The Open Science Framework: Increasing Reproducibility Across the Entire Research Lifecycle

Courtney Soderberg
Statistical & Methodological Consultant
Center for Open Science
@OSFramework
Power failure: why small sample size undermines the reliability of neuroscience

Katherine S. Button\footnote{1,2}, John P. A. Ioannidis\footnote{3}, Claire Mokrysz\footnote{4}, Brian A. Nosek\footnote{5}, Jonathan Flint\footnote{6}, Emma S. J. Robinson\footnote{7} and Marcus R. Munafò\footnote{8}

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 should be a major priority for the field.

Published research findings are sometimes refined by subsequent evidence, with ensuing confusion and disappointment. Refutation and controversy is seen across the range of research designs, from clinical trials to traditional epidemiological studies [1–3] to the most modern molecular research [4,5]. There is increasing concern that in modern research, false findings may be the majority of the results [6–8]. However, this should not be surprising. It can be proven that most claimed research findings are false. Here I will examine the key factors that influence this problem and some corollaries thereof.

Why Most Published Research Findings Are False

John P. A. Ioannidis

Essay

Why Most Published Research Findings Are False

John P. A. Ioannidis

Summary

There is increasing concern that most current published research findings are false. The probability that a research claim is true may depend on study power and bias, the number of other studies on the same topic, and, importantly, the ratio of true to no relationships among the relationships identified in each scientific field. In this framework, a research finding is less likely to be true when the study size is small, when effects are smaller and when there is a greater number of reported relationships. The power of a finding is a function of the field, the study size, and the number of relationships. In a scientific field in which some relationships are true, the probability of a true relationship being reported is lower. The probability of a study finding a true relationship is high if the study is large and the number of relationships is small. The probability of a study finding a true relationship is low if the study is small and the number of relationships is large.

It can be proven that most claimed research findings are false.

should be interpreted based only on p-values. Research findings are defined here as any relationship reaching formal statistical significance, e.g., effective interventions, informative predictors, risk factors, or associations. "Negative" research is also very useful. "Negative" is actually a minorer, and the misinterpretation is widespread. However, here we will target relationships that investigators claim exist, rather than null findings.

As has been shown previously, the probability that a research finding is indeed true depends on the prior probability of it being true (before doing the study), the statistical power of the study, and the level of statistical significance [10,11]. Consider a 2 × 2 table in which research findings are compared against the gold standard of true relationships in a scientific field. In a research field both true and false hypotheses can be made about the presence of relationships. Let R be the ratio of the number of "true relationships" to "false relationships". The Essay section contains opinion pieces on topics relating to neuroscience.
Data and code sharing

• Sharing data and code can help increase computational reproducibility.

• Doesn’t indicate how analyses may have changed over time.

• Mainly for published papers.
Acquire materials

Design study

Develop idea

Search and discover

Publish report

Write report

Interpret findings

Analyze data

Collect data

Store data

Researcher DFs

HARKing
Publication Bias

- Positive results more likely to get published
Figure 1. Positive Results by Discipline.

http://127.0.0.1:8081/plosone/article?id=info:doi/10.1371/journal.pone.0010068
Publication Bias

- Positive results more likely to get published
- The file drawer problem
- Leads to biased accumulation of knowledge through the published literature
Search and discover

Develop idea

Design study

Acquire materials

Collect data

Store data

Analyze data

Write report

Interpret findings

Publication Bias

HARKing

Researcher DFs
Research Degrees of Freedom

• Any data processing and analytical choices made after seeing and interacting with data

• Can severely inflate false positive

• Often occur outside of conscious awareness
How can improve?

- Increase documentation of the workflow
- Document from the beginning
- Make discoverability of all research, published or unpublished, easier
Open Science Framework

https://osf.io
Resources

● Free consulting on reproducible stats and methods
  ○ stats-consulting@cos.io
  ○ https://cos.io/stats_consulting/

● OSF Helpdesk
  ○ support@osf.io