LEIBNIZ OPEN SCIENCE DAY 2025

OCTOBER 27

LEIBNIZ ASSOCIATION BERLIN

(Un)Published: Evidence of Publication Bias from Two German Probability-Based Panel Infrastructures

Caroline Poppa

SHARE Berlin Institute
cpoppa@share-berlin.eu

SHARE BERLIN Institute · GESIS Leibniz Institute for the Social Sciences · SOEP IS · IMPRS on Learning, Institutions, and Future Evolution (LIFE) INTERNATIONAL MAX PLANCK RESEARCH SCHOOL

Outline

Introduction – *What is Publication Bias?*

Assessing Publication Bias – *PubBias Project*

Prevalence of (Un)Published Research – *Descriptive Outcomes*

Evidence of Publication Bias – *Assessing Study and Hypothesis Levels*

Author Survey – *First Insights & Outlook*

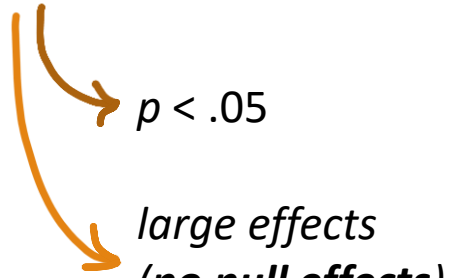
Discussion

What is publication bias?

"(…) *a* ***tendency*** *toward preparation, submission and publication of research findings* ***based on the nature and direction*** *of the research results*" (Dickersin, 2005).

"(…) *whenever the likelihood or time-lag of publication, or the prominence, language, impact factor of journal space or the citation rate of studies depend* ***on the direction and significance*** *of research findings*" (Auspurg & Hinz, 2011).

=> ***Publication bias*** is the **prioritized** and **selective reporting** of scientifically **significant** results.



Different kinds of prioritzed and selective reporting …

a) **Selective reporting**: publish **significant findings only,** without mentioning other results below standard statistical thresholds
(author level)

= "bias at the *hypothesis level*"

b) **Nonpublication** of **entire** ("failed"/nonsignificant) **studies,** due to ...

- not writing up / not submitting
file drawer bias *(author level)*

- *editorial/reviewer rejection*
editorial bias

= "bias at the *study level*"

distorted / biased scientific evidence (public)

time, effort + resources wasted in replications

c) HARKing (Hypothesizing after Results are Known) *(author level)*

=> increased chance for Type I errors

Sciencexpress Reports

Publication bias in the social sciences: Unlocking the file drawer

Annie Franco,[1] Neil Malhotra,[2]* Gabor Simonovits[1]

[1]Department of Political Science, Stanford University, Stanford, CA, USA. [2]Graduate School of Business, Stanford University, Stanford, CA, USA.

*Corresponding author. E-mail: neilm@stanford.edu

We study publication bias in the social sciences by analyzing a known population of conducted studies—221 in total—where there is a full accounting of what is published and unpublished. We leverage TESS, an NSF-sponsored program where researchers propose survey-based experiments to be run on representative samples of American adults. Because TESS proposals undergo rigorous peer review, the studies in the sample all exceed a substantial quality threshold. Strong results are 40 percentage points more likely to be published than null results, and 60 percentage points more likely to be written up. We provide not only direct evidence of publication bias, but also identify the stage of research production at which publication bias occurs—authors do not write up and submit null findings.

Publication bias occurs when “publication of study results is based on the direction or significance of the findings” (*1*). One pernicious form of publication bias is the greater likelihood of statistically significant results being published than statistically insignificant results, holding fixed research quality. Selective reporting of scientific findings is often referred to as the “file drawer” problem (*2*). Such a selection process increases the likelihood that published results reflect Type I errors rather than true population parameters, biasing effect sizes upwards. Further, it constrains efforts to assess the state of knowledge in a field or on a particular topic, since null results are largely unobservable to the scholarly community.

Publication bias has been documented in various disciplines within the biomedical (*3–9*) and social sciences (*10–17*). One common method of detecting publication bias is replicating a meta-analysis with and without unpublished literature (*18*). This approach is limited because much of what is unpublished is unobserved. Other methods solely examine the published literature and rely on assumptions about the distribution of unpublished research by, for example, comparing the precision ing and not worthy of further analysis or publication (*32–35*). One analysis of all IRB-approved studies at a single university over two years found that a majority of conducted research was never submitted for publication or peer-review (*36*).

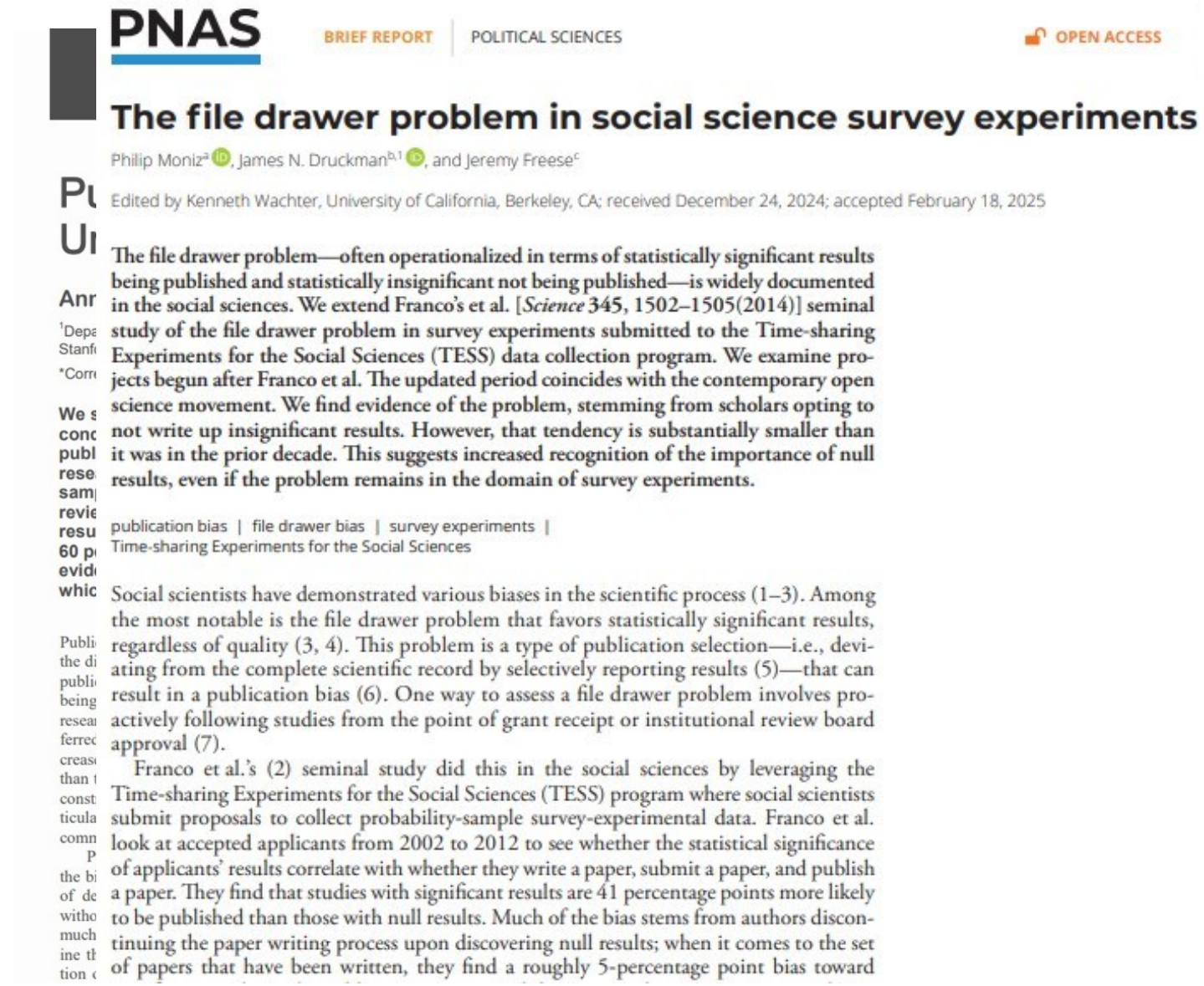
Surprisingly, similar cohort analyses are much rarer in the social sciences. There are two main reasons for this lacuna. First, there is no process in the social sciences of pre-registering studies comparable to the clinical trials registry in the biomedical sciences. Second, even if some unpublished studies could be identified, there are likely to be substantial quality differences between published and unpublished studies that make them difficult to compare. As noted, previous research attempted to identify unpublished results by examining conference papers and dissertations (*37*) and human subjects registries of single institutions (*36*). However, such techniques may produce unrepresentative samples of unpublished research, and the strength of the results may be confounded with research quality. Conference papers, for example, do not undergo a similar process of peer review as journal articles in the social sciences and therefore cannot be used as a comparison set. This paper is unique in the study of publication bias in the social sciences in that it analyzes a known population of conducted studies and all studies in the population exceed a substantial quality threshold.

We leverage TESS (Time-sharing Experiments in the Social Sciences), an NSF-sponsored program established in 2002 where researchers propose survey-based experiments to be run on nationally representative samples. These experiments typically embed some randomized manipulation (e.g., visual stimulus, question wording difference) within a survey questionnaire. Researchers apply to TESS, which then peer reviews the proposals and distributes grants on a competitive basis (*38*). Our basic approach is to compare the statistical results of TESS experiments that eventually got published to the results of those that remain un-

file drawer bias *(not writing up or not submitting studies based on the nature of study results)*

Franco et al.,2014 :

reported by roughly **60%** of authors (studies conducted within TESS **2002 – 2012**)



PNAS BRIEF REPORT | POLITICAL SCIENCES OPEN ACCESS

The file drawer problem in social science survey experiments

Philip Moniz[a], James N. Druckman[b,1], and Jeremy Freese[c]

Edited by Kenneth Wachter, University of California, Berkeley, CA; received December 24, 2024; accepted February 18, 2025

The file drawer problem—often operationalized in terms of statistically significant results being published and statistically insignificant not being published—is widely documented in the social sciences. We extend Franco's et al. [*Science* 345, 1502–1505(2014)] seminal study of the file drawer problem in survey experiments submitted to the Time-sharing Experiments for the Social Sciences (TESS) data collection program. We examine projects begun after Franco et al. The updated period coincides with the contemporary open science movement. We find evidence of the problem, stemming from scholars opting to not write up insignificant results. However, that tendency is substantially smaller than it was in the prior decade. This suggests increased recognition of the importance of null results, even if the problem remains in the domain of survey experiments.

publication bias | file drawer bias | survey experiments | Time-sharing Experiments for the Social Sciences

Social scientists have demonstrated various biases in the scientific process (1–3). Among the most notable is the file drawer problem that favors statistically significant results, regardless of quality (3, 4). This problem is a type of publication selection—i.e., deviating from the complete scientific record by selectively reporting results (5)—that can result in a publication bias (6). One way to assess a file drawer problem involves proactively following studies from the point of grant receipt or institutional review board approval (7).

Franco et al.'s (2) seminal study did this in the social sciences by leveraging the Time-sharing Experiments for the Social Sciences (TESS) program where social scientists submit proposals to collect probability-sample survey-experimental data. Franco et al. look at accepted applicants from 2002 to 2012 to see whether the statistical significance of applicants' results correlate with whether they write a paper, submit a paper, and publish a paper. They find that studies with significant results are 41 percentage points more likely to be published than those with null results. Much of the bias stems from authors discontinuing the paper writing process upon discovering null results; when it comes to the set of papers that have been written, they find a roughly 5-percentage point bias toward

file drawer bias *(not writing up or not submitting studies based on the nature of study results)*

Franco et al.,2014 :

↓

reported by roughly **60%** of authors (studies conducted within TESS **2002 – 2012**)

Moniz et al., 2025 :

↓

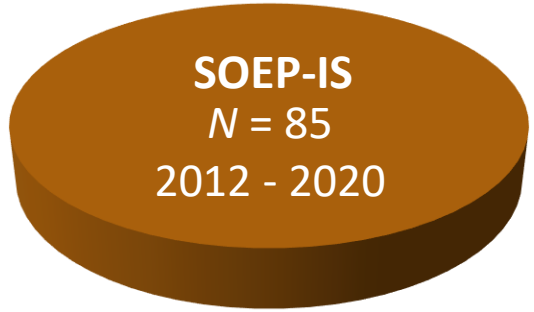
reported by roughly **20%** of authors (studies conducted within TESS **2012 – 2018**)

Opening the File Drawer: *Assessing and Understanding Publication Bias in the Economic, Behavioral and Social Sciences by Utilizing two Probability-Based German Academic Access Panels*

≈ 5,200 respondents

- open **self-administered** probability-based panel
- mixed mode survey (75% web-based, 25% mail)
- quarterly data collection (no proposal deadlines)

- **no privileged access** for study authors
- acceptance rate 46.0% (2013 -2021)

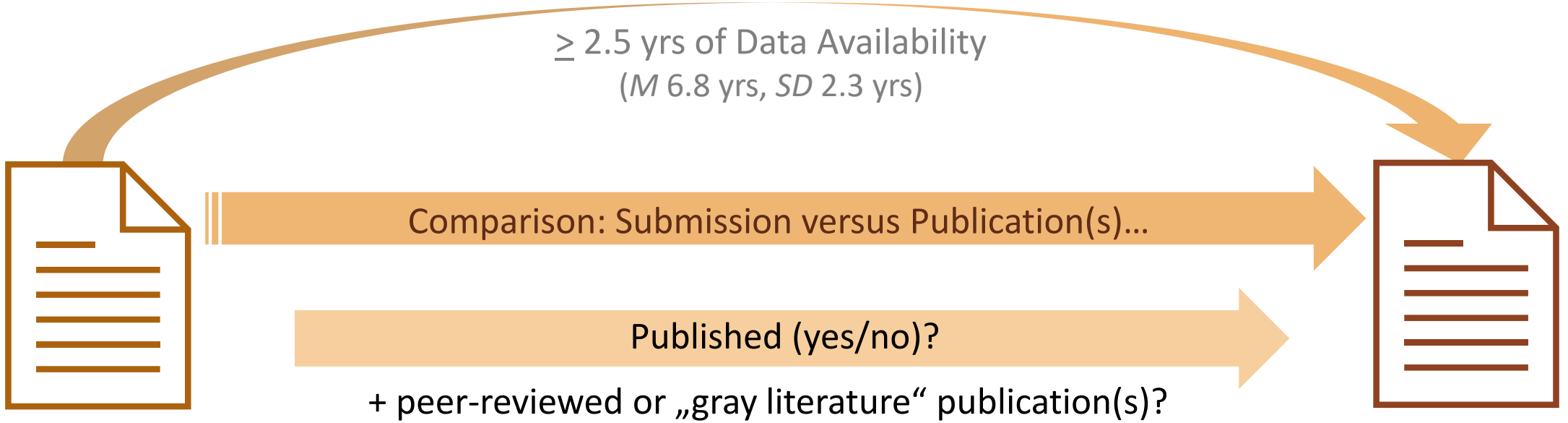


≈ 4,800 respondents

- probability-based **household panel survey**
- interviewer administered (in-person) interviews
- annual data collection

- one year **privileged data access** for proposal authors
- acceptance rate 36.6% (2012 -2019)

⇒ researchers from all social, behavioral or economic sciences can submit their own studies



Hypotheses (explicit or implicit)

Explorative Analyses

Author Information (Seniority, h-index, Gender, etc.)

Allows to assess …
… prevalence of **(un-)published research** (study-level bias)
… **„hit-rate“** (% of supported versus unsupported hyp.)
… **hypothesis-level bias** (inclusion of new hyp., select. rep.)

(Same) Hypotheses/ Explorative Analyses?

Hit-rate (% of Supported Hypotheses)

Change in Author Constellation?

RQ1: *What proportion of studies have resulted in (a) a peer-reviewed journal article, (b) gray literature, or (c) no publication?*

RQ2: *Does **time since data availability** predict publication likelihood?*

H1 (Publication Bias): *We hypothesize that studies **with findings that reject the null hypothesis are more likely to be published than those with null or nonsignificant results**. This bias is expected to manifest at both the study and hypothesis levels:*

H1a (Study-level bias): *A **higher proportion of studies publish findings that predominantly support their hypotheses** compared to studies that report null or nonsignificant findings.*

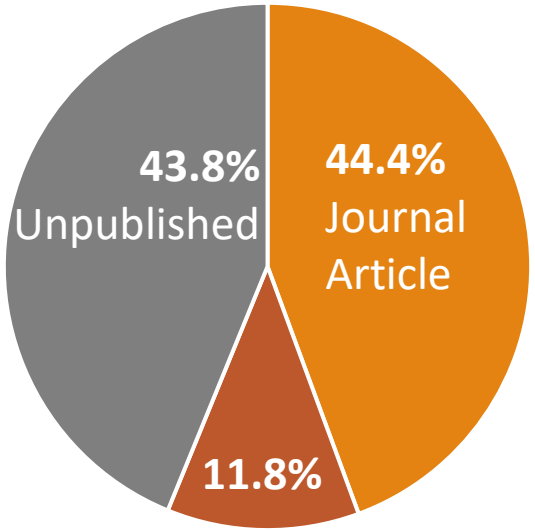
H1a (Hypothesis-level bias): *The proportion of **supported newly added (ad-hoc) hypotheses** is higher than the proportion of supported hypotheses that were submitted.*

Table 1

Time Since Data Availability and Publication Outcomes (Total Sample and by Panel)

	Total	GESIS Panel	SOEP-IS
N of Studies (Years)	178 (2012-2021)	93 (2013-2021)	85 (2012-2020)
Mean Years Since Data Availability *(SD)*	6.8 (2.3)	6.5 (2.5)	7.1 (2.1)
- *Min (Max)*	2.5 (11.0)	2.5 (10.9)	4.0 (11.0)
Mean Weeks Since Data Availability *(SD)*	356.2 (121.4)	340.7 (129.4)	373.1 (110.3)
- *Min (Max)*	129.7 (574.0)	129.7 (567.0)	209.4 (574.0)
N of studies with … publication(s) (%)	100 (56.2)	42 (45.2)	58 (68.2)
- at least one journal article	79 (44.4)	36 (38.7)	43 (50.6)
- exclusively gray literature	21 (11.8)	6 (6.5)	15 (17.6)
- no	78 (43.8)	51 (54.8)	27 (31.8)

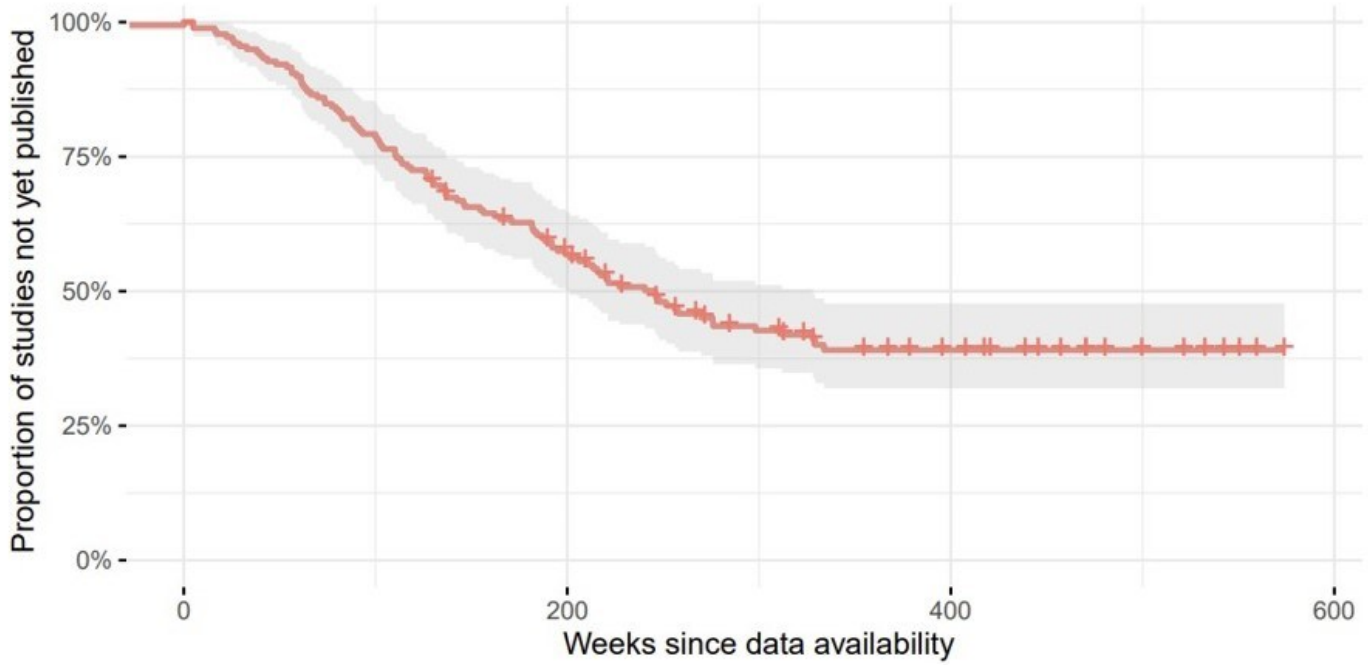
Publication Outcomes After ≥ 2.5 yrs Data Availability



- **55.6**% of studies did not result in a peer-reviewed journal article
- similar to Franco et al. (2014): 57% of studies did not publish in peer-reviewed outlets

Figure S1

Kaplan-Meier Curve (t to any Publication in Weeks), all Studies (Study-level)

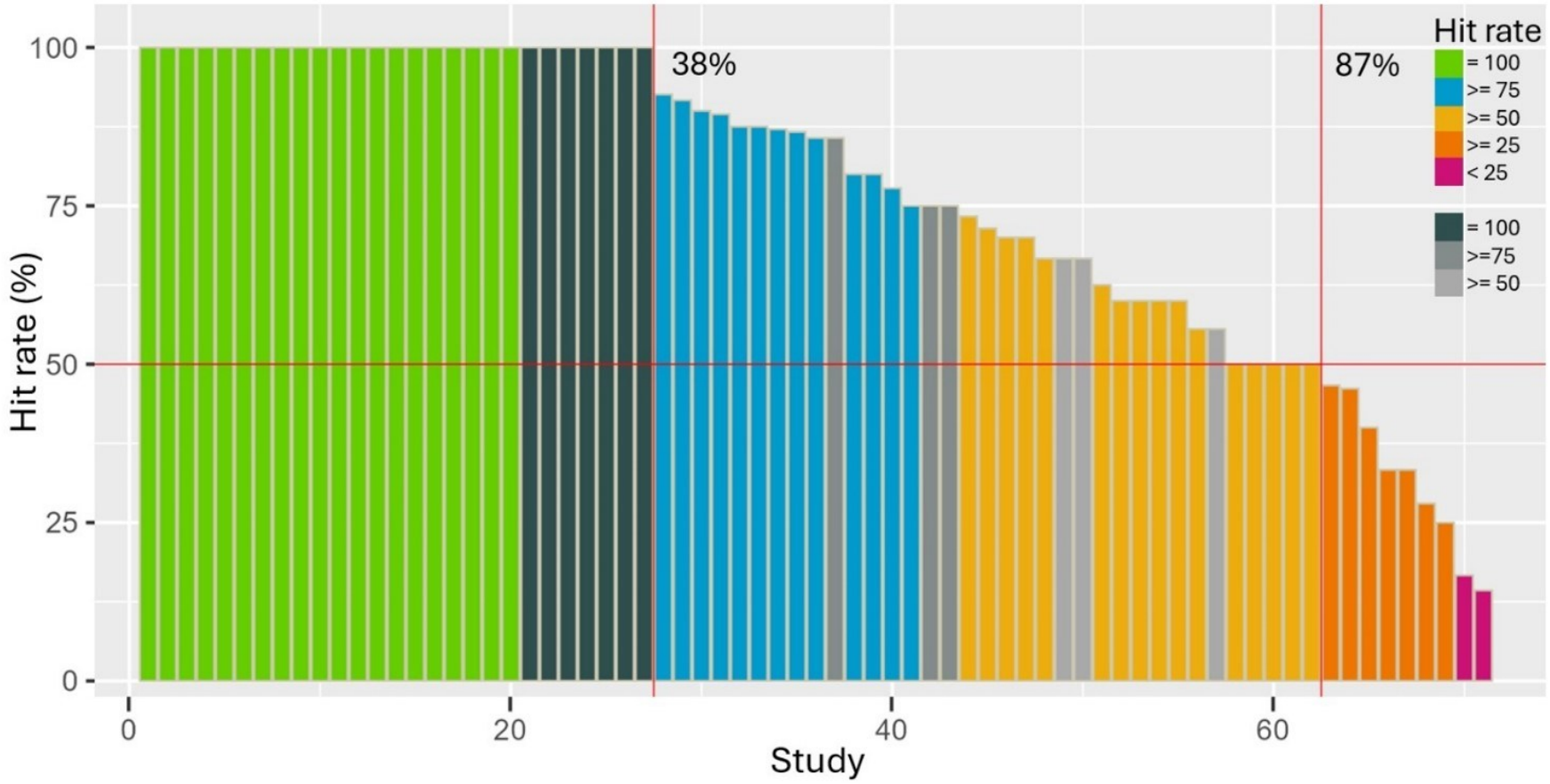


Note. Time to publication in this case describes the minimum number of weeks between data availability and any publication (for all published studies) and weeks between data availability and end of publication search (May 5th, 2025) for all unpublished studies.

- Discrete-time survival analysis, modelling the weekly "hazard" of resulting in a publication.
- Model indicated that **weekly "hazard" of publication slightly decreased over time** (week: b = – 0.002, SE = 0.0009, z = -2.16, p = .031).
- Time since data availability **did not meaningfully predict publication likelihood** in our sample, suggesting study authors had sufficient time to "go through" publication process (if decided to).

Evidence of Publication Bias – H1a (Study-level bias): *A higher proportion of studies publish findings that predominantly support their hypotheses compared to studies that report null or nonsignificant findings.*

Fig. 1a *Hit Rate of Studies Publishing Any Hypotheses (N = 71)*



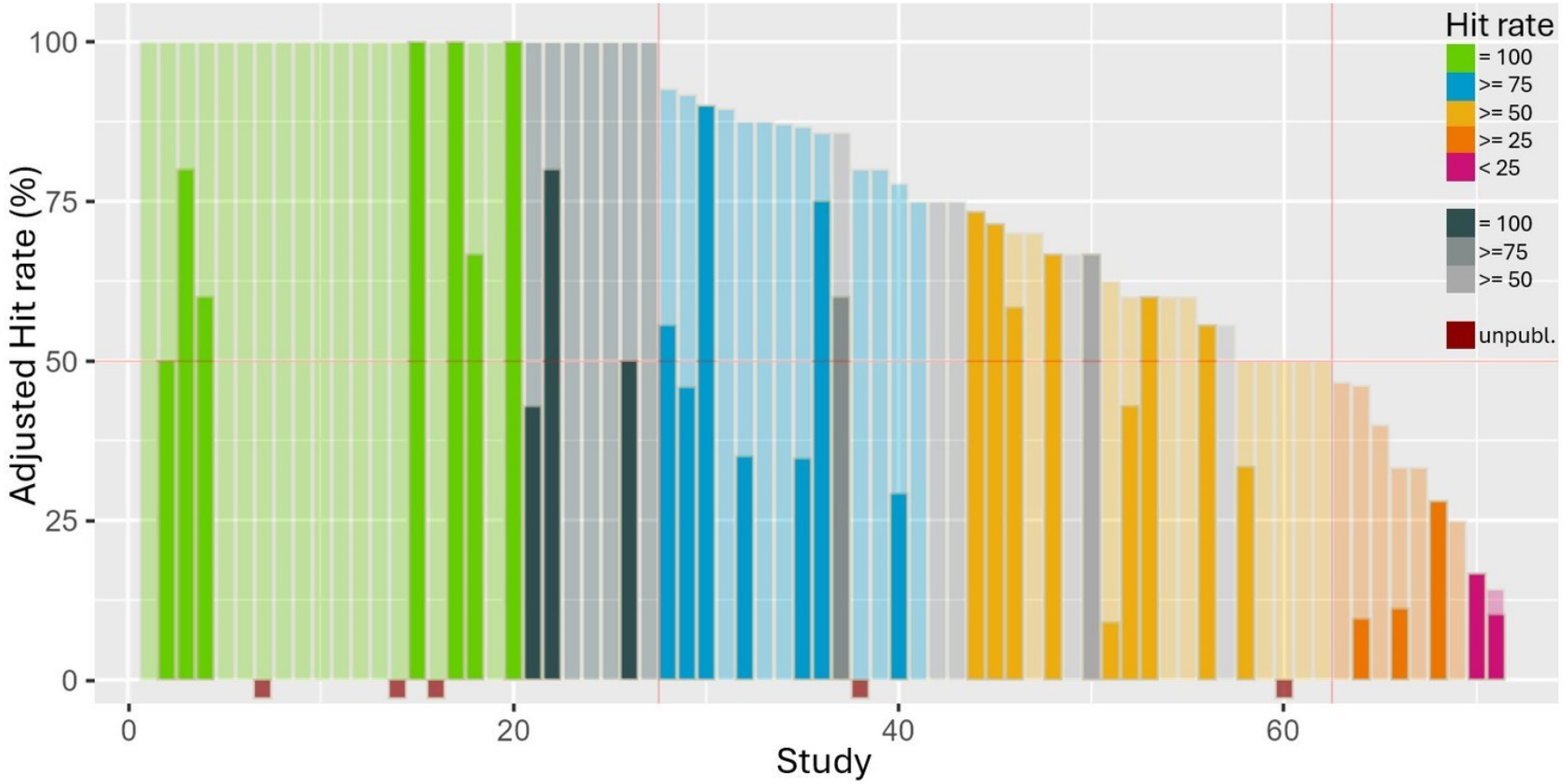
- **87.3%** of published studies (N = 62) **supported at least half their hypotheses**
- More than one third (**38.0%,** N = 27) of studies **supported all their published hypotheses.**

- **Prioritized reporting of supported results.**

  = Publication Bias

Evidence of Publication Bias – H1a (Study-level bias): *A higher proportion of studies publish findings that predominantly support their hypotheses compared to studies that report null or nonsignificant findings.*

Fig. 1b *Adjusted Hit Rate of Studies that Published any Submitted Hypotheses*



- **53.3%** published exclusively non-submitted hypotheses (38 of 71 studies faded out)
- Among studies that published with a 100% hit-rate, **63.0%** published **exclusively new hypotheses**

- **Evidence for selective reporting and HARKing**

  = Publication Bias

Evidence of Publication Bias – H1a (Hypothesis-level bias): *The proportion of **supported newly added (ad-hoc) hypotheses** is higher than the proportion of supported hypotheses that were submitted.*

Table 2 *Total N of Submitted (Published , Unpublished) and New Hypotheses by Support Status*

	N	*N* supported	*N* unsupported
N of submitted hypotheses	395	-	-
… of which published (% of submitted)	119 (30.1)	95 (79.8)	24 (20.2)
- of which completely tested	99	81	18
- of which partially tested (hypothesis change)	20	14	6
… of which unpublished (% of submitted)	276 (69.9)	-	-
N of novel hypotheses published (% of all published)	492 (80.9)	357 (72.6)	135 (27.4)
... Sum of all published (submitted and novel)	611	452	159

No support for H1a:

- **79.8%** of submitted (and published) hypotheses were supported
- **72.6%** of new (ad-hoc) hypotheses were supported

- Solely **30.1% of submitted** hypotheses published (selective reporting), resulting **in new hypotheses comprising >80%** of all **published hypotheses** (potential HARKing).
- **73.9% of all hypotheses supported** => skew towards supported hypotheses

Summary: Evidence of Publication Bias

- *Nonpublication: after, on average, 6.8 yrs of data availability ….*

… 55.6% of studies were not published in a peer-reviewed journal

… **43.8**% of studies remained wholly unpublished

- *t* since data availability did not increase publication „hazard“ – sample consisted of „sufficiently old“ studies

- Prioritized reporting of significant findings (support for H1a):

  >**87%** of studies **supported at least half of their hypotheses** (study-level),

  >**38**% of studies **supported all of their hypotheses** (study-level),

  74% of all (published) hypotheses supported (hypothesis-level)

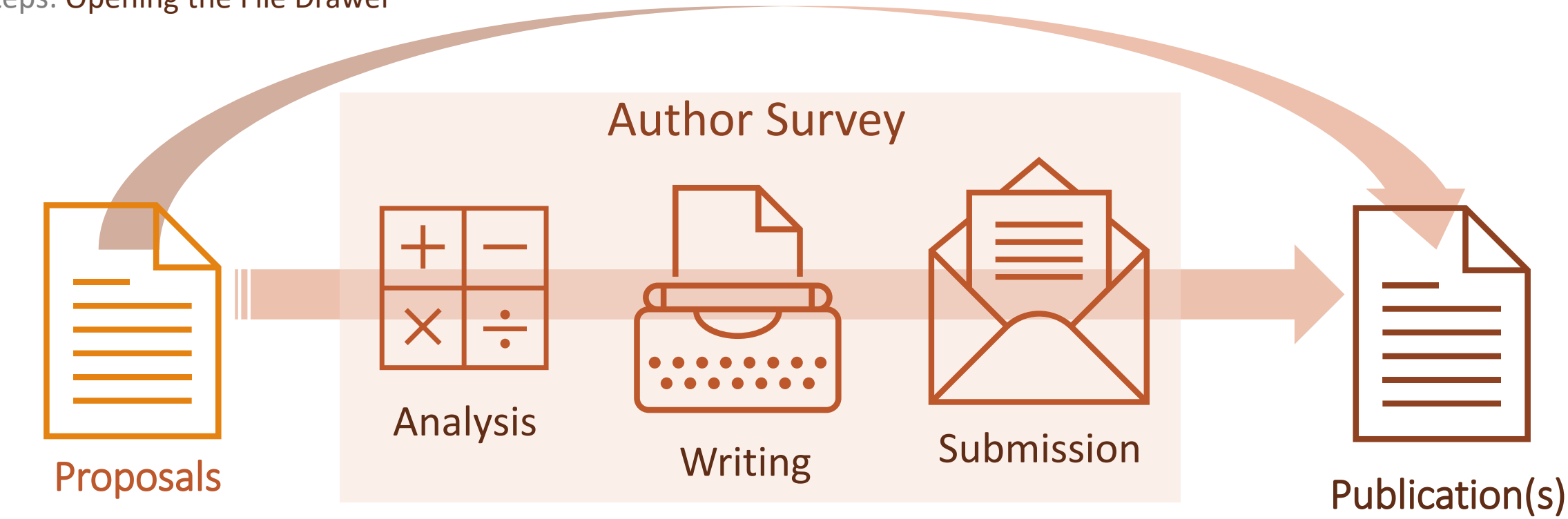
- Selective reporting & potential HARKing:

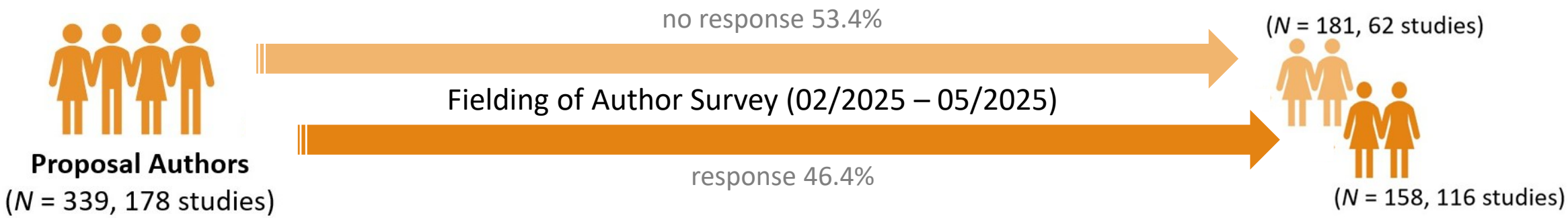
  30.1% of submitted hypotheses published

  >80% of published **hypotheses** were **not submitted**

- **Non-submitted** hypotheses were **not supported more often** than published submitted hypotheses (no support for H1b), suggesting that the prevalent selective reporting among submitted hypotheses “filtered out” nonsignificant findings

Next Steps: Opening the File Drawer





Submission

Author Survey

Publication

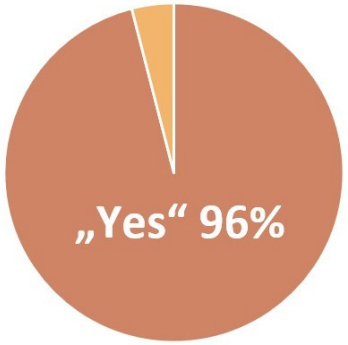
Goal: Assess reasons for (un)published work in detail

Szenario A:
Data not (yet) analyzed

Szenario B:
Data analyzed, not written up

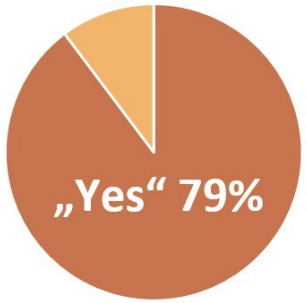
Szenario C:
Data analyzed, written up, not (yet/again) submitted

Data analyzed?



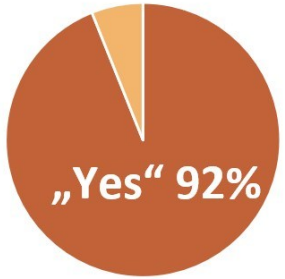
177 responses
(of 107 studies)
82% (*N* = 103) of all survey studies

Written up?



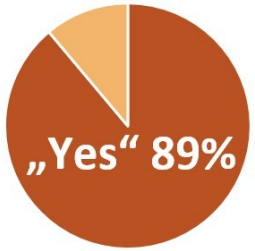
168 responses
(of 103 studies)
71% (*N* = 82) of all survey studies

Submitted?



108 responses
(of 66 studies)
53% (*N* = 62) of all survey studies

Published?



99 responses
(of 62 studies)
47% (*N* = 55) of all survey studies

Author Survey: Reasons for not writing up

What are the reasons why you do not plan to write up the results? Check all that apply.	Responses *N* = 34 (%)
*I was and still am **too busy** with other tasks/research.*	16 (47.1)
*My **co-authors** do **not** have **time** to work on it.*	10 (29.4)
*I have **lost interest** in the project.*	8 (23.5)
*My **research focus** has shifted.*	8 (23.5)
*I (partially) analyzed the data and was planning to report the results, but **due to their nature**, I decided not to.*	8 (23.5)
*I have **moved** to another department/institute/university.*	8 (23.5)
*I am **not working** primarily **in academia** (anymore).*	3 (8.8)
I handed the project over to another colleague.	1 (2.9)

30-47% no time

23% no motivation

23% **file drawer bias**

9 – 23% left project

Author Survey: Reasons for rejections (if provided and remembered)

What were the reasons for the rejection(s)? Check all that apply.	Responses *N* = 46 (%)
Lack of novelty or added value to science	16 (34.8)
Beyond the scope of the journal	11 (23.9)
Methodological approach	12 (26.4)
Theoretical approach	7 (16.2)
Lack of statistical power due to study design (sample size)	2 (4.3)
Flaws in the manuscript	2 (4.3)
Null results	1 (2.2)
Effect sizes too small	1 (2.2)

35% lack of novelty

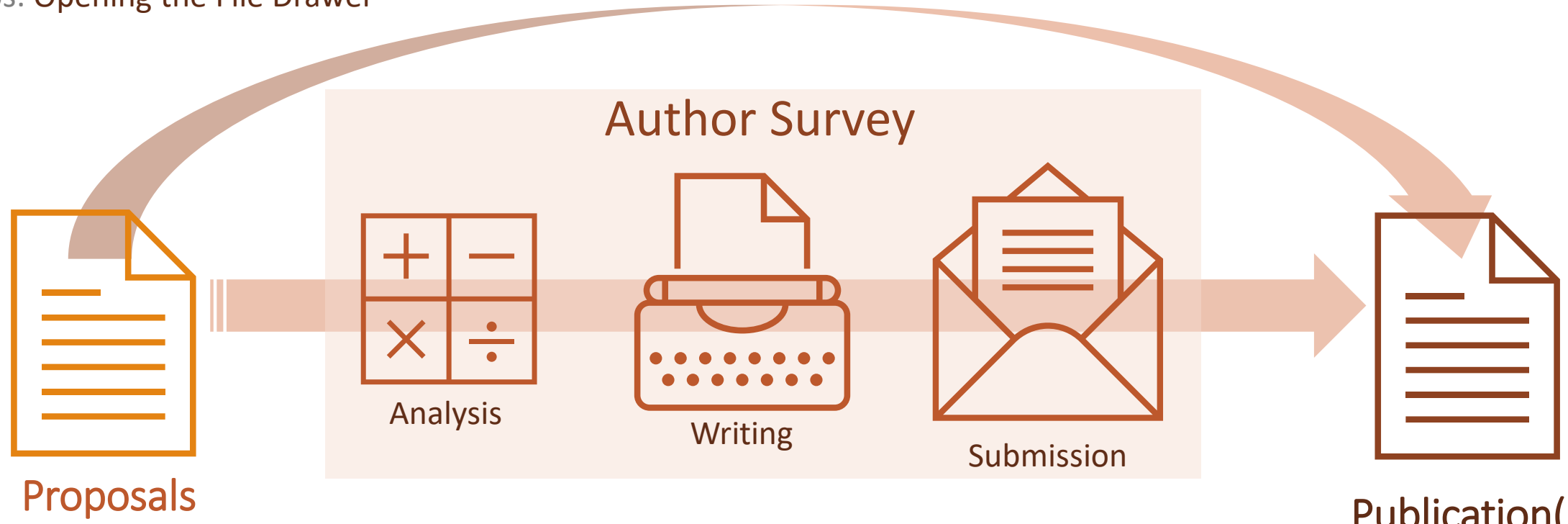
24% „wrong journal“

4 -26% methods, theory sample/study design

2% **nature of results**

Other, namely:
[1] "Are you serious? There are so many reasons why reviewers/editors reject papers. They had nothing or very little to do with the data."
[2] "only the editors knows"

Next Steps: Opening the File Drawer



Assess reasons for (un)published work in more detail, but also:

- beliefs about open science
- perceived pressure to publish
- hypothetical publication szenario
- parenthood and academic productivity
- changes in author constellations (i.e., reasons for leaving academia)

Proposal Authors
(*N* = 339, 178 studies)

no response 53.4%
(*N* = 181, 62 studies)

(*N* = 158, 116 studies)
response 46.4%

Merci!

Merci :)

SOEP IS

GESIS Leibniz Institute for the Social Sciences

ISR INSTITUTE FOR SOCIAL RESEARCH UNIVERSITY OF MICHIGAN

David Richter

Caroline Poppa

cpoppa@share-berlin.eu

Désirée Nießen

Jessica Daikeler

Henning Silber

Bernd Weiß

UNIVERSITÄT DUISBURG ESSEN

SHARE BERLIN Institute

+ happy to answer questions..

QR Code to PubBias Website

IMPRS on Learning, Institutions, and Future Evolution (LIFE)
INTERNATIONAL MAX PLANCK RESEARCH SCHOOL

References

Auspurg, K., & Hinz, T. (**2011**). What fuels publication bias? theoretical and empirical analyses of risk factors using the caliper test. Jahrbücher für Nationalökonomie und Statistik, 231(5-6), 636-660.

Dickersin, K. (**2005**). Publication bias: Recognizing the problem, understanding its origins and scope, and preventing harm. Publication bias in meta-analysis: Prevention, assessment and adjustments, 9-33.

Franco, A., Malhotra, N., & Simonovits, G. (**2014**). Publication bias in the social sciences: Unlocking the file drawer. *Science*, *345*(6203), 1502-1505.

Moniz, P., Druckman, J. N., & Freese, J. (**2025**). The file drawer problem in social science survey experiments. *Proceedings of the National Academy of Sciences*, *122*(12), e2426937122.

PubBias is preregistered:

Nießen, D., Daikeler, J., Silber, H., Weiß, B., Poppa, C., & Richter, D. (**2023**). Opening the file drawer: Assessing and understanding publication bias in the social, economic, and behavioral sciences by utilizing two German academic probabilistic panels. *OSF*. https://doi.org/10.17605/OSF.IO/B4QRP

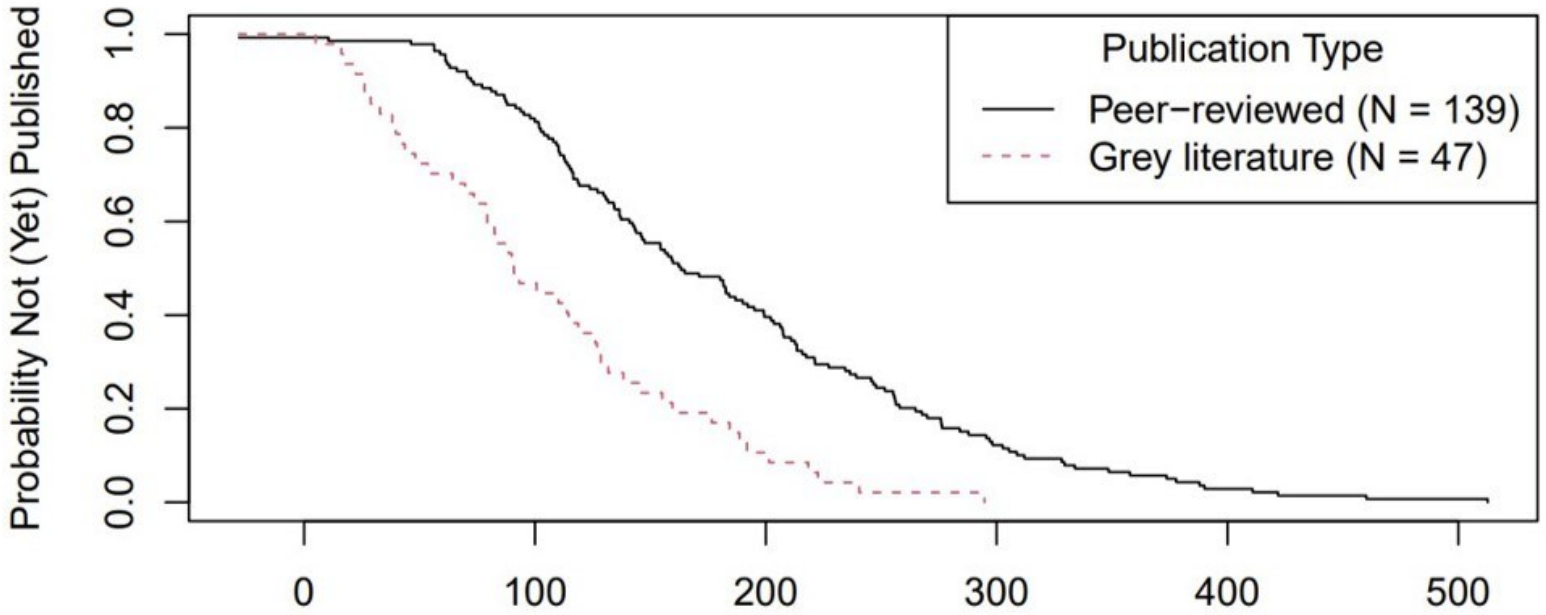
This manuscript is currently in preparation – **please only cite or distribute after approval of first author**

Poppa, C., Nießen, D., Daikeler, J., Henning Silber, Weiß, B., & Richter, D. (**2025**). The Tip of the Iceberg? Insights into the Prevalence of Publication Bias in Two Probability-Based Academic Panels. *[Manuscript in Preparation]*, *SHARE Berlin Institute, Berlin.*

RQ2 *Does time since data availability predict publication likelihood (RQ2a),* ***and are there differences in publication speed between journal articles and gray literature (RQ2b)?***

Figure S2

Kaplan-Meier Curve (t to Publication in Weeks), by Publication Type (Publication Level)



Note. Time to publication in this case describes the weeks between data availability and publication for any publication of each study. If a publication included several waves of a conducted study, the most recent wave was chosen to calculate the minimum timespan.

Table 3

Tracing of Published Hypotheses by Publication Type (Studies Resulting in Journal Articles Versus Studies Solely Producing Gray Literature).

	N = 100 studies that published …		
N (% of studies that published, N = 100)	… only journal articles N = 64	… mixed (both) N = 15	… only gray literature N = 21
… publications including any initial hypotheses N = 33 (33.0)	23 (35.9)	5 (33.3)	5 (23.8)
- of which including changes in initial hypotheses N = 11 (11.0)	8 (12.5)	2 (13.3)	1 (4.8)
- of which omitting parts of the hypothesis N = 10 (10.0)	7 (10.9)	2 (13.3)	1 (4.8)
- of which using different scales/measures N = 2 (2.0)	1 (1.6)	-	1 (4.8)
- of which direction change of hypothesis N = 2 (2.0)	2 (3.1)	-	-

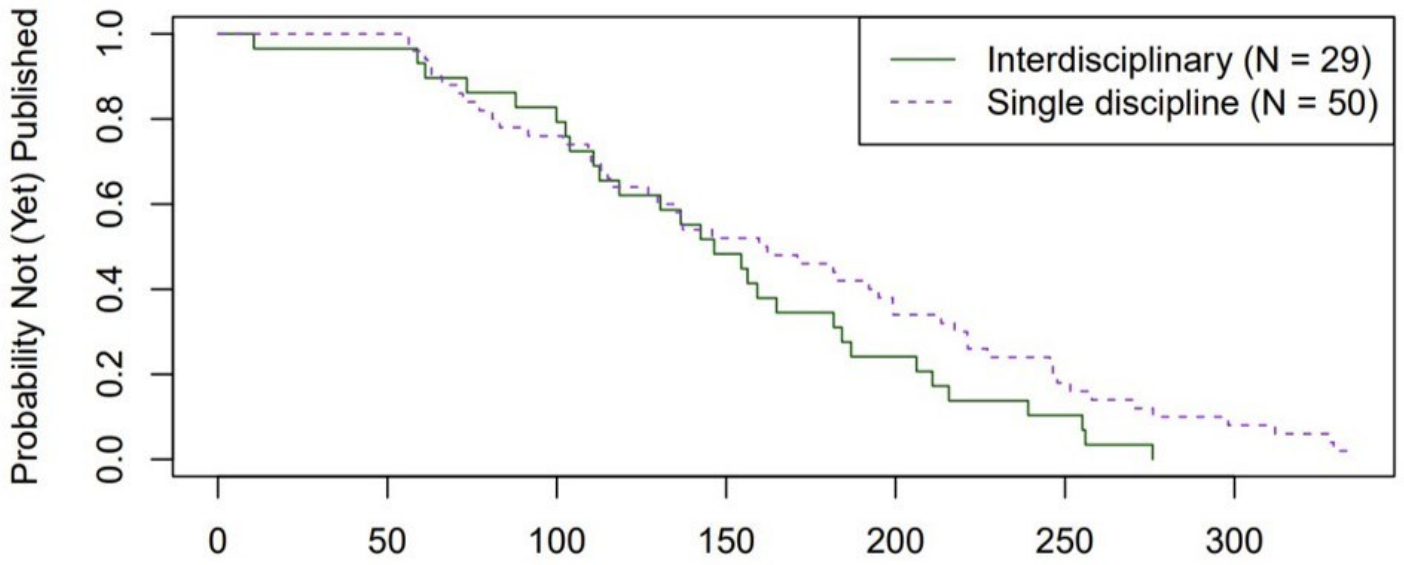
Table 4

Number (and Proportions) of Studies by Discipline, that Resulted in at Least one Journal Article, Gray Literature, or No Publication.

	Psychology $N = 63$	Sociology $N = 45$	Economics $N = 45$	Political Science $N = 42$	Survey Method. $N = 20$	Public Health $N = 14$	Linguistics, Comm. Res. $N = 5$
N of studies that published …							
... at least one journal article (%)	37 (**58.7**)	23 (51.1)	19 (42.2)	17 (40.5)	7 (35.0)	8 (57.1)	3 (60.0)
... gray literature only (%)	4 (6.3)	3 (6.7)	13 (**28.9**)	4 (9.5)	1 (5.0)	*N.A.*	1 (20.0)
... nothing (%)	22 (34.9)	19 (42.2)	13 (28.9)	21 (50.0)	12 (**60.0**)	6 (42.9)	1 (20.0)
N of single discipline studies (%)	42 (66.7)	23 (51.1)	21 (46.7)	21 (50.0)	17 (**85.0**)	4 (28.6)	2 (40.0)
N of total journal articles (Ø per study)	72 (**1.1**)	34 (0.8)	34 (0.8)	29 (0.7)	10 (0.5)	13 (0.9)	5 (1.0)

Figure S4

Kaplan-Meier Curve (t to First Journal Article Publication in Weeks), by Disciplinary Breadth (Study Level)



Note. Time since data availability in this case describes the weeks between data availability and first journal article publication for each study.

- **Interdisciplinary** studies were **published** both significantly **more,** and significantly more often **as journal articles,** than single-discipline studies
- But interdisciplinary research was **not published faster** than single-disciplinary research

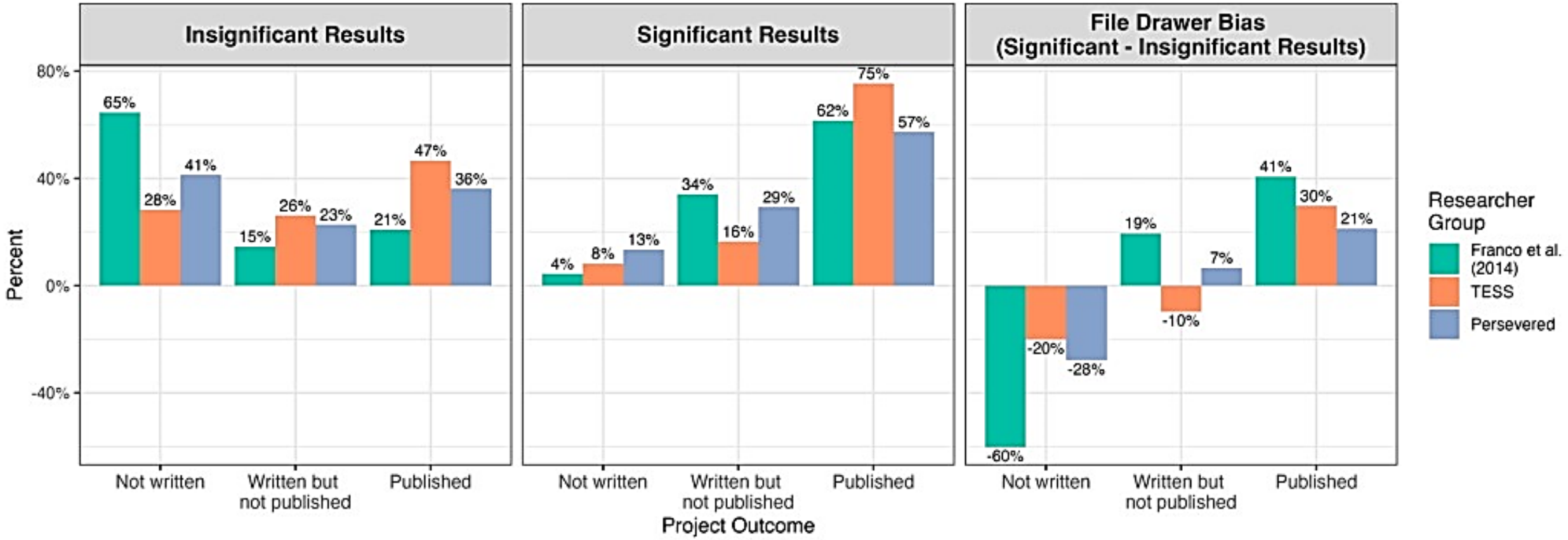


Fig. 1. File drawer bias across researcher groups.

Moniz, P., Druckman, J. N., & Freese, J. (**2025**). The file drawer problem in social science survey experiments. *Proceedings of the National Academy of Sciences*, *122*(12), e2426937122.