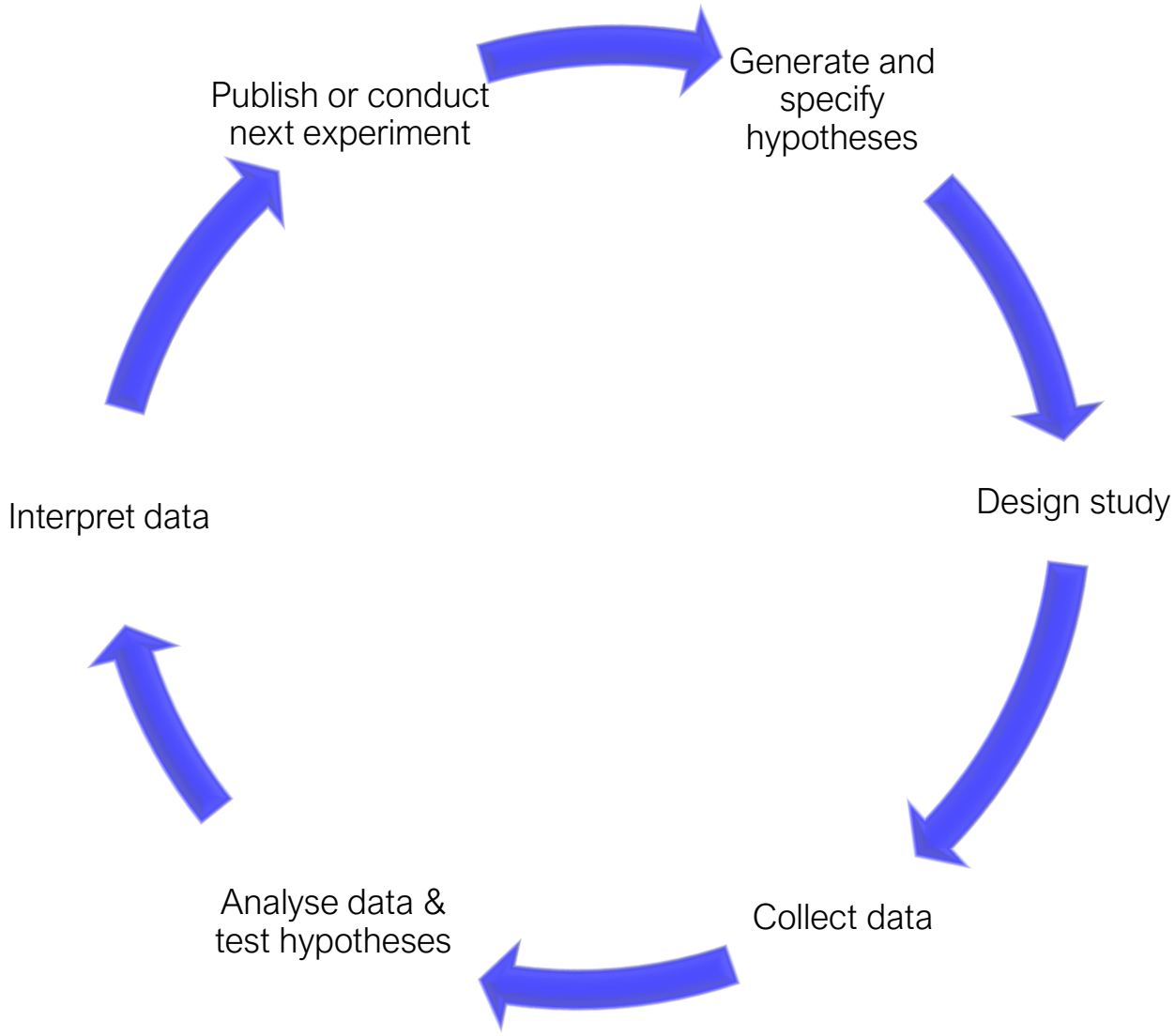
Reproduciblity crisis in science

Thomas Guillemaud, Inrae, ISA & Peer Community In
&
Denis Bourguet, Inrae, CBGP & Peer Community In

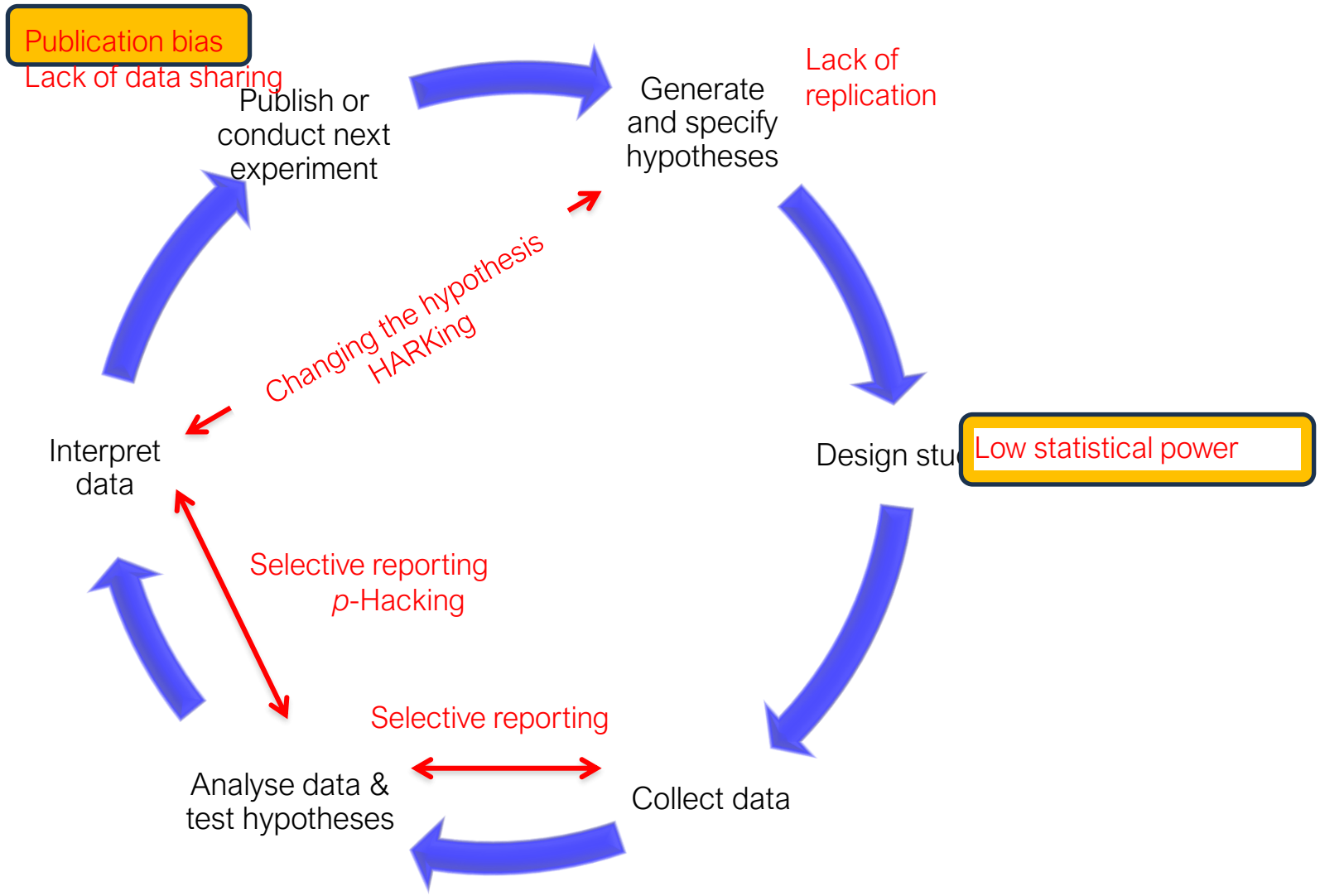
Some slides are inspired from Chris Chambers

School of Psychology, Cardiff University

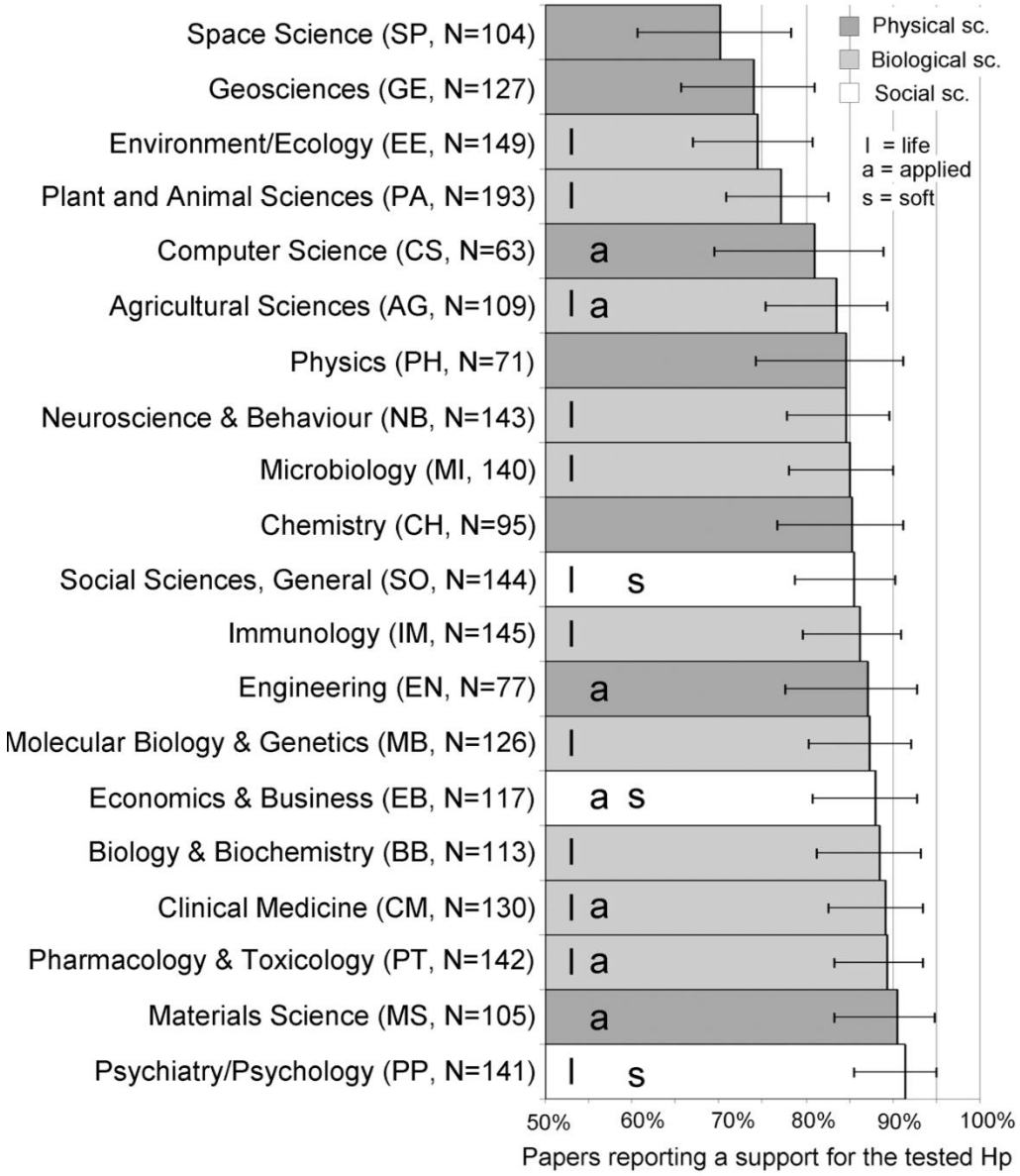
Designing, running and publishing a study



What happens when we put researchers under pressure to get "great results"?



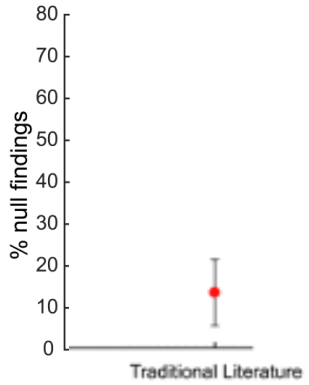
Most published results are positive



- Overall, 2,045 papers (84%) reported a positive or partial support for the tested hypothesis

- Fanelli D (2010) "Positive" Results Increase Down the Hierarchy of the Sciences. *PLOS ONE*, **5**, e10068. https://doi.org/10.1371/journal.pone.0010068

Percentage of null findings



Allen C, Mehler DMA (2019) Open science challenges, benefits and tips in early career and beyond. PLOS Biol 17(5): e3000246. https://doi.org/10.1371/journal.pbio.3000246

PERCENTAGE OF POSITIVE RESULTS

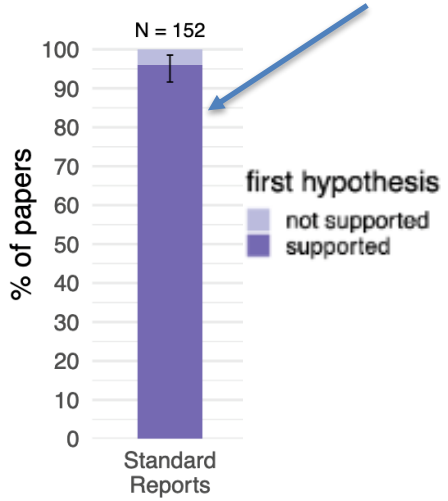
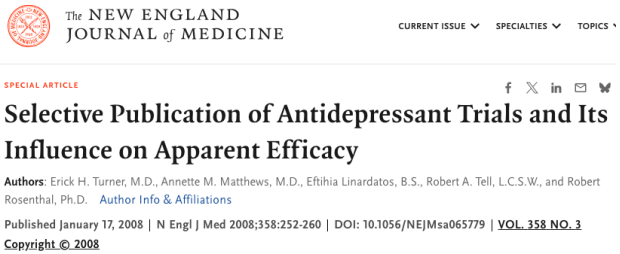


Figure 2. Positive result rates for standard reports and Registered Reports. Error bars indicate 95% confidence intervals around the observed positive result rate.

Scheel, Schijen & Lakens (2021) https://journals.sagepub.com/doi/full/10.1177/25152459211007467


The NEW ENGLAND JOURNAL of MEDICINE

CURRENT ISSUE SPECIALTIES TOPICS

SPECIAL ARTICLE

Selective Publication of Antidepressant Trials and Its Influence on Apparent Efficacy

Authors: Erick H. Turner, M.D., Annette M. Matthews, M.D., Eftihia Linardatos, B.S., Robert A. Tell, L.C.S.W., and Robert Rosenthal, Ph.D. Author Info & Affiliations

Published January 17, 2008 | N Engl J Med 2008;358:252-260 | DOI: 10.1056/NEJMsa065779 | VOL. 358 NO. 3

Copyright © 2008

94% of the trials conducted were positive

etc...

Main factors that make your tests significant

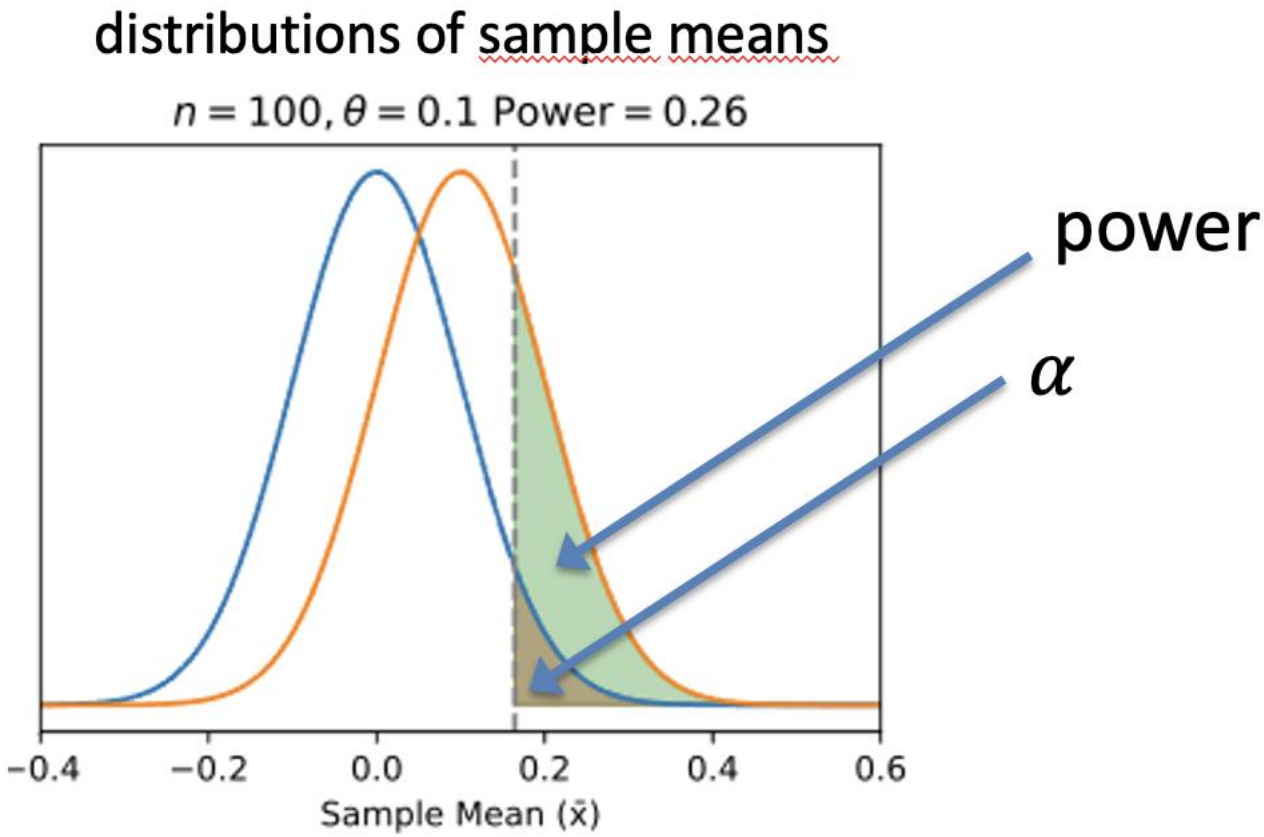
- Effect sizes are real
- Low variability of biological phenomena (genetics, environment, epigenetics, development) and measurements
- Sample size are large enough

Power = probability of rejecting Ho when Ho false

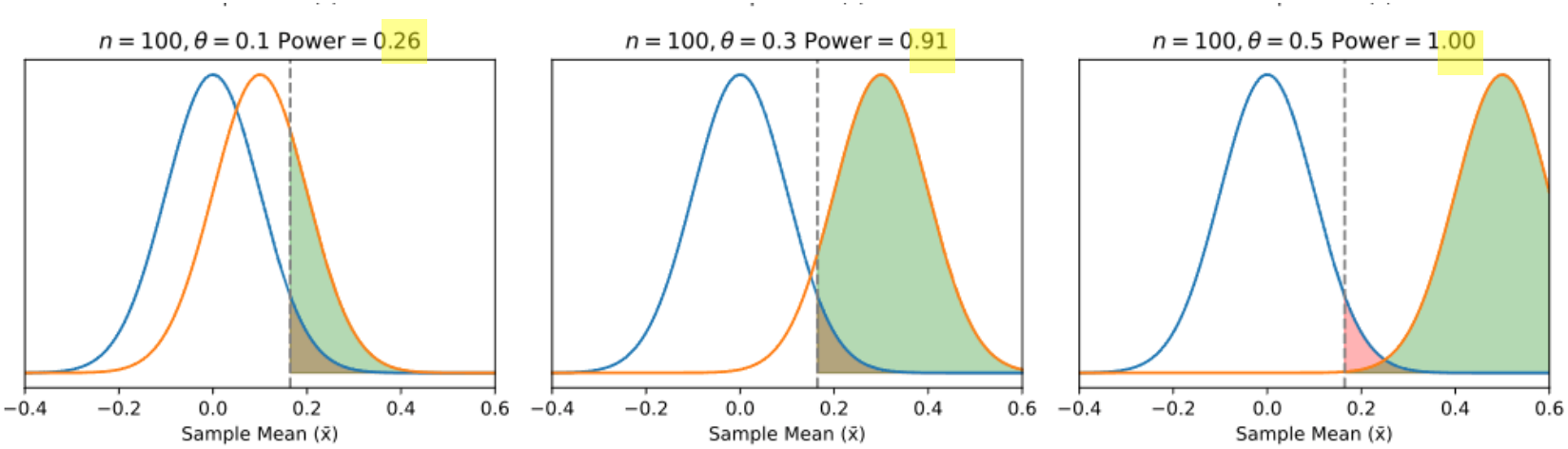
= probability of getting a significant effect when there is an effect

= our capacity to detect effects

= true effect = true positive



distributions of sample means



Distribution of statistical Power in the litterature

Behavioral ecology :

Jennions MD, Møller AP (2003) A survey of the statistical power of research in behavioral ecology and animal behavior. *Behavioral Ecology*, **14**, 438–445. https://doi.org/10.1093/beheco/14.3.438

We estimated the statistical power of the first and last statistical test presented in 697 papers from 10 behavioral journals. First tests had significantly greater statistical power and reported more significant results (smaller p values) than did last tests. This tre
a si
this
effe
valu
pov
Me
bet
effe
sug
s 13–16% to detect
a power of 80%. By
l, medium, or large
. However, mean p
ilar trend for mean
le among journals.
nificant correlation
, $p < .0001$). If true
rn to emerge. This
es with sample size.

Key words: effect size, meta-analysis, publication bias, sample sizes, statistical power. *[Behav Ecol 14:438–445 (2003)]*

in 697 papers form 10 behavioral journals [...] on average, statistical power was 13-16% to detect a small effect and 40-47% to detect medium effect [...] only 2-3%, 13-21% and 37-50% of the test examined had the requisite power to detect a amll, medium or large effect, respectively

Ecology and Bio Evo

Yang Y, Sánchez-Tójar A, O'Dea RE, Noble DWA, Koricheva J, Jennions MD, Parker TH, Lagisz M, Nakagawa S (2023) Publication bias impacts on effect size, statistical power, and magnitude (Type M) and sign (Type S) errors in ecology and evolutionary biology. *BMC Biology*, **21**, 71. https://doi.org/10.1186/s12915-022-01485-y

Abstract

Collaborative efforts to directly replicate empirical studies in the medical and social sciences have revealed alarmingly low rates of replicability, a phenomenon dubbed the 'replication crisis'. Poor replicability has spurred cultural changes targeted at improving reliability in these disciplines. Given the absence of equivalent replication projects in ecology

Ecological and evolutionary studies consistently had low statistical power (15%) with a 4-fold exaggeration of effects on average

in meta-analytic means being over-estimated by (at least) 0.12 standard deviations. The prevalence of publication bias distorted confidence in meta-analytic results, with 66% of initially statistically significant meta-analytic means becoming non-significant after correcting for publication bias. Ecological and evolutionary studies consistently had low statistical power (15%) with a 4-fold exaggeration of effects on average (Type M error rates = 4.4). Notably, publication bias reduced power from 23% to 15% and increased type M error rates from 2.7 to 4.4 because it creates a non-random sample of effect size evidence. The sign errors of effect sizes (Type S error) increased from 5% to 8% because of publication bias. Our research provides clear evidence that many published ecological and evolutionary findings are inflated. Our results highlight the importance of designing high-power empirical studies (e.g., via collaborative team science), promoting and encouraging replication studies, testing and correcting for publication bias in meta-analyses, and adopting open and transparent research practices, such as (pre)registration, data- and code-sharing, and transparent reporting.

Neurosciences

Button KS, Ioannidis JPA, Mokrysz C, Nosek BA, Flint J, Robinson ESJ, Munafò MR (2013) Power failure: why small sample size undermines the reliability of neuroscience. *Nature Reviews Neuroscience*, **14**, 365–376. https://doi.org/10.1038/nrn3475

Table 1 | **Characteristics of included meta-analyses**

First author of study	k	N	Summary effect size			Power	Refs
		Median (range)	Cohen's d	OR	Random or fixed effects	Median (range)	
Babbage	13	48 (24–67)	–1.11		Random	0.96 (0.74–0.99)	24
Bai	18	322 (92–3152)		1.47	Random	0.20 (0.06–1.00)	25
Bjorkhem-Bergman	6	59 (37–72)	–1.20		Random	0.99 (0.94–1.00)	26
Bucossi	21	85 (19–189)	0.41		Random	0.46 (0.13–0.79)	27
Chamberlain	11	53 (20–452)	–0.51		NA	0.54 (0.33–1.00)	28
Chang	56	55 (20–309)	–0.19		Random	0.10 (0.07–0.38)	29
Chang	6	616.5 (157–1492)		0.98	Fixed	0.05 (0.05–0.06)	30
Chen	12	1193 (288–29573)		0.60	Random	0.92 (0.13–1.00)	31
Chung	11	253 (129–703)		0.67	Fixed	0.09 (0.00–0.15)	32
Domellof	14	143.5 (42–5795)		2.12	Random	0.47 (0.00–1.00)	33
Etminan	14	109 (31–753)		0.80	Random	0.08 (0.05–0.23)	34
Feng	4	450 (370–1715)		1.20	Fixed	0.16 (0.09–0.42)	35
Green	17	69 (29–687)	–0.59		Random	0.65 (0.34–1.00)	36
Han	14	212 (40–4190)		1.35	Random	0.12 (0.05–0.95)	37
Hannestad	13	23 (12–100)	–0.13		Random	0.09 (0.07–0.25)	38
Hua	27	468 (114–1522)		1.13	Random	0.09 (0.06–0.22)	39
Lindson	8	257 (48–1100)		1.05	Fixed	0.05 (0.05–0.06)	40
Liu	12	563 (148–1956)		1.04	Fixed	0.05 (0.05–0.07)	41
Lui	6	1678 (1033–9242)		0.89	Fixed	0.15 (0.12–0.60)	42
MacKillop	57	52 (18–227)	0.58		Fixed	0.51 (0.21–0.99)	43
Maneeton	5	53 (22–162)		1.67*	Random	0.13 (0.08–0.35)	44
Ohi	6	674 (200–2218)		1.12	Fixed	0.10 (0.07–0.24)	45
Olabi	14	68.5 (14–209)	–0.40		Random	0.34 (0.13–0.83)	46
Oldershaw	10	65.5 (40–126)	–0.51		Random	0.53 (0.35–0.79)	47
Oliver	7	156 (66–677)		0.86	Fixed	0.07 (0.06–0.17)	48
Peerbooms	36	229 (26–2913)		1.26	Random	0.11 (0.00–0.36)	49
Pizzagalli	22	16 (8–44)	0.92		Random	0.44 (0.19–0.90)	50
Rist	5	150 (99–626)		2.06	Random	0.55 (0.35–0.98)	51
Sexton	8	35 (20–208)	0.43		Fixed	0.24 (0.15–0.98)	52
Shum	11	40 (24–129)	0.89		Fixed	0.78 (0.54–0.93)	53
Sim	2	72 (46–98)		1.23*	Random	0.07 (0.07–0.08)	54
Song	12	85 (32–279)	0.15		NA	0.10 (0.07–0.21)	55
Sun	6	437.5 (158–712)		1.93	Fixed	0.65 (0.14–0.98)	56
Tian	4	50 (32–63)	1.26		NA	0.98 (0.93–1.00)	57
Trzesniak	11	124 (55–279)		1.98	Random	0.27 (0.09–0.64)	58
Veehof	8	58.5 (19–156)	0.37		Fixed	0.26 (0.12–0.60)	59
Vergouwen	24	223 (39–1015)		0.83	Random	0.09 (0.06–0.22)	60
Vieta	10	212 (113–361)		0.68*	Random	0.27 (0.16–0.39)	61
Wisdom	53	137 (20–7895)	–0.14		NA	0.12 (0.06–1.00)	62
Witteman	26	28 (15–80)	–1.41		Random	0.94 (0.66–1.00)	63
Woon	24	30 (8–68)	–0.60		Random	0.36 (0.11–0.69)	64
Xuan	20	348.5 (111–1893)		1.00	Random	0.05 (0.05–0.05)	65
Yang (cohort)	14	296 (100–1968)		1.38*	Random	0.18 (0.11–0.79)	66
Yang (case control)	7	126 (72–392)		2.48	Random	0.73 (0.43–0.93)	66

Neuroimaging

"Our results indicated that the median statistical power of these studies was 8% across 461 individual studies"

Animal model:

19 studies: "Our results indicate that the median statistical power for the water maze studies and the radial maze studies to detect these medium to large effects was 18% and 31%"

Clinical trials

Lamberink HJ, Otte WM, Sinke MRT, Lakens D, Glasziou PP, Tijdink JK, Vinkers CH (2018) Statistical power of clinical trials increased while effect size remained stable: an empirical analysis of 136,212 clinical trials between 1975 and 2014. *Journal of Clinical Epidemiology*, **102**, 123–128. https://doi.org/10.1016/j.jclinepi.2018.06.014

Data from 136.212 clinical trials were available. from 11.852 meta-analyses

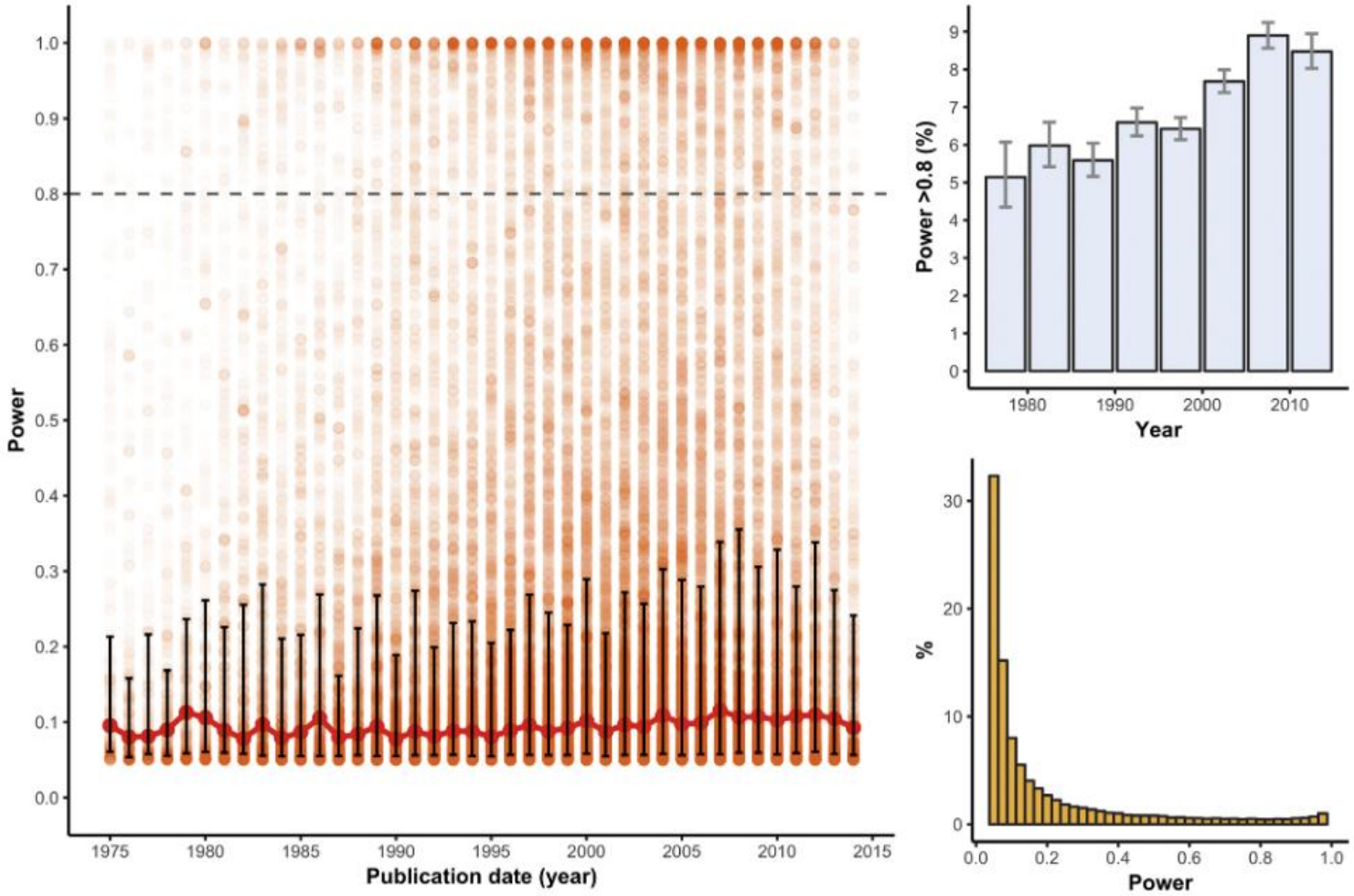


Fig. 1. Statistical power of clinical trials between 1975 and 2014 (left). Individual comparisons are shown as semitransparent dots. Median power is shown in red with interquartile range as error bars. The percentage of adequately powered trial comparisons (i.e., ≥80% power) is increasing over time (top right). The biphasic power distribution of the trials in general is apparent (bottom right). (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

A paradox

- Most publications report positive results

AND

- Most publications have very low power and should not be able to report positive results

➔ False positives?

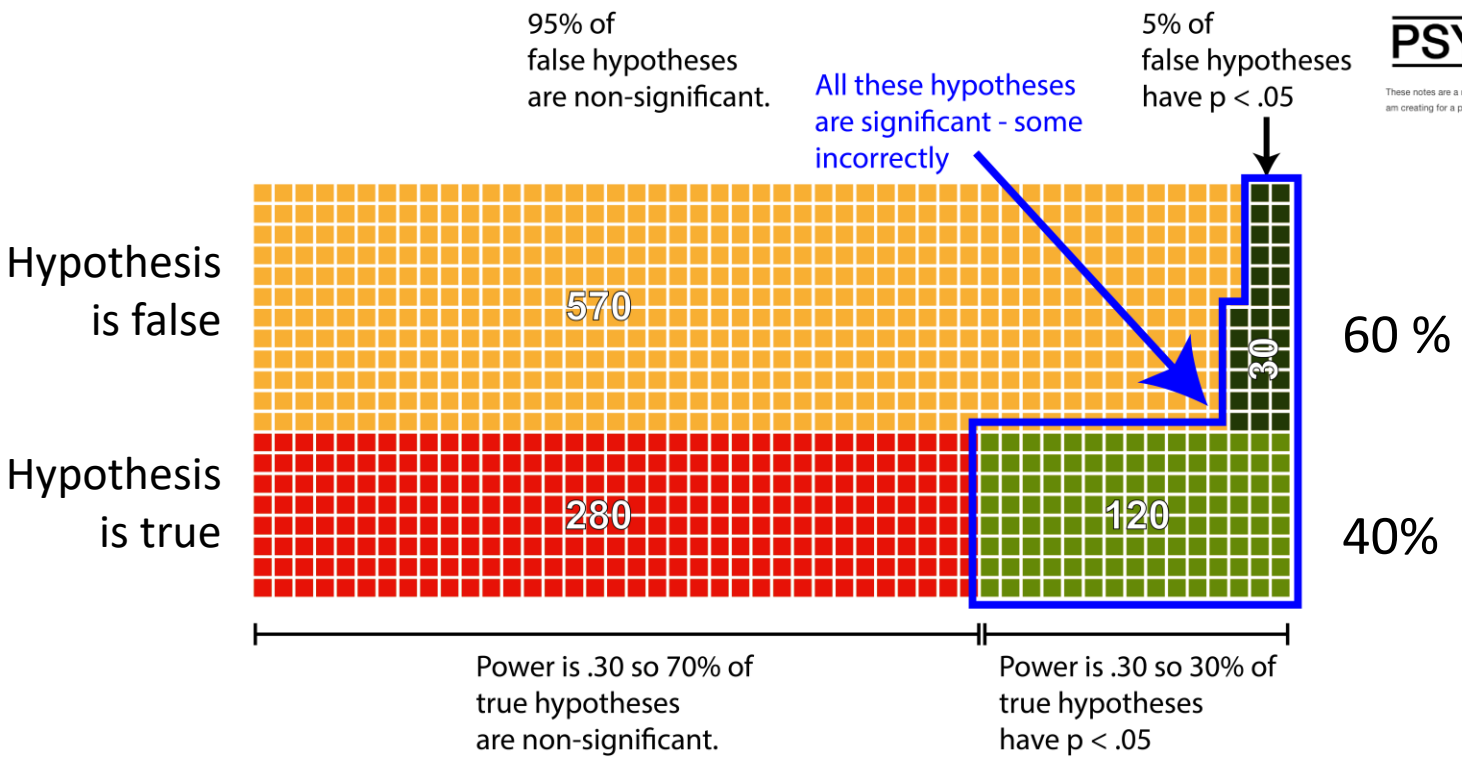
David J. Stanley

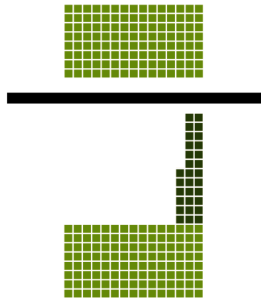
2024-10-29

Course:

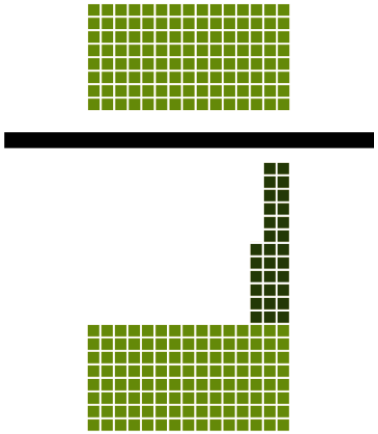
PSYC4780 Advanced Research Methods & Statistics Using R

These notes are a resource for students in my PSYC 4780 class. The notes are a subset of material that I am creating for a public domain statistics book to be released sometime in the distant future.



P(positives are true positives) = PPV =  = true positives / (all positives)

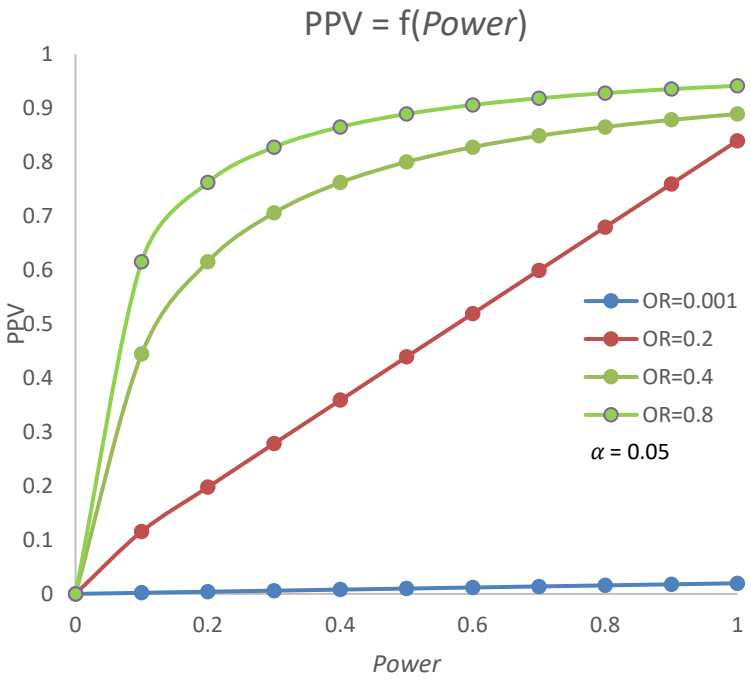
https://dstanley4.github.io/psyc4780bookdown/positive-predictive-value.html

PPV =  = true positives / (all positives) = P(positives are true positives)

$$= \frac{(1-\beta)\,P_{H_1 true}}{(1-\beta)\,P_{H_1 true} + \alpha\,P_{H_0 true}} = \frac{(1-\beta)\,\frac{P_{H_1 true}}{P_{H_0 true}}}{(1-\beta)\,\frac{P_{H_1 true}}{P_{H_0 true}} + \alpha}$$

$$= \frac{(1-\beta)(OR)}{(1-\beta)(OR)+\alpha} = \frac{\textit{Power} \times \textit{OR}}{\textit{Power} \times \textit{OR} + \alpha}$$

P(positives are true positives) = an increasing function of Power



Button KS, Ioannidis JPA, Mokrysz C, Nosek BA, Flint J, Robinson ESJ, Munafò MR (2013) Power failure: why small sample size undermines the reliability of neuroscience. *Nature Reviews Neuroscience*, **14**, 365–376. https://doi.org/10.1038/nrn3475

"A study with low statistical power has a reduced chance of detecting a true effect, but it is less well appreciated that low power also reduces the likelihood that a statistically significant result reflects a true effect."
https://www.projectimplicit.net/nosek/papers/BIMNFRM2013.pdf

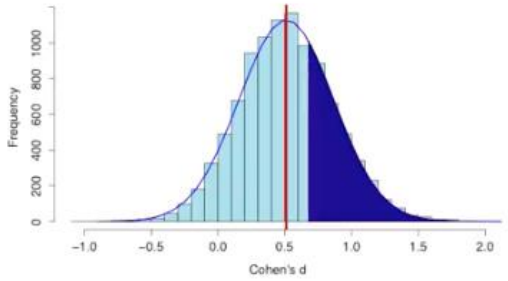
Ioannidis JPA (2005) Why Most Published Research Findings Are False. *PLOS Medicine*, **2**, e124. https://doi.org/10.1371/journal.pmed.0020124

"Simulations show that for most study designs and settings, it is more likely for a research claim to be false than true."

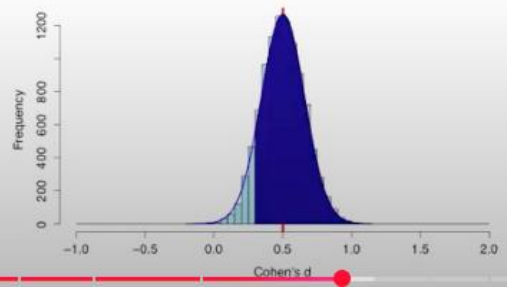
Consequences of a low power?

Increase the probability that positive results are false

Significant effect sizes

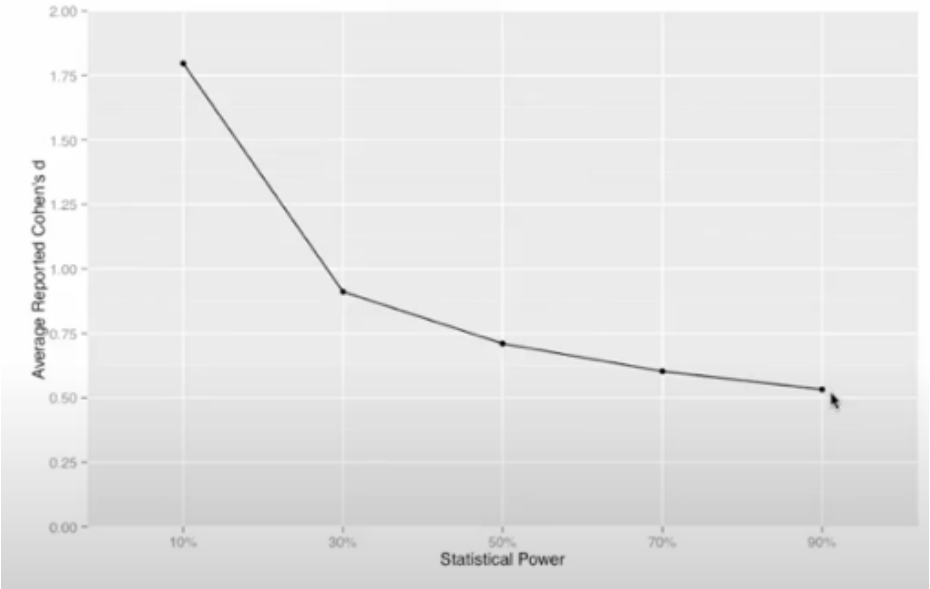


30% Power



90% Power

True Cohen's d = .50



https://youtu.be/7daQRvRO-NE?si=dLKWri9wJjeup_kl

Consequences of a low power?

Over estimation of Effects

So what?

- Most publications show positive results

- Most publications have low power
  - should hardly detect true positive results
  - make large proportion of false positives
  - overestimate true effects

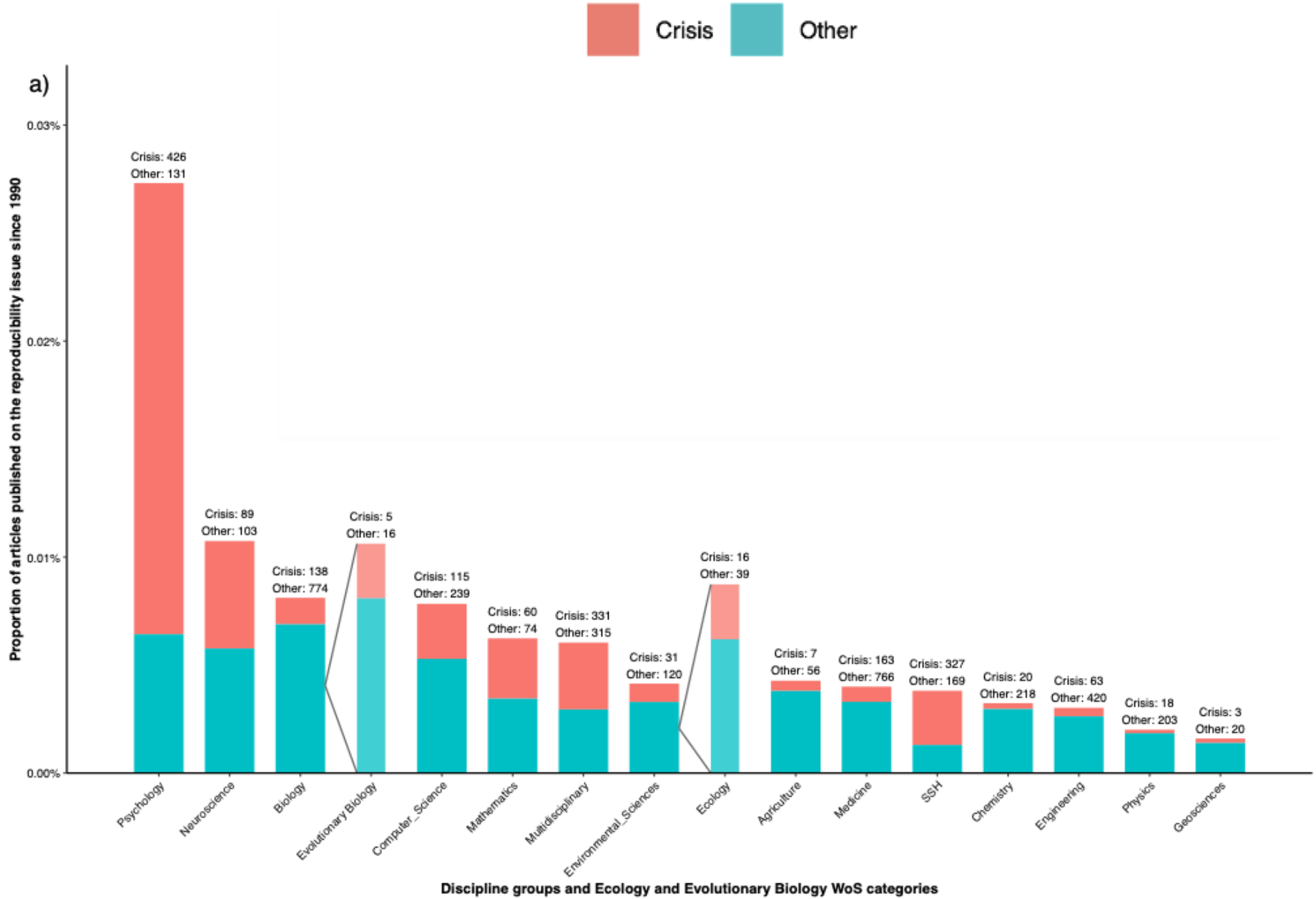
➔ Many results in the literature are probably false (?)

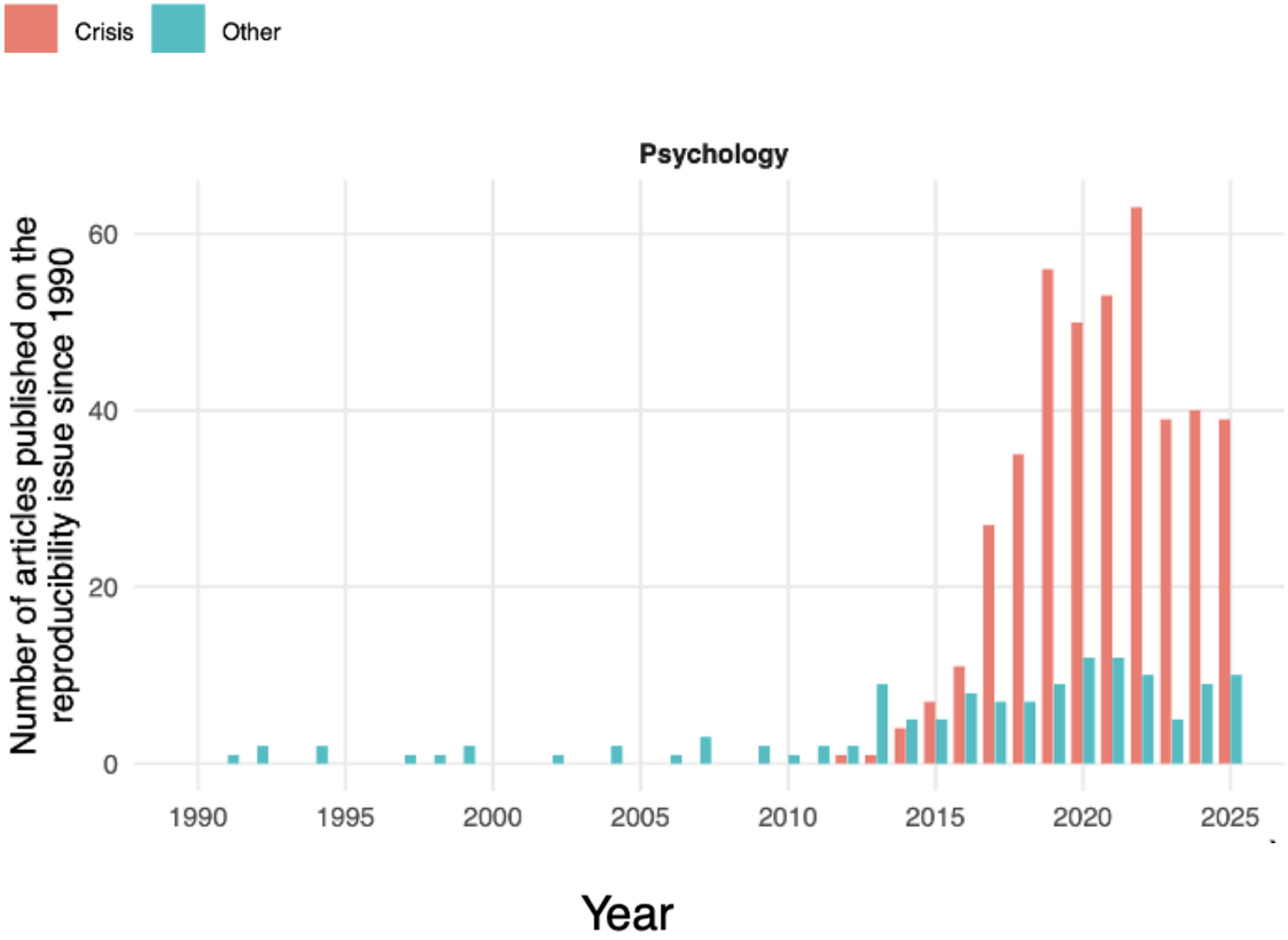
Reproducibility crisis

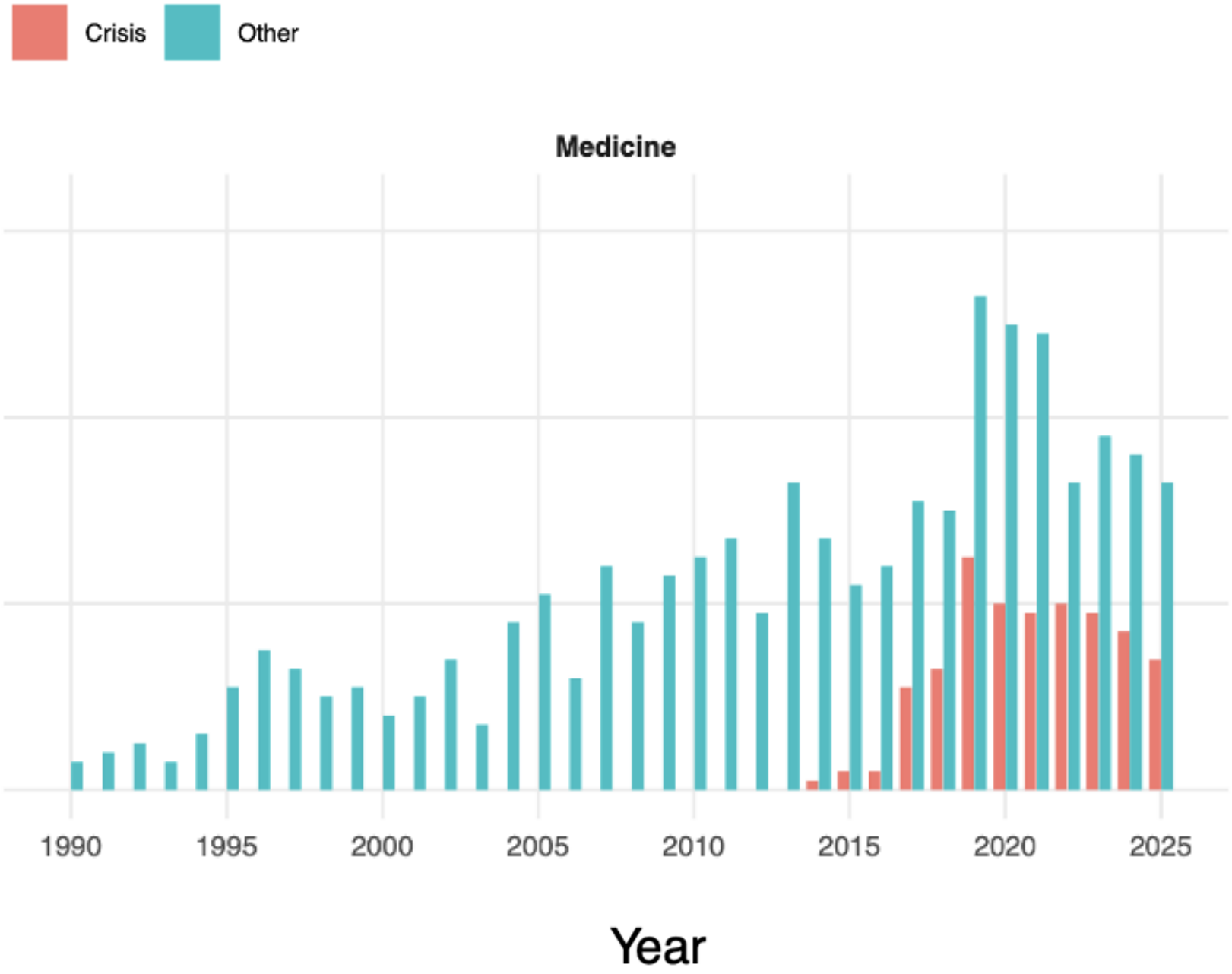
Search in the Web of Science (WoS) using the following keywords:

TOPIC = "reproducibility **crisis**" OR "reproduction **crisis**" OR "replication **crisis**" OR "replicability **crisis**" OR "reproducibility issue" OR "reproduction issue" OR "replication issue" OR "replicability issue" OR "reproducibility problem" OR "reproduction problem" OR "replication problem" OR "replicability problem" OR "reproducibility issue" OR "reproduction issue" OR "replication issue" OR "replicability issue" OR "lack of reproducibility" OR "lack of reproduction" OR "lack of replication" OR "lack of replicability" OR "reproducibility concern" OR "reproduction concern" OR "replication concern" OR "replicability concern" OR "reproducibility failure" OR "reproduction failure" OR "replication failure" OR "replicability failure" OR "reproducibility challenge" OR "reproduction challenge" OR "replication challenge" OR "replicability challenge" OR "irreproducible results" OR "non-reproducible results" OR "non-replicable results".

5,399 articles were published on the reproducibility issue
1,791 using the term "crisis"


Crisis
Other
a)
0.03%
0.02%
0.01%
0.00%
Proportion of articles published on the reproducibility issue since 1990
Crisis: 426
Other: 131
Crisis: 89
Other: 103
Crisis: 138
Other: 774
Crisis: 5
Other: 16
Crisis: 115
Other: 239
Crisis: 60
Other: 74
Crisis: 331
Other: 315
Crisis: 31
Other: 120
Crisis: 16
Other: 39
Crisis: 7
Other: 56
Crisis: 163
Other: 766
Crisis: 327
Other: 169
Crisis: 20
Other: 218
Crisis: 63
Other: 420
Crisis: 18
Other: 203
Crisis: 3
Other: 20
Psychology
Neuroscience
Biology
Evolutionary Biology
Computer_Science
Mathematics
Multidisciplinary
Environmental_Sciences
Ecology
Agriculture
Medicine
SSH
Chemistry
Engineering
Physics
Geosciences
Discipline groups and Ecology and Evolutionary Biology WoS categories


Crisis
Other
Psychology
Number of articles published on the reproducibility issue since 1990
0
20
40
60
1990
1995
2000
2005
2010
2015
2020
2025
Year


Crisis
Other
Medicine
1990
1995
2000
2005
2010
2015
2020
2025
Year

Reproducibility crisis

Systematic reproductions of experiments in life sciences
→ < 50% success

On 67 cancer studies replications, only ~**20–25% were coherent** with original results.

Prinz, F., Schlange, T. & Asadullah, K. Believe it or not: how much can we rely on published data on potential drug targets?. Nat Rev Drug Discov 10, 712 (2011). https://doi.org/10.1038/nrd3439-c1

Out of 53 articles, **results were confirmed for only 6% of cases**

Begley, C., Ellis, L. Raise standards for preclinical cancer research. Nature 483, 531–533 (2012). https://doi.org/10.1038/483531a

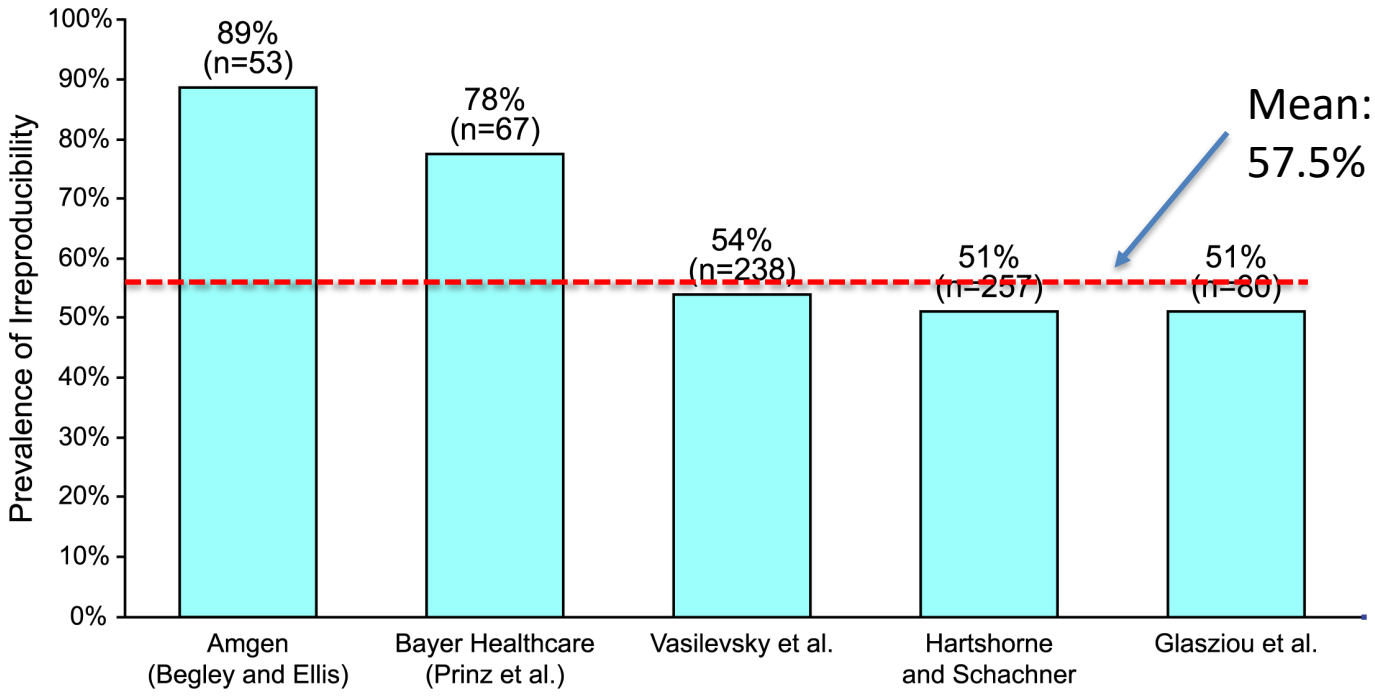
Only **50 experiments out of 193** were reproducible.

The Reproducibility project cancer Biology [eLife/Cos] https://www.cos.io/rpcb

Data analysis from **18 articles** published in *Nature Genetics* in 2005–2006 [...] reproduced the results for **only 2** of them

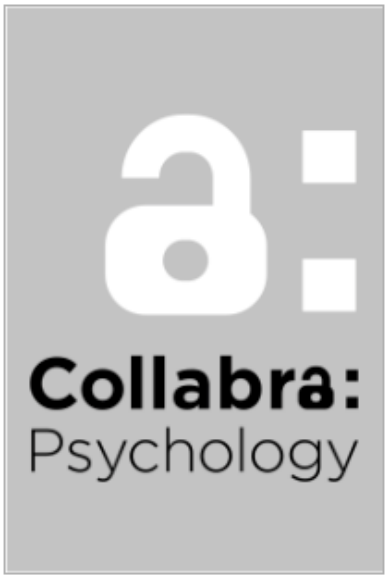
Ioannidis, J., Allison, D., Ball, C. et al. Repeatability of published microarray gene expression analyses. Nat Genet 41, 149–155 (2009). https://doi.org/10.1038/ng.295

How reproducible are published results? Pre-Clinical trials



Freedman LP, Cockburn IM, Simcoe TS (2015) The Economics of Reproducibility in Preclinical Research. PLoS Biol 13(6): e1002165. https://doi.org/10.1371/journal.pbio.1002165

Begley CG, Ellis LM (2012) Drug development: raise standards for preclinical cancer research. Nature 483: 531–533
Glasziou P, Meats E, Heneghan C, Shepperd S (2008) What is missing from descriptions of treatment in trials and reviews? BMJ 336: 1472–1474
Hartshorne JK, Schachner A (2012) Tracking replicability as a method of post-publication open evaluation. Front Comput Neurosci 6: 1–13
Prinz F, Schlange T, Asadullah K (2011) Believe it or not: how much can we rely on published data on potential drug targets? Nat Rev Drug Discov 10: 712–712
Vasilevsky NA, Brush MH, Paddock H, Ponting L, Tripathy SJ, et al. (2013) On the reproducibility of science: unique identification of research resources in the biomedical literature. PeerJ 1: e148



< Previous Article Next Article >

REVIEW ARTICLE | MARCH 22 2022

Sample Size Justification FREE

Collections: Section: Methodology and Research Practice

Daniël Lakens

Editor: Don van Ravenzwaaij

1. d.lakens@tue.nl

Collabra: Psychology (2022) 8 (1): 33267.
https://doi.org/10.1525/collabra.33267 **Article history**

Split-Screen Views PDF Share Tools

An important step when designing an empirical study is to justify the sample size that will be collected. The key aim of a sample size justification for such studies is to explain how the collected data is expected to provide valuable information given the inferential goals of the researcher. In this overview article six approaches are discussed to justify the sample size in a quantitative empirical study: 1) collecting data

Lakens, D. (2022). Sample size justification. *Collabra: Psychology, 8*(1), Article 33267.
https://doi.org/10.1525/collabra.33267

What about authors' guidelines in journals?

Evolution: « »
Ecology Letters: « »
Journal of Evolutionary Biology: « »
Ecology: « *Power analyses (determination of type II error rates)* ***occasionally can be useful*** »

Phytopathology: « »
New Phyotologist: « »

PLoS Biol, *PLoS One*, etc. : « ***If appropriate****, provide sample sizes, along with a description of how they were determined. If a sample size calculation was performed, specify the inputs for power, effect size and alpha* »

Peer Community In: : « A clear justification of sample sizes for empirical studies (see Lakens, 2022; [...]), especially with respect to the following, as relevant: statistical power, Bayes factor thresholds, or estimate precision (quantitative studies) and information power or saturation thresholds (qualitative studies). »

Pilot study (Stage 1 Registered Report)

Frequency of request to justify sample sizes in Author's guidelines) in **scientific journals**

in 4 fields (n = 20 journals randomly screened in each field):

Medicine: 65%
Psychology: 40%
Ecology: 10%
Agronomy:10%

=> Suggest a high variation with the lowest levels of requirements in Ecology and Agronomy

Full study (Stage 2 Registered Report) in progress

So what?

- Let’s try to (scientifically) justify our sample sizes
- Let’s choose the right methods (Hypothesis testing, exploration, etc.)
- Hypothesis testing: power anlaysis (based on litterature or preliminary data : effect size, variance)
- Education on **scientific methods** – not only on scientific knowledge

One way to improve the situation would be…

Research quality to be determined solely on *scientific validity* (question and method), and never on the **results** that studies produce

=> Registered Reports (RR)

Registered Reports 1.0

CORTEX 49 (2013) 609–610

Available online at www.sciencedirect.com

SciVerse ScienceDirect

Journal homepage: www.elsevier.com/locate/cortex

Editorial

Registered Reports: A new publishing initiative at *Cortex*

Christopher D. Chambers

Cardiff University Brain Research Imaging Centre (CUBRIC), School of Psychology, Cardiff University, United Kingdom

Central aspects of the Registered Reports model:

- Researchers decide hypotheses, study procedures, and main analyses ***before*** data collection
- Part of the peer review process takes place ***before*** studies are conducted
- Passing this stage of review virtually guarantees publication

RR: How does it work?

1

Stage 1 RR: -Contex
-Questions/Ho
-Methodology

RR: How does it work?

1

Stage 1 RR: -Context
-Questions/Ho
-Methodology

Stage 1 Peer Review

➔ In principle acceptance (IPA)

RR: How does it work?

1

Stage 1 RR: -Context
-Questions/Ho
-Methodology

2

Experiments
Field work
Analyses

Stage 1 Peer Review

➔ In principle acceptance (IPA)

RR: How does it work?

1

Stage 1 RR: -Context
-Questions/Ho
-Methodology

Stage 1 Peer Review

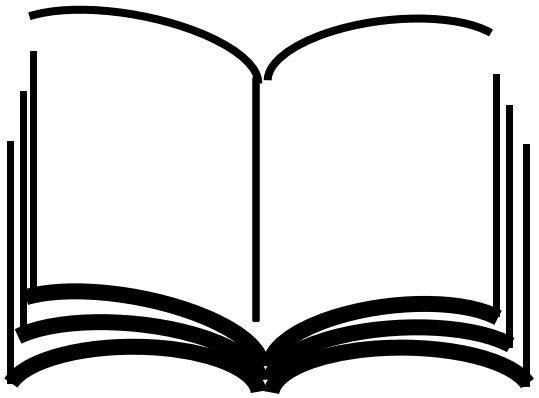
➔ In principle acceptance (IPA)

2

Experiments
Field work
Analyses

3

Stage 2 RR: -Context
-Questions/Ho
-Methodology
-Results
-Discussion



RR: How does it work?

1

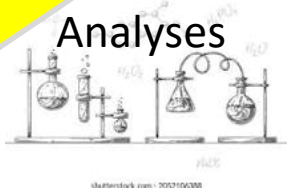
Stage 1 RR: -Context
-Questions/Ho
-Methodology

Stage 1 Peer Review

➔ In principle acceptance (IPA)

2

Experiments
Field work
Analyses

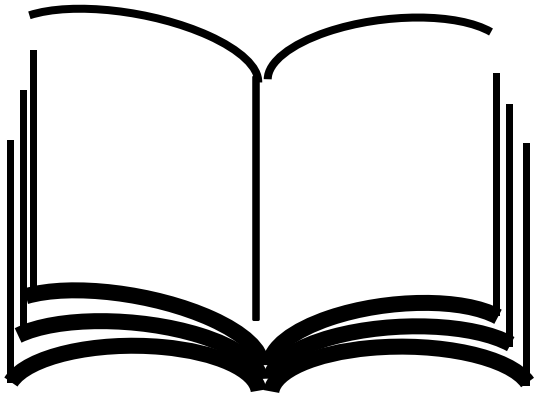


3

Stage 2 RR: -Context
-Questions/Ho
-Methodology
-Results
-Discussion

Stage 2 Peer Review

➔ Publication of the final article(s)



None of these things matter

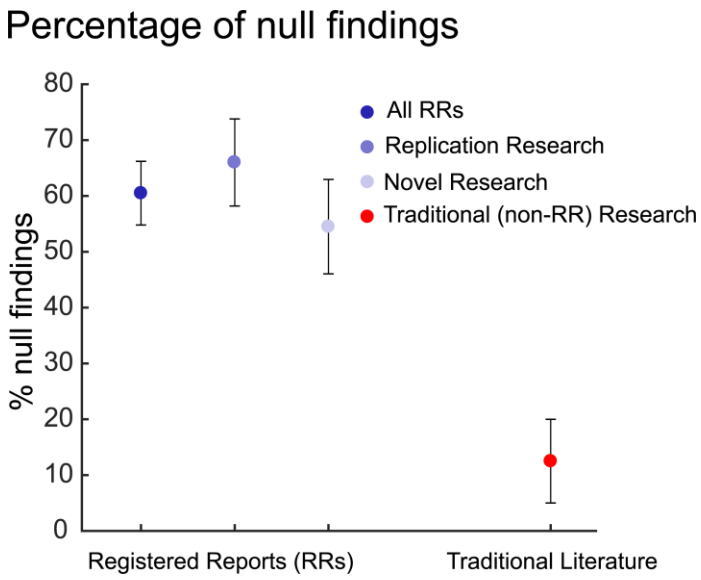
Are Registered Reports working as intended?

NEWS · 24 OCTOBER 2018

First analysis of 'pre-registered' studies shows sharp rise in null findings

Logging hypotheses and protocols before performing research seems to work as intended: to reduce publication bias for positive results.

Matthew Warren



Hypotheses are ~5 times more likely to be **unsupported** in Registered Reports compared with regular articles

Allen C, Mehler DMA (2019) Open science challenges, benefits and tips in early career and beyond. PLOS Biol 17(5): e3000246. https://doi.org/10.1371/journal.pbio.3000246

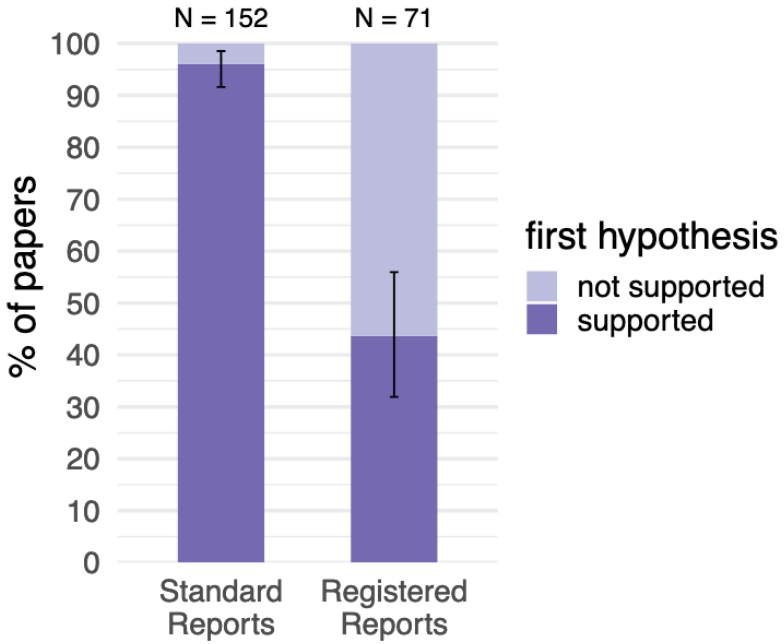


Figure 2. Positive result rates for standard reports and Registered Reports. Error bars indicate 95% confidence intervals around the observed positive result rate.

Same observation in RRs within psychology specifically

Scheel, Schijen & Lakens (2021)
https://journals.sagepub.com/doi/full/10.1177/25152459211007467

Ten years later…

Registered Reports are now mainstream

- **> 350** journals > 1 000 RR published

Generalist journals
Nature, Nature communications, Scientific Report, Plos Biol, Plos One, PNAS nexus, Royal Society Open Science, BMC Biol, Peer Community Journal, Open Research Europe, Peer J

Biology
Conservation Biology, Ecology and Evolution, Ethology, PeerJ Life and Environment, Ecological Solutions and Evidence, Frontiers in Genetics, Frontiers in Plant Science, Journal of Plant Nutrition and Soil Science, Plant Direct, Human Population Genetics and Genomics

Neuroscience
Addiction Research & Theory, Advances in Cognitive Psychology, Advances in Methods and Practices in Psychological Science, Brain and Neuroscience Advances, Collabra: Psychology, Communications in Kinesiology, Cortex, Experimental Psychology, Imaging Neuroscience, International Review of Social Psychology, Journal for Reproducibility in Neuroscience, Journal of Cognition, Meta-Psychology, NeuroImage: Reports, Personality Science, Psychology of Consciousness: Theory, Research, and Practice, Social Psychological Bulletin, Studia Psychologica, Swiss Psychology Open,

etc. linguistics, chemistry, physics, materials, education, medicine

Peer Community In Registered Reports

Free and transparent pre- and post-study recommendations across research fields





Web: https://rr.peercommunityin.org/
Email: contact@rr.peercommunityin.org
https://bsky.app/profile/pci-regreports.bsky.social
https://spore.social/@pcirr

Founders: Corina Logan, Emily Sena, Zoltan Dienes, Chris Chambers, Ben Pujol

- Free for authors and readers – Non commercial
- Evaluation of RR in all disciplines = pipeline for RR
- Recommendation and Evaluations are published (if accepted)
- Publication in *Peer Community Journal* or other journals including 40 PCI RR-friendly journals without additional reviews

Peer Community In Registered Reports

Free and transparent pre- and post-study recommendations across research fields

t pre- and post-study across research fields

Since 2021

186 recommenders (associate editors)

613 submissions of Stage 1 RR

226 submissions of Stage 2 RR

2,256 reviews

301 recommendations of Stage 1 RR

145 recommendations of Stage 2 RR

Psychology, neuroscience, economics, ecology, public health, law

Quantitative and qualitative studies

All with open review

For more info, email contact@rr.peercommunityin.org or chambersc1@cardiff.ac.uk

Slides: https://osf.io/ms9pr

Thanks!

RR and Serendipity

(unplanned fortunate discoveries)

- Unplanned fortunate discoveries occur when there is a plan

➔ Plan does not prevent unplanned discoveries, on the contrary

- Stage 2 RR can include unplanned discoveries, **presented as such**

« We were trying to test hypothesis H or explore situation S when we suddenly discovered the amazing unplanned results R »

RR and Serendipity

(unplanned fortunate discoveries)

- Unplanned fortunate discoveries occur when there is a plan

➔ Plan does not prevent unplanned discoveries, on the contrary

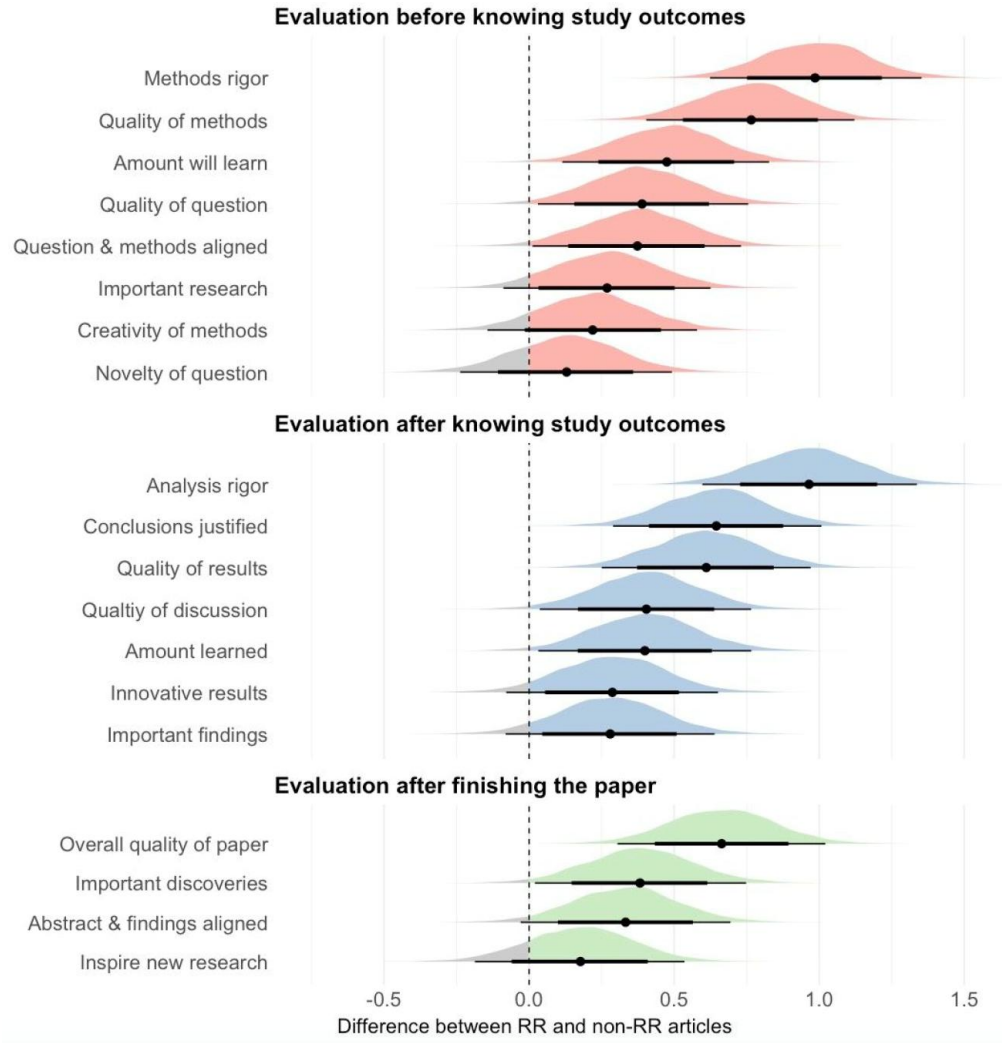
- Stage 2 RR can include unplanned discoveries, **presented as such**

« We were trying to test hypothesis H or explore situation S when we suddenly discovered the amazing unplanned results R »

When using RR, you must present unplanned discoveries as « unplanned discoveries », not as « planned results »

Only difference between using RR and not using RR

Are Registered Reports working as intended?



ARTICLES
https://doi.org/10.1038/s41562-021-01142-4
nature human behaviour
Check for updates

Initial evidence of research quality of registered reports compared with the standard publishing model

Courtney K. Soderberg[1,6], Timothy M. Errington[1,6], Sarah R. Schiavone[2], Julia Bottesini[2], Felix Singleton Thorn[3], Simine Vazire[2,3], Kevin M. Esterling[4] and Brian A. Nosek[1,5]✉

In registered reports (RRs), initial peer review and in-principle acceptance occur before knowing the research outcomes. This combats publication bias and distinguishes planned from unplanned research. How RRs could improve the credibility of research findings is straightforward, but there is little empirical evidence. Also, there could be unintended costs such as reducing novelty. Here, 353 researchers peer reviewed a pair of papers from 29 published RRs from psychology and neuroscience and 57 non-RR comparison papers. RRs numerically outperformed comparison papers on all 19 criteria (mean difference 0.46, scale range −4 to +4) with effects ranging from RRs being statistically indistinguishable from comparison papers in novelty (0.13, 95% credible interval [−0.24, 0.49]) and creativity (0.22, [−0.14, 0.58]) to sizeable improvements in rigour of methodology (0.99, [0.62, 1.35]) and analysis (0.97, [0.60, 1.34]) and overall paper quality (0.66, [0.30, 1.02]). RRs could improve research quality while reducing publication bias and ultimately improve the credibility of the published literature.

Soderberg, C. K., Errington, T. M., Schiavone, S. R., Bottesini, J. G., Singleton Thorn, F., Vazire, S., … Nosek, B. A. (2021). Initial evidence of research quality of registered reports compared with the standard publishing model. *Nature Human Behaviour* https://doi.org/10.1038/s41562-021-01142-4

Well cited – on average, cited same or slightly higher than regular articles

See Hummer, L. T., Singleton Thorn, F., Nosek, B. A. & Errington, T. M. Preprint: https://doi.org/10.31219/osf.io/5y8w7

Example: post doc or PhD student planning to do a series of independent RRs

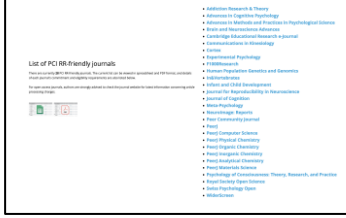
1. Design RRs and **complete Stage 1 Snapshot**

2. **Post Snapshot on the OSF**, either publicly or under private embargo

3. **Submit the snapshot URL** to PCI RR via the "Scheduled Review" track

4. **Select future date for review** (e.g. 8 weeks ahead), and once passed the recommender triage process, set to work writing a full "programmatic RR"

5. While designing & writing the Stage 1 RR, **consult the list of PCI RR-friendly journals** to ensure that you meet any additional requirements for whatever target journals you have in mind (e.g. concerning evidence strength, bias control, etc)



6. **Submit your full Stage 1 manuscript by the due date**. Because review is planned in advance, reviews & an interim recommendation can be expected in ~2 weeks

7. If, likely following revision, you gain in-principle acceptance (IPA), **PCI RR will tell you which journals are eligible outlets** and will auto-endorse the IPA decision. PCI RR makes this decision.

8. With IPA in hand, **you now have an approved programme of multiple Stage 2 RRs accepted in advance** which you can eventually choose to publish in any eligible PCI RR-friendly journal (or you can submit anywhere else as you see fit). Each Stage 2 RR can go in a different journal.

9. **Do research and publish each Stage 2 output as you go** without further peer review, in journal of your choice