SCIENTIFIC ECOSYSTEMS AND RESEARCH REPRODUCIBILITY

Marcus Munafò

University of BRISTOL

MRC Integrative Epidemiology Unit

OPEN ACCESS Freely available online

PLOS | ONE

A Survey on Data Reproducibility in Cancer Research Provides Insights into Our Limited Ability to Translate Findings from the Laboratory to the Clinic

Aaron Mobley[1], Suzanne K. Linder[2], Russell Braeuer[1], Lee M. Ellis[1,3]*, Leonard Zwelling[4]*

1 Department of Cancer Biology, The University of Texas MD Anderson Cancer Center, Houston, Texas, United States of America, **2** Department of General Internal Medicine, The University of Texas MD Anderson Cancer Center, Houston, Texas, United States of America, **3** Department of Surgical Oncology, The University of Texas MD Anderson Cancer Center, Houston, Texas, United States of America, **4** Department of Experimental Therapeutics, The University of Texas MD Anderson Cancer Center, Houston, Texas, United States of America

CORRESPONDENCE

Believe it or not: how much can we rely on published data on potential drug targets?

Florian Prinz, Thomas Schlange and Khusru Asadullah

APS ASSOCIATION FOR PSYCHOLOGICAL SCIENCE

Perspectives on Psychological Science
7(6) 657–660

Reprints and permission:
sagepub.com/journalsPermissions.nav
DOI: 10.1177/1745691612462588
http://pps.sagepub.com
SAGE

An Open, Large-Scale, Collaborative Effort to Estimate the Reproducibility of Psychological Science

Open Science Collaboration[1]

Abstract

Reproducibility is a defining feature of science. However, because of strong incentives for innovation and weak incentives for confirmation, direct replication is rarely practiced or published. The Reproducibility Project is an open, large-scale, collaborative effort to systematically examine the rate and predictors of reproducibility in psychological science. So far, 72 volunteer researchers from 41 institutions have organized to openly and transparently replicate studies published in three prominent psychological journals in 2008. Multiple methods will be used to evaluate the findings, calculate an empirical rate of replication, and investigate factors that predict reproducibility. Whatever the result, a better understanding of reproducibility will ultimately improve confidence in scientific methodology and findings.

University of BRISTOL

MRC Integrative Epidemiology Unit

RESEARCH ARTICLE

PSYCHOLOGY

Estimating the reproducibility of psychological science

Open Science Collaboration*†

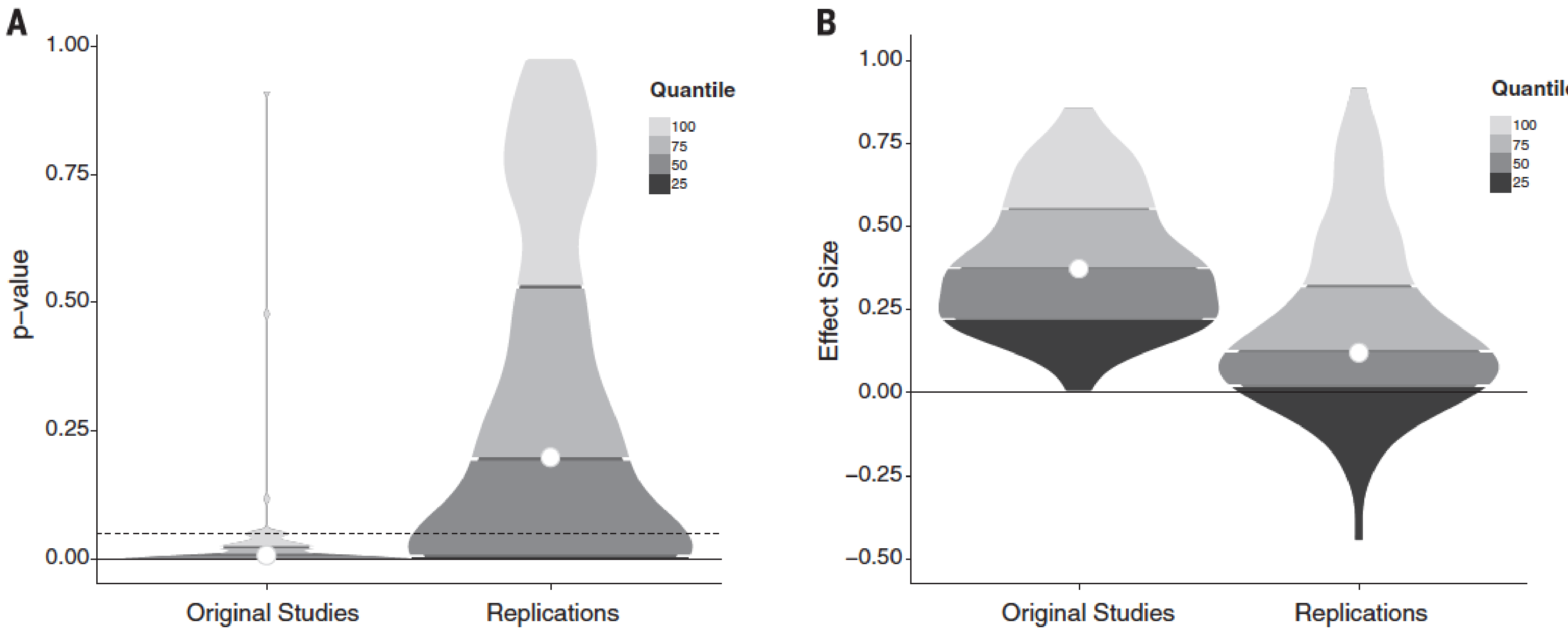
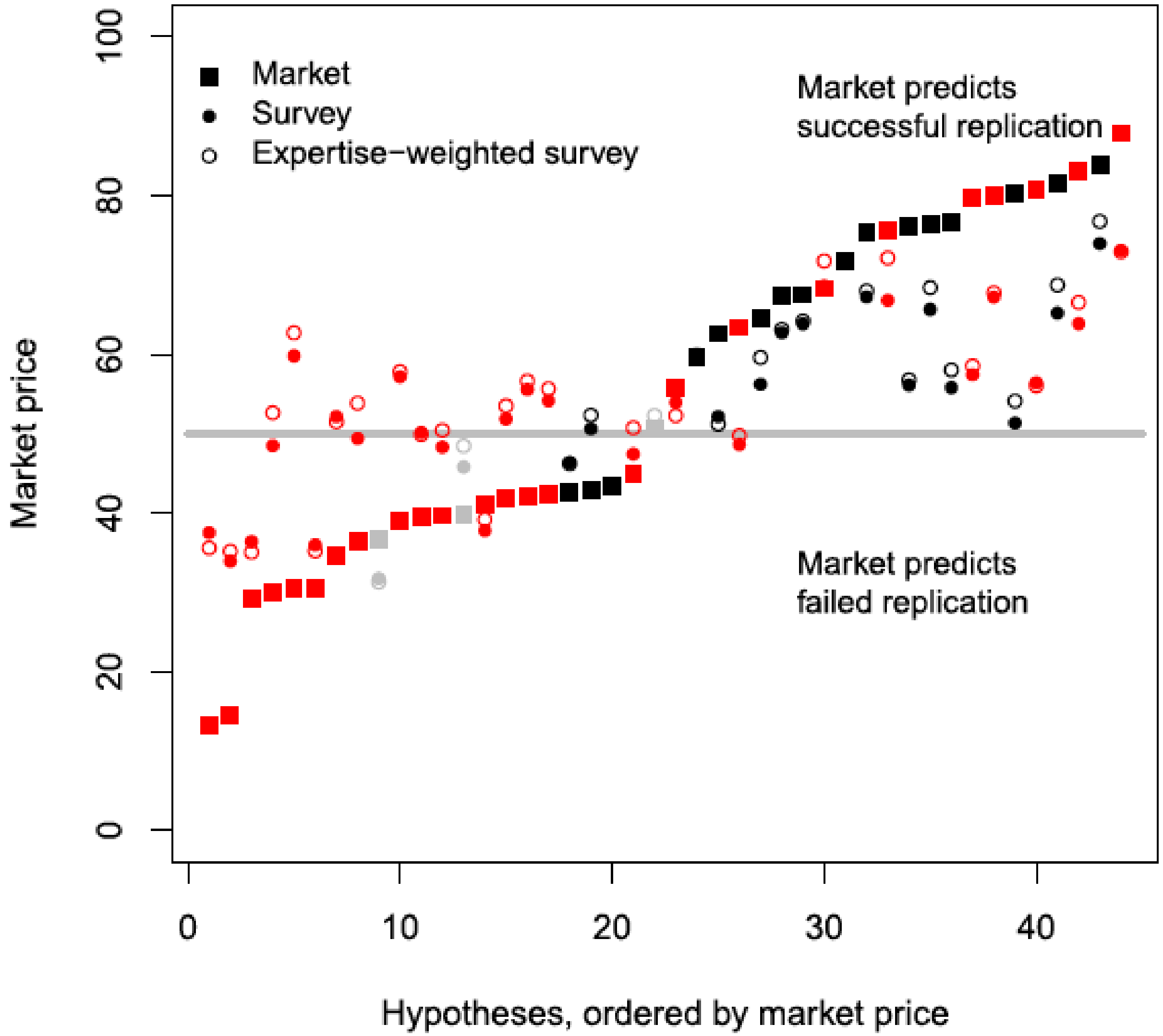


Fig. 1. Density plots of original and replication *P* values and effect sizes. (**A**) *P* values. (**B**) Effect sizes (correlation coefficients). Lowest quantiles for *P* values are not visible because they are clustered near zero.

Open Science Collaboration (2015). Science, 349.

Prediction market on the outcomes of the Reproducibility Project: Psychology



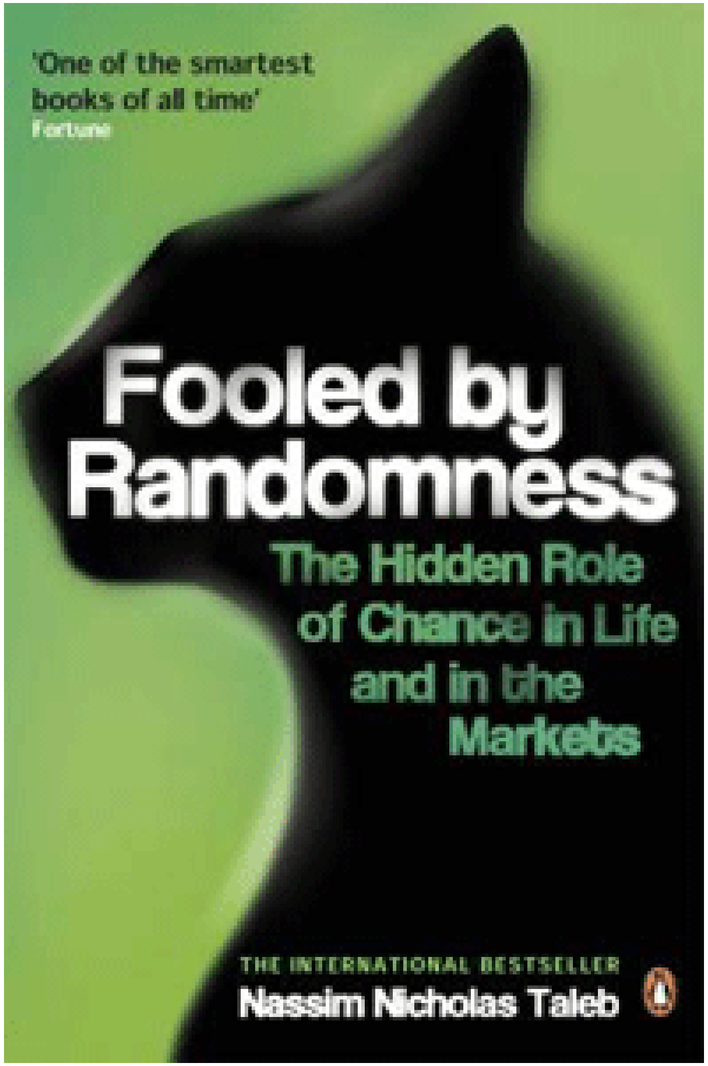
Successful replications are shown in black, unsuccessful replications in red.

Dreber et al. (2015). PNAS, 112, 1534

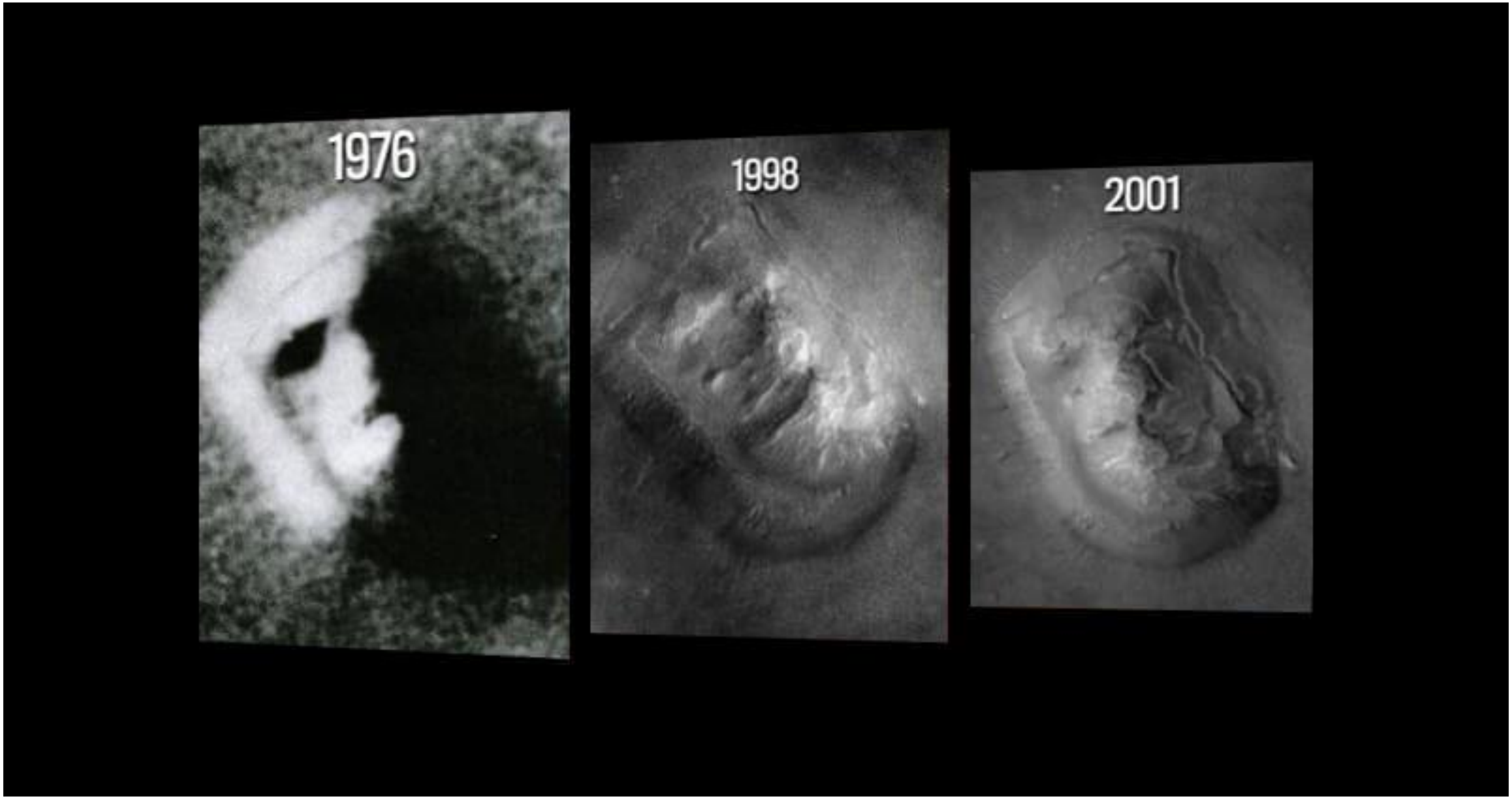
University of BRISTOL

MRC Integrative Epidemiology Unit

“Scientists may be in the business of laughing at their predecessors, but owing to an array of human mental dispositions, few realize that someone will laugh at their beliefs in the (disappointingly near) future”



Taleb (2007). Fooled by Randomness.


1976
1998
2001

Incentive Structures

Perspectives on Psychological Science
7(3) 307–309

DOI: 10.1177/1745691612441216
http://pps.sagepub.com

We Knew the Future All Along: Scientific Hypothesizing is Much More Accurate Than Other Forms of Precognition—A Satire in One Part

Arina K. Bones
University of Darache, Monte Carlo, Monaco

Bones (2012). Perspect Psychol Sci, 7, 307.

COMMENTARY

Scientists behaving badly

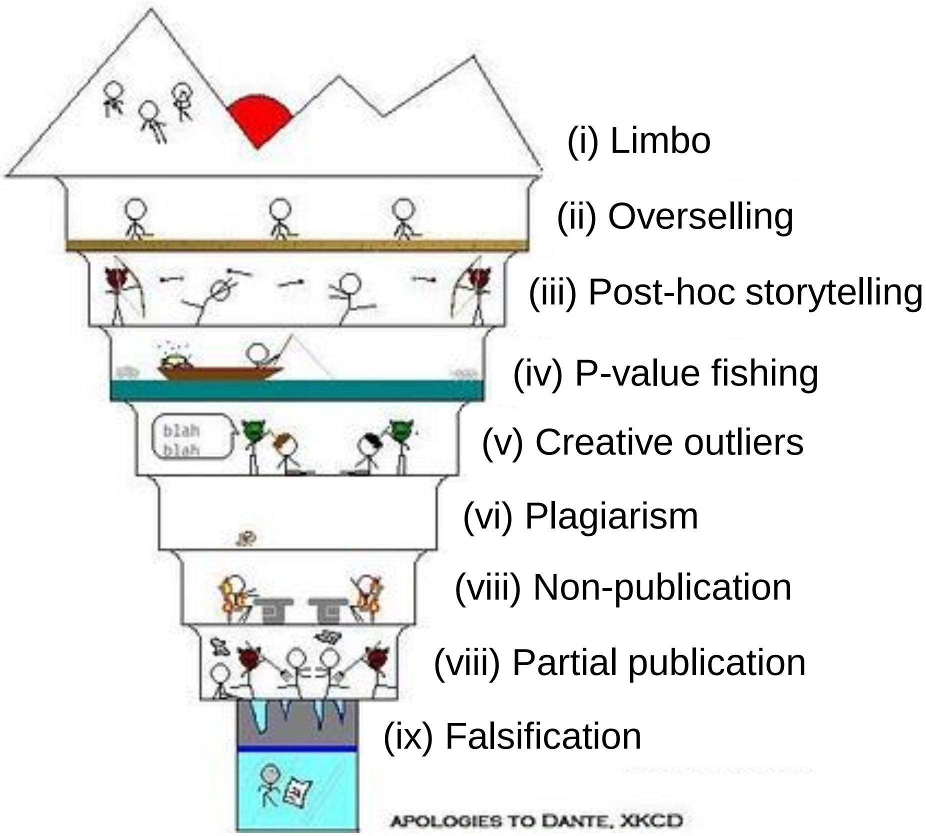
To protect the integrity of science, we must look beyond falsification, fabrication and plagiarism, to a wider range of questionable research practices, argue **Brian C. Martinson**, **Melissa S. Anderson** and **Raymond de Vries**.

"Certain features of the working environment of science may have unexpected and potentially detrimental effects on the ethical dimensions of scientists' work"

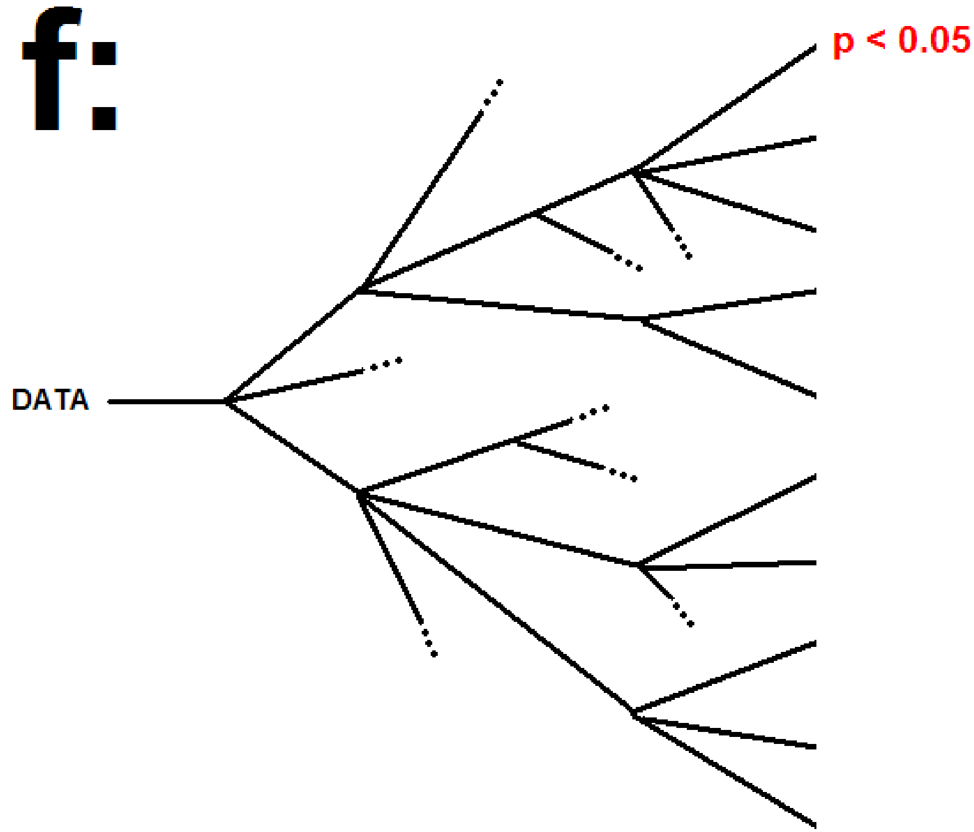
Martinson et al. (2005). Nature, 435, 737-738.

University of BRISTOL

MRC Integrative Epidemiology Unit



Neuroskeptic (2012). Perspect Psychol Sci, 7, 643-644.



http://blogs.discovermagazine.com/neuroskeptic/2013/10/16/the-f-problem

General Article

Psychological Science
22(11) 1359–1366

DOI: 10.1177/0956797611417632
http://pss.sagepub.com
SAGE

False-Positive Psychology: Undisclosed Flexibility in Data Collection and Analysis Allows Presenting Anything as Significant

Joseph P. Simmons[1], Leif D. Nelson[2], and Uri Simonsohn[1]
[1]The Wharton School, University of Pennsylvania, and [2]Haas School of Business, University of California, Berkeley

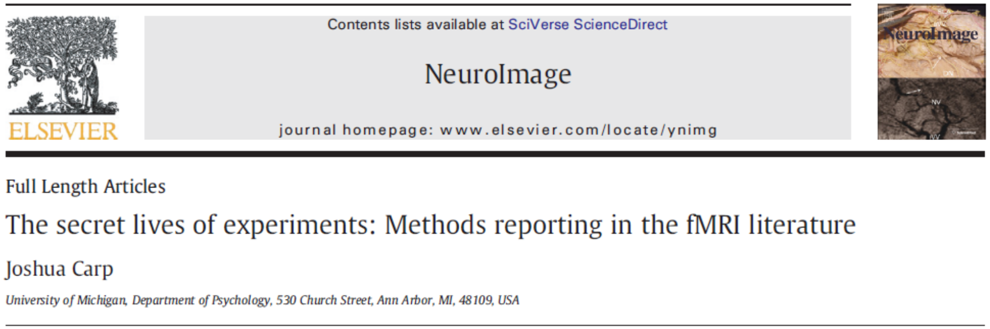
Using the same method as in Study 1, we asked ~~**20**~~ 34 **University of Pennsylvania undergraduates to listen** only **to either "When I'm Sixty-Four" by The Beatles or "Kalimba"** or "Hot Potato" by the Wiggles. We conducted our analyses after every session of approximately 10 participants; we did not decide in advance when to terminate data collection. **Then, in an ostensibly unrelated task, they indicated** only **their birth date (mm/dd/yyyy) and** how old they felt, how much they would enjoy eating at a diner, the square root of 100, their agreement with "computers are complicated machines," **their father's age,** their mother's age, whether they would take advantage of an early-bird special, their political orientation, which of four Canadian quarterbacks they believed won an award, how often they refer to the past as "the good old days," and their gender. **We used father's age to control for variation in baseline age across participants.**

An ANCOVA revealed the predicted effect: According to their birth dates, people were nearly a year-and-a-half younger after listening to "When I'm Sixty-Four" (adjusted M = 20.1 years) rather than to "Kalimba" (adjusted M = 21.5 years), $F(1, 17) = 4.92$, $p = .040$. Without controlling for father's age, the age difference was smaller and did not reach significance (Ms = 20.3 and 21.2, respectively), $F(1, 18) = 1.01$, $p = .33$.

Simmons et al. (2011). Psychol Sci, 22, 1359-1366.

University of BRISTOL

MRC Integrative Epidemiology Unit



Contents lists available at SciVerse ScienceDirect

NeuroImage

journal homepage: www.elsevier.com/locate/ynimg

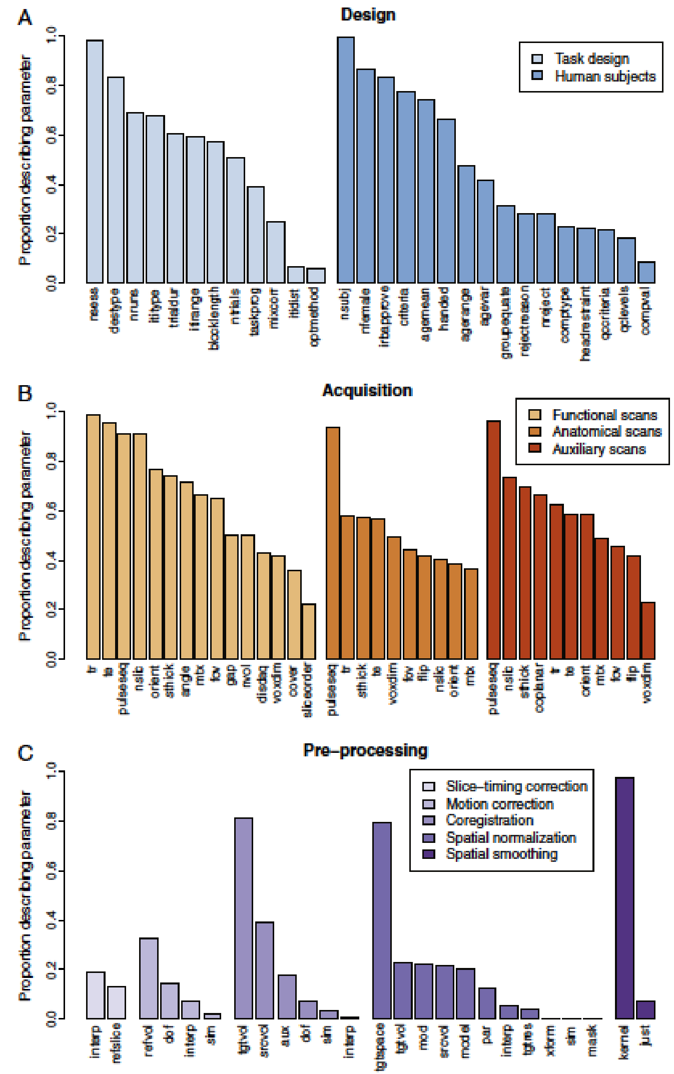
ELSEVIER

Full Length Articles

The secret lives of experiments: Methods reporting in the fMRI literature

Joshua Carp

University of Michigan, Department of Psychology, 530 Church Street, Ann Arbor, MI, 48109, USA



"…nearly as many unique analysis pipelines as there were studies in the sample…"

Carp (2012). Neuroimage, 63, 289-300.

University of BRISTOL

MRC Integrative Epidemiology Unit

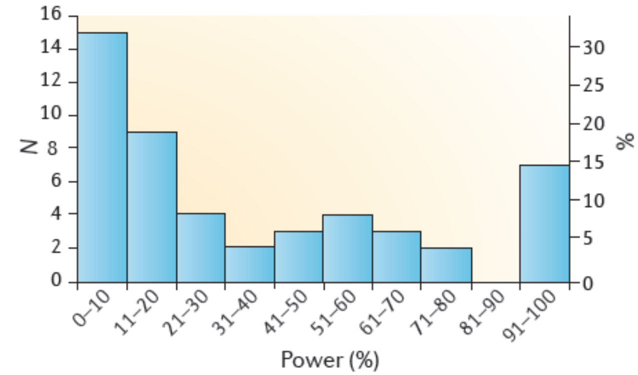
Incentive Structures

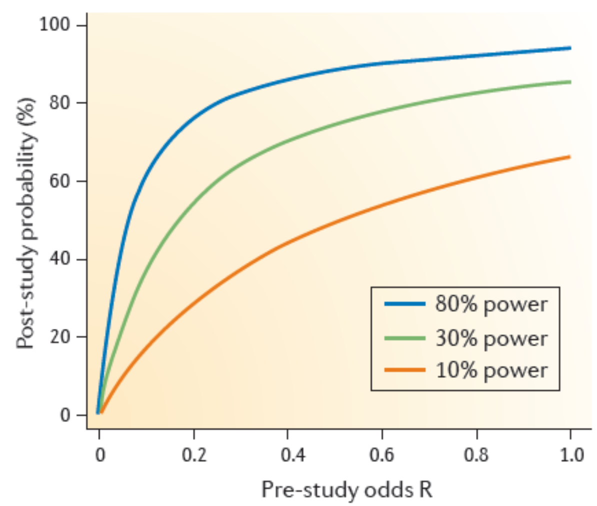
ANALYSIS

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.





Button et al. (2013). Nat Rev Neurosci, 14, 365-376.

Table 1. Reporting of measures to reduce the risk of bias in publications from 2009–2010 that were randomly selected, identified in the context of systematic reviews or from leading UK institutions.

	Randomisation		Blinding		Sample Size Calculation	
	n/*N*	% (95% CI)	*n*/*N*	% (95% CI)	*n*/*N*	% (95% CI)
PubMed	7/14	50 (23–77)	2/14	14 (2–43)	0/14	0 (0–23)
CAMARADES	76/213	36 (29–42)	79/213	37 (30–44)	2/213	1 (0–3)
Institutions	148/1028	14 (12–17)	201/1165	17 (15–20)	16/1168	1 (1–2)

Studies from top-ranked UK institutions perform worse on reporting of measures to reduce the risk of bias than studies selected at random from PubMed…

Macleod et al. (2015). PLOS Biol, 13, e1002301.

US studies may overestimate effect sizes in softer research

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

Predictor	Nonbehavioral (k = 40, n = 566)	Behavioral, all (k = 42, n = 608)	Biobehavioral (k = 20, n = 308)	Behavioral (k = 22, n = 300)
(Intercept)	0.42 [0.40, 0.46]	0.55 [0.51, 0.56]	0.51 [0.47, 0.54]	0.57 [0.50, 0.59]
United States vs. rest	−0.02 [−0.06, 0.00]	0.03 [0.02, 0.06]	0.03 [0.00, 0.07]	0.04 [0.01, 0.07]
Study size (SE)	0.43 [0.27, 0.53]	0.11 [0.07, 0.23]	0.20 [0.11, 0.31]	0.06 [0.01, 0.29]
Pub. order	0.02 [0.00, 0.03]	0.00 [−0.01, 0.01]	0.01 [0.00, 0.05]	0.00 [−0.02, 0.01]
USA*SE	−0.21 [−0.47, 0.22]	−0.19 [−0.31, −0.03]	−0.16 [−0.34, 0.12]	−0.22 [−0.46, −0.02]
USA*pub. order	−0.02 [−0.05, 0.01]	0.00 [−0.02, 0.03]	−0.02 [−0.06, 0.01]	0.01 [−0.02, 0.05]

Fanelli & Ioannidis (2013). PNAS, 5, e10271.

	2008	2009	2010	2011	2012	2013	2014	2015	2016
Nature, Science	$26,212	$26,006	$25,781	$25,365	$33,990	$36,658	$38,908	$43,783	$43,783
PNAS	$3,156	$3,025	$3,353	$3,443	$3,664	$3,619	$3,751	$3,513	$3,513
PLOS One	$1,096	$1,086	$1,035	$994	$991	$915	$941	$984	$984
MIS Quarterly	$2,613	$2,570	$2,553	$2,654	$2,876	$2,861	$2,992	$2,938	$2,938
JASIST	$1,737	$1,758	$1,741	$1,887	$2,066	$2,303	$2,435	$2,488	$2,488
Journal of Documentation	$1,082	$1,087	$1,042	$1,111	$1,167	$1,265	$1,329	$1,408	$1,408
Library Hi Tech	$781	$775	$726	$741	$740	$768	$795	$783	$783
LIBRI	$650	$644	$577	$560	$538	$509	$517	$484	$484

* All the amounts are full amount (in USD) awarded to the first author

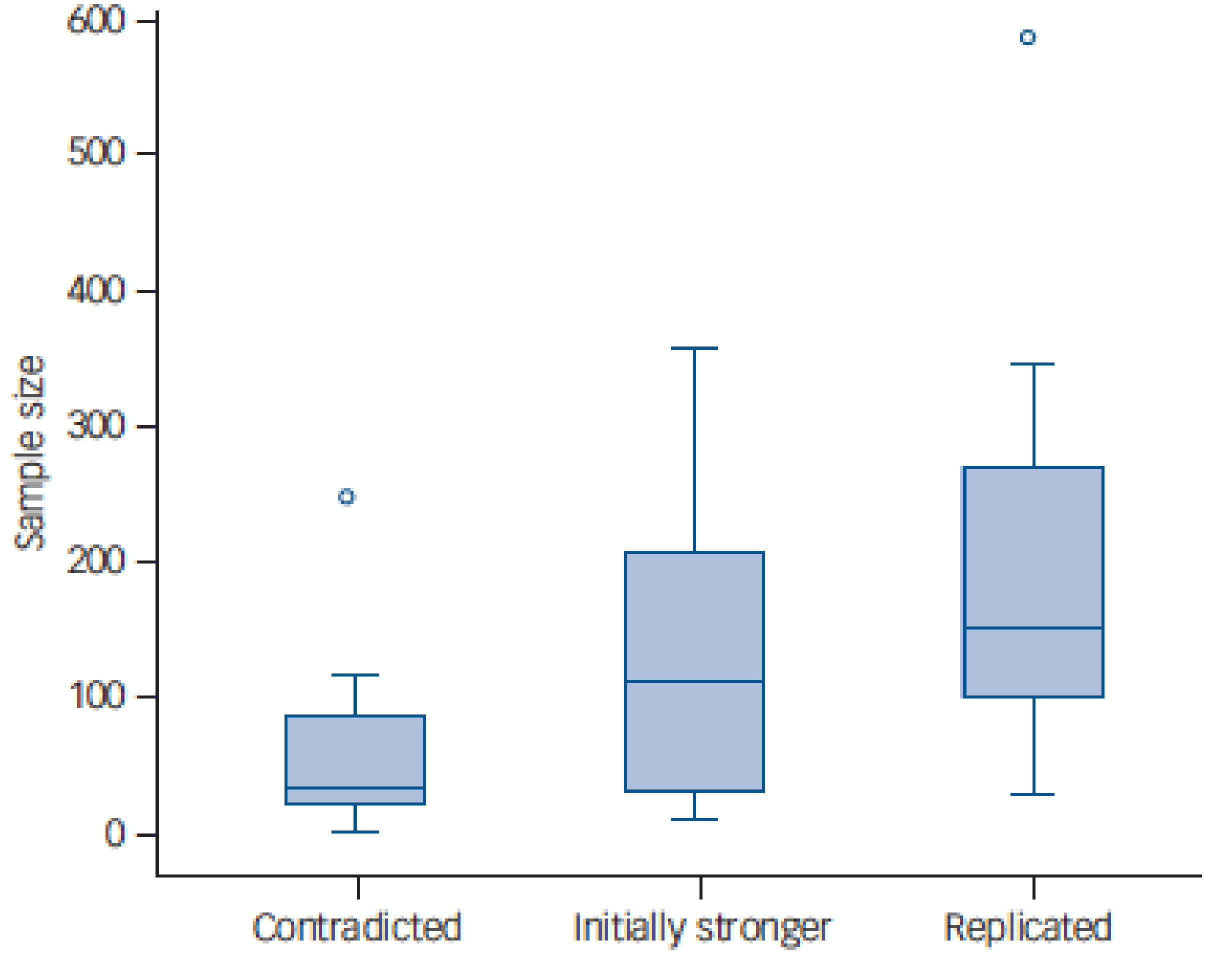
Perspectives on Psychological Science
7(6) 645–654

DOI: 10.1177/1745691612464056
http://pps.sagepub.com

Why Science Is Not Necessarily Self-Correcting

John P. A. Ioannidis
Stanford Prevention Research Center, Department of Medicine and Department of Health Research and Policy, Stanford University School of Medicine, and Department of Statistics, Stanford University School of Humanities and Sciences

Ioannidis (2012). Perspect Psychol Sci, 7, 645-654.



"Among 83 articles recommending effective interventions, 40 had not been subject to any attempt at replication…"

Tajika et al. (2015). Br J Psychiatry, 207, 357-362.

ELSEVIER

Journal of Clinical Epidemiology ■ (2012) ■

Journal of Clinical Epidemiology

ORIGINAL ARTICLE

Primary study authors of significant studies are more likely to believe that a strong association exists in a heterogeneous meta-analysis compared with methodologists

Orestis A. Panagiotou[a], John P.A. Ioannidis[b,c,d,e,*]

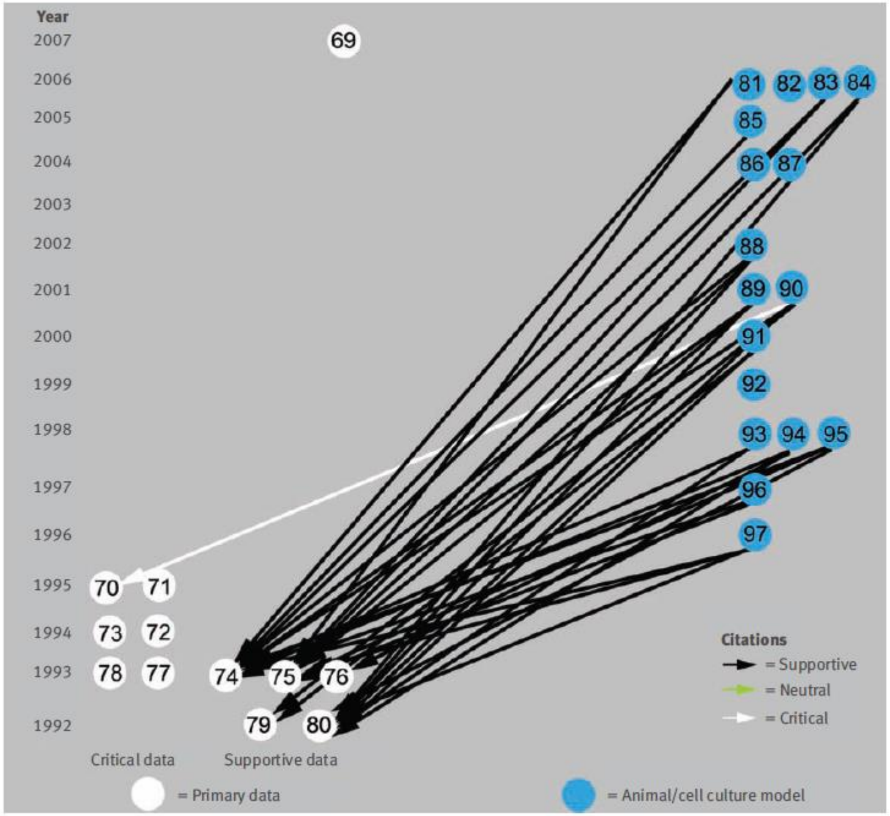
Panagiotou & Ioannidis (2012). J Clin Epidemiol, 65, 740-747.

University of BRISTOL

MRC Integrative Epidemiology Unit

How citation distortions create unfounded authority: analysis of a citation network

Steven A Greenberg, associate professor of neurology



Investigated citation network of papers addressing the belief that B amyloid, a protein accumulated in the brain in Alzheimer’s disease, is produced by and injures skeletal muscle of patients with inclusion body myositis.

Greenberg (2009). Br Med J, 339, b2680.

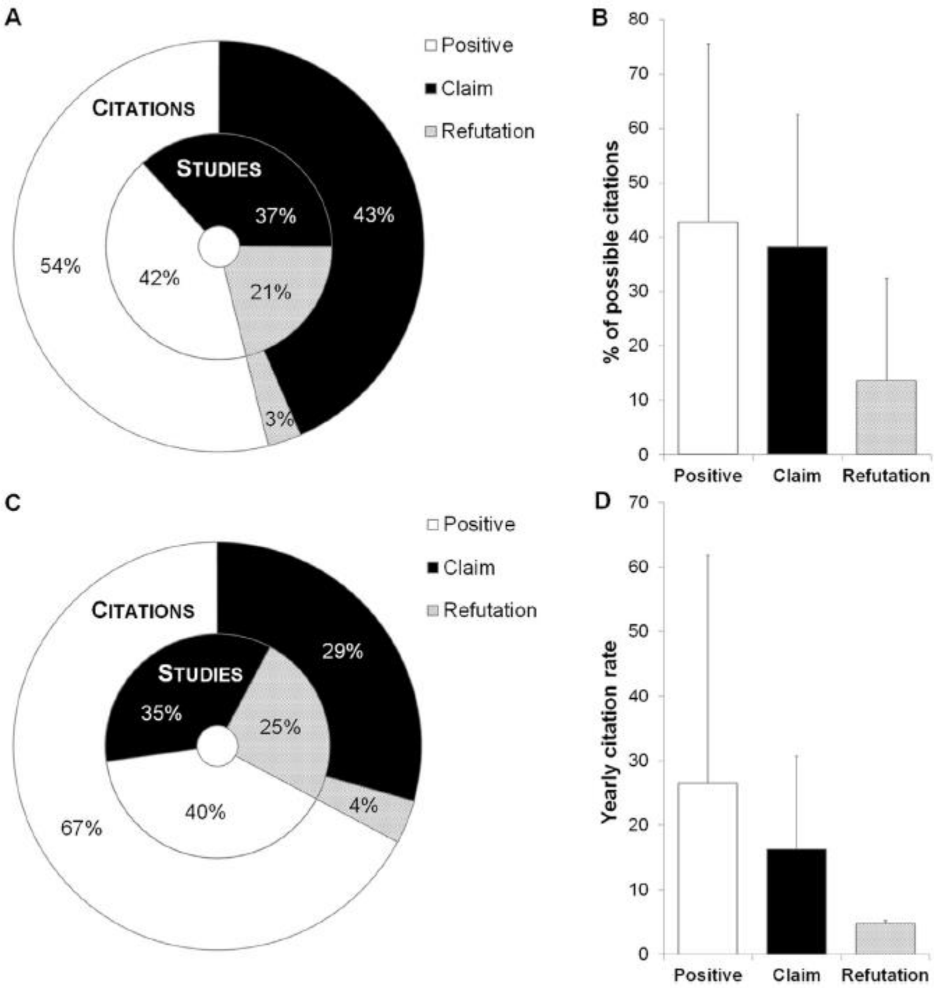
University of BRISTOL

MRC Integrative Epidemiology Unit

Abstracts often "spin" results to give impression that results are positive when they are not.

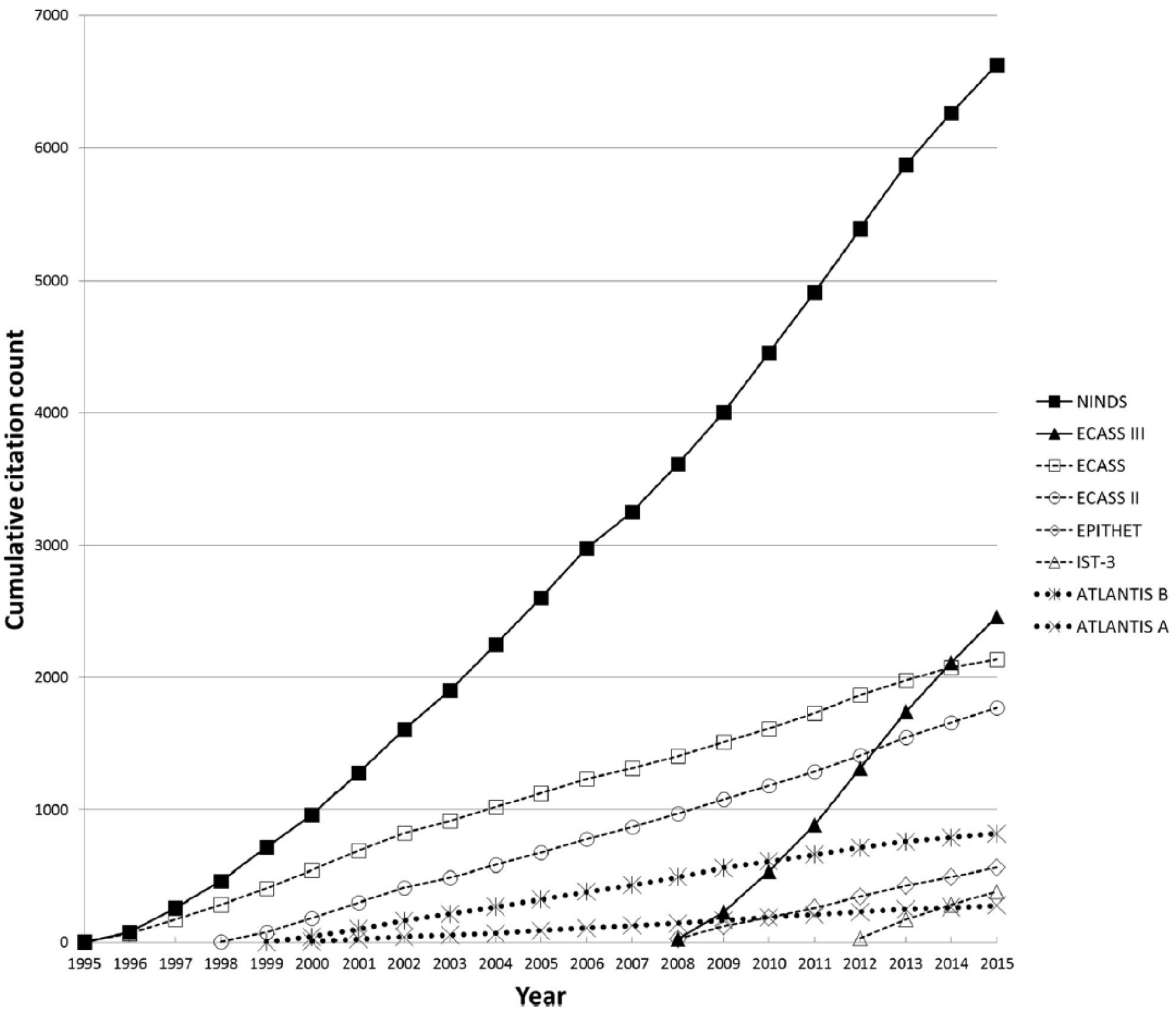
Citation inflation exists for both "positive" studies and "claim" studies in this literature.

True both within this literature (A, B) and in the wider (Web of Science) literature (C, D).



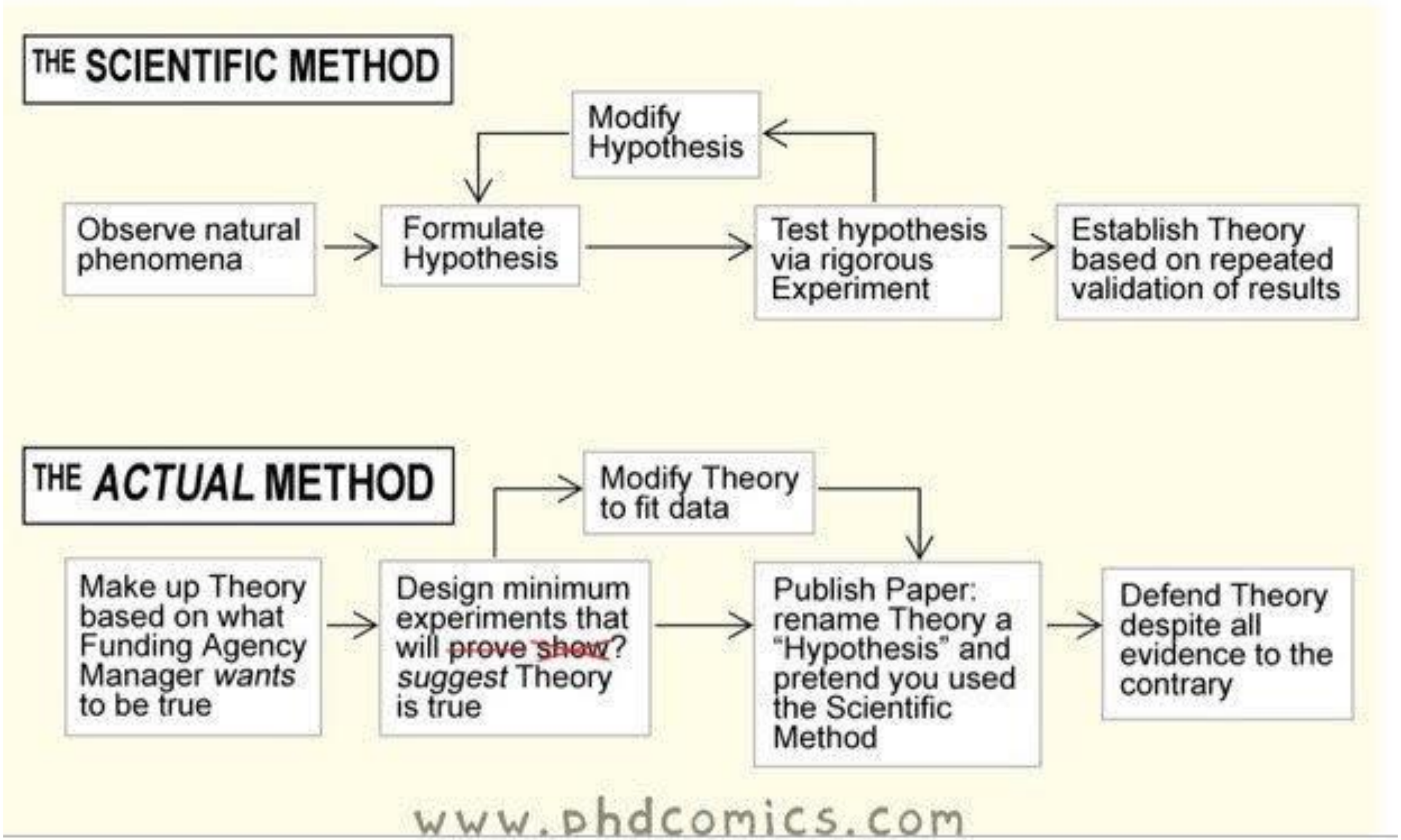
Bastiaansen et al. (2015). Biol Psychiatry, 78, e35-36.

Two positive trials, four neutral trials, two negative trials (stopped early for safety concerns).



Misemer et al. (2016). Trials, 17, 473.

Real Scientific Method



University of BRISTOL

MRC Integrative Epidemiology Unit

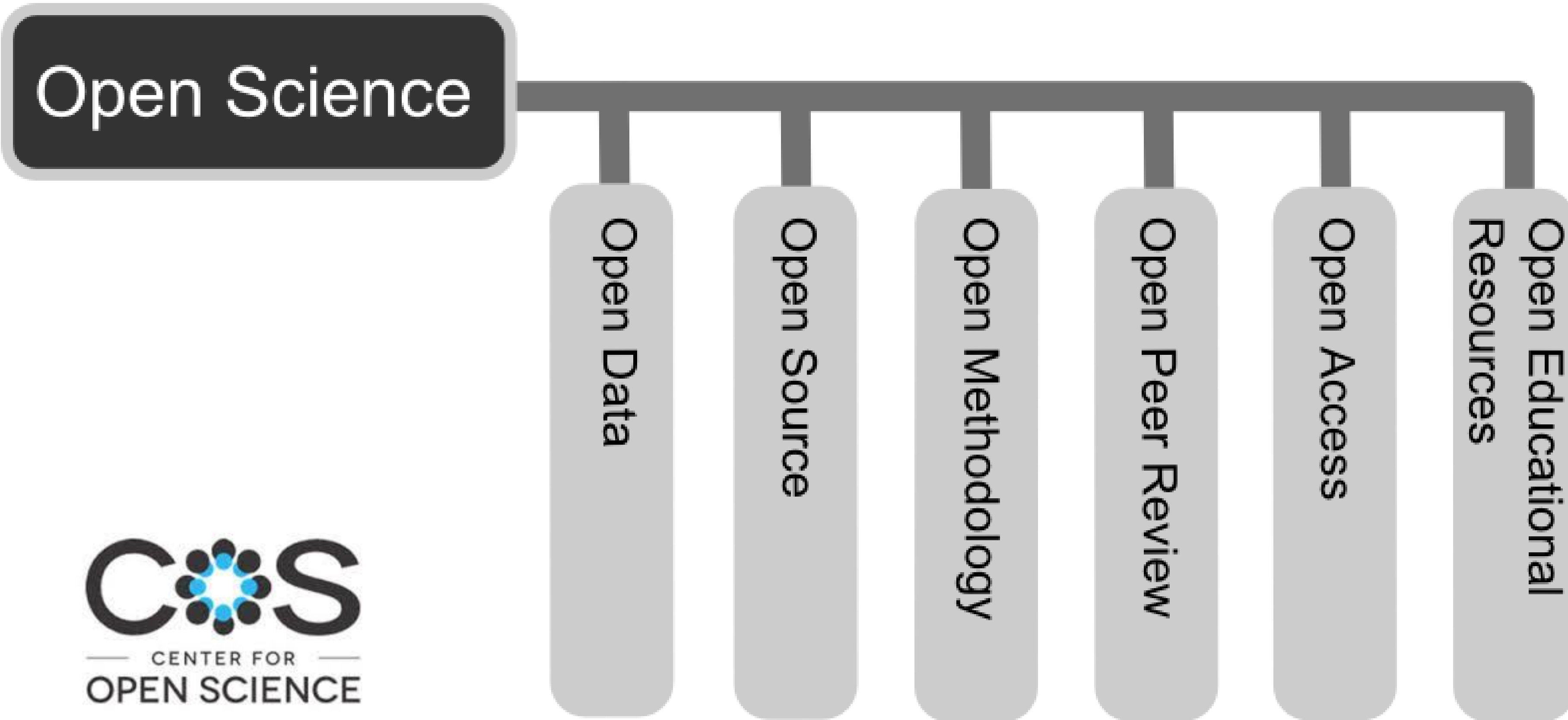
Scientific rigor and the art of motorcycle maintenance

Marcus Munafò, Simon Noble, William J Browne, Dani Brunner, Katherine Button, Joaquim Ferreira, Peter Holmans, Douglas Langbehn, Glyn Lewis, Martin Lindquist, Kate Tilling, Eric-Jan Wagenmakers & Robi Blumenstein

The reliability of scientific research is under scrutiny. A recently convened working group proposes cultural adjustments to incentivize better research practices.

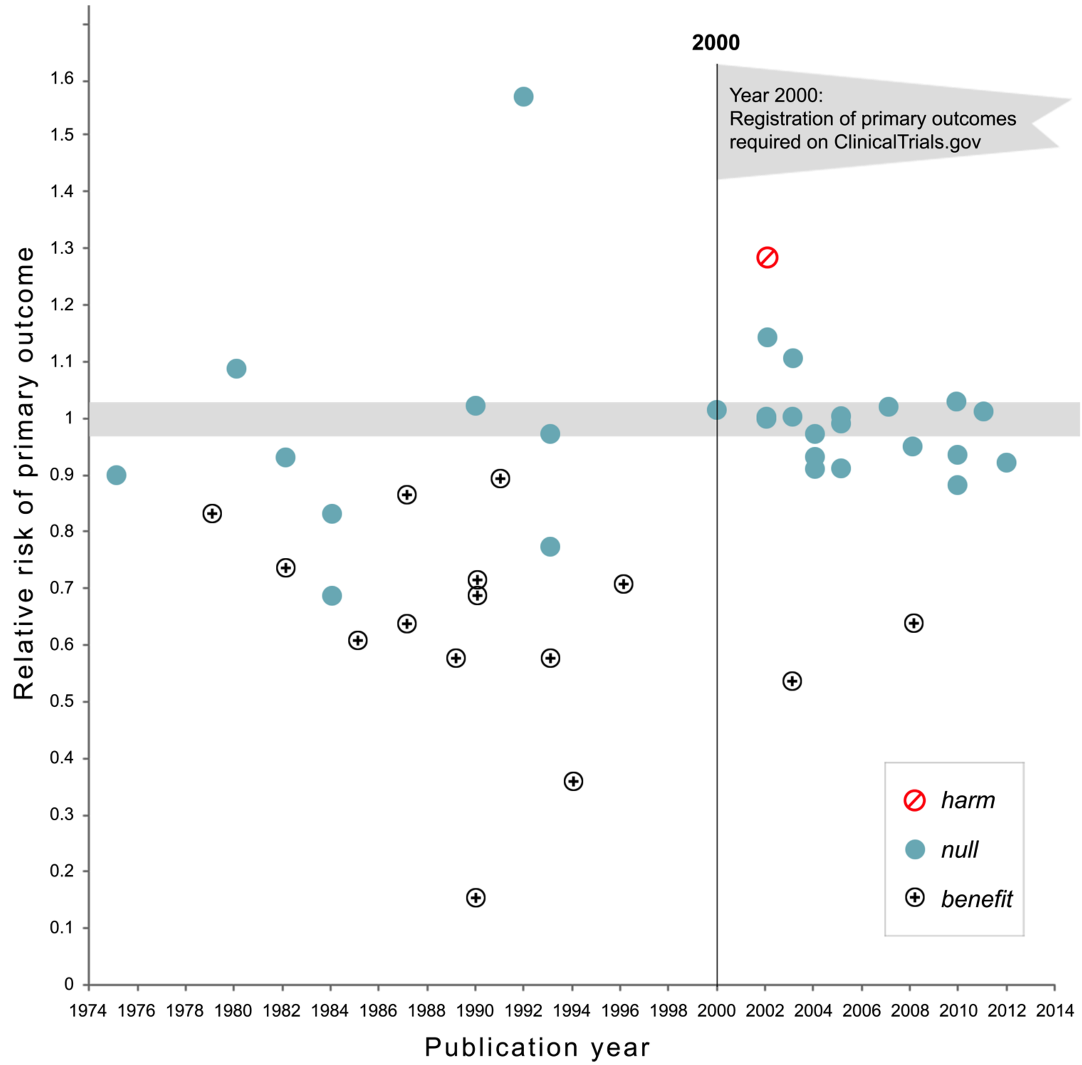
Like auto manufacturing in the 1970s, scientific research is producing too many lemons.

Munafò et al. (2014), Nat Biotech, 32, 871-873.



CENTER FOR OPEN SCIENCE

In 2000 the National Heart Lung, and Blood Institute required the registration of primary outcome on ClinicalTrials.gov for all their grant-funded activity

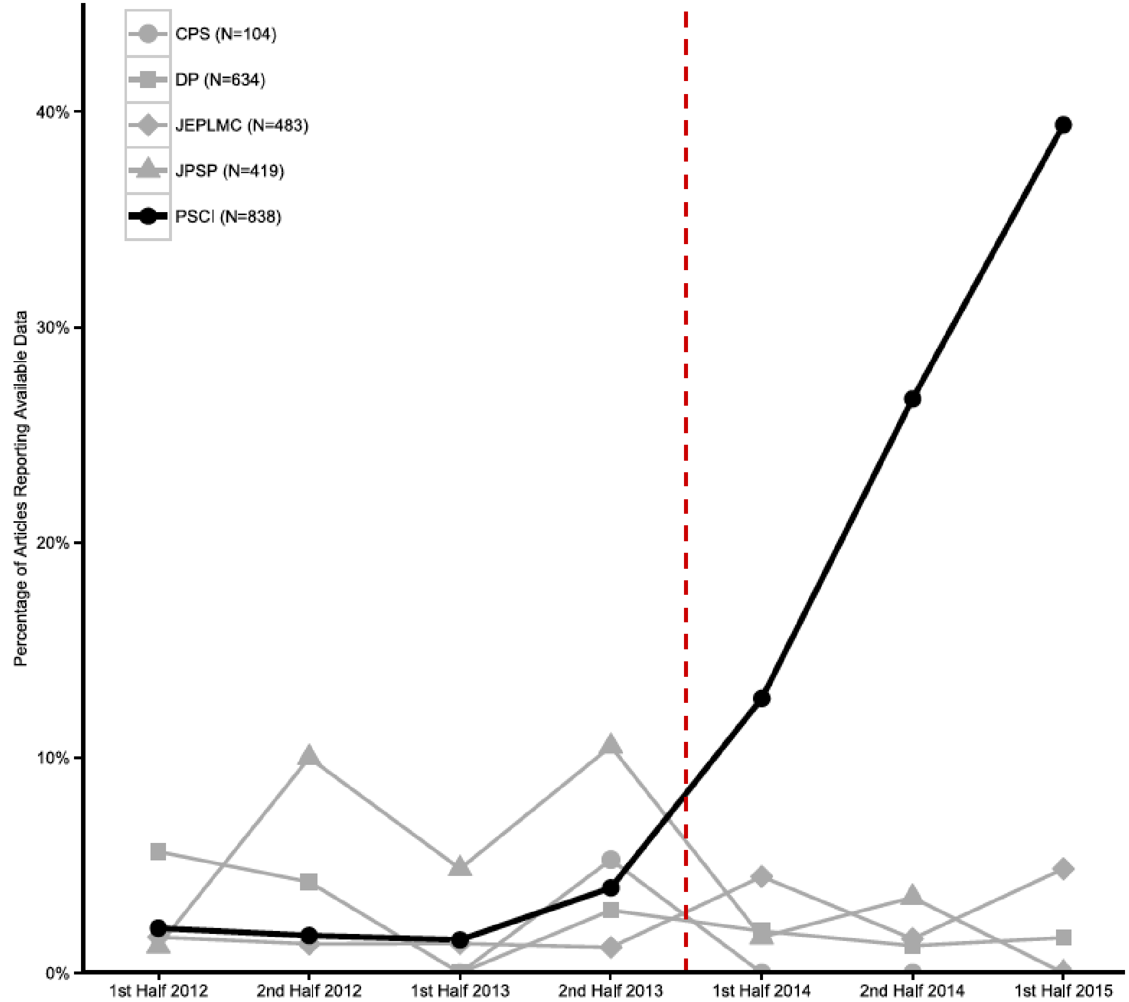


Kaplan & Irvin (2015). PLoS One, 10, e0132382.

University of BRISTOL

MRC Integrative Epidemiology Unit

Introduction of badges for open practices at *Psychological Science* followed by a steep increased in data sharing.



Kidwell et al. (2016). PLoS Biology, 14, e1002456.

University of BRISTOL

MRC Integrative Epidemiology Unit

A manifesto for reproducible science

Marcus R. Munafò[1,2]*, Brian A. Nosek[3,4], Dorothy V. M. Bishop[5], Katherine S. Button[6], Christopher D. Chambers[7], Nathalie Percie du Sert[8], Uri Simonsohn[9], Eric-Jan Wagenmakers[10], Jennifer J. Ware[11] and John P. A. Ioannidis[12,13,14]

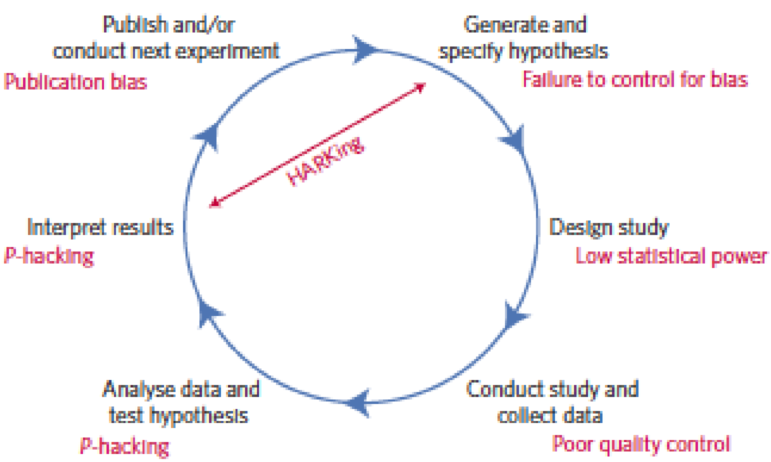


Table 1 | A manifesto for reproducible science.

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.

Munafò et al. (2017). Nat Hum Behav, 1, 0021.

University of BRISTOL

MRC Integrative Epidemiology Unit

UK Reproducibility Network

Understand factors that contribute to poor research reproducibility

Provide training and disseminate best practice

Support and test interventions to improve reproducibility

Ensure coordination with stakeholders

- Launched March 2019
- Local network leads at >40 UK institutions
- Supported by a range of stakeholders

Marcus Munafò | Laura Fortunato | Malcolm MacLeod | Alex Collins | Chris Chambers

Acknowledgements

marcus.munafo@bristol.ac.uk

@MarcusMunafo

@BristolTARG

@UKRepro

http://www.bristol.ac.uk/expsych/research/brain/targ

https://www.bristol.ac.uk/psychology/research/ukrn/

Tobacco and Alcohol Research Group:

Angela Attwood Senior Lecturer

Anna Blackwell	Postdoc
Claire Braboszcz	Postdoc
Laura Brocklebank	Postdoc
Sarah Burrows-Weeks	Administrator
Katie De-Loyde	Research Assistant
Katie Drax	PhD Student
Maddy Dyer	PhD Student
Kayleigh Easey	PhD Student
Andy Eastwood	PhD Student
Jenn Ferrar	Postdoc
Will Gawned	Administrator
Elis Haan	PhD Student
Abigail Jackson	Postdoc
Andy Keavey	Software Developer
Jasmine Khouja	PhD Student
Rebecca Lawn	PhD Student
Liam Mahedy	Postdoc
Osama Mahmoud	Postdoc
Joe Matthews	Research Assistant
Olivia Maynard	Lecturer
Osama Mahmoud	Postdoc
Vicky Rice	PhD Student
Laura Schellhas	PhD Student
Hannah Sallis	Postdoc
Carlos Sillero	Research Assistant
Andy Skinner	Postdoc
Caroline Skirrow	Postdoc
Chris Stone	Research Assistant
Steph Suddell	PhD Student
Jackie Thompson	Postdoc
Sophie Turnbull	Postdoc
Robyn Wootton	Postdoc

University of BRISTOL

MRC Integrative Epidemiology Unit