Pre-registration, Reporting Guidelines, and Publication Patterns in Economics

Fernando Hoces de la Guardia Edward Miguel Gufran Pathan
Viviane Rocha Erik Ø. Sørensen Bertil Tungodden

BITSS Annual Meeting
Berkeley
February 27, 2025

Outline

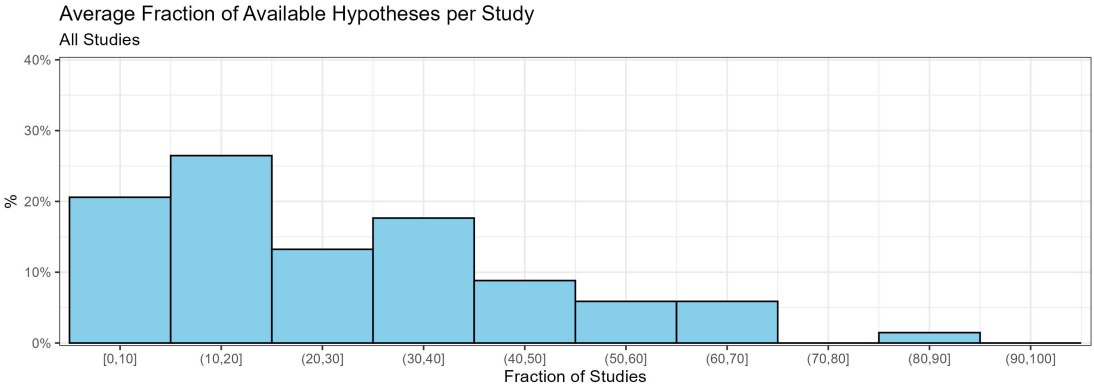
Background and Contribution

- Work measuring publication bias & selective reporting directly and indirectly [Franco et. al., 2014, Akker et. al., 2023, Andrews and Kasy, 2019, Ofosu and Posner, 2023]
- Reporting and Data standards for research in general [CONSORT 2010, 2019, DDI Alliance, 2000] and for RCTs in economics [Cavanagh et. al., 2023]

- **This project** examines how studies on the AEA Registry are reported:
  - Are the results for registered hypotheses publicly **available** 8-9 years later?
  - What fraction of these hypotheses are **null results**?
  - Does the fraction of reported nulls vary across **publication outlets**?

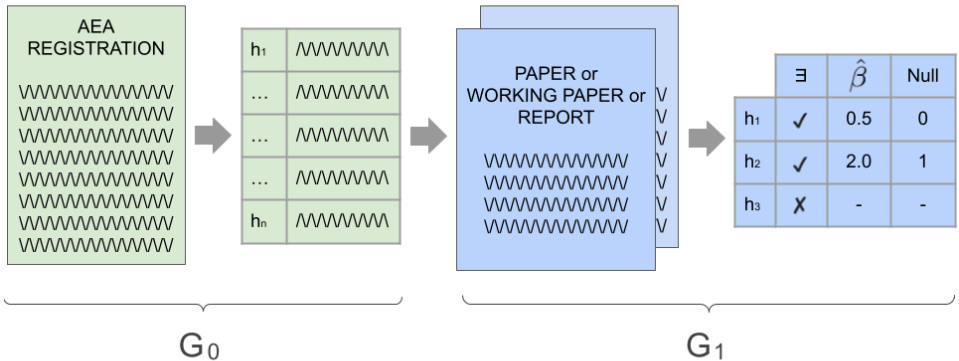
What fraction of hypotheses are available on average? (Forecasts)

What fraction of hypotheses are available on average? (Forecasts)



Outline

Our Approach

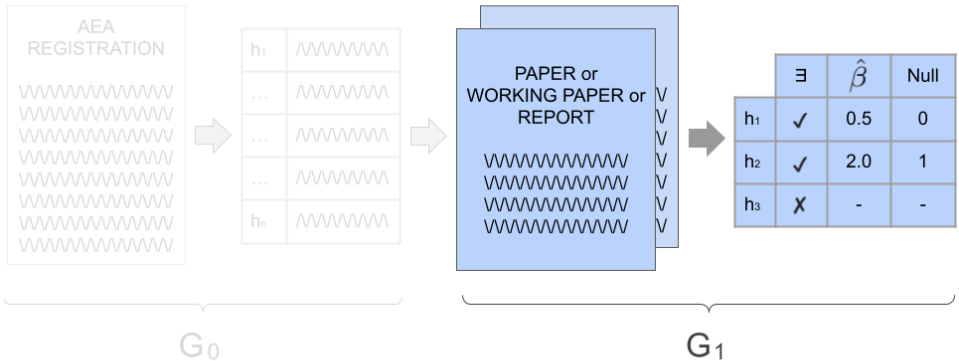


Encoding of Registrations (G_0)

Encoding of Registrations (G_0)

- We developed a tool and encoding system to standardize registrations (based on Cavanagh et al. 2023 and DDI Alliance 2021)
- We added two new elements: hypotheses and heterogeneity tests.
  - Simple difference: $E[Y_i \mid T_i = 1] - E[Y_i \mid T_i = 0] = 0$
  - Double differences: $E[Y_i \mid T_i = 1, t = 1] - E[Y_i \mid T_i = 1, t = 0] - E[Y_i \mid T_i = 0, t = 1] - E[Y_i \mid T_i = 0, t = 0] = 0$
  - Joint tests: $Y_i = \alpha_0 + \alpha_1 X_1 + \alpha_2 X_2 + \alpha_3 X_1 X_2 + \epsilon$, where: $\alpha_1 = \alpha_2 = 0$
  - Regression Estimate: $Y_i = \alpha + \beta \cdot \text{Treatment}_i + \gamma \cdot X_i + \epsilon_i$, where: $\beta = 0$
- Example: "Students' 2013 test scores in Kiswahili are not different for the students in capitation-grant schools compared to that of students in control schools ($E[kiswahili\_2013|input] - E[kiswahili\_2013|control] = 0$)"

Encoding of Papers and/or Reports (G_1)



Encoding of Papers and/or Reports (G_1)

Once the G_0 is encoded, the team searches for academic (published and working papers) and non-academic (policy reports, etc.) output that documents the pre-registered hypotheses' results and records them in a standardized format, i.e. G_1.

For each hypotheses found we extract:

- Arm, outcome, and hypothesis status: available and modification assessments.
- For each hypothesis, we record the estimate, standard errors, p-values (i.e., null), standardized effect size (when available), and total sample size.
- In addition, the encoder provides a subjective assessment of the ease of extraction: easy, medium, and hard.

Encoding Approach

- **Coders:** 3 research assistants (pre-docs) with Masters in Development Economics.
- **Tool Development:** A simple XML-based data entry tool was created to record key fields (G_0).
- **Time per Study:** On average, around 11 hours to encode G_0 and G_1.
- **Sample:** Currently at 230 studies encoded. All of 2015 and part of 2016, minus duplicate main PIs. ▸ Criteria for exclusion based on timeline
- Extensive quality control was conducted. ▸ Details here

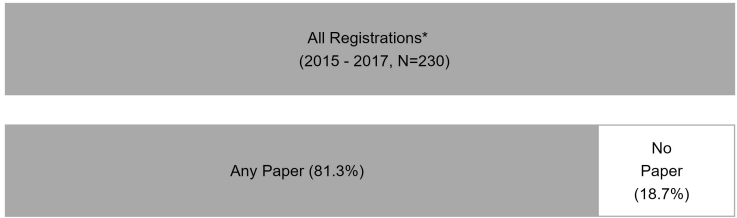
Primary Outcomes

- **Fraction Completely Available** (Y_{1i})**:** Fraction of hypotheses that are completely available per study. Requires:
  - numerical estimate exists for the hypothesis as in G_0
  - anywhere in the paper, appendix, or any other public record
  - regardless of the level of effort spent by the encoder to find it
  - allowing for modification we judge consistent with the pre-registration
- **Fraction of Results Reported Null** (Y_{2i})**:** Fraction of reported hypotheses with $|z_{hi}| \leq 1.96$.
  - Different denominator than Y_{1i}

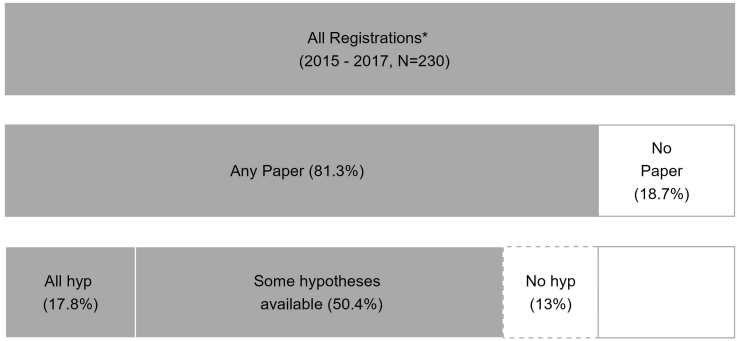
Outline

Measuring The File Drawer

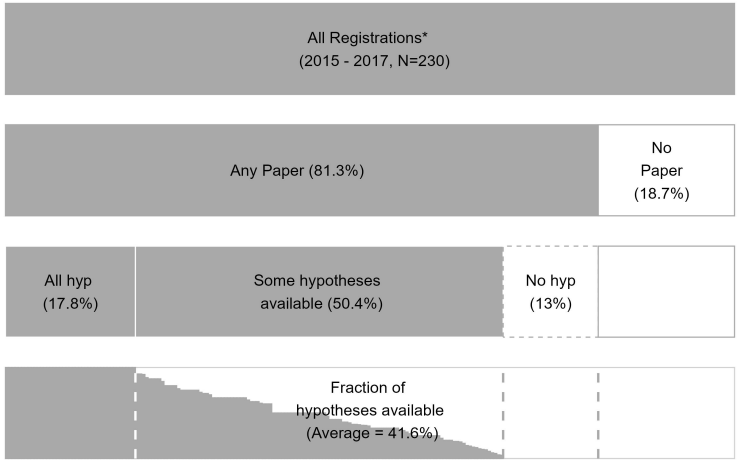
Measuring The File Drawer



Measuring The File Drawer



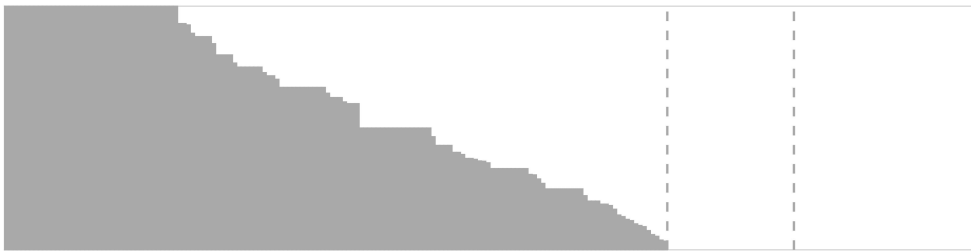
Measuring The File Drawer



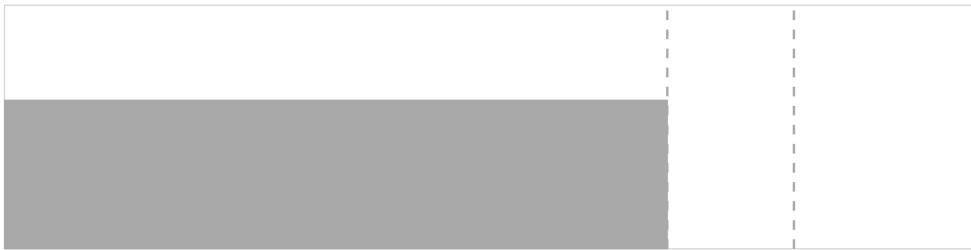
$\Rightarrow$ On average 58.4% of hypotheses are missing per paper.

$\Rightarrow$ If sample is constrained to registrations with paper, this fraction falls to 48.8%.

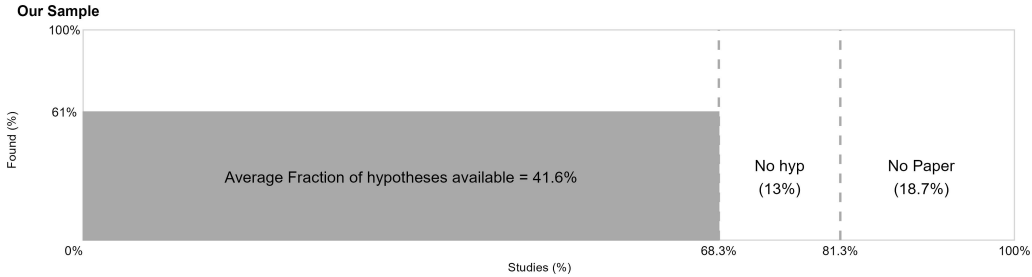
Comparing Estimates with Expert Forecasts



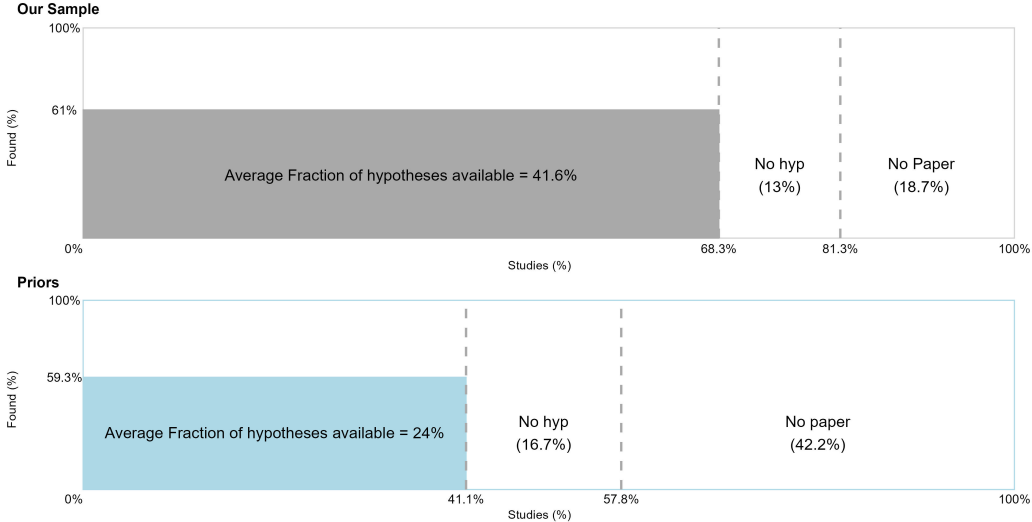
Comparing Estimates with Expert Forecasts



Comparing Estimates with Expert Forecasts: All Results

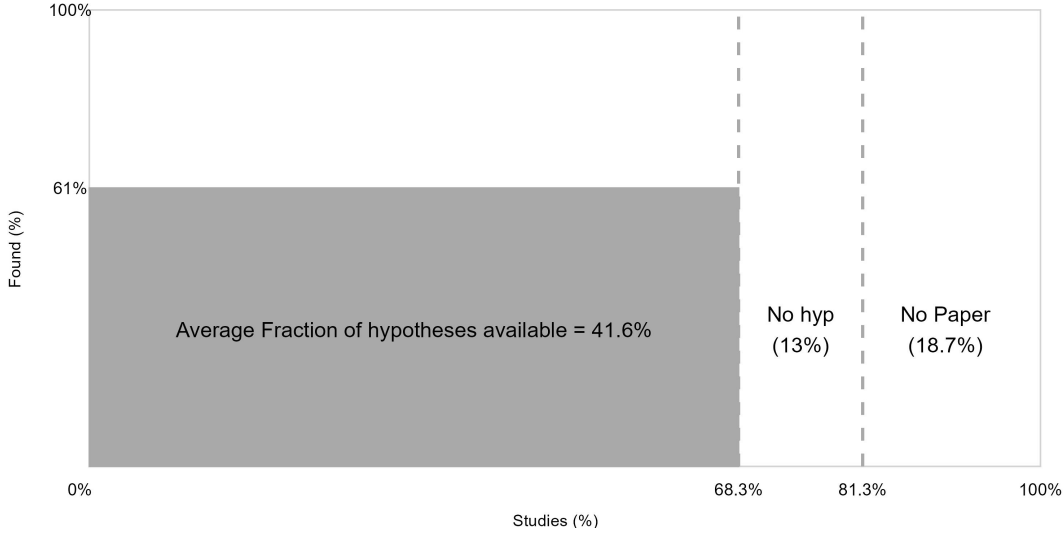


Comparing Estimates with Expert Forecasts: All Results

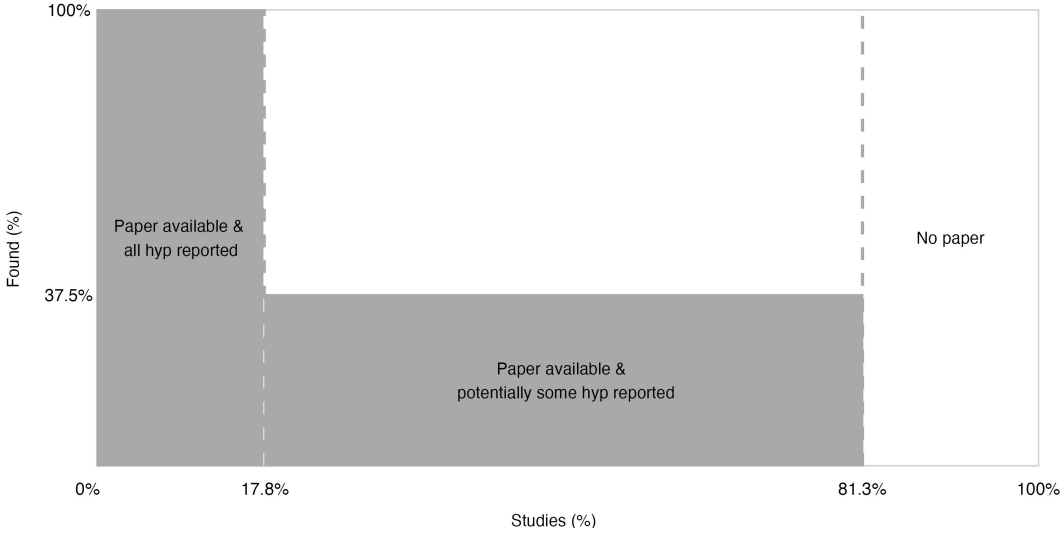


$\Rightarrow$ Forecasters predict that on average 76% of hypotheses will be missing per paper. Step by Step

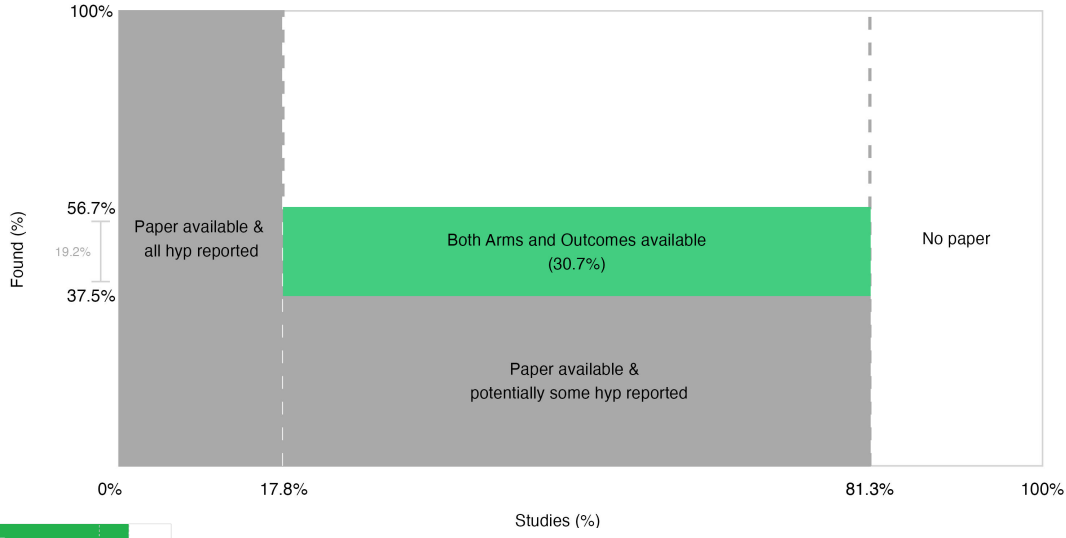
Characterizing Missing Hypotheses



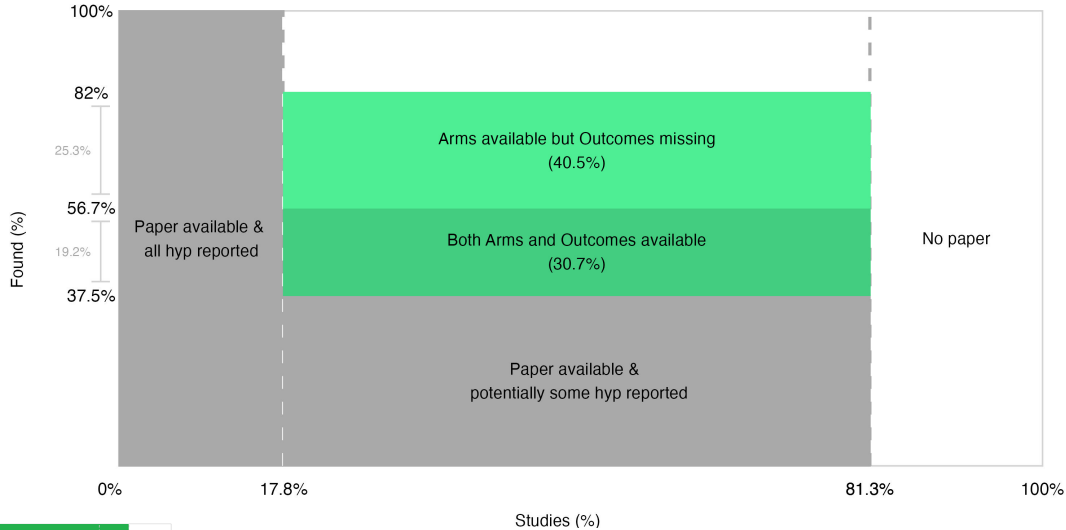
Characterizing Missing Hypotheses



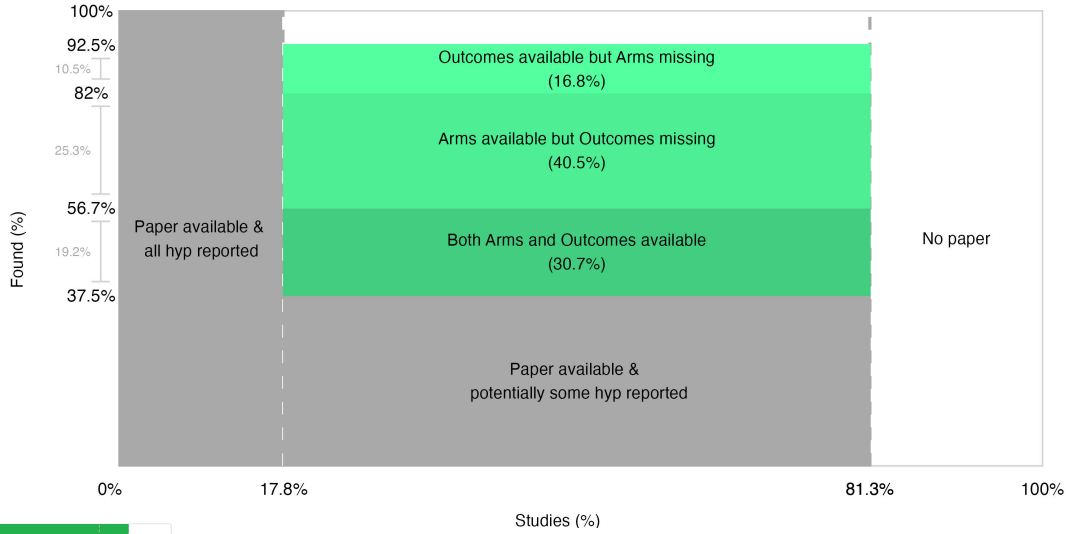
Characterizing Missing Hypotheses



Characterizing Missing Hypotheses



Characterizing Missing Hypotheses



Characterization of Missing Hypotheses With Paper

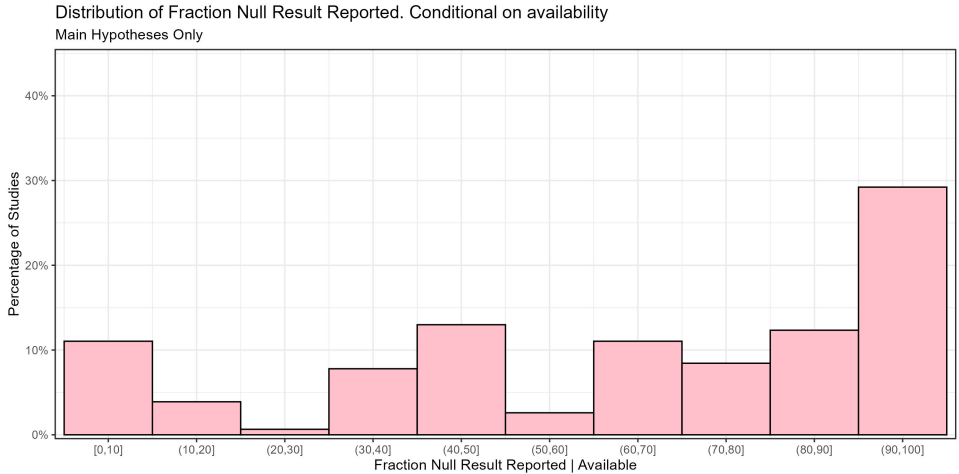
Possible explanations for missing:

- Study was never carried out? Not a big part of the story: 51% of hypotheses are available when a paper is found. For this missing hypotheses:
  - 30.7% have both arms and outcomes available, but pre-registered hypothesis is not reported.
  - Outcome not collected or nor reported? Of the hypotheses that have a study, 40.5% have at least one outcome missing
  - The remaining hypotheses that have a study, 16.8% have at least one arm missing

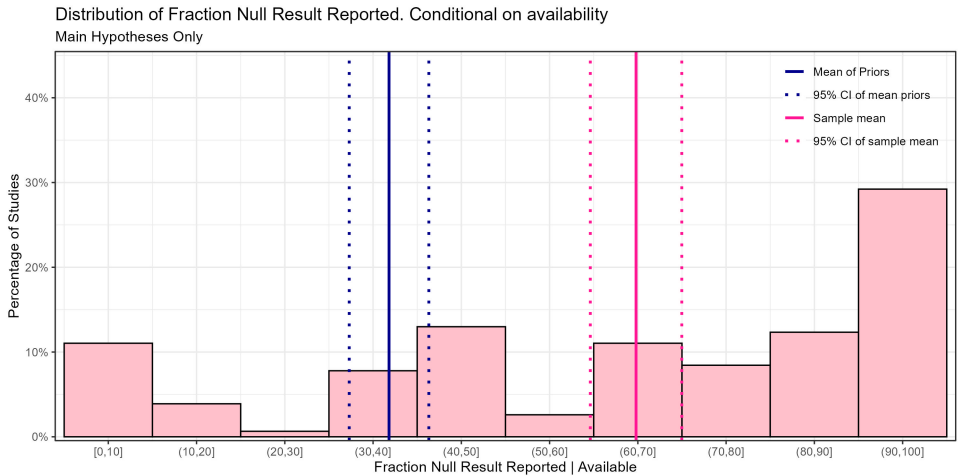
▸ Comparison of Fraction Available for different categories

▸ Alternative definitions of Availability

Distribution of Fraction Null Results Reported



Distribution of Fraction Null Results Reported



Reported Null and Fraction Available by Publication Type

Study	Frac. Null* Hypotheses per Study	Frac Available Hypotheses per Study	N
Not written	–	0%	43
Written, not published	70%	51%	68
Published, not top 5	64%	53%	93
Published top 5	52%	44%	26

* Conditional on hypothesis being available

▸ Comparison of Fraction Null for different categories

Outline

Conclusions

1. A substantial fraction of registered hypotheses have no available results (58%). Although large, this fraction is smaller than predicted by forecasters, at 76%.
2. This gap, between what is registered vs. reported, is not driven by unfinished studies: conditioning on paper availability, 83% of the gap remains (49% of hypotheses are missing).
3. Systematic patterns suggesting publication bias: null results are less likely to be published, and less likely to be published in top-5 economics journals.

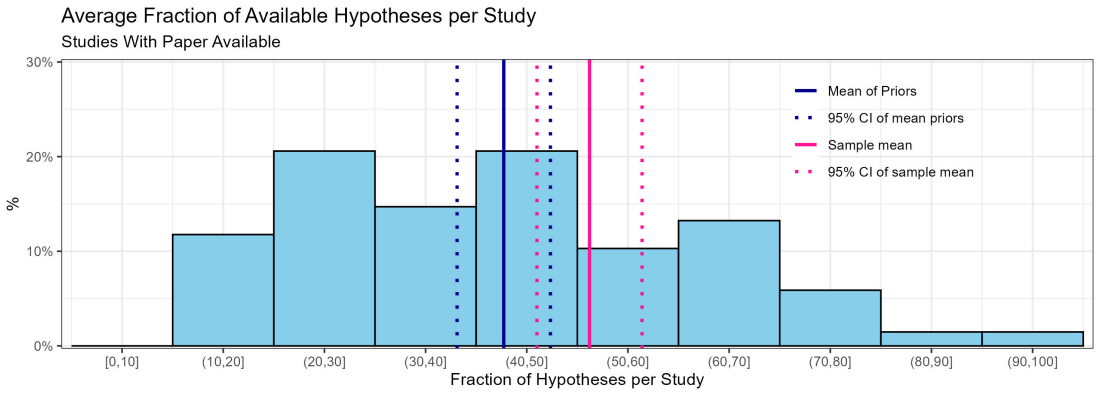
Next Steps

- Write the paper.
- We are currently running an RCT to learn best ways to recover the missing hypotheses
- Later this year: compare our human encodings with LLMs

Thank You.

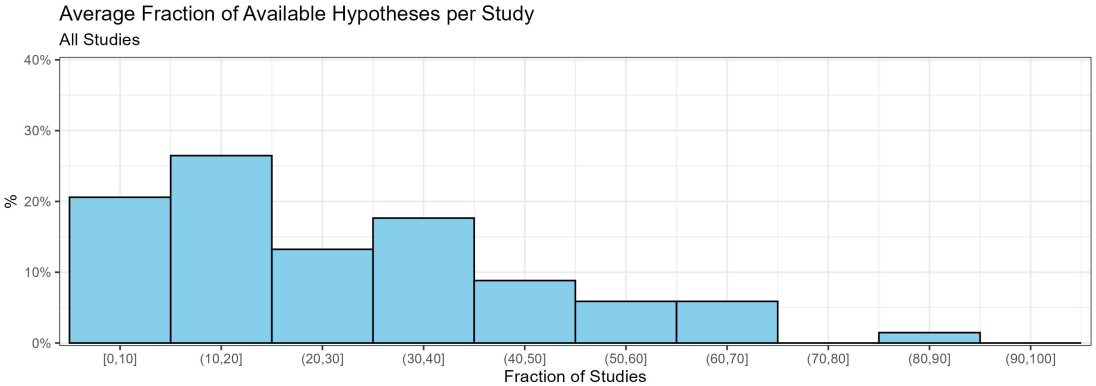
fhoces@berkeley.edu

Conditional On Paper: Average Fraction of Hypotheses Available

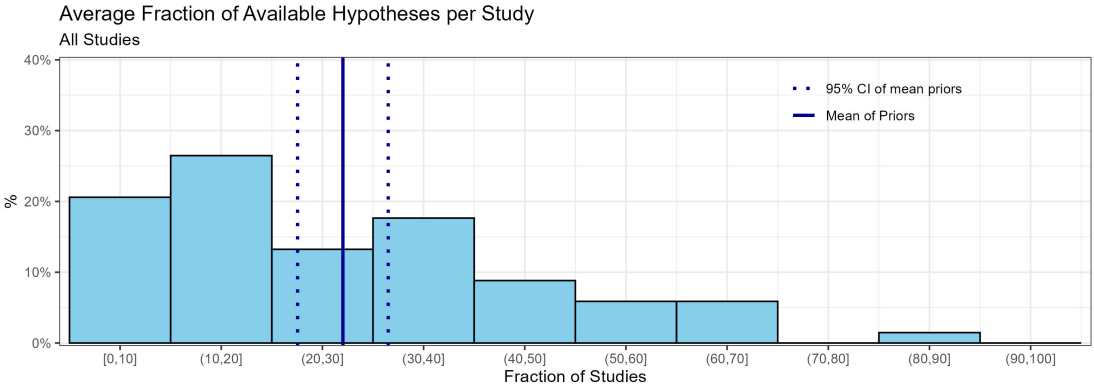


▸ Conditional on At Least One Result Available

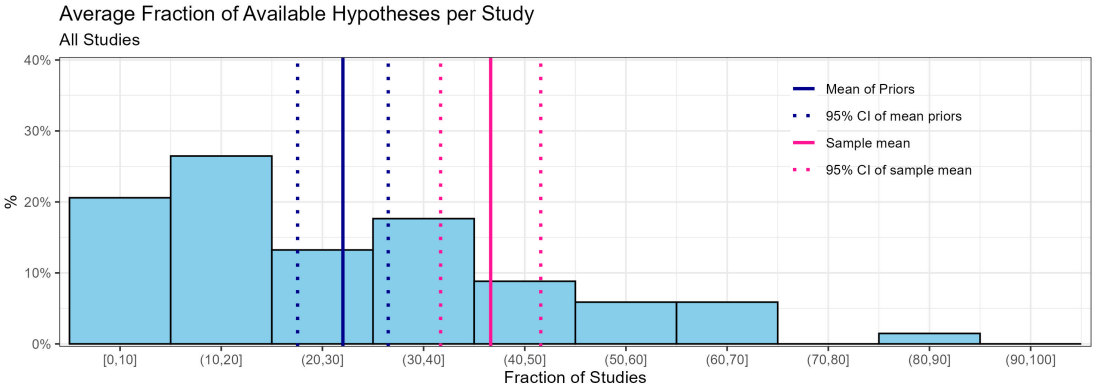
Unconditional: Fraction of Hypotheses Available Per Registration



Unconditional: Fraction of Hypotheses Available Per Registration



Unconditional: Fraction of Hypotheses Available Per Registration



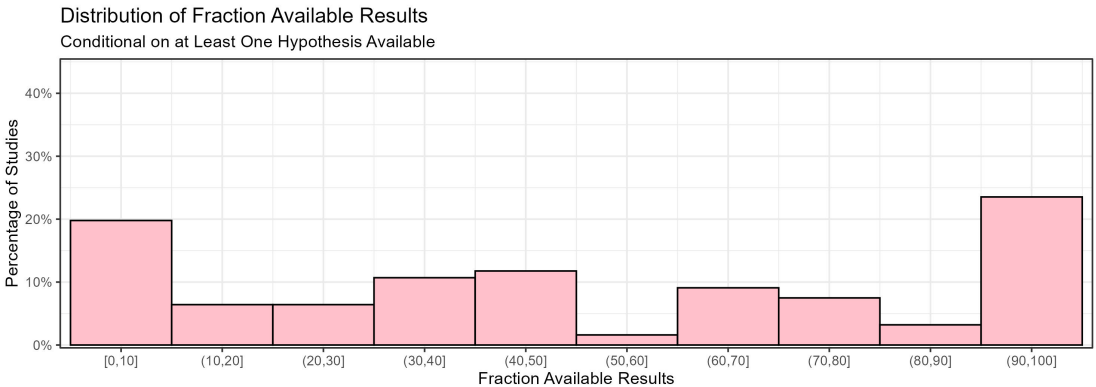
▸ Sample distribution

Priors and Results on Availability

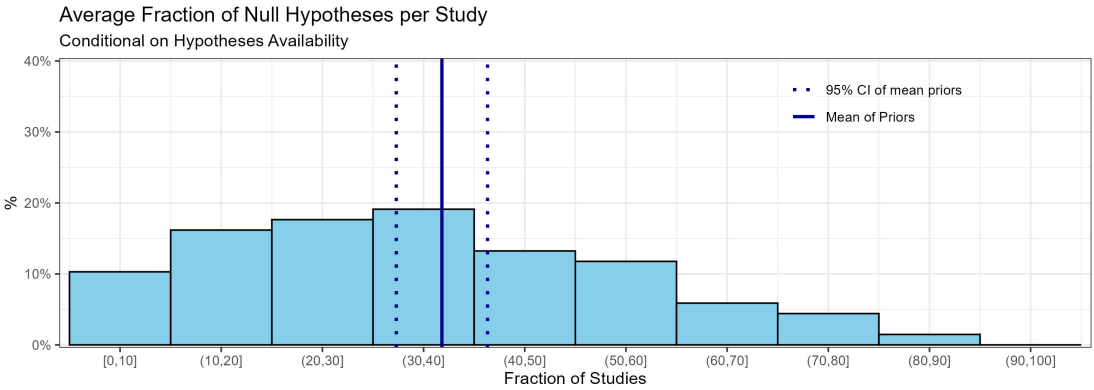
	Literature	Priors	Our Sample		
			Mean	Studies	Hypoth.
Panel A: Studies					
(1) Paper available	80%	57.8%	81.3%	230	2889
(2) At least one result available, conditional on paper available	–	71.2%	84%	187	2406
Panel B: Average fraction of main hypotheses available per study					
(3) Conditional on at least one result available	–	59.3%	61%	157	2067
(4) Conditional on paper available	–	42.7%	51.2%	187	2406
(5) Unconditional (All Studies)	–	27%	41.6%	230	2889

▸ Alternative Definitions of Availability

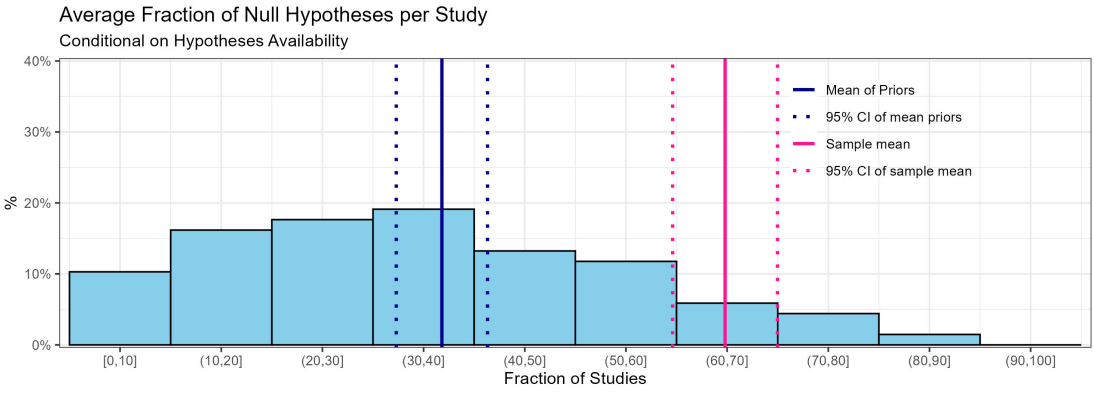
Fraction of Available Hypotheses: Conditional on Paper



Conditional on Available: Fraction of Null Results



Conditional on Available: Fraction of Null Results



▸ Sample distribution

Comparison of Frac. Avail. (Y_1) Across Different Dimensions

Is the fraction of available hypotheses higher, lower, or the same across each of the following dimensions?

Dimension	Forecasts		
	Lower	Same	Higher
1 Results written-up			
2 **Results published**†	0.37	0.28	0.35
3 **Results published in Top 5**‡	0.25	0.46	0.29
4 **Registration has PAP**	0.15	0.16	0.69
5 **Study in LMIC**	0.16	0.56	0.28
6 More than 10 hypothesis			
7 Most hypothesis are detailed			

†: n=187 , ‡: n=119 , All other dimensions: n=230

Back

Comparison of Frac. Avail. (Y_1) Across Different Dimensions

Is the fraction of available hypotheses higher, lower, or the same across each of the following dimensions?

Dimension	Forecasts			Estimates		
	Lower	Same	Higher	Coeff	SE	Mean
1 Results written-up				0.51	0.06	0
2 **Results published**†	0.37	0.28	0.35	-0	0.06	0.51
3 **Results published in Top 5**‡	0.25	0.46	0.29	-0.09	0.08	0.53
4 **Registration has PAP**	0.15	0.16	0.69	0.02	0.05	0.41
5 **Study in LMIC**	0.16	0.56	0.28	-0.13	0.05	0.51
6 More than 10 hypothesis				-0.14	0.05	0.47
7 Most hypothesis are detailed				0.03	0.05	0.4

†: n=187 , ‡: n=119 , All other dimensions: n=230 Back

Comparison of Frac. Null (Y_2) Across Different Dimensions

Is the fraction of null results lower, the same, or higher in each of the following dimensions?

Dimension	Forecasts		
	Lower	Same	Higher
1 **Results published**†	0.54	0.18	0.28
2 **Results published in Top 5**‡	0.54	0.29	0.16
3 **Registration has PAP**	0.18	0.13	0.69
4 **Study in LMIC**	0.19	0.53	0.28
5 More than 10 hypothesis			
6 Most hypothesis are detailed			

Comparison of Frac. Null (Y_2) Across Different Dimensions

Is the fraction of null results lower, the same, or higher in each of the following dimensions?

Dimension	Forecasts			Estimates		
	Lower	Same	Higher	Coeff	SE	Mean
1 **Results published**†	0.54	0.18	0.28	-0.08	0.06	0.7
2 **Results published in Top 5**‡	0.54	0.29	0.16	-0.12	0.08	0.64
3 **Registration has PAP**	0.18	0.13	0.69	0.08	0.05	0.61
4 **Study in LMIC**	0.19	0.53	0.28	-0.05	0.06	0.68
5 More than 10 hypothesis				0.09	0.05	0.61
6 Most hypothesis are detailed				0.05	0.06	0.63

†: n=187 written up, ‡: n=119 published, All other dimensions: n=230

back

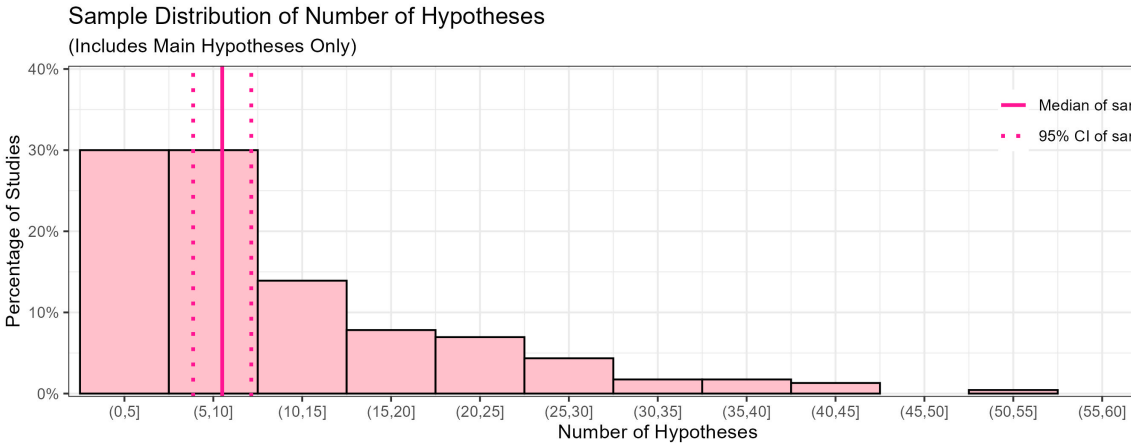
Reported Null and Fraction Available by Publication Type

Study	Frac. Null* Hypotheses per Study	Frac Available Hypotheses per Study	N
Not written	–	0%	43
Written, not published	70%	51%	68
Published, not top 5	64%	53%	93
Published top 5	52%	44%	26

* Conditional on hypothesis being available

▸ Hypotheses Level Table

Sample Distribution: Number of Hypotheses



back

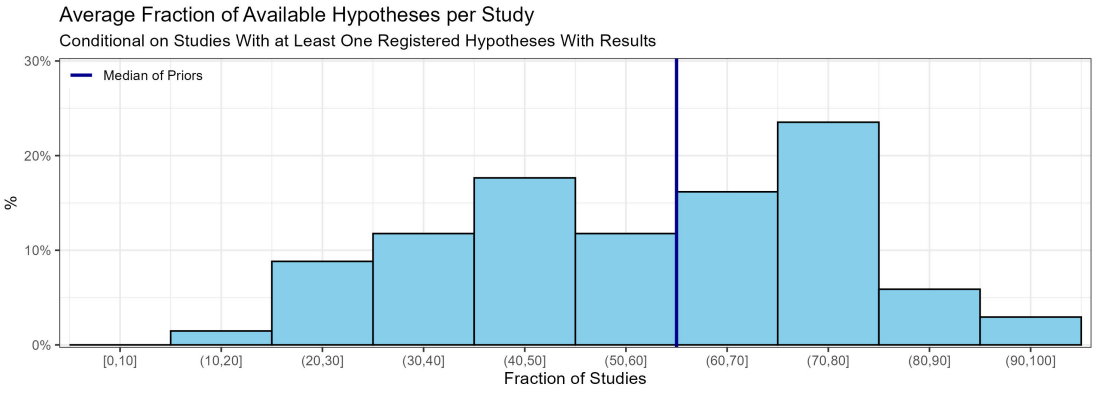
Priors: Number of Hypotheses

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.

What do you think is the **median number of main hypotheses per study** we extracted in the sample of registrations on the AEA Registry?

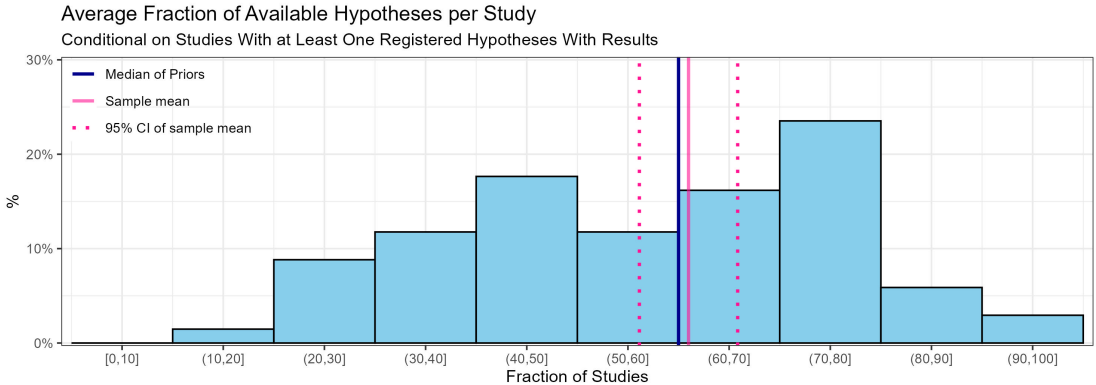
(A note on our process: we generally did not encode heterogeneity or subgroup tests as main hypotheses unless they were explicitly emphasized as primary tests.)

Conditional on at Least One Hypotheses Available: Avg. Fraction Available

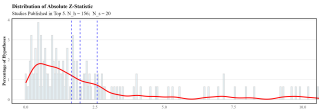


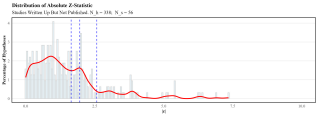
back

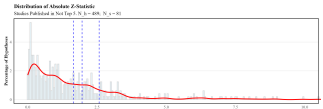
Average Fraction of Hypotheses Available Per Registration



Distribution of Absolute Z-Statistics, By Publication Type

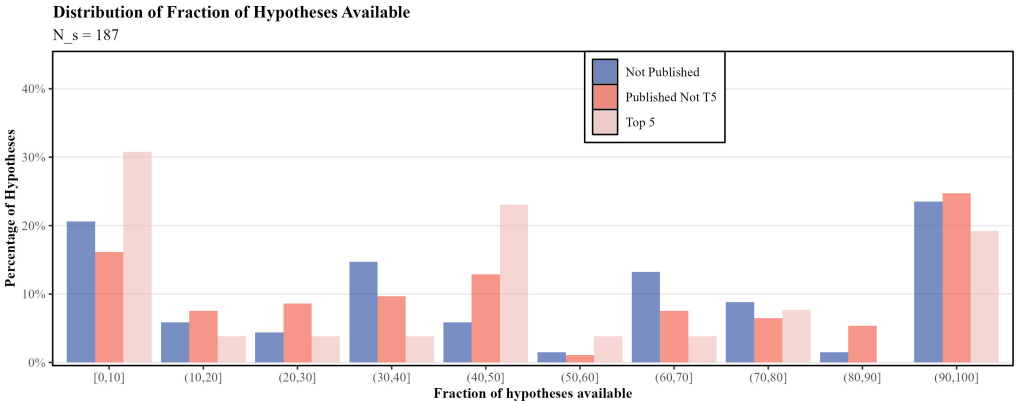






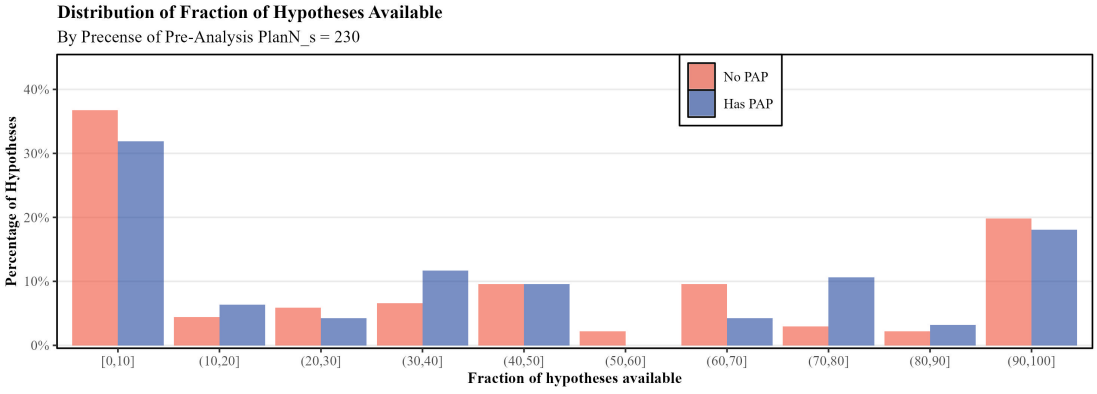
Distribution by group

Distribution of Fraction Available, By Publication Type



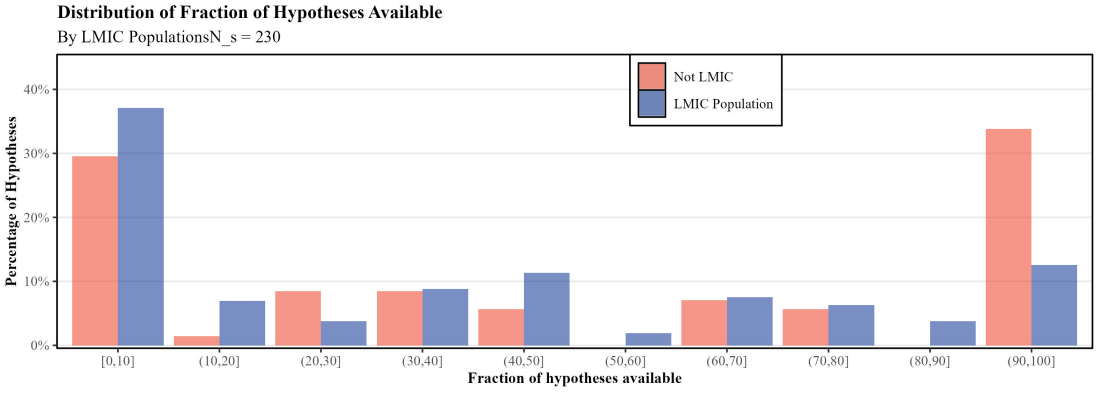
back

Distribution of Fraction Available, By PAP



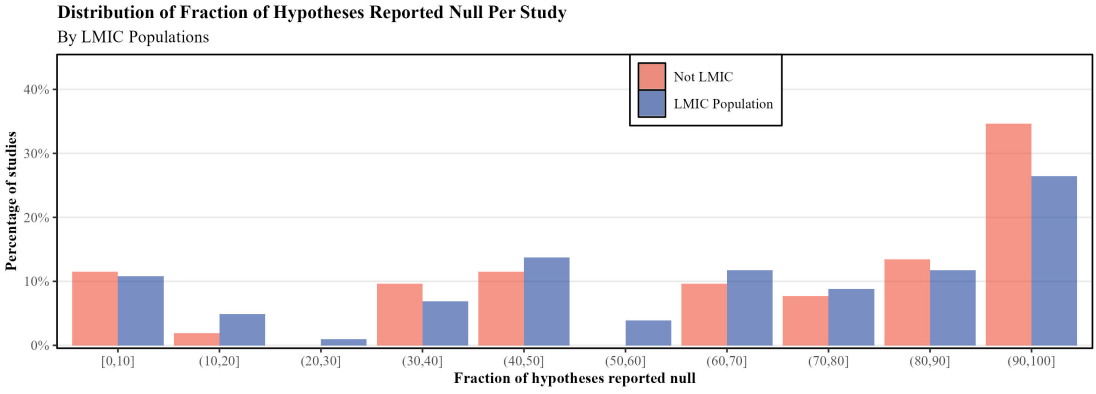
back

Distribution of Fraction Available, By LMIC



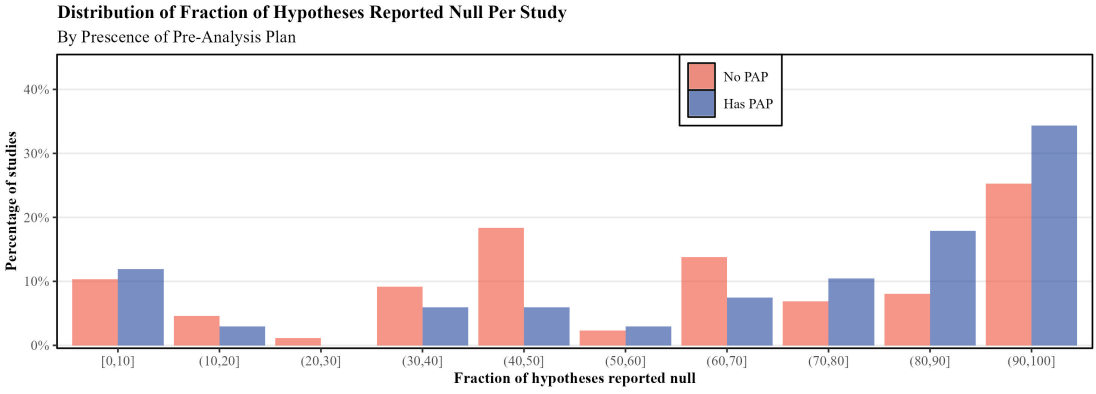
back

Distribution of Fraction Null (Conditional), By LMIC



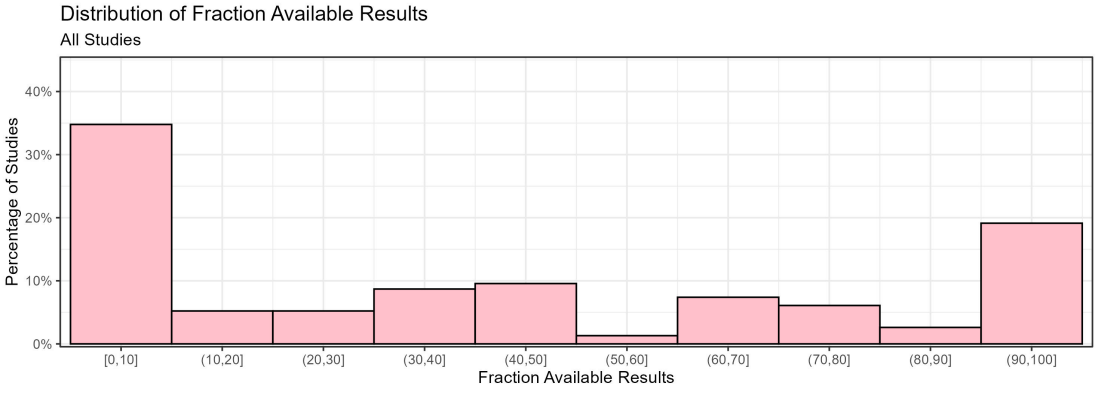
back

Distribution of Fraction Null (Conditional), By PAP



back

Fraction of Available Hypotheses: All Studies



back

Encoding Approach (cont.)

- **Initial set of studies:** 10 studies were double-coded to ensure reliability.
- **Efficiency Choice:** To increase sample size, we opted not to double-code all studies due to the task's labor-intensive nature.
- **Quality Control:** Initial encoding of around 20 studies was closely supervised and revised by principal investigators Fernando Hoces and Erik Sørensen.
- **Supervision:** The next 80 studies were supervised and reviewed by one PI (FH), with discrepancies and questions tracked via Git and GitHub issues. Lessons from these were documented.
- **Ongoing Quality Check:** Starting from study 101, research assistants reviewed each other's work, with an additional quality check by a PI on a random 20% of studies.

back

Heterogeneity - Details

- Only main, well specified, and motivated

Example of what we do not encode:

3. Dimensions of Heterogeneity:
(a) Cognitive ability
(b) Risk and time preferences
(c) Education
(d) Locus of control
(e) Wealth/economic status

Example of what we encode:

Heterogeneity:

- Price increases in the treatment will be most pronounced during harvest season, meaning that the inter-seasonal price variation in the treatment will be dampened.
- The market-level impact of the treatment will be heterogeneous across the distribution of infrastructure quality, as measured by distance from the trading center to the nearest paved road.
- Guarantees will be particularly important at engendering trade across long distances where uncertainty is highest. We can disentangle the specific effect of the match rate by the discontinuities in the insurance coverage across the continuous distance between buyer and seller (see Analysis section for details).

back

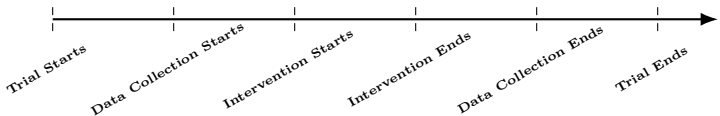
Comparison to Franco et. al. 2014 (Hypotheses Level)

Study	Franco et al.			Our Sample		
	Frac. Null Studies	Frac Null or Mixed	N	Frac. Null Hypotheses*	Frac Available Hypotheses	N
Not written	69%	91%	45	–	0%	686
Written, not published	10%	56%	70	73%	39%	915
Published, not top	7%	51%	71	69%	36%	1908
Published top	14%	40%	35	59%	28%	585

* Conditional on hypothesis being available

back

Inclusion/Exclusion Rule Based on Registration Timeline



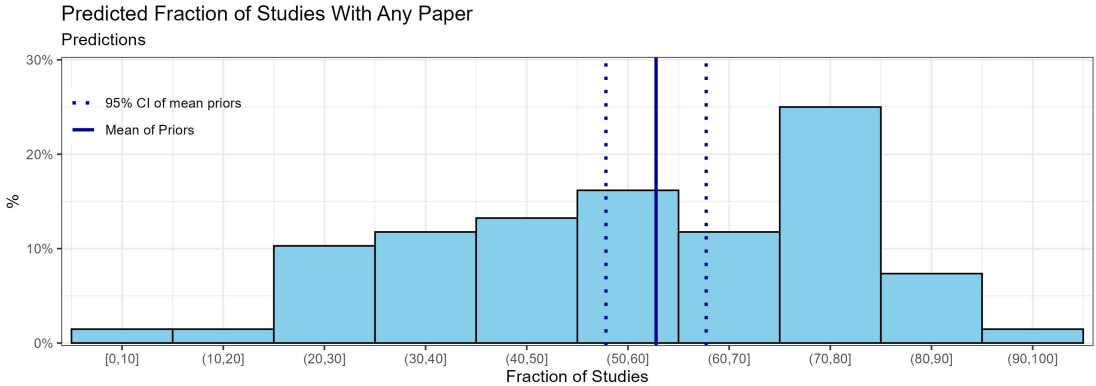
Include if:
- Registration occurs **before** intervention/data collection end date **OR**
- Authors explicitly state no data access prior to registration.

Exclude if:
- Registration occurs **after** intervention/data collection end date **AND**
- No mention of data access/analysis.

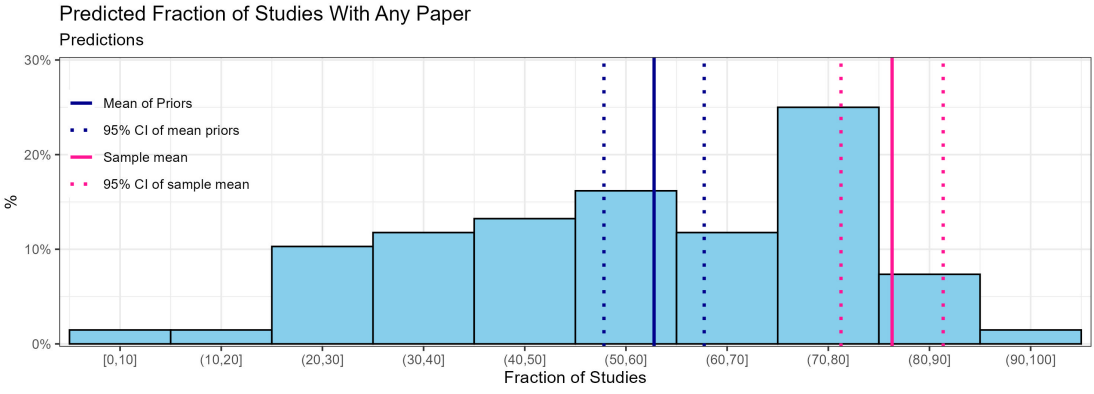
Grey Area:
Registration occurs within 1 month after intervention/data collection end date.

back

Studies With Paper 8-9 Years After Registration

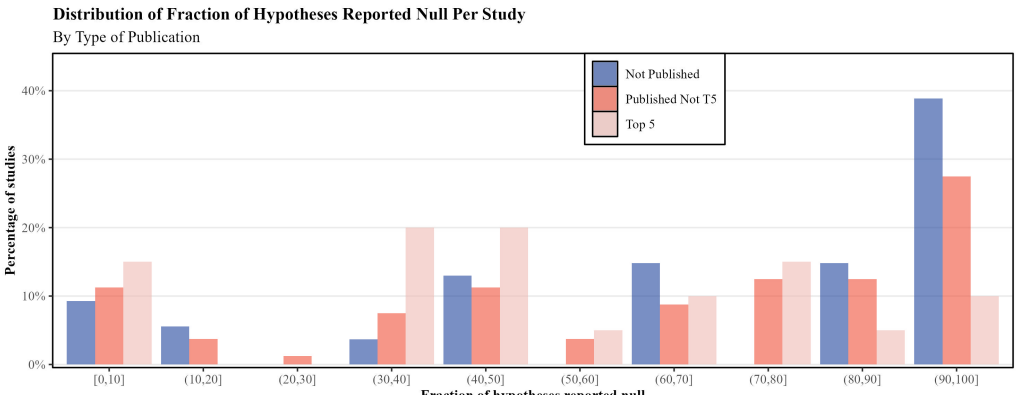


Studies With Paper 8-9 Years After Registration



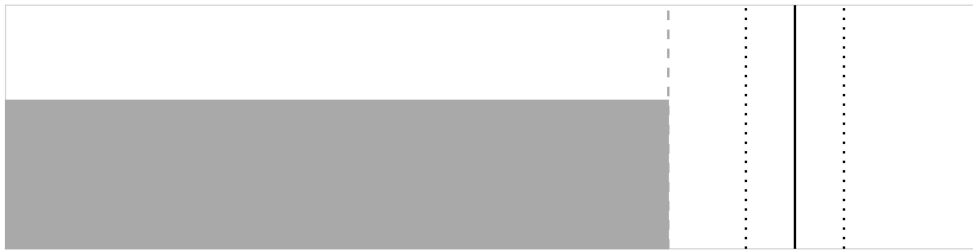
▸ Sample distribution

Distribution of Fraction Null (Conditional), By Publication Type



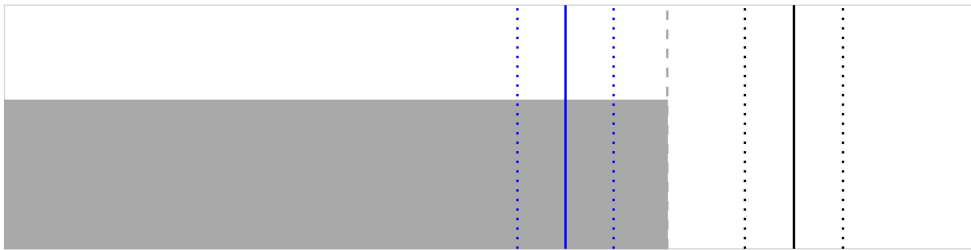
back

Comparing Estimates with Expert Forecasts: No Paper



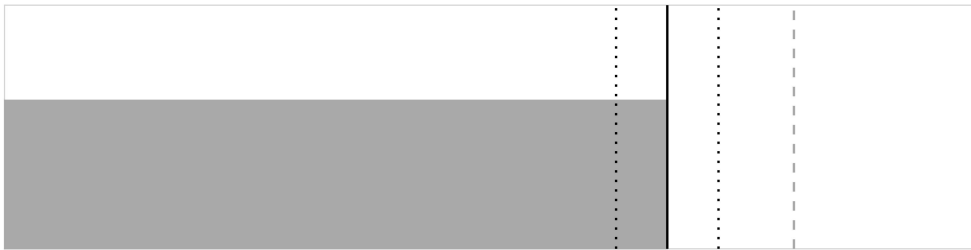
Back

Comparing Estimates with Expert Forecasts: No Paper



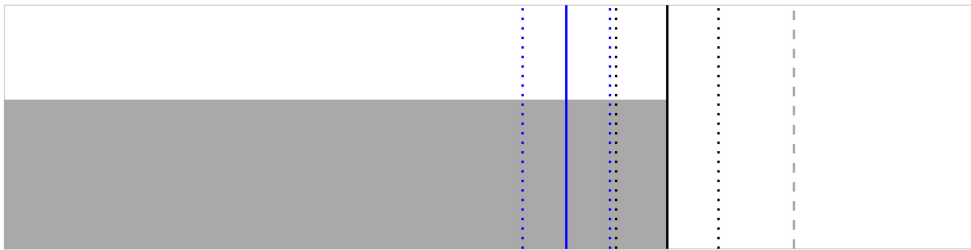
Back

Comparing Estimates with Expert Forecasts: No Hypothesis



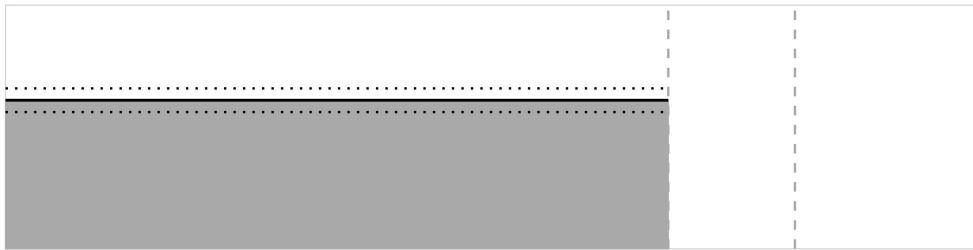
Back

Comparing Estimates with Expert Forecasts: No Hypothesis



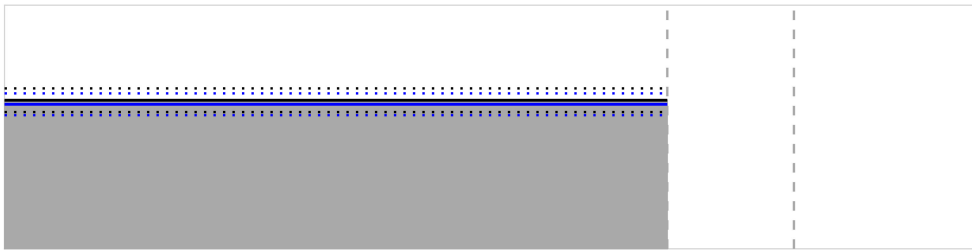
Back

Comparing Estimates with Expert Forecasts: Fraction Available



Back

Comparing Estimates with Expert Forecasts: No Hypothesis



Back

Study and Hypothesis Availability

	%
% of Studies with:	
any results	81.30
at least one estimate	77.39
at least one estimate & consistent modified	68.26
Average % of Hypotheses per study with:	
estimates	60.03
estimates & (at most) consistently modified	41.63
estimates with no modifications	27.24
estimates with no mod & in main body	23.32
estimates with no modifications, main body & not hard to find	18.40

Table: Availability of Studies and Main Hypotheses

back