UC Berkeley
CEGA Working Papers

Title
Beyond Publication Bias: Characterizing and Understanding Missing Results in Economics

Permalink
https://escholarship.org/uc/item/50k4m53n

Authors
Hoces de la Guardia, Fernando
Miguel, Edward
Pathan, Gufran
et al.

Publication Date
2026-11-13

DOI
10.1257/rct.13755-2.0

eScholarship.org Powered by the California Digital Library
University of California

Series Name: WPS
WPS Paper No.: 268
Issue Date: November 13, 2026

Beyond Publication Bias: Characterizing and Understanding Missing Results in Economics

Fernando Hoces de la Guardia, Edward Miguel, Gufran Pathan, Akash Shaji, Viviane Helena Silva da Rocha, Erik Ø. Sørensen, Bertil Tungodden



CEGA

Center for Effective Global Action

Working Paper Series

Center for Effective Global Action
University of California

eScholarship
University of California

This paper is posted at the eScholarship Repository, University of California. http://escholarship.org/uc/cega_wps Copyright © 2023 by the author(s).

The CEGA Working Paper Series showcases ongoing and completed research by faculty affiliates of the Center. CEGA Working Papers employ rigorous evaluation techniques to measure the impact of large-scale social and economic development programs, and are intended to encourage discussion and feedback from the global development community.

Recommended Citation:

Hoces de la Guardia, Fernando; Miguel, Edward; Pathan, Gufran; Shaji, Akash; Rocha, Viviane Helena Silva da; Sørensen, Erik Ø; Tungodden, Bertil. (2026): Beyond Publication Bias: Characterizing and Understanding Missing Results in Economics. Working Paper Series No. WPS-268. Center for Effective Global Action. University of California, Berkeley. https://doi.org/10.1257/rct.13755-2.0.

Beyond Publication Bias: Characterizing and Understanding Missing Results in Economics

Fernando Hoces de la Guardia[*†] Edward Miguel[‡] Gufran Pathan[†]
Akash Shaji[†] Viviane Helena Silva da Rocha[§] Erik Ø. Sørensen[§]
Bertil Tungodden[§]

November 13, 2026

Abstract

What scientists choose not to publish can be as important as what they do publish. We measure missing research results at the hypothesis level and use this approach to study under-reporting in economics. Across 5,057 pre-registered hypotheses from 317 studies registered on the American Economic Association RCT Registry between 2015 and 2017, only 42% have publicly available results eight to nine years after pre-registration. Among the available results, 64% are null, and we find little evidence that null results are disproportionately missing. In an RCT, direct engagement with authors increases the share of hypotheses with available results by 6.2 percentage points, while raising awareness alone has no significant effect and additional research assistance provides no further improvement. Direct engagement also increases explanations for unavailable results, reducing the share of hypotheses that remain unaccounted for by 16.4 percentage points. The findings document extensive under-reporting in economics that does not appear to be driven primarily by the suppression of null results, and suggest that reporting pre-registered results may need to become a standard part of the research process rather than something recovered years later.

[*]We thank Katie Baker, Peter Grinde-Hollevik, Wilson King, Alix Schoback, Sarah Stillman, and Johanna Weech for providing valuable feedback and assistance on this project, and to Michael Walker and seminar audiences for useful feedback. This project was administered by CEGA, U.C. Berkeley and FAIR, NHH, and financially supported by the Research Council of Norway through its Centres of Excellence Scheme, FAIR project No 262675. The study received IRB approval from the U.C. Berkeley Committee on the Protection of Human Subjects (CPHS) and the NHH IRB. The study is registered on the AEA Registry, AEARCTR-0013755, `https://doi.org/10.1257/rct.13755-2.0`.

[†]University of California, Berkeley

[‡]University of California, Berkeley and NBER

[§]NHH Norwegian School of Economics

1 Introduction

Many research findings never become public. Researchers, referees, and editors may all contribute to this selective reporting, particularly of null results (Chopra et al., 2024), which is known as the "file drawer problem" (Rosenthal, 1979). Such missing results can distort scientific conclusions and lead to an inefficient allocation of research effort, as researchers unknowingly pursue hypotheses that others have already examined.

Measuring the full extent of the file drawer problem is difficult because the set of hypotheses researchers intended to test is typically unobservable. One approach is to examine whether research projects lead to publications or other publicly available outputs (Franco et al., 2014). But a study may produce a paper while leaving out results for many, or even all, of the hypotheses it originally set out to test. Using the availability of research outputs to measure the file drawer problem may therefore understate the extent of missing results.

Study registries allow one to go beyond the availability of research outputs and examine whether results are available for each pre-registered hypothesis. In this study, we use the American Economic Association (AEA) RCT Registry to address two main questions. First, to what extent do results for pre-registered hypotheses in economics remain unreported and what role do null findings play in this under-reporting? Second, can interventions designed to address barriers to reporting make more results available?

The sample covers 317 studies pre-registered in the AEA RCT Registry between 2015 and 2017. We examine the availability of results eight to nine years after registration, when most studies would be expected to have reported their findings. To identify the hypotheses, we develop a standardized procedure for encoding individual hypotheses from registry entries and associated pre-analysis plans (PAPs), yielding 5,057 pre-registered hypotheses. We then conduct an extensive search of published papers, working papers, reports, and other publicly accessible materials to identify and document available results for each hypothesis.

We complement the descriptive analysis with a pre-registered randomized controlled trial (RCT) designed to investigate barriers to making results available. We randomly assign the studies to four experimental arms. In the Control arm, authors receive no intervention. Authors in all three treatment arms receive an email emphasizing the scientific value of reporting results for all pre-registered hypotheses. In the Awareness arm, the email is accompanied by an empty template (Results Report) for documenting their findings. In the Engagement arm, authors instead receive a study-specific Results Report pre-filled with their pre-registered hypotheses and the corresponding results we identified, or an indication that a result was not found. They are invited to review, correct, and complete the report. Finally, in the Resources arm, authors receive the same materials as in the Engagement arm and are offered up to 30 hours of research assistance or $1,500 to help document missing results. Approximately six months after the interventions, we implement a new search of publicly accessible materials. This experimental design allows us to assess how much raising awareness, engaging authors directly, and providing additional resources each contribute to making results available.

The first main finding is that a substantial share of pre-registered results remain unavailable nearly a decade after registration. Although 80.8% of studies have produced some publicly available research output, results are available for only 42% of pre-registered hypotheses on average across studies. Thus, 58% of hypotheses lack publicly available results, compared with only 19.2% of studies lacking any research output. Nearly half (48.3%) of the studies have made results avail-

able for some, but not all, of their pre-registered hypotheses. Using a more restrictive definition of availability, the share of missing results is even greater: specifically, only 24.0% of pre-registered hypotheses have results reported in the main body of a paper without modifications to the originally stated hypothesis. These findings show that missing results are a substantial problem in economics, and that focusing on whether studies produce research outputs understates its extent.

The second main finding is that null results account for a large share of the available evidence. On average across studies, 64.5% of available results are not statistically significant at the 5% level. To put this finding in perspective, we elicit forecasts from 68 experts, who substantially underestimate both the availability of results and the share of null findings among available results. Their forecasts are consistent with the conventional emphasis on selective reporting of null results, while our findings suggest that the problem of missing results may be broader. Experts do, however, correctly anticipate that the share of null results among available results is lower in studies published in top-five general interest journals than in other economics journals.

The third main finding is that direct engagement with researchers increases the availability of pre-registered results. Both the Engagement and Resources interventions lead to statistically significant increases in result availability relative to the control group. Pooling these two treatments, we estimate an increase of 6.2 percentage points in the share of hypotheses with available results, corresponding to approximately 11% of initially unavailable results. Simply raising awareness has no statistically significant effect on result availability, and we find no evidence that offering additional research assistance increases reporting beyond the effect of direct engagement.

The results made available during follow-up also shed light on the role of null findings in under-reporting. On average across studies, 68.4% of these results are not statistically significant at the 5% level, compared with 64.1% of results available at baseline, a difference that is not statistically significant ($p = 0.563$). While this does not rule out selective reporting of null results, it suggests that the suppression of non-statistically significant findings may not be the primary explanation for the extensive under-reporting we document.

The interventions also lead authors to explain why some results are unavailable. The Awareness treatment increases the share of hypotheses for which authors provide an explanation, even though it does not increase result availability. The effects of direct engagement are larger. Pooling the Engagement and Resources treatments, we estimate a 10.3 percentage point increase in the share of hypotheses for which authors explain why results are unavailable. Combining newly available and explained results, these interventions reduce the share of hypotheses that remain unaccounted for by 16.4 percentage points, a reduction of approximately 26% relative to the control-group mean.

We find little evidence that the treatment effects differ across subgroups of studies, and the main results are robust to accounting for potential spillovers through co-author networks. Taken together, these findings provide evidence that direct engagement with researchers can substantially reduce the share of hypotheses that remain unaccounted for, while also highlighting how difficult it is to make all results available. Even after the interventions, a substantial share of pre-registered hypotheses (44.8% in the treated studies) remains without either an available result or an explanation.

This paper contributes to the literature on publication bias and research transparency in several ways. Previous work has documented selective publication of studies and results (Franco et al., 2014; Andrews and Kasy, 2019), as well as evidence of inflated statistical significance (Brodeur et al., 2016, 2020). Abrams et al. (2026), provide an assessment of the AEA RCT Registry and document important limitations in registration practices, including incomplete registration, insufficiently detailed preregistrations, and limited post-trial reporting. Relatedly, Leight et al. (2025) examine

the research outputs of early studies on the AEA RCT Registry and the reporting of pre-registered primary outcomes in paper abstracts. We contribute to this literature by more systematically documenting the extent of missing results at the level of individual pre-registered hypotheses, examining the entire paper and appendix for results, exploring spillovers through co-author networks, and by providing experimental evidence on interventions aimed at reducing barriers to making these results available. Finally, we introduce a simple, standardized format for reporting results of pre-registered hypotheses, the Results Report, which may also be useful for registries and journals.

The remainder of the paper is organized as follows. Section 2 describes the sample, research design, data collection, and expert forecast study. Section 3 presents the empirical strategy. Section 4 presents the descriptive analysis and experimental results, including analyses of treatment effect heterogeneity and potential spillovers through coauthor networks, and the final section concludes. The Online Appendix provides supplementary analyses, an overview of how the analysis relates to the pre-analysis plan, details on the encoding procedure, an example Results Report, the experimental materials, and details of the expert forecast study.

2 Study design

This section describes the sample selection, research design, data collection, sample characteristics, and an expert forecast study that benchmarks the findings against researchers' expectations. The study was pre-registered on the AEA Registry.[1]

2.1 Sample selection

We construct a sample of pre-registered RCTs using the AEA Registry, as summarized in panel A of Figure 1.[2] The sampling frame consisted of all registrations between January 2015 and May 2017 ($n = 913$), a period in which RCT pre-registration was becoming increasingly common in economics and which leaves a follow-up window of eight to nine years after each registration, for tracking whether study results became public. To construct the sample, we follow a pre-specified sampling procedure. We exclude trials that were not genuinely pre-registered, i.e. registered only after the intervention or data collection had been completed (350), all but the earliest remaining AEA registration in the study window for each principal investigator (152), studies authored by members of our own research team (7), and registrations that were insufficiently informative (52). We then randomly set aside 35 eligible control studies for possible later encoding. This leaves an analysis sample of 317 pre-registrations.

[FIGURE 1 HERE]

[1]The pre-registration (AEARCTR-0013755) can be found at `https://www.socialscienceregistry.org/trials/13755`. See Appendix B for an overview of deviations from the pre-analysis plan.

[2]The AEA RCT Registry was established in 2013, and registration became mandatory for publication in AEA journals in 2015. Registrations provide time-stamped information on the experimental design, outcomes, study population, sampling strategy, randomization, and sample size; researchers may also submit a more detailed pre-analysis plan (PAP). At the time of writing, the Registry contains more than 13,000 trials (AEA RCT Registry, 2025).

2.2 Research design

Overview. Figure 1, panels B–D, provides an overview of the research design. At baseline, for each pre-registration in the analysis sample, we followed a standardized procedure for coding pre-registered hypotheses and then systematically searched publicly available research outputs for results corresponding to each hypothesis (panel B). For each pre-registered hypothesis, we collected the coefficient estimate, standard error or *p*-value, its location in publicly available research outputs, reasons for missing results, and any relevant heterogeneity estimates. This procedure produced a hypothesis-level baseline dataset that records whether a result is publicly available for each pre-registered hypothesis and, when available, whether the result is statistically significant. An important feature of the design is thus that we track results at the level of individual pre-registered hypotheses rather than at the study or paper level.

The study was implemented sequentially in four intervention waves, starting with the oldest pre-registrations in the analysis sample. Once baseline data collection had been completed for the pre-registrations in a given wave, they were randomized and the interventions were implemented (panel C), while baseline data collection continued for later waves. This process was repeated across all four waves between June 2024 and August 2025. At follow-up, approximately six months after each intervention launch, we again searched for publicly available results for the same pre-registered hypotheses (panel D). Together with the baseline search, this yields hypothesis-level measures of result availability and statistical significance at baseline and follow-up (see Appendix Figure A.1 for a detailed timeline).

The three intervention arms were designed to address potentially important constraints on the public availability of pre-registered results: awareness, engagement, and resources. The *Awareness* intervention addresses the possibility that researchers may not fully appreciate the scientific value of making results for all pre-registered hypotheses publicly available. The *Engagement* intervention additionally involves direct engagement by our research team with the authors' research and reduces the time and effort required for them to organize and report their results. The *Resources* intervention goes one step further by providing research assistance that could be used to generate or document missing results. The interventions are therefore cumulative, with each successive arm addressing an additional potential constraint on result availability.

In the first intervention wave, 140 pre-registrations were assigned in equal proportions to the *Control*, *Awareness*, *Engagement*, and *Resources* arms. In each of the subsequent waves, we planned to assign 5% of pre-registrations to the *Control* arm and the remaining 95% in equal proportions to the three intervention arms, increasing statistical power for comparisons across the interventions.[3] Following the pre-specified randomization procedure, we stratified treatment assignment by whether the pre-registration included a pre-analysis plan (PAP) and whether the study population was in a low- or middle-income country (LMIC) or a high-income country (HIC), with the former serving as a proxy for studies in development economics.[4]

[3]This change in allocation followed a pre-specified contingency. At the time of pre-registration, eight to nine years had already passed. We therefore specified that we could reduce the share assigned to the *Control* arm if the first-wave data showed little change in result availability over time, as they did. The eligible control studies set aside during sample construction could be added to the *Control* arm if substantial changes in result availability were subsequently observed. In practice, 16 (9.0%) of the pre-registrations in the later waves are in the *Control* arm, more than planned, because of rounding in small batches of studies and because some studies intended to be set aside were coded and included.

[4]The randomization was additionally blocked by the research team member responsible for the baseline coding. In the small number of cases in which more than one team member coded a study, one was randomly selected.

Implementation. The different arms were implemented as follows. In the *Control* arm, study authors received no intervention and were generally not contacted. The only exception was when we needed to request a non-public PAP before the intervention, which occurred for 14% of pre-registrations in the *Control* arm.[5]

In the *Awareness* arm, the corresponding author received an email emphasizing the scientific value of making results for all pre-registered hypotheses publicly available (Appendix, Figure E.1). Authors were encouraged to upload their results to the AEA Registry, either in their preferred format or using an empty Results Report developed by the research team to facilitate the reporting of the results of pre-registered hypotheses (Appendix, Figure E.9). The Results Report also included instructions for optionally posting it to the AEA Registry, as well as an example of a completed Results Report from another study.

In the *Engagement* arm, authors received the same email and materials as in the *Awareness* arm (Appendix, Figure E.2), together with a study-specific Results Report pre-populated with the standardized pre-registered hypotheses we had constructed and the corresponding results we had identified at baseline. For each hypothesis, the report indicated whether we had identified a corresponding result and, when available, provided the estimate we had recorded (Appendix, Figures E.12 and E.13). The report reflected our direct engagement with the authors' research, signaling that other researchers had taken an active interest in their studies and the reporting of their results. It also made it easier for authors to make their results publicly available by allowing them to review and, where necessary, correct and complete the information we had provided.

In the *Resources* arm, authors received the same materials as in the *Engagement* arm and were additionally offered up to 30 hours of research assistant support at the University of California, Berkeley or up to $1,500 to fund research assistance at their own institution (Appendix, Figure E.3). The support could be used to generate or document missing results.

The three intervention arms followed the same contact protocol. The initial email was sent to all publicly available email addresses for the corresponding author listed in the AEA Registry. If there was no response after two weeks, we sent a reminder to the same addresses. If we still received no response one week later, we sent a final email to all principal investigators listed in the registration using email addresses identified through an online search.

2.3 Data collection

This section provides details on the data collection at baseline and follow-up.

Baseline data collection. The first step was to construct a standardized set of hypotheses from each pre-registration and any associated PAPs. We used all relevant information available in the AEA Registry, supplemented by PAPs identified through an online search. We first recorded key features of the study population, including the country, sample size (when possible), unit of randomization, and a short description (e.g. "small business owners in city X").[6]

[5] Some authors indicate in the AEA Registry that they have a PAP without making it publicly available. We requested these PAPs to construct comparable baseline data. We view this limited contact as unlikely to affect subsequent reporting behavior.

[6] The AEA Registry contains a dedicated field for the study population, including the study country. Country information was not always recorded consistently and sometimes required encoders to consult other fields or documents.

We next recorded the main outcomes. There was considerable variation in the degree of detail provided by the authors for this field.[7] The most common sources of ambiguity were vaguely defined outcomes (e.g. 'health', 'income', 'education', 'spending', 'trust' rather than specific measures based on clearly noted questions or sources), under-specification of how any indices that combine multiple measures were to be constructed (i.e., simple averages, weighted averages, or something else), and ambiguity about whether individual measures, indices combining those measures, or both were intended to constitute main outcomes.

We then encoded the study interventions and corresponding treatment arms. Multiple interventions were sometimes combined into a single arm, which raised ambiguity about how interventions should be mapped into treatment arms. For example, a study might include two different microfinance interventions but pool them into a single treatment arm for the main comparison with the control group. This ambiguity becomes particularly important in studies with more than two interventions (counting control as one), which account for 65.9% of the sample. In such studies, there may be several possible ways to group interventions into treatment arms for the relevant comparisons, yet registrations often do not clearly specify the intended arms. In these cases, we defined the arms using the design that most closely addressed the research questions stated in the registration, or what appeared to us to be the simplest and most intuitive design.

With these elements in place, we stated each pre-registered hypothesis in a standardized format, recording the following components: the arms to contrast, the statistical test, an alternative hypothesis, and the estimation method (including standard error estimation). The resulting set included all main causal hypotheses based on the experimental design, as well as all pre-registered heterogeneity analyses for which the relevant subgroups or dimensions were clearly motivated and specified. Descriptive and other non-randomization-based analyses were excluded, as were vague or exploratory references to heterogeneity for which the relevant subgroups were not sufficiently defined. Each hypothesis was then classified into one of four categories: (1) a simple difference between conditional expectations, (2) a double difference between conditional expectations, (3) a joint hypothesis test (for instance, across multiple study arms), and (4) other hypotheses that did not fall into the first three categories, such as those involving structurally estimated parameters. Whenever encoding required a judgment call, we recorded the rationale and any relevant precedent.[8]

The final step of baseline data collection was to identify and record publicly available results corresponding to each of these pre-registered hypotheses. For each recorded hypothesis, a systematic online search was conducted across potential sources of dissemination, including published journal articles, working papers, policy reports, authors' websites, and any other publicly accessible materials. This process enabled the construction of a hypothesis-level database linking each pre-registered hypothesis to any related publicly available results. When a result was found, we recorded the coefficient estimate, p-value, standard error, statistical significance, sample size, source location (e.g., the relevant table in the article), whether it appeared only in an appendix, and the effort required to find and extract the information (classified subjectively as easy, medium, or hard). We also recorded any discrepancies between the pre-registered hypothesis and the reported analysis, including differences in outcomes or treatment arms, and whether the authors provided a justifica-

[7] Other authors have made similar observations, including Abrams et al. (2026); Ofosu and Posner (2023) and Leight et al. (2025).

[8] Specifically, an XML schema was developed to encode pre-registrations and hypotheses in a consistent manner, extending the fields defined in DDI Alliance (2021) and Cavanagh et al. (2023). The schema is available at https://github.com/ErikOSorensen/rgtools.

tion for these changes. If no result corresponding to the pre-registered hypothesis could be located in any of the materials, the hypothesis was coded as *not available*. If only qualitative evidence was available (e.g., whether the result was null, mixed, or strong), the hypothesis was coded as *partially available*. For example, this occurred when authors stated that a result was not statistically significant but did not report the corresponding estimate or standard error; in practice, these cases were quite rare (Appendix Table B.1).

Follow-up data collection. Approximately six months after each intervention wave, we repeated the systematic search for publicly available results corresponding to the pre-registered hypotheses in that wave. In constructing the follow-up data, we also incorporated information received from authors in the treatment groups following the intervention, including completed Results Reports, information provided in email correspondence, and pointers to additional publicly available materials. In a small number of cases, information identified during the follow-up period revealed that the baseline encoding of the pre-registered hypotheses needed to be revised. In these cases, we updated the hypothesis coding consistently at both baseline and follow-up. Apart from these cases, the follow-up used the same search and coding procedures as at baseline.

Main outcomes. We construct four main outcome variables at the study level.

The *fraction of hypotheses with available results* (*Available*) is the share of pre-registered hypotheses for which we find a numerical estimate and a measure of uncertainty (p-value, standard error, or qualitative statement about statistical strength). We use an inclusive definition of availability. A result is counted as available if it is found in a paper, appendix, or any other public record, including in the Results Report provided to the study team in the treatment arms. We also count results that differ somewhat from the pre-registered hypothesis as long as the encoders judge the reported analysis to be broadly consistent with it.

The *fraction of hypotheses with an explanation for the missing result* (*Explained*) is the share of all pre-registered hypotheses for which a) no result is available and b) the authors provided an explanation for why there is no result. We grouped these explanations into seven categories: (1) no intervention occurred or the intervention failed; (2) the intervention occurred but no data were collected; (3) the intervention occurred but the data were of poor quality; (4) data were collected but the analysis had not yet been carried out; (5) data were collected but statistical power was inadequate; (6) the authors disagreed with our interpretation that a given hypothesis did not have an available result; and (7) the authors disagreed with our assessment that the hypothesis was pre-registered. In the main analysis, we classify a result as explained only if it falls into the first two categories, which indicate that a result could not be made available because it did not exist, rather than reflecting other reasons for not reporting a result.

Based on *Available* and *Explained*, we also calculate the *fraction of hypotheses with missing results* (*Missing*), defined as the share of pre-registered hypotheses for which no result is available and no explanation for its absence is provided.

Finally, the *fraction of hypotheses with null results* (*Null*) is the share of pre-registered hypotheses for which the associated available result has a p-value that exceeds 0.05. When a p-value is not reported directly, we calculate it, where possible, from the reported estimate and standard error.

Quality control. Given the unstructured nature of the source material, extensive quality control was implemented to ensure consistent encoding. In a pilot, two principal investigators jointly coded five studies with the research assistants who later carried out most of the encoding. This collaborative approach was chosen instead of independent double-coding due to the high cost of encoding: we estimate that the encoding process described above takes approximately 11 hours per study on average.[9] Early involvement by the principal investigators also allowed us to develop common standards and precedents for subsequent encoding decisions, which was particularly important given evidence from related projects of substantial initial differences in coding reliability (van den Akker et al., 2023).

As an additional quality check, the first 10 studies in the analysis sample were independently double-encoded. This exercise revealed several discrepancies, often because coders had started from different versions of a pre-registration. In response, we established a system for tracking and resolving encoding issues through GitHub, and one of the principal investigators reviewed the first 100 encoded studies. Decisions and precedents arising from this process were documented and applied to subsequent studies. Over time, coding became qualitatively and quantitatively similar across coders, and assignment to treatment was blocked on encoder to ensure that any encoder-idiosyncratic tendencies would balance across treatments.

2.4 Sample characteristics

Table 1 presents descriptive statistics for the analysis sample of 317 pre-registrations.

[TABLE 1 HERE]

Panel A shows that 40% of pre-registrations have a PAP (found either on the AEA Registry or elsewhere). About half of the studies have resulted in a published paper, while only a small share have been published in a top-5 economics general interest journal.[10] 65% of the studies are conducted in low- and middle-income countries (LMICs).

Panel B describes the main elements of the pre-registrations. Each study has, on average, around 4 interventions, 5 experimental arms (including any control arms), and 5 main outcomes. Across the 317 pre-registrations, we encode 5,057 hypotheses, with a mean of nearly 16 and a median of 9 hypotheses per study. Of these, 3,731 are main hypotheses and 1,326 are heterogeneity hypotheses. Overall, 43% of studies pre-register more than 10 hypotheses, reflecting the highly skewed distribution of the number of hypotheses across studies. The large number of hypotheses partly reflects the fact that experimental design dimensions combine multiplicatively: even studies with a moderate number of interventions, treatment arms, outcomes, and heterogeneity dimensions can generate a substantial number of distinct hypotheses.

Turning to hypothesis characteristics, Panel C shows that, on average across studies, only 31% of hypotheses were coded as 'detailed', meaning that the registration provided sufficient information for encoding without additional judgment calls by the encoder. This share differs sharply by study type: among studies with a PAP, 65% of hypotheses are considered detailed, compared to only 9% among those without a PAP. The difference underscores the potential importance of PAPs in

[9] Additional details on the encoding process are presented in Section C of the Appendix.

[10] Namely, the *American Economic Review*, *Econometrica*, *Journal of Political Economy*, *Quarterly Journal of Economics*, and *Review of Economic Studies*.

constraining researcher discretion and clarifying pre-registered analytical choices.[11] Most hypotheses involve relatively straightforward experimental comparisons: 92% can be expressed as simple differences and 3% as double differences, while only 5% fall into more complex categories. Thus, most pre-registered studies in the sample feature research designs that can be expressed as direct experimental comparisons rather than, for instance, structural econometric estimates.

2.5 Expert study

To benchmark the findings against researchers' expectations, we conducted a forecast survey using the Social Science Prediction Platform (SSPP). A total of 68 experts participated, including researchers, graduate students, and practitioners in economics and related social science disciplines. Participants first predicted the median number of main hypotheses in a pre-registered study. We then elicited their expectations about result availability in three steps: the share of studies with any publicly available research output eight to nine years after registration; among these studies, the share reporting none of their pre-registered hypotheses; and, among studies reporting at least one hypothesis, the average share of hypotheses reported. We combine these responses to construct each participant's prediction of the average fraction of pre-registered hypotheses with available results. Participants also predicted the average share of reported results that are null (i.e., not statistically significant). Respondents further predicted whether the fraction of available and null results would be higher, lower, or the same across study characteristics, including the presence of a PAP, the study population (LMIC versus HIC), publication in an academic journal, and publication in a top-5 journal. The complete survey instrument is provided in Appendix Figure F.1.

Although the number of forecasters is relatively modest, prediction exercises with similar sample sizes have been shown to provide useful benchmarks for researchers' expectations (DellaVigna et al., 2019). The forecasts allow us to compare economists' expectations with actual reporting practices.

3 Empirical Strategy

We first outline the empirical strategy for estimating the causal effects of the interventions and examining treatment effect heterogeneity, before turning to potential spillovers across studies through overlapping co-author networks.

3.1 Main analysis

The main analysis estimates the causal effect of each treatment on the four main outcomes using two specifications: a difference-in-differences specification and a cross-sectional specification using follow-up outcomes and controlling for baseline availability. The difference-in-differences specification is:

$$Y_{it} = \mu_i + \delta_1 I_{t=1} + \sum_{j=1}^{3} \tau_j \left(T_{ij} \times I_{t=1} \right) + \varepsilon_{it}, \quad t = 0, 1, \tag{1}$$

[11] For related evidence on the effects of pre-analysis plans on research outcomes, see Brodeur et al. (2024).

where Y_{it} is the relevant outcome for study i at baseline ($t = 0$) or follow-up ($t = 1$), and T_{ij} indicates assignment to arm j, where $j = 0, 1, 2, 3$ correspond to Control, Awareness, Engagement, and Resources, respectively. The specification includes study fixed effects (μ_i) and a time fixed effect ($\delta_1 I_{t=1}$). The coefficients τ_j capture the estimated effect of each treatment arm relative to Control.

As an alternative, we estimate treatment effects using the follow-up cross-section, controlling for baseline availability and the variables used to stratify randomization:

$$Y_{i1} = \mu_0 + \sum_{j=1}^{3} \tau_j T_{ij} + \rho\, Y_{i0} + \mathbf{X}_i'\gamma + \varepsilon_{i1}, \tag{2}$$

where Y_{i0} is the fraction of hypotheses with available results at baseline and $\mathbf{X}_i$ includes indicators for LMIC status, PAP status, and encoder.

To examine heterogeneity in treatment effects, let Z_i indicate whether study i belongs to a given subgroup. We extend Equation (1) as follows:

$$Y_{it} = \mu_i + \delta_1 I_{t=1} + \delta_{1Z}\left(Z_i \times I_{t=1}\right) + \sum_{j=1}^{3} \tau_j \left(T_{ij} \times I_{t=1}\right) + \sum_{j=1}^{3} \gamma_j \left(Z_i \times T_{ij} \times I_{t=1}\right) + \varepsilon_{it}, \quad t = 0, 1. \tag{3}$$

The coefficients γ_j capture differences in treatment effects for studies in the subgroup versus others; we examine heterogeneity by PAP status and whether the study population is in an LMIC. The cross-sectional specification is extended analogously.

3.2 Spillover analysis

A potential issue is that the interventions may generate spillovers across studies through co-author networks. To measure such spillovers, let S_{ik} denote the number of co-author links from study i to studies in arm k, where $k = 0, 1, 2, 3$ correspond to Control, Awareness, Engagement, and Resources, respectively: for each other study in arm k, we count the authors it shares with study i, and sum over these studies. We define treatment exposure as $S_i = \sum_{k=1}^{3} S_{ik}$ and total network size (including links to the Control group) as $P_i = \sum_{k=0}^{3} S_{ik}$.

We estimate spillover effects by extending the main difference-in-differences specification. Let D_i indicate assignment of study i to any of the three treatment arms. The specification is:

$$Y_{it} = \mu_i + \delta_1 I_{t=1} + \sum_{j=1}^{3} \tau_j (T_{ij} \times I_{t=1}) + \beta_0 (S_i \times I_{t=1}) + \theta (S_i \times I_{t=1} \times D_i) + \sum_{j=0}^{3} \psi_j (T_{ij} \times I_{t=1} \times P_i) + \varepsilon_{it}. \tag{4}$$

Here, β_0 captures spillovers from treatment exposure onto Control studies, while θ captures whether the effect of such exposure is stronger ($\theta > 0$) or weaker ($\theta < 0$) for treated studies than for Control studies. We control for total co-author network size, P_i, to distinguish treatment exposure from the extent to which a study is connected to other studies in the sample. We also report results from a more flexible specification allowing spillovers to vary by own treatment assignment and the treatment arm to which co-authors are exposed.

4 Results

We first document the extent of missing results in the sample and compare the findings with expert predictions. We then examine whether the interventions increase result availability, before turning to heterogeneity in treatment effects and potential spillovers through co-author networks.

4.1 Missing Results

Figure 2 provides an overview of results reporting at baseline. The analysis sample consists of 317 studies and 5,057 pre-registered hypotheses (Panel A). Eight to nine years after pre-registration, 19.2% of the studies have no publicly available research output, whether a published paper, working paper, or report (Panel B). Focusing on whether a study has any research output, however, masks substantial under-reporting at the hypothesis level. Panel C shows that only 20.2% of studies have made results publicly available for all of their pre-registered hypotheses, and that 12.3% of studies have no available results for any of their pre-registered hypotheses even though they have a paper. Nearly half of the studies have made results available for some, but not all, of their pre-registered hypotheses.

[FIGURE 2 HERE]

Panel D shows the fraction of pre-registered hypotheses with available results for each study. On average across studies, results are available for only 42% of pre-registered hypotheses nearly a decade after registration. To our knowledge, this provides the first estimate of the size of the file drawer in economics at the level of individual pre-registered hypotheses.

Table 2 examines how the share of available results varies across hypothesis types, definitions of availability, and study characteristics. In Panel A, we observe that the share of available results is much lower for heterogeneity hypotheses than for main hypotheses (22.3% vs. 44.7%, $p < 0.001$), suggesting greater selectivity in the reporting of heterogeneity results. This difference also holds when restricting the sample to the 39 studies with heterogeneity hypotheses (22.3% vs. 53.8%, $p < 0.001$). In Panel B, we consider alternative definitions of availability. We find that the average share of pre-registered hypotheses with results reported in the main body of the paper is 34.8%, while the share with results that match the pre-registered hypotheses without modifications is 28.6%. The average share of available results that meet our strictest definition of availability—being reported in the main body of the paper without any modifications to the pre-registered hypotheses—is only 24.0%.

In Panel C, we examine how availability varies across study characteristics. The share of available results is only marginally higher for studies with a pre-analysis plan (43.9% vs. 40.7%,

$p = 0.466$), while it is lower for studies conducted in low- and middle-income countries (37.8% vs. 49.8%, $p = 0.010$). Among studies with a published paper, the share of available results is substantially lower for papers published in top-five journals than for those published in other journals (41.3% vs. 55.3%, $p = 0.046$).

[TABLE 2 HERE]

We next turn to null results. Figure 3 shows the distribution of the share of available results that are null. Null results are remarkably common: on average across studies, 64.5% of available results have a p-value above 0.05, and for almost 30% of studies, at least 90% of available results are null. While there is substantial heterogeneity, the majority (62%) of studies report more null results than non-null results. This pattern runs counter to the common perception that available results in economics are predominantly statistically significant. Table 2 shows that the share of null results is strikingly consistent across hypothesis types, definitions of availability, and study characteristics, with one notable exception: the share is substantially lower in studies published in top-five journals than in other journals (51.3% vs. 65.3%, $p = 0.030$).

[FIGURE 3 HERE]

Figure 4 compares these findings to the expectations of experts. The upper two panels show their predictions for the share of available results and the share of available results that are null. Panel A shows that experts, on average, underestimate the share of available results: 27% compared with 42% ($p < 0.001$). Only a small minority (21%) expected a higher share than we actually observe. The expert forecasts are more closely aligned with the share obtained using our strictest definition of availability (24.0%), suggesting that their expectations may be shaped primarily by results reported in the main body of papers without modifications.[12]

[FIGURE 4 HERE]

Panel B shows that experts, on average, substantially underestimate the share of available results that are null, at 37% compared with 64.5% ($p < 0.001$); overall, 91% of experts underestimate this share. Most experts (69%) also expect the majority of available results to be non-null (i.e., statistically significant), in contrast to what we observe. These expectations are consistent with the view that the research process systematically disfavors studies with null results.

The lower two panels show how experts expect the share of available results (Panel C) and the share of available results that are null (Panel D) to vary across study characteristics, compared with the corresponding regression estimates from the data.[13] Experts expect studies with a pre-analysis plan (PAP) to have both a higher share of available results ($p < 0.001$) and a higher share of null results ($p < 0.001$), whereas we find no statistically significant differences in either ($p = 0.466$ and $p = 0.236$). Experts also tend to expect a higher share of available results for studies conducted in low- and middle-income countries (LMICs), although this difference is not statistically significant ($p = 0.200$). In contrast, we find a lower share of available results for these studies (-12.0

[12] Experts also substantially underestimated the number of hypotheses per study, with a median forecast of 4 compared with an observed median of 9.

[13] See Table A.1 for further details on the expert forecasts by study characteristics.

percentage points, $p = 0.010$). Finally, experts expect a lower share of available null results both in published studies ($p = 0.022$) and in studies published in top-five journals ($p < 0.001$), a pattern we find evidence of only for studies published in top-five journals ($p = 0.030$, compared with $p = 0.333$ for published studies).

4.2 Increasing the Availability of Results

We now examine whether the interventions reduced the share of missing results and increased the availability of null results. We proceed in three steps. First, we examine whether the interventions generated engagement with the information we provided and elicited informative responses from the authors. Second, we document how the follow-up changed the distribution of available, explained, and missing results, as well as the share of null results. Finally, we use regression analysis to estimate the causal effects of the interventions.

The follow-up combined a systematic search for new results with information obtained directly from authors. The interventions generated substantial engagement. Across the three treatment arms, 88% of the studies had at least one author click on the Results Report, and about half of the studies in the Engagement and Resources arms had an author access the detailed information about their own study. The interventions also generated substantial email correspondence, with about two-thirds of the author teams responding. Importantly, the Engagement and Resources arms generated more informative responses than the Awareness arm. We classify a response as informative if it provided new results or an explanation for why results were unavailable. Informative responses were received from 30.7% of the studies in the Engagement arm and 19.3% in the Resources arm, compared to 8.9% in the Awareness arm. As expected, we received no responses from authors in the control group, who had not been contacted directly.[14]

Figure 5 shows how the follow-up changed the availability of results and explanations for unavailable results (Panel A), as well as the availability of null results (Panel B), in the treatment arms. The grey bars show the baseline distribution, while the blue bars represent additional results found in the follow-up and the yellow bars represent explanations for unavailable results.

Panel A shows a substantial increase in both the share of available results and the share of results for which an explanation was obtained. These increases came both from studies that already had some available results at baseline and from studies with no available results. Overall, new results were identified for 27 studies (19 with and 8 without available results at baseline), while explanations for unavailable results were obtained for 30 studies (13 with and 17 without available results at baseline). The average share of available results increased from 43.0% to 47.3%, and explanations accounted for another 7.9%, bringing the total share of results accounted for to 55.2%.

Panel B shows that the follow-up also increased the availability of null results. The additional null results were found in 23 studies with different baseline shares of null results, including 8 studies for which no null results were initially available. Overall, the average unconditional share of null results increased from 28.0% to 31.0%; on average, 68.4% of the newly found results are null.

[FIGURE 5 HERE]

[14] See Table A.2 for further details on the responses from authors and Table A.3 for an overview of their explanations for why results were not available.

Table 3 presents the regression analysis of the treatment effects on available, explained, and missing results. Columns (1)–(2) report difference-in-differences estimates for available results, while columns (3)–(8) report cross-sectional estimates for the three outcomes, including the pre-specified controls and baseline availability.

We first consider the effects on available results. The difference-in-differences estimates in columns (1)–(2) show no significant effect of the Awareness treatment ($p = 0.073$), but substantial positive effects of both the Engagement ($p = 0.001$) and Resources treatments ($p = 0.004$).[15] The effect of Engagement is also significantly larger than that of Awareness ($p = 0.001$), while we find no significant difference between Engagement and Resources ($p = 0.687$). Pooling the latter two treatments, we estimate an increase of 6.2 percentage points in the share of available results ($p < 0.001$). The cross-sectional estimates in columns (3)–(4) yield almost identical results, with a pooled effect of 6.1 percentage points ($p < 0.001$).

Turning to explained results in columns (5)–(6), we find a somewhat different pattern. The Awareness treatment significantly increases the share of results for which an explanation is provided ($p = 0.025$), even though it has no significant effect on the availability of results. This shows that simply making authors aware of missing results is sufficient to elicit some explanations, while increasing the availability of results requires more intensive engagement. The effect of Engagement is larger than that of Awareness, although the difference is only marginally significant ($p = 0.051$), and the difference between Resources and Awareness is not statistically significant ($p = 0.349$). Again pooling Engagement and Resources, we estimate an increase of 10.3 percentage points in the share of explained results ($p < 0.001$).

Finally, columns (7)–(8) combine these two dimensions by examining the share of missing results, defined as results that are neither available nor explained. All three treatments significantly reduce missing results, although the effect of Awareness is smaller and operates almost entirely through explanations. Pooling the Engagement and Resources treatments, we estimate a reduction of 16.4 percentage points in the share of missing results ($p < 0.001$), corresponding to a 26% reduction relative to the control-group mean.[16]

[TABLE 3 HERE]

Table 4 presents the corresponding regression analysis of the treatment effects on the average share of null results across studies. We find the same pattern as for available results: no significant effect of the Awareness treatment ($p = 0.091$), but significant positive effects of both the Engagement ($p = 0.002$) and Resources treatments ($p = 0.015$). The effect of Engagement is also significantly larger than that of Awareness ($p = 0.004$), while we find no significant difference between Engagement and Resources ($p = 0.850$). Pooling the latter two treatments, we estimate an increase of 4.2 percentage points in the average share of null results ($p < 0.001$). The cross-

[15] For the two pre-registered outcomes, Available and Null, p-values are from one-sided tests, with the alternative hypotheses as specified in our pre-registration on the AEA RCT Registry (AEARCTR-0013755): positive effects, and a larger effect of each successive treatment arm (Appendix Table B.5). For Explained and Missing, which were not pre-registered, p-values are from two-sided tests. Appendix Table B.5 lists all pre-registered hypotheses and their p-values with and without correcting for multiple testing.

[16] Table A.4 shows that the main findings are robust to an alternative definition of availability, in which results identified during the follow-up as having been publicly available at baseline are also counted as available. The estimated treatment effects are smaller than in the main analysis, but qualitatively the overall pattern remains unchanged.

sectional estimates in columns (3)–(4) yield almost identical results, with a pooled effect of 4.1 percentage points ($p < 0.001$).[17]

[TABLE 4 HERE]

Importantly, the increase in the share of null results among all pre-specified hypotheses does not necessarily imply that the interventions disproportionately increased the availability of null results. To examine this, we complement the regression analysis in Tables 3 and 4 by comparing the average share of null results among newly available results with the corresponding share among results available at baseline in the treated studies. The share is 68.4% for newly available results, compared to 64.1% at baseline, a difference that is not statistically significant ($p = 0.563$).[18] Thus, despite concerns that null results are disproportionately missing from the empirical economics literature, we find no robust evidence of a difference in the share of null results between newly available results and those available at baseline.[19]

Finally, we conduct a joint test of the three pre-specified treatment comparisons for each of our two pre-specified outcomes, available and null results. The three comparisons are Awareness versus Control, Engagement versus Awareness, and Resources versus Engagement. A joint Wald test strongly rejects the hypothesis that all six comparisons are zero ($\chi^2(6) = 20.52$, $p = 0.002$; Table B.6).

4.3 Heterogeneity analysis

We next examine heterogeneity in treatment effects along two pre-specified dimensions: whether the study had a pre-analysis plan (PAP) and whether it was conducted in a low- or middle-income country (LMIC). Figure 6 presents the estimated treatment effects separately for each subgroup, with Panels A–D showing the effects on available, explained, missing, and null results, respectively. The figure reports point estimates and 95% confidence intervals for each of the three treatments.

Overall, there are similar treatment effects across the different subgroups, with few statistically significant differences. In particular, the positive effect of the Engagement treatment on the availability of results is observed across all subgroups and is of similar magnitude. We find only a few significant differences: the Resources treatment has a larger effect on availability for studies with a

[17] Tables A.5 and A.6 show the findings under two alternative specifications. The first uses an alternative definition of baseline availability, counting results identified during the follow-up as having already been publicly available at baseline. The second accounts for hypotheses with multiple expansions – meaning, in cases where there are multiple possible results corresponding to a single hypothesis, we randomly chose one for the intervention – by averaging over 100 random choices of the expansion that represents each such main hypothesis, rather than relying on the single random choice made for the intervention. Both specifications yield similar patterns, although the estimated effects are somewhat smaller under the alternative definition of baseline availability.

[18] This comparison weights each study equally. Pooling all hypotheses instead (Table A.7), the share of null results is 79.2% among newly available results and 68.9% among results available at baseline ($p = 0.038$, standard errors clustered by study), reflecting a small number of studies with many newly available null results. Under the alternative classification, in which results that authors pointed to in the follow-up as previously available are counted as available at baseline, the difference disappears (70.1% vs. 70.2%, $p = 0.990$).

[19] Consistent with this, a difference-in-differences regression of the share of null results among available results, for the 217 studies with available results in the baseline, gives a pooled effect of the Engagement and Resources treatments of −0.4 percentage points ($p = 0.534$). Because this share is only defined for studies with available results, the estimate conditions on availability and should be interpreted with caution.

PAP ($p = 0.015$), while the Awareness treatment has a larger effect on explained results ($p = 0.021$) and a larger reduction in missing results ($p = 0.016$) for studies conducted in LMICs. None of the remaining differences across subgroups are statistically significant. Moreover, Wald tests do not reject the joint null hypothesis that the effects of all three treatments are the same across subgroups, neither for each of Available, Explained, and Null separately nor for the three independent outcomes jointly (smallest $p = 0.112$ for PAP and $p = 0.142$ for LMIC; see Panel D of Tables A.8 and A.9).

4.4 Spillover Effects

Finally, we examine whether the interventions generated spillovers across studies through co-author networks. Figure 7 illustrates the co-authorship network among the 317 studies in our sample, with each node representing a study and connections indicating that two studies share at least one author. The shape of each node identifies its treatment assignment, while the thickness of each connection indicates the number of shared authors. The figure reveals considerable variation in the extent of co-author overlap. While 182 studies have no connections to other studies in the sample, the remaining 135 studies are connected through 110 study pairs. These connections form 45 separate co-authorship networks, most involving only a few studies (29 consist of just two studies), but some larger networks, with up to 12 studies, include studies assigned to all three treatment arms and the control group.

Figure 8 shows the distribution of the number of co-author links to studies assigned to any of the three treatment arms (Panel A), to studies assigned to the control group (Panel B), and to all studies in the sample (Panel C). Most studies have no or only a few co-author links, but there is variation in exposure to treated studies, with some studies having several such connections. Exposure to control studies is considerably less common, reflecting the smaller share of studies assigned to the control group.

Panel D presents the estimated spillover effects on available, explained, missing, and null results. The pre-specified empirical approach distinguishes between spillovers onto control studies, which capture whether exposure to treated studies affects outcomes in the absence of direct treatment, and spillovers onto treated studies, which capture whether such exposure amplifies or reduces the effects of the interventions. Since none of the control studies' available or null results changed between baseline and follow-up, the estimated spillovers onto control studies are exactly zero for these two outcomes. We find no statistically significant evidence of either type of spillover for any of the four outcomes. Tables A.10 and A.11 report the full set of regression estimates. The results are consistent across specifications, and the estimated direct treatment effects remain broadly similar to the main estimates. A joint Wald test of the unrestricted specification fails to reject the hypothesis of no spillover effects for either available results ($p = 0.703$) or null results ($p = 0.781$).

5 Conclusion

We document a substantial file drawer problem in economics. Nearly a decade after study registration, results are publicly available for only 42% of pre-registered hypotheses, even though 80.8% of studies have produced some research output. This highlights the importance of moving beyond a focus on whether studies produce research outputs to examining whether results for individual pre-registered hypotheses are available.

At the same time, the findings challenge the view that missing results are primarily driven by selective reporting of non-statistically significant findings. Nearly two-thirds of available results are null, and the results made available during follow-up due to the interventions are not significantly more likely to be null than those available at baseline. We cannot rule out that the results that remain unavailable differ systematically, but the large share of null findings among available results suggests that other mechanisms may also be important in determining which results are reported. A possibility is that selective reporting in economics depends on which findings contribute to a compelling research narrative, with both significant and null results potentially helping to tell the "story", i.e., by testing theories or establishing mechanisms. Other findings may be left out because they do not fit. The data do not allow us to test this explanation directly, but the findings suggest that selective reporting may reflect more than researchers hesitating to report null results.

One notable finding is that far fewer pre-registered hypotheses are available among studies published in top-5 economics general interest journals, and that these articles also feature significantly fewer null results. While this pattern is consistent with a greater degree of selective reporting in top-5 papers on average compared to other studies, there are many other potential explanations and confounding factors that cannot be ruled out.

The field experiment provides evidence on how the file drawer can be reduced. Direct engagement with researchers increases both the availability of results and the provision of explanations for missing results. Offering additional research assistance does not significantly improve outcomes beyond direct engagement, while simply raising awareness of the importance of reporting pre-registered results has no significant effect on their availability. The experiment also shows how challenging it can be to address the file drawer problem after research has been completed. Despite substantial efforts to engage authors and facilitate reporting, almost half of the pre-registered hypotheses do not have an available result. This points to the potential importance of establishing stronger expectations for reporting results as a normal part of the research process, along with infrastructure that can facilitate such reporting, rather than relying on efforts to recover missing results years later. The standardized Results Reports developed in this study offer one possible tool for such reporting and could be integrated into existing research registries.

More broadly, this study highlights the distinction between registering research and making results available. Pre-registration makes intended analyses observable, but does not ensure that the results become publicly available. Realizing the full potential of research registries may therefore require greater attention to what happens after registration, including how researchers document and share the results of their pre-registered hypotheses.

References

Abrams, Eliot, Jonathan Libgober, and John A. List (2026). "Research registries and the credibility crisis: An empirical and theoretical investigation." *Economic Journal*, 136(679): 2473–2530.

AEA RCT Registry (2025). "Registrations in the AEA RCT Registry (2013-05-15 through 2025-08-01)." Harvard Dataverse, doi: 10.7910/DVN/T8KQ1L.

van den Akker, Olmo R., et al. (2023). "Selective hypothesis reporting in psychology: Comparing preregistrations and corresponding publications." *Advances in Methods and Practices in Psychological Science*, 6(3): 25152459231187988.

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

Brodeur, Abel, Nikolai Cook, and Anthony Heyes (2020). "Methods matter: *p*-hacking and publication bias in causal analysis in economics." *American Economic Review*, 110(11): 3634–3660.

Brodeur, Abel, et al. (2016). "Star Wars: The Empirics Strike Back." *American Economic Journal: Applied Economics*, 8(1): 1–32.

Brodeur, Abel, et al. (2024). "Do preregistration and preanalysis plans reduce *p*-hacking and publication bias? Evidence from 15,992 test statistics and suggestions for improvement." *Journal of Political Economy Microeconomics*, 2(3): 527–561.

Cavanagh, Jack, et al. (2023). *A Metadata Schema for Data from Experiments in the Social Sciences*. Technical report, World Bank, Development Research Group. Policy Research Working Paper, 10296.

Chopra, Felix, et al. (2024). "The null result penalty." *Economic Journal*, 134(657): 193–219.

DDI Alliance (2021). *Data Documentation Initiative (DDI) Codebook 2.5: XML Schema and Documentation*. Technical report, DDI Alliance.

DellaVigna, Stefano, Devin Pope, and Eva Vivalt (2019). "Predict science to improve science." *Science*, 366(6464): 428–429.

Franco, Annie, Neil Malhotra, and Gabor Simonovits (2014). "Publication bias in the social sciences: Unlocking the file drawer." *Science*, 345(6203): 1502–1505.

Leight, Jessica, Viola Asri, and Taisuke Imai (2025). "Pathways from registration to publication: Evidence from the AEA RCT Registry." Mimeo, International Food Policy Research Institute.

Ofosu, George K. and Daniel N. Posner (2023). "Pre-analysis plans: An early stocktaking." *Perspectives on Politics*, 21(1): 174–190.

Romano, Joseph P. and Michael Wolf (2005). "Exact and approximate stepdown methods for multiple hypothesis testing." *Journal of the American Statistical Association*, 100(469): 94–108.

Romano, Joseph P. and Michael Wolf (2016). "Efficient computation of adjusted *p*-values for resampling-based stepdown multiple testing." *Statistics & Probability Letters*, 113: 38–40.

Rosenthal, Robert (1979). "The file drawer problem and tolerance for null results." *Psychological Bulletin*, 86(3): 638.

Rubin, Donald B. (1987). *Multiple Imputation for Nonresponse in Surveys*. Wiley Series in Probability and Mathematical Statistics. New York: John Wiley & Sons.

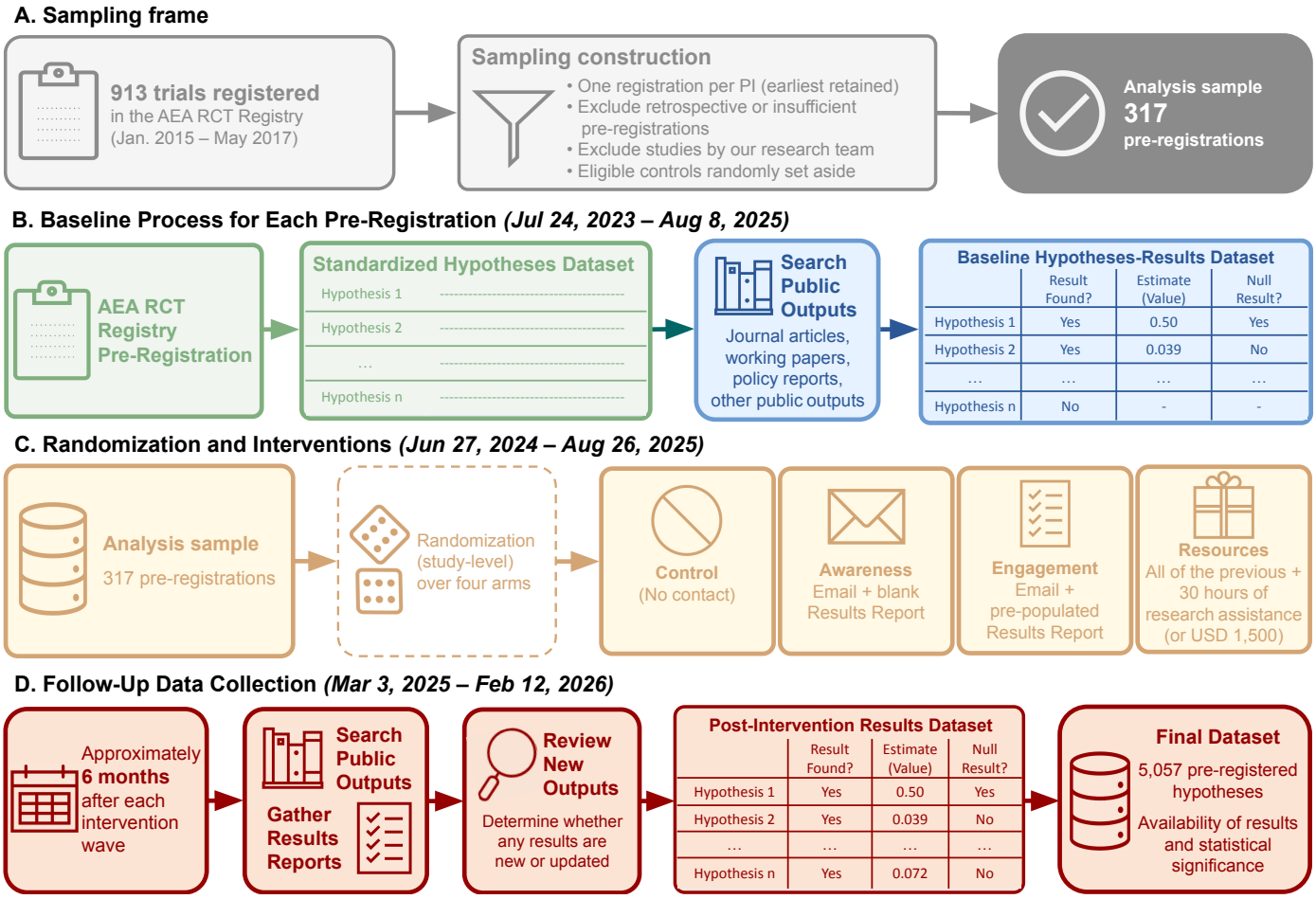


Figure 1: Study Sample, Encoding Process, and RCT Design

Note: This figure summarizes the study design and construction of the hypothesis-level datasets used throughout the paper. Panel A shows the selection of the analysis sample of 317 pre-registrations from the AEA RCT Registry. Panel B illustrates the baseline data collection and encoding process, in which pre-registered hypotheses were standardized and linked to results identified in public outputs. Panel C shows the study-level randomization across the four experimental arms and the interventions. Panel D illustrates the follow-up data collection and encoding process, in which public outputs and submitted Results Reports were reviewed and linked to the same pre-registered hypotheses to construct the final dataset. The timing of interventions across waves is shown in Figure A.1.

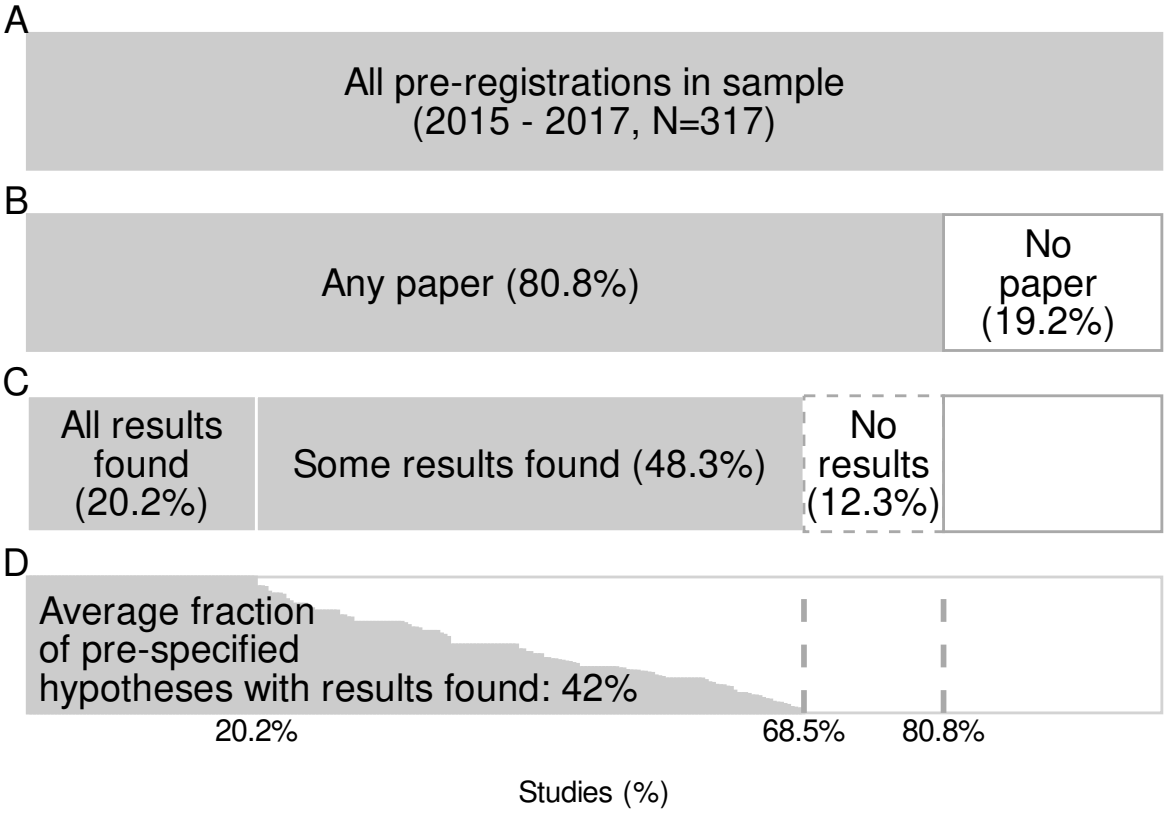


Figure 2: Reporting of Pre-specified Hypotheses in Economics

Note: The volume of the rectangle in each panel represents 100% of the sample. "Any paper" means any publicly available research output (a published paper, working paper, book chapter, or report; Section 4). "All results found" refers to studies in which results were found for all pre-specified hypotheses, "No results" means that no results corresponding to any of the pre-specified hypotheses of those studies were found. "Availability" is defined in Section 2.3. The height of the bottom panel represents the fraction of pre-registered hypotheses with results available by study, sorted on the share available hypotheses in decreasing order from left to right. The construction of the sample is described in Section 2.1.

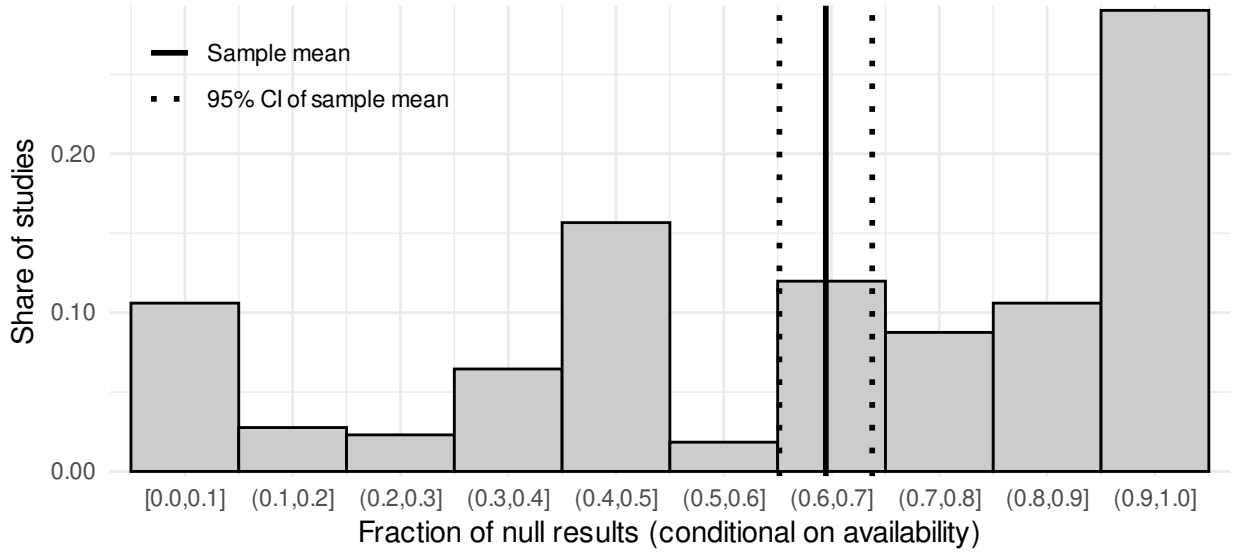


Figure 3: Distribution of Null Results Across Studies

Note: The histogram shows the distribution across studies of the fraction of available results that are null. The solid vertical line indicates the average across studies (64.5%), and the dotted lines indicate its 95% confidence interval.

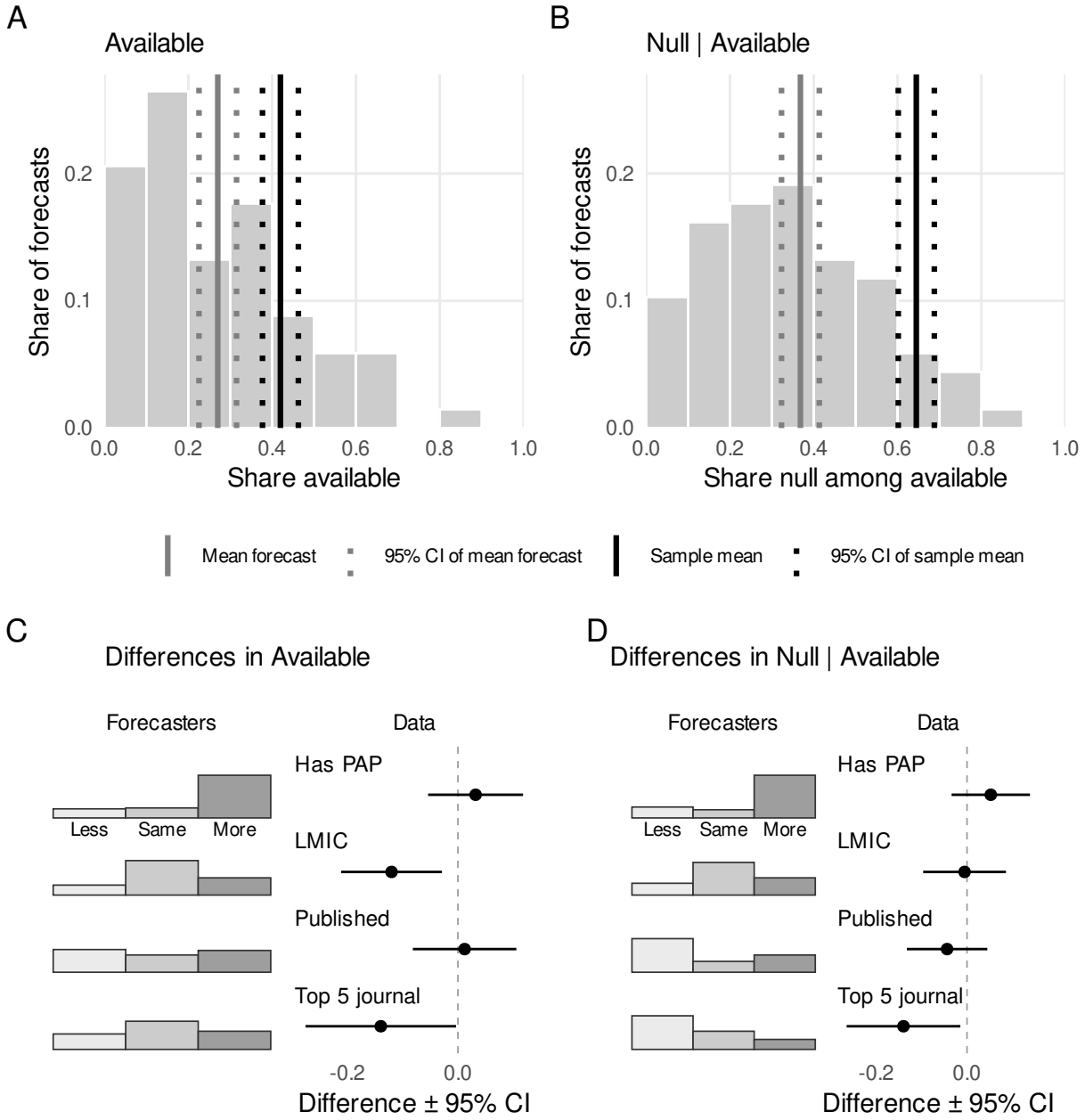


Figure 4: Priors and Data Regarding Pre-Specified Outcomes

Note: Panels A and B show the distribution of expert forecasts (bars), mean forecasts (grey solid lines), and observed sample means (black solid lines), with dotted lines indicating 95% confidence intervals. The outcomes are study-level shares of all pre-registered hypotheses: the share with available results in Panel A and the share of null results among available results in Panel B. Panels C and D show the shares of forecasters predicting lower, equal, or higher outcomes for studies with the indicated characteristic, alongside bivariate OLS estimates (study-level shares of all pre-registered hypotheses) and 95% confidence intervals from the AEA Registry sample. Published refers to studies with a journal publication relative to studies with other public outputs; Top 5 journal refers to publication in *AER*, *QJE*, *Econometrica*, *JPE*, or *ReStud* relative to other published studies. $N = 68$ forecasters; 83.8% economists, 76.5% academics or graduate students.

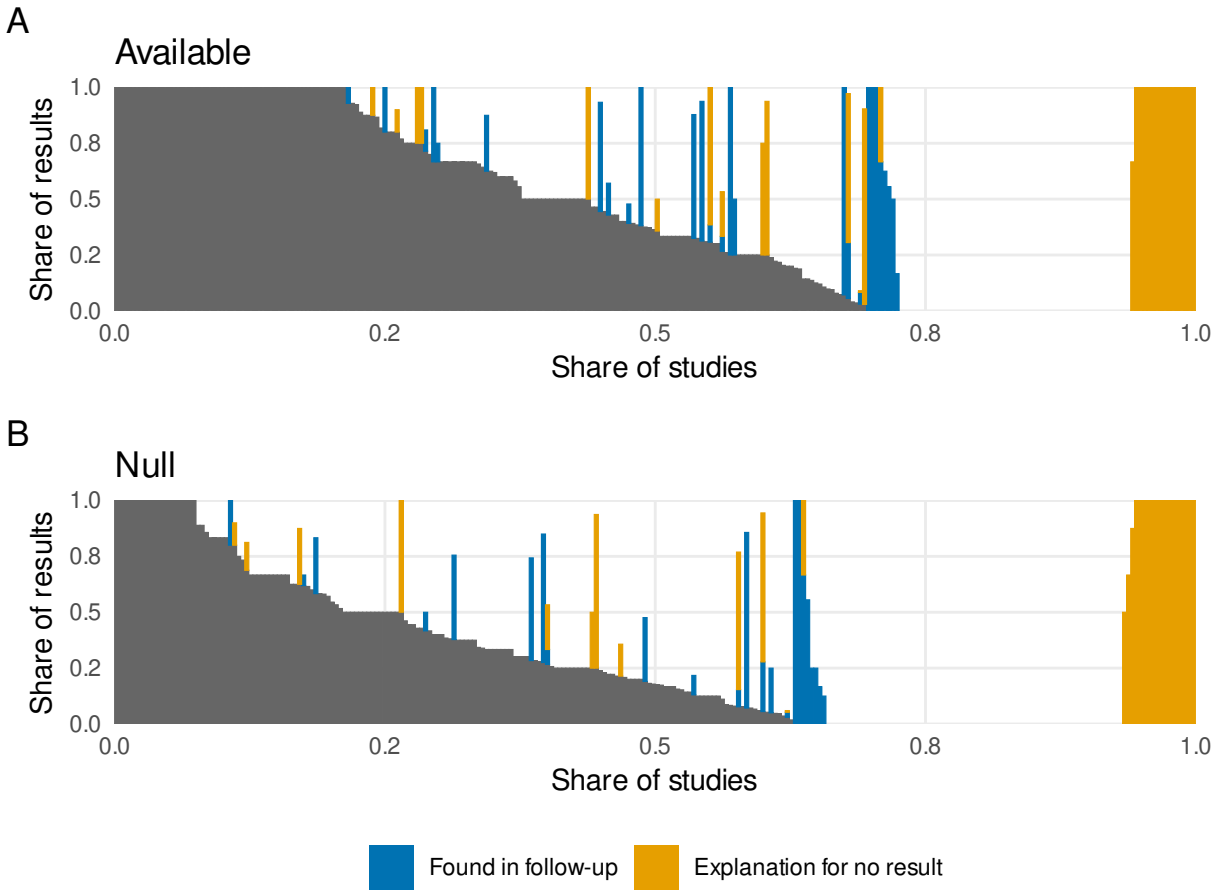


Figure 5: Distribution of main outcomes in the baseline and follow-up

Note: Each panel shows the distribution of an outcome in the baseline (panel A: the share of results that are available, panel B the share of null results (unconditional) in the treatment arms (there were no changes in the control arm, and it is not included). The baseline distribution is shown in dark grey. The blue bars represent results that were found in the follow-up, matched to the baseline numbers and extending the share of results found. The yellow bars correspond to shares of hypotheses that authors explained why there are no results. In these arms, the baseline average share available was 43.0%; 4.3% was found in the follow-up, and for 7.9% of non-available results an explanation was provided, giving 55.2% accounted for after the follow-up. The new information from the follow-up came from 27 studies with new results and 30 with explanations, for a total of 52 out of 266 studies with either new results or explanations in the treated arms.

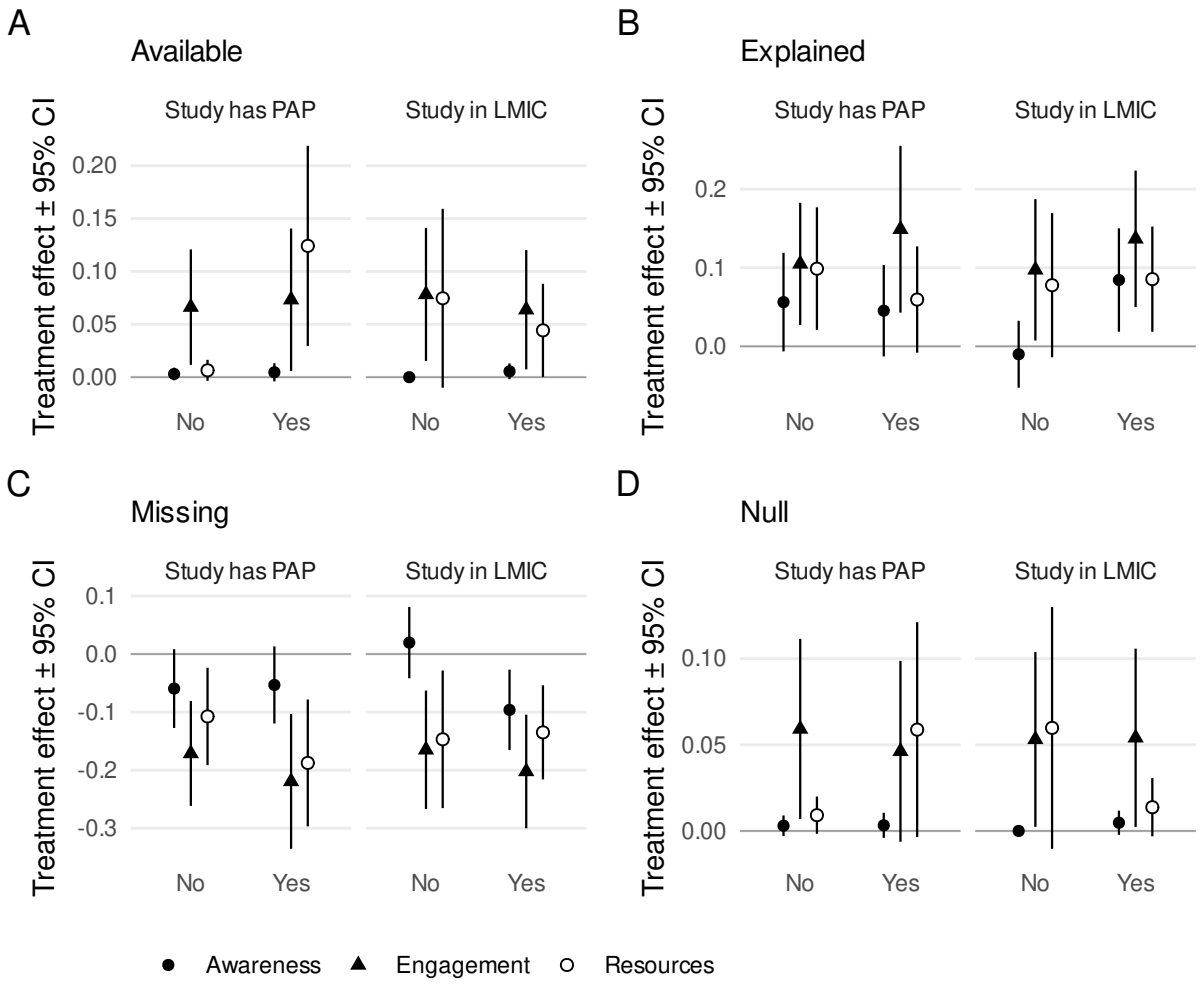


Figure 6: Heterogeneity of treatment effects

Note: Each panel shows the treatment effects on one outcome, estimated separately for two pairs of subgroups: studies with and without a pre-analysis plan (PAP), and studies with and without a study population in a low- or middle-income country (LMIC). For the outcomes 'Available' and 'Null', difference-in-difference estimates are shown; for 'Explained' and 'Missing', cross-sectional-estimates are shown which includes baseline availability (corresponding to specifications in Table 3 and Table 4). The error bars show 95% confidence intervals based on robust standard errors.

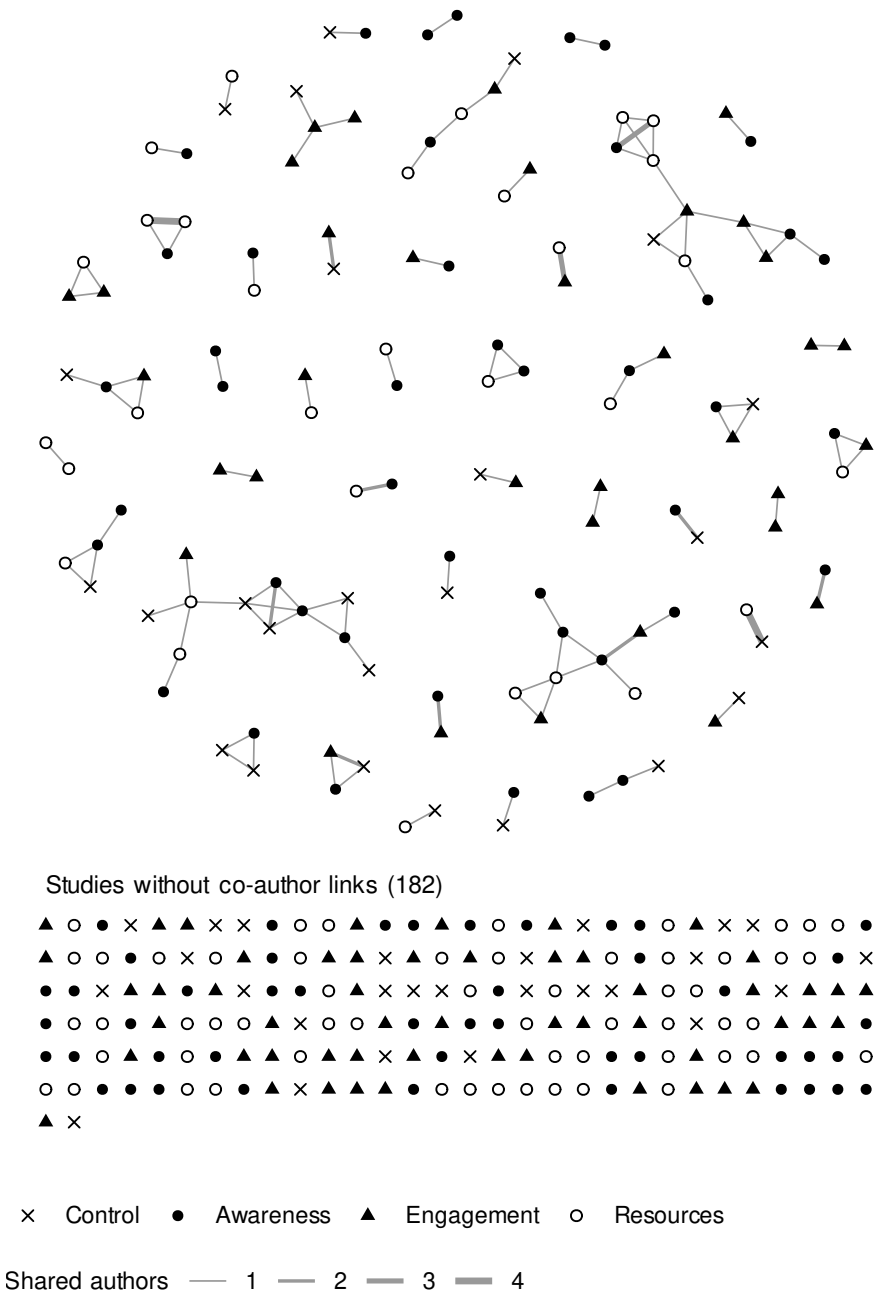


Figure 7: Co-author network graphs

Note: Co-authorship network of the 317 studies in the analysis sample. Each node is a study, with its shape giving the study's treatment arm (Control, Awareness, Engagement, Resources). Two studies are connected when they share at least one author, i.e. a person listed as primary or other primary investigator in the AEA RCT Registry entries of both studies (registry snapshot of September 2024; authors matched by name). Edge weights give the number of shared authors. 135 studies are connected to at least one other study, through 110 study pairs; the remaining 182 studies, shown in the grid below, share no authors with other studies in the sample. Node positions within the network come from a force-directed layout and carry no meaning beyond the connections.

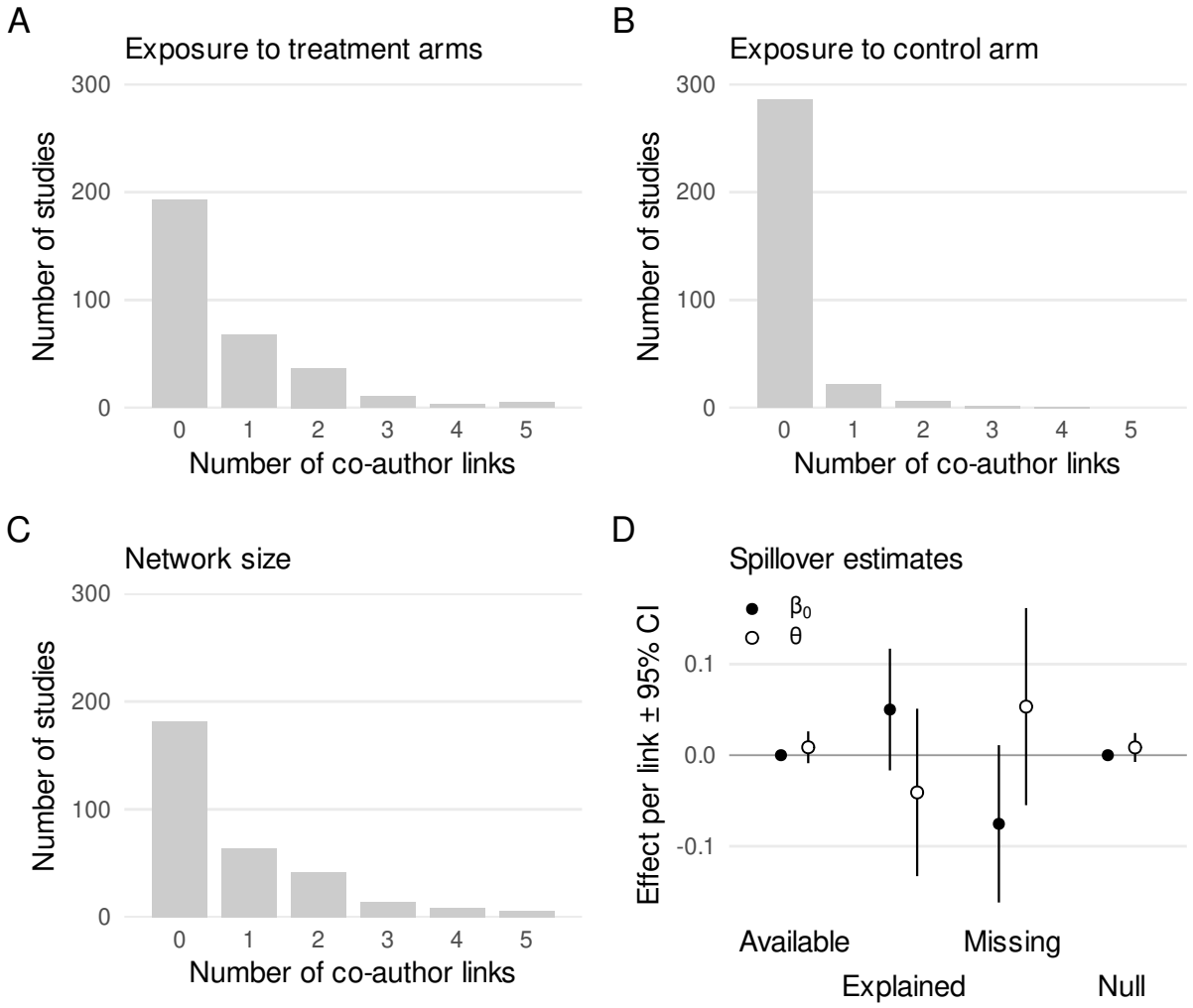


Figure 8: Distribution of co-author network exposure and estimated spillover effects

Note: Panel A shows the distribution of exposure to any treatment arms (S_i), Panel B shows the distribution of exposure to the control ($P_i - S_i$), and Panel C shows the distribution of total network size (P_i). Panel D shows the estimated spillover effects β_0 and θ with 95% confidence intervals for the restricted model reported in the main text, estimated with differences-in-differences (Available and Null) or cross-sectional with baseline Available as control (Explained and Missing). Because there were no changes in the available or null results in the control group, β_0 is exactly zero, with no confidence interval, for Available and Null.

Table 1: Baseline Characteristics of the Analysis Sample

		Quantiles of outcomes				
Dimension	Average	10th	25th	50th	75th	90th
A. Study characteristics						
Registered with a PAP	0.40	0	0	0	1	1
Conducted in LMIC	0.65	0	0	1	1	1
Published, but not in top 5 journal	0.39	0	0	0	1	1
Published, in top 5 journal	0.10	0	0	0	0	0
B. Elements of the pre-registration						
No. Interventions (incl. control)	3.77	2	2	3	4	7
No. Arms	4.79	2	2	3	5	9.4
No. Main Outcomes	5.36	1	2	4	6	10.4
No. Hypotheses (Total) per Study	15.95	2	4	9	18	32
No. Main Hypotheses	11.77	2	4	8	15	25.4
No. Primary Heterogeneity Hypotheses	4.18	0	0	0	0	4
Has more than 10 Hypotheses (indicator)	0.43	0	0	0	1	1
C. Average share of hypotheses per registration that are:						
Specified in detail	0.31	0	0	0	1	1
Type of hypotheses: Simple Difference	0.92	0.67	1	1	1	1
Type of hypotheses: Double Difference	0.03	0	0	0	0	0
Type of hypotheses: 'Other'	0.05	0	0	0	0	0

Note: Averages over the full sample (317 studies). PAP: pre-analysis plan. LMIC: low- and middle-income countries (World Bank definition). 'Published' refers to a final output, not a working paper (e.g. journal article, book chapter,...). The top 5 journals are AER, QJE, Econometrica, JPE and ReStud. 'Specified in detail' refers to whether a hypothesis is specific enough to be encoded in a precise way, without the need for a judgment call on the part of the encoder. A hypothesis is classified as 'not detailed' if any of the following apply: (i) the test is not clearly specified (e.g., interventions are described but comparison arms are absent); (ii) no regression specification is provided (unless the test is otherwise obvious); (iii) the outcome is ambiguous; or (iv) the experimental design is not clearly described. Hypotheses that are not simple or double differences (hence in the 'other' category) are mostly estimated model parameters of various sorts.

Table 2: Availability of Results and Share of Null Results by Hypothesis Type, Definitions of Availability, and Study Characteristics

	Average share		
Category	Available	Null | Available	*N*
A. Hypotheses			
All hypotheses	0.420	0.645	317
Main hypotheses	0.447	0.644	317
Heterogeneity hypotheses	0.223	0.642	39
B. Definitions of availability			
In main body	0.348	0.611	317
No modifications	0.286	0.660	317
In main body and no modifications	0.240	0.625	317
C. Study characteristics			
Registered with a PAP	0.439	0.673	127
Conducted in LMIC	0.378	0.643	206
Published, but not in top 5 journal	0.553	0.653	124
Published, in top 5 journal	0.413	0.513	31

Note: All values are at baseline. 'Available' is the fraction of pre-registered hypotheses with a numerical result in any publicly accessible source; 'Null | Available' is the fraction of available results with $p > 0.05$. Each value is a mean over studies of the study's share; 'Null | Available' is averaged over studies with at least one available result. A: Main hypotheses excludes the heterogeneity hypotheses; Heterogeneity hypotheses covers the heterogeneity hypotheses of the studies that have them. B: In main body excludes results found only in an appendix; No modifications requires the hypothesis, its outcomes and its arms to match the pre-registration exactly; the last row applies both restrictions. C: PAP: pre-analysis plan. LMIC: low- and middle-income countries (World Bank definition). 'Published' refers to a final output, not a working paper (e.g. journal article, book chapter,...). Top 5 journals are AER, QJE, Econometrica, JPE and ReStud.

Table 3: Main Treatment Effects

	Available (DD)		Available (XS)		Explained		Missing	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Awareness	0.004	0.004	0.004	0.004	0.052	0.052	-0.056	-0.057
	(0.003)	(0.003)	(0.007)	(0.007)	(0.023)	(0.023)	(0.025)	(0.026)
Engagement	0.069		0.066		0.123		-0.190	
	(0.022)		(0.020)		(0.033)		(0.037)	
Resources	0.055		0.056		0.083		-0.139	
	(0.021)		(0.022)		(0.028)		(0.035)	
Pooled (Eng. or Res.)		0.062		0.061		0.103		-0.164
		(0.015)		(0.015)		(0.024)		(0.028)
Available (baseline)			0.930	0.929	-0.156	-0.158	-0.774	-0.771
			(0.022)	(0.021)	(0.034)	(0.034)	(0.038)	(0.038)
N	634	634	317	317	317	317	317	317
Study fixed effects	X	X						

Note: Available is the fraction of pre-registered hypotheses for which a numerical result is found in any publicly accessible source. Explained is the fraction of hypotheses whose results are missing for a reason outside the authors' control (failed or no intervention, or intervention without data), as given by the authors; this outcome is only observed post-intervention. Missing is the fraction of pre-registered hypotheses that are neither available nor explained. Columns (1)-(2) report Available estimated with the pre-registered difference-in-differences specification; columns (3)-(8) use the cross-sectional specification, under which Available + Explained + Fraction Missing = 1 holds exactly across coefficients. Control group means (post-intervention): 0.36 (Available), 0.00 (Explained), 0.64 (Fraction Missing). Columns (1)–(2) use the pre-registered difference-in-differences specification; columns (3)–(8) include, but do not report, indicators for whether the study has a PAP, whether the study was undertaken in an LMIC, and for each encoder on the team, and the baseline share Available. The Explained and Missing outcomes were not pre-registered. Robust standard errors in parentheses.

Table 4: Treatment Effects on the Share of Null Results

	Share null results			
	Diff-in-Diff		Cross section	
	(1)	(2)	(3)	(4)
Awareness	0.003	0.003	0.000	0.000
	(0.002)	(0.002)	(0.005)	(0.005)
Engagement	0.054		0.052	
	(0.019)		(0.018)	
Resources	0.029		0.029	
	(0.014)		(0.015)	
Pooled (Eng. or Res.)		0.042		0.041
		(0.012)		(0.012)
Share null (baseline)			0.954	0.953
			(0.019)	(0.020)
N	634	634	317	317
Study fixed effects	X	X		

Note: Null is the fraction of pre-registered hypotheses per study with an available result that is null ($p > 0.05$). Columns (1)-(2) use the pre-registered difference-in-differences specification; columns (3)-(4) use the cross-sectional specification with the pre-specified controls (indicators for whether the study has a PAP, whether the study was undertaken in an LMIC, and for each encoder on the team) and the baseline share null. Columns (2) and (4) pool the Engagement and Resources arms. Control group mean (post-intervention): 0.25. Robust standard errors in parentheses.

Online Appendix

Beyond Publication Bias:

Characterizing and Understanding Missing Results in Economics

Contents

A Supplementary Analysis **S-3**

B Relating the Paper to our Pre-Analysis Plan **S-16**
B.1 Outcomes S-16
B.2 Analysis S-16
B.3 Tests S-17

C Details on Encoding Procedure **S-24**
C.1 Encoding Protocol S-24

D An example Results Report: The Current Study **S-27**

E Materials Sent to the Authors of the Studies in the RCT **S-34**

F Expert study on The Social Science Prediction Platform **S-54**

List of Tables

A.1 Symmetry of Forecasted Differences by Study Characteristic S-3
A.2 Engagement following the intervention email S-4
A.3 Explanations Provided for Why Results are Not Available S-5
A.4 Main Treatment Effects, Alternative Definition of Available at Baseline S-6
A.5 Treatment Effects on the Share of Null Results, Alternative Definition of Available at Baseline S-7
A.6 Treatment Effects on the Share of Null Results, Pooled over Choices of Main-Hypothesis Expansions S-8
A.7 Null Results Among Results Available at Baseline and Follow-up S-9
A.8 Heterogeneity of Treatment Effects by PAP S-10
A.9 Heterogeneity of Treatment Effects by LMIC S-11
A.10 Spillover Effects, Restricted Model S-12
A.11 Spillover Effects, Unrestricted Model S-13
B.1 Completely and Partially Available Results S-18

B.2 Main Treatment Effects, Main Hypotheses Only S-19
B.3 Treatment Effects on the Share of Null Results, Main Hypotheses Only S-20
B.4 Treatment Effects in the Pre-Registered Cross-Sectional Specification S-21
B.5 Tests of the Pre-Registered Main Hypotheses S-22
B.6 Multiple Hypothesis Testing for the Primary Hypotheses S-23

List of Figures

A.1 Timeline of RCT Study by Month (July 2023 – February 2026) S-15
D.1 *This Project's* Results Report, Page 1 of 6 (H1, Available) S-28
D.2 *This Project's* Results Report, Page 2 of 6 (H4, Available) S-29
D.3 *This Project's* Results Report, Page 3 of 6 (H5, Available) S-30
D.4 *This Project's* Results Report, Page 4 of 6 (H1, Null) S-31
D.5 *This Project's* Results Report, Page 5 of 6 (H4, Null) S-32
D.6 *This Project's* Results Report, Page 6 of 6 (H5, Null) S-33
E.1 *Awareness* Treatment Sample Email S-34
E.2 *Engagement* Treatment Sample Email S-35
E.3 *Resources* Treatment Sample Email S-36
E.4 *Awareness* Treatment's Study Consent Form S-37
E.5 *Engagement* Treatment's Study Consent Form, Page 1 of 2 . . . . . S-38
E.6 *Engagement* Treatment's Study Consent Form, Page 2 of 2 . . . . . S-39
E.7 *Resources* Treatment's Study Consent Form, Page 1 of 2 . . . . . S-40
E.8 *Resources* Treatment's Study Consent Form, Page 2 of 2 . . . . . S-41
E.9 *Empty* Results Report Sample Cover S-42
E.10 *Empty* Results Report Sample Page S-43
E.11 *Pre-Filled* Results Report Sample Cover S-44
E.12 *Found* Results Report Sample Page S-45
E.13 *Not Found* Results Report Sample Page S-46
E.14 Extra Heterogeneity Rows Sample Page S-47
E.15 *Not Found* Results Report Sample Page (as Filled by Authors) S-48
E.16 *Found* Results Report Explainer S-49
E.17 *Not Found* Results Report Explainer S-50
E.18 Encoded Pre-Registration (G_0) Sample, Page 1 of 2 . . . . . S-51
E.19 AEA RCT Registry Post Section Update Explainer S-53
F.1 Survey instrument (page 1/10) S-54
F.2 Survey instrument (page 3/10, page 2/10 is blank) S-55
F.3 Survey instrument (page 4/10) S-56
F.4 Survey instrument (page 5/10) S-57
F.5 Survey instrument (page 6/10) S-58
F.6 Survey instrument (page 7/10) S-59
F.7 Survey instrument (page 8/10) S-60
F.8 Survey instrument (page 9/10) S-61
F.9 Survey instrument (page 10/10) S-62

A Supplementary Analysis

Table A.1: Symmetry of Forecasted Differences by Study Characteristic

	Available				Null | Available			
Characteristic	Less	Same	More	*p*-value	Less	Same	More	*p*-value
Has PAP	10	11	47	<0.001	12	9	47	<0.001
LMIC	11	38	19	0.200	13	36	19	0.377
Published	25	19	24	1.000	37	12	19	0.022
Top 5 journal	17	31	20	0.743	37	20	11	<0.001

Note: Numbers of forecasters expecting a lower (Less), the same (Same) or a higher (More) outcome for studies with the characteristic than for studies without it (panels C and D of Figure 4). The *p*-value is from an exact two-sided binomial test that forecasters expecting a difference are equally likely to expect a lower as a higher outcome (Less and More 50/50; Same left out).

Table A.2: Engagement following the intervention email

		Clicked on			Responded with email	
	(1)	(2)	(3)	(4)	(5)	(6)
Arm	*N*	Any	Results	Details.pdf	Yes	Informative
Awareness	90	0.844	0.822	–	0.511	0.089
Engagement	88	0.920	0.920	0.523	0.761	0.307
Resources	88	0.909	0.909	0.477	0.693	0.193
Total	266	0.891	0.883	0.500	0.654	0.195

Note: Columns (2)-(6) are shares of the studies in the arm. Column (2) reports whether at least one author on the study clicked any link in the intervention email. Column (3) reports the click rate for the Results Report link specifically. Column (4) reports the click rate for the encoding details PDF, which was only sent in the Engagement and Resources arms; its Total is over those two arms. Columns (5)-(6) classify email responses: "Yes" includes any reply; "Informative" (6) includes the studies that replied and either have new results found in the follow-up or have an explanation for non-available results (the intervention failed or was not implemented, or no data were collected). Fourteen (14) studies expressed interest in RA support through the Resources arm (2 via external funds and 12 via Berkeley-based RA time); in practice, six took up the Berkeley RA support and none followed up for the funds transfer.

Table A.3: Explanations Provided for Why Results are Not Available

	Share of all	
Explanation type	Hypotheses	Studies
A. No data produced		
Intervention failed or not implemented	0.024	0.047
Intervention ran, no data collected	0.040	0.050
B. Data collected, not reported		
Intervention ran, data issues	0.008	0.009
Data available, analysis pending	0.026	0.044
Data available, underpowered	0.025	0.016
Disagree with authors	0.006	0.009
Any explanation	0.129	0.151

Note: Explanations given by the authors for why pre-registered results are not available after the intervention, as classified by the research team. Hypotheses: share of all 5,057 pre-registered hypotheses with a missing result and an explanation of that type. Studies: share of all 317 studies with at least one such explanation; studies can give explanations of several types. An additional 8 hypotheses (3 studies) received the explanation 'not pre-registered' and are excluded from all rows.

Table A.4: Main Treatment Effects, Alternative Definition of Available at Baseline

	Available (DD)		Available (XS)		Explained		Missing	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Awareness	0.003	0.003	0.005	0.005	0.052	0.052	-0.057	-0.057
	(0.002)	(0.002)	(0.005)	(0.005)	(0.023)	(0.023)	(0.025)	(0.025)
Engagement	0.042		0.043		0.127		-0.170	
	(0.016)		(0.016)		(0.033)		(0.036)	
Resources	0.037		0.040		0.085		-0.126	
	(0.018)		(0.020)		(0.029)		(0.034)	
Pooled (Eng. or Res.)		0.039		0.041		0.107		-0.148
		(0.012)		(0.013)		(0.024)		(0.027)
Available (baseline)			0.953	0.953	-0.160	-0.162	-0.793	-0.791
			(0.019)	(0.019)	(0.034)	(0.035)	(0.038)	(0.038)
N	634	634	317	317	317	317	317	317
Study fixed effects	X	X						

Note: In this table, results that were not found in the baseline (but pointed to by authors in the follow-up) are counted as available at baseline. Available is the fraction of pre-registered hypotheses for which a numerical result is found in any publicly accessible source. Explained is the fraction of hypotheses whose results are missing for a reason outside the authors' control (failed or no intervention, or intervention without data), as given by the authors; this outcome is only observed post-intervention. Missing is the fraction of pre-registered hypotheses that are neither available nor explained. Columns (1)-(2) report Available estimated with the pre-registered difference-in-differences specification; columns (3)-(8) use the cross-sectional specification, under which Available + Explained + Fraction Missing = 1 holds exactly across coefficients. Control group means (post-intervention): 0.36 (Available), 0.00 (Explained), 0.64 (Fraction Missing). Columns (1)–(2) use the pre-registered difference-in-differences specification; columns (3)–(8) include, but do not report, indicators for whether the study has a PAP, whether the study was undertaken in an LMIC, and for each encoder on the team, and the baseline share Available. The Explained and Missing outcomes were not pre-registered. Robust standard errors in parentheses.

Table A.5: Treatment Effects on the Share of Null Results, Alternative Definition of Available at Baseline

	Share null results			
	Diff-in-Diff		Cross section	
	(1)	(2)	(3)	(4)
Awareness	0.002	0.002	0.002	0.002
	(0.002)	(0.002)	(0.003)	(0.003)
Engagement	0.029		0.029	
	(0.013)		(0.014)	
Resources	0.019		0.020	
	(0.012)		(0.013)	
Pooled (Eng. or Res.)		0.024		0.025
		(0.009)		(0.010)
Share null (baseline)			0.972	0.972
			(0.017)	(0.017)
N	634	634	317	317
Study fixed effects	X	X		

Note: In this table, results that were not found in the baseline (but pointed to by authors in the follow-up) are counted as available at baseline. Null is the fraction of pre-registered hypotheses per study with an available result that is null ($p > 0.05$). Columns (1)-(2) use the pre-registered difference-in-differences specification; columns (3)-(4) use the cross-sectional specification with the pre-specified controls (indicators for whether the study has a PAP, whether the study was undertaken in an LMIC, and for each encoder on the team) and the baseline share null. Columns (2) and (4) pool the Engagement and Resources arms. Control group mean (post-intervention): 0.25. Robust standard errors in parentheses.

Table A.6: Treatment Effects on the Share of Null Results, Pooled over Choices of Main-Hypothesis Expansions

	Share null results			
	Diff-in-Diff		Cross section	
	(1)	(2)	(3)	(4)
Awareness	0.003	0.003	0.000	0.000
	(0.002)	(0.002)	(0.005)	(0.005)
Engagement	0.055		0.053	
	(0.020)		(0.019)	
Resources	0.029		0.030	
	(0.014)		(0.015)	
Pooled (Eng. or Res.)		0.042		0.041
		(0.012)		(0.012)
Share null (baseline)			0.953	0.952
			(0.020)	(0.020)
N	634	634	317	317
Study fixed effects	X	X		

Note: When encoders encountered more than one clear candidate for a result, all were recorded (expansions) but only one was randomly selected for the study. In such cases all expansions were *available*, but they could be either null or non-null. In this table, the estimates are pooled over 100 copies of the sample in which the expansion representing each main hypothesis with several expansions is drawn at random (the same in both periods), using Rubin's rules: coefficients are averaged over the copies, and standard errors combine the average robust variance and the variance across copies (Rubin, 1987). Null is the fraction of pre-registered hypotheses per study with an available result that is null ($p > 0.05$). Columns (1)-(2) use the pre-registered difference-in-differences specification; columns (3)-(4) use the cross-sectional specification with the pre-specified controls (indicators for whether the study has a PAP, whether the study was undertaken in an LMIC, and for each encoder on the team) and the baseline share null. Columns (2) and (4) pool the Engagement and Resources arms. Control group mean (post-intervention): 0.25. Robust standard errors in parentheses.

Table A.7: Null Results Among Results Available at Baseline and Follow-up

	Original classification		Alternative classification	
	(1)	(2)	(3)	(4)
	Baseline	Only in Follow-up	Baseline	Only in Follow-up
Non-null	497	46	514	29
Null	1,103	175	1,210	68
Share null	0.689	0.792	0.702	0.701
p-value, equal share null		0.038		0.990

Note: Counts of pre-registered hypotheses (analysis sample) with an available result, by whether the result is null ($p > 0.05$). Baseline: available at baseline. Only in Follow-up: available after the intervention but not at baseline. Columns (3)-(4) also count as available at baseline the results that were not found in the baseline search but that the authors pointed to in the follow-up as previously available. The p-value tests whether the share null differs between the baseline and follow-up results, from a regression of a null indicator on an indicator for the result becoming available in the follow-up, with standard errors clustered by study.

Table A.8: Heterogeneity of Treatment Effects by PAP

	Treatment effects			
	Available	Explained	Missing	Null
	(1)	(2)	(3)	(4)
A. Studies without a PAP				
Awareness	0.003	0.056	-0.060	0.003
	(0.003)	(0.032)	(0.035)	(0.003)
Engagement	0.066	0.105	-0.171	0.059
	(0.028)	(0.040)	(0.046)	(0.027)
Resources	0.006	0.099	-0.107	0.009
	(0.005)	(0.040)	(0.043)	(0.006)
B. Studies with a PAP				
Awareness	0.005	0.045	-0.053	0.003
	(0.004)	(0.030)	(0.034)	(0.004)
Engagement	0.073	0.149	-0.219	0.046
	(0.034)	(0.054)	(0.059)	(0.027)
Resources	0.124	0.059	-0.188	0.059
	(0.048)	(0.034)	(0.056)	(0.032)
C. Tests, p-values for equality by characteristic				
Awareness	0.765	0.799	0.896	0.961
Engagement	0.874	0.501	0.515	0.732
Resources	0.015	0.440	0.237	0.124
D. Joint tests, p-values for equality by characteristic				
All arms	0.112	0.750	–	0.478
All arms, outcomes (1), (2) and (4)		0.705		
N	634	317	317	634
Study fixed effects	X			X

Note: Treatment effects in studies where the characteristic is absent (A) and present (B), from regressions in which the treatment arms are interacted with an indicator for whether the study has a PAP, one regression per outcome. Columns (1) and (4) use the difference-in-differences specification with study fixed effects; columns (2) and (3) the cross-sectional specification with the pre-specified controls (indicators for whether the study has a PAP, whether it was undertaken in an LMIC, and for each encoder) and baseline Available. Robust standard errors in parentheses. Panel C gives the *p*-values of two-sided tests that each arm's effect is the same in A and B. Panel D gives the *p*-values of Wald tests that all three arms' effects are the same in A and B: by outcome, with the regression's robust standard errors, and jointly for Available, Explained and Null, with the three regressions stacked as seemingly unrelated regressions and standard errors clustered by study. Missing, a linear combination of the other outcomes, is not tested separately (dash).

Table A.9: Heterogeneity of Treatment Effects by LMIC

	Treatment effects			
	Available	Explained	Missing	Null
	(1)	(2)	(3)	(4)
A. Studies not in an LMIC				
Awareness	0.000	-0.010	0.020	0.000
	–	(0.022)	(0.031)	–
Engagement	0.078	0.097	-0.165	0.053
	(0.032)	(0.046)	(0.052)	(0.026)
Resources	0.075	0.078	-0.147	0.060
	(0.043)	(0.047)	(0.060)	(0.036)
B. Studies in an LMIC				
Awareness	0.006	0.084	-0.096	0.005
	(0.004)	(0.034)	(0.035)	(0.004)
Engagement	0.064	0.137	-0.202	0.054
	(0.029)	(0.044)	(0.050)	(0.026)
Resources	0.044	0.085	-0.135	0.014
	(0.022)	(0.034)	(0.041)	(0.009)
C. Tests, p-values for equality by characteristic				
Awareness	0.145	0.021	0.016	0.183
Engagement	0.737	0.535	0.604	0.980
Resources	0.533	0.890	0.866	0.211
D. Joint tests, p-values for equality by characteristic				
All arms	0.454	0.142	–	0.343
All arms, outcomes (1), (2) and (4)		0.656		
N	634	317	317	634
Study fixed effects	X			X

Note: Treatment effects in studies where the characteristic is absent (A) and present (B), from regressions in which the treatment arms are interacted with an indicator for whether the study is in an LMIC, one regression per outcome. Columns (1) and (4) use the difference-in-differences specification with study fixed effects; columns (2) and (3) the cross-sectional specification with the pre-specified controls (indicators for whether the study has a PAP, whether it was undertaken in an LMIC, and for each encoder) and baseline Available. Robust standard errors in parentheses. Panel C gives the *p*-values of two-sided tests that each arm's effect is the same in A and B. Panel D gives the *p*-values of Wald tests that all three arms' effects are the same in A and B: by outcome, with the regression's robust standard errors, and jointly for Available, Explained and Null, with the three regressions stacked as seemingly unrelated regressions and standard errors clustered by study. Missing, a linear combination of the other outcomes, is not tested separately (dash). A dash marks a standard error of zero: none of these studies' outcomes changed between the periods.

Table A.10: Spillover Effects, Restricted Model

	Available	Explained	Missing	Null
	(1)	(2)	(3)	(4)
Awareness	0.001	0.050	-0.045	0.002
	(0.004)	(0.027)	(0.030)	(0.003)
Engagement	0.069	0.119	-0.179	0.053
	(0.029)	(0.041)	(0.046)	(0.026)
Resources	0.069	0.062	-0.126	0.036
	(0.028)	(0.031)	(0.039)	(0.018)
Exposure (β_0)	–	0.050	-0.075	–
		(0.034)	(0.044)	
Exposure × treated (θ)	0.009	-0.041	0.053	0.008
	(0.009)	(0.047)	(0.055)	(0.008)
Network size	–	-0.056	0.087	–
		(0.028)	(0.035)	
Network size × Awareness	-0.004	0.042	-0.066	-0.005
	(0.006)	(0.036)	(0.042)	(0.006)
Network size × Engagement	-0.007	0.042	-0.061	-0.006
	(0.019)	(0.046)	(0.054)	(0.018)
Network size × Resources	-0.026	0.063	-0.065	-0.015
	(0.013)	(0.042)	(0.048)	(0.010)
Available (baseline)		-0.159	-0.768	
		(0.035)	(0.039)	
N	634	317	317	634
Study fixed effects	X			X

Note: Restricted spillover model (Equation 4). Exposure is the number of co-author links from a study to studies in the Awareness, Engagement or Resources arms, network size the number of links to studies in any arm, including control; both enter interacted with the post period. β_0 is the spillover effect of exposure, θ its difference for treated studies, so the spillover effect on a treated study is $\beta_0 + \theta$. Columns (1) and (4) use the difference-in-differences specification with study fixed effects; there, β_0 and the network-size main effect are identified only from control studies, whose outcomes do not change between the periods, and are not reported (–). Columns (2)-(3) use the cross-sectional specification with the pre-specified controls (indicators for whether the study has a PAP, whether it was undertaken in an LMIC, and for each encoder) and baseline Available. Standard errors clustered by study in parentheses.

Table A.11: Spillover Effects, Unrestricted Model

	Available	Explained	Missing	Null
	(1)	(2)	(3)	(4)
Treatment effects				
Awareness	0.001	0.054	-0.048	0.002
	(0.004)	(0.029)	(0.032)	(0.003)
Engagement	0.068	0.125	-0.182	0.053
	(0.029)	(0.042)	(0.047)	(0.026)
Resources	0.070	0.066	-0.129	0.036
	(0.029)	(0.031)	(0.040)	(0.019)
Spillover effects				
Exposure to Awareness	–	0.053	-0.064	–
		(0.043)	(0.057)	
Exposure to Engagement	–	0.056	-0.078	–
		(0.040)	(0.055)	
Exposure to Resources	–	0.049	-0.068	–
		(0.033)	(0.044)	
Awareness × exposure to Awareness	0.005	-0.133	0.145	0.003
	(0.007)	(0.073)	(0.085)	(0.006)
Awareness × exposure to Engagement	0.007	-0.080	0.097	0.003
	(0.008)	(0.075)	(0.086)	(0.005)
Awareness × exposure to Resources	0.003	-0.081	0.103	0.001
	(0.004)	(0.055)	(0.063)	(0.003)
Engagement × exposure to Awareness	0.100	0.043	-0.153	0.110
	(0.070)	(0.107)	(0.122)	(0.068)
Engagement × exposure to Engagement	0.052	-0.141	0.099	0.030
	(0.042)	(0.085)	(0.101)	(0.027)
Engagement × exposure to Resources	0.017	0.034	-0.041	0.028
	(0.022)	(0.103)	(0.112)	(0.018)
Resources × exposure to Awareness	-0.015	-0.014	0.026	-0.006
	(0.010)	(0.069)	(0.081)	(0.008)
Resources × exposure to Engagement	-0.012	-0.066	0.095	-0.007
	(0.009)	(0.047)	(0.064)	(0.007)

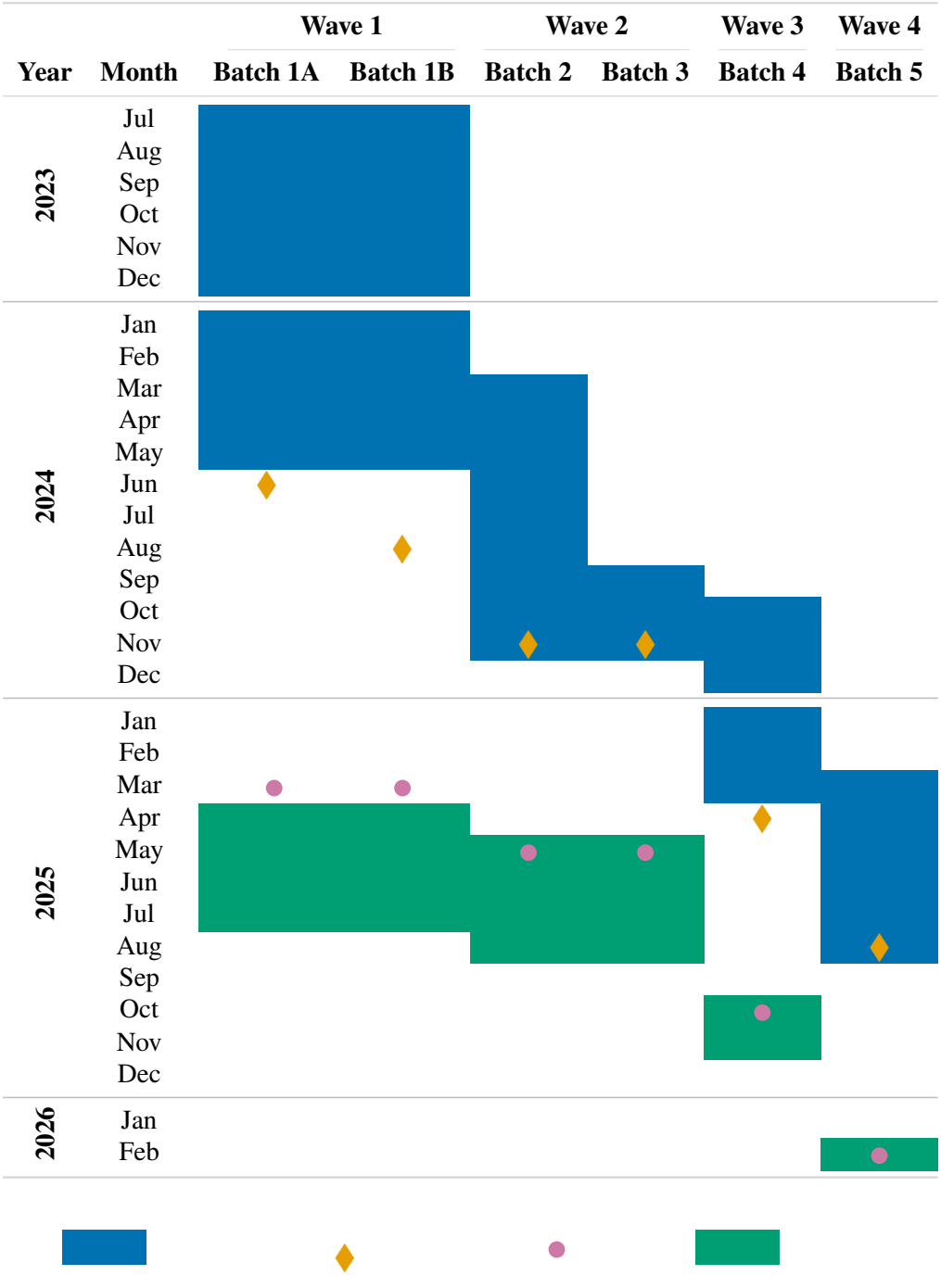
Continued on next page

Table A.11: Spillover Effects, Unrestricted Model (Continued)

	Available	Explained	Missing	Null
	(1)	(2)	(3)	(4)
Resources × exposure to Resources	-0.004	0.017	-0.001	-0.008
	(0.016)	(0.088)	(0.099)	(0.013)
Network size				
Network size	–	-0.059	0.082	–
		(0.031)	(0.041)	
Network size × Awareness	-0.001	0.083	-0.109	0.000
	(0.003)	(0.048)	(0.055)	(0.003)
Network size × Engagement	-0.045	0.022	0.013	-0.045
	(0.020)	(0.075)	(0.081)	(0.017)
Network size × Resources	-0.010	0.039	-0.044	-0.002
	(0.014)	(0.039)	(0.053)	(0.011)
Baseline control				
Available (baseline)		-0.160	-0.768	
		(0.036)	(0.040)	
N	634	317	317	634
Study fixed effects	X			X
Wald test, all spillover effects = 0: χ^2	6.37	14.67	13.44	5.58
Degrees of freedom	9	12	12	9
p-value	0.703	0.260	0.338	0.781

Note: Unrestricted spillover model. Exposure to an arm is the number of co-author links from a study to studies in that arm, entered interacted with the post period. 'Exposure to' rows give the spillover effect on control studies; '× exposure to' rows its difference for studies in the given arm. Network size (the number of links to studies in any arm, including control) enters as in the restricted model. Columns (1) and (4) use the difference-in-differences specification with study fixed effects; there, the effects on control studies and the network-size main effect are identified only from control studies, whose outcomes do not change between the periods, and are not reported (–). Columns (2)-(3) use the cross-sectional specification with the pre-specified controls (indicators for whether the study has a PAP, whether it was undertaken in an LMIC, and for each encoder) and baseline Available. The Wald test is of the hypothesis that all identified spillover effects are zero. Standard errors clustered by study in parentheses.

Figure A.1: Timeline of RCT Study by Month (July 2023 – February 2026)



Note: This figure shows the timeline of baseline encoding, intervention rollout, and follow-up encoding by study batch. Intervention rollout dates by wave: Wave 1 on June 27, 2024 (Batch 1A, N = 12) and August 14, 2024 (Batch 1B, N = 128); Wave 2 on November 19, 2024 (Batches 2 and 3, N = 88); Wave 3 on April 2, 2025 (Batch 4, N = 77); Wave 4 on August 5, 2025 (Batch 5, N = 12). After interventions were first launched, we allowed for one or two reminders if there was no engagement (the second to co-authors of the main PI). Wave sizes are analysis-sample studies per batch (sum = 317; the 35 archived control studies are excluded). When multiple events occur in the same month for a batch, the milestones (symbols) are overlaid on the encoding periods.

B Relating the Paper to our Pre-Analysis Plan

Our study has two registered pre-analysis plans (PAPs), available at `https://www.socialscienceregistry.org/trials/13755`. The first covers the main analysis, while the second, registered later, covers potential spillover effects. In this section, we explain how the analyses reported in the paper relate to these plans.

B.1 Outcomes

The outcomes *Available* and *Null* in the paper follow exactly the definitions of the outcomes *Completely Available* (Y_{1i}) and *Reports Null Result* (Y_{3i}) in the PAP. We also examine the proportion of null results conditional on availability, but because of the selection problem with the conditioning, we do not attempt to make causal statements about this outcome.

Following the unexpectedly large amount of feedback received from authors, we introduced the outcome *Explained*, which captures written explanations for why results could not be calculated. Our preferred definition of this outcome is reported in Panel A of Table A.3. The outcome Missing in the main analysis is defined as one minus the sum of *Available* and *Explained*.

We pre-registered a secondary outcome, *At Least Partially Available* (Y_{2i}), because we anticipated receiving responses that would be difficult to classify precisely. In practice, such responses accounted for only a small share of the results collected, and we therefore decided not to report this outcome alongside the primary Available outcome (see Table B.1). We also pre-registered exploratory analyses of barely significant results, p-values, and standardized effect sizes. However, these measures could only be calculated precisely for a selected subset of our sample, and we do not include these analyses in the paper.

The PAP specified broad and narrow definitions of the outcomes. The broad definition includes both main and heterogeneity (subgroup) hypotheses, while the narrow definition includes only main hypotheses. The broad definition was pre-specified as primary. Since treatment effects are very similar under the two definitions, we focus on the broad definition in the paper. Tables B.2 and B.3 report the main regressions using the narrow definition.

B.2 Analysis

The difference-in-differences regressions in column (1) of Tables 3 and 4 correspond to the main pre-registered analyses. The PAP also specified that the Engagement and Resources arms would be pooled if take-up of the resource offer was low. The corresponding pooled estimates are reported in column (2) of both tables.

We also pre-registered cross-sectional regressions using only post-intervention outcomes, without controlling for baseline outcomes. These specifications are reported in Table B.4. The estimated treatment effects are larger, but the standard errors are also substantially larger. The difference-in-differences and ANCOVA specifications used in the main paper take advantage of the strong persistence in outcomes between baseline and follow-up, particularly in the control group. We therefore focus on these specifications in the main paper.

The heterogeneity analyses reported in Figure 6 and Tables A.8 and A.9 correspond to the main pre-registered dimensions of heterogeneity.

B.3 Tests

While the main paper discusses several pre-registered tests, it does not provide a unified presentation of the registered hypotheses. Table B.5 presents the main hypotheses and corresponding *p*-values for each of the pre-registered tests for the two primary outcomes (PAP Section 3.2.2). Table B.6 reports the multiple-testing-adjusted *p*-values and the joint test pre-registered in Section 3.2.3 of the PAP.

Table B.1: Completely and Partially Available Results

	Baseline		Follow-up	
	Hypotheses	Share	Hypotheses	Share
Completely available (Y_1)	1,600	0.316	1,818	0.360
Partially, not completely available (Y_2, not Y_1)	117	0.023	132	0.026
Not available	3,340	0.660	3,107	0.614
Total	5,057	1.000	5,057	1.000

Note: Pre-registered hypotheses in the analysis sample by availability of their result. Completely available (Y_1): a numerical estimate is found, the main outcome of the paper. At least partially available (Y_2, the secondary outcome of the pre-analysis plan): completely available, or at least a qualitative record of the result's statistical significance is found, e.g. the authors state that the result is null without reporting it. Unlike Y_1, Y_2 does not require a measure of uncertainty (a *p*-value, standard error or significance level). The second row counts the hypotheses that are at least partially but not completely available. Baseline: before the intervention. Follow-up: after the intervention.

Table B.2: Main Treatment Effects, Main Hypotheses Only

	Available (DD)		Available (XS)		Explained		Missing	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Awareness	0.003	0.003	0.004	0.004	0.053	0.053	-0.057	-0.057
	(0.002)	(0.002)	(0.007)	(0.007)	(0.023)	(0.023)	(0.026)	(0.026)
Engagement	0.064		0.062		0.125		-0.187	
	(0.020)		(0.019)		(0.033)		(0.036)	
Resources	0.056		0.057		0.085		-0.143	
	(0.021)		(0.022)		(0.029)		(0.035)	
Pooled (Eng. or Res.)		0.060		0.060		0.106		-0.165
		(0.014)		(0.015)		(0.024)		(0.028)
Available (baseline)			0.931	0.931	-0.159	-0.161	-0.772	-0.770
			(0.021)	(0.021)	(0.034)	(0.034)	(0.038)	(0.037)
N	634	634	317	317	317	317	317	317
Study fixed effects	X	X						

Note: In this table, all outcomes are computed over the main hypotheses only, leaving out the heterogeneity hypotheses. Available is the fraction of pre-registered hypotheses for which a numerical result is found in any publicly accessible source. Explained is the fraction of hypotheses whose results are missing for a reason outside the authors' control (failed or no intervention, or intervention without data), as given by the authors; this outcome is only observed post-intervention. Missing is the fraction of pre-registered hypotheses that are neither available nor explained. Columns (1)-(2) report Available estimated with the pre-registered difference-in-differences specification; columns (3)-(8) use the cross-sectional specification, under which Available + Explained + Fraction Missing = 1 holds exactly across coefficients. Control group means (post-intervention): 0.39 (Available), 0.00 (Explained), 0.61 (Fraction Missing). Columns (1)–(2) use the pre-registered difference-in-differences specification; columns (3)–(8) include, but do not report, indicators for whether the study has a PAP, whether the study was undertaken in an LMIC, and for each encoder on the team, and the baseline share Available. The Explained and Missing outcomes were not pre-registered. Robust standard errors in parentheses.

Table B.3: Treatment Effects on the Share of Null Results, Main Hypotheses Only

	Share null results			
	Diff-in-Diff		Cross section	
	(1)	(2)	(3)	(4)
Awareness	0.002	0.002	-0.001	-0.001
	(0.002)	(0.002)	(0.005)	(0.005)
Engagement	0.050		0.048	
	(0.018)		(0.017)	
Resources	0.032		0.032	
	(0.014)		(0.015)	
Pooled (Eng. or Res.)		0.041		0.040
		(0.011)		(0.011)
Share null (baseline)			0.959	0.958
			(0.019)	(0.019)
N	634	634	317	317
Study fixed effects	X	X		

Note: In this table, all outcomes are computed over the main hypotheses only, leaving out the heterogeneity hypotheses. Null is the fraction of pre-registered hypotheses per study with an available result that is null ($p > 0.05$). Columns (1)-(2) use the pre-registered difference-in-differences specification; columns (3)-(4) use the cross-sectional specification with the pre-specified controls (indicators for whether the study has a PAP, whether the study was undertaken in an LMIC, and for each encoder on the team) and the baseline share null. Columns (2) and (4) pool the Engagement and Resources arms. Control group mean (post-intervention): 0.26. Robust standard errors in parentheses.

Table B.4: Treatment Effects in the Pre-Registered Cross-Sectional Specification

	Cross section, post-intervention			
	Available		Null	
	(1)	(2)	(3)	(4)
Awareness	0.055	0.055	-0.012	-0.012
	(0.066)	(0.066)	(0.053)	(0.053)
Engagement	0.095		0.084	
	(0.066)		(0.056)	
Resources	0.136		0.093	
	(0.067)		(0.056)	
Pooled (Eng. or Res.)		0.115		0.089
		(0.059)		(0.050)
N	317	317	317	317

Note: Cross-sectional specification as pre-registered: the post-intervention outcome regressed on indicators for the treatment arms and the pre-specified controls (indicators for whether the study has a PAP, whether the study was undertaken in an LMIC, and for each encoder on the team), without a control for the outcome at baseline. Available is the fraction of pre-registered hypotheses for which a numerical result is found in any publicly accessible source; Null is the fraction of pre-registered hypotheses with an available result that is null ($p > 0.05$). Columns (2) and (4) pool the Engagement and Resources arms. Control group means (post-intervention): 0.36 (Available), 0.25 (Null). Robust standard errors in parentheses.

Table B.5: Tests of the Pre-Registered Main Hypotheses

PAP	Hypothesis (pre-analysis plan, Section 3.2.2)	*p*-value	
		Available	Null
H1	There is no effect of the informational treatment together with the provision of an empty results report template (T_1): $H_0 : \tau_1 = 0$, $H_A : \tau_1 > 0$.	0.073	0.091
H2	There is no effect of the informational treatment together with the provision of the pre-filled results report (T_2): $H_0 : \tau_2 = 0$, $H_A : \tau_2 > 0$.	0.001	0.002
H3	There is no effect of the informational treatment with the provision of the pre-filled results report and the offer of 30 hours of research assistant (RA) support (T_3): $H_0 : \tau_3 = 0$, $H_A : \tau_3 > 0$.	0.004	0.015
H4	There is no incremental effect of provision of the pre-filled results report (T_2) relative to an empty results report template (T_1): $H_0 : \tau_2 - \tau_1 = 0$, $H_A : \tau_2 - \tau_1 > 0$.	0.001	0.004
H5	There is no incremental effect of the offer of RA support (T_3), given the informational treatment with provision of the pre-filled results report (T_2): $H_0 : \tau_3 - \tau_2 = 0$, $H_A : \tau_3 - \tau_2 > 0$.	0.687	0.850

Note: One-sided *p*-values of the main hypotheses of the pre-analysis plan, numbered as there (H1, H4 and H5 are the primary hypotheses, H2 and H3 secondary), from the pre-registered difference-in-differences specification (equation 3.1 of the plan) with study fixed effects and robust standard errors; τ_j is the effect of arm T_j. Available (Y_1) is the fraction of pre-registered hypotheses with a numerical result in any publicly accessible source; Null (Y_3) the fraction with an available result that is null ($p > 0.05$). *p*-values are not adjusted for multiple testing.

Table B.6: Multiple Hypothesis Testing for the Primary Hypotheses

		Romano–Wolf adjusted p-value	
PAP	Hypothesis	Available	Null
H1	Awareness: $\tau_1 > 0$	0.141	0.178
H4	Engagement vs. awareness: $\tau_2 - \tau_1 > 0$	0.004	0.010
H5	Resources vs. engagement: $\tau_3 - \tau_2 > 0$	0.773	0.863
	Joint Wald test, all six treatment effects = 0: $\chi^2(6)$	20.52	
	p-value	0.002	

Note: Multiple hypothesis testing as in Section 3.2.3 of the pre-analysis plan. The pre-registered difference-in-differences specification is estimated for Available and Null stacked as seemingly unrelated regressions, with outcome-specific study fixed effects, post-period effects and treatment effects, and standard errors clustered by study. The adjusted p-values for the six primary one-sided hypotheses (H1, H4 and H5 for the two outcomes, H_A: the contrast is positive) are Romano–Wolf step-down p-values (Romano and Wolf, 2005, 2016), based on 2,000 bootstrap samples that resample studies within treatment arm. The joint Wald test is of the hypothesis that the effects of all three arms on both outcomes are zero.

C Details on Encoding Procedure

C.1 Encoding Protocol

This appendix describes the protocol used to encode pre-registered hypotheses in the baseline and match them to results reported in research outputs for studies in the AEA RCT Registry. The protocol governs three stages: determining which studies enter the sample, extracting the relevant elements of each pre-registered design, and matching reported results back to pre-registered hypotheses. The protocol establishes consistent rules for handling ambiguities in the identification of pre-registered hypotheses and their corresponding results.

Sample Inclusion

We restrict the sample to studies involving a genuine RCT in which treatment is assigned by the study team. Registrations are excluded when no treatment randomization occurs, when the entry is effectively empty, when results appear in the registry abstract (suggesting that registration occurred after the analysis), or when randomization applies only to survey sampling. Studies confirmed to have been registered only after data analysis are excluded. Studies described as pilots are retained when they implement a fully randomized design. A study registered after the intervention has concluded is retained only if the authors explicitly state that no outcome data had been analyzed prior to registration.

When a registration is modified after trial completion, the operative version is the last one predating the end of the intervention or the beginning of outcome data collection, unless the authors attest that later revisions were made without access to post-treatment data. When multiple PAPs exist, only the first one predating data collection for primary outcomes is used. If a referenced private PAP cannot be obtained despite outreach, encoding proceeds on public registry materials alone. If a PAP is available only in a non-English language, machine translation is used. When a PAP and registry entry conflict, we determine which takes precedence based on their timing relative to the intervention and outcome data collection, preferring the latest specification before intervention or data collcetion.

Baseline Encoding of Hypotheses

We record populations, interventions, arms, outcomes, hypotheses, and pre-specified heterogeneity analyses. Multiple phases are treated as a single study, with later phases coded as additional populations. Missing or vague intervention descriptions are reconstructed from the abstract and design description. Any procedure that would not have occurred in the absence of the study is treated as an intervention; non-randomized components are not.

Phase-in and stepped-wedge designs are treated as one treatment arm and one control arm. Arms are mutually exclusive. In factorial designs, all independently randomized combinations are treated as distinct arms. In audit studies, we assume a factorial design unless clearly indicated otherwise; when the distinction between factorial and parallel designs is ambiguous, we default to parallel. When interactions are clearly pre-specified, the interaction structure is encoded; otherwise, we encode interactions in 2×2 designs but restrict higher-order factorials to parallel contrasts. In cascading designs, hypotheses isolate the marginal effect of each incremental component. When no

explicit control group is defined, the most plausible reference group is inferred and documented; non-randomized comparison groups are not treated as valid arms.

When a registration lists an impractically large number of primary outcomes, we prioritize those most central to the stated research question. Outcome families are encoded as indices when a composite is explicitly defined; otherwise, components are recorded separately. For vaguely specified outcomes, we preserve the registry language and document the coder's interpretation. Mechanism and process outcomes are encoded only when central to the research question, causally identifiable, and explicitly pre-registered. When a registered outcome could correspond to multiple measurement forms, each distinct form is coded separately.

Hypotheses identify the estimand as precisely as possible. In parallel-arm studies, we encode only the minimal comparisons needed—usually each treatment arm relative to control—without adding contrasts or joint tests unless explicitly pre-specified. In difference-in-differences designs, the estimand is the treatment-by-post interaction. ITT is primary when both ITT and LATE are pre-specified. Quantile treatment effects are distinct from average effects. When multiple specifications are pre-specified, we encode the most central one. Contingent hypotheses are encoded once, with the contingency recorded and the results encoding results matched using the realized condition. Hypotheses are marked as *not detailed* when the registration does not identify a specific operationalization, contrast, or estimating equation. Non-experimental comparisons are not encoded. Spillover hypotheses are encoded only when explicitly pre-specified and central to the design. Heterogeneity analyses are included only when subgroup definitions are clear, central, and sufficiently pre-specified; they are not encoded in isolation from the main hypothesis. When the registration distinguishes "different effects" from "separate effects," that distinction is preserved; otherwise, "different effects" is the default.

Results Encoding

Results encoding matches reported results to the corresponding baseline hypotheses, recording any deviations and was done twice: Once in the baseline before randomization and once in the follow-up, 6 months after randomization and interventions. Reported outcomes are classified as *modified consistent* when the alternative operationalization could not plausibly have been chosen ex post, and *modified inconsistent* otherwise. This distinction applies to changes in transformations, functional form, and outcome construction. When several reported measures could correspond to the same underspecified registered outcome, we select the broadest or most aggregated measure; if several are equally plausible, one is selected at random and the choice is documented. When pre-registered outcomes are combined into a single reported index, the component most closely aligned with the index is recorded as modified consistent, while the remaining components are coded as *not found*. If both binary and continuous forms were explicitly pre-registered and the paper reports one, the hypothesis is treated as found and unmodified.

A hypothesis is *completely found* when the treatment effect can be recovered without assumptions beyond what is explicitly reported—whether directly stated, calculated using reported statistics, inferred from treatment-arm means, or read from figures. Standard errors are recovered from confidence intervals by dividing their width by 3.92, or from arm-level statistics by taking the square root of the sum of squared arm-level standard errors. Partially available results are coded as *partially found*. We do not reconstruct results requiring structural assumptions or unreported econometric steps.

When multiple specifications are reported, we select the column most closely matching the baseline hypothesis; in the absence of pre-specified covariates, we prefer the unadjusted specification. If one hypothesis maps to several variants, one is selected at random without regard to statistical significance. Covariate changes are classified as consistent when involving pre-treatment variables and inconsistent when introducing post-treatment controls. Minor changes that do not substantively alter the estimand are not treated as modifications. When a registration implies an endline comparison and the paper estimates difference-in-differences, or when a binary outcome is reported using logit/probit rather than a linear probability model, the result is classified as modified consistent. Explicitly acknowledged PAP deviations are recorded accordingly.

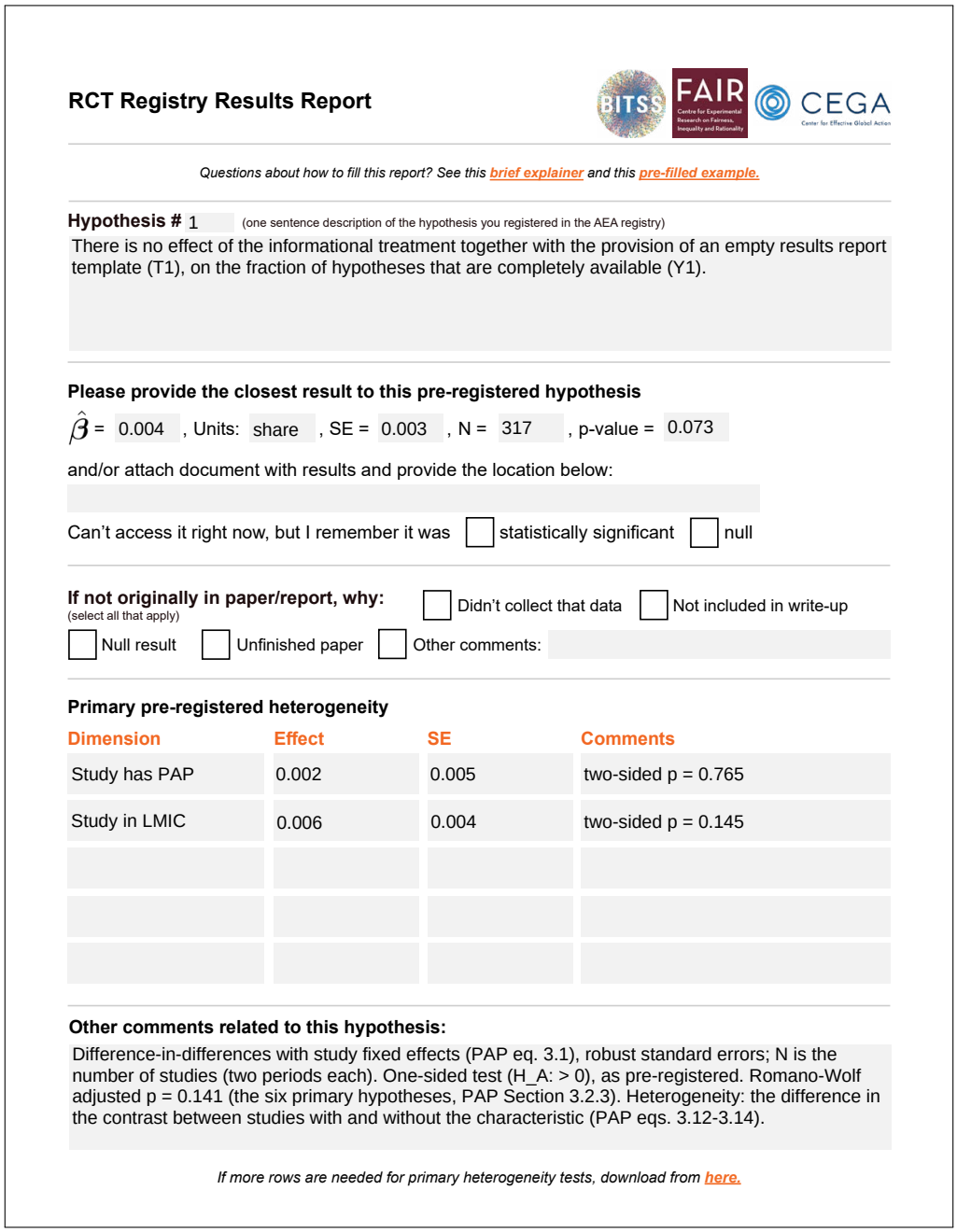
Judgment Calls

Coding decisions not fully resolved by the protocol are documented as judgment calls, describing the ambiguity, the decision made, and the rule or precedent applied. The most common judgment calls concern registration timing, ambiguous arm structure, underspecified outcomes, specification choice, and matching vague pre-registered outcomes to reported measures.

D An example Results Report: The Current Study

The Results Report provided to authors in the *Engagement* and *Resources* arms was study-specific and based on the information collected at baseline. It began with a cover letter identifying the study and the number of pre-registered hypotheses we had constructed (Appendix, Figure E.11), followed by a separate page for each hypothesis. Each page presented the standardized hypothesis statement and linked to documentation of how it had been constructed, including the relevant population, outcome, intervention, treatment arms, and any judgment calls (Appendix, Figure E.18). For hypotheses with results identified at baseline, the report presented the corresponding estimates and asked authors to review and, where necessary, correct the information (Appendix, Figure E.12). For hypotheses without identified results, the report indicated that no result had been found and invited authors to provide the missing information (Appendix, Figure E.13). The report also included instructions for posting results to the AEA Registry and a link to a completed example.

In the following pages, Figure D.1–Figure D.6, we have populated the results report templates with our own pre-registered hypotheses and the corresponding results.



RCT Registry Results Report

Questions about how to fill this report? See this brief explainer and this pre-filled example.

Hypothesis # 1 (one sentence description of the hypothesis you registered in the AEA registry)

There is no effect of the informational treatment together with the provision of an empty results report template (T1), on the fraction of hypotheses that are completely available (Y1).

Please provide the closest result to this pre-registered hypothesis

$\hat{\beta}$ = 0.004 , Units: share , SE = 0.003 , N = 317 , p-value = 0.073

and/or attach document with results and provide the location below:

Can't access it right now, but I remember it was ☐ statistically significant ☐ null

If not originally in paper/report, why:
(select all that apply)

☐ Didn't collect that data ☐ Not included in write-up

☐ Null result ☐ Unfinished paper ☐ Other comments:

Primary pre-registered heterogeneity

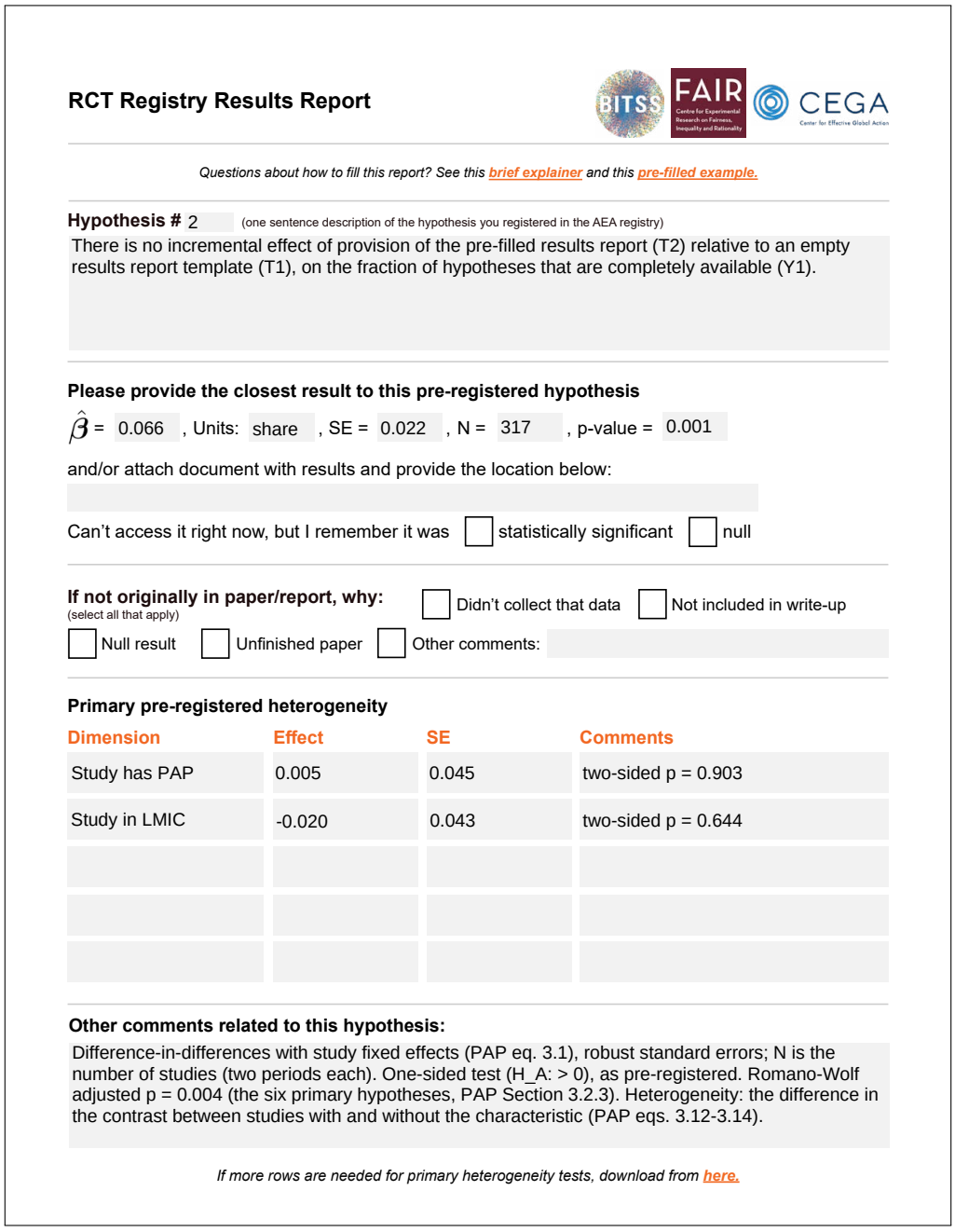
Dimension	Effect	SE	Comments
Study has PAP	0.002	0.005	two-sided p = 0.765
Study in LMIC	0.006	0.004	two-sided p = 0.145

Other comments related to this hypothesis:

Difference-in-differences with study fixed effects (PAP eq. 3.1), robust standard errors; N is the number of studies (two periods each). One-sided test (H_A: > 0), as pre-registered. Romano-Wolf adjusted p = 0.141 (the six primary hypotheses, PAP Section 3.2.3). Heterogeneity: the difference in the contrast between studies with and without the characteristic (PAP eqs. 3.12-3.14).

If more rows are needed for primary heterogeneity tests, download from here.

Figure D.1: *This Project's* Results Report, Page 1 of 6 (H1, Available)



RCT Registry Results Report

Questions about how to fill this report? See this brief explainer and this pre-filled example.

Hypothesis # 2 (one sentence description of the hypothesis you registered in the AEA registry)

There is no incremental effect of provision of the pre-filled results report (T2) relative to an empty results report template (T1), on the fraction of hypotheses that are completely available (Y1).

Please provide the closest result to this pre-registered hypothesis

$\hat{\beta}$ = 0.066 , Units: share , SE = 0.022 , N = 317 , p-value = 0.001

and/or attach document with results and provide the location below:

Can't access it right now, but I remember it was ☐ statistically significant ☐ null

If not originally in paper/report, why:
(select all that apply)

☐ Didn't collect that data ☐ Not included in write-up
☐ Null result ☐ Unfinished paper ☐ Other comments:

Primary pre-registered heterogeneity

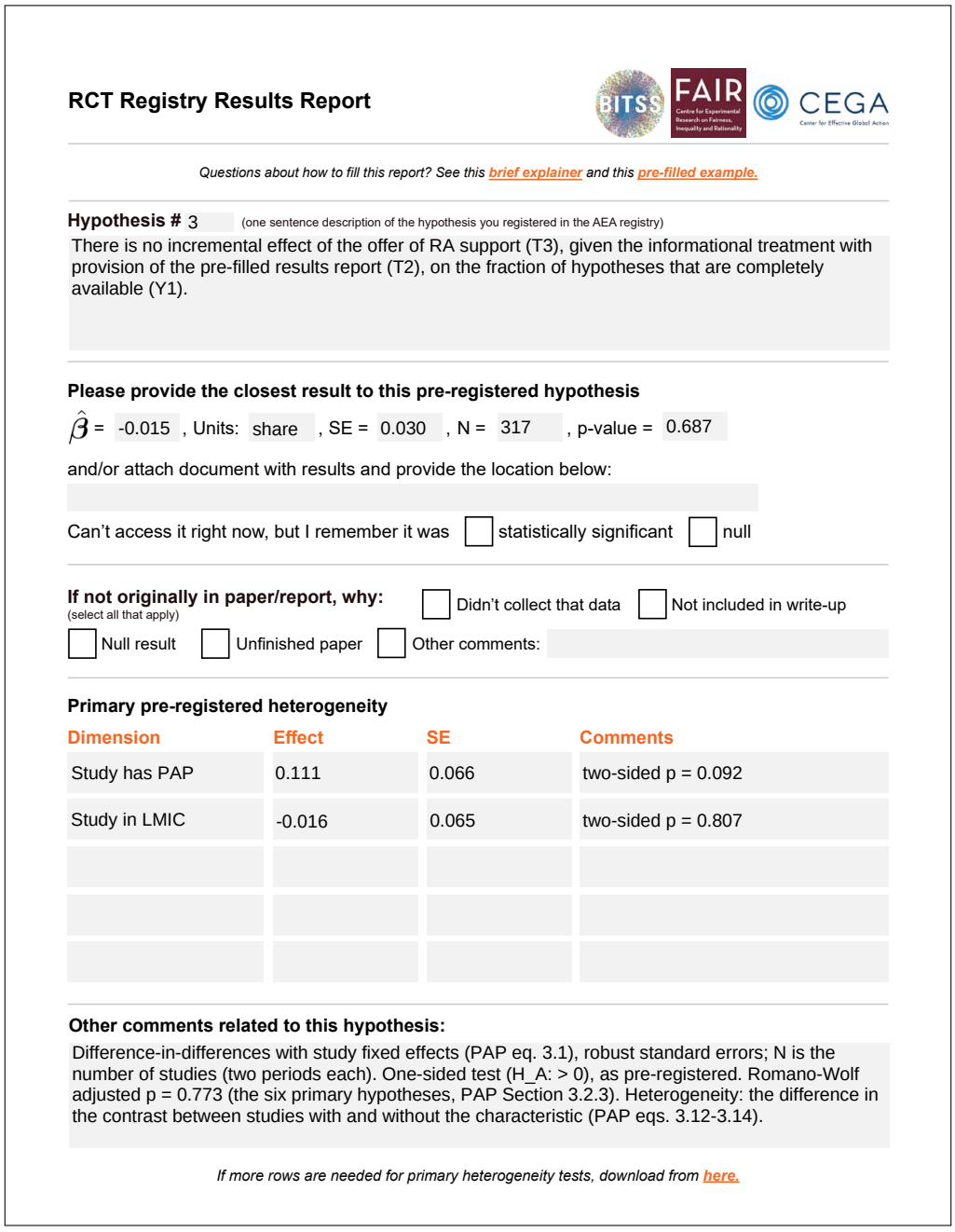
Dimension	Effect	SE	Comments
Study has PAP	0.005	0.045	two-sided p = 0.903
Study in LMIC	-0.020	0.043	two-sided p = 0.644

Other comments related to this hypothesis:

Difference-in-differences with study fixed effects (PAP eq. 3.1), robust standard errors; N is the number of studies (two periods each). One-sided test (H_A: > 0), as pre-registered. Romano-Wolf adjusted p = 0.004 (the six primary hypotheses, PAP Section 3.2.3). Heterogeneity: the difference in the contrast between studies with and without the characteristic (PAP eqs. 3.12-3.14).

If more rows are needed for primary heterogeneity tests, download from here.

Figure D.2: *This Project's* Results Report, Page 2 of 6 (H4, Available)



RCT Registry Results Report

Questions about how to fill this report? See this brief explainer and this pre-filled example.

Hypothesis # 3 (one sentence description of the hypothesis you registered in the AEA registry)

There is no incremental effect of the offer of RA support (T3), given the informational treatment with provision of the pre-filled results report (T2), on the fraction of hypotheses that are completely available (Y1).

Please provide the closest result to this pre-registered hypothesis

$\hat{\beta}$ = -0.015 , Units: share , SE = 0.030 , N = 317 , p-value = 0.687

and/or attach document with results and provide the location below:

Can't access it right now, but I remember it was ☐ statistically significant ☐ null

If not originally in paper/report, why:
(select all that apply)

☐ Didn't collect that data ☐ Not included in write-up
☐ Null result ☐ Unfinished paper ☐ Other comments:

Primary pre-registered heterogeneity

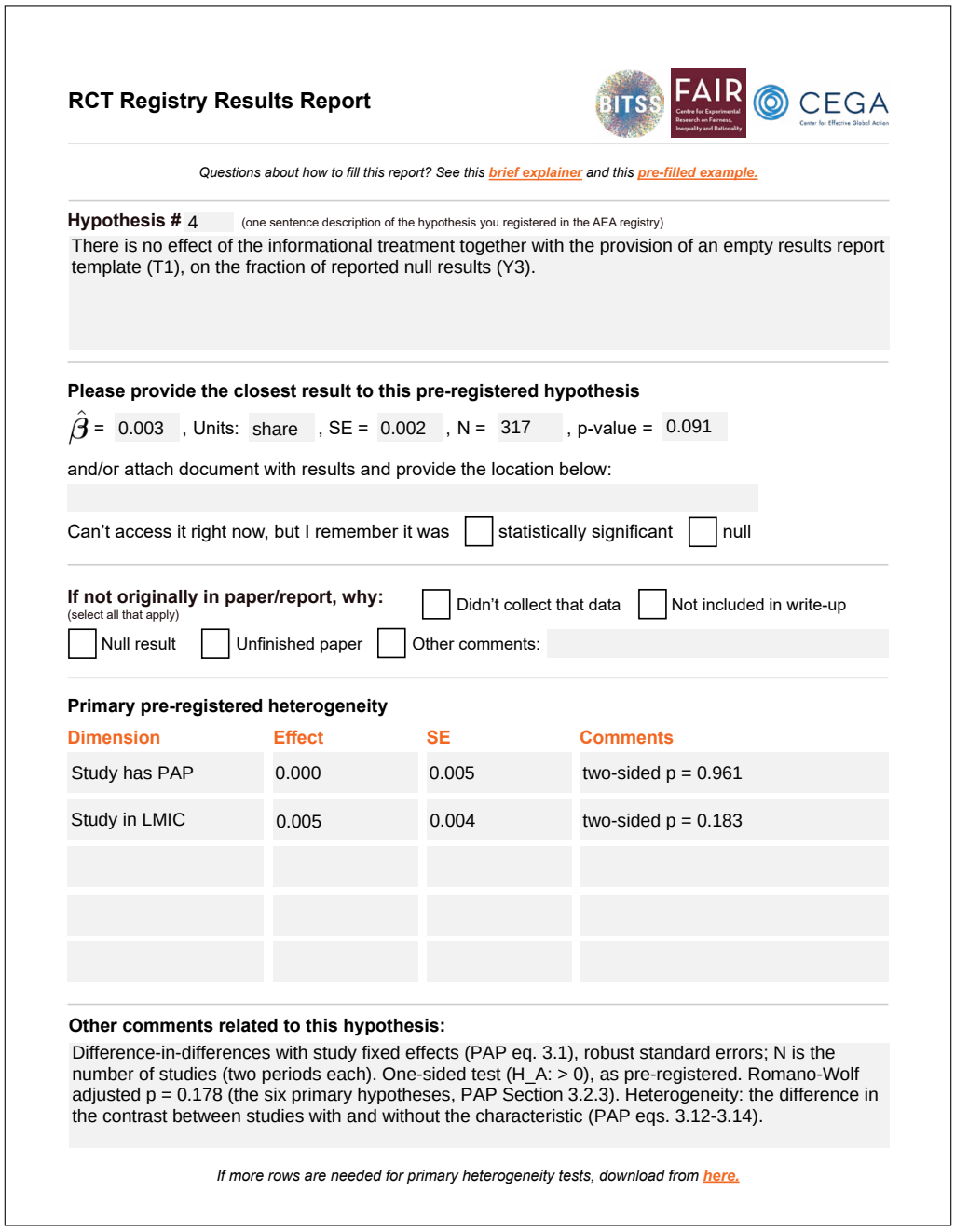
Dimension	Effect	SE	Comments
Study has PAP	0.111	0.066	two-sided p = 0.092
Study in LMIC	-0.016	0.065	two-sided p = 0.807

Other comments related to this hypothesis:

Difference-in-differences with study fixed effects (PAP eq. 3.1), robust standard errors; N is the number of studies (two periods each). One-sided test (H_A: > 0), as pre-registered. Romano-Wolf adjusted p = 0.773 (the six primary hypotheses, PAP Section 3.2.3). Heterogeneity: the difference in the contrast between studies with and without the characteristic (PAP eqs. 3.12-3.14).

If more rows are needed for primary heterogeneity tests, download from here.

Figure D.3: *This Project's* Results Report, Page 3 of 6 (H5, Available)



RCT Registry Results Report

Questions about how to fill this report? See this brief explainer and this pre-filled example.

Hypothesis # 4 (one sentence description of the hypothesis you registered in the AEA registry)

There is no effect of the informational treatment together with the provision of an empty results report template (T1), on the fraction of reported null results (Y3).

Please provide the closest result to this pre-registered hypothesis

$\hat{\beta}$ = 0.003 , Units: share , SE = 0.002 , N = 317 , p-value = 0.091

and/or attach document with results and provide the location below:

Can't access it right now, but I remember it was ☐ statistically significant ☐ null

If not originally in paper/report, why:
(select all that apply)

☐ Didn't collect that data ☐ Not included in write-up
☐ Null result ☐ Unfinished paper ☐ Other comments:

Primary pre-registered heterogeneity

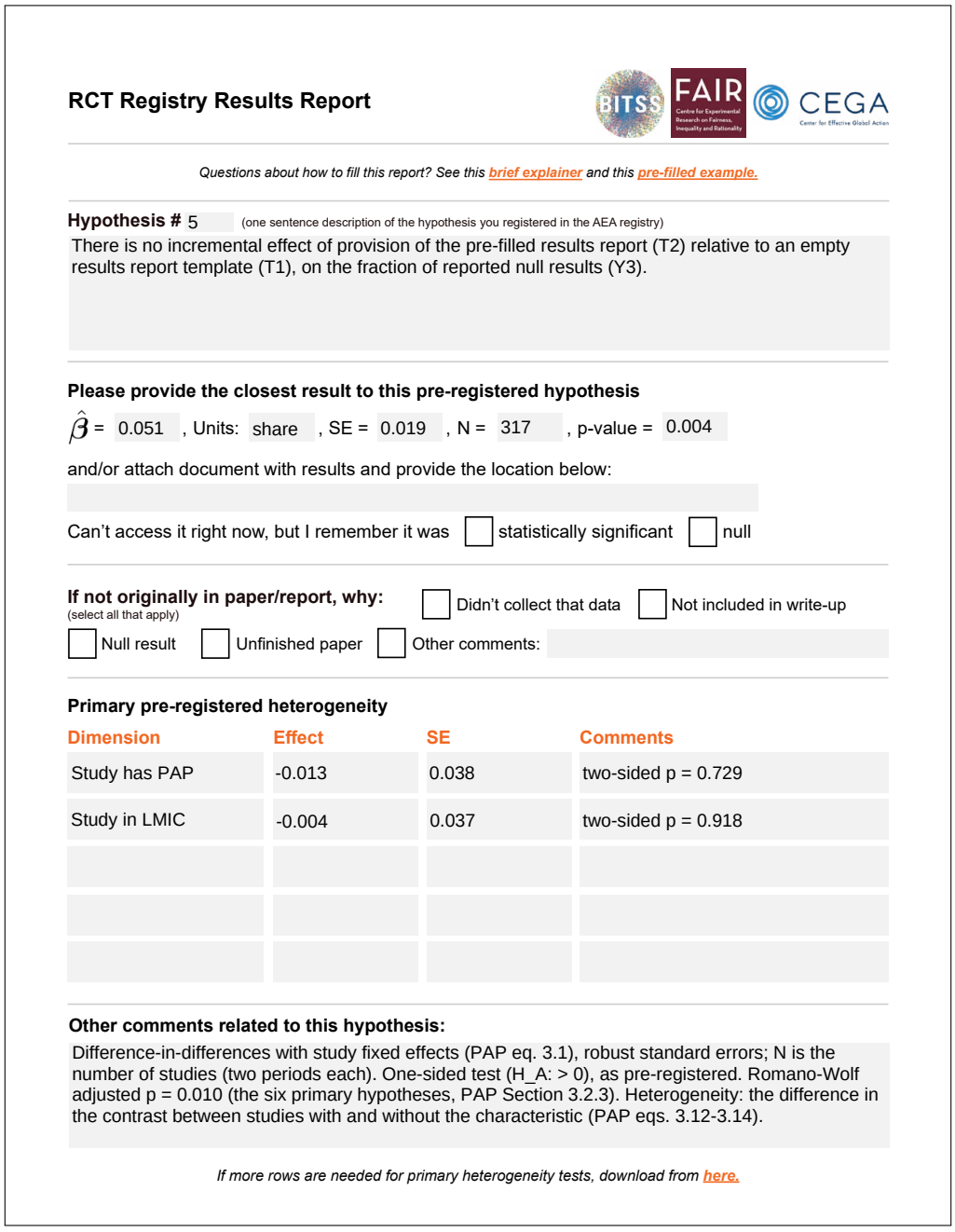
Dimension	Effect	SE	Comments
Study has PAP	0.000	0.005	two-sided p = 0.961
Study in LMIC	0.005	0.004	two-sided p = 0.183

Other comments related to this hypothesis:

Difference-in-differences with study fixed effects (PAP eq. 3.1), robust standard errors; N is the number of studies (two periods each). One-sided test (H_A: > 0), as pre-registered. Romano-Wolf adjusted p = 0.178 (the six primary hypotheses, PAP Section 3.2.3). Heterogeneity: the difference in the contrast between studies with and without the characteristic (PAP eqs. 3.12-3.14).

If more rows are needed for primary heterogeneity tests, download from here.

Figure D.4: *This Project's* Results Report, Page 4 of 6 (H1, Null)



RCT Registry Results Report

Questions about how to fill this report? See this brief explainer and this pre-filled example.

Hypothesis # 5 (one sentence description of the hypothesis you registered in the AEA registry)

There is no incremental effect of provision of the pre-filled results report (T2) relative to an empty results report template (T1), on the fraction of reported null results (Y3).

Please provide the closest result to this pre-registered hypothesis

$\hat{\beta}$ = 0.051 , Units: share , SE = 0.019 , N = 317 , p-value = 0.004

and/or attach document with results and provide the location below:

Can't access it right now, but I remember it was ☐ statistically significant ☐ null

If not originally in paper/report, why:
(select all that apply)

☐ Didn't collect that data ☐ Not included in write-up
☐ Null result ☐ Unfinished paper ☐ Other comments:

Primary pre-registered heterogeneity

Dimension	Effect	SE	Comments
Study has PAP	-0.013	0.038	two-sided p = 0.729
Study in LMIC	-0.004	0.037	two-sided p = 0.918

Other comments related to this hypothesis:

Difference-in-differences with study fixed effects (PAP eq. 3.1), robust standard errors; N is the number of studies (two periods each). One-sided test (H_A: > 0), as pre-registered. Romano-Wolf adjusted p = 0.010 (the six primary hypotheses, PAP Section 3.2.3). Heterogeneity: the difference in the contrast between studies with and without the characteristic (PAP eqs. 3.12-3.14).

If more rows are needed for primary heterogeneity tests, download from here.

Figure D.5: *This Project's* Results Report, Page 5 of 6 (H4, Null)

RCT Registry Results Report

Questions about how to fill this report? See this brief explainer and this pre-filled example.

Hypothesis # 6 (one sentence description of the hypothesis you registered in the AEA registry)

There is no incremental effect of the offer of RA support (T3), given the informational treatment with provision of the pre-filled results report (T2), on the fraction of reported null results (Y3).

Please provide the closest result to this pre-registered hypothesis

$\hat{\beta}$ = -0.024 , Units: share , SE = 0.023 , N = 317 , p-value = 0.850

and/or attach document with results and provide the location below:

Can't access it right now, but I remember it was ☐ statistically significant ☐ null

If not originally in paper/report, why:
(select all that apply)

☐ Didn't collect that data ☐ Not included in write-up
☐ Null result ☐ Unfinished paper ☐ Other comments:

Primary pre-registered heterogeneity

Dimension	Effect	SE	Comments
Study has PAP	0.063	0.050	two-sided p = 0.208
Study in LMIC	-0.047	0.052	two-sided p = 0.368

Other comments related to this hypothesis:

Difference-in-differences with study fixed effects (PAP eq. 3.1), robust standard errors; N is the number of studies (two periods each). One-sided test (H_A: > 0), as pre-registered. Romano-Wolf adjusted p = 0.863 (the six primary hypotheses, PAP Section 3.2.3). Heterogeneity: the difference in the contrast between studies with and without the characteristic (PAP eqs. 3.12-3.14).

If more rows are needed for primary heterogeneity tests, download from here.

Figure D.6: *This Project's* Results Report, Page 6 of 6 (H5, Null)

E Materials Sent to the Authors of the Studies in the RCT

Treatment #1
Subject: Reporting results for your study on the AEA RCT Registry

Dear {{Researcher Name}},

We are a team of researchers studying pre-registration, pre-analysis plans, and the reporting of results in social science research, funded by the Research Council of Norway (#262675) and the Berkeley Initiative for Transparency in the Social Sciences (BITSS).

To have an accurate representation of evidence in our field, it's essential to keep track of the results from pre-registered studies, regardless of how they turn out. For that purpose, we've developed a short **results report** to standardize and streamline the reporting of all research hypotheses for studies on the AEA Registry.

The filled-in results report for your study will serve as a useful resource for the economics research community. We encourage you to upload a results report from your study "{{TITLE OF STUDY}}" to the AEA Registry. In the form, we ask you to briefly describe **the hypotheses recorded in {{AEA-ID}} and their associated results**. To help you navigate this process, we are linking an empty results report, as well as a pre-filled example from another study. The linked results report template is a suggested format, but of course, you are free to share the results in other formats, including in a paper. To submit, please upload via the Post-Trial section on the AEA RCT Registry page for your study (in the "Reports, Papers, and Other Materials" tab), or if you prefer, you can attach the file in a reply to this email.

We plan to use the information you and other scholars report in a study of publication patterns in economics. As a part of this project, study-level information from the results reports – including project information you share with us, such as pre-analysis plans or additional statistical results – may be shared publicly online (though we will not highlight individual projects or authors). Please review the study consent form, and feel free to reach out with any questions.

We understand that providing this information is a non-trivial ask, and we would be very grateful for your participation.

Many thanks,

Edward Miguel (UC Berkeley)
Bertil Tungodden (NHH Norwegian School of Economics)
Erik Ø. Sørensen (NHH Norwegian School of Economics)
Fernando Hoces de la Guardia (UC Berkeley)

Figure E.1: *Awareness* Treatment Sample Email

Treatment #2
Subject: Results report for your study on the AEA RCT Registry

Dear {{Researcher Name}},

We are a team of researchers studying pre-registration, pre-analysis plans, and the reporting of results in social science research, funded by the Research Council of Norway (#262675) and the Berkeley Initiative for Transparency in the Social Sciences (BITSS).

To have an accurate representation of evidence in our field, it's essential to keep track of the results from pre-registered studies, regardless of how they turn out. For that purpose, we've developed a short **results report** to standardize and streamline the reporting of all research hypotheses for studies on the AEA Registry.

Our team has read the materials on the AEA Registry for your study "{{TITLE OF STUDY}}" ({{AEA-ID}}) and encoded the research hypotheses into the {{Hyperlink}}. We've also read any paper(s) associated with this registration and attempted to record the results of the primary hypotheses. **If possible, we ask you to review these materials and provide feedback in two areas:**

1. Did we report the hypotheses (as you recorded them on the AEA Registry) and the results of your study correctly? If you disagree or would have encoded them differently, please note your preferred hypotheses and results using the linked results report.
2. In some cases, we could not find the results for specific research hypotheses recorded on the AEA Registry. Could you fill in these estimates or report why they are not available?

The filled-in results report for your study will serve as a useful resource for the economics research community. We encourage you to upload a results report from your study "{{TITLE OF STUDY}}" to the AEA Registry. To help you navigate this process, we are linking a pre-filled example results report from another study. The linked results report template is a suggested format, but of course, you are free to share the results in other formats, including in a paper. To submit, please upload via the Post-Trial section on the AEA RCT Registry page for your study (in the "Reports, Papers, and Other Materials" tab), or if you prefer, you can attach the file in a reply to this email.

We plan to use the information you and other scholars report in a study of publication patterns in economics. As a part of this project, study-level information from the results reports – including project information you share with us, such as pre-analysis plans or additional statistical results – may be shared publicly online (though we will not highlight individual projects or authors). Please review the study consent form, and feel free to reach out with any questions.

We understand that providing this information is a non-trivial ask, and we would be very grateful for your participation.

Many thanks,

Edward Miguel (UC Berkeley)
Bertil Tungodden (NHH Norwegian School of Economics)
Erik Ø. Sørensen (NHH Norwegian School of Economics)
Fernando Hoces de la Guardia (UC Berkeley)

Figure E.2: *Engagement* Treatment Sample Email

Treatment #3
Subject: Results report for your study on the AEA RCT Registry + RA support

Dear {{Researcher Name}},

We are a team of researchers studying pre-registration, pre-analysis plans, and the reporting of results in social science research, funded by the Research Council of Norway (#262675) and the Berkeley Initiative for Transparency in the Social Sciences (BITSS).

To have an accurate representation of evidence in our field, it's essential to keep track of the results from pre-registered studies, regardless of how they turn out. For that purpose, we've developed a short **results report** to standardize and streamline the reporting of all research hypotheses for studies on the AEA Registry.

Our team has read the materials on the AEA Registry for your study "{{TITLE OF STUDY}}" ({{AEA-ID}}) and encoded the research hypotheses into the {{Hyperlink}}. We've also read any paper(s) associated with this registration and attempted to record the results of the primary hypotheses. **If possible, we ask you to review these materials and provide feedback in two areas:**

1. Did we report the hypotheses (as you recorded them on the AEA Registry) and the results of your study correctly? If you disagree or would have encoded them differently, please note your preferred hypotheses and results using the linked results report.
2. In some cases, we could not find the results for specific research hypotheses recorded on the AEA Registry. Could you fill in these estimates or report why they are not available?

The filled-in results report for your study will serve as a useful resource for the economics research community. We encourage you to upload a results report from your study "{{TITLE OF STUDY}}" to the AEA Registry. To help you navigate this process, we are linking a pre-filled example results report from another study. The linked results report template is a suggested format, but of course, you are free to share the results in other formats, including in a paper. To submit, please upload via the Post-Trial section on the AEA RCT Registry page for your study (in the "Reports, Papers, and Other Materials" tab), or if you prefer, you can attach the file in a reply to this email.

To facilitate this work, we are able to offer you the support of a UC Berkeley-based research assistant (at no charge to you) who has knowledge of common programming software and languages (e.g., STATA and R). We can offer you up to 30 hours of RA work time on this task of obtaining study estimates, specifically, on data cleaning, analysis, and documentation. Alternatively, if this proves to be a logistical challenge, we can offer **up to $1,500 to cover the RA time required on your end to find or produce the estimates for the results report.** Please let us know if you are interested in this support by responding to this email (ucbitss@berkeley.edu) in the next 14 days (by {{Date}}).

We plan to use the information you and other scholars report in a study of publication patterns in economics. As a part of this project, study-level information from the results reports – including project information you share with us, such as pre-analysis plans or additional statistical results – may be shared publicly online (though we will not highlight individual projects or authors). Please review the study consent form, and feel free to reach out with any questions.

We understand that providing this information is a non-trivial ask, and we would be very grateful for your participation.

Many thanks,

Edward Miguel (UC Berkeley)
Bertil Tungodden (NHH Norwegian School of Economics)
Erik Ø. Sørensen (NHH Norwegian School of Economics)
Fernando Hoces de la Guardia (UC Berkeley)

Figure E.3: *Resources* Treatment Sample Email

UNIVERSITY OF CALIFORNIA AT BERKELEY

BERKELEY DAVIS IRVINE LOS ANGELES MERCED RIVERSIDE SAN DIEGO SAN FRANCISCO SANTA BARBARA SANTA CRUZ

CONSENT TO PARTICIPATE IN RESEARCH

Pre-Registration, Reporting Guidelines and Publication Patterns in Economics

Introduction: My name is Fernando Hoces De La Guardia and I am working with Edward Miguel, a Professor of Economics at the University of California, Berkeley to conduct a research study, which we invite you to take part in.

Purpose: The purpose of this research study is *to learn about publication patterns in the social sciences through the use of results reports.*

Procedures: If you agree to be in this study, you will be sent a results report template over email with a brief message on the importance of reporting results from pre-registered studies, regardless of how they turn out, and encouraging you to fill up and upload the results report to the AEA Registry. The time it will take to populate and upload this results report is variable. All activities can take place remotely.

Benefits: There are no direct benefits to participating in this study. We hope that the information gained from the study will help improve understanding of results reporting and publication patterns in the social sciences.

Risks/Discomforts: Given that study-level data may be shared publicly, there is a chance that the level of adherence of a published study to a pre-analysis plan could be linked to you and your co-authors.

Future use of study data: Study-level data, including non-public sources or elements you share with us, may be shared publicly online.

Compensation and Costs: You will not receive any monetary compensation for your participation in this study, nor will you be charged for any of the study activities.

Rights: Participation in research is completely voluntary. You have the right to decline to participate or to withdraw at any point in this study without penalty or loss of benefits to which you are otherwise entitled.

Questions: If you have any questions or concerns about this study, you may contact Professor Edward Miguel at emiguel@berkeley.edu. If you have any questions or concerns about your rights and treatment as a research subject, you may contact the office of UC Berkeley's Committee for the Protection of Human Subjects, at 510-642-7461 or subjects@berkeley.edu.

If you do not agree to participate, please reply to this email letting us know.

CPHS # 2023-08-16633Page **1** of **1**

Figure E.4: *Awareness* Treatment's Study Consent Form

CONSENT TO PARTICIPATE IN RESEARCH
Pre-Registration, Reporting Guidelines and Publication Patterns in Economics

Introduction: My name is Fernando Hoces De La Guardia and I am working with Edward Miguel, a Professor of Economics at the University of California, Berkeley to conduct a research study, which we invite you to take part in.

Purpose: The purpose of this research study is *to learn about publication patterns in the social sciences through the use of results reports.*

Procedures: If you agree to be in this study, you will be sent a results report template over email pre-populated with information from your study's AEA Registry page with a brief message on the importance of reporting results from pre-registered studies, regardless of how they turn out, and encouraging you to upload the report to the AEA Registry. The time it will take to review and upload this results results report is variable. All activities can take place remotely.

Benefits: There are no direct benefits to participating in this study. We hope that the information gained from the study will help improve understanding of results reporting and publication patterns in the social sciences.

Risks/Discomforts: Given that study-level data may be shared publicly, there is a chance that the level of adherence of a published study to a pre-analysis plan could be linked to you and your co-authors.

Future use of study data: Study-level data, including non-public sources or elements you share with us, may be shared publicly online.

Compensation and Costs: You will not receive any monetary compensation for your participation in this study, nor will you be charged for any of the study activities.

Rights: Participation in research is completely voluntary. You have the right to decline to participate or to withdraw at any point in this study without penalty or loss of benefits to which you are otherwise entitled.

Questions: If you have any questions or concerns about this study, you may contact Professor Edward Miguel at emiguel@berkeley.edu. If you have any questions or concerns

Figure E.5: *Engagement* Treatment's Study Consent Form, Page 1 of 2

about your rights and treatment as a research subject, you may contact the office of UC Berkeley's Committee for the Protection of Human Subjects, at 510-642-7461 or subjects@berkeley.edu.

If you do not agree to participate, please reply to this email letting us know.

Figure E.6: *Engagement* Treatment's Study Consent Form, Page 2 of 2

CONSENT TO PARTICIPATE IN RESEARCH
Pre-Registration, Reporting Guidelines and Publication Patterns in Economics

Introduction: My name is Fernando Hoces De La Guardia and I am working with Edward Miguel, a Professor of Economics at the University of California, Berkeley to conduct a research study, which we invite you to take part in.

Purpose: The purpose of this research study is *to learn about publication patterns in the social sciences through the use of results reports.*

Procedures: If you agree to be in this study, you will be sent a results report template over email pre-populated with information from your study's AEA Registry page with a brief message on the importance of reporting results from pre-registered studies, regardless of how they turn out, and encouraging you to upload the results report to the AEA Registry. You will also be offered up to 30 hours of RA support to obtain the study estimates. Alternatively, if this proves to be a logistical challenge, we can offer up to $1,500 to cover the RA time on your end. The time it takes to coordinate with the RAs, review the output, and upload the report is variable. All activities can take place remotely.

Benefits: You will be offered up to 30 hours of RA support to obtain the study estimates. Alternatively, if this proves to be a logistical challenge, we can offer up to $1,500 to cover the RA time on your end. We hope that the information gained from the study will help improve understanding of results reporting and publication patterns in the social sciences.

Risks/Discomforts: Given that study-level data may be shared publicly, there is a chance that the level of adherence of a published study to a pre-analysis plan could be linked to you and your co-authors.

Future use of study data: Study-level data, including non-public sources or elements you share with us, may be shared publicly online.

Compensation and Costs: You will not receive any monetary compensation for your participation in this study, nor will you be charged for any of the study activities.

Figure E.7: *Resources* Treatment's Study Consent Form, Page 1 of 2

Rights: Participation in research is completely voluntary. You have the right to decline to participate or to withdraw at any point in this study without penalty or loss of benefits to which you are otherwise entitled.

Questions: If you have any questions or concerns about this study, you may contact Professor Edward Miguel at emiguel@berkeley.edu. If you have any questions or concerns about your rights and treatment as a research subject, you may contact the office of UC Berkeley's Committee for the Protection of Human Subjects, at 510-642-7461 or subjects@berkeley.edu.

If you do not agree to participate, please reply to this email letting us know.

Figure E.8: *Resources* Treatment's Study Consent Form, Page 2 of 2



Reporting results for your study on the AEA RCT Registry

We are a team of researchers studying pre-registration, pre-analysis plans, and the reporting of results in social science research. Our research team has developed a short template to facilitate reporting all pre-registered hypotheses for a given study.

It would be extremely valuable for the research community if you could provide results. We also encourage you to post these results on the AEA RCT Registry page for your study (in the "Reports, Papers, and Other Materials" section).

We thank you for your valuable participation.

Edward Miguel (UC Berkeley)
Bertil Tungodden (NHH Norwegian School of Economics)
Erik Ø. Sørensen (NHH Norwegian School of Economics)
Fernando Hoces de la Guardia (UC Berkeley)

Figure E.9: *Empty* Results Report Sample Cover

RCT Registry Results Report

Questions about how to fill this report? See this brief explainer and this pre-filled example.

Hypothesis # (one sentence description of the hypothesis you registered in the AEA registry)

Please provide the closest result to this pre-registered hypothesis

$\hat{\beta}$ = , Units: , SE = , N = , p-value =

and/or attach document with results and provide the location below:

Can't access it right now, but I remember it was ☐ statistically significant ☐ null

If not originally in paper/report, why:
(select all that apply)

☐ Didn't collect that data ☐ Not included in write-up

☐ Null result ☐ Unfinished paper ☐ Other comments:

Primary pre-registered heterogeneity

Dimension	Effect	SE	Comments

Other comments related to this hypothesis:

If more rows are needed for primary heterogeneity tests, download from here.

Figure E.10: *Empty* Results Report Sample Page

Note: Sent to all three treatment arms.



Reporting results for your study on the AEA RCT Registry:

Study title goes here

We are a team of researchers studying pre-registration, pre-analysis plans, and the reporting of results in social science research. Our research team has developed a short template to facilitate reporting all pre-registered hypotheses for a given study.

We have extracted hypotheses from your study registered on the AEA RCT registry and attempted to find results for each of them.

We extracted x main hypotheses from the registration. We found results for y main hypotheses.

It would be extremely valuable for the research community if you could provide results that we were unable to find and review the ones that we found. We also encourage you to post these results via the Post-Trial section on the AEA RCT Registry page for your study (in the "Reports, Papers, and Other Materials" tab). If you would like to use an empty results report template, you can access it by clicking here.

We thank you for your valuable participation.

Edward Miguel (UC Berkeley)
Bertil Tungodden (NHH Norwegian School of Economics)
Erik Ø. Sørensen (NHH Norwegian School of Economics)
Fernando Hoces de la Guardia (UC Berkeley)

Figure E.11: *Pre-Filled* Results Report Sample Cover

Note: Sent only to the *Engagement* and *Resources* treatment arms.

RCT Registry Results Report

Main Results Found

Questions about how to fill in this report? See this brief explainer and this pre-filled example.

Here is your original registration, and the attachment details_lastname.pdf contains the details of how your registration was encoded.

Hypothesis # 10 (as interpreted from pre-registration)

Outcome 5 is not different in Treatment 3 relative to Treatment 2.

Mostly agree with this statement? ☐ Yes ☐ No **If not, please say why and add correct one:**

Closest result to this pre-registered hypothesis (as found in article, publication, or write-up)

$\hat{\beta}$ = 12345 , SE = 54321

Location: Page X, Table Y, col W, row Z

Agree with result? ☐ Yes ☐ No **If not, please tell us why here:**

Primary pre-registered heterogeneity Agree with below? ☐ Yes ☐ No

If you disagree with this interpretation of estimates and/or believe that some dimensions of primary pre-registered heterogeneity are missing, add or modify them below. We will keep track of your edits.

Dimension	Effect	SE	Comments
Heterogeneity A	12345	12345	
Heterogeneity B	12345	12345	

If more rows are needed for primary heterogeneity tests, download from here.

Figure E.12: *Found* Results Report Sample Page

Note: Sent only to the *Engagement* and *Resources* treatment arms.

RCT Registry Results Report

Main Results Not Found

Questions about how to fill in this report? See this brief explainer and this pre-filled example.

Here is your original registration, and the attachment details_lastname.pdf contains the details of how your registration was encoded.

Hypothesis # 10 (as interpreted from pre-registration)

Outcome 5 is not different in Treatment 3 relative to Treatment 2.

Mostly agree with this statement? ☐ Yes ☐ No **If not, please say why and add correct one:**

Please provide closest result to this pre-registered hypothesis. ***We could not find this result.***

$\hat{\beta}$ = , Units: , SE = , N = , p-value =

and/or attach document with results and provide the location below:

Can't access it right now, but I remember it was ☐ statistically significant ☐ null

Why not originally in paper/report: (select all that apply)

☐ Didn't collect that data ☐ Not included in write-up

☐ Null result ☐ Unfinished paper ☐ Missed by reviewer team

Other comments:

Primary pre-registered heterogeneity Agree with below? ☐ Yes ☐ No

If you disagree with this interpretation of estimates and/or believe that some dimensions of primary pre-registered heterogeneity are missing, add or modify them below. We will keep track of your edits.

Dimension	Effect	SE	Comments
Heterogeneity A	Not found	Not found	
Heterogeneity B	Not found	Not found	

If more rows are needed for primary heterogeneity tests, download from here.

Figure E.13: *Not Found* Results Report Sample Page

Note: Sent only to the *Engagement* and *Resources* treatment arms.

Questions about how to fill in this report card? See this brief explainer and this pre-filled example.

Additional Primary Heterogeneity Tests

Hypothesis #

Dimension	Effect	SE	Comments

Figure E.14: Extra Heterogeneity Rows Sample Page

RCT Registry Results Report

Main Results Not Found

Questions about how to fill in this report? See this brief explainer and this pre-filled example.

Here is your original registration, and the attachment details_lastname.pdf contains the details of how your registration was encoded.

Hypothesis # 10 (as interpreted from pre-registration)

Outcome 5 is not different in Treatment 3 relative to Treatment 2.

Mostly agree with this statement? ☑ Yes ☐ No **If not, please say why and add correct one:**

Please provide closest result to this pre-registered hypothesis. ***We could not find this result.***

$\hat{\beta}$ = 12345 , Units: kg/ha , SE = 54321 , N = 12345 , p-value = 54321

and/or attach document with results and provide the location below:

Can't access it right now, but I remember it was ☐ statistically significant ☐ null

Why not originally in paper/report: (select all that apply)

☐ Didn't collect that data ☑ Not included in write-up

☐ Null result ☐ Unfinished paper ☐ Missed by reviewer team

Other comments: Authors would write here the reason why it was not included in the write-up

Primary pre-registered heterogeneity Agree with below? ☑ Yes ☐ No

If you disagree with this interpretation of estimates and/or believe that some dimensions of primary pre-registered heterogeneity are missing, add or modify them below. We will keep track of your edits.

Dimension	Effect	SE	Comments
Heterogeneity A	12345	12345	
Heterogeneity B	12345	12345	

If more rows are needed for primary heterogeneity tests, download from here.

Figure E.15: *Not Found* Results Report Sample Page (as Filled by Authors)

Note: Sent only to the *Engagement* and *Resources* treatment arms.

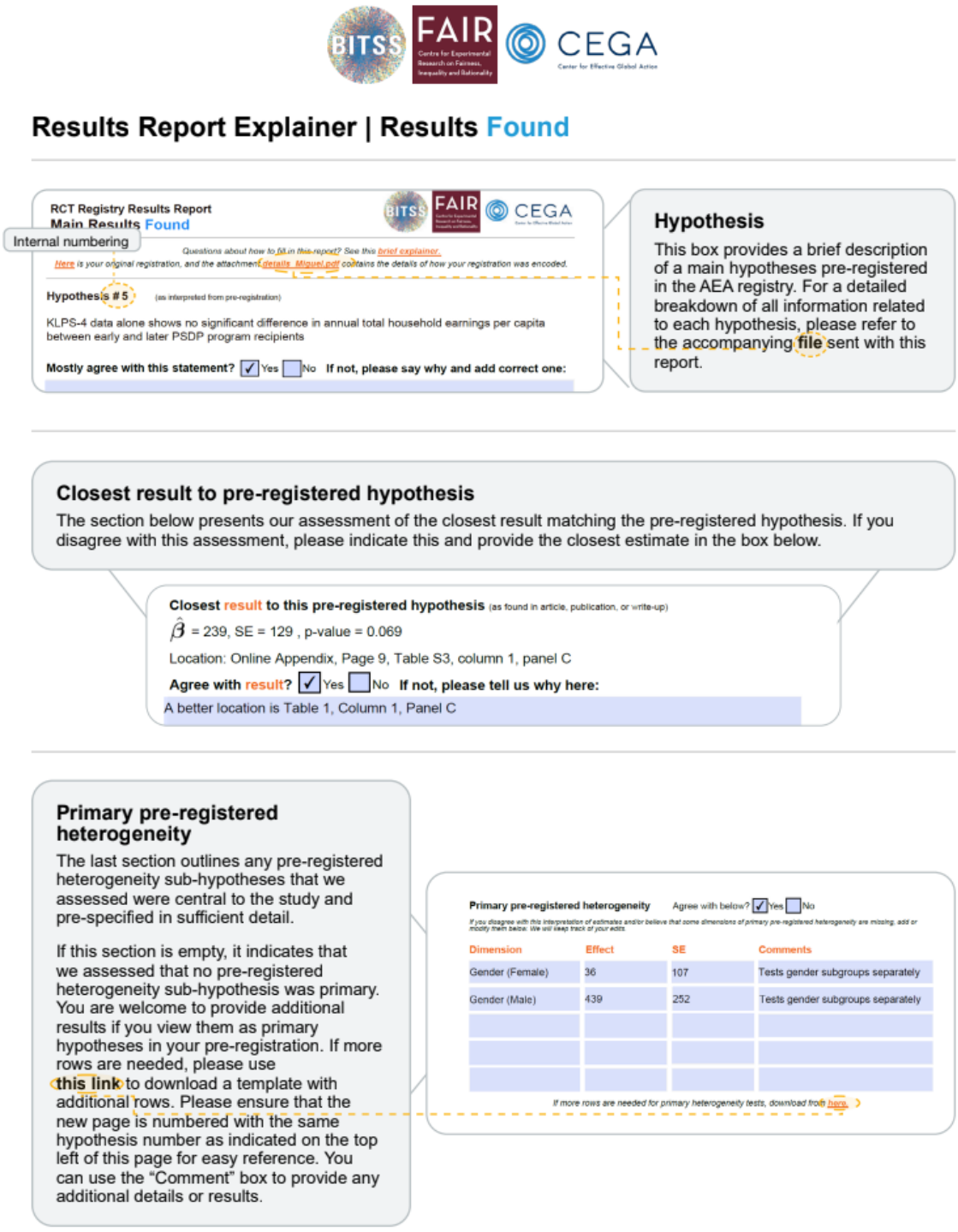


Figure E.16: *Found* Results Report Explainer

Results Report Explainer | Results Not Found

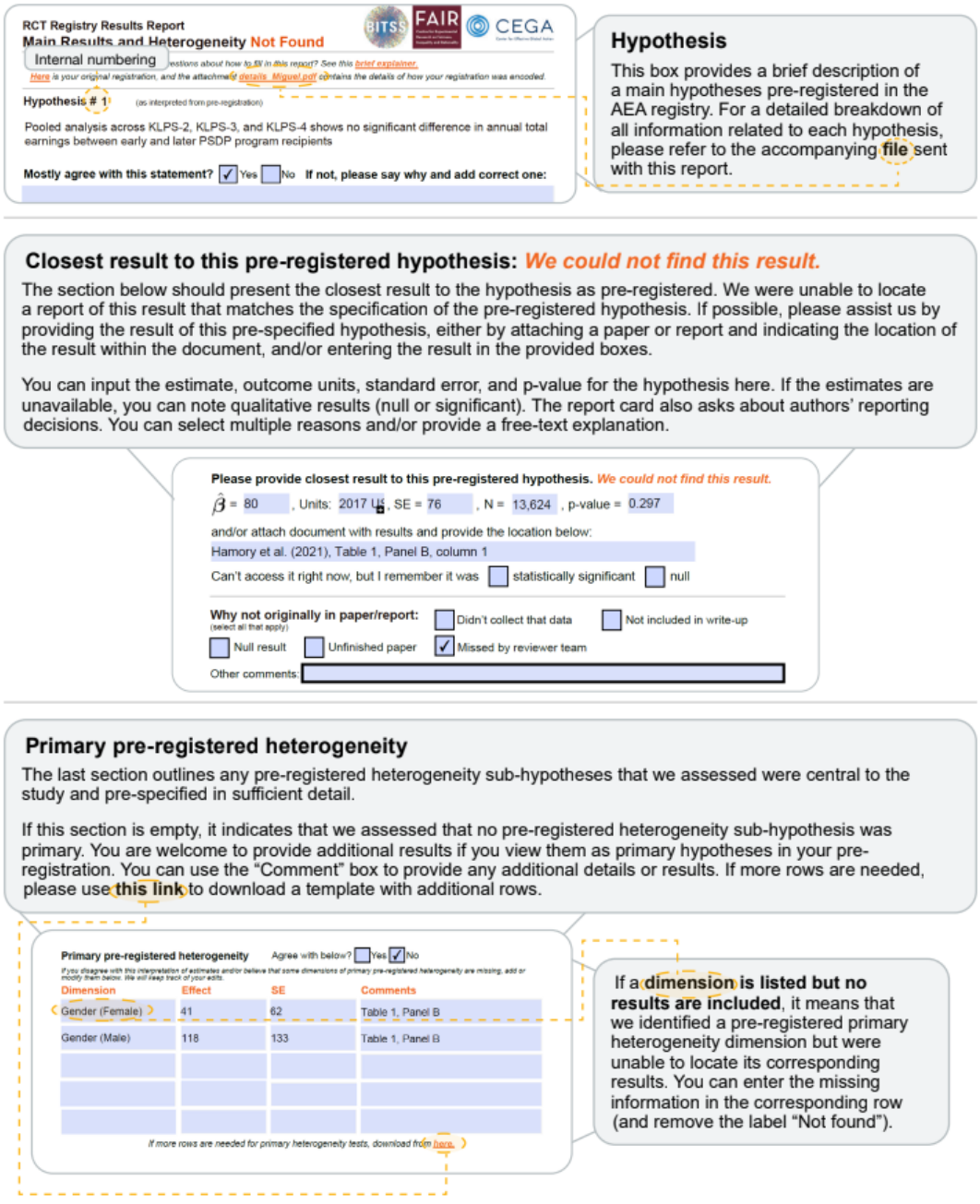


Figure E.17: *Not Found* Results Report Explainer

AEA RCT Trial Registration Summary # 0000

Title: The title goes here

https://www.socialscienceregistry.org/trials/0000

1 Populations

	label	country	Randomization	N	coverage
1	population1	GLO	Insert unit here	12345	The description of population 1 goes here

2 Outcomes

	label	unit_original	unit_analytical	description
1	outcome1	Insert unit here	Insert unit here	The description of outcome 1 goes here
2	outcome2	Insert unit here	Insert unit here	The description of outcome 2 goes here
3	outcome3	Insert unit here	Insert unit here	The description of outcome 3 goes here
4	outcome4	Insert unit here	Insert unit here	The description of outcome 4 goes here
5	outcome5	Insert unit here	Insert unit here	The description of outcome 5 goes here

3 Interventions

	label	description
1	intervention1	The description of intervention 1 goes here
2	intervention2	The description of intervention 2 goes here
3	intervention3	The description of intervention 3 goes here

4 Arms

	label	population	intervention
1	control	population1	NaN
2	treatment1	population1	intervention1
3	treatment2	population1	intervention2
4	treatment3	population1	intervention3

5 Arm Groups

	label	armlabel

Figure E.18: Encoded Pre-Registration (G_0) Sample, Page 1 of 2

Note: Each study in the *Engagement* and *Resources* arms received a version of this document for its own study, linked as "details-lastname.pdf" at the top of each Results Report page.

6 Hypotheses

	label	description
1	hypothesis1	The description of hypothesis 1 goes here $E[outcome1\|treatment1] - E[outcome1\|control] = 0$
2	hypothesis2	The description of hypothesis 2 goes here $E[outcome2\|treatment1] - E[outcome2\|control] = 0$
3	hypothesis3	The description of hypothesis 3 goes here $E[outcome3\|treatment1] - E[outcome3\|control] = 0$
4	hypothesis4	The description of hypothesis 4 goes here $E[outcome4\|treatment1] - E[outcome4\|control] = 0$
5	hypothesis5	The description of hypothesis 5 goes here $E[outcome5\|treatment1] - E[outcome5\|control] = 0$
6	hypothesis6	The description of hypothesis 6 goes here $E[outcome1\|treatment3] - E[outcome1\|treatment2] = 0$
7	hypothesis7	The description of hypothesis 7 goes here $E[outcome2\|treatment3] - E[outcome2\|treatment2] = 0$
8	hypothesis8	The description of hypothesis 8 goes here $E[outcome3\|treatment3] - E[outcome3\|treatment2] = 0$
9	hypothesis9	The description of hypothesis 9 goes here $E[outcome4\|treatment3] - E[outcome4\|treatment2] = 0$
10	hypothesis10	The description of hypothesis 10 goes here $E[outcome5\|treatment3] - E[outcome5\|treatment2] = 0$

7 Heterogeneity

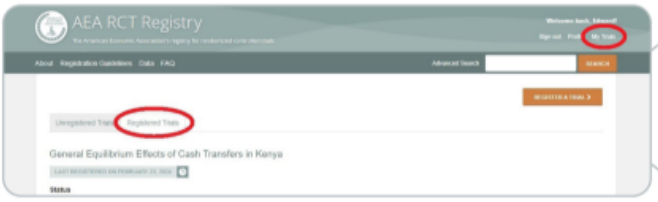
	subgroups	effect_type	hypothesis_id
1	Heterogeneity A	different_effects	hypothesis1, hypothesis2, hypothesis3, hypothesis4, hypothesis5, hypothesis6, hypothesis7, hypothesis8, hypothesis9, hypothesis10
2	Heterogeneity B	different_effects	hypothesis1, hypothesis2, hypothesis3, hypothesis4, hypothesis5, hypothesis6, hypothesis7, hypothesis8, hypothesis9, hypothesis10

8 Judgment Calls

	concerning	call
1	hypotheses	The description of the judgment call goes here

Figure E.18: Encoded Pre-Registration (G_0) Sample, Page 2 of 2

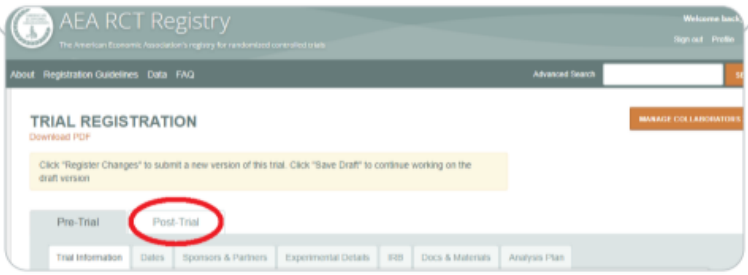
Post Trial Update Explainer



1. Open your preferred browser and go to https://www.socialscienceregistry.org/trials/.
2. Once you log in with your credentials, click on "My Trials" on the upper-right-hand-corner.
3. **Click on the "Registered Trials" tab (see screenshot).**
4. Select the study for which you wish to upload the results report. The corresponding page of the selected study should open.

5. On the right side, click on "Edit Trial". An editable version of the "Trial Registration" page should open.
6. **Select the "Post-Trial" tab (see screenshot).**

If the "Post-Trial" tab is **not** selectable, follow these steps:
(i). Under the "Pre-Trial" tab & "Dates" subtab, ensure that the "Trial End Date" is in the past (i.e. prior to current date)
(ii). Now click on the "Pre-Trial" tab & "Trial Information" subtab. Under the "Additional Trial Information" section, change the "Trial Status" to "Completed".
(iii) Re-do step 6, i.e. click on the "Post-Trial" tab (see screenshot).



7. Under the "Post-Trial" tab, select the "Reports, Papers & Other Materials" subtab.
8. **Under the "Reports & Other Materials" section, select "+ Upload a document" (see screenshot).**
9. Upload your results report.
10. On the bottom of this page, click on "Register Changes" to save the edits.

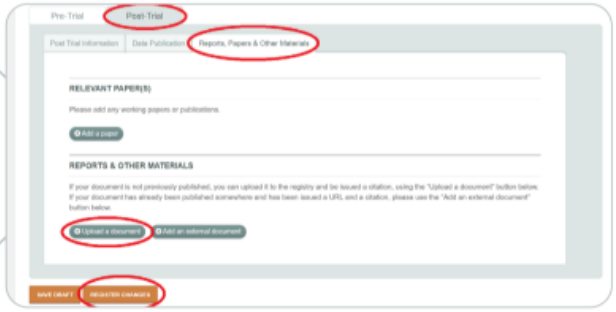
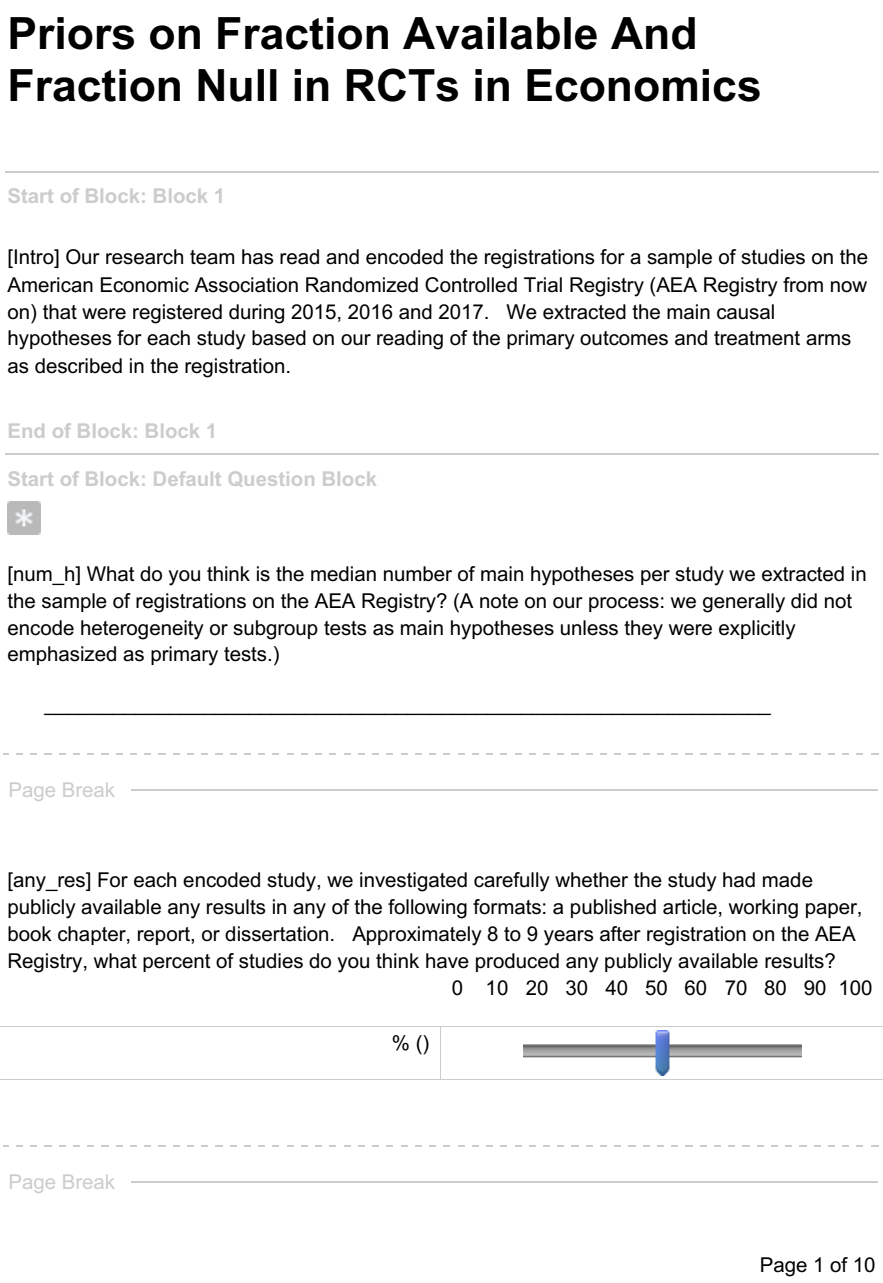


Figure E.19: AEA RCT Registry Post Section Update Explainer

F Expert study on The Social Science Prediction Platform



Priors on Fraction Available And Fraction Null in RCTs in Economics

Start of Block: Block 1

[Intro] Our research team has read and encoded the registrations for a sample of studies on the American Economic Association Randomized Controlled Trial Registry (AEA Registry from now on) that were registered during 2015, 2016 and 2017. We extracted the main causal hypotheses for each study based on our reading of the primary outcomes and treatment arms as described in the registration.

End of Block: Block 1

Start of Block: Default Question Block

[num_h] What do you think is the median number of main hypotheses per study we extracted in the sample of registrations on the AEA Registry? (A note on our process: we generally did not encode heterogeneity or subgroup tests as main hypotheses unless they were explicitly emphasized as primary tests.)

Page Break

[any_res] For each encoded study, we investigated carefully whether the study had made publicly available any results in any of the following formats: a published article, working paper, book chapter, report, or dissertation. Approximately 8 to 9 years after registration on the AEA Registry, what percent of studies do you think have produced any publicly available results?

0 10 20 30 40 50 60 70 80 90 100

% ()

Page Break

Page 1 of 10

Figure F.1: Survey instrument (page 1/10)

[at_least_one] It is possible that even when results have been produced for a study, none of the available results correspond to the main hypotheses that were registered on the AEA Registry. (For example, the publicly available results may use a different primary outcome and not present any of the ones that were registered.) Among the studies that did produce any results (you answered ${any_res/TotalSum}%), what percent do you think **do not** report results for any of the registered main hypotheses 8-9 years after the study was registered?

0 10 20 30 40 50 60 70 80 90 100

% ()

Page Break

[Fraction available] Your responses imply that $e{round(${any_res/TotalSum} * (100 - ${at_least_one /TotalSum}) /100, 1)}% of the registered studies have made publicly available results for **at least one** of the registered hypotheses (${any_res/TotalSum}% * (100% - ${at_least_one /TotalSum}%) = $e{round(${any_res/TotalSum} * (100 - ${at_least_one /TotalSum}) /100, 1)}%). Among the studies that do report results for at least one of the registered main hypotheses, what do you believe is the percent of registered hypotheses per study that is reported on average?

0 10 20 30 40 50 60 70 80 90 100

% ()

Page Break

Figure F.2: Survey instrument (page 3/10, page 2/10 is blank)

[Fraction Null] Think about the same group of studies as the previous question (the $e{round(${any_res/TotalSum} * (100 - ${at_least_one /TotalSum}) /100, 1)}% that you estimated reported results for **at least one** hypothesis). Now focus on the registered hypotheses where a result was reported. What percent of these reported results per study do you think were null results (by which we mean a p-value >= 0.05) on average? For a made-up example, if a study registered 15 main hypotheses, reported results for 9 of them, and x of those were null results, then the percent of reported null results for this study would be (x/9)*100 (x of 9 reported results).

0 10 20 30 40 50 60 70 80 90 100

% ()



Page Break

[Summary] Here is a summary of your answers about what you think is publicly available from a study 8-9 years after registration: Percent of studies reporting any result: ${any_res/TotalSum}% Percent of studies reporting results for at least one registered hypothesis: $e{ round(${any_res/TotalSum} * (100 - ${at_least_one /TotalSum}) /100, 1)}% (= ${any_res/TotalSum}% * (100% - ${at_least_one /TotalSum}%)) Average percent of hypotheses reported: $e{round(${Fraction available/TotalSum} * ${any_res/TotalSum}/100 * (100 - ${at_least_one /TotalSum}) /100 , 1)}% (= ${Fraction available/TotalSum}% * $e{round(${any_res/TotalSum}/100 * (100 - ${at_least_one /TotalSum}) ,1)}%) Conditional on a result being reported: Average percent of hypotheses reported as null results: $e{round(${Fraction Null/TotalSum}, 1) }% of those studies that report at least one registered hypothesis (your $e{ round(${any_res/TotalSum} * (100 - ${at_least_one /TotalSum}) /100, 1)}%)

Please confirm or go back and revise your answers.

Page Break

Figure F.3: Survey instrument (page 4/10)

[PAP] Comparing studies that have a Pre-Analysis Plan (PAP) in the AEA registry with studies that do not have one. Do you think that there will be systematic differences in: Average percent of hypotheses reported? Conditional on a result being reported: Average percent of hypotheses reported as null results?

	Percent with PAP is **lower** than percent without a PAP (1)	No systematic differences (2)	Percent with PAP is **higher** than percent without a PAP (3)
Reported hypotheses (1)	○	○	○
Hypotheses reporting null results (2)	○	○	○

[LMIC] Comparing studies from development economics, i.e., that focus on a study population in low and middle income countries (using the World Bank definition), with studies focusing on populations in high income countries. Do you think that there will be systematic differences in: Average percent of hypotheses reported? Conditional on a result being reported: Average percent of hypotheses reported as null results?

	Percent for studies in LMIC is **lower** than percent not in LMIC (1)	No systematic differences (2)	Percent for studies in LMIC is **higher** than percent not in LMIC (3)
Reported hypotheses (1)	○	○	○
Hypotheses reporting null results (2)	○	○	○

Figure F.4: Survey instrument (page 5/10)

[Published] Comparing studies that have been published in a peer-reviewed journal with studies that have not been published (yet) but have a manuscript available (i.e., a working paper, dissertation or report). Do you think that there will be systematic differences in: Average percent of hypotheses reported? Conditional on a result being reported: Average percent of hypotheses reported as null results?

	Percent in studies with publication **lower** than percent in studies with unpublished manuscripts (1)	No systematic differences (2)	Percent in studies with publication **higher** than percent in studies with unpublished manuscripts (3)
Reported hypotheses (1)	○	○	○
Hypotheses reporting null results (2)	○	○	○

[Top 5] Comparing studies that have been published in a top-5 economics journal with studies that have been published in other peer-reviewed journals. Do you think that there will be systematic differences in: Average percent of hypotheses reported? Conditional on a result being reported: Average percent of hypotheses reported as null results?

	Percent in studies with top-5 publication **lower** than percent in studies with other non-top-5 publication (1)	No systematic differences (2)	Percent in studies with top-5 publication **higher** than percent in studies with other non-top-5 publication (3)
Reported hypotheses (1)	○	○	○
Hypotheses reporting null results (2)	○	○	○

Page Break

Figure F.5: Survey instrument (page 6/10)

End of Block: Default Question Block

Start of Block: Block 3

Q35 Have you ever posted a study in the AEA Registry?

- ○ No (1)
- ○ Yes (2)

Display this question:
If Have you ever posted a study in the AEA Registry? = Yes

Q36 If yes, what was the first year in which you posted on the Registry?

_____

Q37 Have you seen a presentation by the research team on this study? The team is: Fernando Hoces de la Guardia, Edward Miguel, Viviane Silva da Rocha, Gufran Pathan, Erik O. Sørensen, Bertil Tungodden.

- ○ No (1)
- ○ Yes (2)

End of Block: Block 3

Start of Block: Block 2

Q39 This final section contains questions about your background, which are optional.

Figure F.6: Survey instrument (page 7/10)

Q22 What describes you best?

- Graduate student (Master level or PhD) (1)
- Faculty (2)
- Post-doc (3)
- Non-academic researcher (4)
- Other (5)

Display this question:
If What describes you best? = Other

Q36 If other, please describe:

_____

Display this question:
If What describes you best? = Graduate student (Master level or PhD)
Or What describes you best? = Faculty
Or What describes you best? = Post-doc

Q24 Discipline:

- Economics (1)
- Political Science (2)
- Psychology (3)
- Sociology (4)
- Other (5)

Figure F.7: Survey instrument (page 8/10)

Display this question:
If Discipline: = Other

Q37 If other, please describe:

_____

Display this question:
If What describes you best? = Graduate student (Master level or PhD)
Or What describes you best? = Faculty
Or What describes you best? = Post-doc

Q25 Title:

- ○ Professor (1)
- ○ Assistant Professor (2)
- ○ Post-doc (3)
- ○ Researcher (4)
- ○ PhD Student (5)
- ○ Masters Student (6)
- ○ Associate Professor (7)
- ○ Other (8)

Display this question:
If Title: = Other

Q38 If other, please describe:

_____

Figure F.8: Survey instrument (page 9/10)

Display this question:
If What describes you best? = Non-academic researcher
Or What describes you best? = Other

Q28 What type of organization do you currently work for?

- ○ Academia/research institute (1)
- ○ Government (2)
- ○ International organization (3)
- ○ NGO (4)
- ○ Private sector (5)
- ○ Other (6)

Display this question:
If What type of organization do you currently work for? = Other

Q29 If other, please describe:

_____

Display this question:
If What describes you best? = Non-academic researcher
Or What describes you best? = Other

Q31 What is the highest level of education you have attained?

- ○ Bachelor's degree (1)
- ○ Masters degree (2)
- ○ PhD (3)
- ○ Other graduate/professional degree (4)

End of Block: Block 2

Page 10 of 10

Figure F.9: Survey instrument (page 10/10)