What’s Open Science?
Why Open Science?

- Collaboration
- Citations
- Reuse
- Impact
- Transparency
- Validations
- Accountability
Empirically analyzing empirical evidence

One of the central goals in any scientific endeavor is to understand causality. Experiments that seek to demonstrate a cause/effect relation most often manipulate the postulated causal factor. Aarts et al. describe the replication of 100 experiments reported in papers published in 2008 in three high-ranking psychology journals. Assessing whether the replication and the original experiment yielded the same result according to several criteria, they find that about one-third to one-half of the original findings were also observed in the replication study.

Science, this issue 10.1126/science.aac4716
Evaluating the replicability of social science experiments in *Nature* and *Science* between 2010 and 2015

Colin F. Camerer, Anna Dreber, Felix Holzmeister, Teck-Hua Ho, Jürgen Huber, Magnus Johannesson, Michael Kirchler, Gideon Nave, Brian A. Nosek, Thomas Pfeiffer, Adam Altmejd, Nick Buttrick, Taizan Chan, Yiling Chen, Eskil Forsell, Anup Gampa, Emma Heikensten, Lily Hummer, Taisuke Imai, Siri Isaksson, Dylan Manfredi, Julia Rose, Eric-Jan Wagenmakers & Hang Wu


We find a significant effect in the same direction as the original study for 13 (62%) studies, and the effect size of the replications is on average about 50% of the original effect size. Replicability varies...
Correspondence

*Nature Reviews Drug Discovery* **10**, 712 (September 2011) | doi:10.1038/nrd3439-c1

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

See also: [News and Analysis by Arrowsmith](#)

Florian Prinz¹, Thomas Schlange² & Khusru Asadullah³

![Pie chart showing data]

- **43 (65%)** Inconsistencies
- **14 (21%)** Not applicable
- **5 (7%)** Literature data are in line with in-house data
- **2 (3%)** Main data set was reproducible
- **3 (4%)** Some results were reproducible
Why do we care?
Ethics

• Waste of animal lives

• Waste of patients’ time, hope and lives

• Waste of money (up to 85% of research funding, Chalmers and Glasziou, Lancet 2009

https://www.thelancet.com/journals/lancet/article/PIIS0140-6736(09)60329-9)
Trust in evidence

Donald J. Trump
@realDonaldTrump

Healthy young child goes to doctor, gets pumped with massive shot of many vaccines. AUTISM

Donald J. Trump
@realDonaldTrump

The concept of global warming was created by and for the Chinese in order to make U.S. manufacturing non-competitive.
Five selfish reasons to work reproducibly

Florian Markowetz
How to work reproducibly?
Good Research Practice

Publishing

Hypotheses

Design

Interpretation

Data analysis

Data collection

All images on CC0 license from https://pixabay.com/
Ask the right question

• Do a proper literature review, look for systematic reviews and meta-analyses, where are the gaps?

• Talk to your colleagues

• Speak to clinicians

• Involve patients and the public e.g. James Lind Alliance, social media, Citizen Science projects

=> Be open and collaborate
Good Research Practice

Hypotheses

Design
Do you need to collect new data?
Study design

Best Practices in Data Analysis and Sharing in Neuroimaging using MRI

Reporting guideline provided for?
(i.e. exactly what the authors state in the paper)
Reporting in neuroimaging using magnetic resonance imaging (MRI).

Full bibliographic reference

Language
English

Relevant URLs
Full-text access to this reporting guideline and available checklists is available at:
http://www.humanbrainmapping.org/COBIDASreport

http://www.humanbrainmapping.org/COBIDAS

http://www.equator-network.org/
Study design

Bias?

Confounder

Cause

Effect

Chance?

All images on CC0 license from https://pixabay.com/
Statistical power and sample size

Analysis

Nature Reviews Neuroscience 14, 365-376 (May 2013) | doi:10.1038/nrn3475
Corrected online: 15 April 2013

There is an Erratum (May 1 2013) associated with this article.

Power failure: why small sample size undermines the reliability of neuroscience
See also: Correspondence by Quilan | Correspondence by Ashton | Correspondence by Bacchetti | Author's reply by Button et al.

Katherine S. Button1,2, John P. A. Ioannidis3, Claire Mokrysz1, Brian A. Nosek4, Jonathan Flint5, Emma S. J. Robinson6 & Marcus R. Munafò1. About the authors
Statistical power

• Median statistical power in neuroscience: 21%
• Median statistical power in animal studies: between 18 – 31%
• Median statistical power in neuroimaging: 8%

=> chance of false negative = 1-power

=> “the likelihood that any nominally significant finding actually reflects a true effect is small” (low positive predictive value)
Sample size calculation - example

- Effect size = Cohen’s d of 0.5 (medium effect)
- Statistical power to find effect = 90%
- alpha = 0.05
- One sample one-tailed t test
  ⇒ 36 participants
- Independent groups two-tailed t test
  ⇒ 86 participants per group
Pre-register

https://osf.io/tvyxz/wiki/home/
Pre-registration

• Write up your hypothesis, study design and detailed analysis pipeline (introduction and methods)

• Pre-register it online (e.g. on OSF) or use registered report

• Registered report means your study will be published IRRESPECTIVE of results, purely based on scientific merit

=> you get feedback before you collect the data

=> you can put it on your CV before you have finished the study

• More info on pre-registration: https://tomstafford.staff.shef.ac.uk/?p=573 and registered reports: https://cos.io/rr/
Pre-registration

# Pre-registration of analysis pipeline

Pre-registration of analysis pipeline

<table>
<thead>
<tr>
<th>Processing step</th>
<th>Reason</th>
<th>Options [suboptions]</th>
<th>Number of plausible options</th>
</tr>
</thead>
<tbody>
<tr>
<td>Motion correction</td>
<td>Correct for head motion during scanning</td>
<td>• 'Interpolation' [linear or sinc]</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• 'Reference volume' [single or mean]</td>
<td></td>
</tr>
<tr>
<td>Slice timing correction</td>
<td>Correct for differences in acquisition timing of different slices</td>
<td>'No', 'before motion correction'</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td></td>
<td>or 'after motion correction'</td>
<td></td>
</tr>
<tr>
<td>Field map correction</td>
<td>Correct for distortion owing to magnetic susceptibility</td>
<td>'Yes' or 'no'</td>
<td>2</td>
</tr>
<tr>
<td>Spatial smoothing</td>
<td>Increase SNR for larger activations and ensure assumptions of GRF theory</td>
<td>‘FWHM’ [4 mm, 6 mm or 8 mm]</td>
<td>3</td>
</tr>
<tr>
<td>Spatial normalization</td>
<td>Warps an individual brain to match a group template</td>
<td>‘Method’ [linear or nonlinear]</td>
<td>2</td>
</tr>
<tr>
<td>High-pass filter</td>
<td>Remove low-frequency nuisance signals from data</td>
<td>‘Frequency cut-off’ [100 s or 120 s]</td>
<td>2</td>
</tr>
<tr>
<td>Head motion regressors</td>
<td>Remove remaining signals owing to head motion via statistical model</td>
<td>'Yes' or 'no' [if yes: 6/12/24 parameters or single time point 'scrubbing' regressors]</td>
<td>5</td>
</tr>
<tr>
<td>Haemodynamic response</td>
<td>Account for delayed nature of haemodynamic response to neuronal activity</td>
<td>• 'Basis function' ['single-gamma' or 'double-gamma']</td>
<td>6</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• 'Derivatives' ['none', 'shift' or 'dispersion']</td>
<td></td>
</tr>
<tr>
<td>Temporal autocorrelation</td>
<td>Model for the temporal autocorrelation inherent in fMRI signals</td>
<td>'Yes' or 'no'</td>
<td>2</td>
</tr>
<tr>
<td>model</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Multiple-comparison</td>
<td>Correct for large number of comparisons across the brain</td>
<td>'Voxel-based GRF', 'cluster-based GRF', 'FDR' or 'non-parametric'</td>
<td>4</td>
</tr>
<tr>
<td>correction</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total possible workflows</td>
<td></td>
<td></td>
<td>69,120</td>
</tr>
</tbody>
</table>

FDR, false discovery rate; FWHM, full width at half maximum; GRF, Gaussian random field; SNR, signal-to-noise ratio.

Poldrack et al., Nat Rev Neuroscience, 2017, 18:115-126
Good Research Practice

- Hypotheses
- Design
- Data collection
- Data analysis
Do you need a new method?

protocols.io

Make your science more reproducible
protocols.io is the #1 open access repository for science methods

JOIN US - IT'S FREE!
Reproducible measures

• Validity (Can I get the right answer?)
• Reliability (Can I get the same answer twice?)

Reliable  Valid  Not reliable  Not valid
Not Valid  Not reliable  Not valid  Both reliable and valid
Reproducible measures

• How reliable and valid are your tests?
  • Can you compare with gold standard or well-established tests?

• Are you doing any quality control of your tools (experimental setup, acquired data, etc.)

• Analysis pipelines
  • Use well-established tools
  • Follow good programming practice
  • Test your code using simulations
There are no miracles!

“I think you should be more explicit here in step two.”

Sidney Harris, *What’s so funny about Science?* (1977)
Reproducible workflows

<table>
<thead>
<tr>
<th>ELN software</th>
<th>Language base</th>
<th>License</th>
<th>Hosted on</th>
<th>Other info</th>
</tr>
</thead>
<tbody>
<tr>
<td>eLabFTW</td>
<td>PHP/MySQL</td>
<td>AGPL v3</td>
<td>Easy to install on your own server</td>
<td>Runs in Docker containers</td>
</tr>
<tr>
<td>OpenWetWare</td>
<td>PHP</td>
<td>GPL v2</td>
<td>MIT</td>
<td>Based on MediaWiki</td>
</tr>
<tr>
<td>sciNote[17]</td>
<td>Ruby on Rails 4.2.3</td>
<td>MPL 2.0</td>
<td>Amazon</td>
<td>JS Framework: Bootstrap 3, jQuery, DB: PostgreSQL 9.4</td>
</tr>
<tr>
<td>ELOG</td>
<td>C</td>
<td>GPL v3 or later</td>
<td>Install on a personal computer or server</td>
<td>Developed at PSI for the MEG experiment. It can be used to create both personal and common logbooks.</td>
</tr>
<tr>
<td>Electronic Collaboration Logbook (ECL)</td>
<td>python 2.7 apache PostgreSQL Django</td>
<td>BSD?</td>
<td>install on a server</td>
<td>Developed at Fermilab</td>
</tr>
<tr>
<td>Jupyter</td>
<td>Python plus over 40 supported languages</td>
<td>BSD</td>
<td>Install on a personal computer or server</td>
<td>As opposed to the experiment documentation of an ELN, the pages created in the Jupyter Notebook act as a &quot;computational record of [an interactive computing] session, interleaving executable code with explanatory text, mathematics, and rich representations of resulting objects.&quot;[18]</td>
</tr>
</tbody>
</table>

Good statistical practice

Erroneous analyses of interactions in neuroscience: a problem of significance

Sander Nieuwenhuis, Birte U Forstmann & Eric-Jan Wagenmakers

doi:10.1038/nn.2886

Published: 26 August 2011

In theory, a comparison of two experimental effects requires a statistical test on their difference. In practice, this comparison is often based on an incorrect procedure involving two separate tests in which researchers conclude that effects differ when one effect is significant ($P < 0.05$) but the other is not ($P > 0.05$). We reviewed 513 behavioral, systems and cognitive neuroscience articles in five top-ranking journals (*Science, Nature, Nature Neuroscience, Neuron* and *The Journal of Neuroscience*) and found that 78 used the correct procedure and 79 used the incorrect procedure. An additional analysis suggests that incorrect analyses of interactions are even more common in cellular and molecular neuroscience. We discuss scenarios in which the erroneous procedure is particularly beguiling.
Good Research Practice

Hypotheses

Design

Data collection

Data analysis

Interpretation

Publishing
Open reporting

• Publish ALL the analyses you did (pre-registered and exploratory)

• Publish ALL results (not just “significant” ones)

• Publish according to best practice guidelines

• Use preprints
Preprints

Contradictory Results

An empirical, 21st century evaluation of phrenology

Oiwi Parker Jones, Fidel Alfaro-Almagro, Saad Jbabdi
doi: https://doi.org/10.1101/243089

• bioRxiv: https://www.biorxiv.org/
• OSF: https://osf.io/preprints/
• PsyArXiv: https://psyarxiv.com/
Open reporting

• Publish ALL the analyses you did (pre-registered and exploratory)
• Publish ALL results (not just “significant” ones)
• Publish according to best practice guidelines
• Be honest about your biases and conflicts of interest
• Use preprints
• Publish Open Access
Publish data and materials

https://osf.io/tvyxz/wiki/home/
Robust Research - summary

• Open Research
  • Data
  • Materials
  • Reporting (and pre-registration)

• Good Research Practice
  • Relevant research question
  • Robust study design
  • Reproducible measures
  • Reproducible workflows
https://en.wikipedia.org/wiki/The_Scream
Change the system

- Researchers
- Institutions
- Professional Societies
- Publishers
- Funders
- Industry
- The Public

Scientific Ecosystem
Reproducible Research Oxford

• Started initiative in September 2017
• Mainly early career researchers
• Disciplines:
  • experimental psychology
  • biomedical sciences (preclinical to clinical)
  • social sciences (archaeology, anthropology)
  • bioethics
Journal Club, @ReproducibilitT

Seminar series on Reproducibility and Open Research

Software/ Data Carpentry workshops

Berlin-Oxford summer school: https://www.bihealth.org/de/aktuell/berlin-oxford-summer-school/

Provide speakers for lab meetings

Develop skills training for DTCs, MSc programmes, etc.
Education

- Ethics of reproducible research
- Open data/ materials/ reporting/ publishing/ workflows
- Experimental design (incl. bias and confounding)
- Statistics
- Programming skills
- Critical thinking and peer review
- Pre-registration of research projects
Next: the institution

- Incentives - HR policies
- Infrastructure
- Research ethics review
- And many other plans
And the UK

The **UK Reproducibility Network (UKRN)** is a peer-led consortium that aims to ensure the UK retains its place as a centre for world-leading research, by investigating the factors that contribute to robust research, providing training and disseminating best practice, and working with stakeholders to ensure coordination of efforts across the sector.
Acknowledgements

Reproducible Research Oxford (core people):
Dorothy Bishop, Laura Fortunato, David Gavaghan, Amy Orben, Sam Parsons, Thees Spreckelsen, Jackie Thompson

Chris Chambers (Cardiff, UKRN)
Marcus Munafò (Bristol, UKRN)
Uli Dirnagl and the QUEST team (Berlin)
Thanks for your attention!

Get in touch:
verena.heise@ndph.ox.ac.uk