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

Courtney Soderberg
Statistical & Methodological Consultant
Center for Open Science
@OSFramework

CENTER FOR
OPEN SCIENCE

COMMUNITY

METASCIENCE

INFRASTRUCTURE

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

Florian Prinz, Thomas Schlange and Khusru Asadullah

A recent report by Arrowsmith noted that the success rates for new development projects in Phase II trials have fallen from 28% to 18% in

to 'feasible/marketable', and the financial costs of pursuing a full-blown drug discovery and development programme for a particular tar-

results that are published duce. However, there is an this apparently widespread public recognition (for exa and the surprisingly few tions dealing with this top knowledge, so far there has in-depth, systematic analy reproduced results with pu wet-lab experiments related tion and validation.

Early research in the pha try, with a dedicated budget mainly work on target val

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

Katherine S. Button[1,2], John P. A. Ioannidis[3], Claire Mokrysz[1], Brian A. Nosek[4], Jonathan Flint[5], Emma S. J. Robinson[6] and Marcus R. Munafò[1]

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

Open access, freely available onli

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 question, and, importantly, the ratio of true to no relationships among the relationships probed in each scientific field. In this framework, a research finding is less likely to be true when the studies conducted in a field are smaller; when effect sizes are smaller; when there is a greater number and lesser preselection of tested relationships; where there is greater flexibility in designs, definitions, outcomes, and analytical modes; when there is greater financial and other interest and prejudice; and when more teams are involved in a scientific field in chase of statistical significance. Simulations show that for most study designs and settings, it is more likely for a research claim to be false than true. Moreover, for many current scientific fields, claimed research findings may often be simply accurate measures of the prevailing bias. In this essay, I discuss the implications of these problems for the conduct and interpretation of research.

Published research findings are sometimes refuted by subsequent evidence, with ensuing confusion and disappointment. Refutation and controversy is seen across the range of research designs, from clinical trials and 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 or even the vast majority of published research claims [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.

The Essay section contains opinion pieces on topics

Modeling the Framework for False Positive Findings

Several methodologists have pointed out [9–11] that the high rate of nonreplication (lack of confirmation) of research discoveries is a consequence of the convenient, yet ill-founded strategy of claiming conclusive research findings solely on the basis of a single study assessed by formal statistical significance, typically for a *p*-value less than 0.05. Research is not most appropriately represented and summarized by *p*-values, but, unfortunately, there is a widespread notion that medical research articles

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 misnomer, 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 "no relationships"

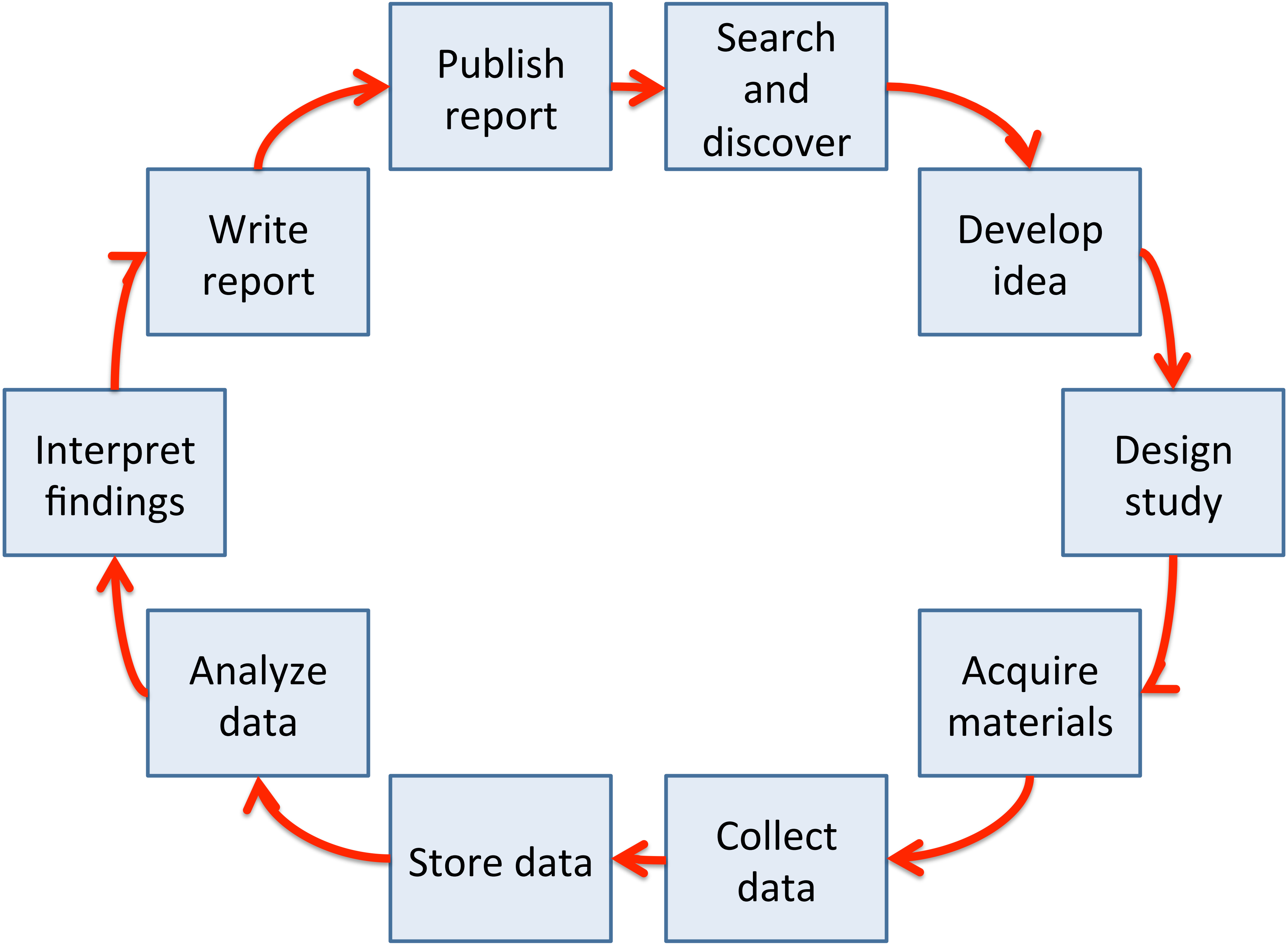
is characteristic of the field and can vary a lot depending on whether the field targets highly likely relationship or searches for only one or a few true relationships among thousands and millions of hypotheses that may be postulated. Let us also consider, for computational simplicity, circumscribed fields where either the is only one true relationship (among many that can be hypothesized) or the power is similar to find any of the several existing true relationships. Th pre-study probability of a relationship being true is $R/(R+1)$. The probabil of a study finding a true relationship reflects the power $1-\beta$ (one minus the Type II error rate). The probabil of claiming a relationship when none truly exists reflects the Type I error rate, α. Assuming that c relationships are being probed in the field, the expected values of the 2 × 2 table are given in Table 1. After a research finding has been claimed based on achieving formal statistical significanc the post-study probability that it is tru is the positive predictive value, PPV. The PPV is also the complementary probability of what Wacholder et al. have called the false positive report probability [10]. According to the 2 × 2 table, one gets PPV = $(1-\beta)R/(R$ $-\beta R+\alpha)$. A research finding is thus

Citation: Ioannidis JPA (2005) Why most publishec research findings are false. PLoS Med 2(8): e124.

Abbreviation: PPV, positive predictive value

John P. A. Ioannidis is in the Department of Hygien and Epidemiology, University of Ioannina School o Medicine, Ioannina, Greece, and Institute for Clinica Research and Health Policy Studies, Department o Medicine, Tufts-New England Medical Center, Tufts University School of Medicine, Boston, Massachuse United States of America. E-mail: jioannid@cc.uoi.g

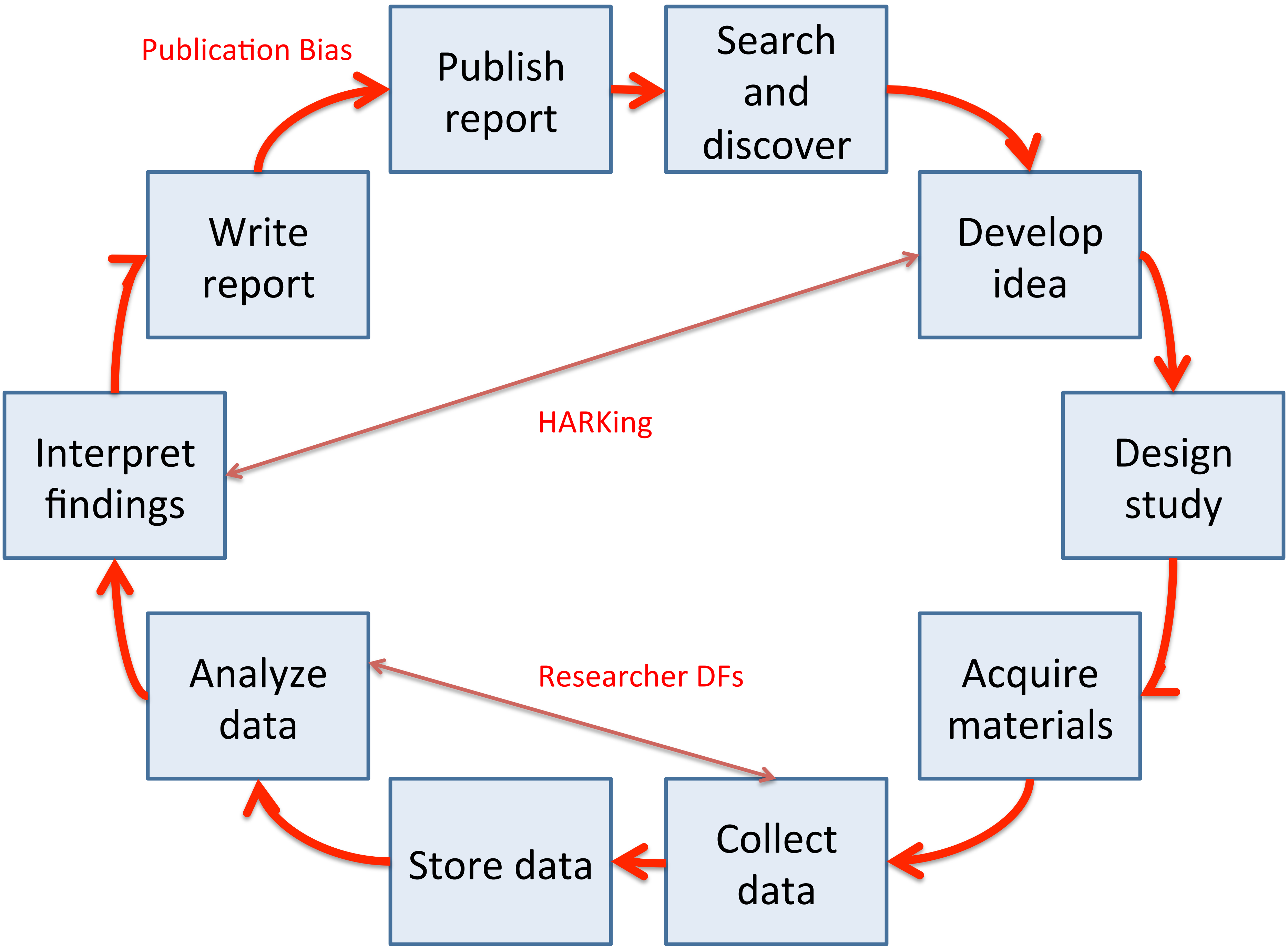
Competing Interests: The author has declared tha no competing interests exist.



Publish report
Search and discover
Develop idea
Design study
Acquire materials
Collect data
Store data
Analyze data
Interpret findings
Write report

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

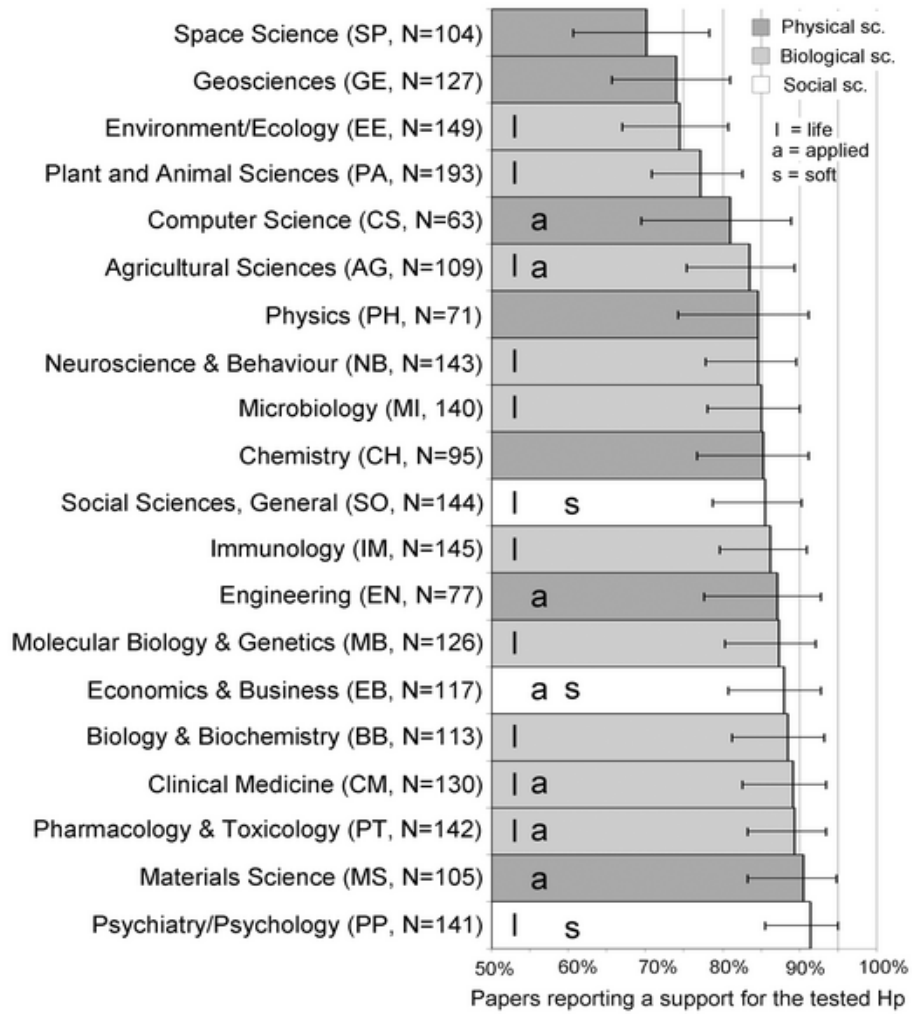


Publication Bias
Publish report
Search and discover
Develop idea
Design study
Acquire materials
Collect data
Store data
Analyze data
Interpret findings
Write report
HARKing
Researcher DFs

Publication Bias

- Positive results more likely to get published

Figure 1. Positive Results by Discipline.

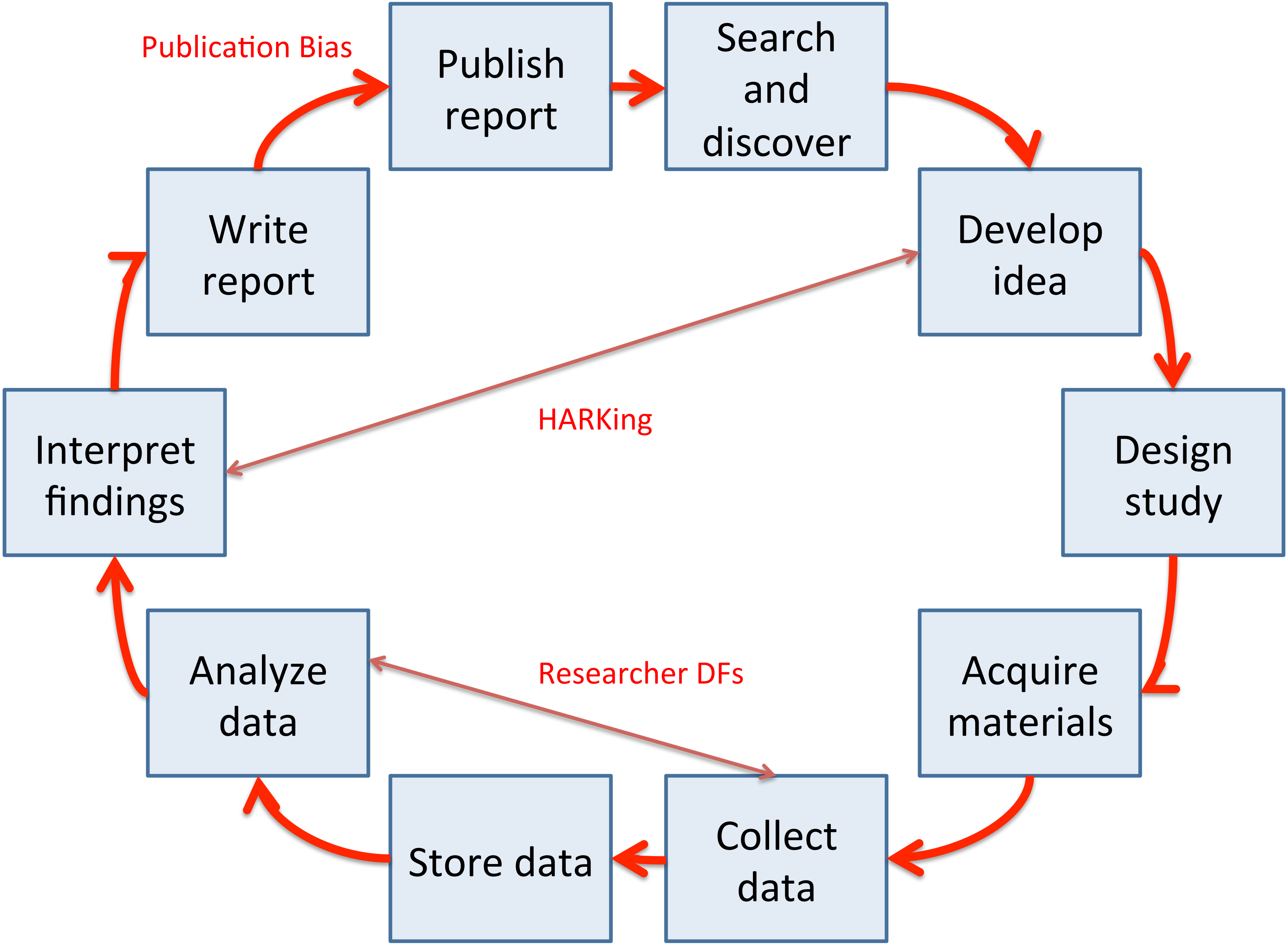


Fanelli D (2010) “Positive” Results Increase Down the Hierarchy of the Sciences. PLoS ONE 5(4): e10068. doi:10.1371/journal.pone.0010068
http://127.0.0.1:8081/plosone/article?id=info:doi/10.1371/journal.pone.0010068

PLOS | ONE

Publication Bias

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


Publication Bias
Publish report
Search and discover
Develop idea
Design study
Acquire materials
Collect data
Store data
Analyze data
Interpret findings
Write report
HARKing
Researcher DFs

Research Degrees of Freedom

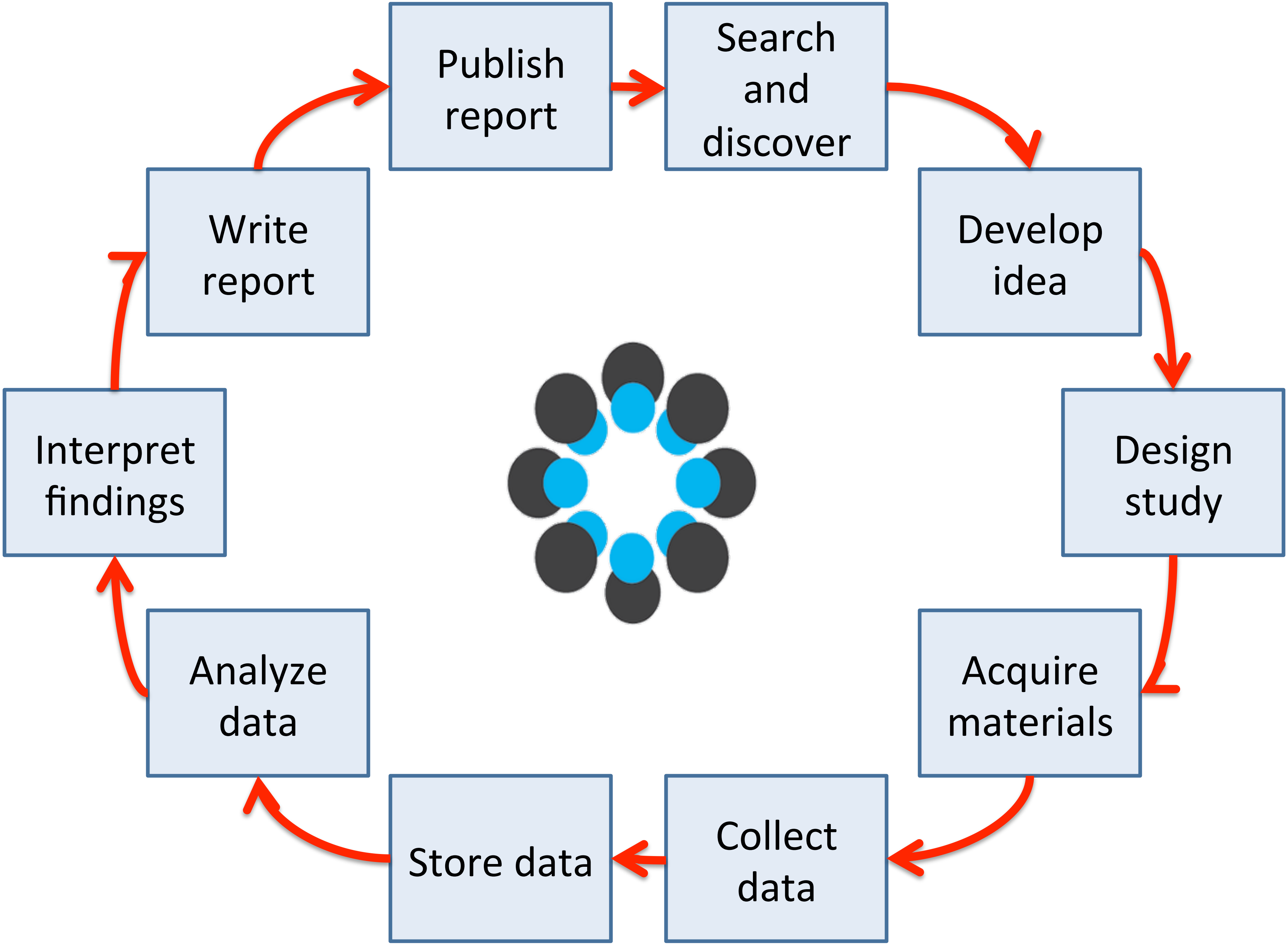
- Any data processing and analytical choices made after seeing and interacting with data
- Can severely inflate false positive
- Often occur outside of conscious awareness

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


Publish report
Search and discover
Develop idea
Design study
Acquire materials
Collect data
Store data
Analyze data
Interpret findings
Write report

Resources

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

- OSF Helpdesk
  - support@osf.io