistical inference, and interpretability. However, critics in the broader social sciences have recently called for the experimental movement to proceed more cautiously. An active debate has emerged over claims that experiments face a "credibility crisis."[2] This charge follows from the fact that data are ultimately finite, so that researchers must choose which hypotheses to test, report, and trumpet in a system where publication incentives imply that not all results are equally likely to get published. Economists, along with researchers in other empirical disciplines, have recognized that these limitations could lead to a departure from socially optimal experiment conduct. A growing literature provides policy prescriptions aimed at improving the experimental approach.[3]

This paper conducts an empirical and theoretical examination of one of the most significant policy prescriptions—the establishment of research registries for randomized controlled trials (RCTs). These registries provide a venue for researchers to document their experiment setup, execution, and results in a venue that is searchable by external audiences. In principle, if used appropriately, research registries can tackle key issues in the credibility crisis.[4]

We choose to examine the extent to which research registries address two concrete issues that have received particular attention—and, to the best of our knowledge, form the primary motivation

[1] See Harrison and List (2004)

[2] See Jennions and Møller (2003), Ioannidis (2005), Nosek, Spies and Motyl (2012), Bettis (2012), Maniadis, Tufano and List (2014), and Dreber et al. (2015)

[3] See Glaeser (2008), Coffman, Niederle and Wilson (2017), Christensen and Miguel (2018), Al-Ubaydli, List and Suskind (2019), Dufwenberg and Martinsson (2019), McCloskey and Michaillat (2020), and Abadie (Forthcoming)). For related meta-analyses on the credibility crisis see Young (2018), Vivalt (2018), Andrews and Kasy (2019), Meager (2019), and Vivalt (Forthcoming).

[4] We acknowledge that the credibility crisis applies to empirical research broadly. However, discussions of the crisis and policy prescriptions (including research registries) tend to focus on RCTs. We believe that this is because RCTs are seen as low hanging fruit——each RCT is ostensibly designed to test a small set of interventions and has an explicit start and end date. One notable exception is the Open Science Framework (OSF) Registries Network. The OSF advocates for open collaboration in science research and their registries network permits the registration of observational studies. Other web services, such as AsPredicted, also facilitate recording any research hypothesis. However, unlike research registries, AsPredicted does not provide a way to search the recorded hypotheses.

for the establishment of registries in the first place:[5]

- *The file drawer problem*, namely that many studies are never made public (and so relegated to the proverbial "file drawer").

- *Scope for p-hacking*, namely that researchers often make adaptive decisions on the basis of initial analyses that are not accounted for or reported in the final analysis.[6]

A registry can address the file drawer problem for RCTs to the extent that researchers record all RCTs started and their outcomes. A registry can address p-hacking in RCTs to the extent that researchers use the registry to document their initial experimental design and analysis plan along with changes to these over time.[7]

We focus our examination on the American Economic Association's registry for randomized controlled trials (the AEA RCT Registry) and ClinicalTrials.gov. Though still relatively new, the AEA RCT Registry is the most commonly used registration database in economics (see Appendix C.1). The AEA RCT Registry lists 2,444 studies across 133 countries as of April 1, 2019. ClinicalTrials.gov is maintained by the National Institutes of Health and is the largest research registry overall. It contains 302,850 medical trial registrations from 208 countries as of April 1, 2019. An existing literature (reviewed in Appendix D) has assessed the mixed effectiveness of ClinicalTrials.gov. To the best of our knowledge, we are the first to provide a systematic assessment of the AEA RCT Registry.

We make two specific contributions to the literature on policy prescriptions for the credibility crisis. First, we empirically evaluate whether the AEA RCT Registry has been effective at solving the file drawer and p-hacking problems for economics field experiments both in absolute terms

[5]We focus on these two issues due to their concreteness and since they are, as far as we are aware, the most salient issues registries are designed to address. While other issues are certainly interesting (such as transparency and hypothesis selection broadly defined), we leave empirical examinations of these to future work. See Christensen and Miguel (2018) for a notable discussion of transparency in economics research.

[6]The extent to which results of empirical studies are manipulated in practice has been studied by Brodeur et al. (2016), Brodeur, Cook and Hayes (2020), Vivalt (2018), and Elliott, Kudrin and Wüthrich (2021).

[7]Nosek et al. (2018) note that despite registries' ability to reduce p-hacking in theory, there is limited empirical support for their effectiveness in practice. Kaplan and Irvin (2015) provide suggestive evidence that registration with ClinicalTrials.gov is positively correlated with an increase in the publication of null findings in a sample of large NHLBI supported RCTs.

and in comparison to ClinicalTrials.gov.[8] Second, we advance a model of registration that suggests alternative registry designs that could improve registry effectiveness broadly. Our theoretical analysis focuses on one concrete design issue, namely that both the AEA RCT Registry and ClinicalTrials.gov accommodate *late registration*. While typical motivations for promoting registration rely upon the assumption that it is done prior to experimentation,[9] the AEA RCT Registry and ClinicalTrials.gov permit the registration of completed trials.[10]

Unfortunately, our empirical analysis suggests that the AEA RCT Registry does not yet sufficiently address either the file drawer or p-hacking problems for economics field experiments. A theme that emerges from our analysis is that, despite the formal requirements discussed below, 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 design beyond the mandatory registration requirements. Insofar as the formal requirements are fairly weak (which was 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 field economics experiments is unobserved, we are able to perform a census of papers conducting field experiments published in select economics journals and working paper series. Roughly 50% of the field experiments published in top journals (both general interest and field) in economics—between 2017 and the end of 2019 Q2—are registered. We note that this is a selected sample which likely *overstates* the overall fraction of started field experiments that have registered.[11] Perhaps more telling is that only 20% of working papers on field experiments in the New Economics Papers Report on Experimental Economics (NEP-EXP) are registered. On the same theme, we also show that the AEA RCT

[8] A previous version of this paper considered both field and lab experiments. However, as we explain in Appendix C, the AEA RCT Registry is primarily targeted at the registration of field experiments, making it more difficult to interpret any success or failure in the registration of lab experiments. We now focus on field experiments in order to prevent this distinction from interfering with the interpretation of our conclusions.

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

[10] The AEA RCT Registry chose to allow late registration primarily to facilitate the registration of RCTs that started prior to the registry's establishment. However, our understanding is that there is no plan to revisit this design choice now. ClinicalTrials.gov generally allows late registration although several categories of medical experiments are required to preregister by law. As far as we know, no existing laws require either the registration or preregistration of economics experiments.

[11] For instance, field experiments may register late in order to be considered by AEA Journals, and, in fact, we find this is the case for most papers published in these journals as we discuss in Section 2.2.

Registry is not currently effective at capturing RCT outcomes. Only one-third of registered trials follow-up with any outcome data as of April 1, 2019.[12]

Regarding the p-hacking problem, we again cannot directly assess the extent to which the AEA RCT Registry succeeds at limiting the scope for p-hacking since we do not observe the research process itself. However, we can examine whether the preregistrations include a supplemental pre-analysis plan and whether the primary outcomes reported in the preregistrations themselves are specific enough to tie researchers to one experimental design. We find that only 10% of the preregistrations include a supplemental pre-analysis plan and that the primary outcomes reported in the preregistrations themselves are generally *not* binding.

Our assessment that the primary outcomes are generally not binding is made from a detailed study of the outcomes reported by 300 randomly chosen preregistrations.[13] We find that 241 of the 300 preregistrations are for field experiments. Focusing on this subset, the average preregistration reports just over 3 primary outcomes, but even the most detailed of these outcomes fails to specify either a specific variable construction or measurement timeframe. As we discuss in more detail 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 match working papers to 100 of the field experiment preregistrations.[14] Encouragingly, we find that 90% of the primary outcomes in the average working paper match their preregistered construction. That said, nearly 20% of working papers add at least one primary outcome (i.e., highlight an unregistered variable in their abstract, introduction, or conclusion) with the average working paper adding 0.46 primary outcomes.

When interpreting these results, it is important to note that a preregistration is only a statement on the initial plans for the experiment. Plans can and do change both before and during execution. Correspondingly, researchers can update the registration to reflect how and why the initial plans

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

[13]We had one set of RAs assess the specificity of the primary outcomes reported by each preregistration. We then had a second set of RAs label each preregistration as either a lab or field experiment according to the definitions in Harrison and List (2004). The AEA RCT Registry does not collect data on whether a given RCT is a field experiment or lab experiment.

[14]There are no associated published papers.

changed.[15] Put another way, a preregistration is not a prohibition against altering the experiment in response to realized hurdles or in order to explore unanticipated paths. For this reason, the observed low preregistration rate and the lack of specificity of preregistered primary outcomes is very surprising.

Assessments of ClinicalTrials.gov provide a useful benchmark for our results on the AEA RCT Registry. We extend the existing literature on ClinicalTrials.gov by examining the restrictiveness and fidelity of primary outcomes reported by 300 randomly chosen preregistrations from the first five years of ClinicalTrials.gov.[16] We find that the ClinicalTrials.gov preregistrations are only slightly more restrictive than the AEA RCT Registry preregistrations. We also find that papers associated with the ClinicalTrials.gov preregistrations and the AEA RCT Registry preregistrations have similar fidelity to the registered primary outcomes. This result combined with the literature review in Appendix D suggests that there is little reason to be optimistic that existing research registries will significantly dent the credibility crisis in economics.

In an effort to improve research registry designs, we next construct a simple model of registration.[17] The model features a researcher endowed with an experiment on an underlying hypothesis, whose payoffs improve as an "outsider" becomes more optimistic that the researcher's hypothesis is true. The researcher first chooses whether to preregister and conduct the experiment. The researcher then chooses whether to register late. Finally, the researcher receives a payoff based on the outsider's updated belief about the underlying hypothesis after seeing the registration decision and the experiment outcome. Preregistration allows researchers to signal confidence in their hypotheses, for instance from strong intuition based on prior work or domain expertise. But late registration is tempting due to option value—there is a chance that registration is not worth it ex-post since there are costs associated with registration.[18]

The value of the model is threefold. First, we use the model to scrutinize how registration

[15] Alternatively, researchers can explain any changes in the research paper itself.

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

[17] 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 related discussions of researcher experiment choice see Di Tillio, Ottaviani and Sorenson (2019), Libgober (2020), Al-Ubaydli, List and Suskind (2019), Tetenov (2016), and Anderson and Magruder (2017).

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

influences the decision to experiment, particularly for researchers on the margin.[19] Second, we articulate how addressing p-hacking and the file drawer problem can further scientific progress via an explicit social welfare function. Third, we provide comparative statics that examine how counterfactual policies influence registration decisions.

The most notable policy we examine with the model is a late registration ban. Perhaps surprisingly, our analysis identifies plausible conditions under which a ban on late registration could increase the *total* number of registrations (and improve welfare as well). One might find it natural to conjecture that allowing late registration would only increase the number of registered experiments by providing researchers more opportunities to register. However, we highlight that the option value associated with late registration gives researchers on the margin (between registering and not) an added incentive to delay their registration decision. Therefore, banning late registration always increases registration rates for the marginal experiment. And since not all researchers who delay their registration decision will find it worthwhile to ultimately register, this effect can be sufficiently strong to overturn the conjecture.

We use a calibration exercise to argue that this insight is empirically relevant for the AEA RCT Registry. Generally, the comparison between registration rates with and without a late registration ban is ambiguous due to the competing effects identified in the previous paragraph. Our calibration exercise gives some suggestions about which way this may resolve in practice. Under parameter values that match current registration rates, we show that banning late registration would strictly increase total registrations for the AEA RCT Registry. As such, we recommend the AEA RCT Registry consider banning late registration in order to more effectively address both the file drawer problem and p-hacking.[20]

So where do we go from here? Our first recommendation is to explore prohibiting late registration, while simultaneously providing incentives for researchers to preregister their work (such as mandating preregistration as a condition for publication). To the extent that the ultimate goal of the registry is to maximize the total number of preregistrations, this dual approach can move

[19]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).

[20]The file drawer problem is mitigated due to the increase in total registrations and p-hacking is mitigated due to all new registrations being preregistrations.

us in that direction. That said, our analysis also highlights that (1) registration costs can reduce experimentation and (2) current registrations lack the specificity needed to significantly address p-hacking. Our second recommendation tackles these latter issues. Since RCTs generally require Institutional Review Board (IRB) approval, we propose having researchers submit their IRB materials directly to the registry. While IRB materials are admittedly heterogeneous across schools, in our experience they contain enough uniformity and detail to provide a check on p-hacking. Further, this policy avoids large additional costs since researchers can simply upload the IRB forms that they have already completed.

The remainder of our paper is organized as follows. Section 2 presents our empirical assessment of whether the registry is currently solving the file drawer and p-hacking problems for economics field experiments. Section 3 compares the AEA RCT Registry to ClinicalTrials.gov and discusses other registration venues as well. Section 4 presents our model of a researcher's registration decision. Section 5 concludes. We highlight that Appendix C contains background information on the AEA RCT Registry relevant for our analysis and that Appendix D surveys past work on ClinicalTrials.gov. Apart from these, 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.[21] 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."

[21] 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.

Together, these two effects can undermine public trust in empirical research and cause inefficient resource allocations.

We start by empirically examining the extent to which the AEA RCT Registry is currently capturing the universe of economics RCTs (i.e., addressing the file drawer problem) and the extent to which it succeeds in pre-committing researchers to assessing a specific set of outcome variables (i.e., addressing p-hacking). We consider the AEA RCT Registry from its launch on May 15, 2013 up through April 5, 2019. The AEA RCT Registry is primarily targeted at the registration of field experiments, making it more difficult to interpret any success or failure in the registration of lab experiments. As such, we focus our analysis on field experiments when possible in order to prevent this distinction from interfering with the interpretation of our conclusions.[22]

2.1 File Drawer Problem

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

1. Every field experiment that is started is added to the registry
2. Experiment results are added to the registry at the conclusion of the experiment

Because the universe of started field experiments is unknown, we cannot determine the fraction of RCTs that register with accuracy.[23] As such, we start by examining the registration rate for field experiments published in leading economics journals and working paper series. Table I presents the registration rates for RCTs appearing in the following outlets over 2017, 2018, and the first two quarters of 2019:

- American Economic Review (AER)
- American Economic Journal: Microeconomics (AEJ-Mic)
- 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)

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

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

- Experimental Economics (EE)
- New Economics Papers Report on Experimental Economics (NEP-EXP)

We find that field experiment registration rates across journals are heterogeneous and overall quite low. The AER, QJE, and AEJ-AE have the highest field experiment registration rates—75%, 69%, and 62% respectively—over this period. The remaining journals have registration rates below 60%—with only 14% of the field experiments published in the JPE registering. Of note, the AEA journals require that field experiments be registered as a condition for publication as of January 2018 (this requirement does not apply to lab experiments).[24] Table II investigates the effectiveness of this requirement by reporting the registration rates for field experiments by journal and year. Despite the requirement, we find a registration rate of only 63% for field experiments published in the AEA journals considered here over 2018 and the first two quarters of 2019. This could be due to ambiguity in the designation of a given RCT as a field or lab experiment.

The second step in addressing the file drawer problem is reporting outcomes. The registry data speaks immediately to whether outcomes are added to the registry at the conclusion of the RCT. We note that the AEA RCT Registry does not collect data on whether a given RCT is a field experiment or lab experiment. As such, we examine outcome reporting across all registered trials. We find that 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.

2.1.1 Late Registration

The AEA RCT Registry allows researchers to register RCTs even after completion. Allowing late registration might help solve the file drawer 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

[24] See Appendix C for additional background.

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 can come at the cost of diminishing preregistration.

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.[25] 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 exacerbate the file drawer 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.[26] Table III 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 the registry itself opened in May 2013 and that it is infeasible to subset this analysis to field experiments because the registry does not collect data on whether a given trial is a field or lab experiment. 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.

[25] 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).

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

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.

2.2 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. RCTs register before the intervention begins, i.e. preregister
2. Registrations sharply specify at least the primary outcome variables
3. Published or working papers on the RCTs match the registrations or explain any differences

We examine each of these issues in turn. When interpreting these results, it is important to remember that a preregistration is a statement of the initial plans for the experiment. A preregistration does not prohibit altering the experiment to navigate realized hurdles or explore unanticipated paths. Plans can and do change both before and during execution. Correspondingly, researchers can update the registration to reflect how and why the initial plans changed or explain any such changes in the research paper itself.

2.2.1 Preregistration

We start by examining the fraction of RCTs that preregister. As the registry does not collect data on whether a given trial is a field or lab experiment, we are again unable to subset this analysis to field experiments. 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.[27] Of these trials, only 47% registered before their intervention began. Another 30% registered before the intervention ended.[28] 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. These figures illustrate that while the fraction of RCTs that are preregistered has been weakly growing over time, the registry is still dominated by late registrations. We also conduct an ANOVA test to examine the

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

[28]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.

drivers of preregistration. Table IV assesses whether quarter, the first trial topic keyword listed, the first JEL code listed (two-digit classification), the first trial location listed, and primary investigator affiliation predict the preregistration status. We find that quarter is the strongest predictor followed by affiliation and keyword.

Finally, we verify that preregistration is still uncommon among published papers. Table V presents the preregistration rates for field experiments published in the journals considered above. We find that only one-third of these papers preregistered their trial. 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. And Columns 7-9 report the preregistration rates for this subset. Notably, of the 12 registered RCTs published by AEA journals that were started after 2013, only 5 preregistered. This suggests the registry is still largely capturing studies that are already going through the publication process, which limits the registry's impact on experiment design and p-hacking.

2.2.2 Restrictiveness

We next examine the extent to which the preregistrations specify the experimental design. Note that researchers are required to describe their:

- Primary outcomes[29]
- Randomization method
- Planned number of observations and treatment arms

Researchers are not required to submit a supplemental pre-analysis plan (and correspondingly, only 11% of the 1,792 trials post a pre-analysis plan).

As a first pass, we focus on whether researchers specify their primary outcomes in enough detail to tie their hands to particular variable constructions. We randomly sampled 300 preregistrations and had one set of RAs assess the specificity of the primary outcomes reported by each preregistration. We instructed these RAs to independently count the number of primary outcomes listed and score the outcome descriptions on a scale of 0 (not specific) to 5 (very specific). We gave the following example scale: "Mark "health" as a 0, "nutritional intake" as a 1, "number of

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

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 F provides the full instructions.[30] We then had a second set of RAs label each preregistration as either a lab or field experiment according to the definitions in Harrison and List (2004).[31] We find that 241 of the 300 preregistrations are for field experiments.

We find that the field experiment preregistrations leave significant latitude.[32] Table VI reports the assessed restrictiveness. The average preregistration specified just over 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 timeframe. 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 timeframe. 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 timeframe; quoted price and price paid are missing both a specification of the products to be considered (likely the primary outcomes of interest will actually be price indexes) and a timeframe; and number of "missed" clients is missing both a specification of

[30]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%. Our analysis is based on the average of the two assessments.

[31]The AEA RCT Registry does not collect data on whether a given RCT is a field experiment or lab experiment.

[32]Qualitatively, the results are largely unchanged when including the lab experiment preregistrations.

how missed will be measured and a timeframe. 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 pre-registration to explore if any primary outcome or sample specification changed. The last two rows in Table VI report the results. We find that 4% of the 241 assessed field experiments changed one of their primary outcomes after the preregistration. Similarly, 4% of the assessed field experiments changed some aspect of their sample specification after the preregistration.

2.2.3 Fidelity

We finally 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. 227 of the 241 field experiment preregistrations listed an intervention end date and 182 ended before June 2019. However, only 7 of the preregistrations provided 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 100 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 VII reports the assessed fidelity of the working papers associated with the field experiment preregistrations. In the average working paper, 90% of the primary outcomes match their preregistered construction. This figure is encouraging, but may be somewhat misleading because the vast majority of preregistered primary outcomes are unspecific—to use Delecourt and Ng's example, there are many ways to construct a variable that reports the "price paid" for products sold by micro-businesses in India. More troubling, roughly 20% of the working papers report an additional primary outcomes (i.e. highlight an unregistered variable in their abstract, introduction, or conclusion—see Appendix F). The average working paper reports 0.5 additional primary out-

comes. Similarly, roughly 20% of the working papers fail to report at least one primary outcome with the average working paper under-reporting 0.4 primary outcomes.

3 Analysis of ClinicalTrials.gov

We now conduct a new survey of ClinicalTrials.gov which serves to more precisely benchmark our results on the restrictiveness and fidelity of AEA RCT Registry preregistrations. We emphasize that our analysis of ClinicalTrials.gov here builds on a large literature and provide a survey of this literature in Appendix D. That said, we are not aware of any previous comparisons to the AEA RCT Registry, which is the main contribution of this section. Of note, Appendix C.1 examines whether economists use other research registries in addition to or in place of the AEA RCT Registry. There we find that the AEA RCT Registry is indeed the dominant registry for economists (verifying the consensus view).

We proceed in the same manner as Section 2.2 and focus on the launch of ClinicalTrials.gov in order to provide a reasonable comparison. We find that preregistrations from the first five years of ClinicalTrials.gov are somewhat more restrictive than the AEA RCT Registry preregistrations. We also find that published and working papers associated with the ClinicalTrials.gov preregistrations and with the AEA RCT Registry preregistrations have similar fidelity to the registered primary outcomes. This result combined with the literature review in Appendix D suggests that if ClinicalTrials.gov gives a sign of where the AEA RCT Registry is headed, then there is little reason to be optimistic that the current approach will significantly dent the credibility crisis in economics.

More precisely, we randomly sampled 300 trials that preregistered with ClinicalTrials.gov between March 1, 2000 and July 1, 2005. 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 then employed a new set of 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.[33] Appendix G repeats this analysis for preregistrations with ClinicalTrials.gov after the implementation of the Final Rule for Clinical Trials Registration and Results Information Submission and reaches similar conclusions.

Table VIII reports the assessed restrictiveness of the 300 randomly selected ClinicalTrials.gov preregistrations. The average preregistration specified 2 primary outcomes—1 less than the average AEA RCT preregistration. The average minimumly restrictive outcome is classified as a 2.8, the average median restrictive outcome as 3, and the average maximumly restrictive outcome as 3.4—each roughly 1 unit more restrictive than the equivalent value for the AEA RCT preregistrations. Put another way, the median primary outcome from a ClinicalTrials.gov preregistration is roughly as specific as "number of fruits consumed at school per week." In contrast, the median primary outcome from an AEA RCT Registry preregistration is just "number of fruits consumed."[34]

We were able to associate published or working papers with 278 of the 300 ClinicalTrials.gov preregistrations. Table IX reports the assessed fidelity of the primary outcomes reported in these papers to those in the registration. In the average paper, 80% of the primary outcomes matched their registered construction—as compared to 90% for the AEA RCT Registry.[35] However, as with the AEA RCT Registry results, this figure may be misleading because the vast majority of registered primary outcomes are vague enough to match with multiple possible variable constructions. Perhaps more telling, the average paper reported 0.4 primary outcomes that were not registered and failed to report 0.4 registered primary outcomes. These values closely match those found for the AEA RCT Registry.

[33]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.

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

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

4 A Model of Registration

We now introduce a simple model that articulates a researcher's incentives to register and the implications of the registration timing decision. In the model, the researcher faces a dynamic decision of when to register their experiment; in equilibrium, early registration signals confidence in the researcher's hypothesis. The model provides two sets of results regarding registration patterns. 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 a registration (explained in detail below). Here, we show that increasing informativeness 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 posits a 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 (who we think of as a representative member of the research community) 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:[36]

- 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*.

[36] 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.

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, $s_1 \in [\underline{s}_1, \overline{s}_1]$ and $s_2 \in [\underline{s}_2, \overline{s}_2]$, 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 in first-order stochastic dominance (FOSD) 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 larger in the FOSD if $\theta = T$ than if $\theta = F$.

We interpret $\gamma \in \Gamma$ as parameterizing the informativeness of the second period signal following registration. This may reflect the ability to credibly disprove p-hacking (discussed below), but this need not be the only way in which registration could positively influence an experiment. For instance, 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 registration at $t = 2$ respectively). The outsider therefore updates p_0 from the observation of d and s_2. We denote the updated 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)). While such information may influence the researcher's beliefs, it is not made observable by registration.[37] In contrast, s_2 reflects the experimental findings, which can be conveyed verifiably. Notably, registration influences the second signal but not the first. 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 E). All signals are also assumed to have full support.

In our analysis below, we focus on the case where registration can increase experiment informativeness. Since preregistrations can be edited after they have been completed, it is less clear to us how delaying registration would yield *more* informative experiments versus preregistered ones. At the very least, our view is that the dominant case is one where preregistration increases informativeness. Ultimately, however, what matters for our results is that: (1) There is no asymmetric information outside of s_1, so that registration does not signal other relevant private information; and (2) outsiders *perceive* preregistered experiments as more informative. These assumptions avoids unnecessary complexity which obfuscates the main message; and the latter seems plausible based on our conversations with colleagues, particularly since delaying registration may at least give an *option* for additional p-hacking—whether or not this is actually done.

4.1.2 Researcher Payoffs

The researcher incurs a cost of $c_E \geq 0$ when conducting the experiment and also incurs a cost $c_R \geq 0$ whenever registering the experiment (whether registration is early or late). The researcher receives a payoff of 0 if the experiment is not conducted. Otherwise, the payoff 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 impose the following assumptions on the payoff functions, throughout our main analysis:

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

[37] Recall that registration only involves researchers basic properties of their upcoming experiment.

decision, $b_R(p) - b_N(p)$, is strictly increasing in p.

That payoffs are increasing in the beliefs reflects a preference for positive results; for instance, the payoff could reflect a benefit from publication, where the probability of publication is increasing in the outsider's belief that $\theta = T$ (see Brodeur et al. (2016) and Andrews and Kasy (2019) for empirical evidence suggestive of this preference). The increasing difference assumption says that the gain to registration is higher when the outsider's belief is more optimistic. Equivalently, this assumption says that additional optimism benefits the researcher more following registration, suggesting complementarities between beliefs and registration. We emphasize that researcher payoffs as a function of beliefs may arise from a variety of sources (e.g., reputational considerations). In Appendix E, 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.[38]

4.1.3 Equilibrium of Interest

In order to speak meaningfully about the tradeoffs of the marginal researcher (which drives the key intuition for our results), we will focus on the following class of equilibria:

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

Note that $s^*_{1,\emptyset} > s^*_{1,R}$ cannot hold in equilibrium, since not experimenting yields payoff 0, whereas registering an experiment and not conducting it yields negative payoff. Partitional equilibria are convenient to work with because the threshold signal is indifferent between actions on each side of the threshold—that is, a researcher with signal $s^*_{1,R}$ should be indifferent between preregistration

[38] In Appendix E, we show that, under Assumption 1, the second period registration decision does not convey information regarding the first period signal. This natural property can fail more generally.

and not, and likewise for other signals in this definition.[39] We are able to derive the comparative statics results described below by studying these indifference conditions.[40]

Together with an added assumption on the informational environment, our use of partitional equilibrium will also ensure our analysis is not sensitive to choices of off-path beliefs, a common difficulty in signalling models (particularly dynamic ones). To see how this concern might emerge, notice that Definition 1 allows $s^*_{1,R} = s^*_{1,\emptyset}$. This edge case, where all registration is early, does not appear to us to be the empirically relevant one. But to avoid it theoretically, we must ensure all registration decisions can be on-path. The following assumption delivers this end:[41]

Assumption 2. *Consider the informational environment where **every** conducted experiment must be preregistered (and is otherwise identical). Let $\tilde{s}_1 > \underline{s}$ denote the lowest signal in this alternative setting that obtains nonzero payoff when the outsider conjectures all $s_1 > \tilde{s}_1$ experimented. Then if the option to experiment without registering were introduced, a researcher with signal $\tilde{s}_1$ would have a profitable deviation to register if this induced the outsider to believe that $s_1 = \tilde{s}_1$.*

The assumption says that if all experimenting researchers were to register early, the experimenting researcher with the least promising signal would prefer to not register if doing so revealed the signal. Intuitively, we have in mind that registration reveals the interval s_1 belongs to; we are not interested in cases where equilibria are supported by excessively negative inferences following off-path actions. Indeed, an equilibrium where all registration is early might be supportable by the (off-path) belief that $s_1 = \underline{s}_1$ whenever $d \in \{\emptyset, 2\}$.

We briefly describe how we compute the threshold $s^*_{1,R}$ when Assumption 2 holds. The pro-

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

[40] 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.

[41] Note that this assumption can be checked directly from model primitives and is not an assumption on equilibrium.

cedure is simple, and illustrated in Figure VII.[42] The idea is the following: For each signal s_1, we conjecture that $s^*_{1,R} = s_1$, and then given this conjecture, compute the payoff from (a) preregistering and (b) experimenting without registration. Assumption 2 states that the payoff corresponding to (b) is higher than the payoff corresponding to (a) at the lowest signal. So, we compute $s^*_{1,R}$ to be the intersection of these lines. And indeed, the intersection point of the two lines in Figure VII is unique and therefore pins down a unique interior $s^*_{1,R}$ for this example.

4.2 Model Summary and Discussion

To summarize, our model features the following timing of moves (illustrated in Figure VI):

- A state $\theta \in \{T, F\}$ is determined, with $\mathbb{P}[\theta = T] = p_0$
- The researcher observes $s_1 \sim f(\cdot \mid \theta)$, decides whether to conduct the experiment, and if the experiment is conducted, decides whether or not to register.
- A public signal $s_2 \sim g_{\tilde{\gamma}}(\cdot \mid \theta)$ is drawn, where $\tilde{\gamma} \in \Gamma$ if the experiment is registered and $\tilde{\gamma} = 0$ if it is not.
- Depending on whether the experiment is unregistered ($d = \emptyset$), registered early ($d = 1$) or registered late ($d = 2$), the outsider updates beliefs to $\hat{p}_d(s_2)$.
- The researcher obtains an added benefit of $b_R(\hat{p}_d(s_2))$ if the experiment is registered and $b_N(\hat{p}_d(s_2))$ if it is not, in addition to the associated costs of registration (c_R) and experimentation (c_E).

Furthermore, we focus on equilibria where the researcher's decisions are *partitional*; that is, the equilibrium is characterized by a pair of thresholds in the first period and a single threshold in the second period, with registration in a given period occurring if the signal in that period belongs to the highest interval.

Before presenting our results, we provide some extended discussion which may clarify our contribution.

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

4.2.1 On p-hacking

Despite the ability for registration to influence the distribution over s_2 in many ways, our framework accommodates the possibility that a researcher who does not register early may be able to p-hack as a special case. To make this clear, we discuss how this would fit within our general framework explicitly. Under this interpretation, γ could parameterize the *scope* of an experiment:

- First, the researcher decides how many possible hypothesis tests may be considered (e.g., by considering different covariates), and
- Then, the researcher reports the most optimistic among all the considered tests.

In this case, a larger γ would by synonymous with an experiment with less scope. G_γ would then be the distribution of tests emerging when the scope is γ, and G_0 would be the distribution of tests when the researcher is free to p-hack.

Note that an implicit assumption is that the selection induced by p-hacking decreases the resulting informativeness of this experiment. While our sense is that this is generally accepted colloquially, theoretically it requires assumptions. For instance, suppose we view each test as an independent observation of $s_i = h(\theta) + \varepsilon_i$ for $\varepsilon_i \sim H$, $h : \{T, F\} \to \mathbb{R}$, and $i \in \{1, \ldots, K\}$, with the researcher only reporting the largest of the s_i. For this setting, Di Tillio, Ottaviani and Sorenson (2019) show that informativeness decreases in K only if $-\log H$ is logconvex; if this distribution is logconcave, then increasing K *increases* informativeness. Intuitively, assuming the number of samples is known, informativeness comparisons require comparing the distribution of ε_1 to the distribution of $\max\{\varepsilon_1, \ldots, \varepsilon_K\}$; in general, the order of these two variables in terms of belief dispersion is ambiguous (as the authors show formally). However, in a model of p-hacking where the independence of different signals is dropped—for instance, if a p-hacked experiment always gives the same, fixed and biased distribution over signals—this result would not hold.[43]

Another caveat is that, under this interpretation, taking g_0 to be common knowledge does implicitly assume that any p-hacking would be anticipated and correctly inferred by the outsider. If there is uncertainty about how much p-hacking occurred, then not observing this would lower the

[43] See also Section 4.2 of Libgober (2020) for a discussion of how the impact of p-hacking on the distribution over outcomes depends on properties of the experimentation environment and the precise form p-hacking takes.

informativeness of the experiment. Accommodating the potential for asymmetric information over p-hacking would take us too far afield with (in our view) little added relevant insight for the topic at hand. Lastly, by taking the hypothesis to be exogenous, we are implicitly assuming that all such p-hacking is "vertical," i.e., relating to the same hypothesis (e.g., resampling data). In principle, a researcher may p-hack by drawing multiple hypotheses. Though we could accommodate this somewhat by changing researcher payoffs below in other ways (corresponding to a different choice of experiment), it is less clear whether registration prevents this "horizontal" p-hacking, nor how this should influence the resulting inference.

4.2.2 Social Welfare

Our main results discuss how preregistration rates and choice to experiment varies with policies related to registration. To translate these into welfare statements, we briefly describe a particular social welfare function which organizes our thoughts about how policymakers may seek to design the registry. We have in mind that registration not only interacts with the researcher's decision, but has an added social benefit of increasing the probability that the findings are disseminated—i.e., solving the file drawer problem.

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

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

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

With this social welfare function, how might registration improve welfare? The short answer is

by addressing p-hacking (i.e., increasing informativeness) and the file drawer problem (i.e., wider dissemination of results independent of publication). By increasing the informativeness of the experiment, welfare may be increasing in $\tilde{\gamma}$, as society gains more from more informative experiments. On the other hand, it may be that the value of $I_{\tilde{\gamma}}(s_2)$ does not align with the probability of publication—for instance, if negative results are valuable in aggregate despite the fact that publication is more likely for positive results. In this case, $r(\hat{p}_d(s_2))$ might be larger when $\hat{p}_d(s_2)$ is smaller, correcting a loss when negative results are not be seen. Thus, effectively addressing the file drawer problem could create a welfare improvement. In our view, this welfare function captures the first order issues relevant to the evaluation of registry design; while we do not doubt other issues are important in maximizing the advancement of science, it is harder to see how a registry is well-suited to address them.[44]

4.2.3 On Partitional Equilibria

Notice that the informational content of pre-registration is determined endogenously in equilibrium, depending on the researcher's strategy.[45] Using the increasing differences property in Assumption 1, we can show (using standard arguments) that the second period registration decision must be partitional. However, 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.[46] As a result, indifference at the partitional thresholds is not enough to ensure that the thresholds determine an equilibrium. This is an issue for our numerical calibration, where we seek to identify particular equilibria by calculating the partition thresholds. To address this, we identify a sufficient condition in Appendix E which gives that

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

[45]This property contrasts our dynamic signalling model with costly disclosure; in these models, an agent pays a cost to reveal their type (or at least, some signal of their type). The key difference is that, in these models, the informational content of registration is exogenous. Here, the interpretation of pre-registration depends on the gain to registration, which in turn depends on what it means about s_1.

[46]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 non-monotonicities in signalling games. Note that the latter paper shows single-crossing by itself does not ensure monotonicity.

the local indifference condition ensures equilibrium holds globally. This condition states that as researchers grow more optimistic that $\theta = T$, their preference for preregistration over late registration increases as well. For the class of informational environments in our numerical calibration, we verify that it holds and therefore that the global conditions for equilibrium are in fact satisfied given indifference at the threshold signals. Therefore, these issues do not influence our conclusions, and thus we omit the technical details from the main text.

4.3 General Results

We now present our general results. We begin with our results on the impact of a late registration ban, with our main result being that banning late registration increases preregistration, and may *increase* the total number of experiments that register. The latter result holds despite the initial impact of providing fewer opportunities for registration. This suggests a late registration ban could make the registry more effective at combating the file drawer problem, by capturing even more studies, even in a worst-case scenario where unregistered studies are never published. We then turn to our results on how researcher decisions change with their experimentation problem. This illustrates that one way of increasing pre-registrations is to make them more informative in low-cost ways. As an aside, we also articulate a subtle trade-off between incentivizing preregistration and incentivizing experimentation, reflecting and formalizing similar concerns related to the social costs of pre-analysis plans by Duflo et al. (2020). One of our contributions is therefore to formalize this observation, which to the best of our knowledge has not been done elsewhere.

4.3.1 Result 1: Banning Late Registration Can Increase the Total Number of Registrations

We now consider the effect of banning late registration, i.e. prohibiting registration 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 because

researchers substitute between early and late registration. The following proposition identifies 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:[47]

Proposition 1 (Implications of a Late Registration Ban on Total Number of Registrations)**.** *Suppose $g_0 = g_\gamma$ and $c_E = 0$, and fix all other parameters besides c_R and the distribution over s_1.*[48] *Let $\hat{p}_{s_1}(s_2)$ be the posterior belief that $\theta = T$ given signals s_1 and s_2. Banning late registration increases overall registration whenever:*

- *The distribution over s_1 yields $\max_{s_2} \hat{p}_{\overline{s}_1}(s_2) - \hat{p}_{\underline{s}_1}(s_2) := I$ (that is, if the initial signal is sufficiently uninformative)*
- *Experiment costs are in some interval $[\underline{c}_R(I), \overline{c}_R(I)]$.*

The intuition is as follows. Recall that registration decisions are driven by two factors in our model: On the one hand, delaying registration preserves option value, in that it may not be worth it (ex-post) to have registered the experiment if the results are unpromising. On the other hand, registering early signals confidence in the hypothesis, which leads to an endogenous increase in the belief of the outsider. In the particular benchmark described in Proposition 1, the second factor is small, and so option value is significant and the incentive to delay registration is powerful. Of course, late registration still occurs in equilibrium—but only when the second signal is sufficiently positive. On the other hand, if registration costs are not too large, researchers would still find it beneficial to register if only able to do so at time 1. In this case, a ban on late registration induces (most, if not all) researchers to register early instead of late, by eliminating the possibility of option value.

The last part of the intuition, then, is to note that for a given signal s_1, a study that is not *pre*registered is only registered at all if $s_2 > s^*_{2,R}$. By contrast, registering early ensures the study is registered with probability 1. So, if a late registration ban leads to enough researchers substituting from late to early registration, then the overall registration probability will increase as well.

[47] This proposition does not require the equilibrium to be partitional, in contrast to those that follow.

[48] This result remains true if $c_E > 0$, provided it is sufficiently small.

We briefly mention that, under the social welfare function outlined in Section 4.2.2, a late registration ban also increases welfare in the environments described by Proposition 1. This follows from the combined observations that registration has a negligible impact on $\hat{p}_d(s_2)$, and that when $c_E = 0$ there is no selection on which experiments are conducted. Therefore, the impact on welfare is dominated by the increase in the probability of observing s_2.

4.3.2 Result 2: Banning Late Registration Increases Preregistration But Can Discourage Experimentation

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 equilibrium described in this proposition is one which is partitional and involves beliefs consistent with nonzero rates of both early and late registration (which is indeed what we observe). The proof and intuition are straightforward and come from considering the incentives of the marginal researcher indifferent between actions (i.e. experimentation and registration). Again, 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 will now 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. While the proposition does not rule out an increase in $s^*_{1,R}$ when there are multiple equilibria—for instance, it need not be that *all* $s^*_{1,R}$ thresholds with

a late registration ban are lower than *all* $s^*_{1,R}$ thresholds without a late registration ban—this intuition suggests the *most plausible* change following a late registration ban would be a decrease in $s^*_{1,R}$.[49] The effect of diminishing the value of experimentation without registration highlights a potential trade-off between inducing preregistration and inducing experimentation. This could potentially cut against the welfare gains of increasing registration, highlighting the importance of understanding whether the registry is at least effectively combatting the file drawer problem.

4.3.3 Results 3: Making Preregistrations More Informative Can Encourage Preregistration

Our last finding articulates conditions under which increasing the informativeness of registrations causes an increase in preregistrations (which, in isolation, would increase social welfare). 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.*[50] *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).*

The particular equilibrium considered coincides with the one in the previous proposition. Note that convexity is necessary in order to ensure that researchers gain from having more informative experiments.[51]

Insofar as registries prefer *pre*-registration to delayed registration (since, as discussed previously, the main benefits to registration implicitly assume it is done before the experiment is under-

[49] To be more precise about why this is most plausible, note that if the outsider maintained a conjecture that the equilibrium $s^*_{1,R}$ were unchanged following a ban, then researchers with s_1 just below $s^*_{1,R}$ would have a strictly profitable deviation to preregister. Understanding that it is the *lower* s_1 signals with the strict incentive to preregister suggests the outsider should conjecture a lower threshold.

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

[51] 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 $\theta = T$.

taken), 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.

4.4 Numerical Calibration

We conclude by using a numerical calibration of the above model to explore the impact of banning late registration for the AEA RCT Registry as suggested by Propositions 1 and 2.[52] These calculations allow us to speak to the relative costs and benefits of a late registration ban. In our view, it is a priori unclear which specifications of the parameters are most compelling. We therefore seek to be permissive in the specifications we consider, while focusing on an information acquisition technology which allows us to tractably vary signal informativeness. Specifically, 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 $s_t \sim f(s_t \mid F) \propto (1 - s_t), s_t \in [\underline{s}_t, 1 - \underline{s}_t]$

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 next assume that the payoff functions are linear—taking $b_R(\hat{p}) = \hat{p}$ and $b_N(\hat{p}) = 0.8\hat{p}$. This choice reflects that registration is required for publication

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

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. We take as given that all researchers experiment, and so set $c_E = 0$.[53] Then 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 X presents the results. Columns 1 through 3 report the input $\underline{s}_1$, p_0, and c_R. Column 4 gives the percent of RCTs that preregister in equilibrium. Column 5 confirms that this value matches the percent of RCTs that register late. Column 6 displays the total registration rate. Note that the total registration rate is increasing in $\underline{s}_1$. That is, the registration rate is decreasing in the informativeness of the first period signal.

Of interest, Table X Column 7 reports the registration 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 $\underline{s}_1 = 0.33$ and $\underline{s}_1 = 0.35$, banning late registration causes a small decline in overall registration. Whereas when the first period signal is $\underline{s}_1 = 0.38$ and $\underline{s}_1 = 0.4$, banning late registration causes a large increase in overall registration.

These results demonstrate the empirical relevance of Proposition 1 for the AEA RCT Registry. 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. That said, insofar as early registrations may be especially valuable, our the-

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

oretical analysis shows that a late ban is unambiguously beneficial if the relative preference for preregistrations is sufficiently large.

5 Conclusion

This paper provides a relatively sobering assessment of research registries—suggesting that so far they have not been transformative in tackling the major issues at hand. In the case of the AEA RCT Registry, most field experiments do not register and many registrations are done for experiments that are already at the journal submission phase. Perhaps most disconcerting is that even when preregistrations are completed, they often do not provide enough information to attenuate p-hacking concerns. These findings are particularly surprising because preregistrations are statements of intentions. They are not prohibitions against altering experiments to navigate realized hurdles or explore unanticipated paths.

By introducing an economic lens that clarifies the costs and benefits inherent in this knowledge creation market, we are able to provide two policy recommendations. First, we recommend prohibiting late registration and simultaneously providing incentives for preregistration. Second, we propose having researchers submit their IRB materials directly to registries. The first recommendation takes a dual approach towards maximizing total preregistrations, and the second recommendation addresses two other issues that our analysis highlights, namely that: (1) current registrations lack the specificity needed to limit p-hacking and (2) registration costs can reduce experimentation. In our experience, IRB materials contain enough detail to provide a check on p-hacking and, as these materials are already prepared, are nearly costless to submit.

Importantly, we find it valuable to acknowledge that much of the behavior regarding registration is undoubtedly guided by norms. In our economic 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, then our analysis suggests that this feature alone could induce a higher bar for undertaking an experiment in the first place and a lower bar for registration. We suspect that this trade-off is something policymakers are cognizant of, but which our analysis formalizes.

Finally, while data generation via field experiments represents one of the strongest growth areas in the social sciences, greater confidence in the received results is possible. Where will the registry approach to publication bias lead 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, as the credibility of our science is critically linked to key features of the registry.

References

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

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

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

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.

Brodeur, Abel, Nikolai Cook, and Anthony Hayes. 2020. "Methods Matter: P-Hacking and Publication Bias in Causal Analysis in Economics." *American Economic Review*, Forthcoming.

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

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

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.

Dufwenberg, Martin, and Peter Martinsson. 2019. "Sealed Envelope Submissions Foster Research Integrity." *Revue économique*, 70(6): 919–926.

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

Elliott, Graham, Nikolay Kudrin, and Kaspar Wüthrich. 2021. "Detecting p-hacking." *Econometrica*, Forthcoming.

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.

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.

Kaplan, Robert M., and Veronica Irvin. 2015. "Likelihood of Null Effects of Large NHLBI Clinical Trials has Increased Over Time." *PloS One*, 10(8): 1–12.

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.

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, Charles R Ebersole, Alexander C DeHaven, and David T Mellor. 2018. "The preregistration revolution." *Proceedings of the National Academy of Sciences*, 115(11): 2600–2606.

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.

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 papers with an RCT published over our timeframe by journal and experiment type

	Count		Fraction Registered	
	Field	Lab	Field	Lab
Journal				
AEJ-Mic	0	7	NaN	0.00
AEJ-EP	5	1	0.60	0.00
AEJ-AE	29	1	0.62	0.00
AER	16	3	0.75	0.33
JDE	39	2	0.44	0.00
EE	5	70	0.00	0.00
JPE	7	2	0.14	0.00
NEP-EXP	44	62	0.20	0.00
QJE	16	1	0.69	0.00
ReStud	10	9	0.50	0.11

Table II: Number of published papers with a field experiment and fraction that registered with the AEA RCT Registry

	Number Published			Fraction Registered		
	2017	2018	2019*	2017	2018	2019*
Journal						
AEJ-EP	1	2	2	0.00	0.50	1.00
AEJ-AE	9	13	7	0.67	0.62	0.57
AER	5	6	5	1.00	0.50	0.80
JDE	11	19	9	0.36	0.42	0.56
EE	2	3	0	0.00	0.00	NaN
JPE	3	2	2	0.00	0.50	0.00
NEP-EXP	14	11	19	0.07	0.27	0.26
QJE	6	7	3	0.50	0.71	1.00
ReStud	2	4	4	0.50	0.50	0.50

* Data for the first two quarters of 2019

Table III: 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 IV: ANOVA test of predictors of preregistration status

	Partial Sum of Squares	DF	F Stat	PR(>F)
Intercept	0.01	1	0.06	0.81
Quarter	68.03	99	4.42	0.00
Keyword	3.44	13	1.70	0.05
JEL Code	15.37	83	1.19	0.12
Trial Country	8.28	93	0.57	1.00
Affiliation	94.27	444	1.37	0.00
Residual	267.54	1722		

Notes: ANOVA test of whether quarter, the first trial topic keyword listed, the first JEL code listed (two-digit classification), the first trial location listed, and primary investigator affiliation predict the preregistration status of trials in the AEA RCT Registry. Quarter, keyword, and affiliation prove strongly predictive of preregistration.

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		
	2017	2018	2019*	2017	2018	2019*	2017	2018	2019*
Journal									
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	2	2	0	1	1	NaN	1.00	0.00

* Data for the first two quarters of 2019

Table VI: Assessment of the extent to which 241 randomly chosen AEA RCT Registry field experiment preregistrations precisely specify their primary outcomes

	Mean	Std	Min	10%	25%	50%	75%	90%	Max
Number of Outcomes	3.26	2.43	0.0	1.0	2.0	3.00	4.0	6.0	20.5
Maximumly Restrictive Outcome	2.48	1.03	0.0	1.0	1.5	2.50	3.0	4.0	4.5
Minimumly Restrictive Outcome	1.94	1.09	0.0	0.5	1.0	2.00	2.5	3.5	4.5
Median Restrictive Outcome	2.22	0.99	0.0	1.0	1.5	2.25	3.0	3.5	4.5
Outcome Changed (Yes/No)	0.04	0.16	0.0	0.0	0.0	0.00	0.0	0.0	1.0
Sample Changed (Yes/No)	0.04	0.17	0.0	0.0	0.0	0.00	0.0	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 F. Percentiles are computed using linear interpolation.

Table VII: 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	0.91	1.0	1.0	1.00	1.0
Number of Additional Outcomes	0.46	1.05	0.0	0.00	0.00	0.0	0.5	1.65	7.0
Number of Missing Outcomes	0.43	0.90	0.0	0.00	0.00	0.0	0.5	1.10	4.0

Notes: Working papers were found for 100 of the 241 field experiment preregistrations. Percentiles are computed using linear interpolation.

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

	Mean	Std	Min	10%	25%	50%	75%	90%	Max
Number of Outcomes	1.95	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. Percentiles are computed using linear interpolation.

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

	Mean	Std	Min	10%	25%	50%	75%	90%	Max
Fraction of Matching Outcomes	0.80	0.41	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. Percentiles are computed using linear interpolation.

Table X: Equilibrium registration rates for various model specifications

$\underline{s}_1$	p_0	c_R	% Preregister	% Register Late	% Register	% Preregister (Late Ban)
0.33	0.124	0.100	3.87	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

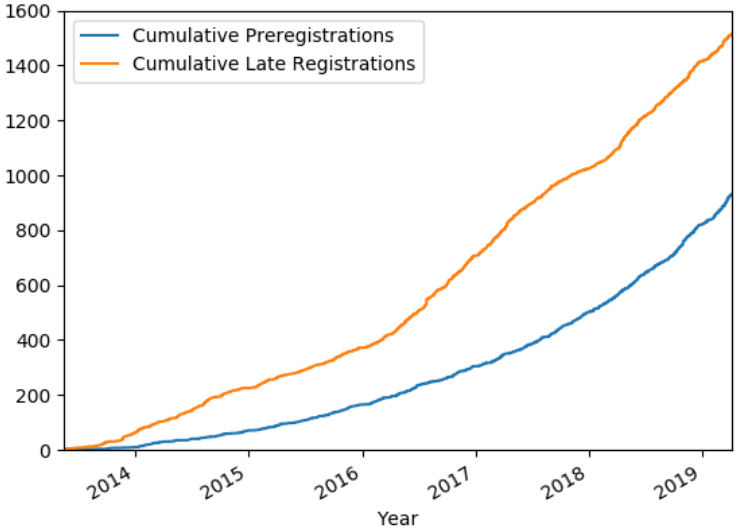


Figure III: Number of AEA RCT preregistrations and late registrations by quarter

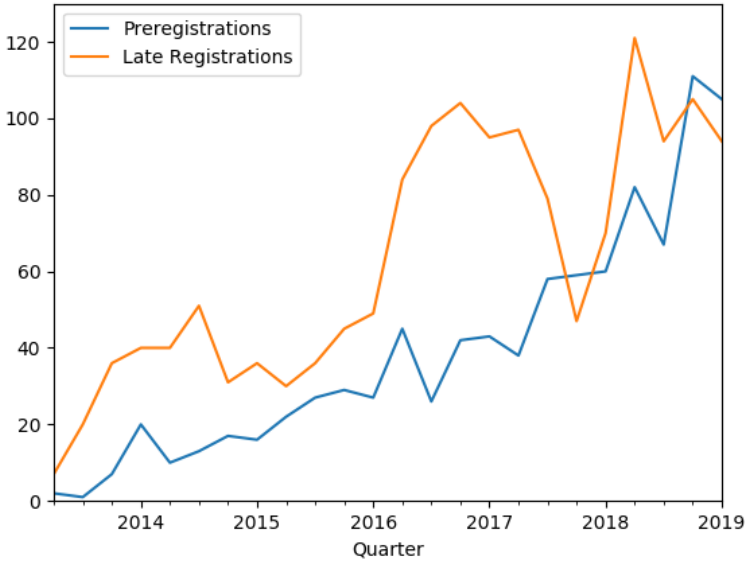


Figure IV: Days between intervention start and AEA RCT registration for RCTs started after January 1, 2014. Positive values indicate that the intervention began after the registration.

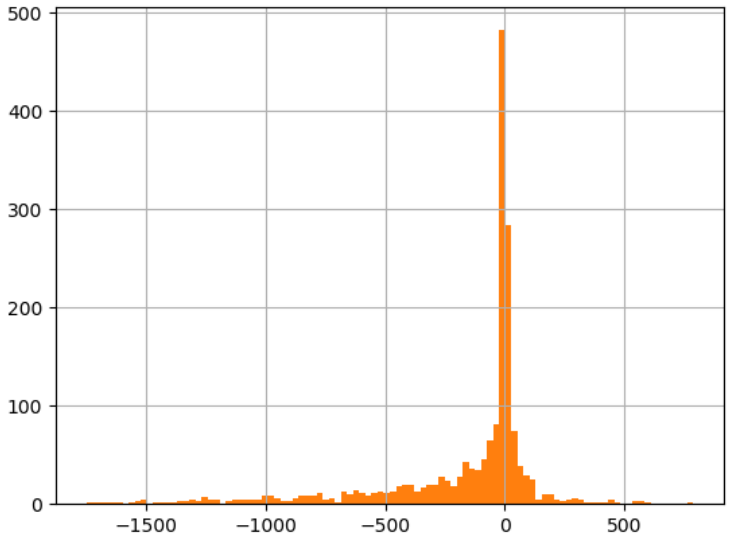


Figure V: Registration in the Registry for International Development Impact Evaluations (RIDIE) and in the Evidence in Governance and Politics (EGAP) registry over time. Solid lines display the total number of registrations while dashed lines present economics registrations

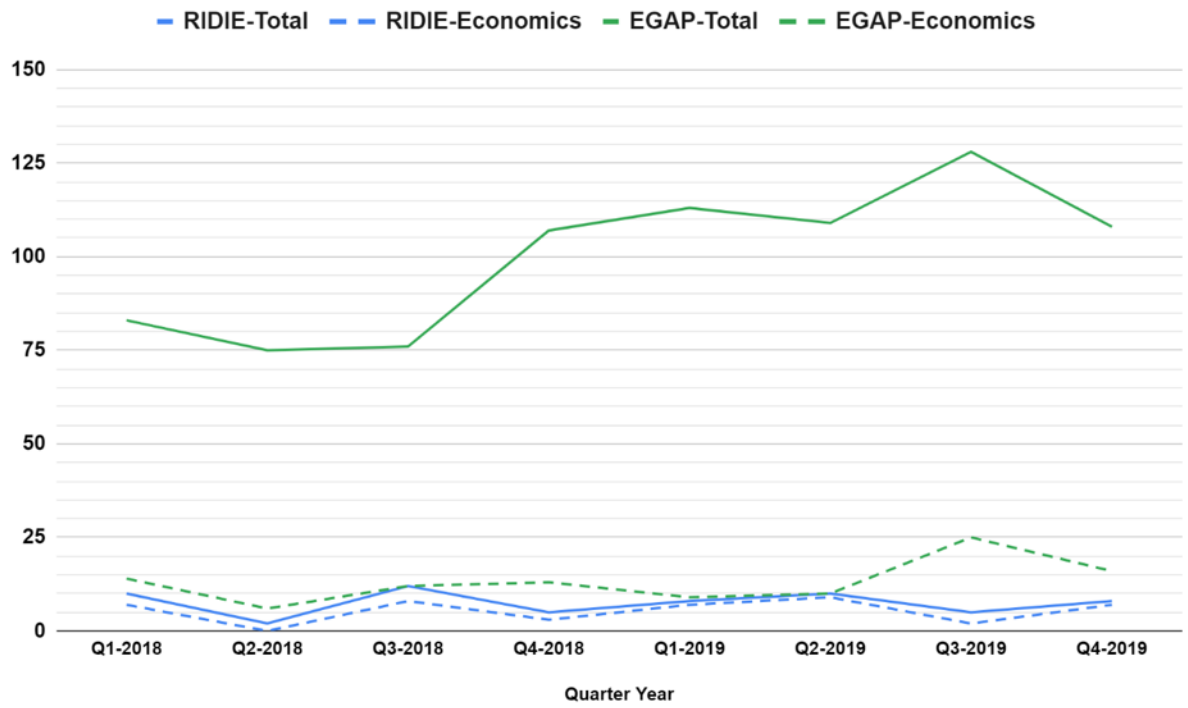


Figure VI: Timing of moves in the model

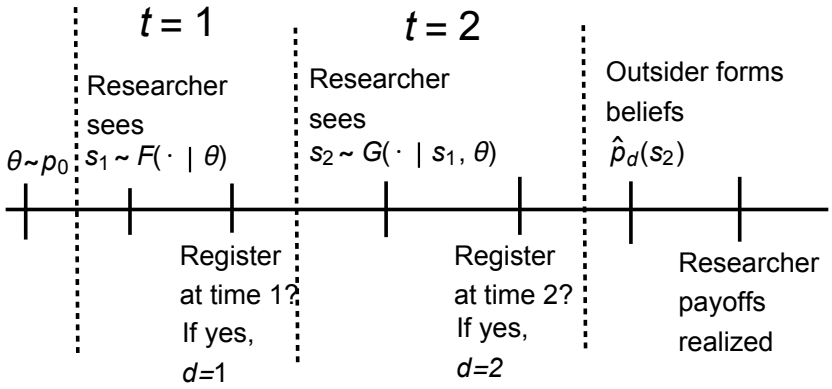
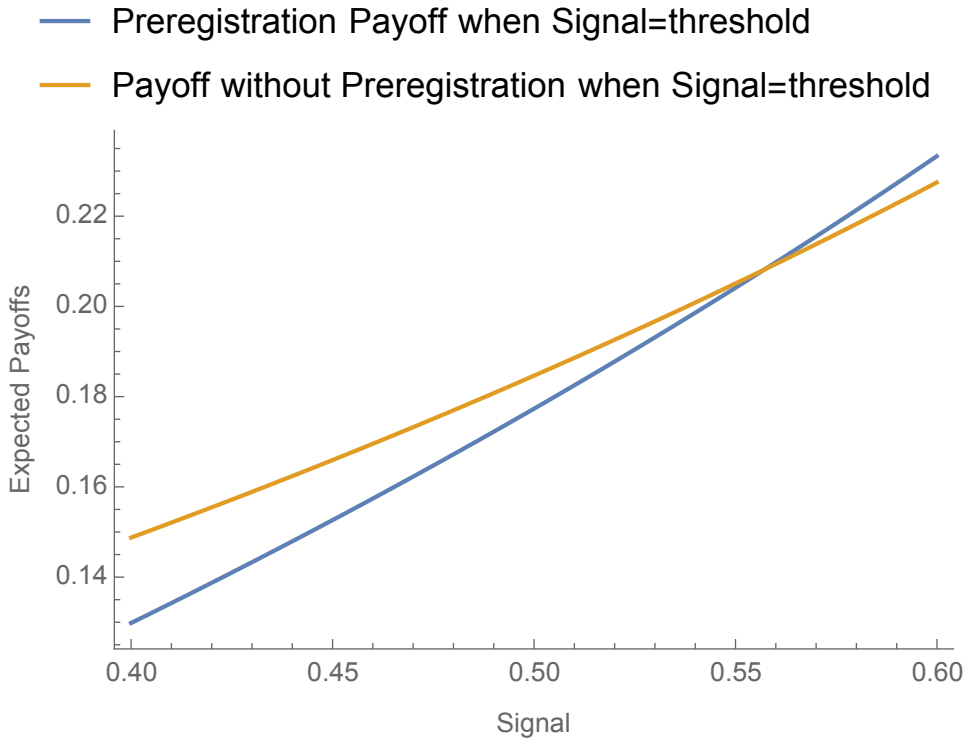


Figure VII: Researcher payoff upon receiving signal $s_1 = s^*_{1,R}$ assuming the equilibrium registration threshold is conjectured by the outsider to be $s^*_{1,R}$



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 X, with $\underline{s}_1 = .4$, $c_R = .1$, and $p_0 = .25$.

C Background on the AEA RCT Registry

The AEA launched the AEA RCT Registry in May 2013 to capture ongoing, completed, and terminated RCTs in economics and other social sciences (see About the Registry on the AEA registry webpage).[54] At the time, existing registries, such as ClinicalTrials.gov, focused on medical trials.[55] The AEA chose to implement a streamlined registration process to encourage participation—registration only requires answering a few questions and researchers are able to register at any time *even after the RCT is completed*. The required questions ask for a title, short abstract, start date, primary outcomes, treatment arms, and IRB approval number.[56] The AEA decided to focus on RCTs since experiments have a fairly distinct beginning and end. While one could imagine allowing non-RCT projects to register as well, it is difficult to implement a credible registration approach for research on pre-existing data.[57]

We use the term preregistration to denote a registration that occurs before the start of data collection. Note that a preregistration is distinct from a pre-analysis plan though in our experience they are often conflated. A preregistration (or 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).

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

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

[56] 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 drawer 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.

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

Crucially, neither a preregistration nor a pre-analysis plan prevent the researcher from altering the experiment to navigate realized hurdles or explore unanticipated paths. Plans can and do change both before and during execution. Correspondingly, registries, including the AEA RCT Registry, generally allow (and even encourage) researchers to update the registration or analysis plan to reflect and explain any changes to the initial experiment design.

Though not explicitly stated on the website, the AEA RCT Registry is primarily focused on capturing economics field experiments. While many lab experiments have chosen to register, some of the registry questions are less natural for certain lab experiments. In the same spirit, as of January 2018, the AEA journals require that field experiments, but not necessarily lab experiments, be registered as a condition for publication.[58] In any case, no economics journal requires that any experiment preregister—instead allowing registration to be done at the time of submission.[59] In contrast, most medical journals require preregistration of clinical trials.

The timing of an AEA RCT Registry registration can be determined from its listing in the registry database. 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). That said, it is not clear to us whether this distinction is salient or appreciated by consumers of research (or referees and editors). Unfortunately, we are not able to precisely study the extent to which the time of registration is distinguishable to someone who searches the registry. Our own conjecture is that the distinction is minor,[60] though researchers may emphasize that a study was preregistered in the corresponding written paper.

Finally, two other aspects of the AEA RCT Registry prove important in practice. First, the registry sends automatic reminders to encourage researchers to complete fields that become relevant during and after the RCT. For example, after the trial has concluded, researchers are asked to link to

[58] 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."

[59] 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.**"

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

any data, program files, or results that they have made public. Second, researchers are able to hide several fields in the registration from public view until later dates (specifically, the trial's location, intervention description, experimental design, names of any sponsors or partners, and supporting documents). On this last point, we stress that, as is, allowing these fields to be temporarily hidden does not fully eliminate the possibility that registering could invalidate an RCT's experimental design. One oversight is that the registry does not permit hiding the researchers' names, experiment title, and start date. This oversight is problematic for any RCT where identification relies on the intervention's occurrence being undisclosed to participants in the control and/or treatment arms.[61]

C.1 Other Research Registries

An open question is whether economists register field experiments in other research registries 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 more than 25 economics registrations in either registry. We also examine the Open Science Framework (OSF) Registries Network, which permits the registration of *both* RCTs and observational studies. As of June 13, 2021, just 12 registrations are tagged with the keywords "Economics" or "economics" and a general search for these words returns only 945 registrations many of which are, on inspection, observational studies conducted by psychologists or sociologists. This exercise provides some evidence that economists primarily use the AEA RCT Registry.

We separately 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 F 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. Economists running

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

lab experiments generally do not register (or preregister) anywhere.

D Survey of ClinicalTrials.gov Literature

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

First, ClinicalTrials.gov, by itself, does not capture a census of all relevant trials. In an early survey of industry-sponsored phase III drug trials, Manheimer and Anderson (2002) found that 25% of prostate cancer drug trials and 40% of colon cancer drug trials failed to register with ClinicalTrials.gov (or any other applicable registry). Dickersin and Rennie (2003) raised similar concerns for academic trials. In response to this issue, the International Committee of Medical Journal Editors (ICMJE) mandated that clinical trials register before the onset of patient enrollment as a condition of consideration for publication.[62] 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.[63]

Second, many trials that do register do not provide sufficient information. Zarin et al. (2011) examine the primary outcome measures from 100 randomly selected non–phase I trials that registered with ClinicalTrials.gov in August 2010 and find that 61% lacked either a specific metric and/or time frame. Zarin et al. (2017) repeat this analysis for 80 articles published in the New Eng-

[62] 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).

[63] 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.

land 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.[64] 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.[65] 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.[66] Adda, Decker and Ottaviani (2020) show an excess in the number of significant results in Phase III investigation relative to Phase II investigations for small industry sponsors; they argue this is consistent with the selective reporting of results.

Finally, when registered trials do report results these often differ from the published results. Hartung et al. (2014) explore these inconsistencies by taking a 10% random sample of Phase III and IV trials that both proceeded to publication and reported results on ClinicalTrials.gov before January 1, 2009. The authors find that 80% were inconsistent in the number of secondary outcomes considered, 35% inconsistently stated the number of individuals with a serious adverse event, 20% had inconsistencies in a primary outcome value, and 15% described a primary outcome inconsis-

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

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

[66]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.

tently. Becker et al. (2014) similarly find that nearly all trials published in high-impact journals that reported results on ClinicalTrials.gov had a least one significant discrepancy. Perhaps more ominously, Earley, Lau and Uhlig (2013) highlight differences between the number of deaths reported on ClinicalTrials.gov and in corresponding published papers.

E 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.

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

Proof of Lemma E.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 E.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:

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

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 E.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; in particular, since registration is deterministic as a function of s_2 alone, this registration decision does not carry information about s_1. Thus, $\sigma(d \mid s_1, s_2)$ is independent of s_1 on the support of the signals that delay registration. Upon inspecting (1), we can therefore factor out $\sigma(d \mid s_1, s_2)$ from both the numerator and the denominator, and further observe that the second period registration decision does not influence the distribution over s_1. Putting these observations together, we have that the registration decision does not influence beliefs if $d \neq 1$. □

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

Proof of Lemma E.3. Replicating the proof of Lemma E.1, Lemma E.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. □

The following remark highlights the features of our model which drive the result that the decision to register late (versus not at all) cannot depend on s_1. Roughly speaking, this requires the increasing differences assumption outlined in Assumption 1, as we describe in more detail below:

Remark 1. *The previous result show that the first period signal s_1 does not influence whether a researcher registers late if they did not pre-register. The reason this result holds is due to the increasing differences assumption in Assumption 1. This is a key technical novelty we highlight, and its necessity is quite subtle. This feature, to our knowledge, plays no role in prior work. Without Assumption 1, mixed strategy equilibria may emerge and cannot be ruled out immediately. The rest of this remark sketches some intuition for why this observation holds, and why Lemma E.2 may fail without it.*

Suppose the increasing differences 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, we sketch an informational environment and strategy which illustrates why 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 immediately 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 E.2 to future work.*

The following Lemma shows that $\hat{p}_d(s_2)$ is higher for all s_2 when $d = 1$ compared to when $d \in \{\emptyset, 2\}$.

Lemma E.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. □

E.2 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:

$$-c_R + \int_{-\infty}^{\infty} b_R(\hat{p}_{\overline{s}_1}(s_2))\mathbb{E}_\theta[g(s_2 \mid \theta) \mid s_1]ds_2. \tag{2}$$

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:

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

Under the assumption that $\hat{p}_{\bar{s}_1}(s_2) - \hat{p}_{\underline{s}_1}(s_2) < \delta$, for some δ, continuity of b_i gives us that we can find some ε such that (3) is greater than:

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

If this equation is larger than (2), 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 (2) yields:

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

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 informa-

tiveness 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. □

E.3 Partitional Equilibria

In this Section, we walk through some additional details on partitional equilibria and conditions which ensure the indifference condition determines an equilibrium. These results are primarily useful in ensuring that our numerical calibration yields valid equilibria, but also clarifies when partitional equilibria *must* emerge. 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 E.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:*

$$\int_{-\infty}^{\infty} b_R(\hat{p}_1(s_2))\mathbb{E}[g_\gamma(s_2 \mid \theta) \mid s_1] - b_R(\hat{p}_2(s_2))\mathbb{E}[g_0(s_2 \mid \theta) \mid s_1]ds_2 \tag{4}$$

is increasing in s_1.

Note that this condition does not depend on b_N, and in fact does not depend on the researcher's registration threshold at $t = 2$; it says that given a researcher *will* register, it is even better to register early rather than late when the initial signal is higher. This reflects our intuition that researchers with more favorable results are generally more eager to register, whenever they may be on the margin. The main usefulness of this condition is that it simplifies the verification that our candidate thresholds (particularly in the calibration exercise) determine an equilibrium, in that no s_1 would profitably from deviate from the prescribed registration decision. As described below, we use this to show that the registration thresholds determine a partitional equilibrium in the informational environment we focus on in the calibration exercise.

Proposition 5. *Suppose a registration strategy satisfies increasing gains to early registration. Then under Assumption 1, then the first period registration 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:

$$-c_R + \int_0^1 b_R(\hat{p}_1(s_2))\left(\tilde{p}f(s_2 \mid T) + (1-\tilde{p})f(s_2 \mid F)\right)ds_2. \tag{5}$$

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:

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

Rewriting this slightly, we wish to show that if:

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

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,[67] 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

[67] 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.

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 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, (5), as well as late registration, 5. 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 E.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, since it allows us to avoid computing the second period signal threshold to check that a conjectured strategy is in fact an equilibrium. In contrast, the increasing gains to early registration condition is straightforward to check across ranges of parameters. In general, we note that the second period threshold will depend on the first period threshold, and visa versa, since the former influences the payoffs of registering early and the latter influences the beliefs of the outsider.

E.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.

E.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.

F RA Instructions

F.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

F.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 ___

F.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 three registries (https://ridie.3ieimpact.org/, https://egap.org/registry/, and https://cos.io/prereg/) for the paper title and for the authors

G ClinicalTrials.gov after the Final Rule

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

FDAAA 801 and the Final Rule may have increased the restrictiveness of trials' preregistrations with ClinicalTrials.gov along with the fidelity of trials' results to their preregistrations. To examine this hypothesis, we repeat our assessment of ClinicalTrials.gov for the year following the effective date of the Final Rule.[68] Specifically, we randomly sample 100 ClinicalTrials.gov preregistrations that occurred between January 18, 2017 and December 31, 2017. We then instruct two RAs to assess each preregistration and associated working or published paper using the same rubric as in Section 3.

Table XI reports our assessment of the extent to which these preregistrations precisely specify their primary outcomes. We find that the preregistrations over January 18, 2017 to December 31, 2017 are slightly *less* specific than over the March 1, 2000 to July 1, 2005 period examined in Section 3. We also find that fewer trials changed their outcome or sample—likely due to the shorter follow-up period here.

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

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

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

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

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

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

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

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