Reproducibility & replicability in science

Perspective from a cross-disciplinary journal editor

Véronique Kiermer

Executive Editor, PLOS

Public Library of Science

NASEM Committee on Reproducibility and Replicability of Science 1st meeting | Dec 2017

PLOS

Agenda

1. Measures of R&R challenges across disciplines
2. Cross-disciplinary approaches by journals:
    1. Editorial policy interventions to improve reporting: two experiments at Nature and PLOS
    2. Open science: PLOS data availability policy as an example
3. The incentive system as an underlying issue

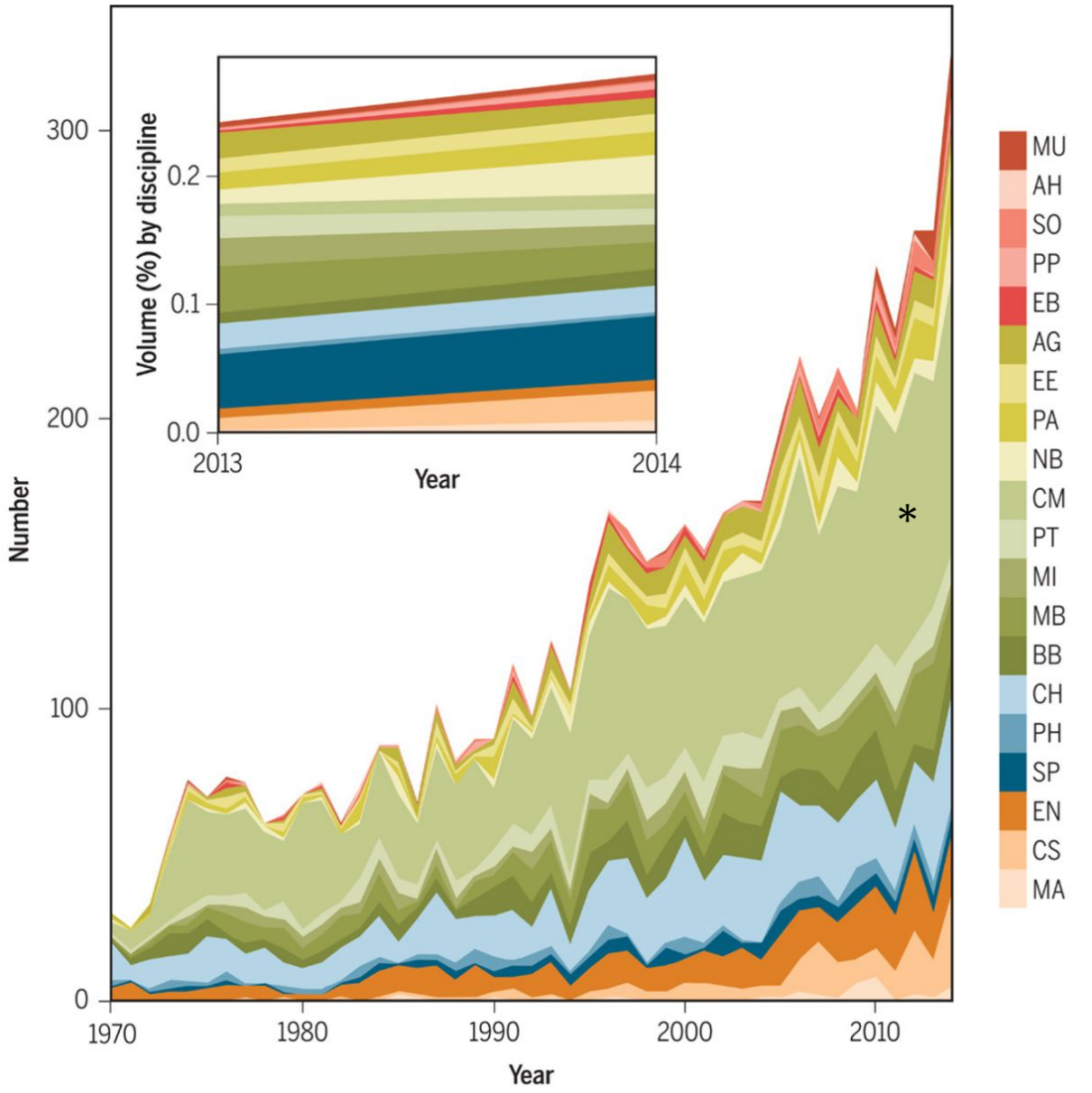
PLOS

R&R challenges and measures across disciplines and in multi-disciplinary research

- Documented evidence of R&R challenges is discipline-specific
- Increased interest in R&R in many disciplines
- Normative variations between disciplines
- Challenges of assembling editorial and peer-review expertise for multi-disciplinary studies.

PLOS

Reproducibility studies across the sciences



Multi-disciplinary

Social sciences, Psychology

Biological & Biomedical
* Clinical medicine

Chemistry,
Physics, Space sciences

Engineering,
Computational,
Mathematics

PLOS

Goodman, Fanelli, Ioannidis, Sci. Transl. Med. 2016 | doi: 10.1126/scitranslmed.aaf5027

Examples of differences that affect the approach to reproducibility in distinct scientific domains.

	Degree of determinism
★	Signal to measurement-error ratio
★	Complexity of designs and measurement tools
	Closeness of fit between hypothesis and experimental design or data
★	Statistical or analytic methods to test hypotheses
	Typical heterogeneity of experimental results
★	Culture of replication, transparency, and cumulating knowledge
	Statistical criteria for truth claims
	Purposes to which findings will be put and consequences of false conclusions

PLOS

Goodman, Fanelli, Ioannidis, Sci. Transl. Med. 2016 | doi: 10.1126/scitranslmed.aaf5027

Editorial policy interventions:
two experiments with ‘checklists’ to improve reporting

- Implementation challenges for journals
- Tension between comprehensive, specialist requirements and non-exhaustive, generalist requirements.

PLOS

Editorial policy interventions to improve reporting of *in vivo* research

- ***PLOS ONE*** mandate of the ARRIVE checklist
  - Normal editorial process
  - Randomized controlled trial
- ***Nature journals*** mandate of their own reporting checklist applicable across all life sciences
  - Part of larger policy change, increased scrutiny
  - Retrospective, control cohort study

Impact measured independently by research teams of Emily Sena and Malcolm Macleod, University of Edinburgh, CAMARADES collaboration.

PLOS

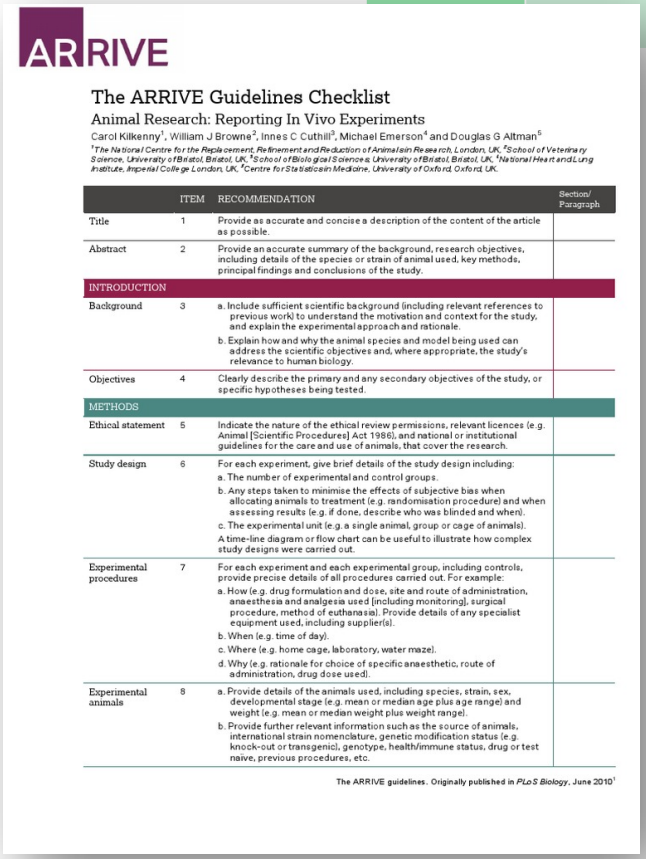
OPEN ACCESS

PERSPECTIVE

Improving Bioscience Research Reporting: The ARRIVE Guidelines for Reporting Animal Research

Carol Kilkenny, William J. Browne, Innes C. Cuthill, Michael Emerson, Douglas G. Altman

Published: June 29, 2010 • https://doi.org/10.1371/journal.pbio.1000412



ARRIVE

The ARRIVE Guidelines Checklist

Animal Research: Reporting In Vivo Experiments

Carol Kilkenny[1], William J Browne[2], Innes C Cuthill[3], Michael Emerson[4] and Douglas G Altman[5]

[1]The National Centre for the Replacement, Refinement and Reduction of Animals in Research, London, UK, [2]School of Veterinary Science, University of Bristol, Bristol, UK, [3]School of Biological Sciences, University of Bristol, Bristol, UK, [4]National Heart and Lung Institute, Imperial College London, UK, [5]Centre for Statistics in Medicine, University of Oxford, Oxford, UK.

	ITEM	RECOMMENDATION	Section/ Paragraph
Title	1	Provide as accurate and concise a description of the content of the article as possible.	
Abstract	2	Provide an accurate summary of the background, research objectives, including details of the species or strain of animal used, key methods, principal findings and conclusions of the study.	
INTRODUCTION			
Background	3	a. Include sufficient scientific background (including relevant references to previous work) to understand the motivation and context for the study, and explain the experimental approach and rationale. b. Explain how and why the animal species and model being used can address the scientific objectives and, where appropriate, the study's relevance to human biology.	
Objectives	4	Clearly describe the primary and any secondary objectives of the study, or specific hypotheses being tested.	
METHODS			
Ethical statement	5	Indicate the nature of the ethical review permissions, relevant licences (e.g. Animal [Scientific Procedures] Act 1986), and national or institutional guidelines for the care and use of animals, that cover the research.	
Study design	6	For each experiment, give brief details of the study design including: a. The number of experimental and control groups. b. Any steps taken to minimise the effects of subjective bias when allocating animals to treatment (e.g. randomisation procedure) and when assessing results (e.g. if done, describe who was blinded and when). c. The experimental unit (e.g. a single animal, group or cage of animals). A time-line diagram or flow chart can be useful to illustrate how complex study designs were carried out.	
Experimental procedures	7	For each experiment and each experimental group, including controls, provide precise details of all procedures carried out. For example: a. How (e.g. drug formulation and dose, site and route of administration, anaesthesia and analgesia used [including monitoring], surgical procedure, method of euthanasia). Provide details of any specialist equipment used, including supplier(s). b. When (e.g. time of day). c. Where (e.g. home cage, laboratory, water maze). d. Why (e.g. rationale for choice of specific anaesthetic, route of administration, drug dose used).	
Experimental animals	8	a. Provide details of the animals used, including species, strain, sex, developmental stage (e.g. mean or median age plus age range) and weight (e.g. mean or median weight plus weight range). b. Provide further relevant information such as the source of animals, international strain nomenclature, genetic modification status (e.g. knock-out or transgenic), genotype, health/immune status, drug or test naive, previous procedures, etc.	

The ARRIVE guidelines. Originally published in *PLoS Biology*, June 2010[1]

ARRIVE guidelines:

- Developed by UK NC3R
- 20-point checklist specific to *in vivo* research

Kilkenny et al, PLOS Biology, June 2010
https://doi.org/10.1371/journal.pbio.1000412

PERSPECTIVE

doi:10.1038/nature11556

A call for transparent reporting to optimize the predictive value of preclinical research

Story C. Landis[1], Susan G. Amara[2], Khusru Asadullah[3], Chris P. Austin[4], Robi Blumenstein[5], Eileen W. Bradley[6], Ronald G. Crystal[7], Robert B. Darnell[8], Robert J. Ferrante[9], Howard Fillit[10], Robert Finkelstein[1], Marc Fisher[11], Howard E. Gendelman[12], Robert M. Golub[13], John L. Goudreau[14], Robert A. Gross[15], Amelie K. Gubitz[1], Sharon E. Hesterlee[16], David W. Howells[17], John Huguenard[18], Katrina Kelner[19], Walter Koroshetz[1], Dimitri Krainc[20], Stanley E. Lazic[21], Michael S. Levine[22], Malcolm R. Macleod[23], John M. McCall[24], Richard T. Moxley III[25], Kalyani Narasimhan[26], Linda J. Noble[27], Steve Perrin[28], John D. Porter[1], Oswald Steward[29], Ellis Unger[30], Ursula Utz[1] & Shai D. Silberberg[1]

- Developed following NINDS stakeholder meeting June 2012
- 4 critical elements (aka "Landis 4")

BOX 1

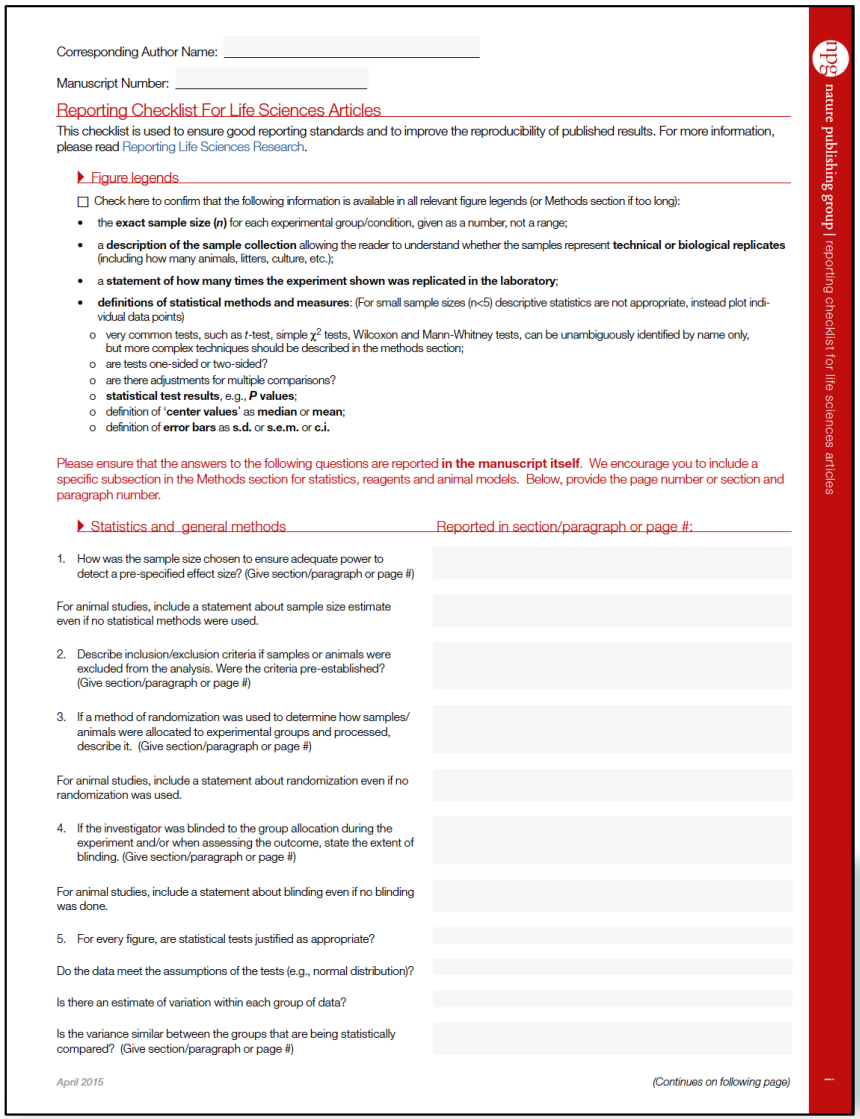
A core set of reporting standards for rigorous study design

- Randomization
- Blinding
- Sample-size estimation
- Data handling

PLOS

Landis et al., Nature. October 2012 doi: 10.1038/nature11556.

Nature journals checklist



Corresponding Author Name: _____

Manuscript Number: _____

Reporting Checklist For Life Sciences Articles

This checklist is used to ensure good reporting standards and to improve the reproducibility of published results. For more information, please read Reporting Life Sciences Research.

▸ Figure legends

☐ Check here to confirm that the following information is available in all relevant figure legends (or Methods section if too long):

- the **exact sample size (*n*)** for each experimental group/condition, given as a number, not a range;
- a **description of the sample collection** allowing the reader to understand whether the samples represent **technical or biological replicates** (including how many animals, litters, culture, etc.);
- a **statement of how many times the experiment shown was replicated in the laboratory**;
- **definitions of statistical methods and measures**: (For small sample sizes (n<5) descriptive statistics are not appropriate, instead plot individual data points)
  - o very common tests, such as *t*-test, simple χ^2 tests, Wilcoxon and Mann-Whitney tests, can be unambiguously identified by name only, but more complex techniques should be described in the methods section;
  - o are tests one-sided or two-sided?
  - o are there adjustments for multiple comparisons?
  - o **statistical test results**, e.g., ***P* values**;
  - o definition of '**center values**' as **median** or **mean**;
  - o definition of **error bars** as **s.d.** or **s.e.m.** or **c.i.**

Please ensure that the answers to the following questions are reported **in the manuscript itself**. We encourage you to include a specific subsection in the Methods section for statistics, reagents and animal models. Below, provide the page number or section and paragraph number.

▸ Statistics and general methods	Reported in section/paragraph or page #:
1. How was the sample size chosen to ensure adequate power to detect a pre-specified effect size? (Give section/paragraph or page #)	
For animal studies, include a statement about sample size estimate even if no statistical methods were used.	
2. Describe inclusion/exclusion criteria if samples or animals were excluded from the analysis. Were the criteria pre-established? (Give section/paragraph or page #)	
3. If a method of randomization was used to determine how samples/animals were allocated to experimental groups and processed, describe it. (Give section/paragraph or page #)	
For animal studies, include a statement about randomization even if no randomization was used.	
4. If the investigator was blinded to the group allocation during the experiment and/or when assessing the outcome, state the extent of blinding. (Give section/paragraph or page #)	
For animal studies, include a statement about blinding even if no blinding was done.	
5. For every figure, are statistical tests justified as appropriate?	
Do the data meet the assumptions of the tests (e.g., normal distribution)?	
Is there an estimate of variation within each group of data?	
Is the variance similar between the groups that are being statistically compared? (Give section/paragraph or page #)	

April 2015 (Continues on following page)

nature publishing group | reporting checklist for life sciences articles

1. Checklist of reporting standards
   - 18-point checklist across all life sciences
   - Including "Landis 4"
2. Eliminated length limits for methods sections
3. Increased scrutiny of statistics
4. Re-emphasized data sharing

nature.com/authors/policies/checklist.pdf

PLOS

International Congress on
Peer Review and Scientific Publication

Enhancing the quality and credibility of science

September 10-12, 2017 | Chicago, USA

Impact of an Intervention to Improve Compliance With the ARRIVE Guidelines for the Reporting of In Vivo Animal Research

Emily Sena,[1] for the Intervention to Improve Compliance With the ARRIVE Guidelines (IICARus) Collaborative Group

PLOS ONE submissions
n = 845 intervention ; n = 844 control
March 2015 – June 2015
Authors, academic editors, and peer reviewers were blinded to the study and allocation

- **No published article achieved full compliance**
- **Significant improvement in reporting of only 2 sub-items**

New Results

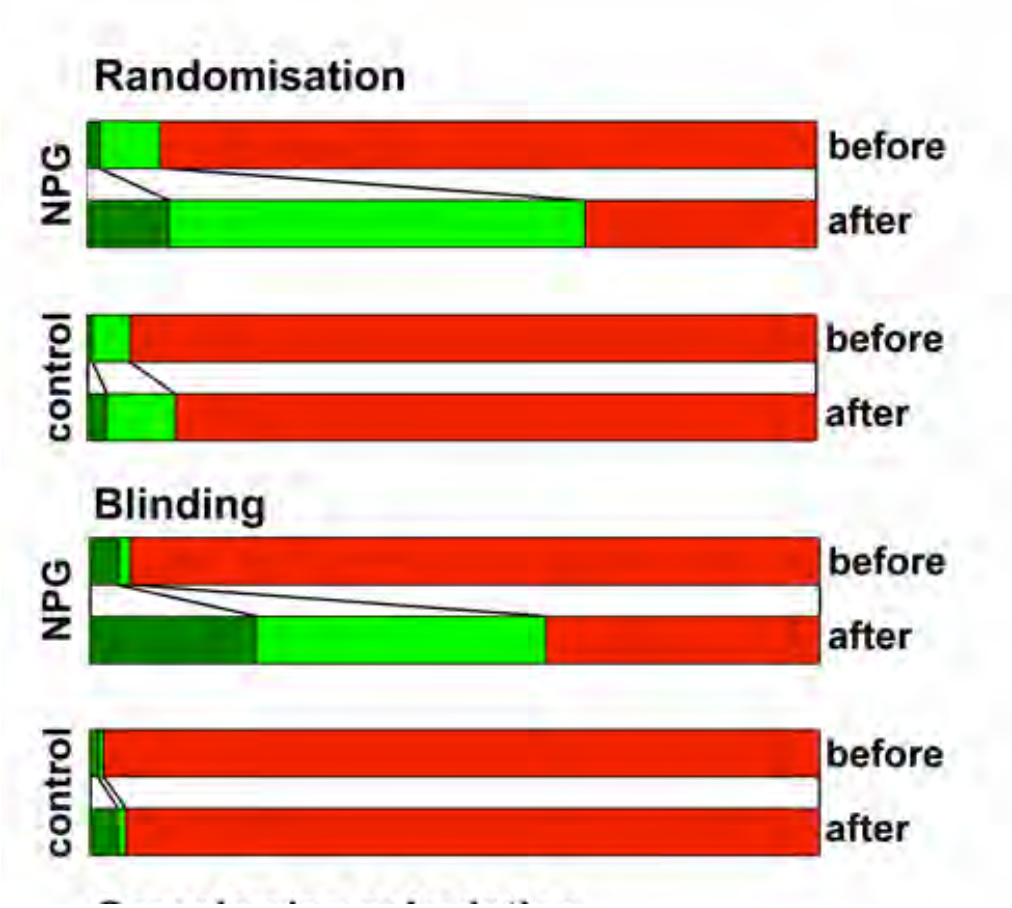
Findings of a retrospective, controlled cohort study of the impact of a change in Nature journals' editorial policy for life sciences research on the completeness of reporting study design and execution

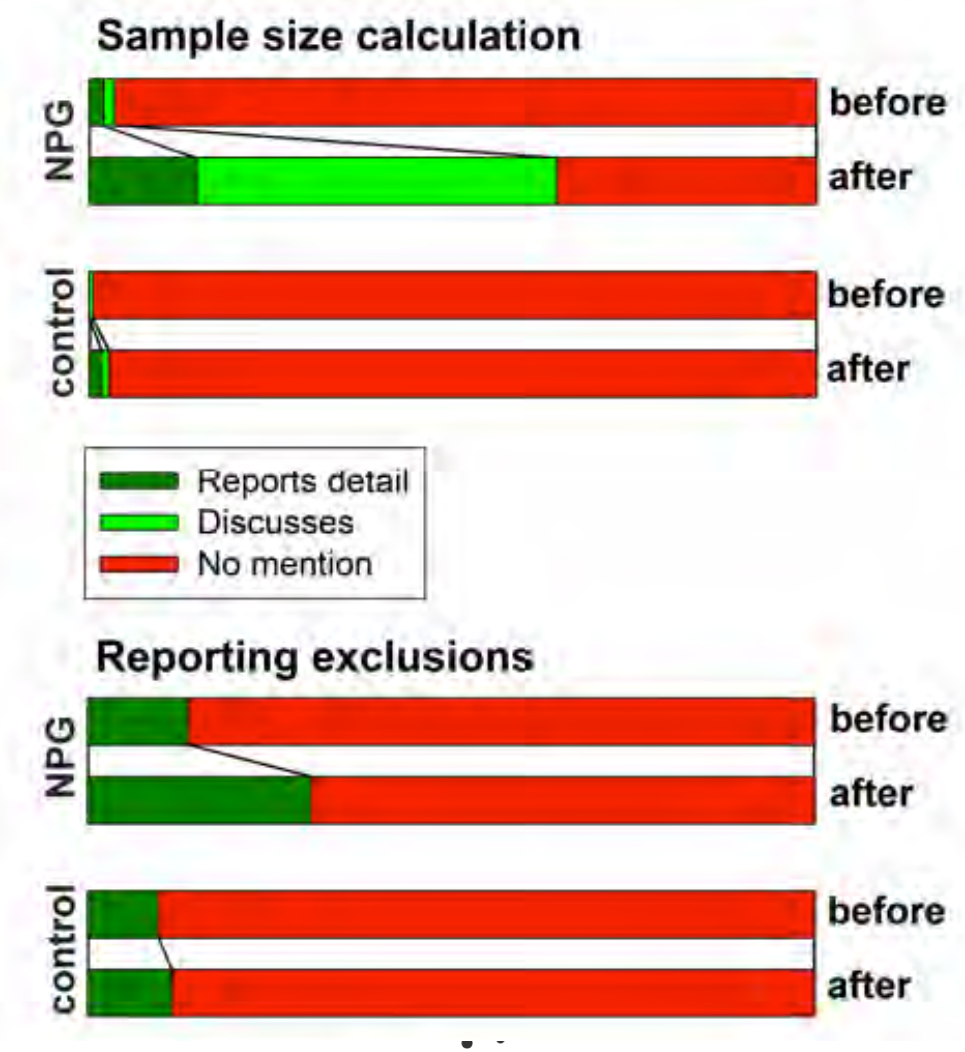
Malcolm Robert Macleod, The NPQIP Collaborative group

doi: https://doi.org/10.1101/187245

n = 394 intervention (NPG); n = 353 matching control (non-NPG)

Substantial improvement in reporting of risk bias in *in vivo* research





Macleod and the NPQIP Collaborative Group
bioRxiv Sep 2017 doi:10.1101/187245

What have we learned?

- Simply asking authors to fill out a checklist is not sufficient to improve reporting.
- Focused attention on fewer key items appears more effective.
- Even engaged editorial attention does not lead to full compliance.
- Compliance improves over time.
- Reporting improvements ≠ study design improvements.
- Substantial ambiguity in checklists formulation.
- Ongoing next steps: editorial collaboration between journals to establish minimal standard checklist as foundation.

PLOS

Open Science

as a cross-disciplinary approach

Open Science, as defined by the availability of key elements of study design, raw results and analysis methods, fosters transparency and facilitates replication.

- Data
- Code
- Methodologies
- Reagents

PLOS

An example: PLOS Data Availability Policy

- PLOS journals require authors to make all data underlying the findings described in their manuscript available at publication.
- Since March 2014, PLOS has published >87,000 articles with a Data Availability Statement describing compliance with this policy.
- Exceptions to public availability are made occasionally for ethical or legal reasons.
- <0.1% rejection for unwillingness or inability to share data

PLOS

FAIR data principles

- **Findable** – unique persistent identifier, rich metadata
- **Accessible** – retrievable (incl. authorization where necessary)
- **Interoperable** – broadly applicable language, qualified references
- **Reusable** – metadata, usage license, provenance, domain-relevant community standards.

Importance of FAIR principles as open access to data is not always sufficient and not always necessary to address R&R challenges.

PLOS

Underlying issues affecting reproducibility across disciplines

"Given finite resources, the importance placed on novel findings, and the emphasis on a relatively small number of publications, scientists wishing to accelerate their career progression should conduct a large number of exploratory studies, each of which will have low statistical power."

Higginson and Munafo, PLOS Biology, 2016, doi: 10.1371/journal.pbio.2000995

"As competition for jobs and promotions increases, the inflated value given to publishing in a small number of so-called "high impact" journals has put pressure on authors to rush into print, cut corners, exaggerate their findings, and overstate the significance of their work."

Alberts, Kirschner, Tilghman and Varmus, PNAS 2014, vol. 11 (16), 5773. doi: 10.1073/pnas.1404402111

Limitations of the scientific literature

- Insufficient distinction between **exploratory** and **hypothesis-testing** research, in particular in broadly defined biomedical research.
- **Publication bias** affects ability to draw reliable conclusions about phenomena.

Publishing solutions exist to counter these limitations:

✓ Broad-scope, inclusive journals with selection criteria focuses on rigor instead of perceived importance (PLOS ONE, Scientific Reports,…)

✓ Increasing momentum of preprint servers.

Better incentives are needed to promote best practices and prevent damaging behaviors.

PLOS

In conclusion

- Open Science is a cross-disciplinary approach to tackle reproducibility and replicability at the publication level.
- Reporting transparency is a pre-requisite across disciplines, but standards need to be defined with domain expertise.
- Journals editorial policy interventions can help but implementation is challenging.
- Better incentives are needed to promote open science, facilitate and reward replication and validation, counter publication bias, and improve rigor and reproducibility.

PLOS