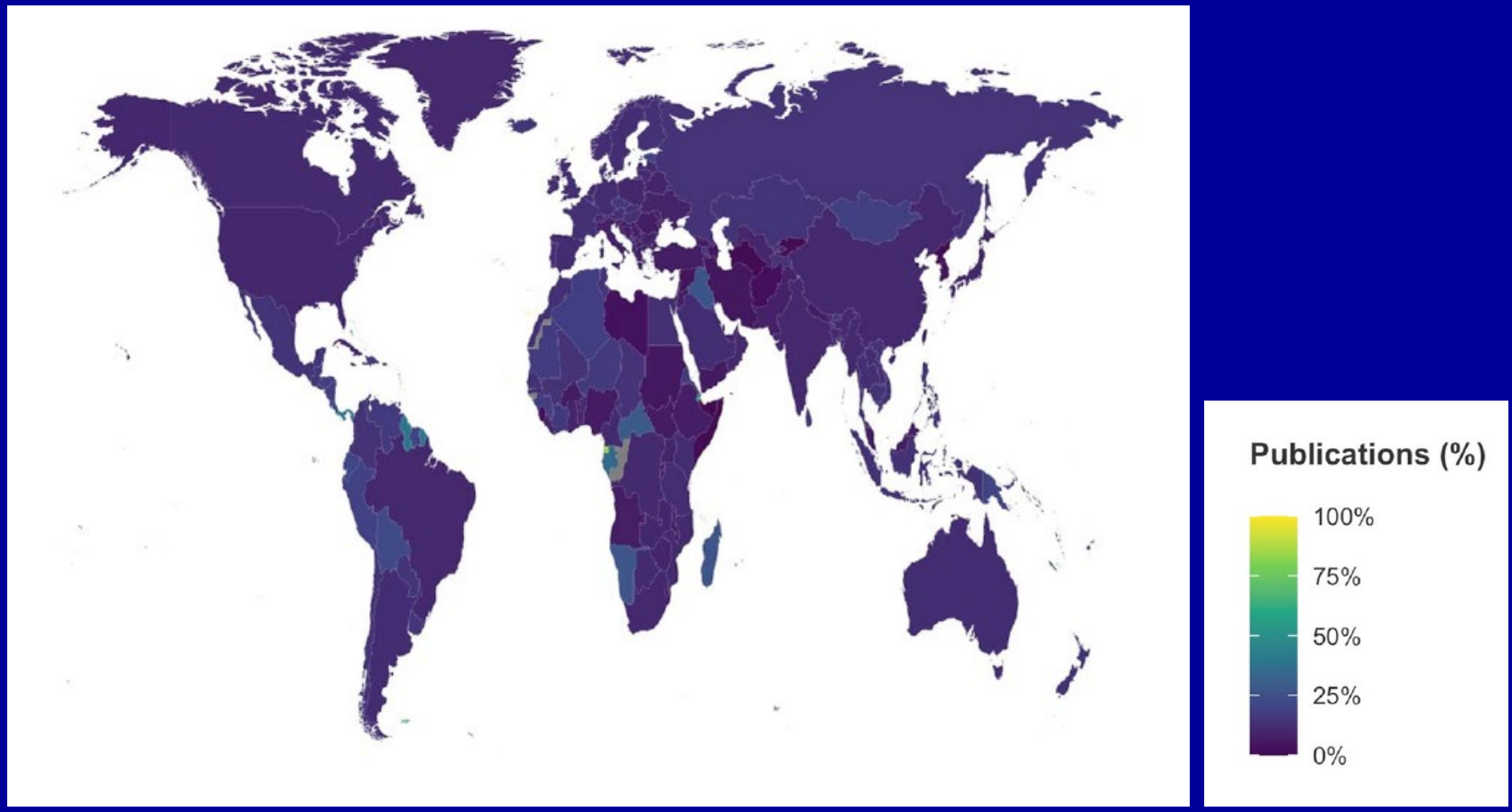
Strategic Council for Research Excellence, Integrity and Trust

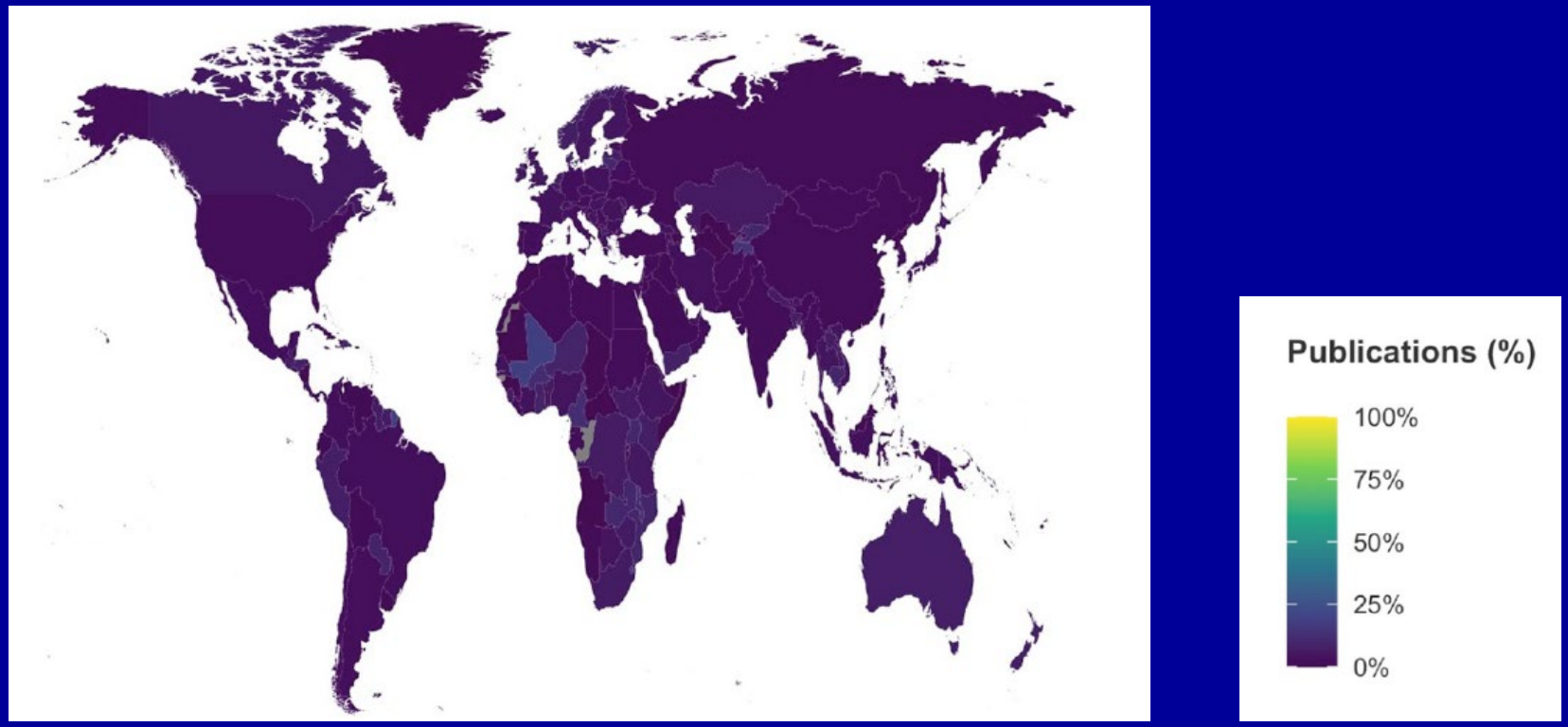
John P.A. Ioannidis, MD, DSc

Professor of Medicine, of Epidemiology and Population Health, and (by courtesy) of Biomedical Data Science, and of Statistics
Stanford University
Co-Director, Meta-Research Innovation Center at Stanford (METRICS)

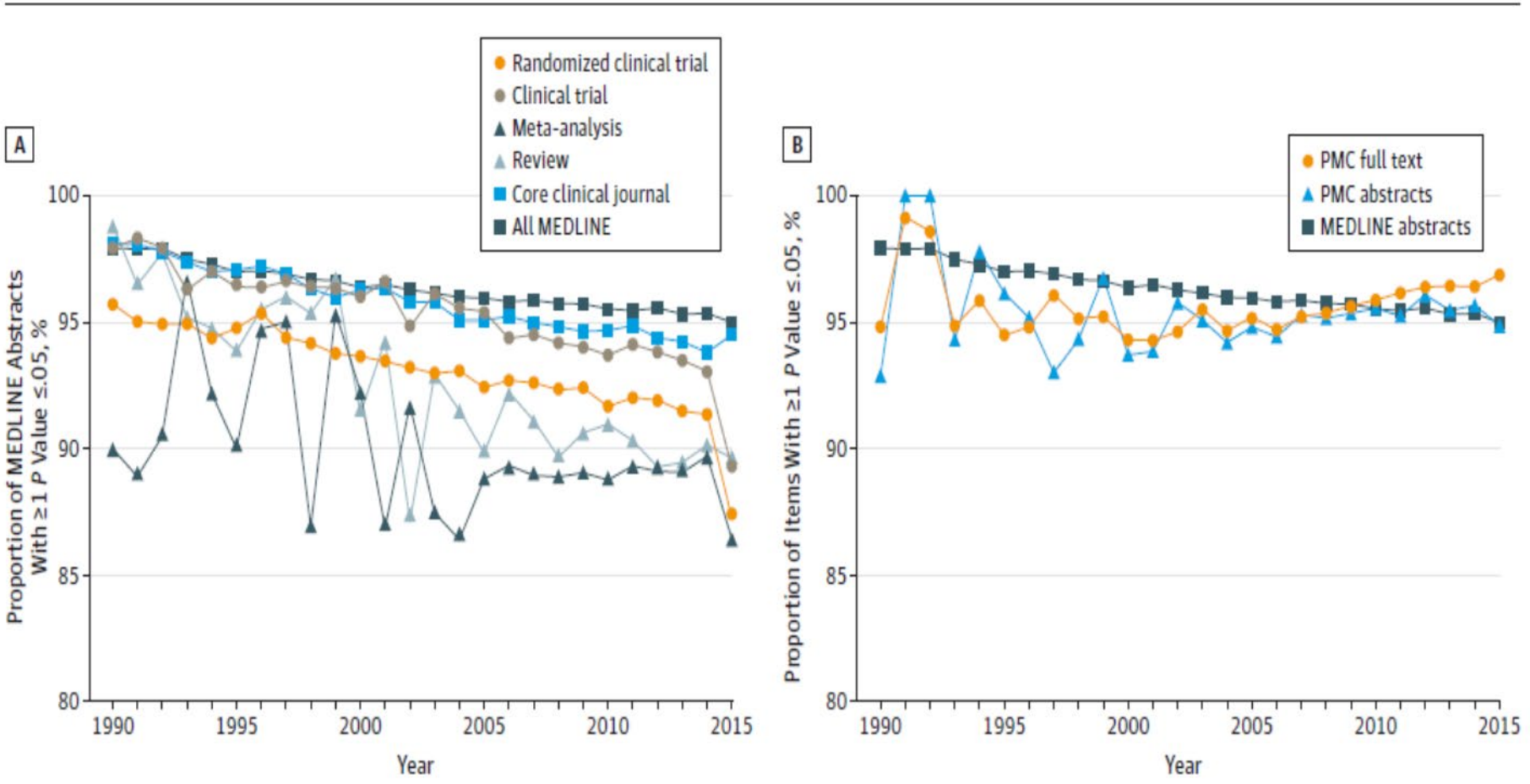
Publications sharing data, PubMed Central 1990-2020



Publications having a registered protocol, PubMed Central 1990-2020



Algorithmic text mining of the entire biomedical literatiure Prevalence of P-values and P<0.05: 96% of the biomedical literature claims significant results



Chavalarias, Wallach, Li, Ioannidis, JAMA 2016

Most studies are very small

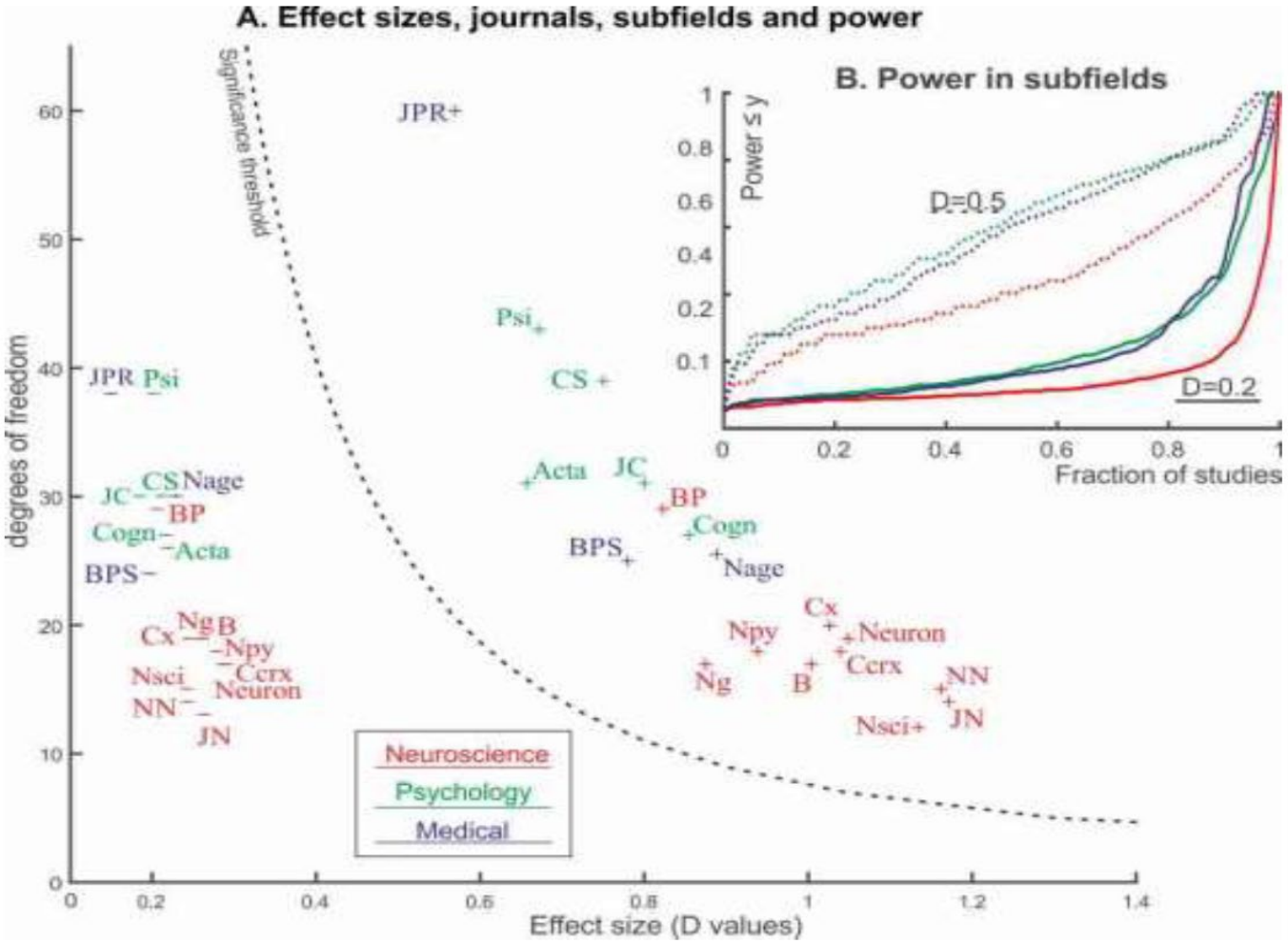


Figure 2. Power in journals and subfields. (A) Median effect sizes and degrees of

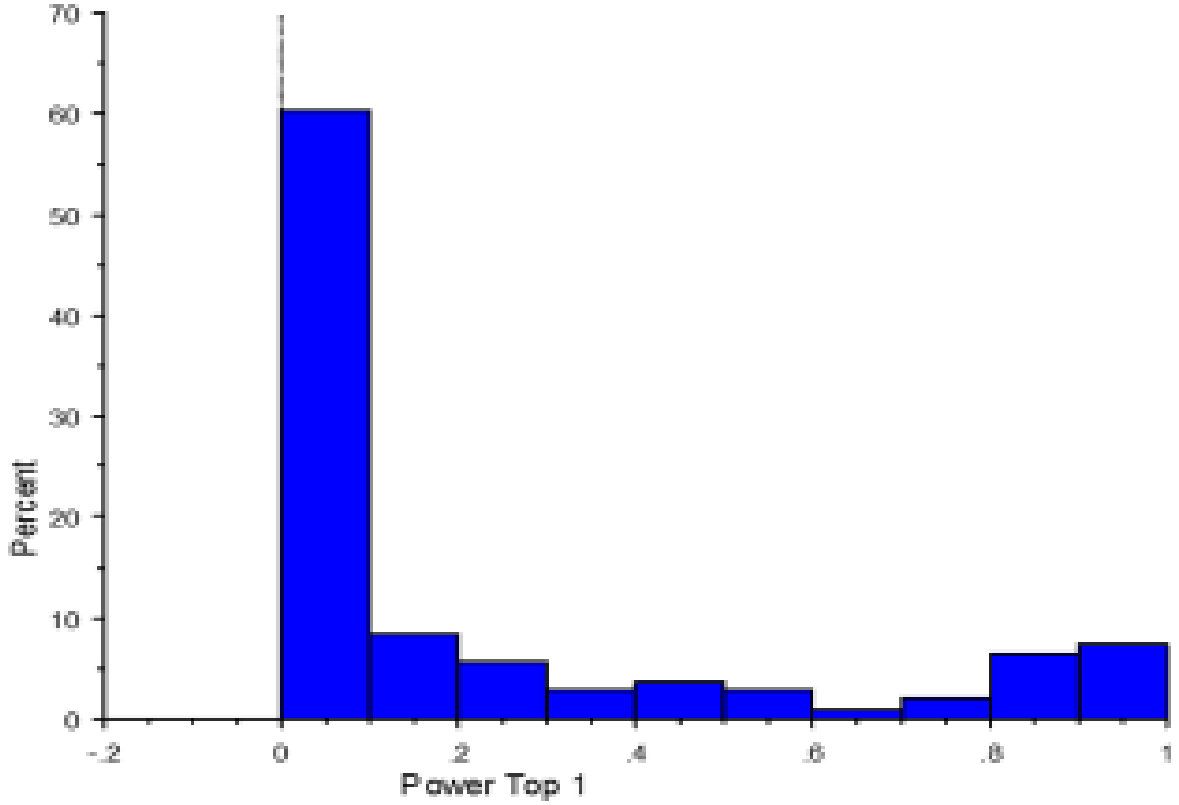
Szucs and Ioannidis, PLoS Biology 2017

Power failure: why small sample size undermines the reliability of neuroscience

Katherine S. Button[1,2], John P. A. Ioannidis[3], Claire Mokrysz[1], Brian A. Nosek[4], Jonathan Flint[5], Emma S. J. Robinson[6] and Marcus R. Munafò[1]

Abstract | 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. Here, we show that the average statistical power of studies in the neurosciences is very low. The consequences of this include overestimates of effect size and low reproducibility of results. There are also ethical dimensions to this problem, as unreliable research is inefficient and wasteful. Improving reproducibility in neuroscience is a key priority and requires attention to well-established but often ignored methodological principles.

Power in 130 economics topics (>10,000 studies with >70,000 effect estimates)



Ioannidis, Stanley, Doucouliagos, Economic Journal 2017

Big Data, Big Noise, Big Error

MEDICINE

Big data meets public health

Human well-being could benefit from large-scale data if large-scale noise is minimized

By **Muin J. Khoury**[1] *and*
John P. A. Ioannidis[2]

In 1854, as cholera swept through London, John Snow, the father of modern epidemiology, painstakingly recorded the locations of affected homes. After long, laborious work, he implicated the Broad Street water pump as the source of the outbreak, even without knowing that a Vibrio organism caused cholera. Today, Snow might have crunched Global Positioning System information and disease prevalence data, solving the problem within hours (*1*). That is the potential impact of "Big Data" on the public's health. But the promise of Big Data is also accompanied by claims that "the scientific method itself is becoming obsolete" (*2*), as next generation computers, such as IBM's Watson (*3*), sift through the digital world to provide predictive models based on massive information. Separating the true signal from the gigantic amount of noise is neither easy nor straightforward, but it is a challenge that must be tackled if information is ever to be translated into societal well being.

The term "Big Data" refers to volumes of large, complex, linkable information (*4*). Beyond genomics and other "omic" fields, Big Data includes medical, environmental, financial, geographic, and social media information. Most of this digital information was unavailable a decade ago. This swell of data will continue to grow, stoked by sources that are currently unimaginable. Big Data stands to improve health by providing insights into

From validity to utility. Big Data can improve tracking and response to infectious disease outbreaks, discovery of early warning signals of disease, and development of diagnostic tests and therapeutics.

For non-genomic associations, false alarms due to confounding variables or other biases are possible even with very large-scale studies, extensive replication, and very strong signals (*9*). Big Data's strength is in finding associations, not in showing whether these associations have meaning. Finding a signal is only the first step.

Even John Snow needed to start with a plausible hypothesis to know where to look, i.e., choose what data to examine. If all he had was massive amounts of data, he might well have ended up with a correlation as spurious as the honey bee-marijuana connection. Crucially, Snow "did the experiment." He removed the handle from the water pump and dramatically reduced the spread of cholera, thus moving from correlation to causation and effective intervention.

How can we improve the potential for Big Data to improve health and prevent disease? One priority is that a stronger epidemiological foundation is needed. Big Data analysis is currently largely based on convenient samples of people or information available on the Internet. When associations are probed between perfectly measured data (e.g., a genome sequence) and poorly measured data (e.g., administrative claims health data), research accuracy is dictated by the weakest link. Big Data are observational in nature and are fraught with many biases such as selection, confounding variables, and lack of generalizability. Big Data analysis may be embedded in epidemiologically well-characterized and representative populations. This epide-

Khoury, Ioannidis, Science 2014

Bird's eye views of bias across scientific domains

Meta-assessment of bias in science

Daniele Fanelli[a,1], Rodrigo Costas[b], and John P. A. Ioannidis[a,c,d,e]

Table 1. Summary of each bias pattern or risk factor for bias that was tested in our study, parameters used to test these hypotheses via meta-regression, predicted direction of the association of these parameters with effect size, and overall assessment of results obtained

Hypothesis type	Hypothesis tested	Specific factor tested	Variables measured to test the hypothesis	Predicted association with effect size	Result
Postulated bias patterns	Small-study effect		Study SE	+	S
	Gray literature bias		Gray literature (any type) vs. Journal article	–	S
	Decline effect		Year order in MA	–	P
	Early extremes		Year order in MA, regressed on absolute effect size	–	N
	Citation bias		Total citations to study	+	S
	US effect		Study from author in the US vs. Any other country	+	P
	Industry bias		Studies with authors affiliated with private industry vs. Not	+	P
Postulated risk factors for bias	Pressures to publish	Country policies	Cash incentive	+	N
			Career incentive	+	N
			Institutional incentive	+	N
		Author's productivity	(First/last) author's total publications, publications per year	+	N
		Author's impact	(First/last) total citations, average citations, average normalized citations, average journal impact, % top10 journals	+	N
	Mutual control		Team size	–	S
			Countries/author, average distance between addresses	+	S
	Individual risk factors	Career level	Years in activity (first/last) author	–	S
		Gender	(First/last) author's female name	–	N
		Research integrity	(First/last) author with ≥1 retraction	+	P

Symbols indicate whether the association between factor and effect size is predictive to be positive (+) or negative (–). Conclusions as to whether results indicate that the hypothesis was fully supported (S), partially supported (P), or not supported (N) are based on main analyses as well as secondary and robustness tests, as described in the main text.

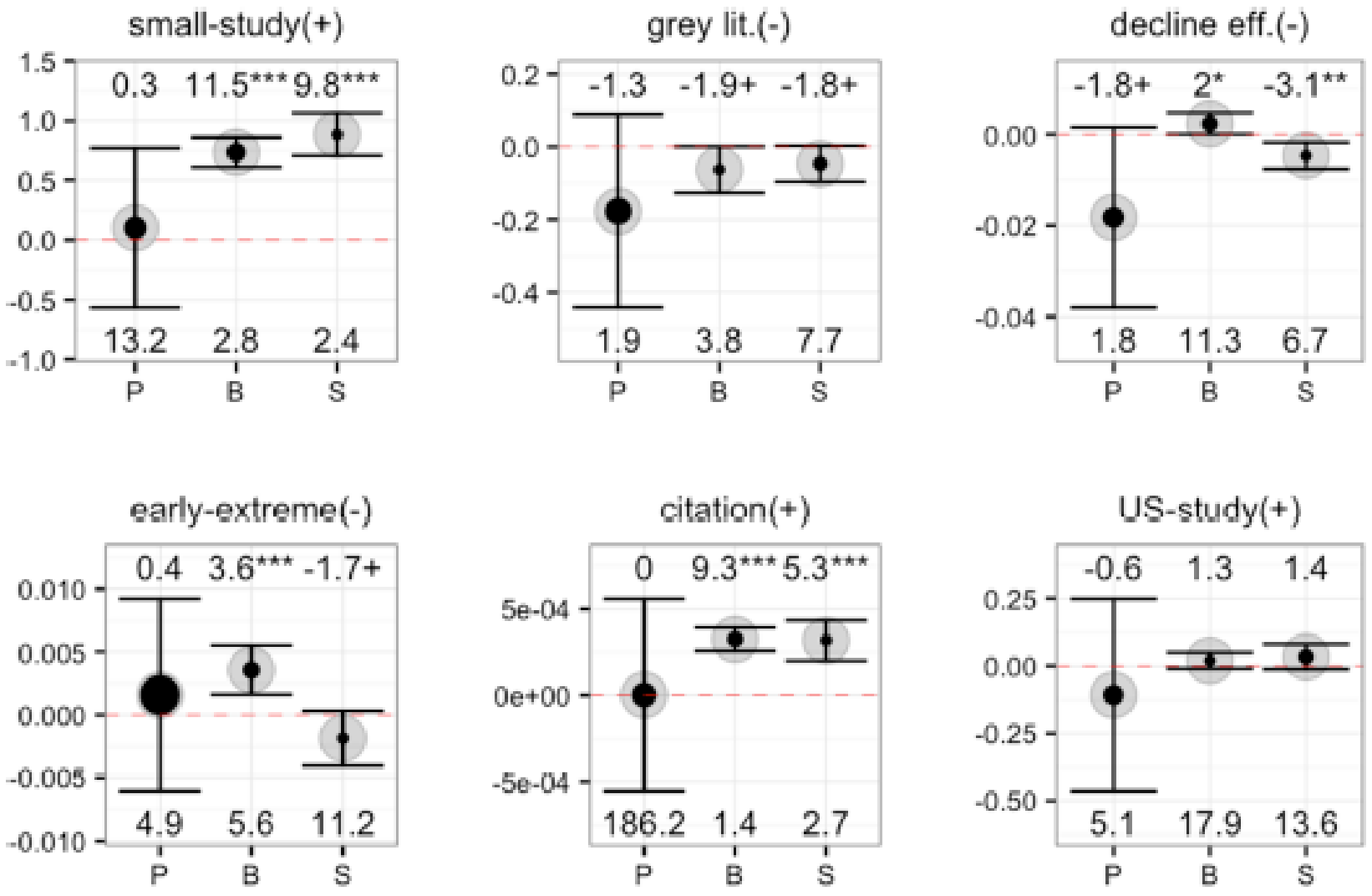
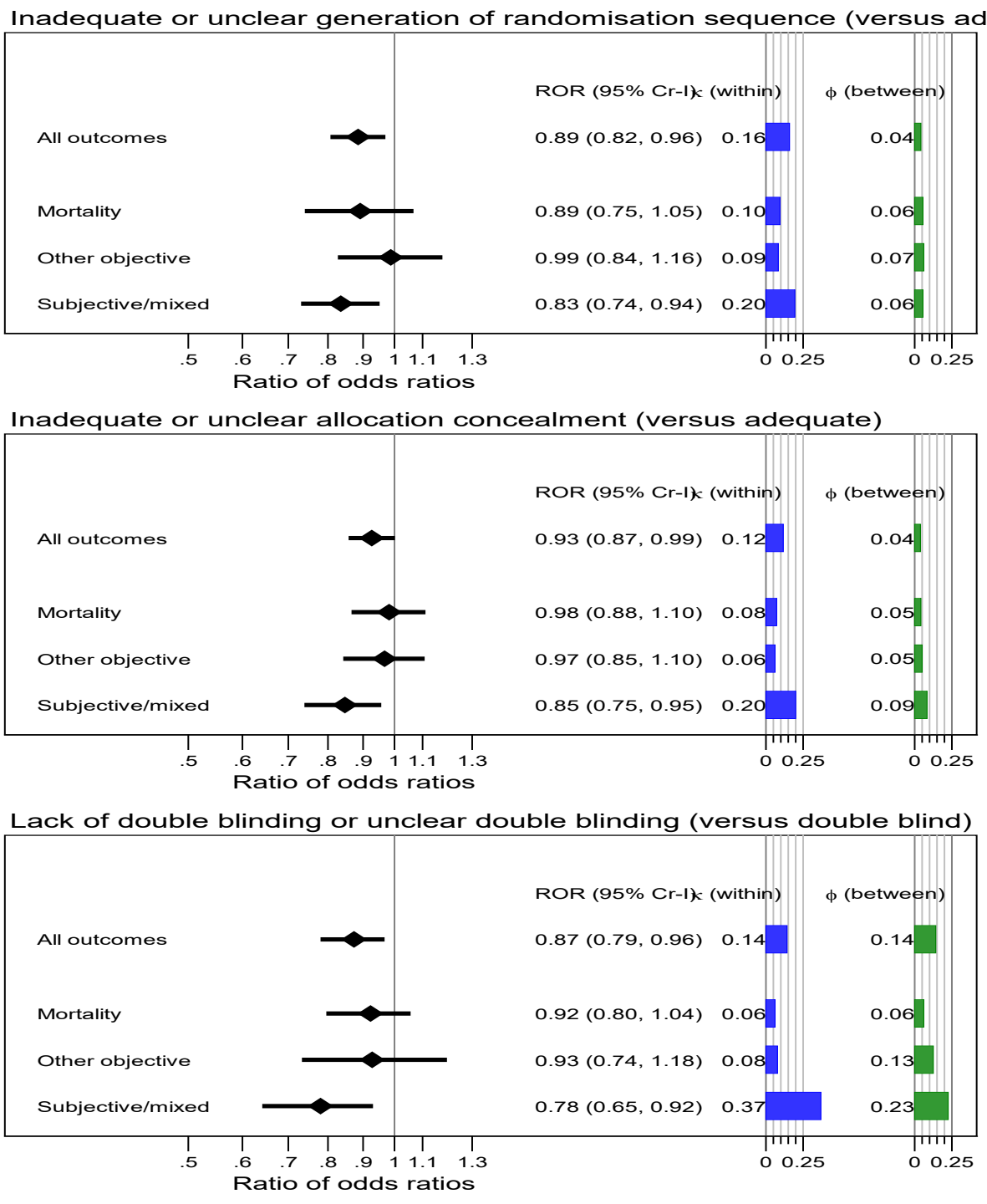


Fig. 2. Bias patterns partitioned by disciplinary domain. Each panel re-

Meta-meta-epidemiology Association of effects with reported quality



Large scale assessments of the scientific workforce

e.g. Covidization of research

720,000 authors publishing on COVID-19

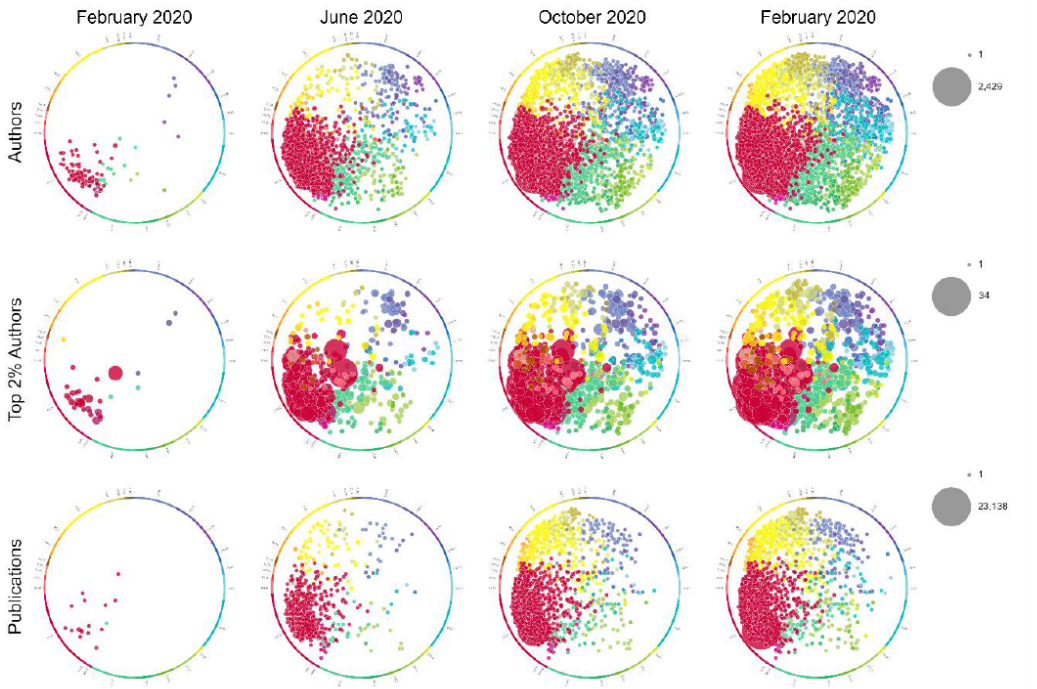


Figure 1 Topics of prominence for COVID-19 authors and publications. The columns represent the progress of the spread at 3 different measuring points: by end of February 2020, end of June 2020, end of October 2020 and end of February 2021. The first row represents the spread of authors of COVID-19 papers. The authors are assigned to their most dominant topic in their career. The data is filtered to include only topics with >=5 authors assigned. The second row shows similarly the topics of the top 2% authors by field according to a composite citations indicator. Only topics with 2 or more authors are displayed. The third row displays the spread of COVID-19 publications across topics. The minimum threshold for a topic to be displayed is set to 5 COVID-19 publications. Of note,

Enhancing data sharing: Data Arks

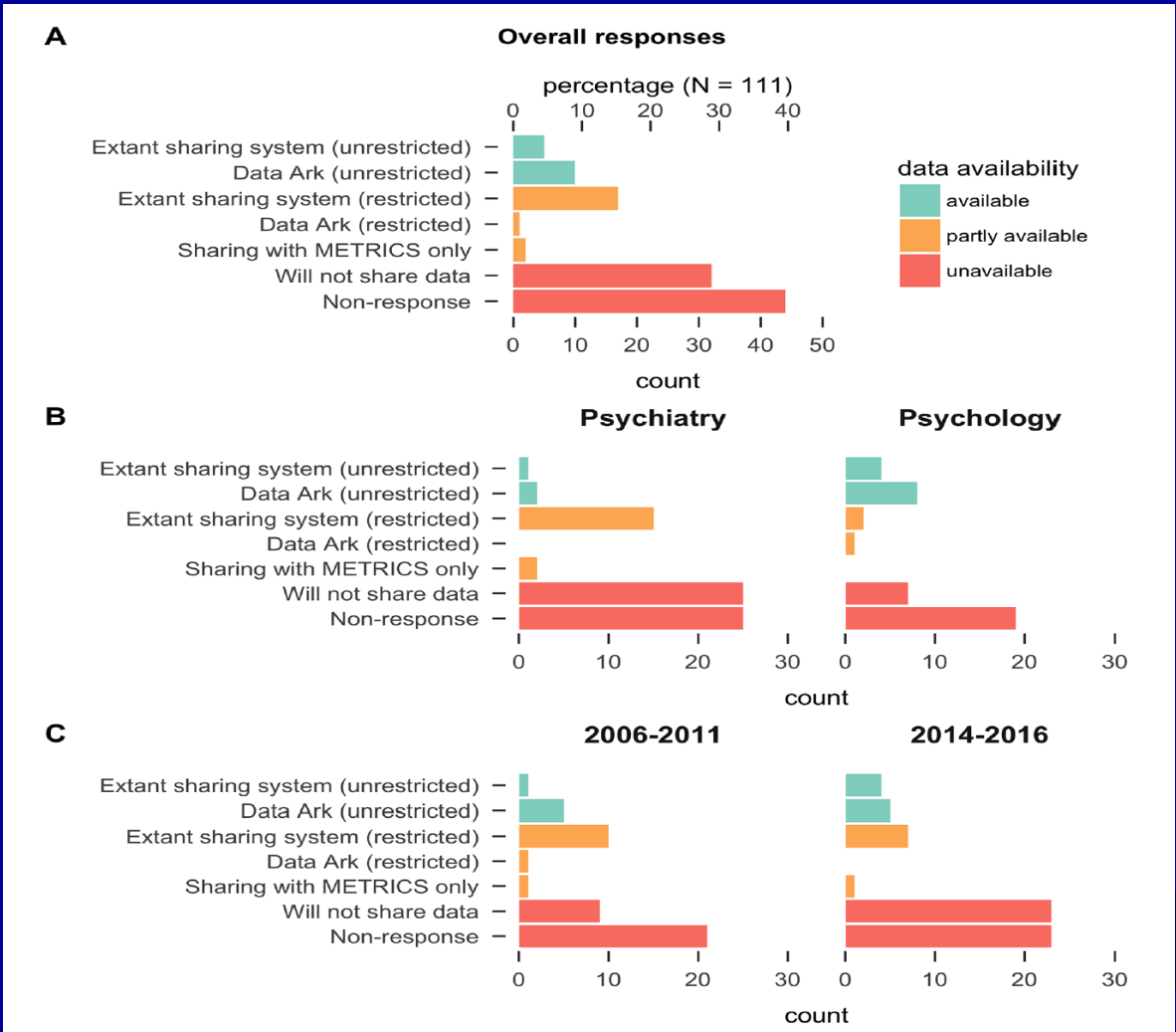


Fig 1. Responses to data request, overall (Panel A) and broken down by field (Panel B) and sample time period (Panel C).

https://doi.org/10.1371/journal.pone.0201856.g001

Hardwicke, Ioannidis 2018

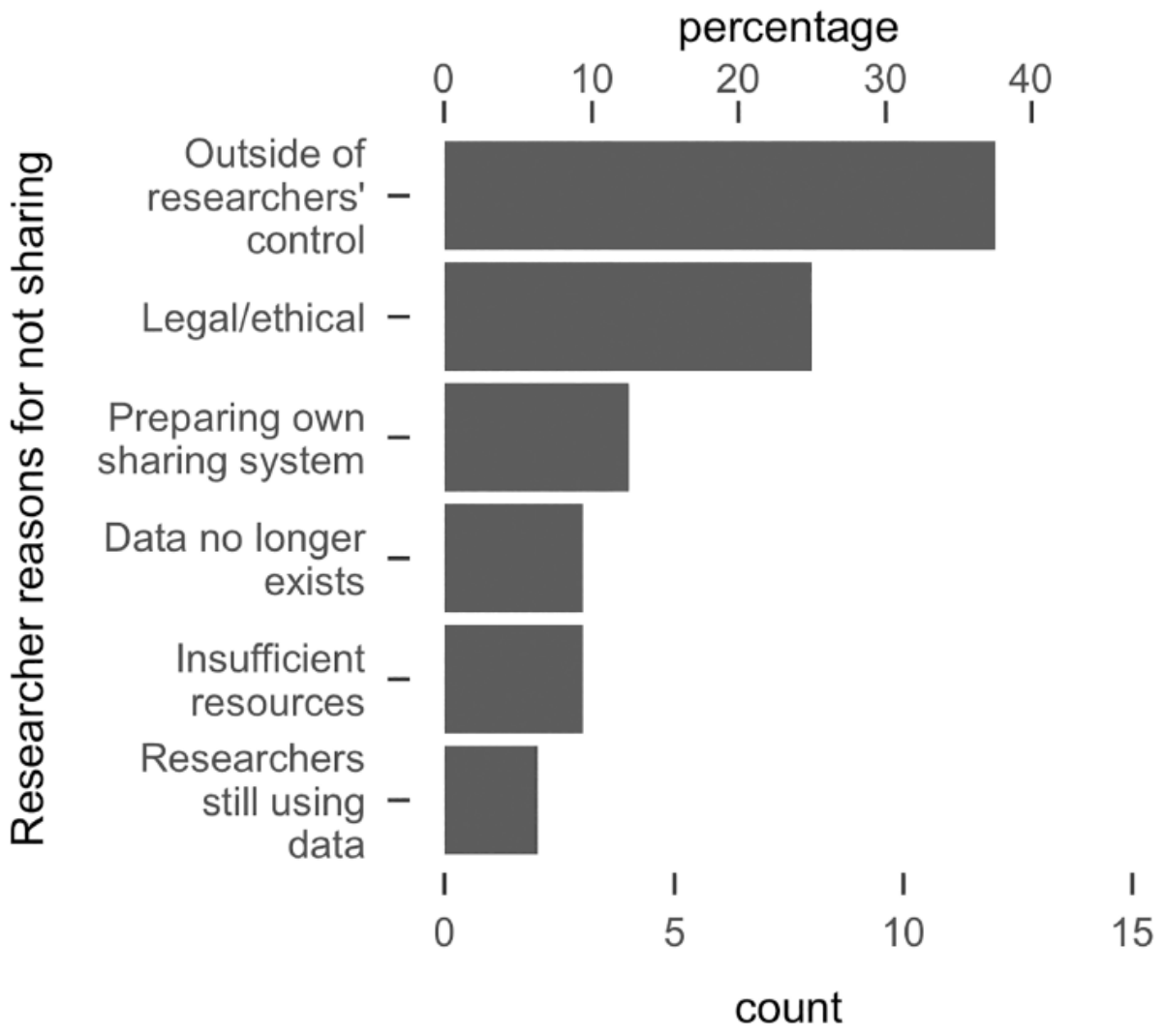


Fig 2. Reasons provided by researchers for not sharing. X-axes represent counts and percentages (of n = 32 who responded that they would not share).

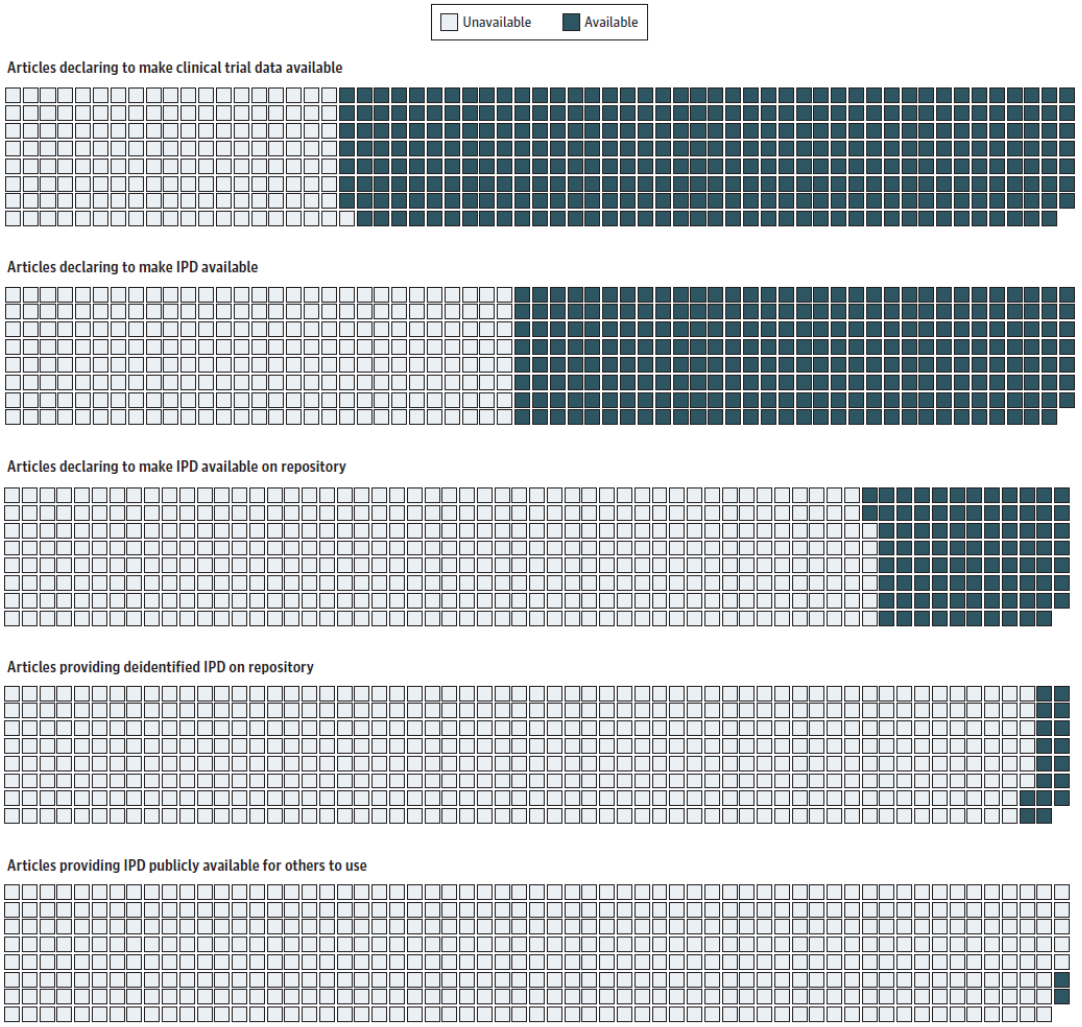
https://doi.org/10.1371/journal.pone.0201856.g002

Original Investigation | Health Policy

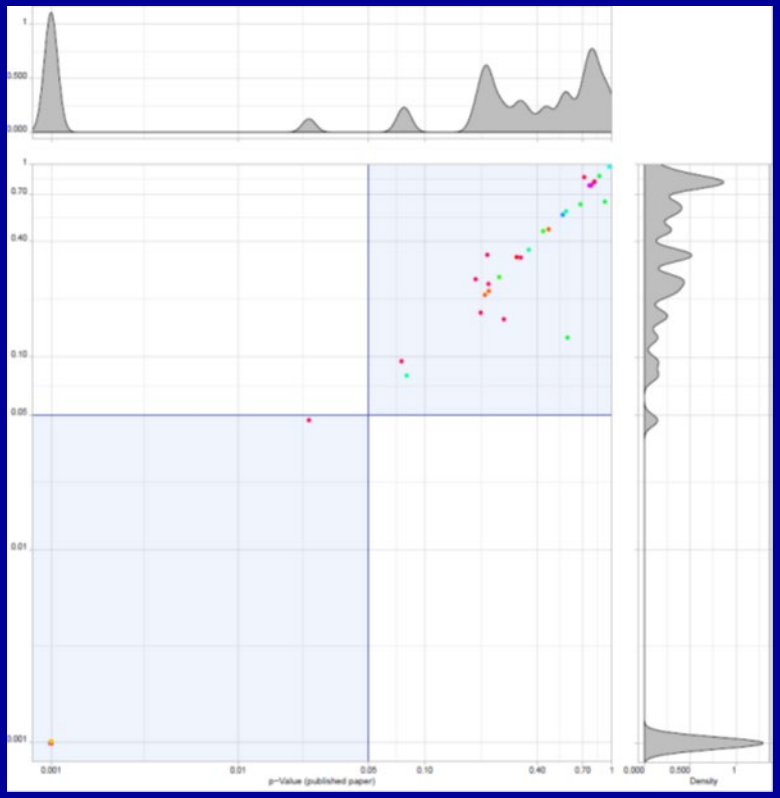
Evaluation of Data Sharing After Implementation of the International Committee of Medical Journal Editors Data Sharing Statement Requirement

Valentin Danchev, DPhil; Yan Min, MD; John Borghi, PhD; Mike Baiocchi, PhD; John P. A. Ioannidis, MD, DSc

Figure 3. Indicators of Declared and Actual Clinical Trial Individual-Participant Data (IPD) Availability as of April 10, 2020



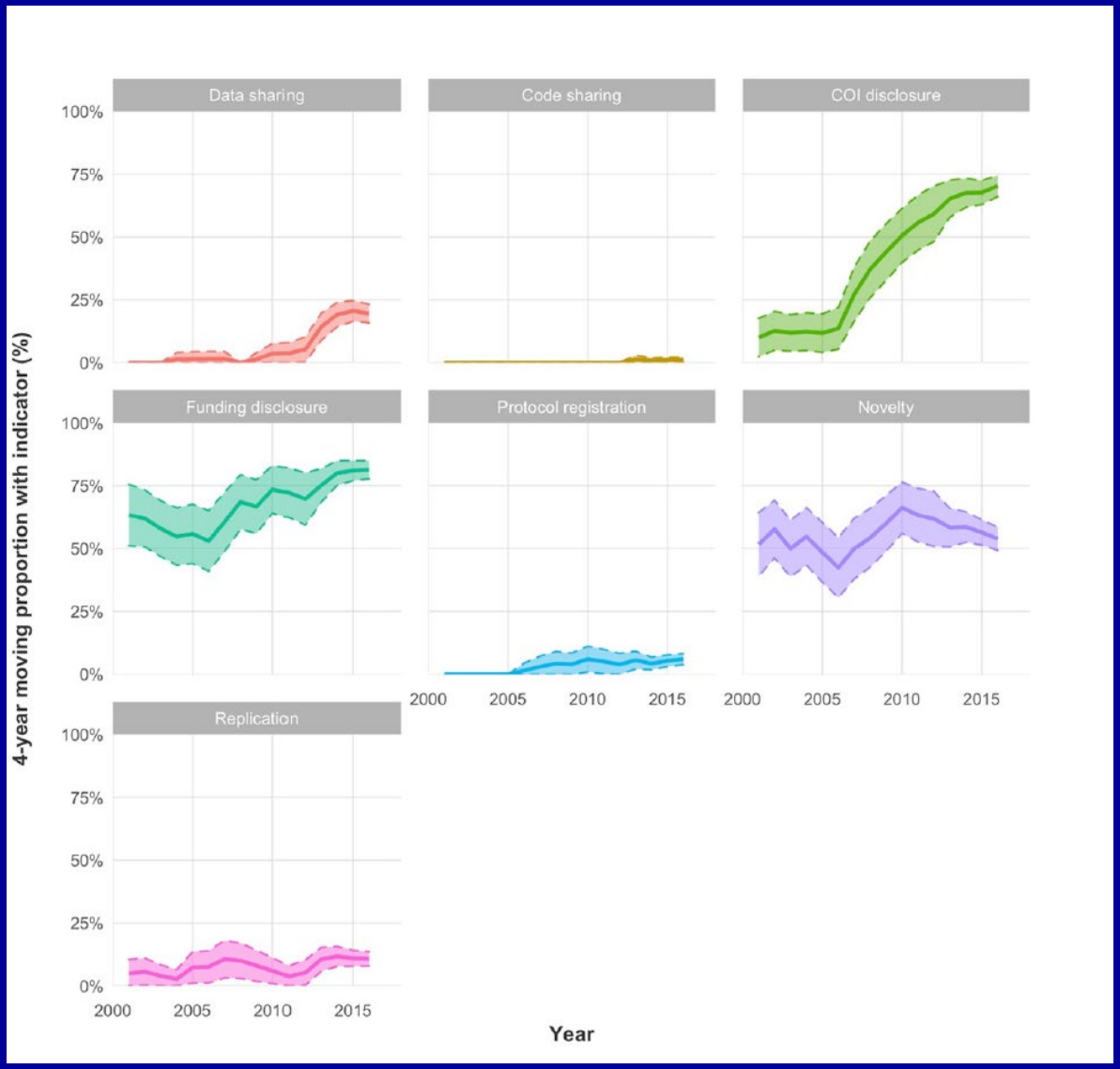
46% retrieval rate for raw data of randomized trials under full data



Naudet et al, BMJ 2018

Assessment of transparency indicators across the biomedical literature: How open is open?

Stylianos Serghiou[1,2], Despina G. Contopoulos-Ioannidis[3], Kevin W. Boyack[4], Nico Riedel[5], Joshua D. Wallach[6], John P. A. Ioannidis[1,2,7,8,9]*



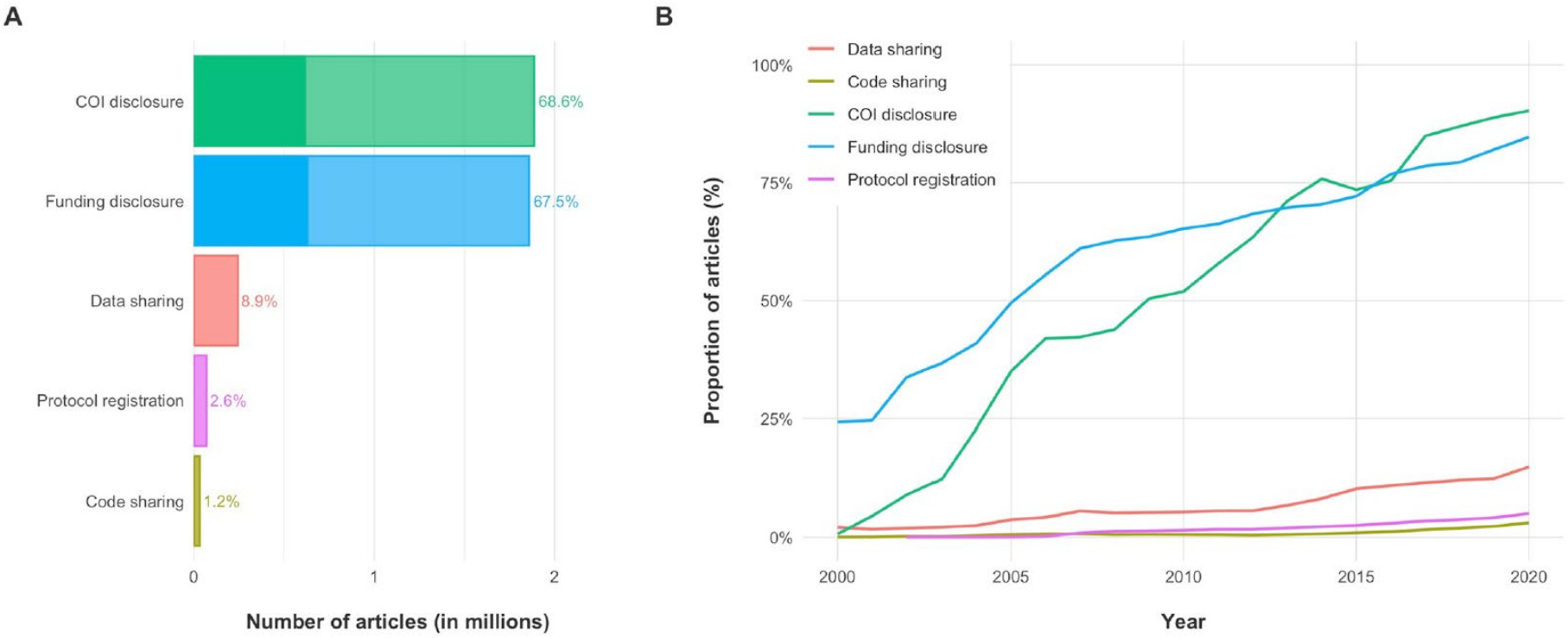
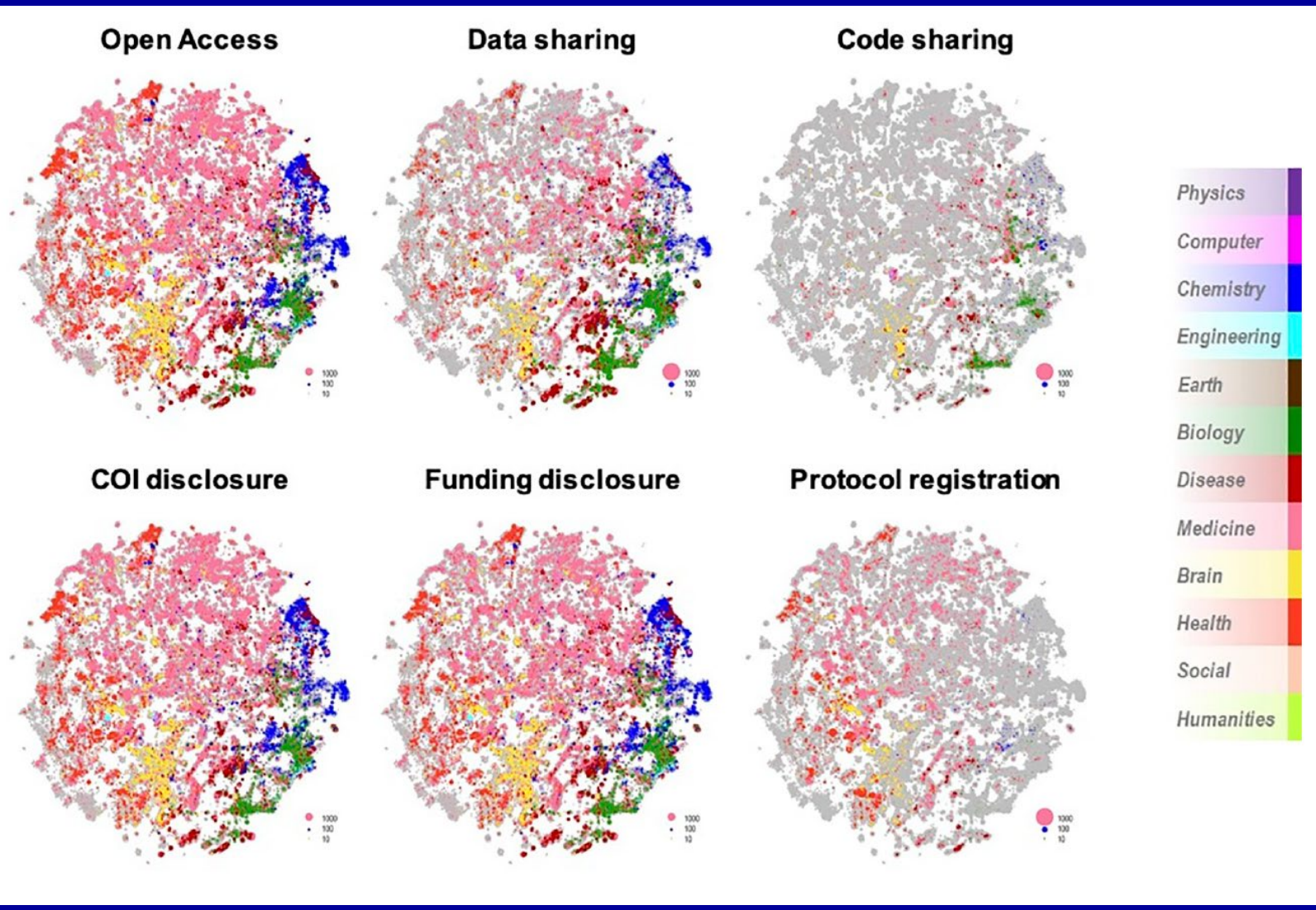
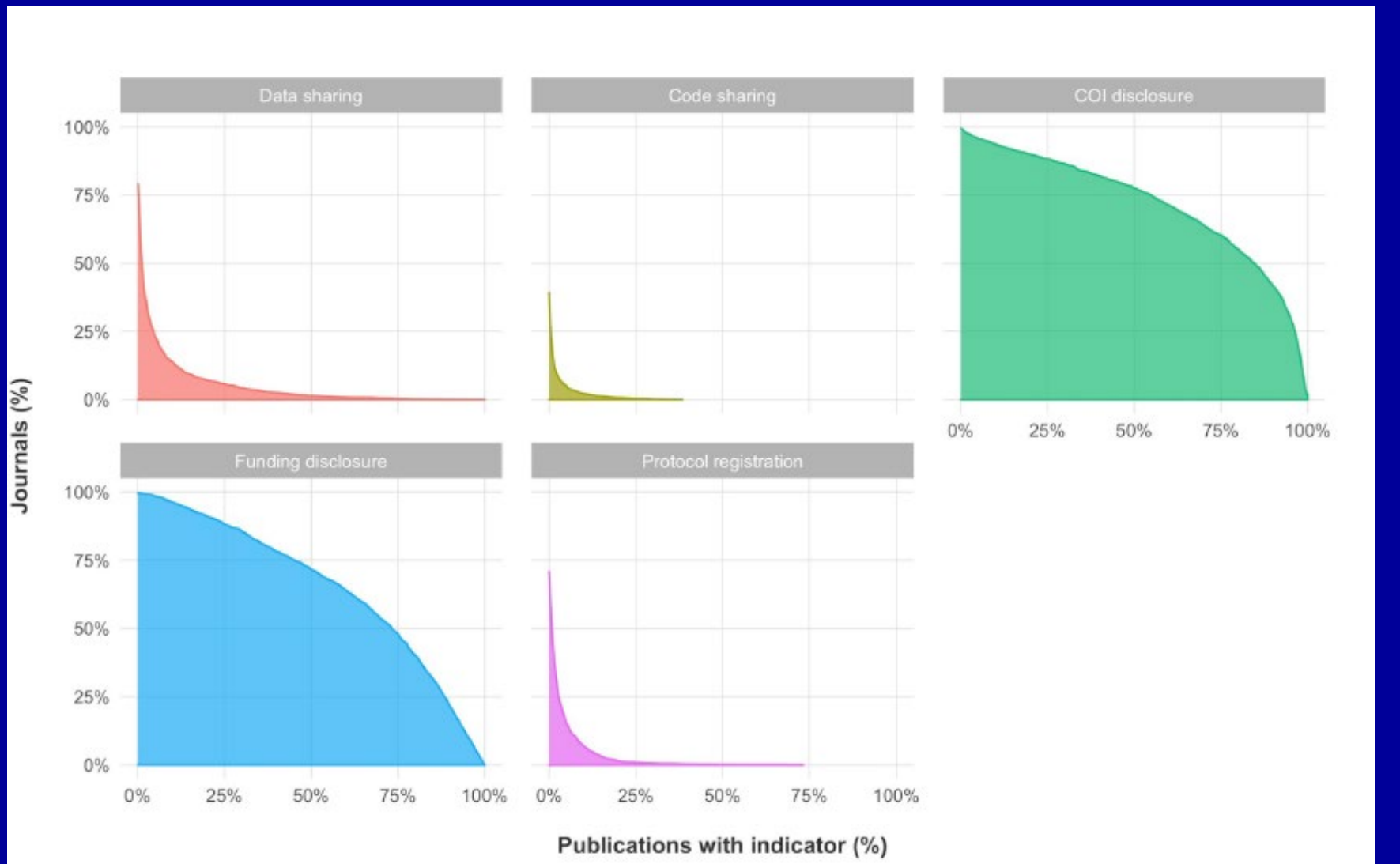


Fig 3. Indicators of transparency across the entire open biomedical literature on PMC (PMCOA) and time. (A) Most open biomedical articles report COI


Open Access
Data sharing
Code sharing
COI disclosure
Funding disclosure
Protocol registration
Physics
Computer
Chemistry
Engineering
Earth
Biology
Disease
Medicine
Brain
Health
Social
Humanities


Data sharing
Code sharing
COI disclosure
Funding disclosure
Protocol registration
100%
75%
50%
25%
0%
Journals (%)
Publications with indicator (%)

PLOS | BIOLOGY

COMMUNITY PAGE

Meta-research: Evaluation and Improvement of Research Methods and Practices

John P. A. Ioannidis*, Daniele Fanelli, Debbie Drake Dunne, Steven N. Goodman

Meta-Research Innovation Center at Stanford (METRICS), Stanford University, Stanford, California, United States of America

* jioannid@stanford.edu

Re-engineering the reward system

Table. PQRST Index for Appraising and Rewarding Research

Item in PQRST Index	Operationalization	
	Example	Data Source
P (productivity)	Number of publications in the top tier % of citations for the scientific field and year	ISI Essential Science Indicators (automated)
	Proportion of funded proposals that have resulted in ≥1 published reports of the main results	Funding agency records and automated recording of acknowledged grants (eg, PubMed)
	Proportion of registered protocols that have been published 2 y after the completion of the studies;	Study registries such as ClinicalTrials.gov for trials
Q (quality of scientific work)	Proportion of publications that fulfill ≥1 quality standards	Need to select standards (different per field/design) and may then automate to some extent; may limit to top-cited articles, if cumbersome
R (reproducibility of scientific work)	Proportion of publications that are reproducible	No wide-coverage automated database currently, but may be easy to build, especially if limited to the top-cited pivotal papers in each field.
S (sharing of data and other resources)	Proportion of publications that share their data, materials, and/or protocols (whichever items are relevant)	No wide-coverage automated database currently, but may be easy to build, eg, embed in PubMed at the time of creation of PubMed record and update if more is shared later
T (translational impact of research)	Proportion of publications that have resulted in successful accomplishment of a distal translational milestone, eg, getting promising results in human trials for intervention tested in animals or cell cultures, or licensing of intervention for clinical trials	No wide-coverage automated database currently, would need to be curated by appraiser (eg, funding agency) and may need to be limited to top-cited papers, if cumbersome

Ioannidis and Khoury, JAMA 2014

Academic criteria for promotion and tenure in biomedical sciences faculties: cross sectional analysis of international sample of universities

Danielle B Rice,[1,2] Hana Raffoul,[2,3] John P A Ioannidis,[4,5,6,7] David Moher[8,9]

ABSTRACT

OBJECTIVE

To determine the presence of a set of pre-specified traditional and non-traditional criteria used to assess scientists for promotion and tenure in faculties of biomedical sciences among universities worldwide.

DESIGN

Cross sectional study.

SETTING

International sample of universities.

PARTICIPANTS

170 randomly selected universities from the Leiden ranking of world universities list.

MAIN OUTCOME MEASURE

Presence of five traditional (for example, number of publications) and seven non-traditional (for example, data sharing) criteria in guidelines for assessing assistant professors, associate professors, and professors and the granting of tenure in institutions with biomedical faculties.

RESULTS

A total of 146 institutions had faculties of biomedical sciences, and 92 had eligible guidelines available for review. Traditional criteria of peer reviewed publications, authorship order, journal impact factor, grant funding, and national or international reputation were mentioned in 95% (n=87), 37% (34), 28% (26), 67% (62), and 48% (44) of the guidelines, respectively. Conversely, among non-traditional criteria, only citations (any mention in 26%; n=24) and accommodations for employment leave (37%; 34) were relatively commonly mentioned. Mention of alternative metrics for sharing research (3%; n=3) and data sharing (1%; 1) was rare, and three criteria (publishing in open access mediums, registering research, and adhering to reporting guidelines) were not found in any guidelines reviewed. Among guidelines for assessing promotion to full professor, traditional criteria were more commonly reported than non-traditional criteria (traditional criteria 54.2%, non-traditional items 9.5%; mean difference 44.8%, 95% confidence interval 39.6% to 50.0%; P=0.001). Notable differences were observed across continents in whether guidelines were accessible (Australia 100% (6/6), North America 97% (28/29), Europe 50% (27/54), Asia 58% (29/50), South America 17% (1/6)), with more subtle differences in the use of specific criteria.

CONCLUSIONS

This study shows that the evaluation of scientists emphasises traditional criteria as opposed to non-traditional criteria. This may reinforce research practices that are known to be problematic while insufficiently supporting the conduct of better quality research and open science. Institutions should consider incentivising non-traditional criteria.

STUDY REGISTRATION

Open Science Framework (https://osf.io/26ucp/?view_only=b80d2bc7416543639f577c1b8f756e44).

Ninth international congress on peer review and scientific publication—call for abstracts

Share your important work on peer review, publication, and the conduct of scientific research

John P A Ioannidis, [1,2] Michael Berkwits, [3] Annette Flanagin, [3,3] Fiona Godlee, [4] Theodora Bloom[4]

1 Meta-Research Innovation Center at Stanford, Stanford University, Stanford, California, USA

2 Meta-Research Innovation Center—Berlin, Berlin, Germany

3 and the JAMA Network, Chicago, Illinois, USA

4 , London, England, UK

Correspondence to: A Flanagin, annette.flanagin@jamanetwork.org

Cite this as: *BMJ* 2021;374:n2252

http://dx.doi.org/10.1136/bmj.n2252

Published: 20 September 2021

In 2019, before the covid-19 pandemic, we highlighted the unprecedented promise and peril surrounding the quantity, quality, and integrity of scientific research.[1] The pandemic has been a crash test for scientific publishing, emphasising the great successes and failures, and the promise and perils of current systems. In 2020 because of the pandemic, we announced a postponement of the ninth international challenge, and improve the standards of peer review and scientific publication. Meaningful improvements are more likely to happen in the current volatile environment, which is hopefully more receptive to change.

Abstracts summarising original, high quality research on any aspect of peer review and publication and the

Concluding comments

- The availability of 200 million scholarly articles, large numbers of systematic reviews and meta-analyses and an increasing amount of raw data allows the conduct of empirical research on research
- Transparency and reproducibility and use of optimal methods and research practices varies widely across fields and scientific applications
- There is plenty of room for improvement
- Reward system incentives may be key in improving what really matters