



Reproducible Research through Open Science

November 20, 2019

Supporting materials at
<https://osf.io/vgjpq/>



Jason Pither (UBC Okanagan, Biology) & Mathew Vis-Dunbar (UBC SMP Librarian)

With additional materials from the Open Science Ambassadors Program
at the Center for Open Science





- [+ New](#)
- [Project Wiki Pages](#)
- [Home](#)
- + [Component Wiki Pages](#)

[View](#)

Wiki Version: (Current) Jason Pither: 2019-09-20 15:50:47+00:00 UTC

UBC Open Science

This OSF component links to resources referred to in Open Science seminars and workshops delivered at the University of British Columbia.

Presentations

- Presentations delivered at UBC are available [here](#).

Articles openly available as PDFs

- Some OS-related resources are provided as PDF documents, which are available for download from the [files page](#), and links to others are provided below.

Articles open online

Smaldino, P.E. and McElreath, R. 2016. The natural selection of bad science. Royal Society Open Science.

Ali-Khan, S.E., Jean, A., and Gold, R.E. 2018. Identifying the challenges in implementing Open Science. MNI Open Research.

Campbell et al. 2019. Early career researchers embrace data sharing. Trends Ecol. Evol. 24:95-98.

Nelson et al. 2018. Psychology's renaissance. Ann. Rev. Psych. 69:511-534.

Nosek et al. 2017. The pre-registration revolution. PNAS.

Nosek et al. 2015. Estimating the reproducibility of psychological science. Science, 349.

Simmons et al. 2011. False-positive psychology: undisclosed flexibility in data collection and analysis allows presenting anything as significant. Psych. Sci. 22(11):1359-1366.

Plessner. 2018. Reproducibility vs. Replicability: a brief history of a confused terminology. Frontiers in Neuroinformatics. 11:1-4.

*Lazic et al. 2018. What exactly is "n" in cell culture and animal experiments? *please be sure to read the [commentary](#) that corrects serious flaws with the Lazic et al. article

Reproducibility bibliography

- Librarians at the University of Minnesota maintain excellent discipline-specific [bibliographies](#) pertaining to reproducibility.

General OS resources

- [Transparency and Openness Promotion Guidelines](#)



openscience.ubc.ca

Open Science @ UBC

Home Page About OSF ▾ Instructors Learn Open Science in Practice ▾ Publish Blog

Open Science and Open Research. Fostering knowledge diffusion and reproducibility. Enhancing access and discoverability.

Research

Open up your research with Open Science Framework (OSF).



Instruct

Integrate Open principles and practices into your classroom.



Learn

Workshops, events, and educational resources in support of Open.



Open Science in Practice

Calling all UBC researchers and educators engaged in Open Science:
Share your story here!



Publish

Background

Problem

- Irreproducible research
- Sources of the problem

Solutions

- Practices that improve reproducibility – Open Science
- Other benefits of Open Science best practices
- Tools that facilitate Open Science best practices

Open Science initiatives at UBC

- Excellence fund strategic initiative
- Okanagan pilot project

Discussion / Q&A

Terminology

One of many possible definitions of **Open Science**:

Scientific research conducted and communicated in an honest, accessible, and transparent way, such that independent researchers can reproduce the results

OPEN LETTER

Identifying the challenges in implementing open science [version 1; referees: 2 approved]

Sarah E. Ali-Khan  ^{1,2}, Antoine Jean¹, E. Richard Gold  ^{1,3}

MNI Open Research 2018, 2:5 Last updated: 04 FEB 2019

Open science (OS) comprises a set of institutional policies, infrastructure and relationships related to open access publication, open data and scientific resources, and lack of restrictive intellectual and other proprietary rights with the goal of increasing the quality and credibility of scientific outputs, increasing efficiency, and spurring both discovery and innovation.

Terminology

Computational Reproducibility:

If we took your data and code/analysis scripts and reran it, we can reproduce the numbers/graphs in your paper

Methods Reproducibility:

We have enough information to rerun the experiment or survey the way it was originally conducted

Results Reproducibility/Replicability:

We reproduce the methods (as above), collect new data, and get the same statistical conclusion

Terminology

Computational Reproducibility:

If we took your data and code/analysis scripts and reran it, we can reproduce the numbers/graphs in your paper

Methods Reproducibility:

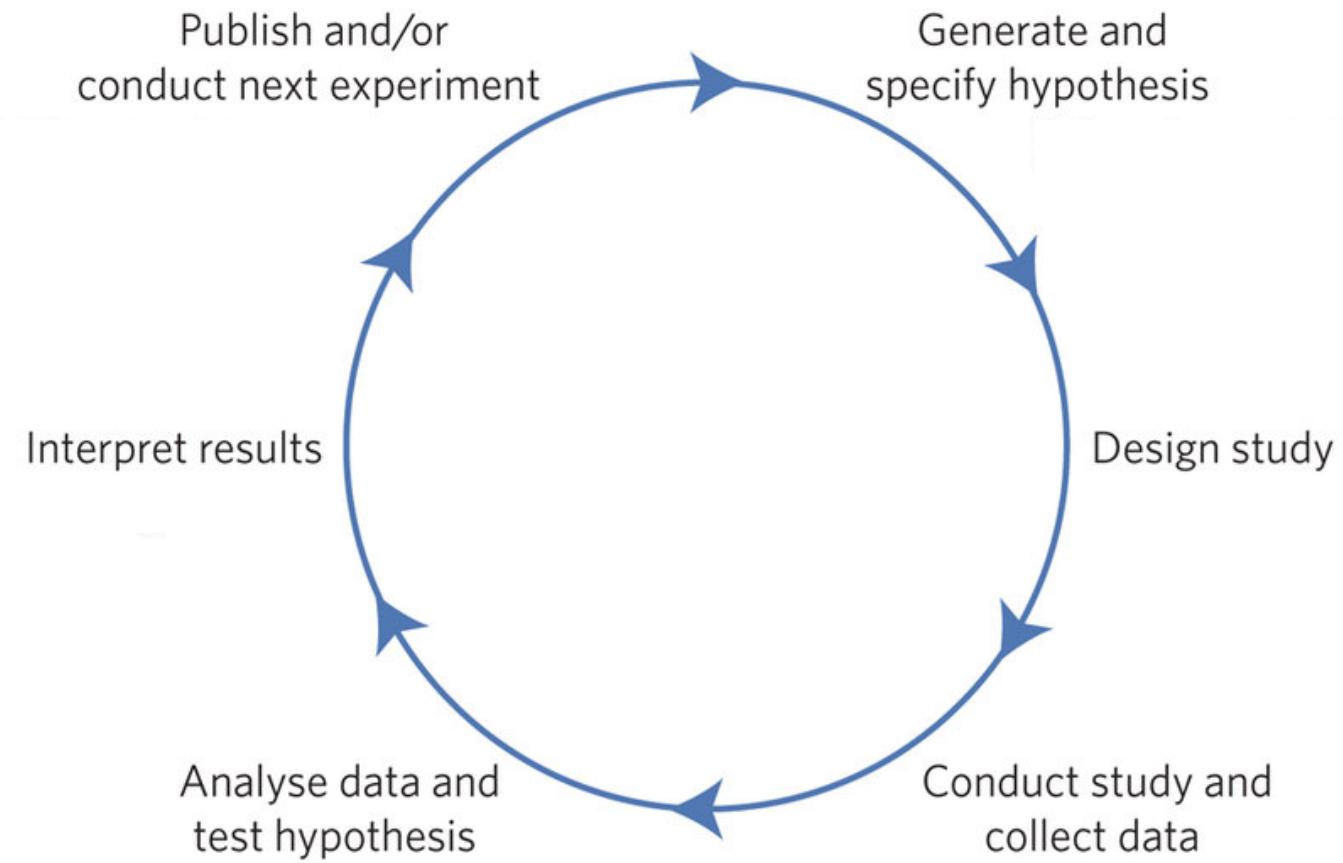
We have enough information to rerun the experiment or survey the way it was originally conducted

Results Reproducibility/Replicability:

We reproduce the methods (as above), collect new data, and get the same statistical conclusion

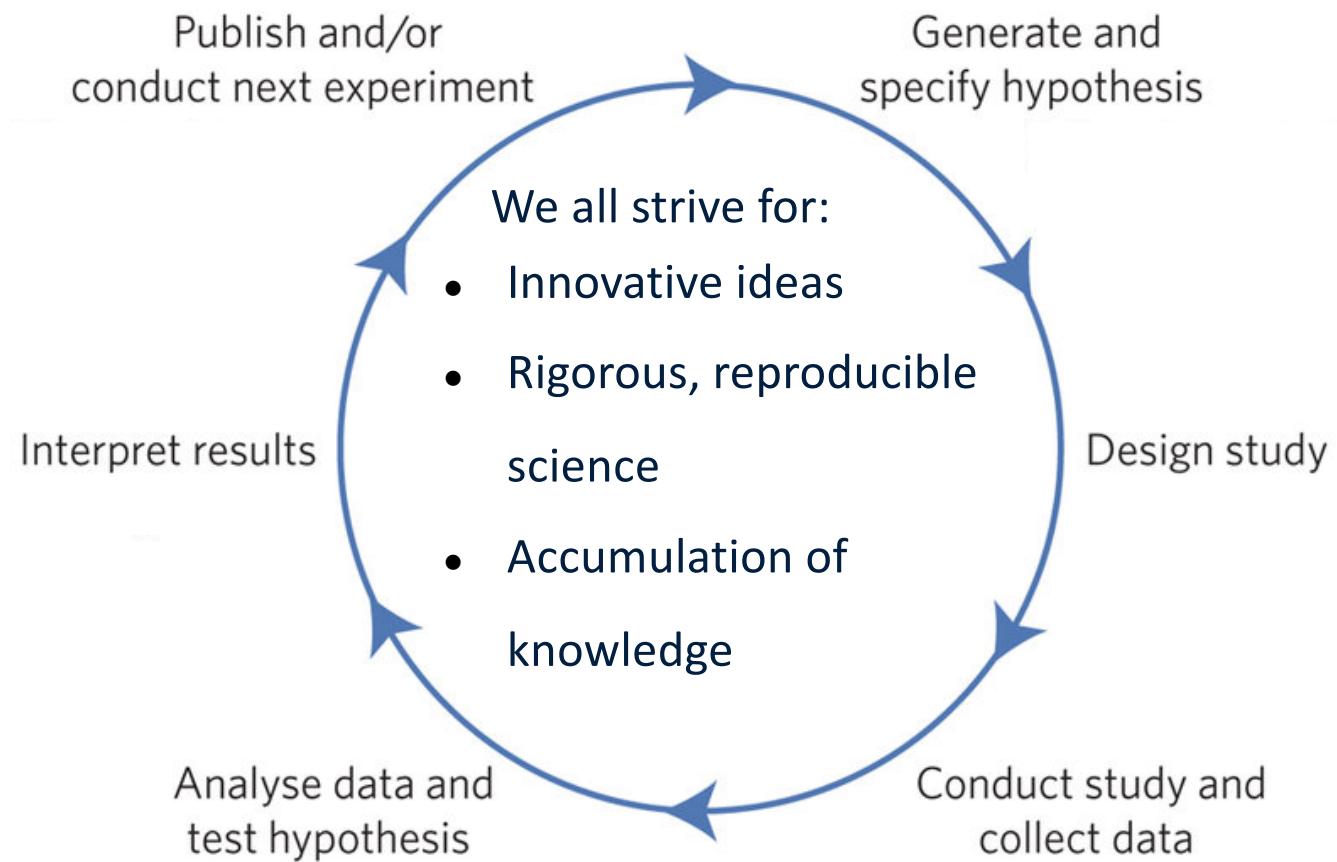
Foundational to science

Hypothetico-Deductive Model of the Scientific Method



Note distinction between Confirmatory studies vs Exploratory studies

Hypothetico-Deductive Model of the Scientific Method



The problem

Questionable methods infiltrate the research workflow at various stages (typically unwittingly!), ultimately yielding irreproducible research.

Science is broken?

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

Katherine S. Button^{1,2}, John P. A. Ioannidis³, Claire Mokrysz¹, Brian A. Nosek⁴, Jonathan Flint⁵, Emma S. J. Robinson⁶ and Marcus R. Munafò¹

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 is a key priority and requires attention to well-established but often ignored methodological principles.

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

NATURE REVIEWS | DRUG DISCOVERY

Florian Prinz, Thomas Schlange and Khusru Asadullah

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

The Essay section contains opinion pieces on topics of broad interest to a general medical audience.

factors that influence this problem and some corollaries thereof.

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” among those tested in the field. *R*

is characteristic of the field and can vary a lot depending on whether the field targets highly likely relationships 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 there 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. The pre-study probability of a relationship being true is *R*/*R* + 1. The probability of a study finding a true relationship reflects the power $1 - \beta$ (one minus the Type II error rate). The probability of claiming a relationship when none truly exists reflects the Type I error rate, α . Assuming that ϵ 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 significance, the post-study probability that it is true 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 $PPV = (1 - \beta)/R/(R - \beta + \alpha)$. A research finding is thus

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

Copyright: © 2005 John P.A. Ioannidis. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Abbreviation: PPV, positive predictive value

John P.A. Ioannidis is in the Department of Hygiene and Epidemiology, University of Ioannina School of Medicine, Ioannina, Greece, and Institute for Clinical Research and Health Policy Studies, Department of Medicine, Tufts-New England Medical Center, Tufts University School of Medicine, Boston, Massachusetts, United States of America. E-mail: ioannidis@ccs.uoa.gr

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

DOI: 10.1371/journal.pmed.0020124

August 2005 | Volume 2 | Issue 8 | e124

Science is broken?

Unreliable research

Trouble at the lab

Scientists like to think of science as self-correcting. To an alarming degree, it is not



The
Economist

THE WEEK

Big Science is broken



Pascal-Emmanuel Gobry

<https://theweek.com/articles/618141/big-science-broken>

SCIENCE

Science Is Broken. How Much Should We Fix It?

More rigor in research could stamp out false positive results. It might also do more harm than good.

By DANIEL ENGBER

SLATE
* THE WORKIEST WORKISTS OF ALL

MAY 05, 2017 • 5:56 AM

Reproducibility is not the norm

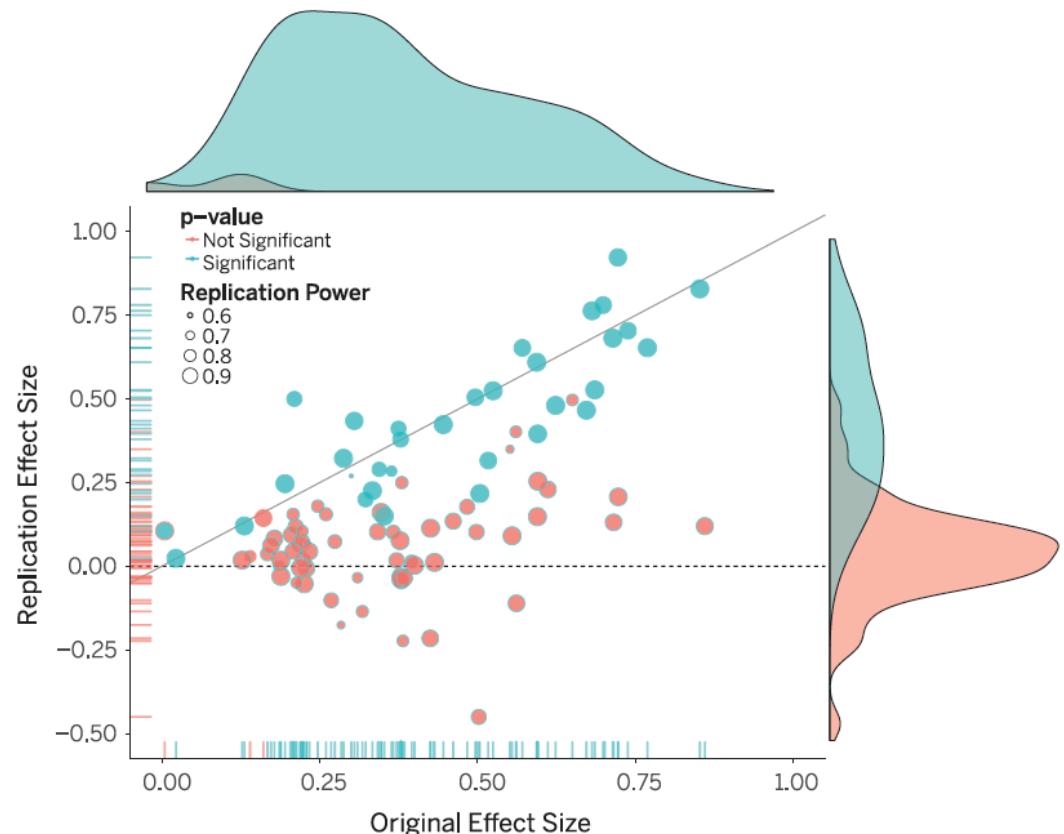
RESEARCH ARTICLE SUMMARY

PSYCHOLOGY

Estimating the reproducibility of psychological science (2015)

Open Science Collaboration*

- 100 studies published in 3 psychology journals
- Used high-powered designs and original materials
- 36% of studies were able to be reproduced



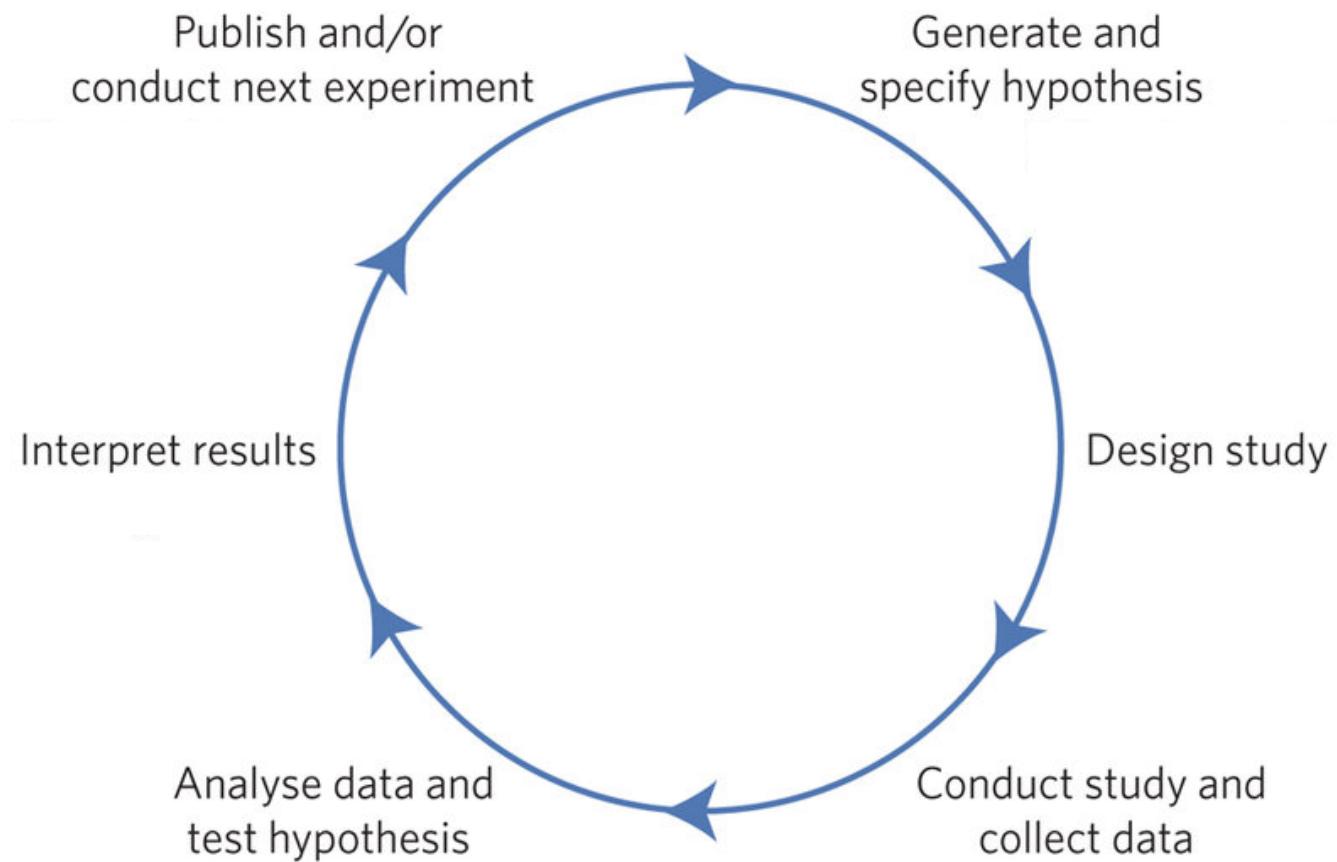
Original study effect size versus replication effect size (correlation coefficients). Diagonal line represents replication effect size equal to original effect size. Dotted line represents replication effect size of 0. Points below the dotted line were effects in the opposite direction of the original. Density plots are separated by significant (blue) and nonsignificant (red) effects.

Reproducibility is not the norm

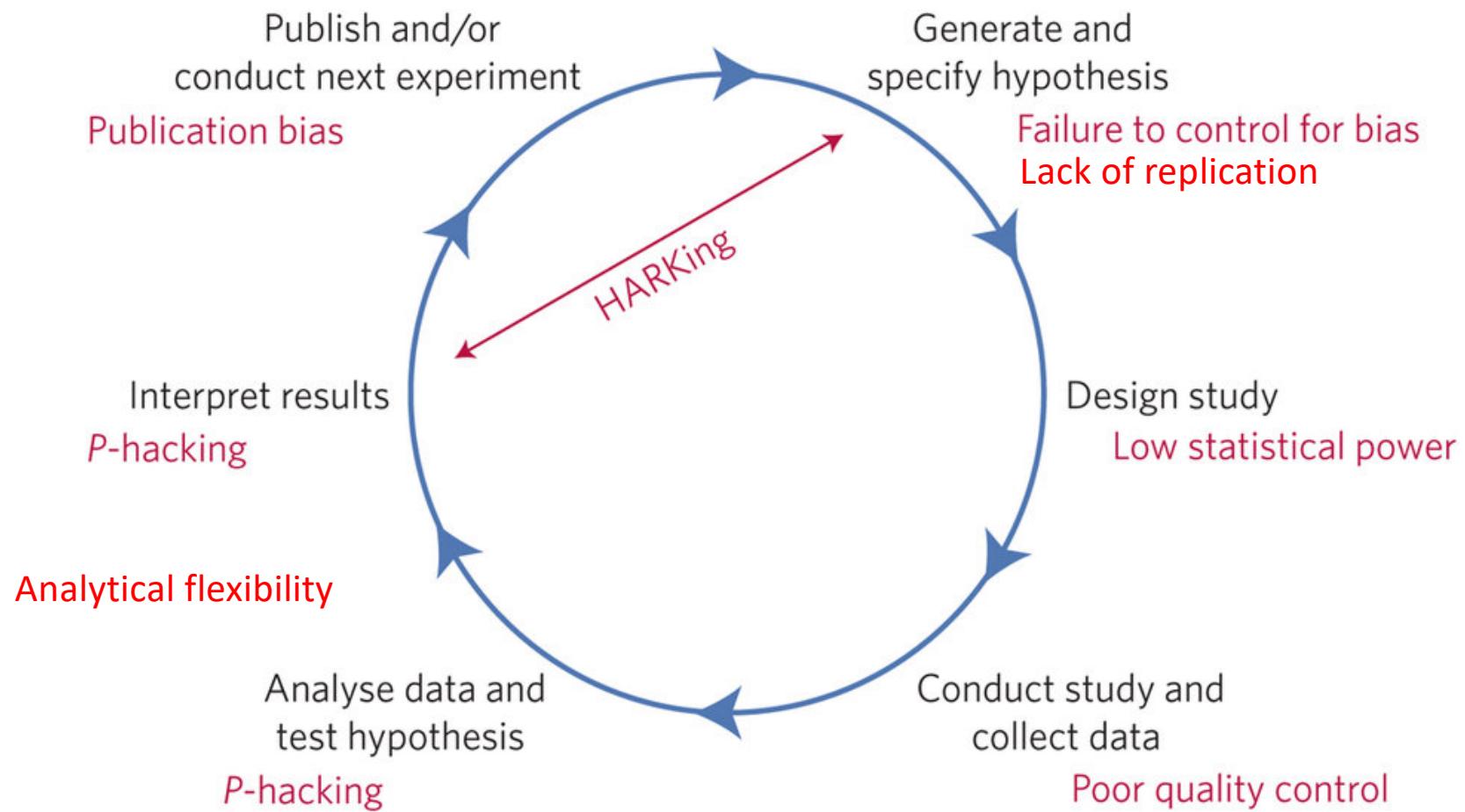
The screenshot shows the eLIFE website. At the top, there is a navigation bar with links for 'ABOUT', 'COMMUNITY', 'SUBMIT MY RESEARCH' (which is highlighted in blue), and 'LOG IN/REGISTER'. Below the navigation bar, the 'eLIFE' logo is on the left, followed by a menu with 'HOME MAGAZINE INNOVATION'. On the right side of the header is a search icon. In the center, there is a category link 'CANCER BIOLOGY' and a download icon. The main title of the page is 'Reproducibility in Cancer Biology: The challenges of replication', described as an 'EDITORIAL' dated 'Jan 19, 2017'. Below the title are social media sharing icons for Facebook, Twitter, Email, and Google+. The entire screenshot is enclosed in a thin black border.

To date four of the studies have reproduced important parts of the original papers; four of the studies have reproduced parts of the original papers but also contain results that could not be interpreted or are not consistent with some parts of the original paper; two of the studies could not be interpreted; and two studies did not reproduce the parts of the original papers that they attempted to reproduce.

Sources of the problem



Sources of the problem

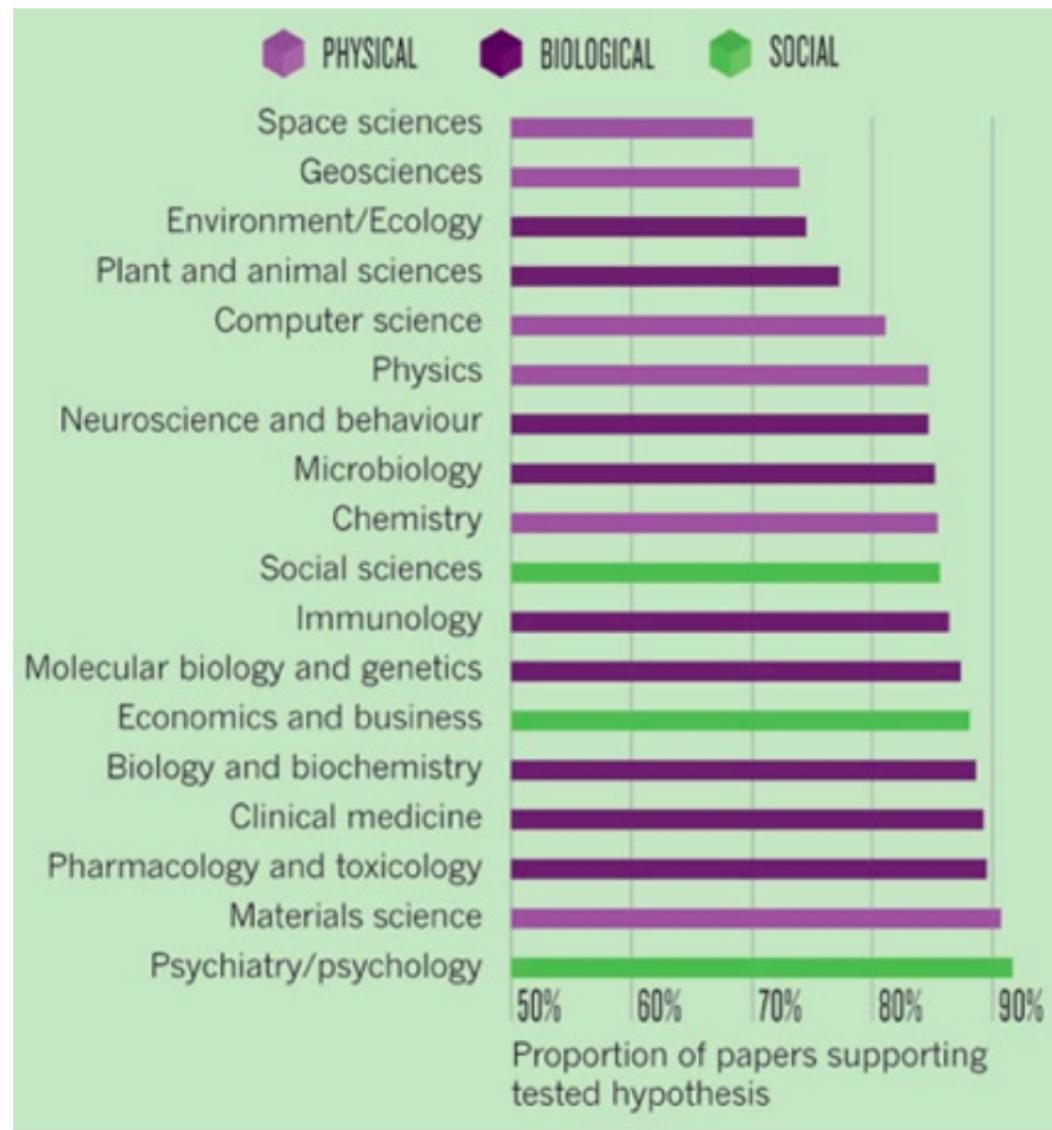


Publication bias: positive results

Positive Results by Discipline

Fanelli D (2010) "Positive" Results Increase Down the Hierarchy of the Sciences.

PLOS ONE 5(4): e10068. [CC-BY doi:10.1371/journal.pone.0010068](https://doi.org/10.1371/journal.pone.0010068)



Publication bias: file drawer

 [MENU](#) ▾

news [Top Stories](#) [Local](#) [The National](#) [Opinion](#) [World](#) [Canada](#) [Politics](#)

Health

[Learn more about the new look.](#)



'We've been deceived': Many clinical trial results are never published

Canadian universities and research hospitals are among the worst offenders, according to new online tool

Darryl Hol · CBC News · Posted: Nov 24, 2016 5:00 PM ET | Last Updated: November 24, 2016

The National
Missing clinical trial results



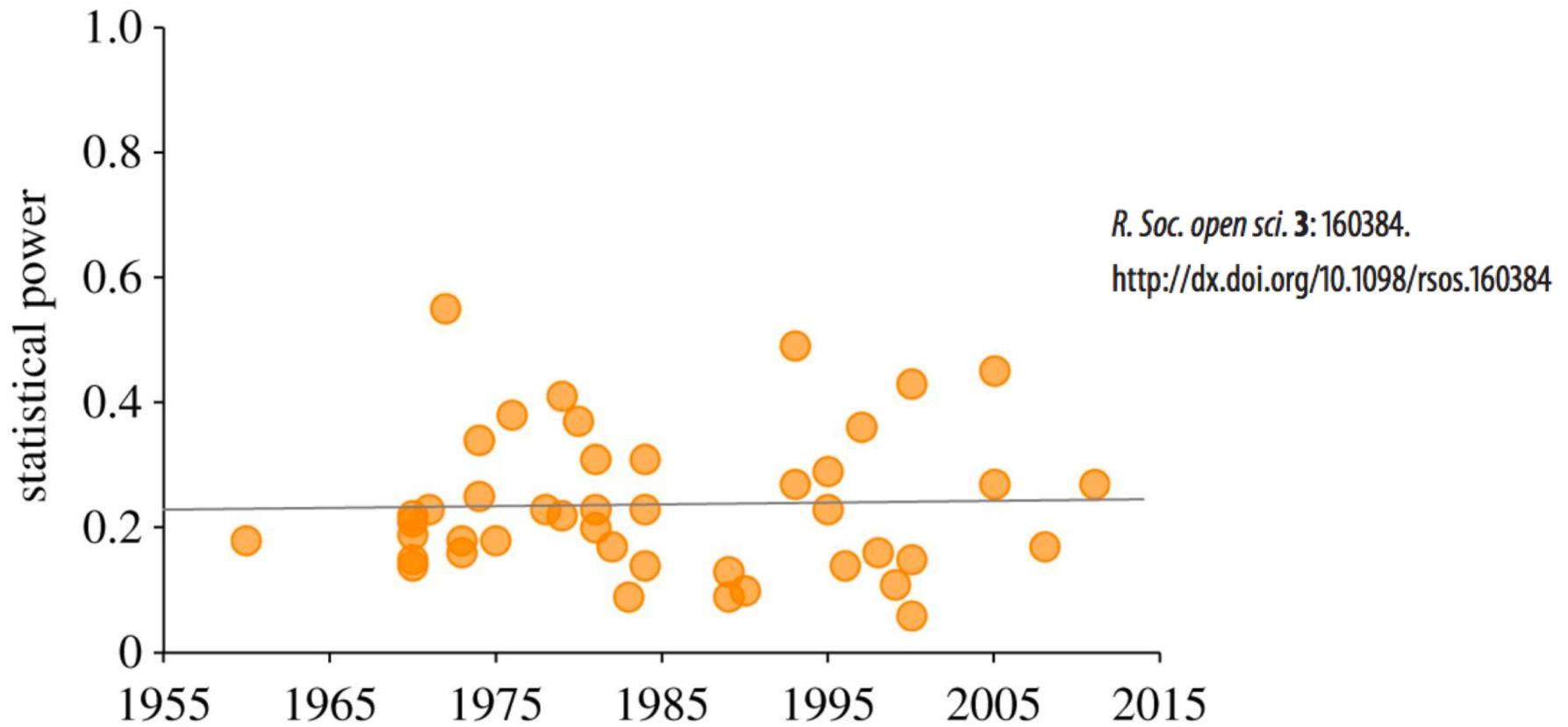
5.7k Shares

00:00 / 02:27

Publication bias: file drawer

University or Institution	Proportion of registered trials missing results (%)
U. Health Network	76.3
U. of Calgary	75.8
U. of British Columbia	72.4
Queen's University	70.2
U. of Manitoba	67.2
U. of Alberta	66.7
U. of Toronto	65.5
Ottawa Hospital Research Institute	63.8
Hospital for Sick Children	62.1
	69.9

Low power



Average statistical power from 44 reviews of papers published in journals in the social and behavioural sciences between 1960 and 2011. Data are power to detect small effect sizes ($d=0.2$), assuming a false-positive rate of $\alpha=0.05$, and indicate both very low power (mean=0.24) but also no increase over time ($R^2=0.00097$).

Low power

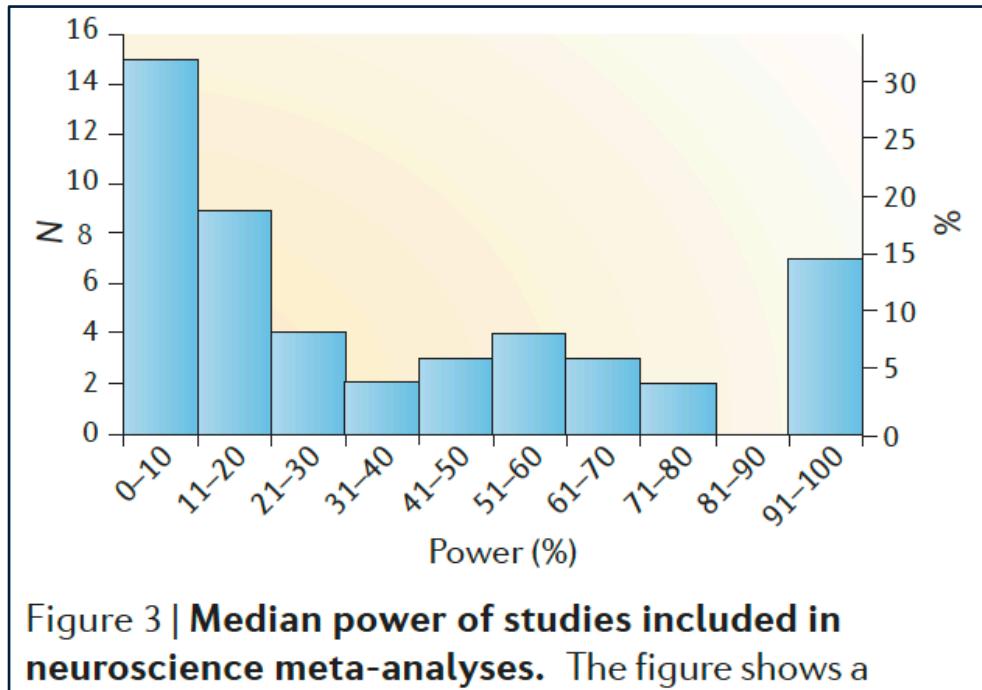


Figure 3 | Median power of studies included in neuroscience meta-analyses. The figure shows a

N = 49 studies

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

Katherine S. Button^{1,2}, John P. A. Ioannidis³, Claire Mokrysz¹, Brian A. Nosek⁴, Jonathan Flint⁵, Emma S. J. Robinson⁶ and Marcus R. Munafò¹

Low power

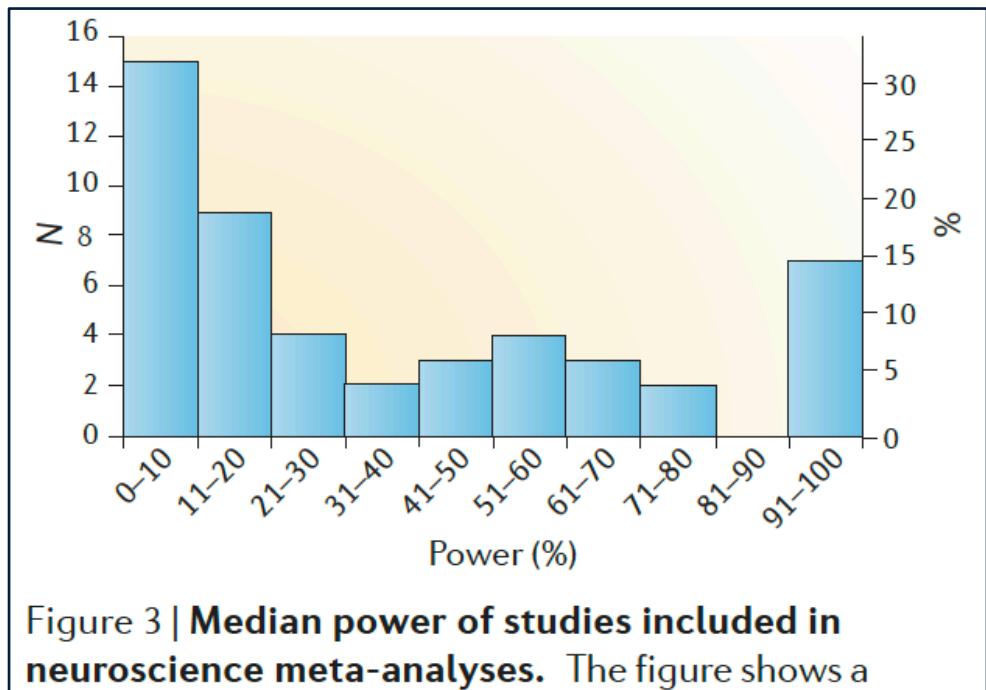


Figure 3 | Median power of studies included in neuroscience meta-analyses. The figure shows a

N = 49 studies

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

Katherine S. Button^{1,2}, John P. A. Ioannidis³, Claire Mokrysz¹, Brian A. Nosek⁴, Jonathan Flint⁵, Emma S. J. Robinson⁶ and Marcus R. Munafò¹

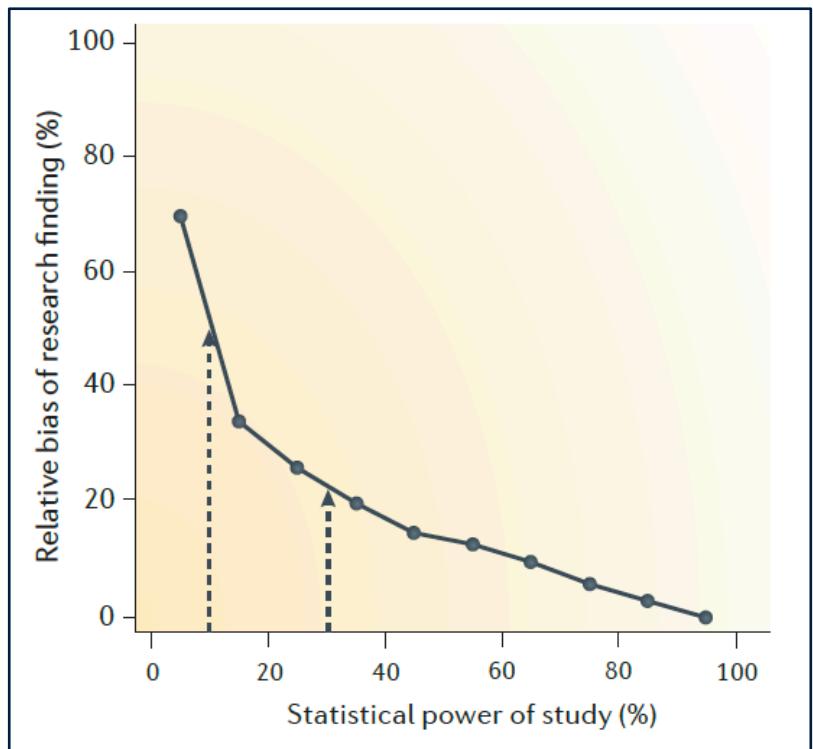


Figure 5 | The winner's curse: effect size inflation as a function of statistical power. The winner's curse

Low power

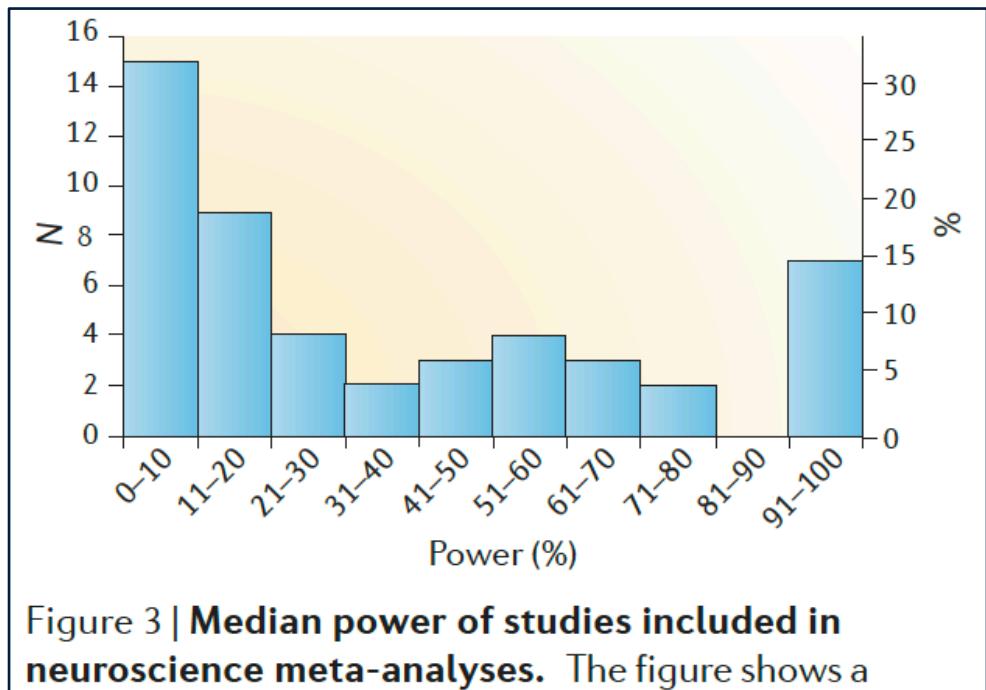


Figure 3 | Median power of studies included in neuroscience meta-analyses. The figure shows a

N = 49 studies

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

Katherine S. Button^{1,2}, John P. A. Ioannidis³, Claire Mokrysz¹, Brian A. Nosek⁴, Jonathan Flint⁵, Emma S. J. Robinson⁶ and Marcus R. Munafò¹

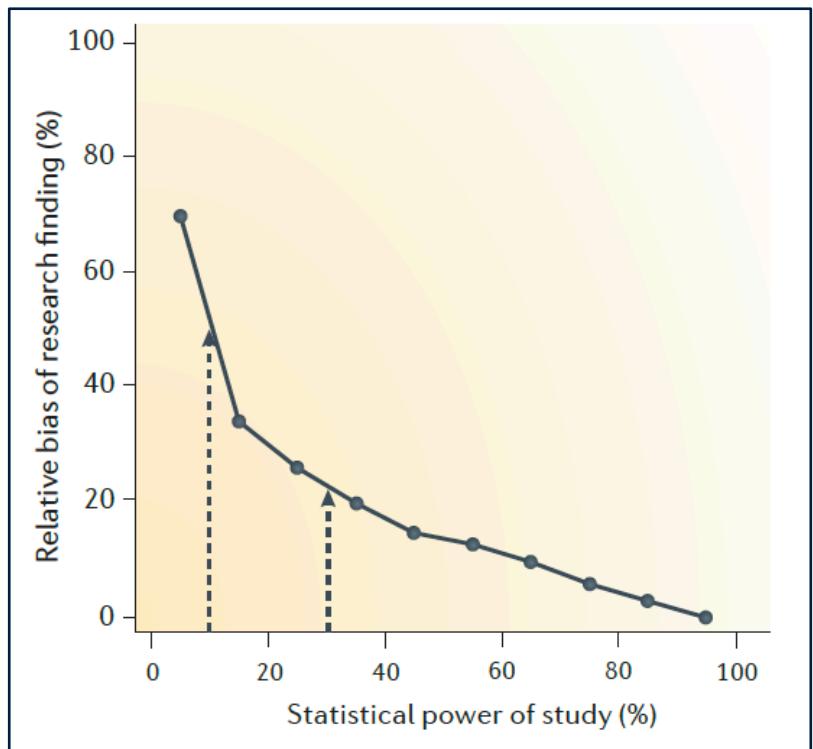


Figure 5 | The winner's curse: effect size inflation as a function of statistical power. The winner's curse

Low power studies are prone to inflated effect sizes (even if no true effect is present)!

Low power

Type “M” error:

Quantifies the proportion by which the critical value must exceed the effect size in order to achieve statistical significance

Researcher degrees of freedom

All data processing and analytical choices made after seeing and interacting with your data

Results become data dependent, and no longer adhere to the original hypothesis testing model

Researcher degrees of freedom

- Should more data be collected?
- Should I exclude this "outlier"? Is it an "outlier"?
- These extra variables I measured, should they be used?
- This analysis seems to behave better than the other...
- Which variable should I use as my main dependent variable?
- How should I treat this covariate?

Researcher degrees of freedom

- Should more data be collected?
- Should I exclude this "outlier"? Is it an "outlier"?
- These extra variables I measured, should they be used?
- This analysis seems to behave better than the other...
- Which variable should I use as my main dependent variable?
- How should I treat this covariate?

Often feel very reasonable and logical in the moment, but...

Inflates false-positive rates and P -values become uninformative

Researcher degrees of freedom

Table I. Likelihood of Obtaining a False-Positive Result

Researcher degrees of freedom	Significance level		
	$p < .1$	$p < .05$	$p < .01$
Situation A: two dependent variables ($r = .50$)	17.8%	9.5%	2.2%
Situation B: addition of 10 more observations per cell	14.5%	7.7%	1.6%
Situation C: controlling for gender or interaction of gender with treatment	21.6%	11.7%	2.7%
Situation D: dropping (or not dropping) one of three conditions	23.2%	12.6%	2.8%
Combine Situations A and B	26.0%	14.4%	3.3%
Combine Situations A, B, and C	50.9%	30.9%	8.4%
Combine Situations A, B, C, and D	81.5%	60.7%	21.5%

Joseph P. Simmons¹, Leif D. Nelson², and Uri Simonsohn¹

¹The Wharton School, University of Pennsylvania, and ²Haas School of Business, University of California, Berkeley

Psychological Science
22(11) 1359–1366
© The Author(s) 2011

Inaccessible data

META-RESEARCH ARTICLE

PLOS Biology | <https://doi.org/10.1371/journal.pbio.2006930> November 20, 2018

Reproducible research practices,
transparency, and open access data in the
biomedical literature, 2015–2017

N = 149 randomly selected articles

Joshua D. Wallach^{1,2}, Kevin W. Boyack³, John P. A. Ioannidis^{4,5,6,7,8 *}

Inaccessible data

META-RESEARCH ARTICLE

PLOS Biology | <https://doi.org/10.1371/journal.pbio.2006930> November 20, 2018

Reproducible research practices,
transparency, and open access data in the
biomedical literature, 2015–2017

N = 149 randomly selected articles

Joshua D. Wallach^{1,2}, Kevin W. Boyack³, John P. A. Ioannidis^{4,5,6,7,8 *}

Among the 104 articles with empirical data in which protocols or data sharing would be pertinent, **19** (18.3% [11.6% to 27.3%]) discussed publicly available data; only one (1.0% [0.1% to 6.0%]) included a link to a full study protocol.

Analytical flexibility

Same Data, Different Conclusions

Twenty-nine research teams were given the same set of soccer data and asked to determine if referees are more likely to give red cards to dark-skinned players. Each team used a different statistical method, and each found a different relationship between skin color and red cards.

Referees are
three times as
likely to give red
cards to
dark-skinned
players

Statistically
significant results
showing referees are
more likely to give red
cards to dark-skinned
players

Twice as likely

Equally likely

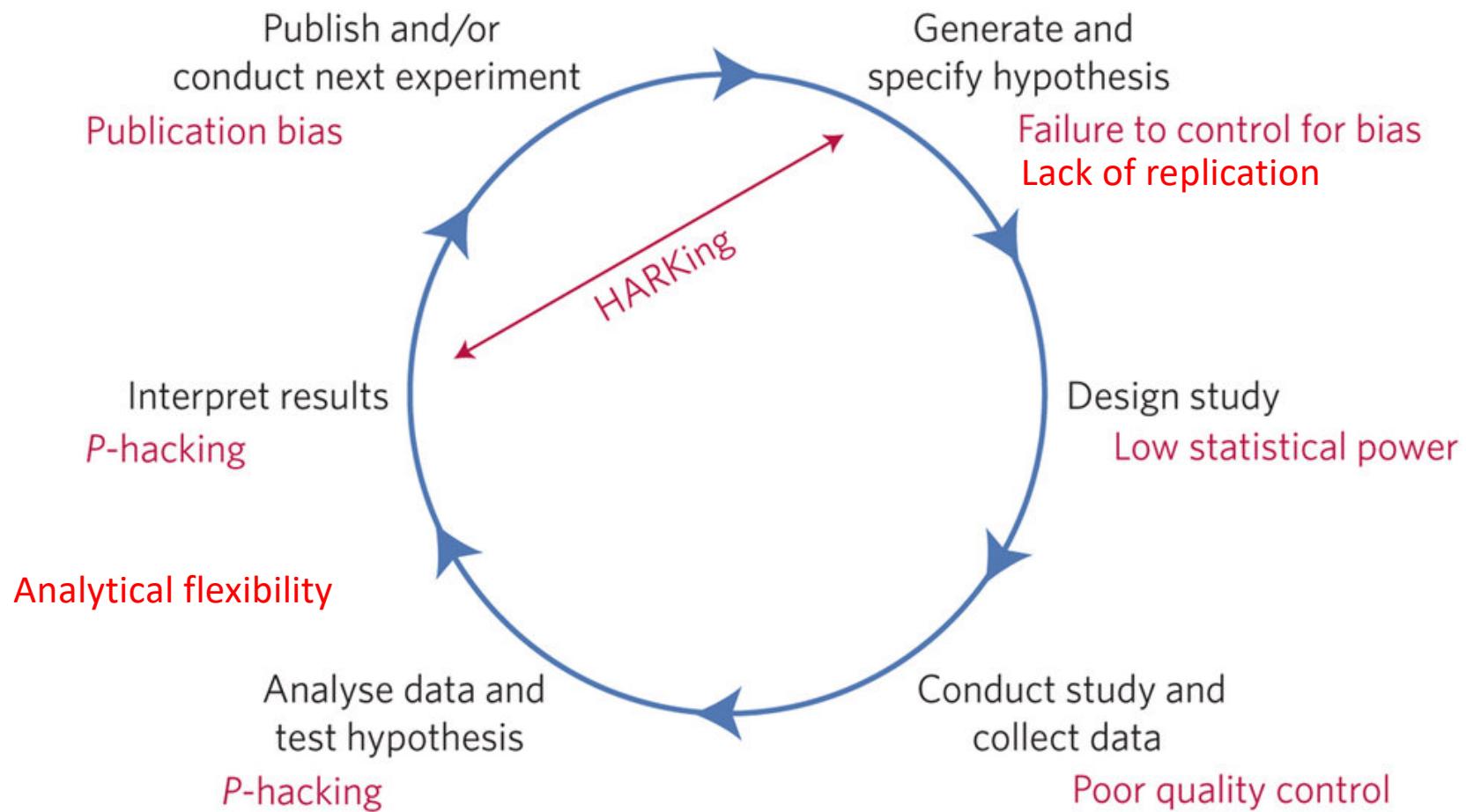
ONE RESEARCH TEAM

95% CONFIDENCE INTERVAL

Non-significant
results

Note distinction between Confirmatory studies vs Exploratory studies

Sources of the problem



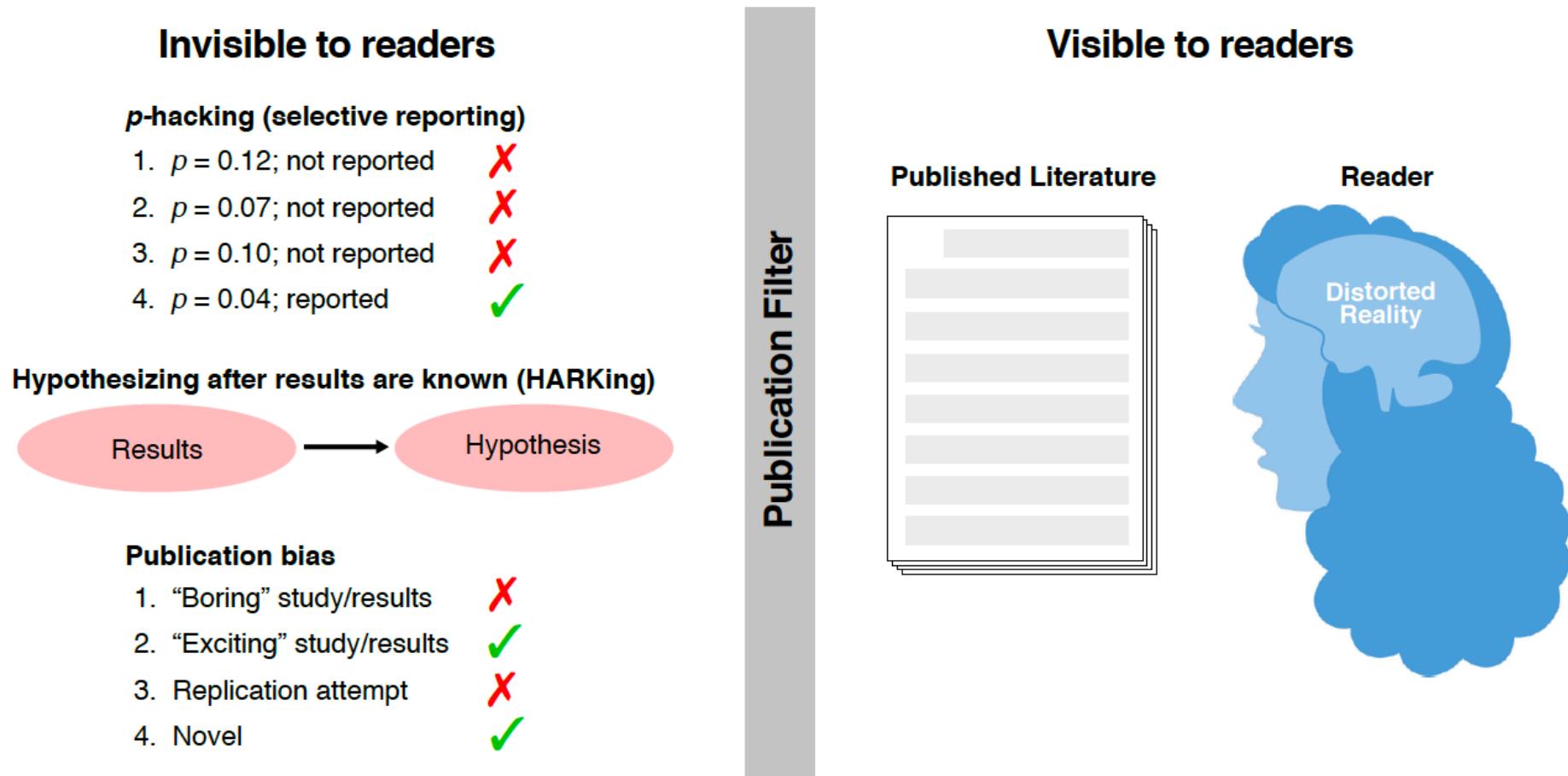


Figure 1: *Researchers' Distorted View of Reality*.

Researchers carry out numerous studies and perform many statistical tests, but not all of them are reported or published. Moreover, those results that are reported are not necessarily hypothesized *a priori*. These biases act as a filter, which distorts the findings present in the published literature, providing readers (researchers) with a distorted view of reality.



A solution: practice Open Science



The workflow is key

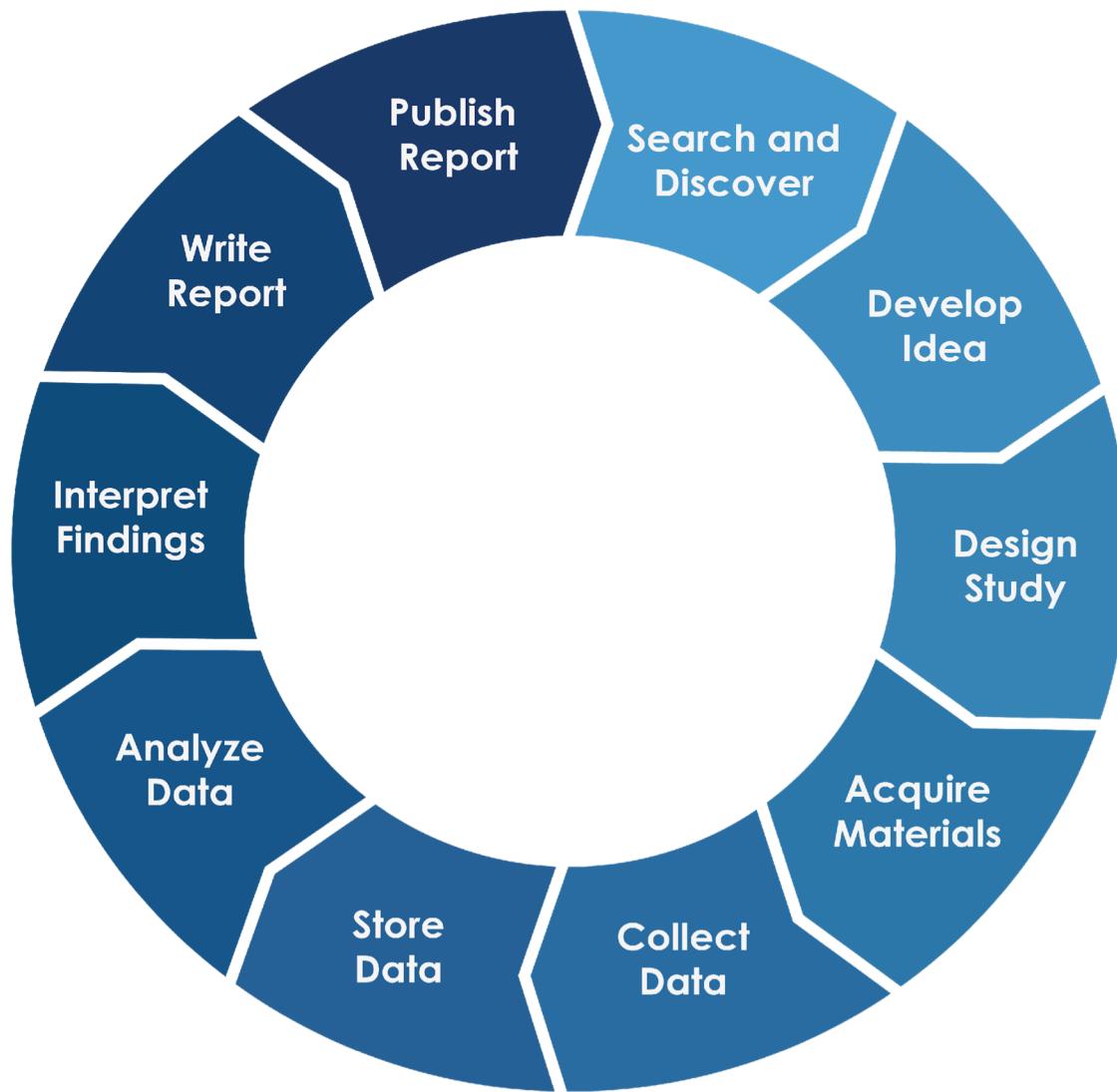
Vast majority of the scientific workflow is obscured*

As a result:

- Hard to reproduce others' work
- Hard to reproduce our own work!
- Difficult to accumulate unpublished knowledge, or to use published results for additional analyses (e.g. meta-analyses)

* Some scientific disciplines (e.g. physics) better than others

Typical workflow



Ideal workflow

Stage 2
Peer Review



Stage 1
Pre-registration &
Peer Review
"Registered report"

Ideal workflow

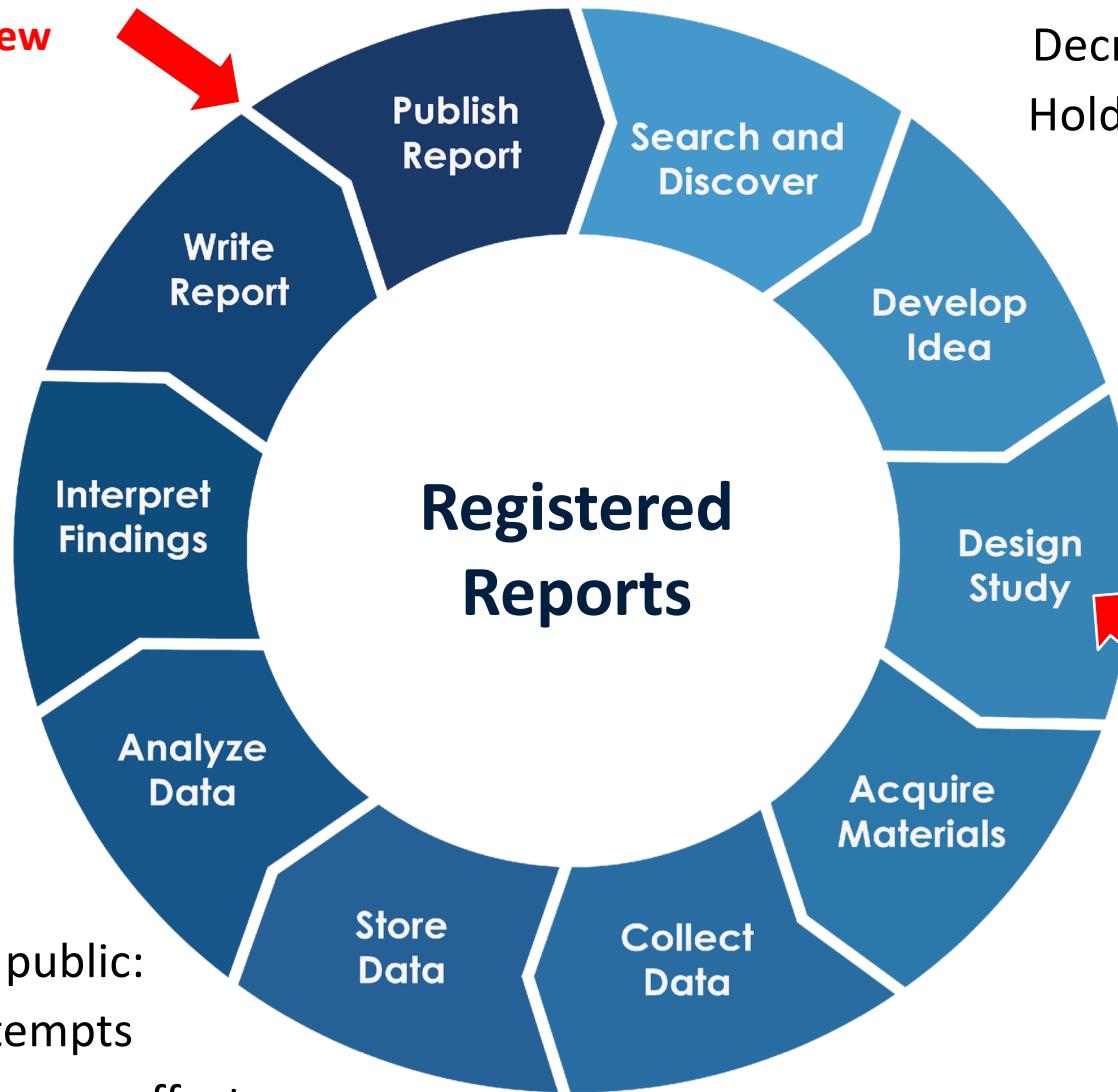
Stage 2
Peer Review



Decreases researcher d.f.
Holds you accountable to
yourself and others

Ideal workflow

Stage 2
Peer Review



Decreases researcher d.f.
Holds you accountable to
yourself and others

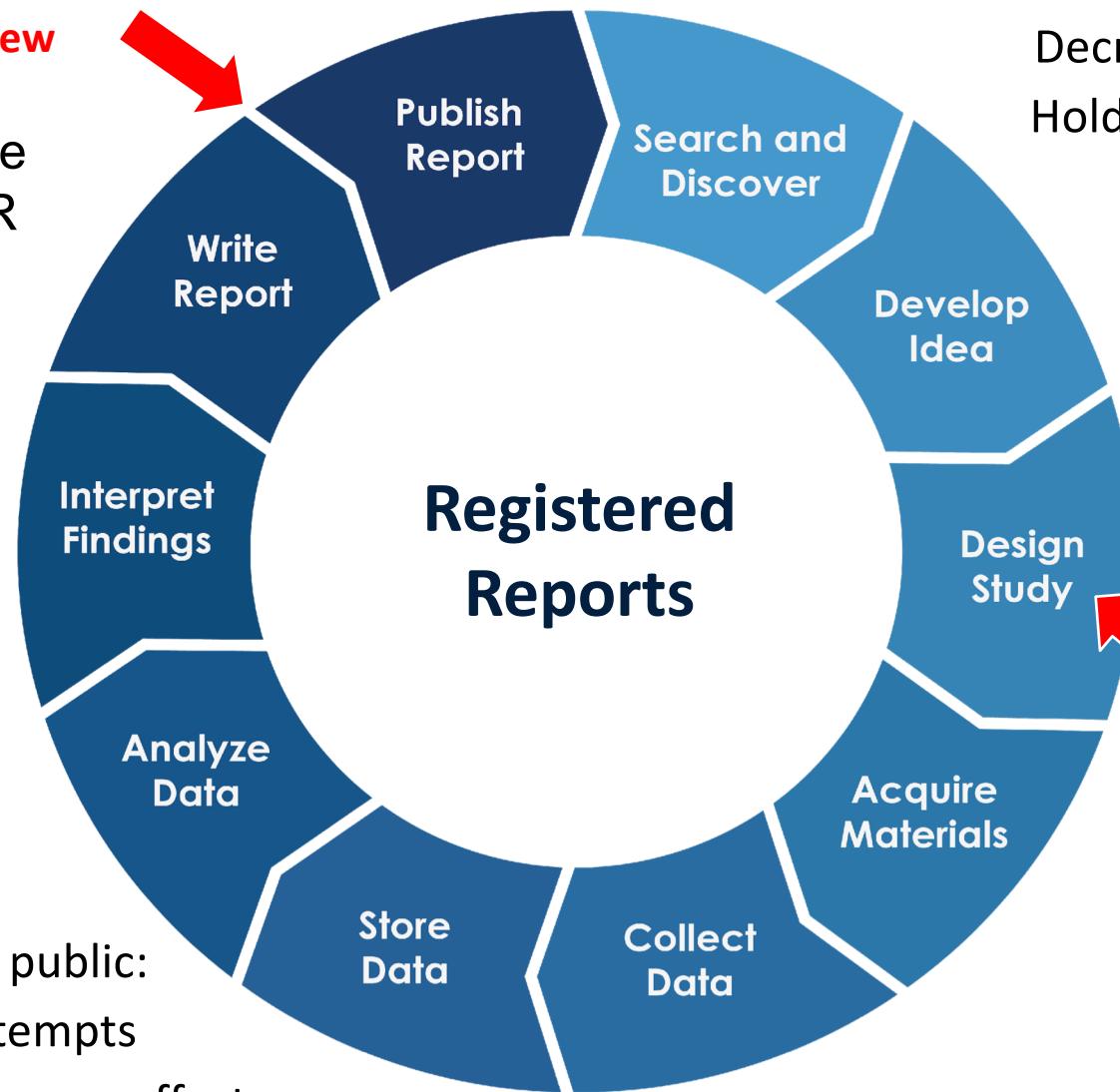
If registration made public:
Helps replication attempts
Decreases the file drawer effect

Ideal workflow

Stage 2

Peer Review

207 journals have
some form of RR
adoption!



Decreases researcher d.f.
Holds you accountable to
yourself and others

If registration made public:
Helps replication attempts
Decreases the file drawer effect

Ideal workflow

See [here](#) for list of current journals offering
some form of registered report



Example registered reports

**ROYAL SOCIETY
OPEN SCIENCE**

royalsocietypublishing.org/journal/rsos

Registered report



Cite this article: Przybylski AK, Weinstein N. 2019 Violent video game engagement is not associated with adolescents' aggressive behaviour: evidence from a registered report. *R. Soc. open sci.* **6**: 171474.

<http://dx.doi.org/10.1098/rsos.171474>

Violent video game engagement is not associated with adolescents' aggressive behaviour: evidence from a registered report

Andrew K. Przybylski^{1,2} and Netta Weinstein³

¹Oxford Internet Institute, University of Oxford, Oxford OX1 3JS, UK

²Department of Experimental Psychology, University of Oxford, Oxford, UK

³School of Psychology, Cardiff University, Cardiff, UK

AKP, 0000-0001-5547-2185

Example registered reports

[Reproducibility project: Cancer Biology](#)

[Estimating the reproducibility of psychological science](#)

All contributing studies include registered reports.

Open science lite

Central to most definitions of open science:

- Research workflow is transparent, open, and thoroughly documented from the start
- Annotated scripts are shared
- The data themselves are open*

* This is not strictly necessary to ensure reproducibility

Open science lite

Central to most definitions of open science:

- Research workflow is transparent, open, and thoroughly documented from the start
- Annotated scripts are shared
- The data themselves are shared

On their own, these features go a long way towards ensuring reproducibility of research.

Open science lite

Central to most definitions of open science:

- Research workflow is transparent, open, and thoroughly documented from the start
- Annotated scripts are shared
- The data themselves are shared or at least accessible

On their own, these features go a long way towards ensuring reproducibility of research.

Even better:

- Pre-register study (different from full registered report)

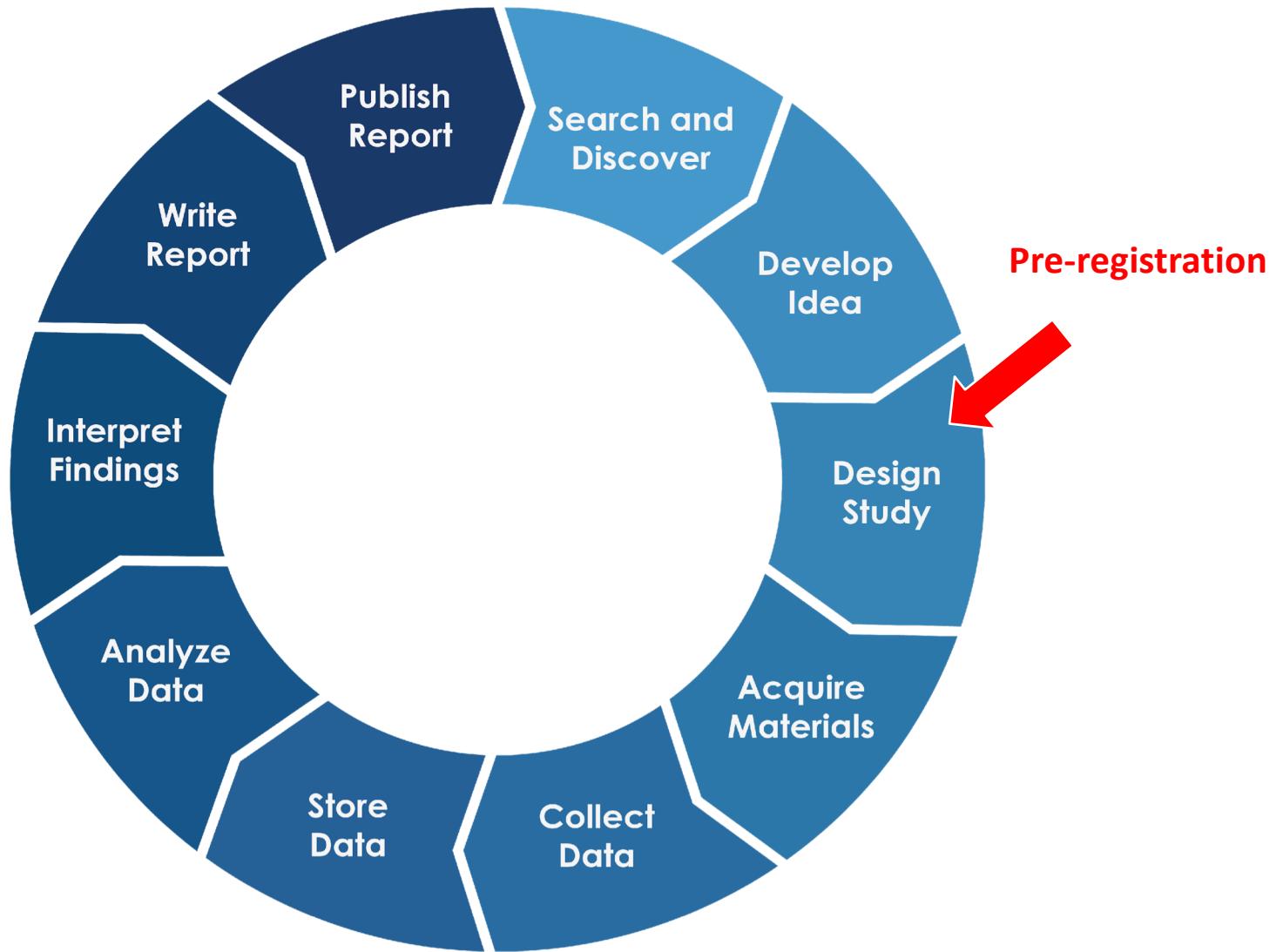
Pre-registration

Documenting your research plan in a read-only public repository before you conduct the study

At a minimum, a pre-registration should include the “what” of a study:

- Research question
- **Target population and sample size based on power analyses or, better, estimates of Type M error**
- General design
- Variables to be measured / dataset you’ll be using

Workflow for the masses



Pre-registration

Helps reduce publication bias, specifically the “file drawer effect”, by increasing discoverability of unpublished studies.

Helps you as a researcher plan ahead more carefully, and fosters greater awareness of pitfalls associated with researcher degrees of freedom.

Focuses attention on effect sizes and reducing type M errors



What good is Open Science to me personally?



POINT OF VIEW

McKiernan et al. eLife 2016;5:e16800. DOI: 10.7554/eLife.16800

How open science helps researchers succeed

Abstract Open access, open data, open source and other open scholarship practices are growing in popularity and necessity. However, widespread adoption of these practices has not yet been achieved. One reason is that researchers are uncertain about how sharing their work will affect their careers. We review literature demonstrating that open research is associated with increases in citations, media attention, potential collaborators, job opportunities and funding opportunities.

These findings are evidence that open research practices bring significant benefits to researchers relative to more traditional closed practices.

DOI: 10.7554/eLife.16800.001

**ERIN C MCKIERNAN*, PHILIP E BOURNE, C TITUS BROWN, STUART BUCK,
AMYE KENALL, JENNIFER LIN, DAMON McDougall, BRIAN A NOSEK,
KARTHIK RAM, COURTNEY K SODERBERG, JEFFREY R SPIES, KAITLIN THANNEY,
ANDREW UPDEGROVE, KARA H WOO AND TAL YARKONI**

Benefits of adopting Open Science

Improves Efficiency

- Saves time and \$\$
- Re-use methods / code
- Avoids duplication while enabling replication
- Facilitates meta-analyses
- Promotes accurate discovery

Benefits of adopting Open Science

- Funding agencies (e.g. Tri-Council) are increasingly demanding aspects of OS (e.g. open data, open access)
- OS practices are increasingly being favoured at academic institutions
- Many aspects of OS will eventually be required

Tools that can help

- Open Science Framework (<https://osf.io>)
- Scripting (e.g. using R instead of Excel! R Markdown)
- Git / Github (see lesson [here](#))
- Data management plan tools e.g. [here](#)

Scripting using R / R Markdown, Python etc...

Open Science Framework





https://people.ok.ubc.ca/jpither/modules/biol202_home.html

Starter Tutorials

Tutorial 00: Preparing and formatting assignments for submission

Tutorial 01: Visualizing and describing a single variable

Tutorial 02: Visualizing associations between two variables

Tutorial 03: Calculating descriptive statistics for a numeric variable grouped by a categorical variable

Tutorial 04: Sampling, Estimation, and Uncertainty

Tutorial 05: Random trials

Tutorial 06: Hypothesis_testing

Tutorial 07: Estimating proportions

Tutorial 08: Binomial distribution and binomial test

Tutorial 09: Goodness of fit tests

Tutorial 10: Odds Ratio

Tutorial 11: Contingency Analysis

Tutorial 12: The normal distribution

Tutorial 13: Comparing one mean to a hypothesized value

Tutorial 14: Comparing two means

Tutorial 15: Comparing means among more than two groups using ANOVA

Tutorial 16: Correlation analysis

Tutorial 17: Regression analysis

Markdown files for the tutorials

Extra tutorials

Master list of functions by lab

BIOL202: Introduction to Biostatistics

This page was last updated on September 17, 2019.

It is continually being updated, so be sure to refresh the page in your browser each time you visit!

You should download copies of these helpful **cheatsheets** and have them on hand:

- [RStudio cheatsheet](#)
- [R Markdown cheatsheet](#)

These cheat sheets deal with packages that will increasingly be used in the tutorials:

- [ggplot2 cheatsheet](#)
- [dplyr cheatsheet](#)

Starter Tutorials

These tutorials teach you the fundamentals of R and R Markdown, and should be completed prior to attempting any subsequent tutorials.

Introduction to R & RStudio

- What is R and RStudio?
- Installing R and RStudio
- How do I code in R?
- What are "packages"?
- Additional resources for learning R and RStudio

Reproducible R with R Markdown **Updated:** Friday Sept 7, 3pm: I clarified instructions on creating new projects in RStudio

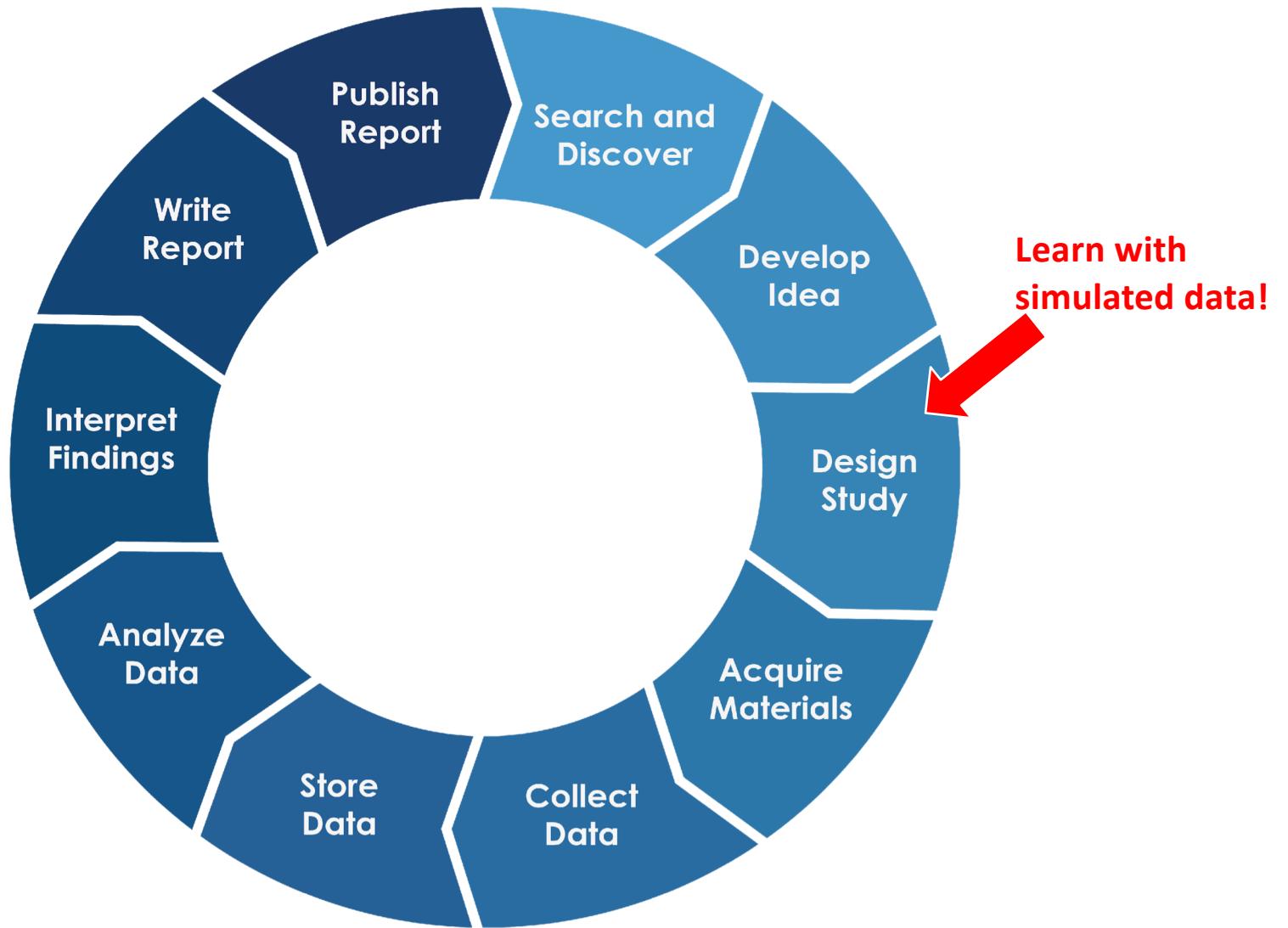
- Creating a reproducible lab report
- What is R Markdown?
- Workflow: create a project and R Markdown document
- More R Markdown information

Importing data into R **Work in progress**

Tutorial 00: Preparing and formatting assignments for submission

This tutorial provides instructions on how to prepare your assignments for submission.

Tutorial 01: Visualizing and describing a single variable





OSF HOME ▾

BLERF_lab(OSF)

Files Wiki Analytics Registrations Contributors Add-ons Settings

My Quick Files My Projects Search Support Donate Jason Pither



Private Make Public P 0 ...

BLERF_lab(OSF)

Contributors: Jason Pither, Carmen Chelick, Bruno Carturan, Brian Muselle, Faye Manning, Corrina Thomsen, Emma Corbett, Mackenzie Repovs, Nicholas Tochor, Kayleigh Nielson

Date created: 2017-01-20 04:22 PM | Last Updated: 2019-02-13 02:06 PM

Category: Project

Description:

For all BLERF lab activities

License: Add a license

Wiki

This OSF site hosts all materials and information relevant to the activities of the Biodiversity and Landscape Ecology Research Facility (BLERF) at the Okanagan campus of the University of British Columbia.

All active members of BLERF are free to add / edit / update material on this site. For instance, members are encouraged to post useful new RStudio tools or amazing new publications on this sit...

[Read More](#)

Files

Click on a storage provider or drag and drop to upload



Name ▾ ▾

Modified ▾ ▾

- BLERF_lab(OSF)
- OSF Storage (United States)
- + Lab meetings
- + Lab R tutorials
- + Lab Resources
- + Lab questions
- Lab Live Documents
- OSF Storage (United States)
 - BLERF_student_presentations.docx 2017-09-14 11:31 AM
- + Data_sharing
- Collaborations

Citation

Components

Lab meetings

Pither, Chelick, Carturan & 7 more

Lab R tutorials

Pither, Chelick, Carturan & 7 more

This holds subdirectories to hold material for each tutorial

Lab Resources

Pither, Chelick, Carturan & 13 more

Here you'll find general resources relevant to career development or lab activities

Lab questions

Pither, Chelick, Carturan & 7 more

A word document with running list of questions to be addressed

Lab Live Documents

Pither, Chelick, Carturan & 7 more

This is where BLERF members will update presentation / communication / publication information.

Data_sharing

Pither

Collaborations

Pither & Estaki

A web app to support:

Collaboration Documentation Archiving

OSF HOME ▾

Materials needing supervisor feedback Files Wiki Analytics Registrations Contributors Add-ons Settings

Notice: The site will undergo maintenance between Apr 19, 2018 6:00 PM and Apr 19, 2018 8:00 PM (-0700 UTC). Thank you for your patience. X

Faye Manning - Research Proposal.docx (Version: 28)

Check out Delete Download View Revisions

Revisions						
Version ID	Date	User	Download	MD5	SHA2	
28	2018-04-18 03:19 PM	Faye Manning	0 	 065ad8c2c0ef6f9329b1	 61446b6f2b088bd2f73	
27	2018-04-13 01:03 PM	Faye Manning	0 	 632331ea46d2e05097e	 a3dfe75689914d78007	
26	2018-04-12 03:51 PM	Faye Manning	0 	 434940d4b6dada3283	 d546e426b54b58546f2	
25	2018-04-12 03:18 PM	Faye Manning	0 	 4ef064f17b32110689e	 502ffc31976759f9cecc1	
24	2018-03-24 09:58 AM	Jason Pither	0 	 d860bcfafe7f875aa98c	 e01a5a55180a5bd53c0	
23	2018-03-06 04:48 PM	Faye Manning	0 	 846da1c16f2b8aae592	 c2c4221adfb70b3c12b	
	2018-02-26 02:07 PM	Faye Manning	0 	 ce5feeee890118d14365	 a6f485322c6ced7a8e4	

Materials needing supervisor feedback OSF Storage Faye Manning - Research Pro... Tags Add a tag to enhance discoverability

Automatic file versioning

Wiki



This is a public repository containing all information and data relevant to the following article that has been accepted for publication at the **New Phytologist**:

Belowground biotic interactions moderated the post-glacial range dynamics of trees

Jason Pither, Brian J. Pickles, Suzanne W. Simard, Alejandro Ordóñez, John W. Williams

Citation

osf.io/vuhe8 ^

APA

Pither, J., & Pickles, B. J. (2018, April 12). Belowground biotic interactions moderated the post-glacial range dynamics of trees.
<http://doi.org/10.17605/OSF.IO/VUHE8>

MLA

Pither, Jason, and Brian J Pickles. "Belowground Biotic Interactions Moderated the Post-Glacial Range Dynamics of Trees." Open Science Framework, 12 Apr. 2018. Web.

... .

Persistent citable identifiers

The screenshot shows the OSF (Open Science Framework) homepage. At the top, there is a navigation bar with links for "OSF HOME", "My Quick Files", "My Projects", "Search", "Support", "Donate", and a user profile for "Jason Pither". Below the navigation bar, there is a search bar and a message about maintenance. The main content area displays a project titled "Promoting School Belongingness and Academic Performance: A Multisite Effectiveness Trial of a Scalable Student Mindset Intervention". On the right side of the project page, there is a large button labeled "Registrations". A red arrow points from the "Registrations" link in the top navigation bar to this button. The "Registrations" section contains a "Citation" link (osf.io/p485e) and a "Recent Activity" section showing a file addition by Geoffrey Borman.

Promoting School Belongingn... Files Wiki Analytics Registrations

Notice: The site will undergo maintenance between Apr 19, 2018 6:00 PM and Apr 19, 2018 8:00 PM (-0700 UTC). Thank you for your patience. X

Promoting School Belongingness and Academic Performance: A Multisite Effectiveness Trial of a Scalable Student Mindset Intervention

Registrations

Contributors: [Geoffrey Borman](#), [Jon Baron](#)

Date created: 2016-10-12 12:32 PM | Last Updated: 2017-02-21 11:22 AM

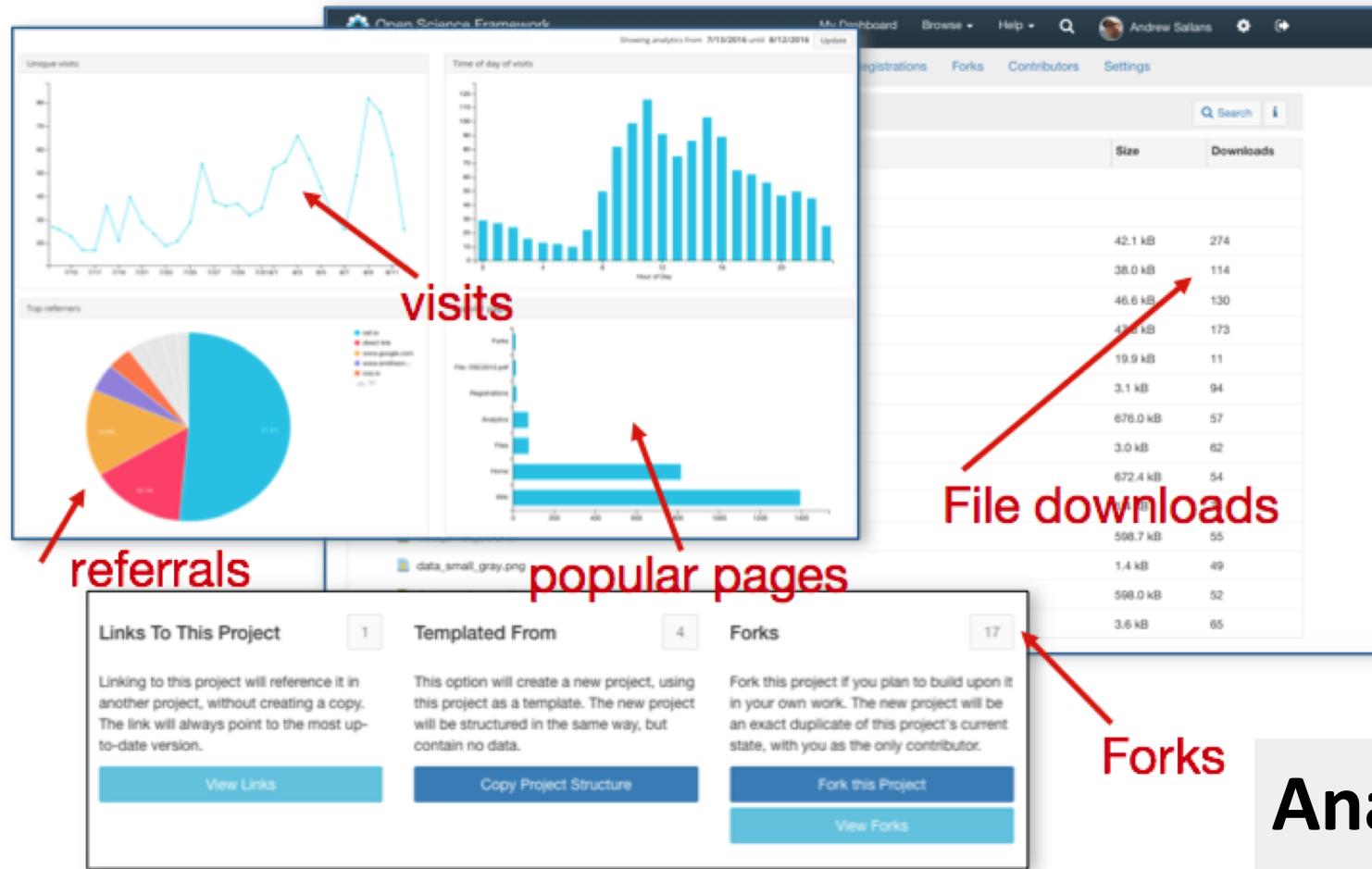
Category: [Project](#)

Files	
Filter i	
Name	Modified
- Promoting School Belongingness an...	
- OSF Storage	
OSF Borman Final Study Plan....	2016-10-12 01:3...

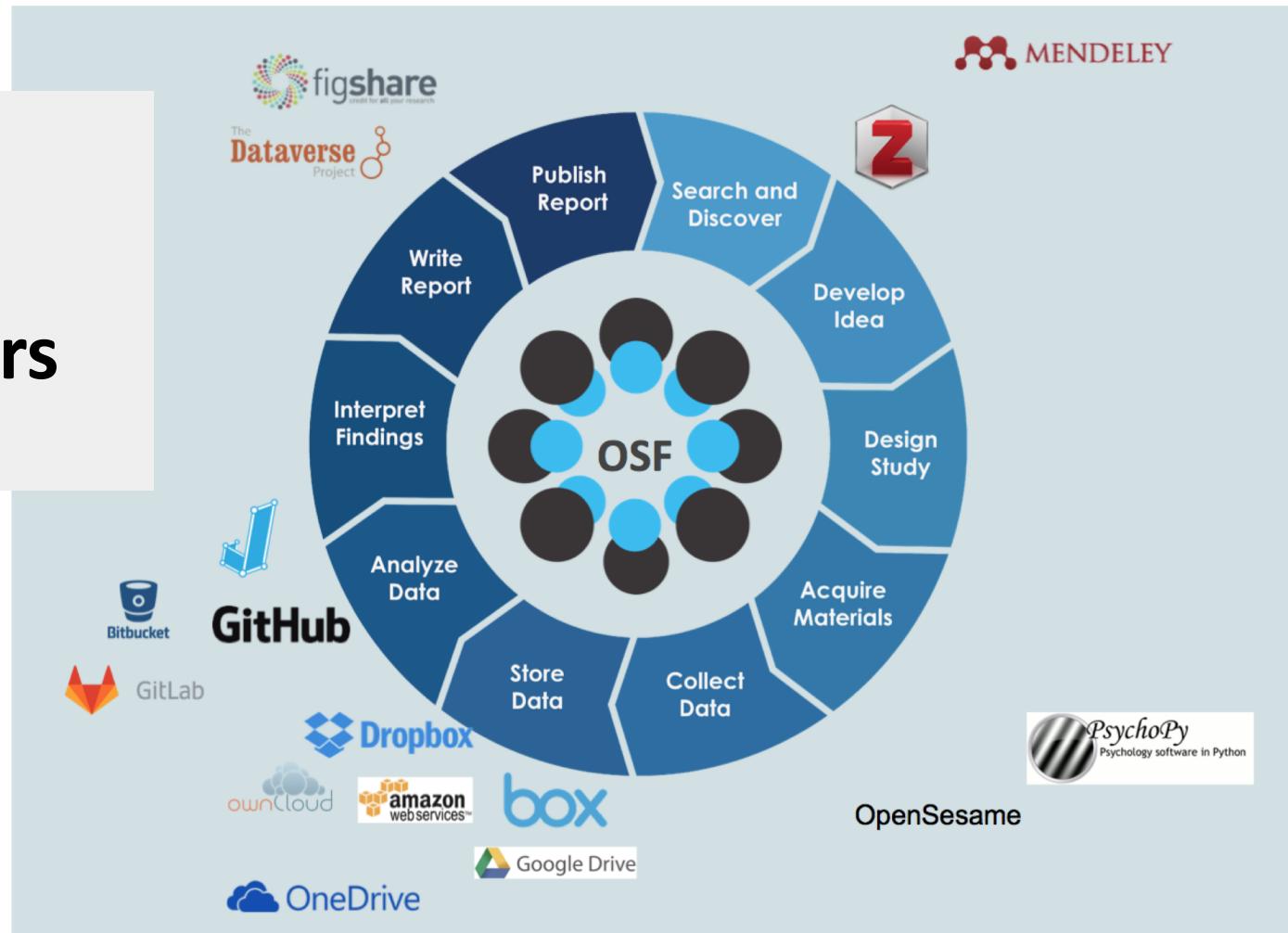
Citation	
osf.io/p485e	

Recent Activity

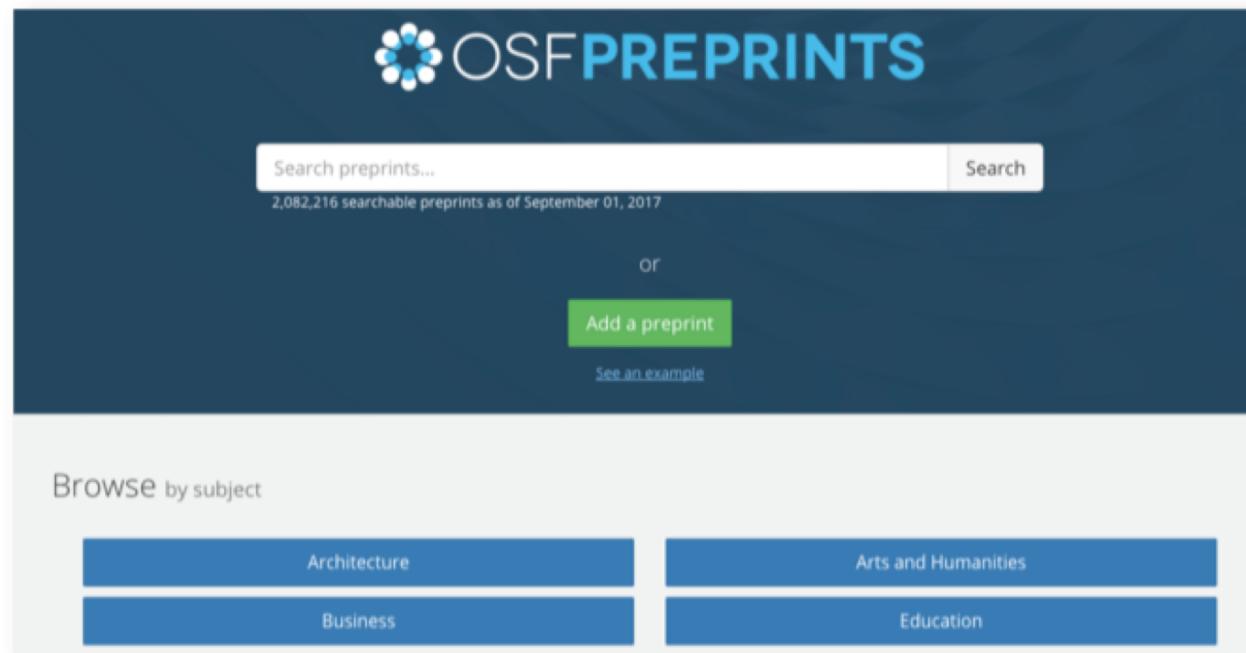
Geoffrey Borman added file [OSF student measures2.docx](#) to OSF Storage in Promoting School Belongingness and Academic Performance: A Multisite Effectiveness Trial of a Scalable Student Mindset Intervention 2017-02-21 11:22 AM



Connects services researchers use



Preprint services enhance research visibility





Open Science

@



Excellence fund strategic initiative:

FOSTERING OPEN SCIENCE AT UBC

This project targets research support and educational initiatives at the undergraduate and graduate levels in order to integrate Open Science practices. Open Science represents an approach to research that emphasizes reproducibility, transparency and accessibility. This project has two strategic goals: first, to establish the expertise and infrastructure required to foster the practice of Open Science among willing members of the entire UBC research community at both campuses. Secondly, to ensure that the core tenets of Open Science are of second nature to graduates of UBC's undergraduate and graduate programs.

Related strategy(ies): Strategy 8: Student Research, Strategy 9: Knowledge Exchange, Strategy 10: Research Culture, Strategy 12: Program Redesign and Strategy 14: Practical Learning

Project sponsor: Provost and Vice-President Academic UBCV; Provost and Vice-Principal Academic UBCO; Associate Vice-President Research & Vice-Principal Research

Target end date: Winter 2020

Progress: On track

UBC-wide initiatives

- Funded by UBC Excellence Fund: strategic initiative
- Promoting best practices in OS throughout UBC community
- Library and Advanced Research Computing developing sustainable infrastructure and expertise to support Open Science and **OSF**
- Establishment of OSF Institutions for UBC (coming soon!)
- Providing regular workshops, seminars, outreach at both campuses
- Developing instructional material for credit - BIOLOGY
- Engaging with Faculty Association and administration to encourage collaborative policy development
- openscience.ubc.ca

Organize and increase visibility of your institutional research.

OSF Institutions is a scholarly web tool that enhances transparency, fosters collaboration, and increases the visibility of research outputs at leading research institutions. The [Center for Open Science](#) partners with these institutions to create central hubs for research projects on a branded, dedicated OSF page.

Single sign-on authentication creates a seamless and integrated framework that accommodates custom research workflows and streamlines institutional data management. You can focus your efforts on generating and sharing research, not on building and maintaining research infrastructure.

Key Benefits for Your Institution

- Collaborate at all levels of a project within the institution as well as outside collaborators
- Conveniently share and make whole projects or just parts of them public, and retain security and privacy for project elements that are not shareable because of ethical or proprietary considerations
- Archive and cite projects easily
- Connect 3rd party tools and services & eliminate silos or redundancy
- Provide visibility for ongoing and unpublished research across the entire institution
- Gain insight and data on research collaboration throughout your organization
- Evaluate impact of research investment beyond citations

[Get Started Now](#)

At UBC's Okanagan campus

Pilot project underway

Integrating best practices in Open Science into the Biology undergraduate program. Starting with small steps...

- Developing and deploying introductory instructional material regarding OS (currently in 1st year BIOL)
- Uploading / archiving digital photos of lab book notes to Canvas
- Teaching R Markdown in 2nd year core Biostats course (BIOL202)
- Encouraging use of Markdown for upper-year lab assignments
- Encouraging instructors to reward transparency and honesty

Longer-term goal

- Accreditation for undergrads and grads

Acknowledgments

- Dr. Eric Eich (Vice-Provost and AVP Academic Affairs)
- Heather Berringer (Chief Librarian, Okanagan campus)
- Sharon Hanna (UBCO librarian helping with OS initiatives)
- Centre for Open Science (<https://osf.io>)
- Wade Klaver (UBC ARC)
- Carmen Chelick and Brian Muselle (MSc Biology students)
- Department of Biology (Okanagan campus)
- Steve Cundy and UBC's Advanced Research Computing team
- UBC Excellence Fund



Extra slides

Discussion points:

- Early career researchers
- Promotion & Tenure
- How to effect change
- Lab practices
- Exploratory vs confirmatory data analysis
- Open Access publications
- eLife interactive articles