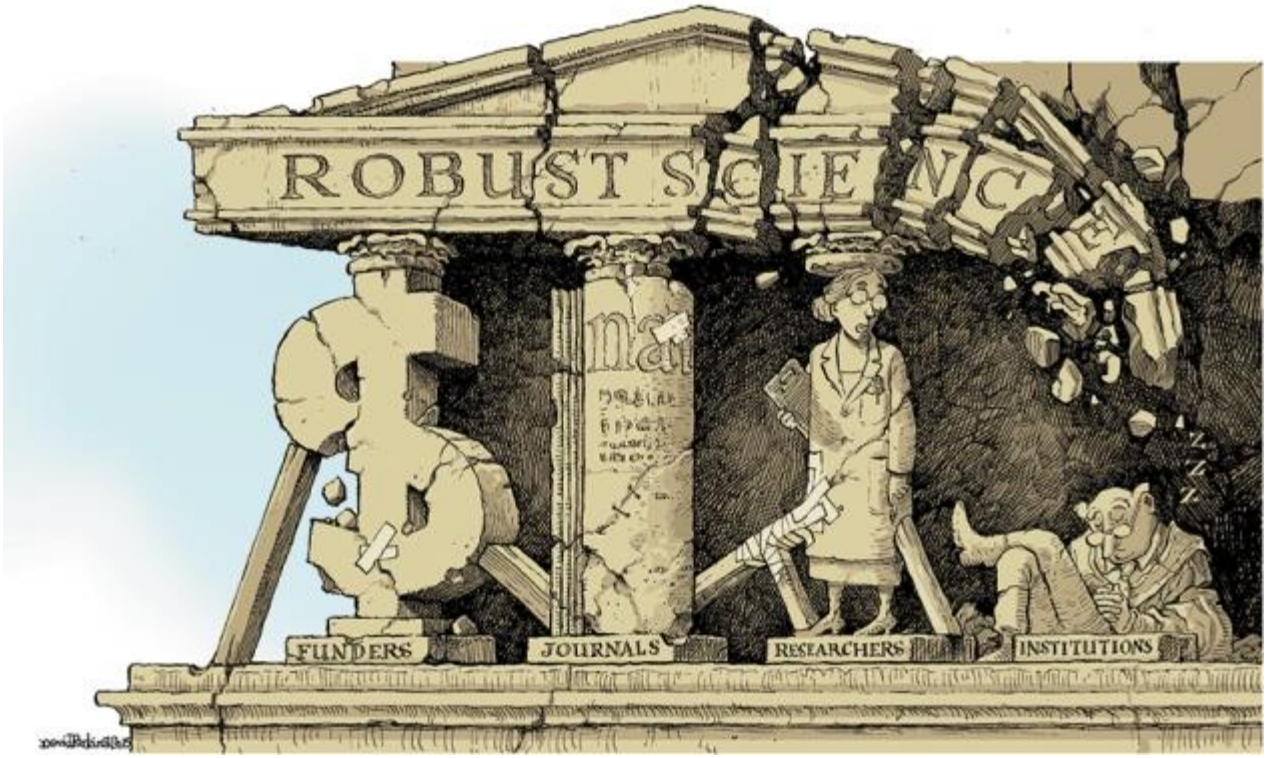
Reproducibilidad e Integridad Científica



UNIVERSITY OF MIAMI
MILLER SCHOOL OF MEDICINE
INSTITUTE FOR BIOETHICS

Curso Internacional de Integridad Científica para Investigadores de EsSalud
Lima, Perú
Octubre 18 de 2017
Sergio G. Litewka M.D. M.P.H

Essay

Why Most Published Research Findings Are False

John P. A. Ioannidis

Summary

There is increasing concern that most current published research findings are false. The probability that a research claim is true may depend on study power and bias, the number of other studies on the same question, and, importantly, the ratio of true to no relationships among the relationships probed in each scientific field. In this framework, a research finding is less likely to be true when the studies conducted in a field are smaller; when effect sizes are smaller; when there is a greater number and lesser preselection of tested relationships; where there is greater flexibility in designs, definitions, outcomes, and analytical modes; when there is greater financial and other interest and prejudice; and when more teams are involved in a scientific field in chase of statistical significance. Simulations show that for most study designs and settings, it is more likely for a research claim to be false than true. Moreover, for many current scientific fields, claimed research findings may often be simply accurate measures of the prevailing bias. In this essay, I discuss the implications of these problems for the conduct and interpretation of research.

Published research findings are sometimes refuted by subsequent evidence, with ensuing confusion and disappointment. Refutation and controversy is seen across the range of research designs, from clinical trials and traditional epidemiological studies [1–3] to the most modern molecular research [4,5]. There is increasing concern that in modern research, false findings may be the majority or even the vast majority of published research claims [6–8]. However, this should not be surprising. It can be proven that most claimed research findings are false. Here I will examine the key factors that influence this problem and some corollaries thereof.

The Essay section contains opinion pieces on topics of broad interest to a general medical audience.

Modeling the Framework for False Positive Findings

Several methodologists have pointed out [9–11] that the high rate of nonreplication (lack of confirmation) of research discoveries is a consequence of the convenient, yet ill-founded strategy of claiming conclusive research findings solely on the basis of a single study assessed by formal statistical significance, typically for a p-value less than 0.05. Research is not most appropriately represented and summarized by p-values, but, unfortunately, there is a widespread notion that medical research articles

It can be proven that most claimed research findings are false.

should be interpreted based only on p-values. Research findings are defined here as any relationship reaching formal statistical significance, e.g., effective interventions, informative predictors, risk factors, or associations. "Negative" research is also very useful. "Negative" is actually a misnomer, and the misinterpretation is widespread. However, here we will target relationships that investigators claim exist, rather than null findings.

As has been shown previously, the probability that a research finding is indeed true depends on the prior probability of it being true (before doing the study), the statistical power of the study, and the level of statistical significance [10,11]. Consider a 2 × 2 table in which research findings are compared against the gold standard of true relationships in a scientific field. In a research field both true and false hypotheses can be made about the presence of relationships. Let R be the ratio of the number of "true relationships" to "no relationships" among those tested in the field. R is characteristic of the field and can vary a lot depending on whether the field targets highly likely relationships or searches for only one or a few true relationships among thousands and millions of hypotheses that may be postulated. Let us also consider, for computational simplicity, circumscribed fields where either there is only one true relationship (among many that can be hypothesized) or the power is similar to find any of the several existing true relationships. The pre-study probability of a relationship being true is $R/(R + 1)$. The probability of a study finding a true relationship reflects the power $1 - \beta$ (one minus the Type II error rate). The probability of claiming a relationship when none truly exists reflects the Type I error rate, α. Assuming that c relationships are being probed in the field, the expected values of the 2 × 2 table are given in Table 1. After a research finding has been claimed based on achieving formal statistical significance, the post-study probability that it is true is the positive predictive value, PPV. The PPV is also the complementary probability of what Wacholder et al. have called the false positive report probability [10]. According to the 2 × 2 table, one gets $PPV = (1 - \beta)R/(R - \beta R + \alpha)$. A research finding is thus

Citation: Ioannidis JPA (2005) Why most published research findings are false. PLoS Med 2(8): e124.

Abbreviation: PPV, positive predictive value

John P. A. Ioannidis is in the Department of Hygiene and Epidemiology, University of Ioannina School of Medicine, Ioannina, Greece and Institute for Clinical Research and Health Policy Studies, Department of Medicine, Tufts-New England Medical Center, Tufts University School of Medicine, Boston, Massachusetts, United States of America. E-mail: jioannid@cc.uoi.gr

Competing Interests: The author has declared that no competing interests exist.

DOI: 10.1371/journal.pmed.0020124

Ioannidis J. Why Most Published Research Findings are false. PLOS Medicine, 2005. Vol 2;8 .e 124

“Por que la mayoría de los resultados de investigaciones son incorrectos

- Simulaciones matemáticas muestran que en la mayoría de los diseños de investigación , es mas factible que los hallazgos no sean correctos.
- Muchos estudios están pobremente diseñados y sesgados hacia un resultado favorable a las ideas del investigador
- Demasiadas variables medidas en estudios incrementan la posibilidad de errores
- Análisis estadísticos interpretan de manera inapropiada los procedimientos o la naturaleza de los datos
- Datos interpretados por sesgo personal
- Los resultados pueden ser solo interpretaciones de los sesgos prevalentes

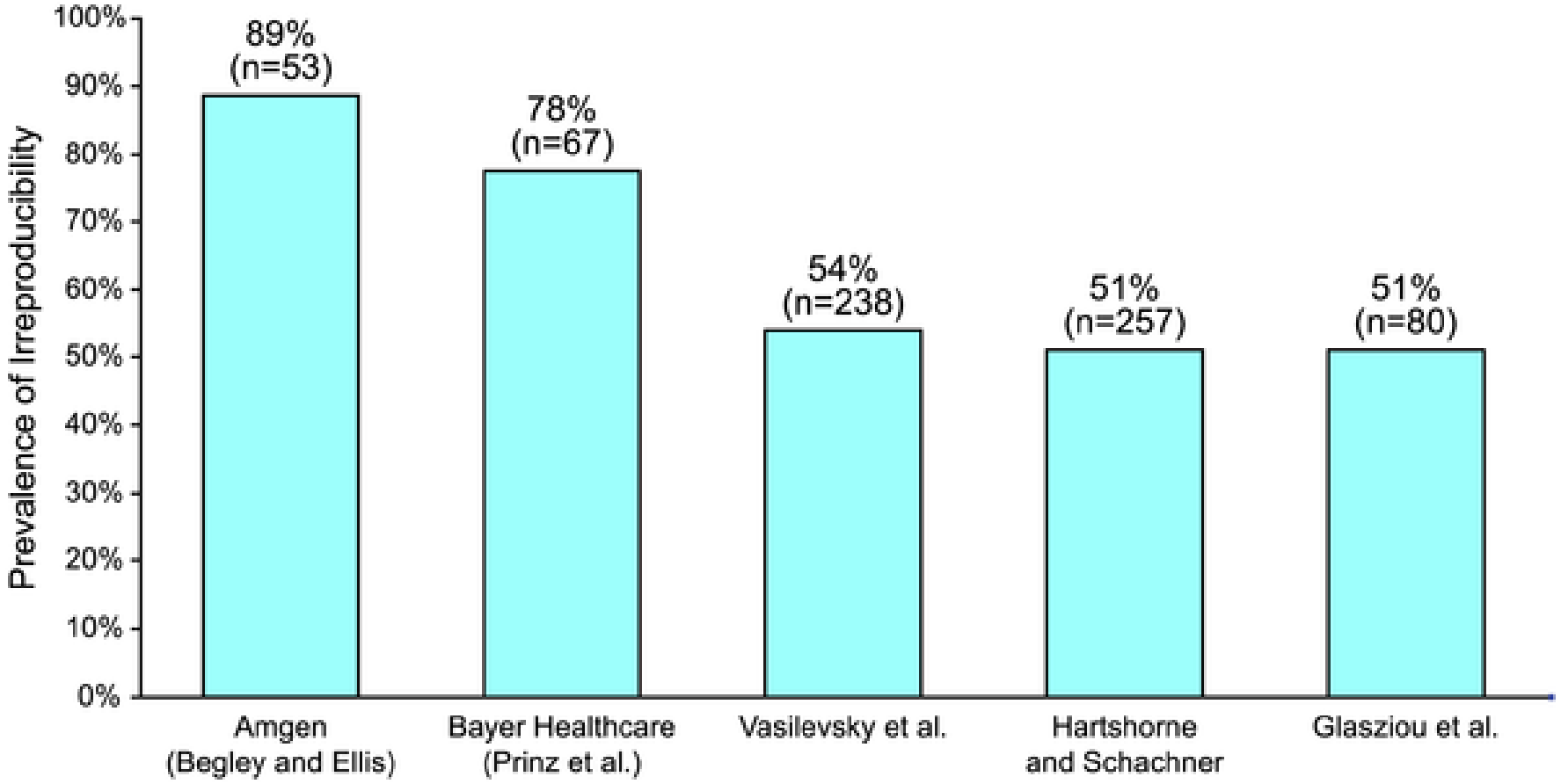
John P. A. Ioannidis

PLOS Medicine August 30, 2005

Reproducibilidad- Repetitividad

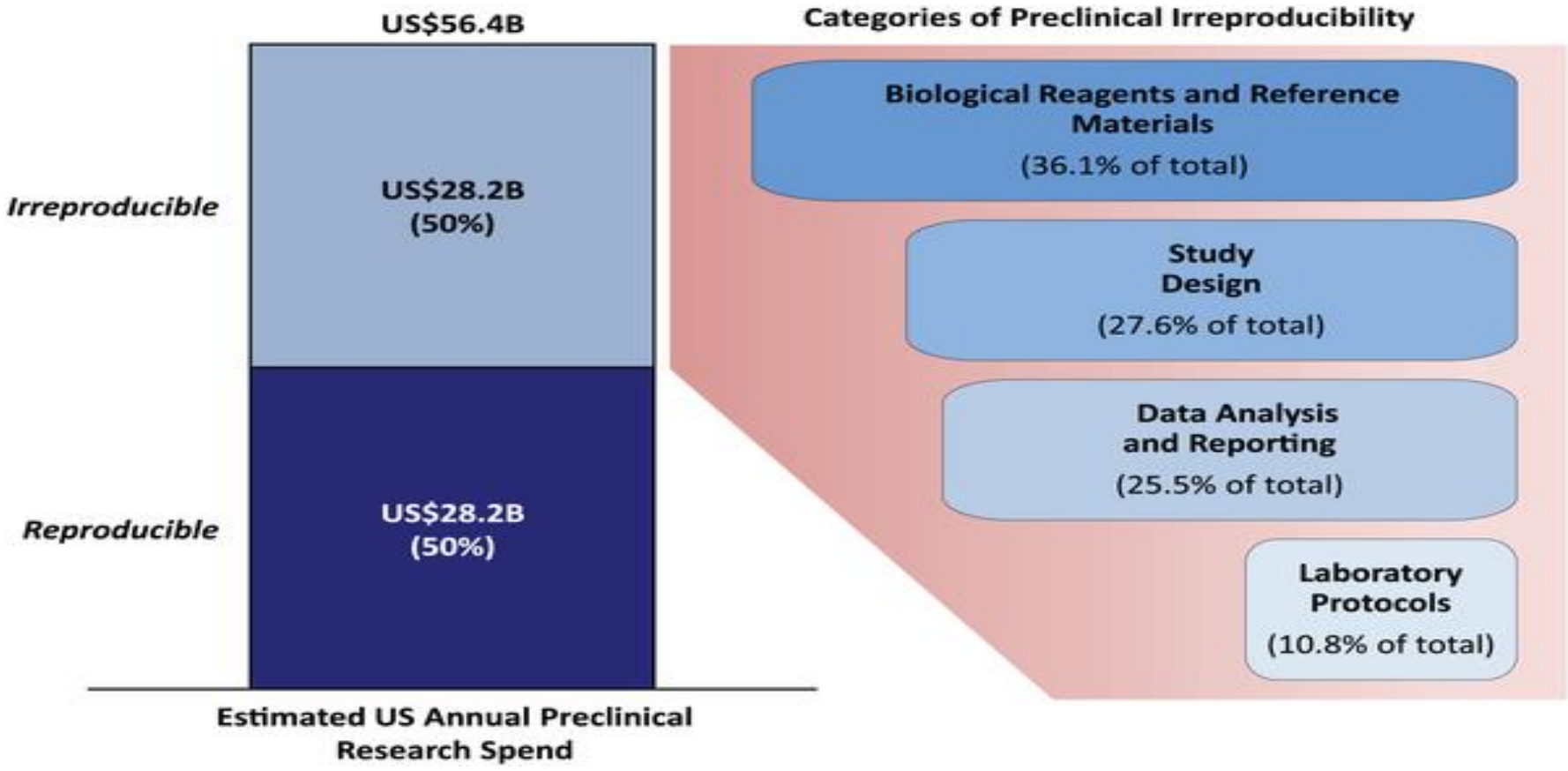
- Repetitividad : Cuando el estudio original genera los mismos hallazgos con el mismo observador, la misma tecnología, los mismos reactivos
- Reproducibilidad: Cuando el meta análisis de los grupos de datos hecho por investigadores independientes da los mismos hallazgos
- Validación : Cuando se mantiene la consistencia con datos de laboratorio o clínicos o con mediciones predictivas
  - Algunos estiman que solo el 22 al 32% de los estudios biomédicos publicados pueden ser validados. Otros hablan de porcentajes menores
    - Ioannidis J. How not to be wrong . New Scientist 2014, 22;32-33
    - Ioannidis J. Improving Validation Pratices in " omics" Research. Science 2011; 334: 1230-1232

Estudios sobre Prevalencia de Ensayos no Reproducibles



Freedman LP, Cockburn IM, Simcoe TS (2015) The Economics of Reproducibility in Preclinical Research. PLoS Biol 13(6): e1002165. doi:10.1371/journal.pbio.1002165
http://journals.plos.org/plosbiology/article?id=info:doi/10.1371/journal.pbio.1002165

Impacto Económico en Estudios Pre- clínicos No Reproducibles



Freedman LP, Cockburn IM, Simcoe TS (2015) The Economics of Reproducibility in Preclinical Research. PLoS Biol 13(6): e1002165. doi:10.1371/journal.pbio.1002165
http://journals.plos.org/plosbiology/article?id=info:doi/10.1371/journal.pbio.1002165

PLOS | BIOLOGY

Table 1: Reproducibility of research findings
Preclinical research generates many secondary publications, even when results cannot be reproduced.

From

Drug development: Raise standards for preclinical cancer research

C. Glenn Begley & Lee M. Ellis

Nature **483**, 531–533 (29 March 2012) | doi:10.1038/483531a

◂ back to article

Table 1: Reproducibility of research findings

Preclinical research generates many secondary publications, even when results cannot be reproduced.

Journal impact factor	Number of articles	Mean number of citations of non-reproduced articles*	Mean number of citations of reproduced articles
>20	21	248 (range 3–800)	231 (range 82–519)
5–19	32	169 (range 6–1,909)	13 (range 3–24)

Results from ten-year retrospective analysis of experiments performed prospectively. The term 'non-reproduced' was assigned on the basis of findings not being sufficiently robust to drive a drug-development programme.

*Source of citations: Google Scholar, May 2011.

Tables index

Table 1: Reproducibility of research findings
Preclinical research generates many secondary publications, even when results cannot be reproduced.

From

Drug development: Raise standards for preclinical cancer research

C. Glenn Begley & Lee M. Ellis

Nature **483**, 531–533 (29 March 2012) | doi:10.1038/483531a

back to article

Table 1: Reproducibility of research findings

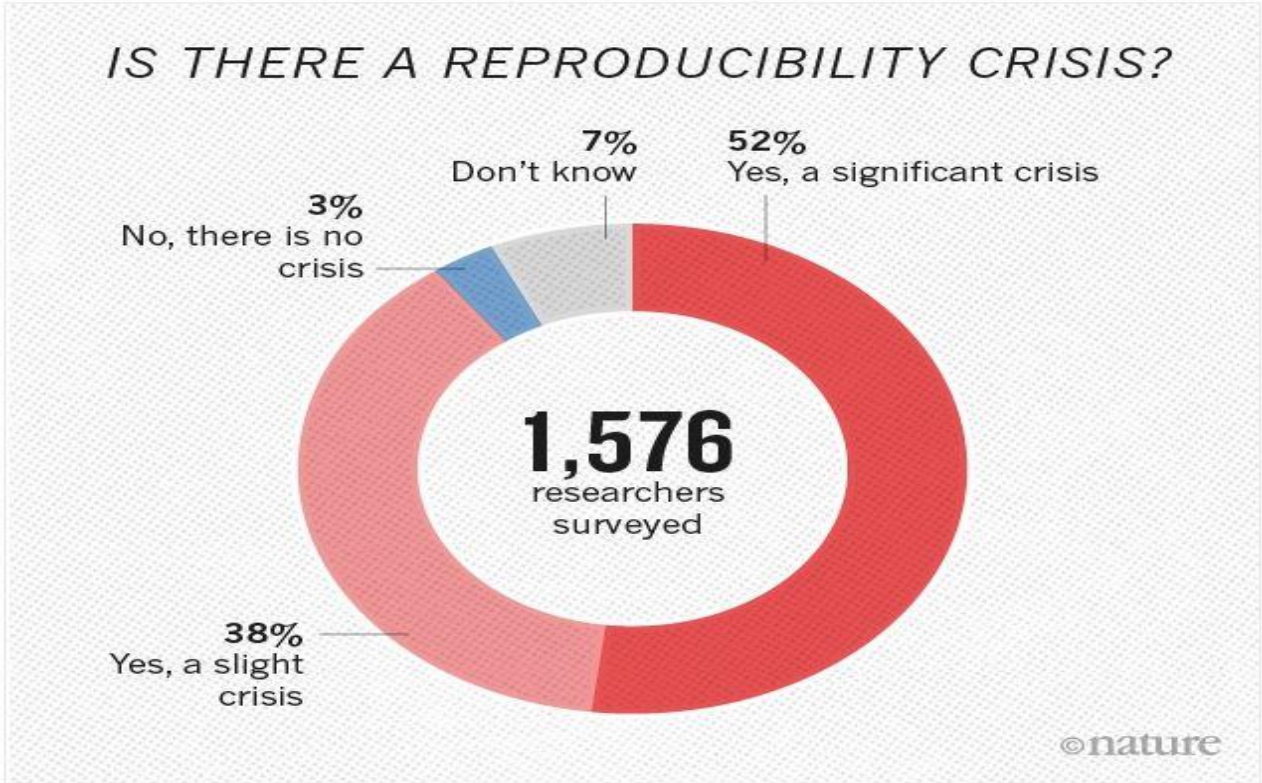
Preclinical research generates many secondary publications, even when results cannot be reproduced.

Journal impact factor	Number of articles	Mean number of citations of non-reproduced articles*	Mean number of citations of reproduced articles
>20	21	248 (range 3–800)	231 (range 82–519)
5–19	32	169 (range 6–1,909)	13 (range 3–24)

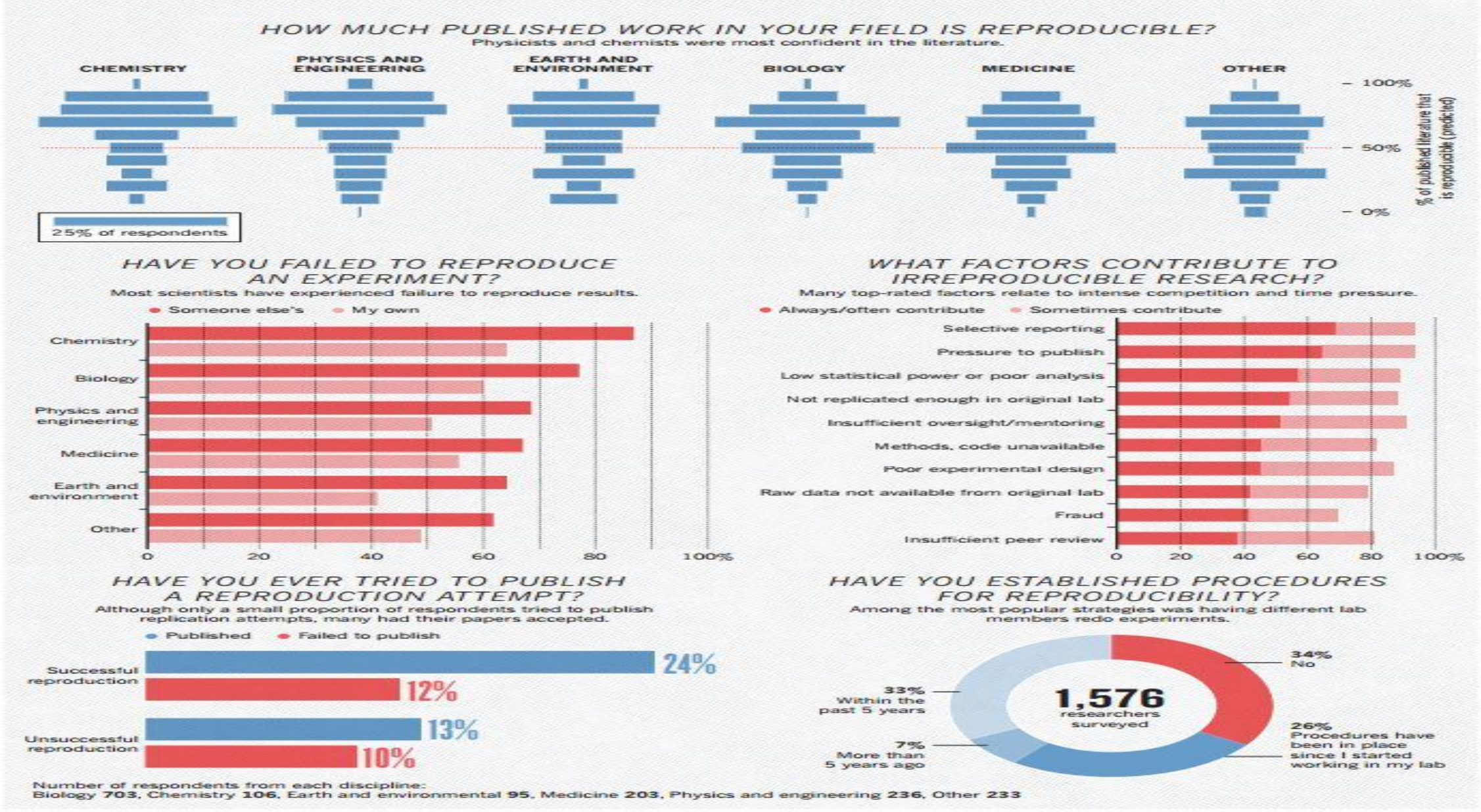
Results from ten-year retrospective analysis of experiments performed prospectively. The term 'non-reproduced' was assigned on the basis of findings not being sufficiently robust to drive a drug-development programme.

More than 70% of researchers have tried and failed to reproduce another scientist's experiments, and more than half have failed to reproduce their own experiments. Those are some of the telling figures that emerged from *Nature*'s survey of 1,576 researchers who took a brief online questionnaire on reproducibility in research.

The data reveal sometimes-contradictory attitudes towards reproducibility. Although 52% of those surveyed agree that there is a significant 'crisis' of reproducibility, less than 31% think that failure to reproduce published results means that the result is probably wrong, and most say that they still trust the published literature.

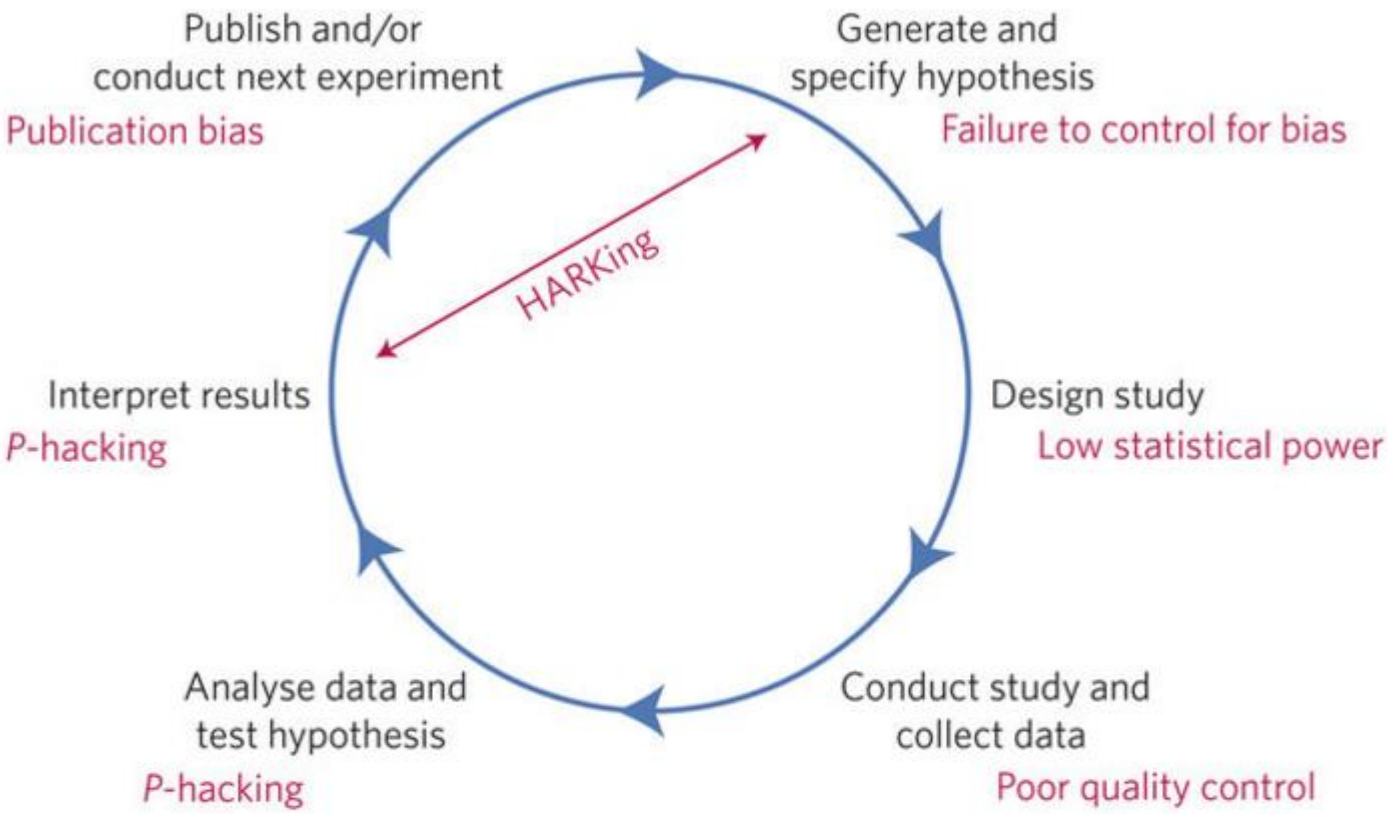


Baker, M. Is there a reproducibility crisis. Nature Vol 533; 2016



BakM. Is there a reproducibility er, crisis. Nature Vol 533; 2016

Figure 1: Threats to reproducible science.



An idealized version of the hypothetico-deductive model of the scientific method is shown. Various potential threats to this model exist (indicated in red), including lack of replication[5], hypothesizing after the results are known (HARKing)[7], poor study design, low statistical power[2], analytical flexibility[51], *P*-hacking[4], publication bias[3] and lack of data sharing[6]. Together these will serve to undermine the robustness of published research, and may also impact on the ability of science to self-correct.

nature human behaviour

A manifesto for reproducible science

Marcus R. Munafò, Brian A. Nosek, Dorothy V. M. Bishop, Katherine S. Button, Christopher D. Chambers, Nathalie Percie du Sert, Uri Simonsohn, Eric-Jan Wagenmakers, Jennifer J. Ware & John P. A. Ioannidis

Nature Human Behaviour **1**,
Article number: 0021 (2017)
doi:10.1038/s41562-016-0021

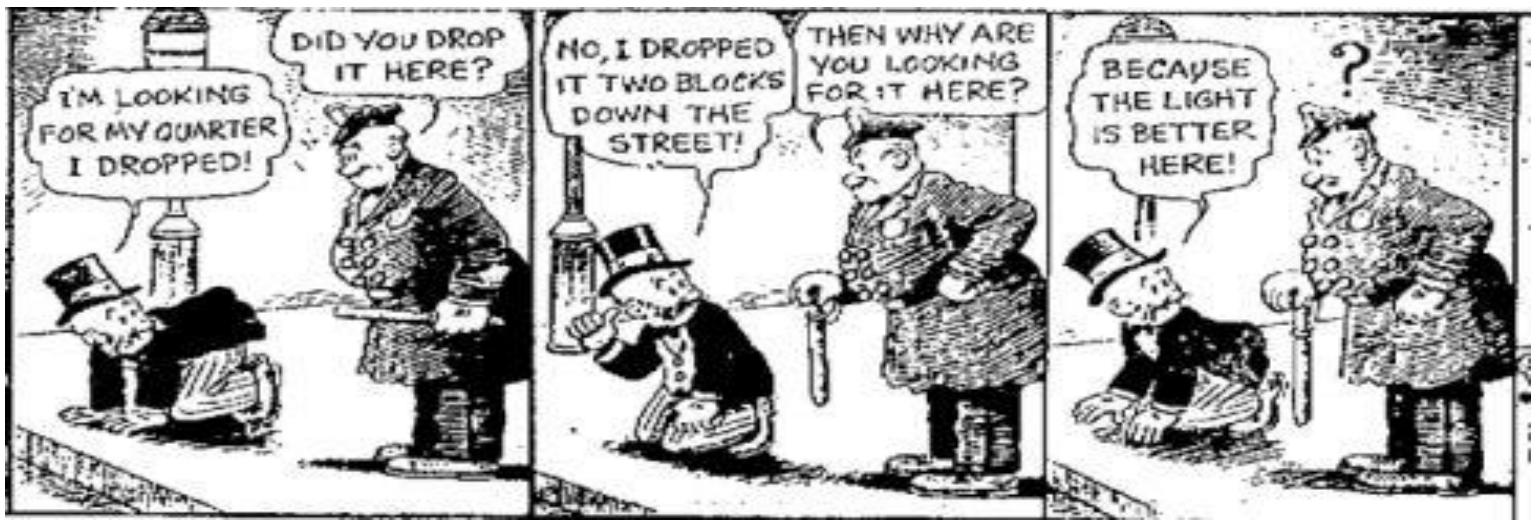
Published online: 10 January 2017

FROM THE JULY-AUGUST 2010 ISSUE

Why Scientific Studies Are So Often Wrong: The Streetlight Effect

Researchers tend to look for answers where the looking is good, rather than where the answers are likely to be hiding.

By David H. Freedman | Friday, December 10, 2010



Box 1: Recommendations: *Improving the reliability of preclinical cancer studies*

From

Drug development: Raise standards for preclinical cancer research

C. Glenn Begley & Lee M. Ellis

Nature **483**, 531–533 (29 March 2012) | doi:10.1038/483531a

Box 1: Recommendations: *Improving the reliability of preclinical cancer studies*

We recommend the following steps to change the culture of oncology research and improve the relevance of translational studies:

- There must be more opportunities to present negative data. It should be the expectation that negative preclinical data will be presented at conferences and in publications. Preclinical investigators should be required to report all findings, regardless of the outcome. To facilitate this, funding agencies, reviewers and journal editors must agree that negative data can be just as informative as positive data.
- Journal editors must play an active part in initiating a cultural change. There must be mechanisms to report negative data that are accessible through PubMed or other search engines. There should be links to journal articles in which investigators have reported alternative findings to those in an initial (sometimes considered landmark) publication. One suggestion is to include 'tags' that report whether the key findings of a seminal paper were confirmed.
- There should be transparent opportunities for trainees, technicians and colleagues to discuss and report troubling or unethical behaviours without fearing adverse consequences.
- Greater dialogue should be encouraged between physicians, scientists, patient advocates and patients. Scientists benefit from learning about clinical reality. Physicians need better knowledge of the challenges and limitations of preclinical studies. Both groups benefit from improved understanding of patients' concerns.
- Institutions and committees should give more credit for teaching and mentoring: relying solely on publications in top-tier journals as the benchmark for promotion or grant funding can be misleading, and does not recognize the valuable contributions of great mentors, educators and administrators.
- Funding organizations must recognize and embrace the need for new cancer-research tools and assist in their development, and in providing greater community access to those tools. Examples include support for establishing large cancer cell-line collections with easy investigator access (a simple, universal material-transfer agreement); capabilities for genetic characterization of newly derived tumour cell lines and xenografts; identification of patient selection biomarkers; and generation of more robust, predictive tumour models. **C.G.B. and L.M.E.**

NIH plans to enhance reproducibility

Francis S. Collins and **Lawrence A. Tabak** discuss initiatives that the US National Institutes of Health is exploring to restore the self-correcting nature of preclinical research.

A growing chorus of concern, from scientists and laypeople, contends that the complex system for ensuring the reproducibility of biomedical research is failing and is in need of restructuring[1,2]. As leaders of the US National Institutes of Health (NIH), we share this concern and here explore some of the significant interventions that we are planning.

Science has long been regarded as 'self-

shorter term, however, the checks and balances that once ensured scientific fidelity have been hobbled. This has compromised the ability of today's researchers to reproduce others' findings.

Let's be clear: with rare exceptions, we have no evidence to suggest that irreproducibility is caused by scientific misconduct. In 2011, the Office of Research Integrity of the US Department of Health and Human Ser-

that insist on data access.

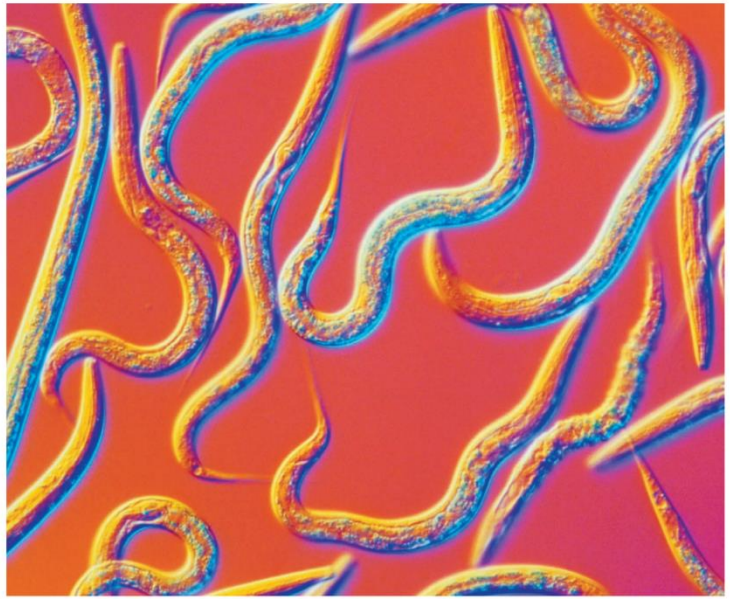
PRECLINICAL PROBLEMS

Reproducibility is potentially a problem in all scientific disciplines. However, human clinical trials seem to be less at risk because they are already governed by various regulations that stipulate rigorous design and independent oversight — including randomization, blinding, power estimates, pre-registration of outcome measures in standardized, public databases such as ClinicalTrials.gov and oversight by institutional review boards and data safety monitoring boards. Furthermore, the clinical trials community has taken important steps towards adopting standard reporting elements[7].

Preclinical research, especially work that uses animal models[1], seems to be the area that is currently most susceptible to reproducibility issues. Many of these failures have simple and practical explanations: different animal strains, different lab environments or subtle changes in protocol. Some irreproduc-

Nature 505, 612–613 (30 January 2014)
doi:10.1038/505612a

SINCLAIR STAMMERS/SPL



The free-living roundworm *Caenorhabditis elegans* is about 1 millimetre long.

A long journey to reproducible results

Replicating our work took four years and 100,000 worms but brought surprising discoveries, explain **Gordon J. Lithgow**, **Monica Driscoll** and **Patrick Phillips**.

nature International weekly journal of science

Home | News & Comment | Research | Careers & Jobs | Current Issue | Ar

Archive > Volume 548 > Issue 7668 > Comment > Article

NATURE | COMMENT

A long journey to reproducible results

Gordon J. Lithgow, **Monica Driscoll** & **Patrick Phillips**

22 August 2017

compound found in red wine — could extend lifespan in lab animals.

The possibility of drugs that stall ageing launched companies and a scientific subfield, but work in the field brought the realization that robust longevity outcomes could be challenging to replicate. Ageing research has long battled to distance itself from pseudoscientific claims. Irreproducible results from respected labs raised the spectre of yet more false promises. This had a chilling effect: some researchers (including G.J.L.) paused work on pharmacological compounds for years.

Nonetheless, scores of publications continued to appear with claims about compounds that slow ageing. There was little effort at replication. In 2013, the three of us were charged with that unglamorous task.

We have certainly not resolved discrepancies in the literature. But, by tracking the individual lifespans of more than 100,000 worms, we have found how crucial it is to understand sources of variability between labs and experiments. We even see hints of new biology that may explain discrepancies.

BROADER PROBLEM

Improved reproducibility often comes from pinning down methods. Scientists studying autophagy — the process by which cells remove degraded components — have coordinated efforts to craft and update extensive guidelines on, for instance, how to quantify that a component has been engulfed or how to verify that a gene is involved in the process[5]. In another, now-famous example, two cancer labs spent more than a year trying to understand inconsistencies[6]. It took scientists working side by side on the same tumour biopsy to reveal that small differences in how they isolated cells — vigorous stirring versus prolonged gentle rocking —

ROYAL SOCIETY OPEN SCIENCE

rsos.royalsocietypublishing.org

Opinion piece

Cite this article: Morey RD *et al*. 2016 The Peer Reviewers' Openness Initiative: incentivizing open research practices through peer review. *R. Soc. open sci.* **3**: 150547. http://dx.doi.org/10.1098/rsos.150547

Received: 10 October 2015
Accepted: 1 December 2015

Subject Category:
Research

Subject Areas:
psychology

Keywords:
science, transparency, open research, peer review

The Peer Reviewers' Openness Initiative: incentivizing open research practices through peer review

Richard D. Morey[1], Christopher D. Chambers[1], Peter J. Etchells[2], Christine R. Harris[3], Rink Hoekstra[4], Daniël Lakens[5], Stephan Lewandowsky[6,7], Candice Coker Morey[8], Daniel P. Newman[9], Felix D. Schönbrodt[10], Wolf Vanpaemel[11], Eric-Jan Wagenmakers[12] and Rolf A. Zwaan[13]

[1]Cardiff University, Cardiff, UK
[2]Bath Spa University, Bath, UK
[3]University of California, San Diego, CA, USA
[4]University of Groningen, Groningen, The Netherlands
[5]Eindhoven University of Technology, Eindhoven, The Netherlands
[6]University of Bristol, Bristol, UK
[7]University of Western Australia, Perth, Australia
[8]University of Edinburgh, Edinburgh, UK
[9]Monash University, Melbourne, Australia

Table 1 | A manifesto for reproducible science.

Theme	Proposal	Examples of initiatives/potential solutions (extent of current adoption)	Stakeholder(s)
Methods	Protecting against cognitive biases	All of the initiatives listed below (* to ****) Blinding (**)	J, F
	Improving methodological training	Rigorous training in statistics and research methods for future researchers (*) Rigorous continuing education in statistics and methods for researchers (*)	I, F
	Independent methodological support	Involvement of methodologists in research (**) Independent oversight (*)	F
	Collaboration and team science	Multi-site studies/distributed data collection (*) Team-science consortia (*)	I, F
Reporting and dissemination	Promoting study pre-registration	Registered Reports (*) Open Science Framework (*)	J, F
	Improving the quality of reporting	Use of reporting checklists (**) Protocol checklists (*)	J
	Protecting against conflicts of interest	Disclosure of conflicts of interest (***) Exclusion/containment of financial and non-financial conflicts of interest (*)	J
Reproducibility	Encouraging transparency and open science	Open data, materials, software and so on (* to **) Pre-registration (**** for clinical trials, * for other studies)	J, F, R
Evaluation	Diversifying peer review	Preprints (* in biomedical/behavioural sciences, **** in physical sciences) Pre- and post-publication peer review, for example, Publons, PubMed Commons (*)	J
Incentives	Rewarding open and reproducible practices	Badges (*) Registered Reports (*) Transparency and Openness Promotion guidelines (*) Funding replication studies (*) Open science practices in hiring and promotion (*)	J, I, F

Estimated extent of current adoption: *, <5%; **, 5–30%; ***, 30–60%; ****, >60%. Abbreviations for key stakeholders: J, journals/publishers; F, funders; I, institutions; R, regulators.

A manifesto for reproducible science

Marcus R. Munafò, Brian A. Nosek, Dorothy V. M. Bishop, Katherine S. Button, Christopher D. Chambers, Nathalie Percie du Sert, Uri Simonsohn, Eric-Jan Wagenmakers, Jennifer J. Ware & John P. A. Ioannidis

Nature Human Behaviour **1**,
Article number: 0021 (2017)
doi:10.1038/s41562-016-0021

Published online: 10 January 2017

“Seamos claros: Excepto casos puntuales, no tenemos evidencia para sugerir que los problemas de irreproducibilidad son causados por mala conducta científica. "

Francis Collins MD, PhD.
Director de los Institutos de Salud de los EE UU