

Introduction to Reproducibility

Galin Jones

School of Statistics
University of Minnesota
galin@umn.edu
@JonesGalin

DSMMA Retreat 2020

Reproducibility

The screenshot shows the PLOS Medicine website interface. At the top, the browser address bar displays the URL: journals.plos.org/plosmedicine/article?id=10.1371/journal.pmed.0020124. The PLOS Medicine logo is prominently displayed in the header. Navigation links for BROWSE, PUBLISH, and ABOUT are visible. A search bar with the text 'SEARCH' and a magnifying glass icon is present, along with a link to 'advanced search'. Below the header, the article title 'Why Most Published Research Findings Are False' is shown, along with the author 'John P. A. Ioannidis' and the publication date 'Published: August 30, 2005'. A table of metrics is displayed on the right side of the article, showing 68,436 Saves, 3,690 Citations, 3,097,623 Views, and 10,664 Shares. A 'Download PDF' button is also visible. The article abstract is partially visible, starting with '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'.

Why Most Published Research Findings Are False

John P. A. Ioannidis

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

Article	Authors	Metrics	Comments	Media Coverage
68,436 Save	3,690 Citation	3,097,623 View	10,664 Share	

Download PDF

Print Share

Check for updates

Related PLOS Articles

has COMPANIONS

Why Current Publication Practices May Distort Science

[View Page](#) [PDF](#)

Why Most Published Research Findings Are False: Author's Reply to [Criticism and Controversy](#)

Figure 1: Your work doesn't count if it can't be reproduced.

Reproducibility

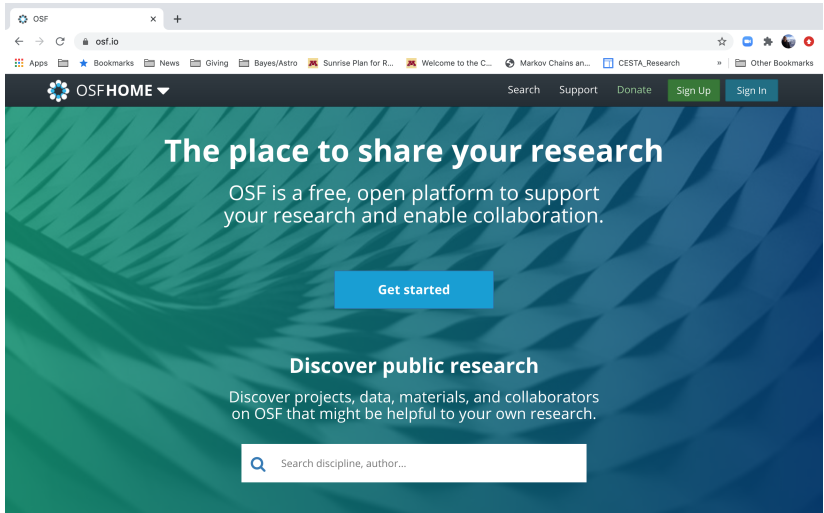
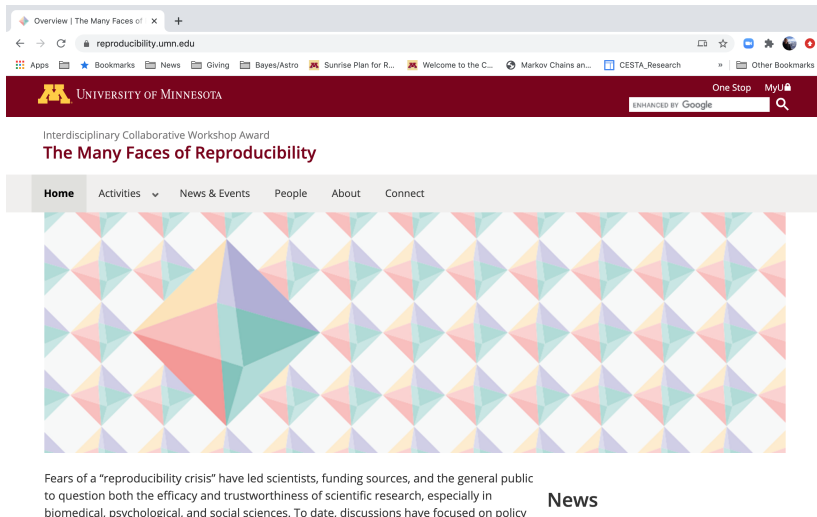


Figure 2: Open science—share everything

Reproducibility



The screenshot shows a web browser window with the URL `reproducibility.umn.edu`. The browser's address bar and tabs are visible at the top. Below the browser window is a dark red header bar featuring the University of Minnesota logo and name on the left, and navigation links for "One Stop" and "MyU" on the right. A search bar with the text "ENHANCED BY Google" is also present. The main content area has a light gray navigation bar with links for "Home", "Activities", "News & Events", "People", "About", and "Connect". Below this is a large banner image with a geometric pattern of overlapping triangles in shades of red, yellow, green, and purple. To the right of the banner, the word "News" is displayed in a bold, dark font. Below the banner, a paragraph of text discusses the "reproducibility crisis" and its impact on scientific research, particularly in biomedical, psychological, and social sciences.

Overview | The Many Faces of x +

← → ↻ reproducibility.umn.edu

Apps Bookmarks News Giving Bayes/Astro Sunrise Plan for R... Welcome to the C... Markov Chains an... CESTA_Research Other Bookmarks

UNIVERSITY OF MINNESOTA One Stop MyU

ENHANCED BY Google

Interdisciplinary Collaborative Workshop Award

The Many Faces of Reproducibility

Home Activities News & Events People About Connect

Fears of a "reproducibility crisis" have led scientists, funding sources, and the general public to question both the efficacy and trustworthiness of scientific research, especially in biomedical, psychological, and social sciences. To date, discussions have focused on policy

News

Figure 3: Everyone is welcome to participate.

Computational Reproducibility

A hallmark of the modern data science environment (industry and academia) is sharing code and data. Make everything available that is required to reproduce your computational results *exactly*.

It begins by writing your code so that it is easily read by someone else (future you for example). Free advice:

- Document your code.

- Use descriptive variable names.

- Format it to be easily read.

- Order functions for linear reading.

- Be consistent.

Computational Reproducibility

Reproducible data analysis and version control

Git/GitHub

Emacs/RStudio/Spyder,

Reproducible data

Data repositories

Dataverse

Reproducible dynamic report generation

R markdown/R Notebook/Jupyter/Pandoc

R Markdown

Here is a nice tutorial on markdown

<https://commonmark.org/help/tutorial/>

and here is the definitive guide

<https://bookdown.org/yihui/rmarkdown/>

Further Reading

R Open Science

<http://ropensci.github.io/reproducibility-guide/>

Dataverse <https://dataverse.org/>

Code and Data for the Social Sciences: A Practitioner's Guide

<http://web.stanford.edu/~gentzkow/research/CodeAndData.pdf>