The Benefits of Preregistration and Registered Reports

Daniel Lakens

TU/e EINDHOVEN UNIVERSITY OF TECHNOLOGY

Study registries exist to reveal the existence of studies, published or not, to investigators and systematic reviewers.

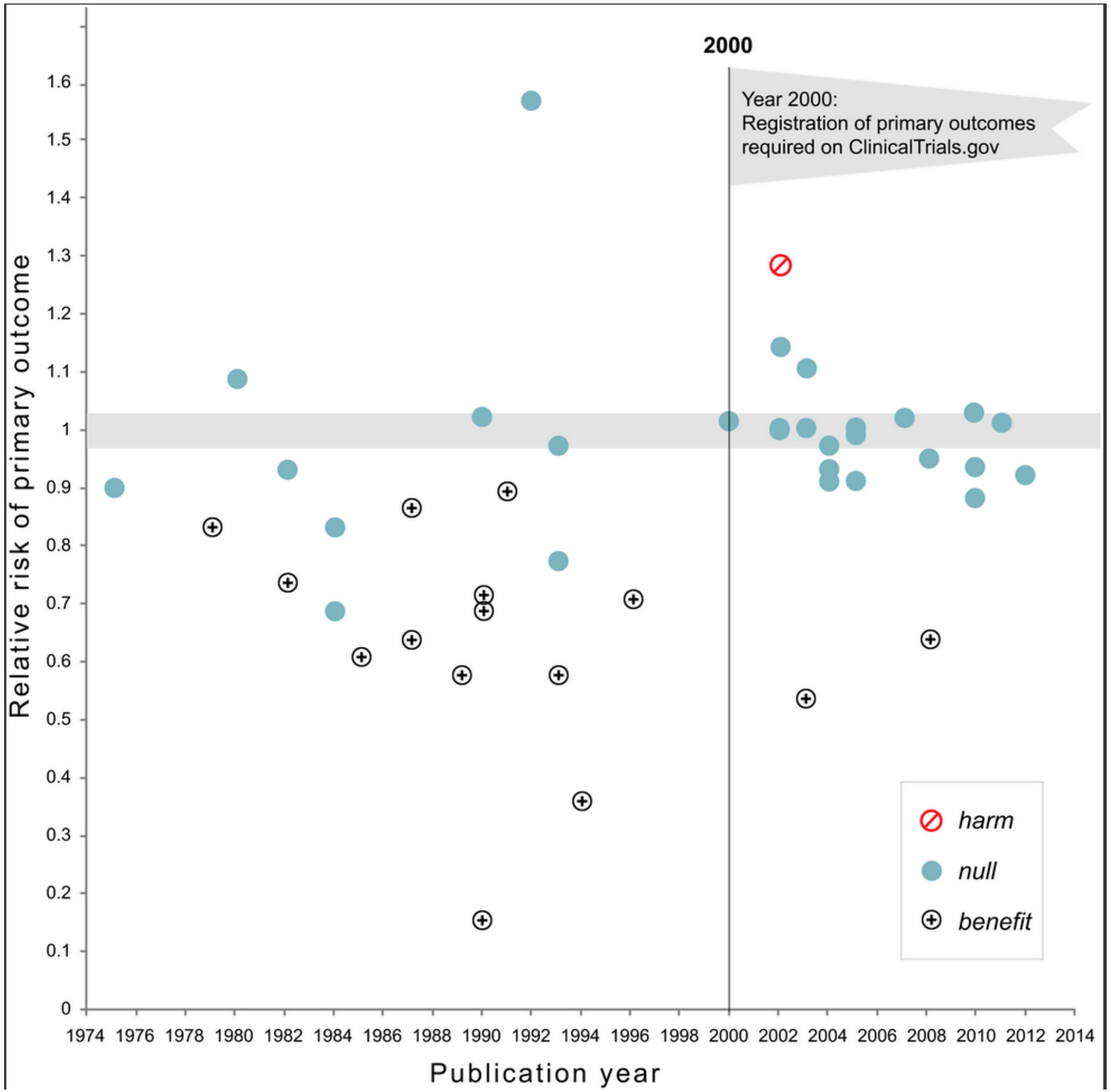
Clinicaltrials.gov was established in 1997, but not widely accepted or used.

Reasons they became more widely used: Legal claims for damages, and required by journals.

Glaxo-SmithKline (GSK) was sued in 2004 for failing to reveal the results of trials that showed an antidepressant might be harmful.

Also in 2004, the International Committee of Medical Journal Editors announced their (prestigious) journals would not publish reports of trials unless they had been registered.

Study registries can require a **preregistered** sampling and analysis plan.



Before 2000 17/30 large national Heart Lung and Blood Institute funded clinical trials showed a significant (+) effect. After pre-registration, only 2/25 showed a significant effect.

Kaplan & Irvin, 2015

If you plan a one-sided test, and find a p = 0.04, do you think peers might doubt that you planned the one-sided test from the beginning?

Now here there was a difficulty. The test of significance is not nearly so automatic an inference process as had been thought. It is manifestly contingent on the decision of the investigator as to whether to run a one- or a two-tailed test. And somehow, making the decision *after* the data were collected and the means computed, seemed like "cheating." How should this be handled?

Bakan, 1966

Should there be some central registry in which one registers one's decision to run a one- or two-tailed test before collecting the data? Should one, as one eminent psychologist once suggested to me, send oneself a letter so that the postmark would prove that one had pre-decided to run a one-tailed test?

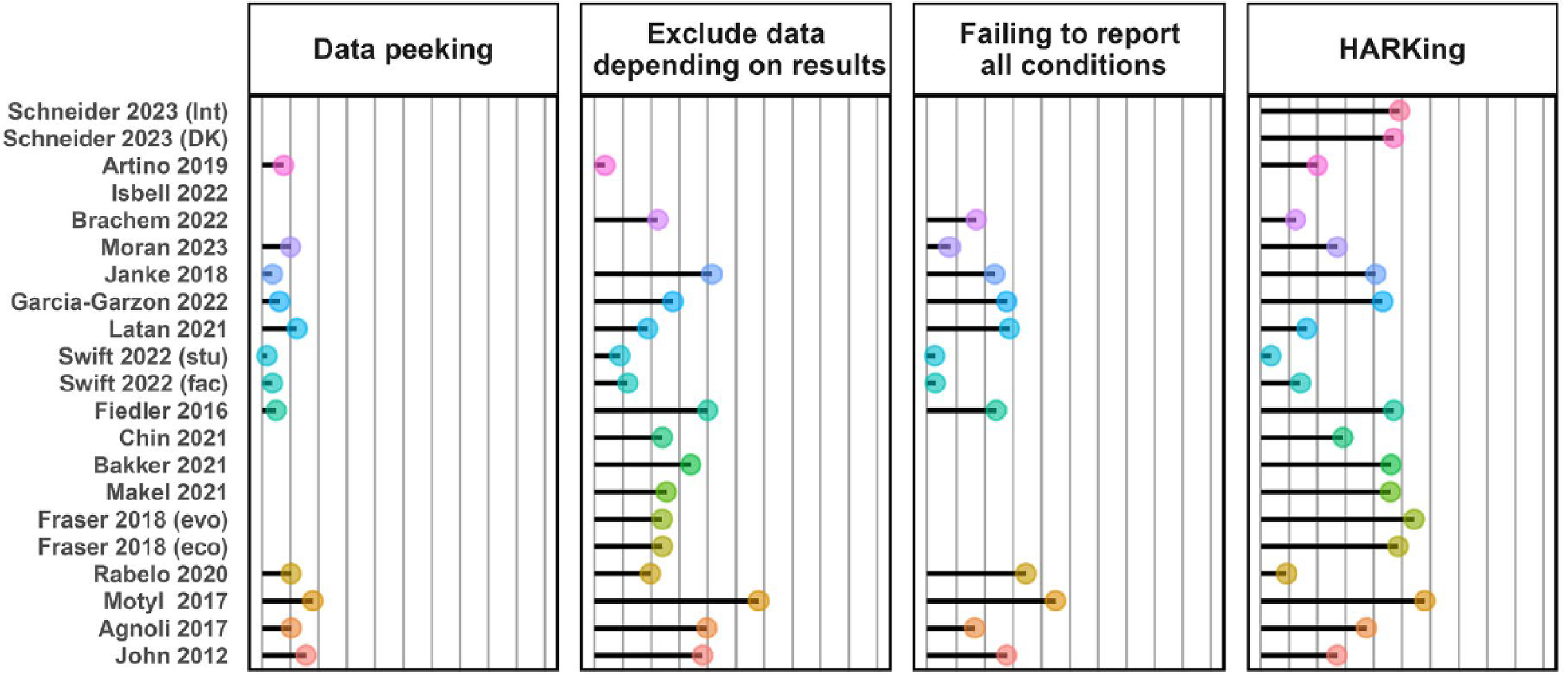
Bakan, 1966

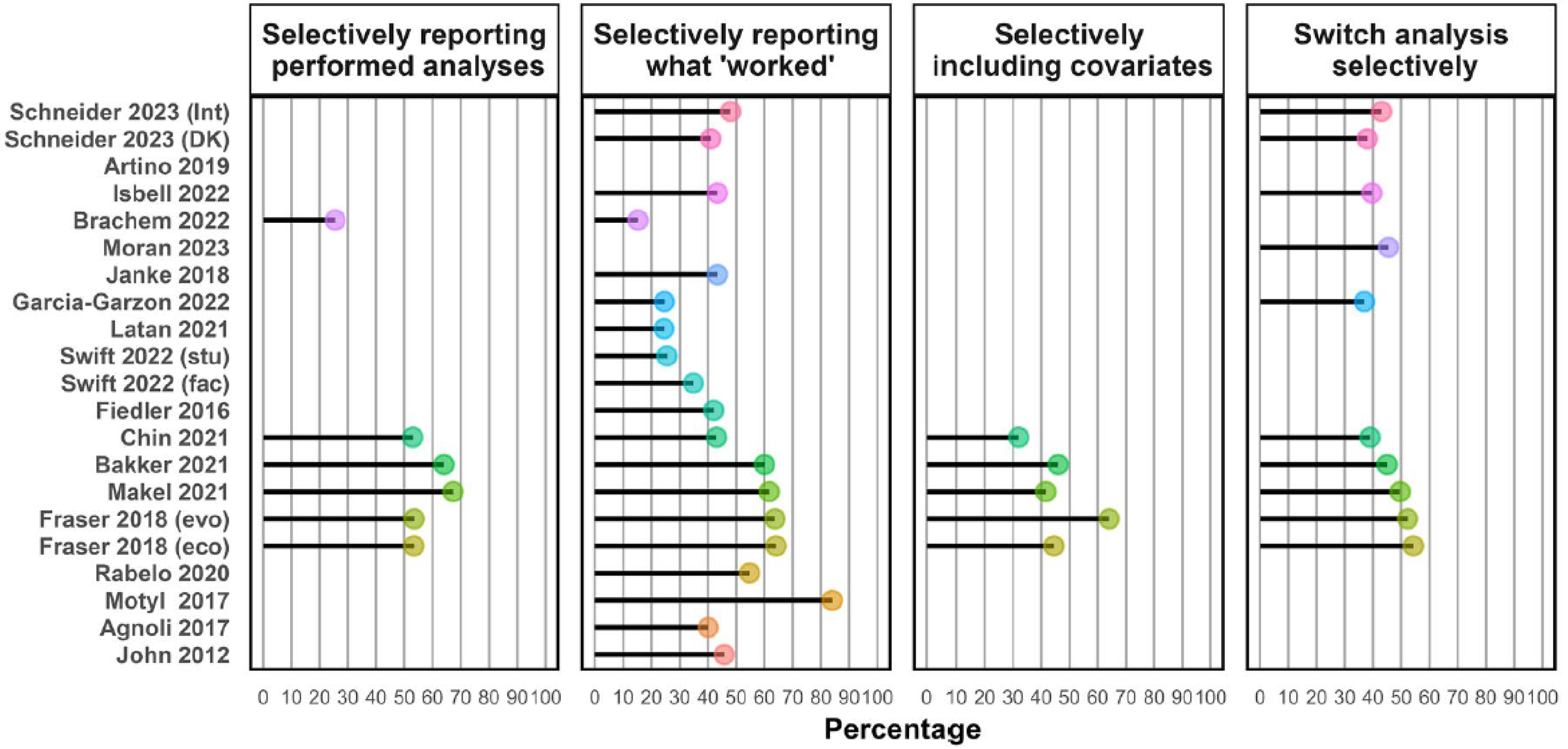
Over the last decade the field of meta-science has revealed many scientists self-admit to flexibly analyze their data.

When flexibility is opportunistically misused to select analyses that support a desired result this is called a ‘questionable research practice’.

In 15 surveys across disciplines many researchers admit to HARKing, excluding data to improve results, or selectively reporting results that ‘worked’.

Self-admission rates of engaging in each practice at least once in their career




Selectively reporting performed analyses
Selectively reporting what 'worked'
Selectively including covariates
Switch analysis selectively
Schneider 2023 (Int)
Schneider 2023 (DK)
Artino 2019
Isbell 2022
Brachem 2022
Moran 2023
Janke 2018
Garcia-Garzon 2022
Latan 2021
Swift 2022 (stu)
Swift 2022 (fac)
Fiedler 2016
Chin 2021
Bakker 2021
Makel 2021
Fraser 2018 (evo)
Fraser 2018 (eco)
Rabelo 2020
Motyl 2017
Agnoli 2017
John 2012
0 10 20 30 40 50 60 70 80 90 100
Percentage

But what exactly is the problem with HARKing?

Researchers can preregister their statistical analysis plan to allow others to **transparently evaluate how severely any claims were tested**.

Lakens, 2019

However, researchers also too easily assumed preregistered studies are always more compelling:

"This is particularly important if one wants to convince a skeptical audience of a controversial claim: After all, **confirmatory studies are much more compelling than exploratory studies**."

Wagenmakers, Wetzels, Borsboom, van der Maas, 2012

Taken together, these practices *[reducing p-hacking and publication bias, and power analysis]* will ensure that articles published as *Registered Reports* have a **substantially higher truth value** than regular studies. Such articles can therefore be expected to be more replicable and have a greater impact on the field.

Chambers, NeuroChambers blog, 2012

Preregistration clarifies the distinction between planned and unplanned research by reducing unnoticed flexibility. **This improves credibility of findings** and calibration of uncertainty.

Nosek, Beck, Campbell, Flake, Hardwicke, Mellor, van ‘t Veer, Vazire, 2019

In practice, confirmatory tests might be **much more compelling**, have **improved credibility of findings**, and **higher truth value.**

They might also not.

Preregistration adds value for people who, based on their philosophy of science, increase their trust in claims that are supported by severe tests and predictive successes.

Lakens, 2019

Preregistration itself does not make a study better or worse compared to a non-preregistered study – **as long as researchers are perfectly honest**.

There are strong indications that in some research lines QRP’s are one cause that makes it difficult to replicate published claims.

One example comes from research into a theory of self-control, known as *ego-depletion*. Let's look at four important studies.

In **2010** a meta-analysis of 198 independent tests of the 'ego-depletion effect' was published claiming there was a medium-to-large effect.

Hagger et al., 2010

Then, in **2014** re-analysis identified strong bias in the studies included in the meta-analysis, and estimated there might be no effect at all.

Carter & McCullough, 2014

In **2016** a preregistered replication study with 2141 participants found a non-significant ego-depletion effect very close to zero.

Hagger et al., 2016

In **2021** a preregistered replication study with 3531 participants, performed by original authors, also found a non-significant effect very close to zero.

Vohs et al., 2021

Preregistration is useful because it can prevent researchers from opportunistically abusing flexibility in the data analysis.

Preregistration transparently communicates which claims are made with a controlled error rate, and which not.

So far, we have focused on how preregistration prevents bias due to flexibility during the analysis. But there are other benefits.

Researchers who preregister report that preregistration resulted in an analysis plan that was more carefully thought-through, and some benefits for the experimental design and the research hypothesis.

Sarafoglou et al., 2022

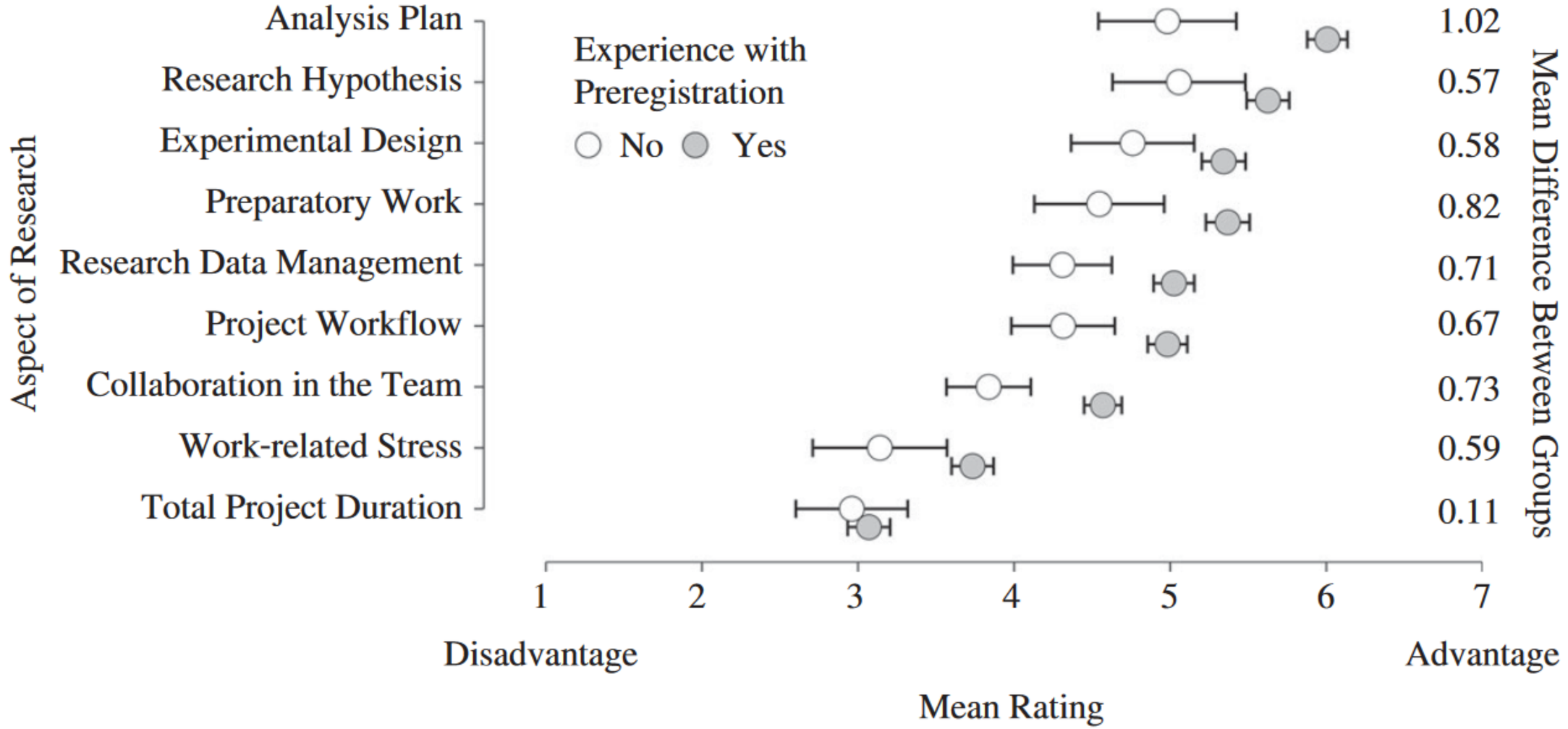


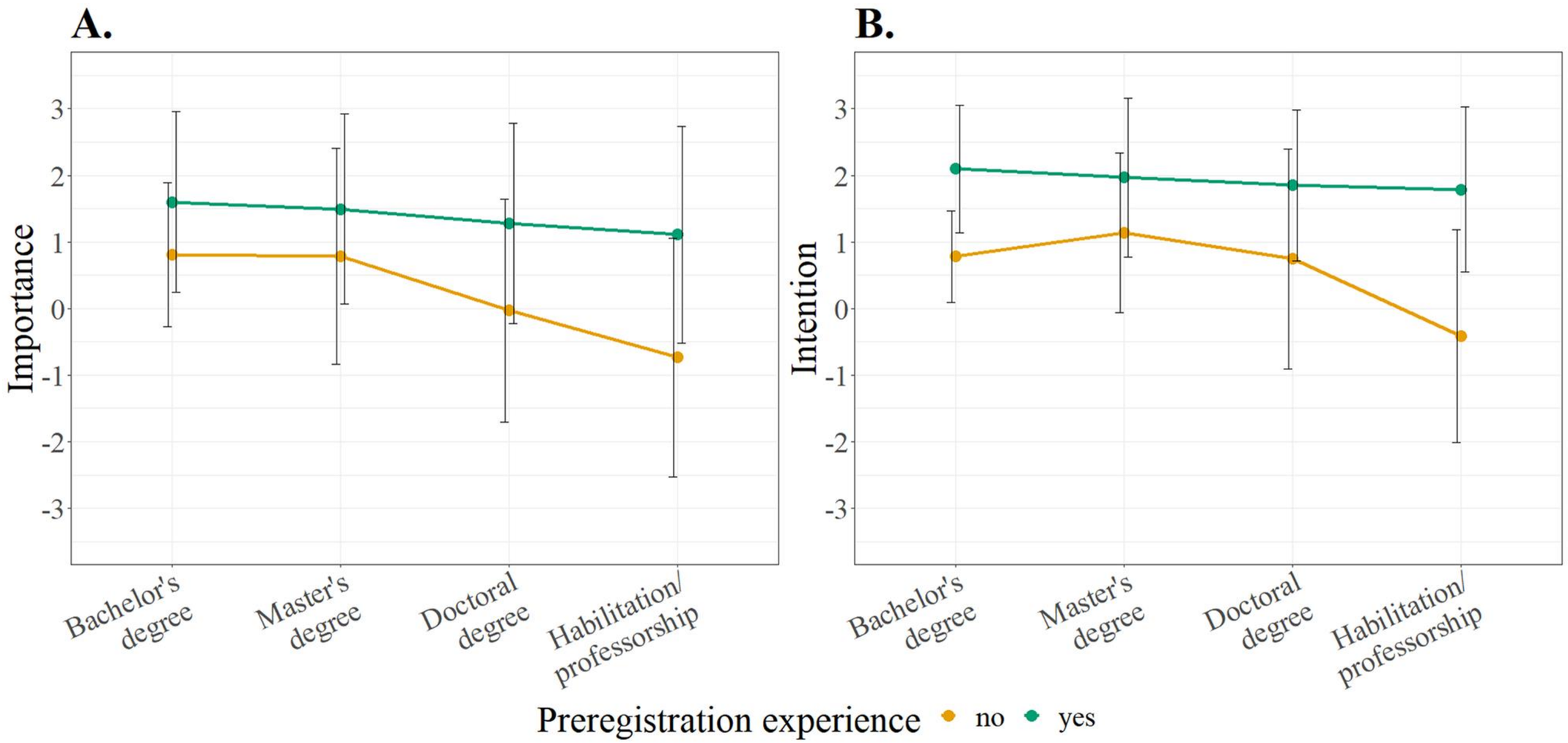
Figure 1. Respondents' opinion on how preregistration influenced different aspects of the research process. Grey dots represent the mean ratings from respondents who have experience with preregistration and white dots represent the mean ratings from respondents who have no experience with preregistration. The square skewers represent 95% confidence intervals. Ratings above and below 4 indicate that preregistration helped and harmed a certain research aspect, respectively.

Sarafoglou et al., 2022

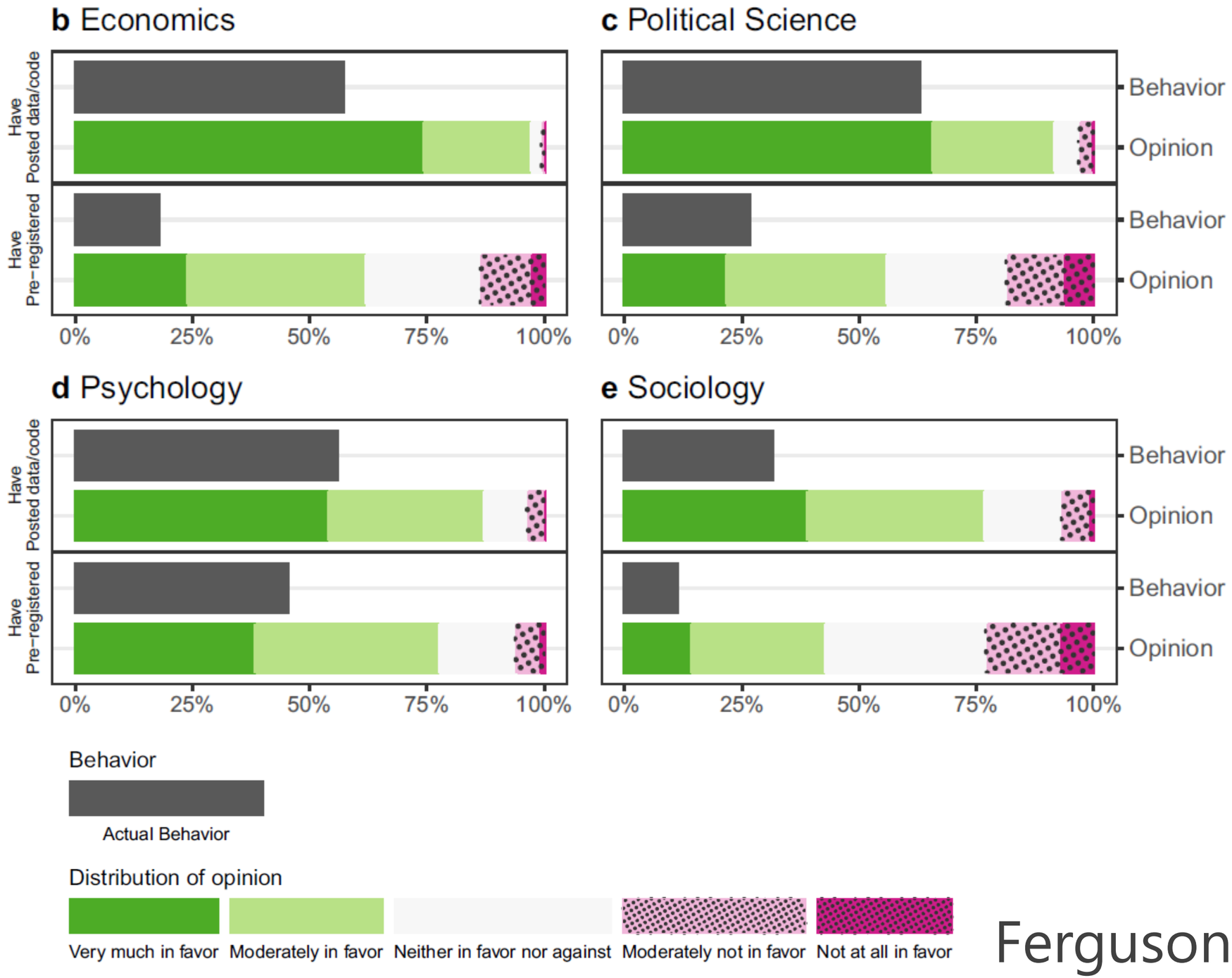
In this general sample (n = 288 responses), 61.81% of researchers indicated having used preregistration in the past. Main benefits are better planning and transparency. Main barrier is time.

Spitzer & Mueller, 2021

Perceived importance (A) and intention (B) to preregister:



Spitzer & Mueller, 2021



Ferguson et al., 2023

There has also been criticism on preregistration. Regrettably, most of it consists of “half-baked criticisms, raising issues that have already been fully addressed” (Syed 2024).

Some argue criticism is not needed, because “statistical problems become irrelevant because theories, not random selection, dictate what comparisons are necessary” (Szollosi et al. 2020).

Other argue against preregistration for the exact opposite reason: “rather than advocating preregistration as a means to foster more falsification-oriented, confirmatory research, it may be more realistic and productive to simply acknowledge that most consumer research is largely exploratory, thus limiting the epistemological value of traditional falsificationism.” (Pham & Oh, 2021)

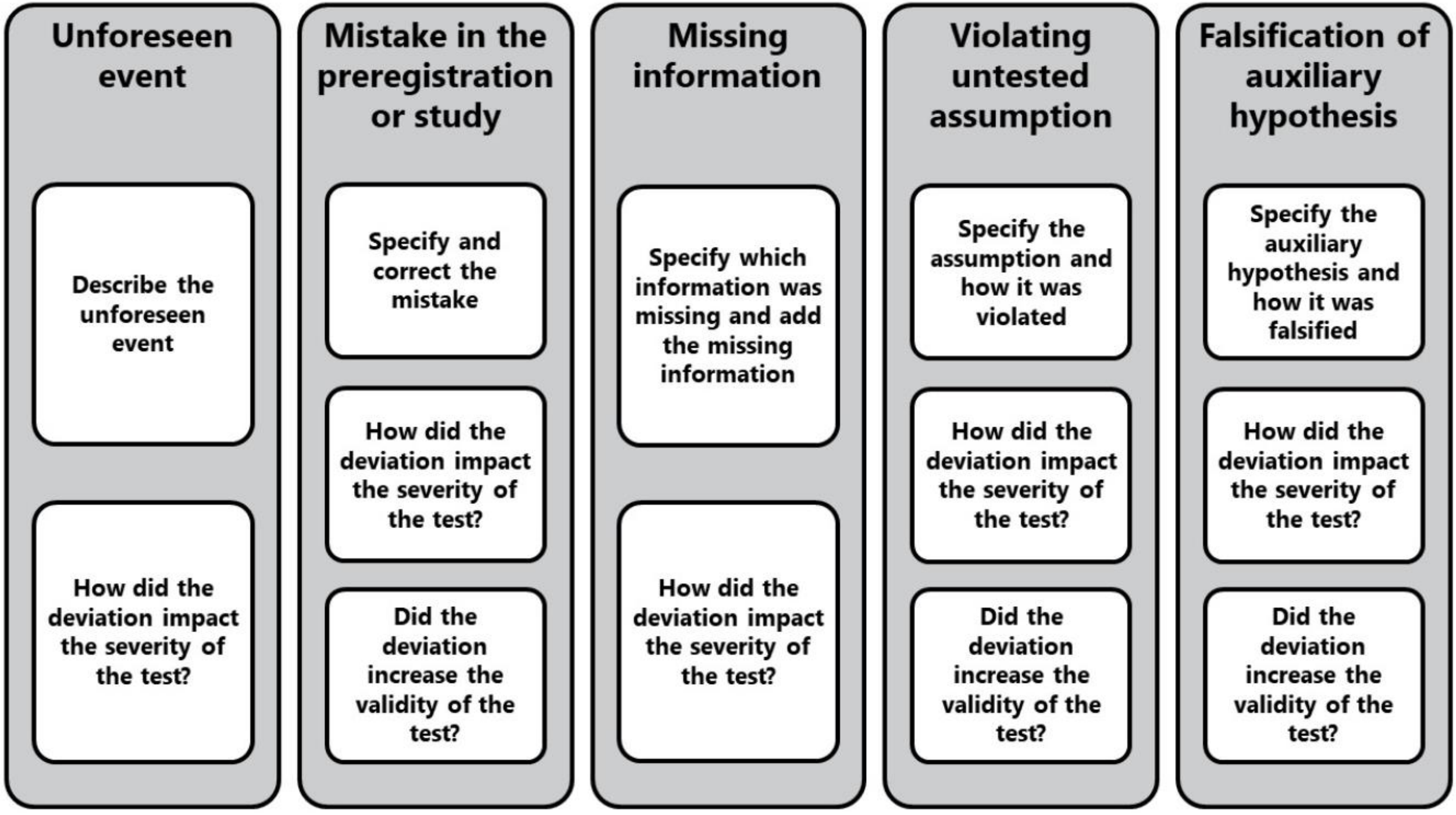
People warn preregistration will become a mindless heuristic to evaluate the quality of studies, it might prevent exploration, or people will stick to bad preregistered analysis plans.

As researchers have started to preregister, it turns out they often preregister uninformed predictions, and change their analysis plan.

Deviations can be improvements (as Meehl says: Don’t make a mockery of honest ad-hocery). Deviations trade guaranteed error control against a subjective evaluation of higher *validity*.

Table 1. Examples of reporting practices that lead to tests with higher or lower severity and claims with higher or lower validity.

	Lower validity	Higher validity
Lower severity	Selectively reporting one out of five variables that measure a construct of interest because only this test yields $p < .05$.	Deviating from a preregistration to exclude observations not caused by processes related to the research question.
Higher severity	Following a preregistered analysis of all data even though 15% of respondents did not follow the instructions.	Following a preregistered statistical analysis plan with high construct and statistical validity.


Unforeseen event
Describe the unforeseen event
How did the deviation impact the severity of the test?
Mistake in the preregistration or study
Specify and correct the mistake
How did the deviation impact the severity of the test?
Did the deviation increase the validity of the test?
Missing information
Specify which information was missing and add the missing information
How did the deviation impact the severity of the test?
Violating untested assumption
Specify the assumption and how it was violated
How did the deviation impact the severity of the test?
Did the deviation increase the validity of the test?
Falsification of auxiliary hypothesis
Specify the auxiliary hypothesis and how it was falsified
How did the deviation impact the severity of the test?
Did the deviation increase the validity of the test?

Should we preregister qualitative research, and secondary data analysis?

I think not.

Open lab notebooks might be a more coherent method to communicate transparently.

Preregistration is a specific tool, with a specific goal.

Evaluation of Statistical Hypotheses

11 September, 2020

Kinship and Prosocial Behaviour Postregistration

- DeBruine, Lisa M.
  - roles: Conceptualization, Data curation, Software, Writing - original draft, Writing - review & editing
  - email: lisa.debruine@glasgow.ac.uk
- Lakens, Daniël
  - roles: Conceptualization, Formal analysis, Writing - original draft, Writing - review & editing

Abstract

A reanalysis of data from DeBruine (2002) Facial Resemblance Enhances Trust, PRSLB.

Results

Hypothesis 1: self_pref

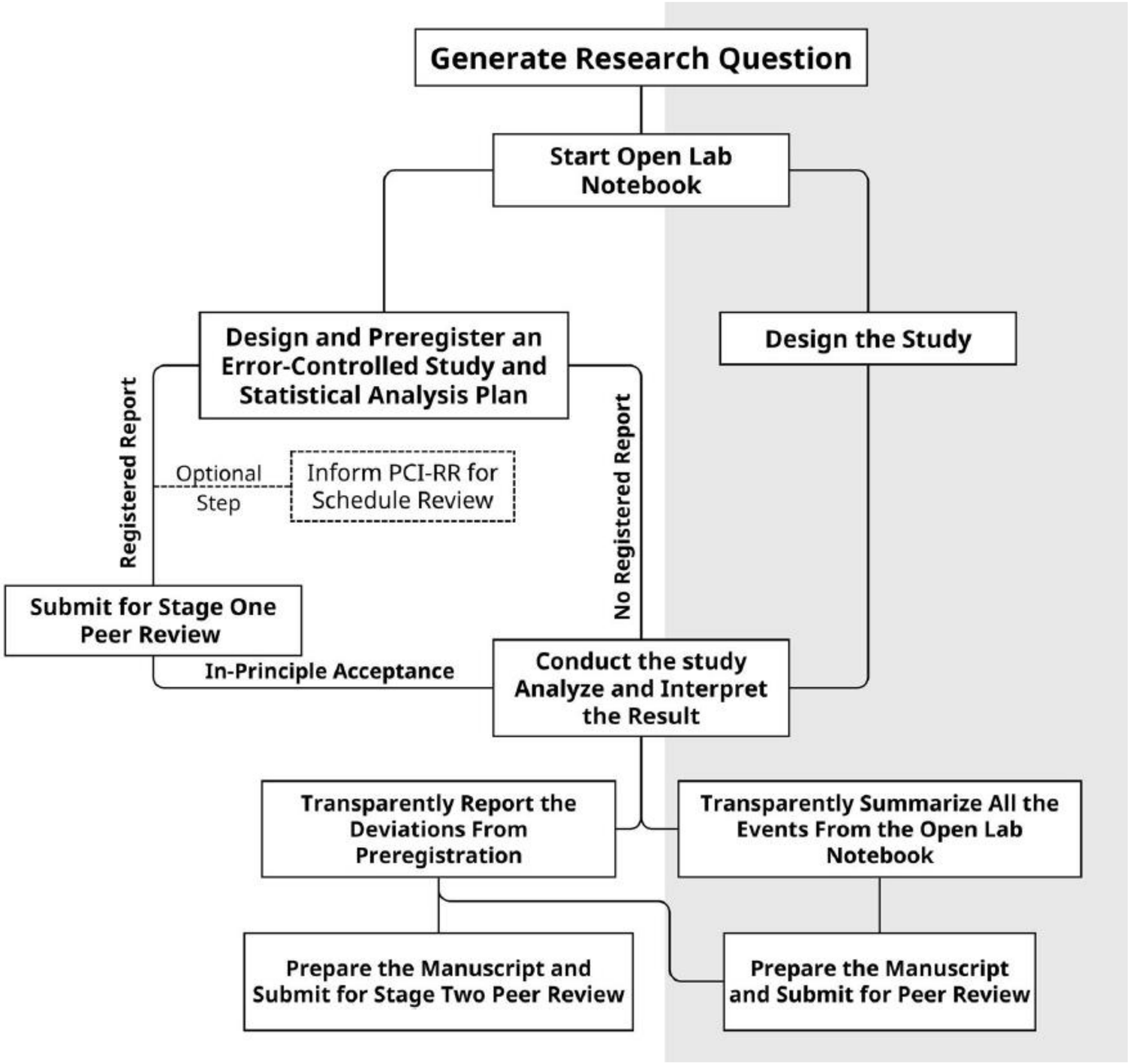
Cues of kinship will increase prosocial behaviour. Cues of kinship will be manipulated by morphed facial self-resemblance. Prosocial behaviour will be measured by responses in the trust game. The prediction is that the number of trusting AND/OR reciprocating moves will be greater to self morphs than to other morphs.

- `t_lo` is confirmed if analysis trust yields conf.int[1] > 0 The result was conf.int[1] = 0.021 (TRUE)
- `t_hi` is confirmed if analysis trust yields conf.int[2] > 0.2 The result was conf.int[2] = 0.979 (TRUE)
- `r_lo` is confirmed if analysis recip yields conf.int[1] > 0 The result was conf.int[1] = -0.509 (FALSE)
- `r_hi` is confirmed if analysis recip yields conf.int[2] > 0.2 The result was conf.int[2] = 0.426 (TRUE)

Lakens & DeBruine, 2021

A novel publication format, known as Registered Reports, offers the opportunity to get peer reviews before analyzing data, and can guarantee the publication of well-designed studies - regardless of whether results are significant or not.

Registered Reports are a novel article publication format that takes place in four steps:



Lakens, Mesquida, Rasti, & Ditroilo, 2024

Step 1: Authors develop the study rationale, design, and analysis plan.

- If predictions are tested the error rates are controlled.
- If authors want to explore data they specify which tests lead to claims without error control.

Step 2: Stage 1 Peer review leads to either a rejection, revisions, or in principle acceptance.

- The peer review process is not influenced by the results, as no data has been collected
- After in principle acceptance the journal commits to publishing the article as long as researchers follow the peer reviewed plan.

Step 3: Data collection and/or preparation, analysis, and writing the complete manuscript.

- If unforeseen circumstances arise authors can contact the editor to discuss deviations from their plan.
- If necessary, peer reviewers will be consulted, and changes are approved (or not). Researchers can update their preregistration to log any changes.

Step 4: Stage 2 Peer review, final acceptance and publication.

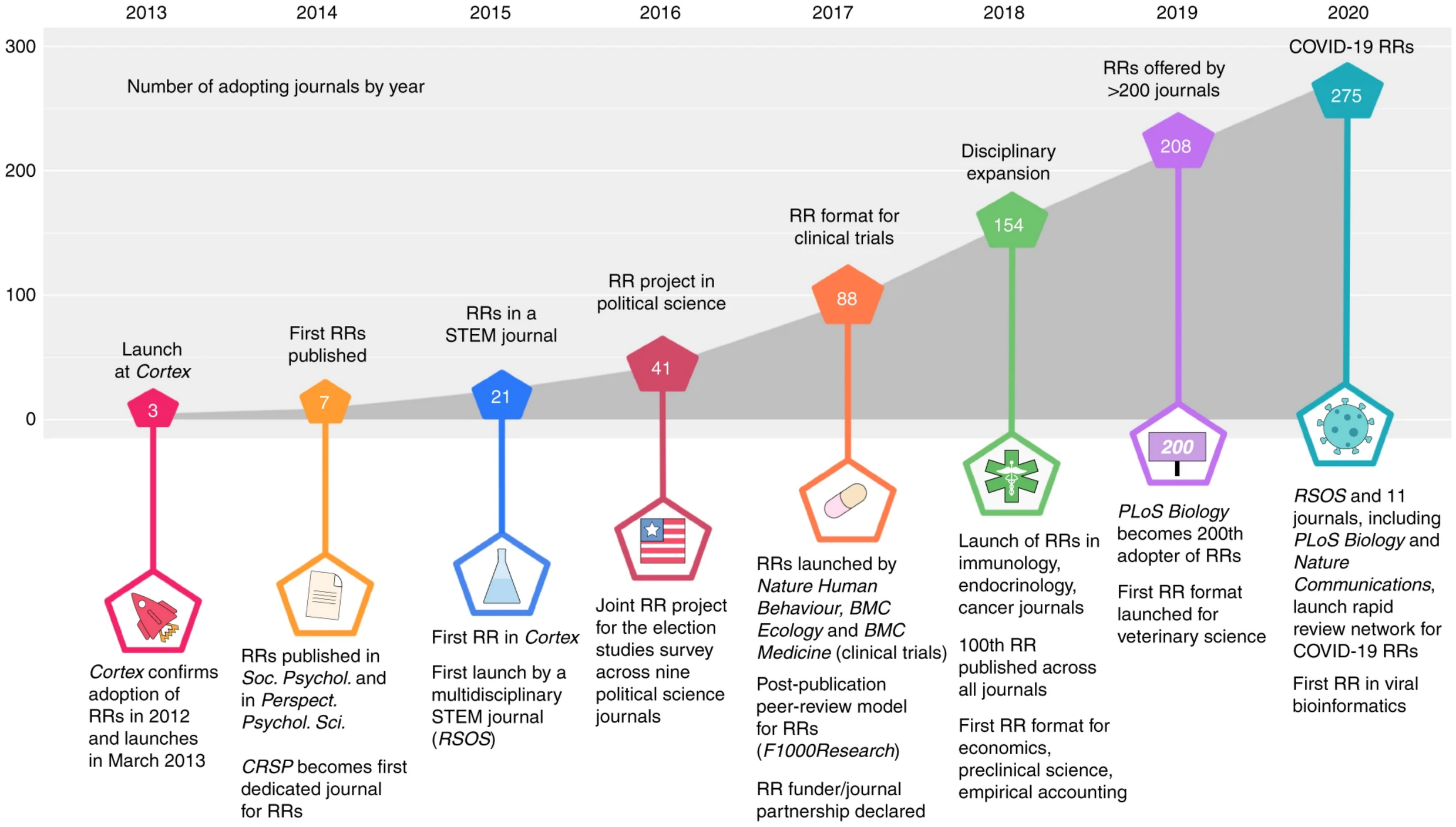
- Peer reviewers check if conclusions follow from the data, and if the analysis plan is followed (or deviations clearly justified).
- Rejection can happen in extreme cases where certain quality checks show methodological problems lead to an uninformative study.

Peer review before data collection has the benefit that any issues identified by peers can be improved before it is too late.

Data collection can only start after the editorial decision is made, which requires planning.

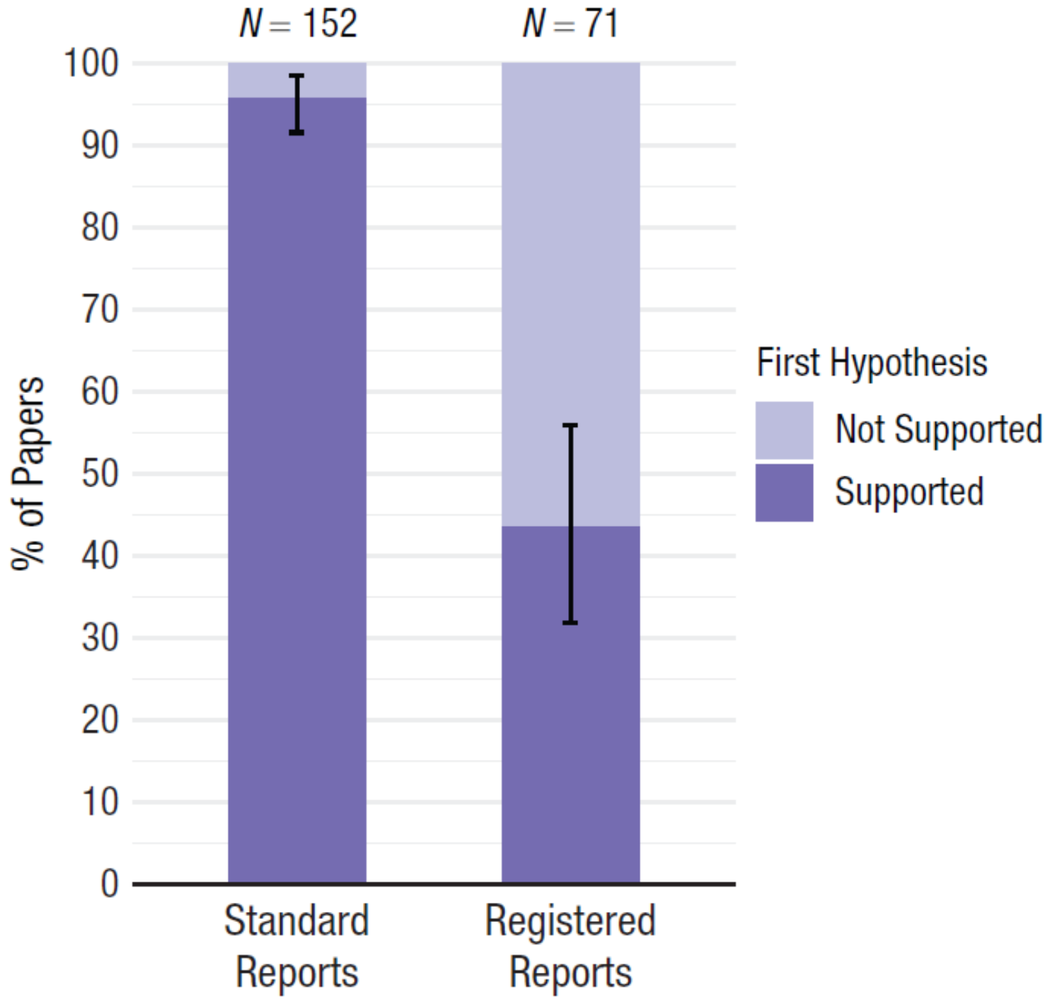
The first Registered Reports were published in 2014. More than 300 journals now offer Registered Reports.

Nosek & Lakens, 2014


2013
2014
2015
2016
2017
2018
2019
2020
300
200
100
0
Number of adopting journals by year
3
7
21
41
88
154
208
275
200
Launch at *Cortex*
First RRs published
RRs in a STEM journal
RR project in political science
RR format for clinical trials
Disciplinary expansion
RRs offered by >200 journals
COVID-19 RRs
Cortex confirms adoption of RRs in 2012 and launches in March 2013
RRs published in *Soc. Psychol.* and in *Perspect. Psychol. Sci.*
CRSP becomes first dedicated journal for RRs
First RR in *Cortex*
First launch by a multidisciplinary STEM journal (*RSOS*)
Joint RR project for the election studies survey across nine political science journals
RRs launched by *Nature Human Behaviour, BMC Ecology* and *BMC Medicine* (clinical trials)
Post-publication peer-review model for RRs (*F1000Research*)
RR funder/journal partnership declared
Launch of RRs in immunology, endocrinology, cancer journals
100th RR published across all journals
First RR format for economics, preclinical science, empirical accounting
PLoS Biology becomes 200th adopter of RRs
First RR format launched for veterinary science
RSOS and 11 journals, including *PLoS Biology* and *Nature Communications*, launch rapid review network for COVID-19 RRs
First RR in viral bioinformatics

Scheel et al. (2021) show that, in one of the most replicable findings in science, there is widespread publication bias, with 96% of traditional publications yielding significant results.

While in the standard literature 96% of reported hypotheses are confirmed, In Registered Reports only 44% of tested hypotheses confirm predictions.

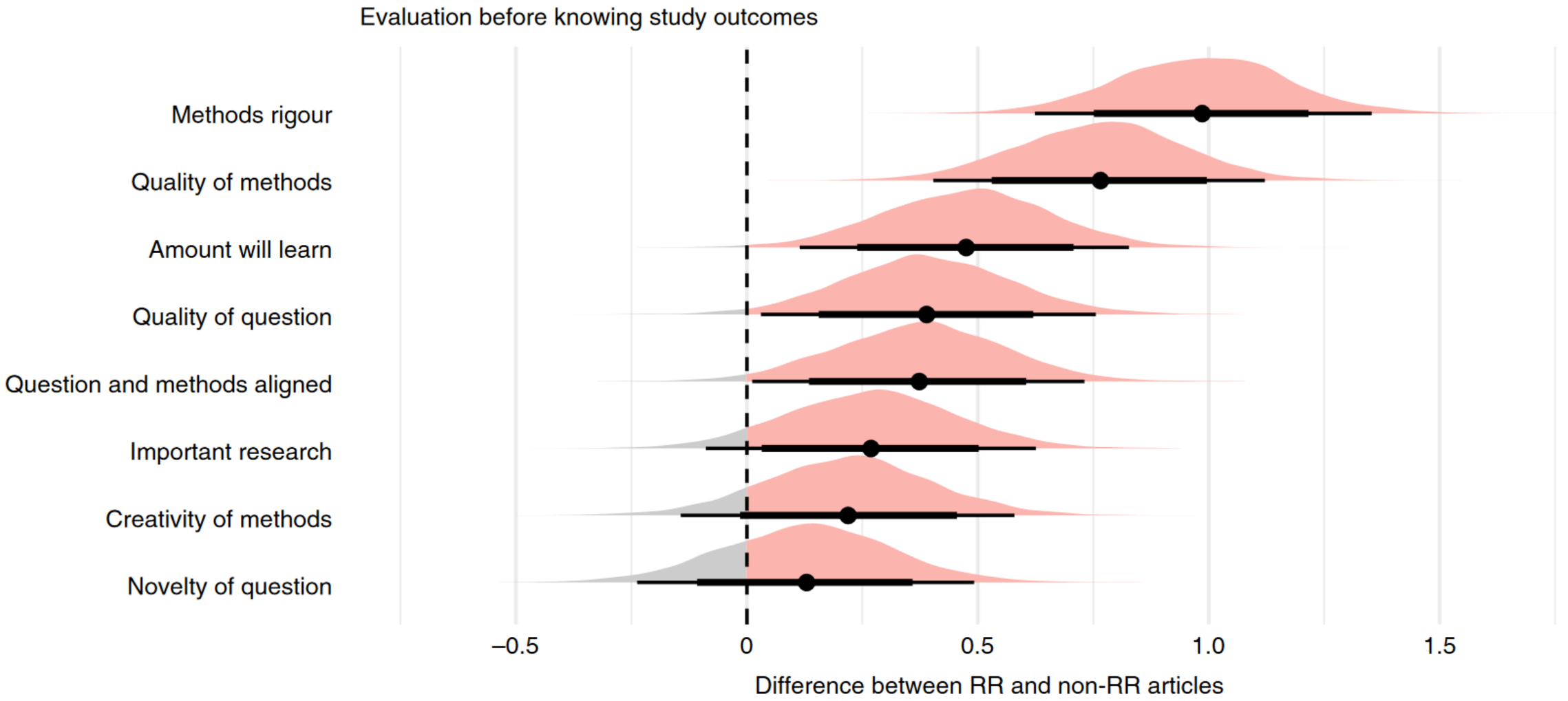


Scheel et al., 2021

This suggests that Registered Reports are a useful publication format to increase the number of non-significant results in the scientific literature.

It also shows that null-results are surprisingly common, even if we don’t always see them! Not finding support for your prediction is a part of doing good science.

Initial meta-scientific research shows that peer reviewers evaluate the quality of Registered Reports more positively than the quality of Standard Reports.



Soderberg et al., 2023

Registered Reports combine preregistration (reducing opportunistic *flexibility in the analysis)* with a journal article format that reduces publication bias.

Peer Community In Registered Reports review RR’s outside of the journal system, and can be scheduled to make it fast.

If you test hypotheses, I strongly advise to preregister. If you want peer feedback before publication, or think null results are difficult to publish, try a Registered Report.

Grazi!

https://osf.io/ejqa2/