SCIENTIFIC ECOSYSTEMS AND RESEARCH REPRODUCIBILITY

Marcus Munafò
The Reproducibility Crisis

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

Aaron Mobley1, Suzanne K. Linder3, Russell Braeuer1, Lee M. Ellis1,2,*, Leonard Zwelling4*

1Department of Cancer Biology, The University of Texas MD Anderson Cancer Center, Houston, Texas, United States of America. 2Department of General Internal Medicine, The University of Texas MD Anderson Cancer Center, Houston, Texas, United States of America. 3Department of Surgical Oncology, The University of Texas MD Anderson Cancer Center, Houston, Texas, United States of America. 4Department 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 Schilange and Khursru Asadullah

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

Open Science Collaboration1

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.
Questionable Practices

(i) Limbo
(ii) Overselling
(iii) Post-hoc storytelling
(iv) P-value fishing
(v) Creative outliers
(vi) Plagiarism
(viii) Non-publication
(viii) Partial publication
(ix) Falsification

Questionable Practices

http://blogs.discovermagazine.com/neuroskeptic/2013/10/16/the-f-problem
Using the same method as in Study 1, we asked 2034 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 ($M$s = 20.3 and 21.2, respectively), $F(1, 18) = 1.01, p = .33$.
“…nearly as many unique analysis pipelines as there were studies in the sample…”

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

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

But Science is Self-Correcting!

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

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

But Science is Self-Correcting!

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\textsuperscript{a}, John P.A. Ioannidis\textsuperscript{b,c,d,e,*}

But Science is Self-Correcting!

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.

But Science is Self-Correcting!

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

Crisis or Opportunity?

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

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

A manifesto for reproducible science

Marcus R. Munafò1,2*, Brian A. Nosek3,4, Dorothy V. M. Bishop5, Katherine S. Button6, Christopher D. Chambers7, Nathalie Percie du Sert8, Uri Simonsohn9, Eric-Jan Wagenmakers10, Jennifer J. Ware11 and John P. A. Ioannidis12,13,14

Table 1 | A manifesto for reproducible science.

<table>
<thead>
<tr>
<th>Theme</th>
<th>Proposal</th>
<th>Examples of initiatives/potential solutions (extent of current adoption)</th>
<th>Stakeholder(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Methods</td>
<td>Protecting against cognitive biases</td>
<td>All of the initiatives listed below (C to ****)</td>
<td>J,F</td>
</tr>
<tr>
<td></td>
<td>Improving methodological training</td>
<td>Rigorous training in statistics and research methods for future researchers (C)</td>
<td>J,F</td>
</tr>
<tr>
<td></td>
<td>Independent methodological support</td>
<td>Involvement of methodologists in research (C)</td>
<td>J,F</td>
</tr>
<tr>
<td></td>
<td>Collaboration and team science</td>
<td>Multi-site studies/distributed data collection (C)</td>
<td>J,F</td>
</tr>
<tr>
<td></td>
<td>Reporting and dissemination</td>
<td>Promoting study pre-registration</td>
<td>J,F</td>
</tr>
<tr>
<td></td>
<td>Improving the quality of reporting</td>
<td>Open Science Framework (C)</td>
<td>J,F</td>
</tr>
<tr>
<td></td>
<td>Protecting against conflicts of interest</td>
<td>Disclosure of conflicts of interest (C)</td>
<td>J,F</td>
</tr>
<tr>
<td>Reproducibility</td>
<td>Encouraging transparency and open science</td>
<td>Open data, materials, software and so on (C to ****)</td>
<td>J,F,R</td>
</tr>
<tr>
<td>Evaluation</td>
<td>Diversifying peer review</td>
<td>Peer reviews (C)</td>
<td>J,F,R</td>
</tr>
<tr>
<td>Incentives</td>
<td>Rewarding open and reproducible practices</td>
<td>Pre-registration (**** for clinical trials, * for other studies)</td>
<td>J,F,R</td>
</tr>
</tbody>
</table>

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

marcus.munafo@bristol.ac.uk
@MarcusMunafo
@BristolTARG
http://www.bristol.ac.uk/expsych/research/brain/targ/

Tobacco and Alcohol Research Group:

Olivia Abrams
Angela Attwood
Anna Blackwell
Alex Board
Emily Crowe
Maddy Dyer
Kayleigh Easey
Andy Eastwood
Meg Fluharty
Suzi Gage
Meryem Grabski
Eleanor Kennedy
Jasmine Khouja
Glenda Lassi
Rebecca Lawn
Jim Lumsden
Olivia Maynard
Hannah Sallis
Carlos Sillero
Andy Skinner
Chris Stone
Amy Taylor
Gemma Taylor
David Troy

Research Assistant
Postdoc
Research Assistant
Research Assistant
PhD Student
PhD Student
PhD Student
PhD Student
PhD Student
PhD Student
PhD Student
PhD Student
PhD Student
PhD Student
Postdoc
Postdoc
PhD Student
PhD Student
PhD Student
Postdoc
Postdoc
Research Assistant
Postdoc
Postdoc
Postdoc
PhD Student

UKCTAS
UK Centre for Tobacco & Alcohol Studies

University of BRISTOL

MRC Integrative Epidemiology Unit
Scientific Ecosystems

LETTER

Modelling the effects of subjective and objective decision making in scientific peer review

In-Uck Park¹,², Mike W. Peacey³,⁵ & Marcus R. Munafò⁴,⁵,⁶

PERSPECTIVE

Current Incentives for Scientists Lead to Underpowered Studies with Erroneous Conclusions

Andrew D. Higginson¹ *, Marcus R. Munafò²,³ *

¹ Centre for Research in Animal Behaviour, College of Life and Environmental Sciences, University of Exeter, Exeter, United Kingdom, 2. MRC Integrative Epidemiology Unit (IEU) at the University of Bristol, Bristol, United Kingdom, 3. UK Centre for Tobacco and Alcohol Studies, School of Experimental Psychology, University of Bristol, Bristol, United Kingdom

* a.higginson@exeter.ac.uk (ADH); marcus.munafo@bristol.ac.uk (MRM)

The natural selection of bad science

Paul E. Smaldino¹ and Richard McElreath²

¹Cognitive and Information Sciences, University of California, Merced, CA 95343, USA
²Department of Human Behavior, Ecology, and Culture, Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany

PES, 0000-0002-7133-5620; RME, 0000-0002-0367-5377
Incentive Structures

ANALYSIS

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

Katherine S. Button1,2, John P. A. Ioannidis3, Claire Mokrysz4, Brian A. Nosek4, Jonathan Flint5, Emma S. J. Robinson6 and Marcus R. Munafò7

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.

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…

Incentive Structures

US studies may overestimate effect sizes in softer research

Daniele Fanelli\textsuperscript{a,1} and John P. A. Ioannidis\textsuperscript{b,c,d}

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

**Incentive Structures**

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Nature, Science</td>
<td>$26,212</td>
<td>$26,006</td>
<td>$25,781</td>
<td>$25,365</td>
<td>$33,990</td>
<td>$36,658</td>
<td>$38,908</td>
<td>$43,783</td>
<td>$43,783</td>
</tr>
<tr>
<td>PLOS One</td>
<td>$1,096</td>
<td>$1,086</td>
<td>$1,035</td>
<td>$994</td>
<td>$991</td>
<td>$915</td>
<td>$941</td>
<td>$984</td>
<td>$984</td>
</tr>
<tr>
<td>MIS Quarterly</td>
<td>$2,613</td>
<td>$2,570</td>
<td>$2,553</td>
<td>$2,654</td>
<td>$2,876</td>
<td>$2,861</td>
<td>$2,992</td>
<td>$2,938</td>
<td>$2,938</td>
</tr>
<tr>
<td>JASIST</td>
<td>$1,737</td>
<td>$1,758</td>
<td>$1,741</td>
<td>$1,887</td>
<td>$2,066</td>
<td>$2,303</td>
<td>$2,435</td>
<td>$2,488</td>
<td>$2,488</td>
</tr>
<tr>
<td>Journal of Documentation</td>
<td>$1,082</td>
<td>$1,087</td>
<td>$1,042</td>
<td>$1,111</td>
<td>$1,167</td>
<td>$1,265</td>
<td>$1,329</td>
<td>$1,408</td>
<td>$1,408</td>
</tr>
<tr>
<td>Library Hi Tech</td>
<td>$781</td>
<td>$775</td>
<td>$726</td>
<td>$741</td>
<td>$740</td>
<td>$768</td>
<td>$795</td>
<td>$783</td>
<td>$783</td>
</tr>
<tr>
<td>LIBRI</td>
<td>$650</td>
<td>$644</td>
<td>$577</td>
<td>$560</td>
<td>$538</td>
<td>$509</td>
<td>$517</td>
<td>$484</td>
<td>$484</td>
</tr>
</tbody>
</table>

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