The Reproducibility Manifesto

Catriona Fennell, Director Publishing Services, Elsevier
Researcher to Reader, February 25th, 2020
Evidence of a Reproducibility crisis

2005: **Why Most Published Research Findings Are False**

John P. A. Ioannidis

Published: August 30, 2005 • https://doi.org/10.1371/journal.pmed.0020124

2011: **Bayer’s** experimental data does not match literature in 65% of 67 target-validation projects. Results initially confidential.

2012: Amgen was unable to reproduce findings in 47 of 53 “landmark” cancer papers. Results initially confidential.

2015: Reproducibility Initiative found that 97% (vs. 36%) of original studies (vs. replication attempts) showed statistically significant effects

2016: **Nature survey** of 1576 researchers: 52% agree that there is a **significant** crisis; 38% say there is a **slight** crisis

2018: **D Fanelli**: “Is science really facing a reproducibility crisis, and do we need it to?”
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) | Cite this article

41k Accesses | 595 Citations | 2587 Altmetric | Metrics
<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><strong>Methods</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Protecting against cognitive biases</td>
<td>All of the initiatives listed below (* to ****)</td>
<td>J, F</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Blinding (**)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Improving methodological training</td>
<td>Rigorous training in statistics and research methods for future researchers (*)</td>
<td>I, F</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Rigorous continuing education in statistics and methods for researchers (*)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Independent methodological support</td>
<td>Involvement of methodologists in research (***)</td>
<td>F</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Independent oversight (*)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Collaboration and team science</td>
<td>Multi-site studies/distributed data collection (*)</td>
<td>I, F</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Team-science consortia (*)</td>
<td></td>
</tr>
<tr>
<td><strong>Reporting and dissemination</strong></td>
<td>Promoting study pre-registration</td>
<td>Registered Reports (*)</td>
<td>J, F</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Open Science Framework (*)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Improving the quality of reporting</td>
<td>Use of reporting checklists (**)</td>
<td>J</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Protocol checklists (*)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Protecting against conflicts of interest</td>
<td>Disclosure of conflicts of interest (****)</td>
<td>J</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Exclusion/containment of financial and non-financial conflicts of interest (*)</td>
<td></td>
</tr>
<tr>
<td><strong>Reproducibility</strong></td>
<td>Encouraging transparency and open science</td>
<td>Open data, materials, software and so on (* to **)</td>
<td>J, F, R</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Pre-registration (**** for clinical trials, * for other studies)</td>
<td></td>
</tr>
<tr>
<td><strong>Evaluation</strong></td>
<td>Diversifying peer review</td>
<td>Preprints (* in biomedical/behavioural sciences, **** in physical sciences)</td>
<td>J</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Pre- and post-publication peer review, for example, Publons, PubMed Commons (*)</td>
<td></td>
</tr>
<tr>
<td><strong>Incentives</strong></td>
<td>Rewarding open and reproducible practices</td>
<td>Badges (*)</td>
<td>J, I, F</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Registered Reports (*)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Transparency and Openness Promotion guidelines (*)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Funding replication studies (*)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Open science practices in hiring and promotion (*)</td>
<td></td>
</tr>
</tbody>
</table>

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

- Reproducibility means something different in every field: even its very definition!
- Lack of **incentives** for open science
- **Methods** lacking in transparency/rigour
- Difficult to **validate** reproducibility during peer review
- Insufficient rigour in **statistics**
- Researchers ambivalent about **sharing data**

---

*Munafo et al 2017 Nature Human Behaviour 21*
*C. Neylon 2016 Science, 349(6251)*
*R. Giner-Sorolla, Jnl of Exper Soc Psych*
Providing incentives
Solutions to the Incentive challenge

- Invest in diverse & innovative journal types
- Offer new article types that provide a ‘reward’ for sharing of data and software
- Actively invite replication studies from established ‘mainstream’ journals
- Provide better CRedit for specialised contributions

New journal will publish results from all scientifically sound clinical trials

Why publishing negative and “less impactful” results and methods is important for researchers and patients

By Paul-André Genest, PhD    Posted on 26 June 2015

www.elsevier.com/connect/how-elsevier-is-breaking-down-barriers-to-reproducibilitys-to-reproducibility
Solutions to the Incentive challenge

- Offer researchers an easy channel to publish their research output, receive credit, and make research objects discoverable
- Specialized, fast and transparent peer-review process
- Co-submission service via regular journals for data articles
- Primarily Open Access, lower APCs
Researcher attitudes to replication studies

➢ There is a disconnect between generic drivers and action: safeguarding scientific knowledge is not a concrete driver for individual researchers

Individual researchers need concrete drivers to conduct replication studies such as:

**New experiments:** To prepare starting materials for future studies

**New variables:** To test different variables (e.g. a larger sample)

**Training students:** To learn how to run experiments and use different methodologies

“One may replicate certain steps of a published experiment to get to the starting point for a new experiment.” – David, Corp

“They have merit if they include at least one variable that is changed and leads to new perspective.” – Branka, A&G
“Unless the original study appears to be flawed in some important way, why dedicate the time and resources to an experimental program which appears to be well executed and well interpreted?”

“Replication studies in my opinion are becoming more and more important, especially since there is a big pressure from the different journals to only publish new results…The thing that the journals do not want to do is not to publish work that disproves [prior] work, but work that confirms [prior] work.”

“I think that publishers should explicitly say that replication is a good thing and that papers won't be rejected on the basis of this”
Reducing publication bias: negative results/replication studies

• Very few submissions received, apparently due to:
  • Lack of funding
  • Authors don’t want to take effort writing up a non-career-advancing paper
  • ECR’s don’t want to be associated with these types of papers

CALL FOR PAPERS - REPRODUCIBILITY STUDIES
Contributor Roles Taxonomy: CRediT

- Simple, open standard of 14 defined author contributions types, suitable for all subject areas
  - Conceptualization; **Data curation**; **Formal analysis**; Funding acquisition; Investigation; **Methodology**; Project administration; Resources; **Software**; Supervision; Validation; Visualization; Writing – original draft; Writing – review & editing
- Developed collaboratively by researchers, funders, publishers together with standards organisations NISO & CASRAI

**CRediT authorship contribution statement**

**Silvana Petzel-Witt**: Conceptualization, Methodology, Investigation, Project administration, Data curation, Validation, Writing - original draft. **Sylvia I. Meier**: Formal analysis, Writing - review & editing. **Manfred Schubert-Zsilavecz**: Conceptualization, Supervision, Writing - review & editing. **Stefan W. Toennes**: Conceptualization, Methodology, Formal analysis, Resources, Writing - original draft.

Giving CRediT for roles that contribute to reproducibility

• CRediT is **fairer, richer & more transparent** than traditional author lists

• More **visibility & recognition** for contributions in **methodology, statistical analysis**

• Has potential to help editors find **statistical and other specialized reviewers**

---

2012

Harvard workshop with funders, academic institutions & other publishers

2013-14

Draft CRediT definitions developed & tested with 230 authors

2015

Cell Press introduces CRediT to all journals

2016

Aries’ Editorial Manager integrates CRediT

2018

Elsevier expands CRediT to 150 further journals

2019

CRediT expanding to 1000 journals
Validation
Solutions to the Validation challenge

- Detection of fraud & ethical issues
- Innovative approaches to peer review, e.g. Registered Reports; Results Masked Reviewer; use of machine-learning
Detection of duplicated/manipulated images

• “Dr Figures” tool: Daniel Acuna, Syracuse University, focused on duplication
• HEADT, Humboldt University, funded by Elsevier. Focused on creation of training set which emerging tools can build from
• Harvard Medical School, partially sponsored by Elsevier; focused on forensic analysis
• Commercial tools:
  • Proofig
  • ImaChek from Lpixel
Registered Reports- acceptance unrelated to outcome

- Registered Reports aim to eliminate publication bias
- Journals offering RRs agree to review study protocols before experiments are conducted.
- If protocols have merit the journal commits, in advance, to publishing the outcomes.
- Authors can perform the research knowing that results will not determine article's publication.

*COS (Center for Open Science)*
diagram showing peer review steps for Registered Reports submissions
Progress since launch

• 14 Elsevier journals & 200 journals in total offer RRJs since launched by Cortex
• For the 14 Elsevier journals:
  • 45 Registered Reports accepted
  • 2018: 3% of Cortex submissions were Stage 1 Registered Reports
• Author concerns are wrt time and reduced experiment flexibility
• Most well received in psychology/neuroscience
Results Masked Review

• Launched in 2016
• Piloted in *Journal of Vocational Behavior, The Leadership Quarterly* and *Infant Behavior and Development*
• Change to the peer review process: papers sent for review without the results or discussion
• Initial decision made for publication based on question, hypothesis and methods only
• 11 journals currently offer RMRs across different publishers
Positive reaction- actual uptake is very low so far

- “This is just so awesome!!!!” – editorial board member
- “We appreciated the feedback during the formulation of the argument … it was helpful receive feedback regarding potential flaws in the analytical process.” – RMR author
- “… a step in the right direction - but … authors still have the possibility to change their hypotheses, analyses … problems could still emerge.” – RMR reviewer
- “I enjoy reviewing RMR articles … an opportunity to provide actionable feedback to authors to help them bring out the best in their work.” – RMR reviewer
- “I love the idea behind it. I just think some kinks need to be worked out. I also worry that it will involve more revisions for the author” – RMR reviewer
Potential of “AI” to aid human peer review
More transparent methodology
STAR Methods

- Launched by Cell Press in 2016 to increase rigor and reproducibility, under expansion to all Elsevier journals
- Methods in journal articles often lack sufficient detail
- Future researchers cannot replicate the work- or even the future authors themselves!

"STAR Methods are organized logically."
"STAR Methods have all the information I need."
"STAR Methods are easy to access & comprehend."
"STAR Methods are key to good science."
ML-assisted creation of Key Resources Table (KRT)

<table>
<thead>
<tr>
<th>REAGENT or RESOURCE</th>
<th>SOURCE</th>
<th>IDENTIFIER</th>
</tr>
</thead>
<tbody>
<tr>
<td>Anti-α-tubulin</td>
<td>Cell Signaling Technology</td>
<td>Cat #3217</td>
</tr>
<tr>
<td>Rabbit monoclonal anti-α-tubulin</td>
<td>Cell Signaling Technology</td>
<td>Cat #3594</td>
</tr>
<tr>
<td>Mouse monoclonal anti-Tuj1 (clone DMTA)</td>
<td>Sigma-Aldrich</td>
<td>C2192</td>
</tr>
<tr>
<td>Rabbit polyclonal anti-IMMA1</td>
<td>This paper</td>
<td>N/A</td>
</tr>
</tbody>
</table>

Labeled with secondary immunoreactions with a mixture of Alexa Fluor 594-conjugated donkey anti-rabbit IgG (for rabbit anti-HAP1) and Alexa Fluor 594-conjugated donkey anti-goat IgG (for goat anti-ChAT) or Alexa Fluor 594-conjugated donkey anti-goat IgG (for goat anti-HAP1) and anti-mouse IgG (Molecular Probes, Eugene, OR, USA; 1:500) for 3h at 20°C.
## Machine Learning generated Key Resources Table (KRT)

<table>
<thead>
<tr>
<th>Type</th>
<th>Resource</th>
<th>Source</th>
<th>Identifier</th>
<th>Context</th>
</tr>
</thead>
<tbody>
<tr>
<td>CellLine</td>
<td>human EndoC-8H1 beta-cells</td>
<td>Ravassard et al., 2011</td>
<td>N/A</td>
<td></td>
</tr>
<tr>
<td>Chemical</td>
<td>2-mercaptoethanol</td>
<td>Sigma-Aldrich</td>
<td>Cat#6369</td>
<td></td>
</tr>
<tr>
<td>Chemical</td>
<td>a-linolenic acid</td>
<td>Sigma-Aldrich</td>
<td>Cat#12376</td>
<td></td>
</tr>
<tr>
<td>Chemical</td>
<td>arachidic acid</td>
<td>Larodan AB</td>
<td>Cat#10-2000</td>
<td></td>
</tr>
<tr>
<td>Chemical</td>
<td>behenic acid</td>
<td>Larodan AB</td>
<td>Cat#10-2200</td>
<td></td>
</tr>
<tr>
<td>BSA</td>
<td>Serologicals Proteins</td>
<td>Cat#80037</td>
<td></td>
<td></td>
</tr>
<tr>
<td>BSA (Fatty acid frsc)</td>
<td>Serva</td>
<td>Cat#11945</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chemical</td>
<td>cis-11,14,17-eicosatrienoic acid</td>
<td>Larodan AB</td>
<td>Cat#10-2003</td>
<td></td>
</tr>
<tr>
<td>Chemical</td>
<td>cis-11,14-eicosadienoic acid</td>
<td>Larodan AB</td>
<td>Cat#10-2002</td>
<td></td>
</tr>
<tr>
<td>Chemical</td>
<td>cis-11-eicosenoic acid</td>
<td>Larodan AB</td>
<td>Cat#10-2001</td>
<td></td>
</tr>
<tr>
<td>Chemical</td>
<td>cis-11-octadecenoic acid</td>
<td>Larodan AB</td>
<td>Cat#10-1512</td>
<td></td>
</tr>
<tr>
<td>Chemical</td>
<td>cis-13,16,19-docosatrienoic acid</td>
<td>Larodan AB</td>
<td>Cat#10-2203</td>
<td></td>
</tr>
<tr>
<td>Chemical</td>
<td>cis-11,16-docosadienoic acid</td>
<td>Larodan AB</td>
<td>Cat#10-2000</td>
<td></td>
</tr>
</tbody>
</table>
KRT- Antibody disambiguation

All measurements were performed in triplicates and normalized against the housekeeping genes β-actin, TATA-box binding protein, and α-tubulin using the abseFLUOS data analysis software (Biozol, Züle, Belgium).

The Antibody Registry

Showing 1 - 20 results out of 79 with the query: α-tubulin

<table>
<thead>
<tr>
<th>Antibody ID</th>
<th>Antibody Name</th>
<th>Target Antigen</th>
<th>Vendor</th>
<th>Cat Num</th>
<th>Proper Citation</th>
<th>Reference</th>
<th>Clonality</th>
</tr>
</thead>
<tbody>
<tr>
<td>AB_2315517</td>
<td>Tu1 (beta III tubulin) antibody</td>
<td>A. Frankfurter, University of Virginia; Department of Biology</td>
<td>Tu1 (beta III tubulin)</td>
<td>PMID:21858821</td>
<td>unknown</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AB_2315519</td>
<td>Tu1 (beta tubulin Tu1 class III) antibody</td>
<td>A. Frankfurter, University of Virginia; Department of Biology</td>
<td>Tu1 (beta tubulin Tu1 class III)</td>
<td>PMID:17120279</td>
<td>unknown</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Improving statistical rigour

• Dedicated (human!) statistics reviewers, extend best practice from medicine

• StatCheck: developed by the META Centre, Tilburg University, being piloted by Journal of Experimental Social Psychology

• StatReviewer: checks methodology & statistical reporting, particularly medicine
Research data
Enabling data storage, access and re-use
Elsevier journals encourage & enable data sharing

- TOP-aligned policies on data sharing in >1,800 journals
- Transparent and clear to authors journal Guides for Authors
- Integrated directly into journal workflow, making it easy for authors to share data
The power of making it easy to do the right thing

% articles with (links to) data increased with 50% after full integration into submission process
The challenges

• No common definition across all fields
• Lack of *incentives* to conduct replications
• Researchers ambivalent about *sharing data*
• *Methods* lacking in transparency/rigour
• Difficult to *validate* reproducibility during peer review
• Insufficient rigour in *statistics*

Munafo et al 2017 Nature Human Behaviour 21
C. Neylon 2016 Science, 349(6251)
R. Giner-Sorolla, Jnl of Exper Soc Psych
Thank you

To: 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

Colleagues especially IJJ Aalbersberg & Jennifer Wood
And all the editors who make innovation in publishing possible.