

Transparency, Reproducibility, and the Credibility of Economics Research

Garret Christensen¹
Edward Miguel²

¹US Census Bureau

²University of California Berkeley

Any opinions and conclusions expressed herein are those of the author(s) and do not necessarily represent the views of the U.S. Census Bureau.

AFSE, June 2019

Slides available online at <http://www.github.com/garretchristensen/AFSEJune2019>

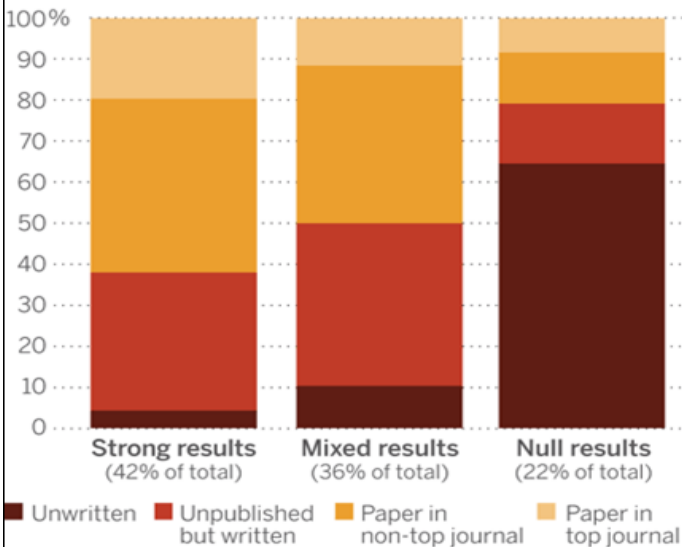
Reason to be Concerned

Publication Bias in All Fields

- ▶ Medicine: Turner et al. (2008)
- ▶ Social Sciences: Franco, Malhotra, Simonovits (2014)*
- ▶ Economics: Brodeur et al. (2016)*
- ▶ Sociology: Gerber and Malhotra (2008)
- ▶ Political Science: Gerber and Malhotra (2008)

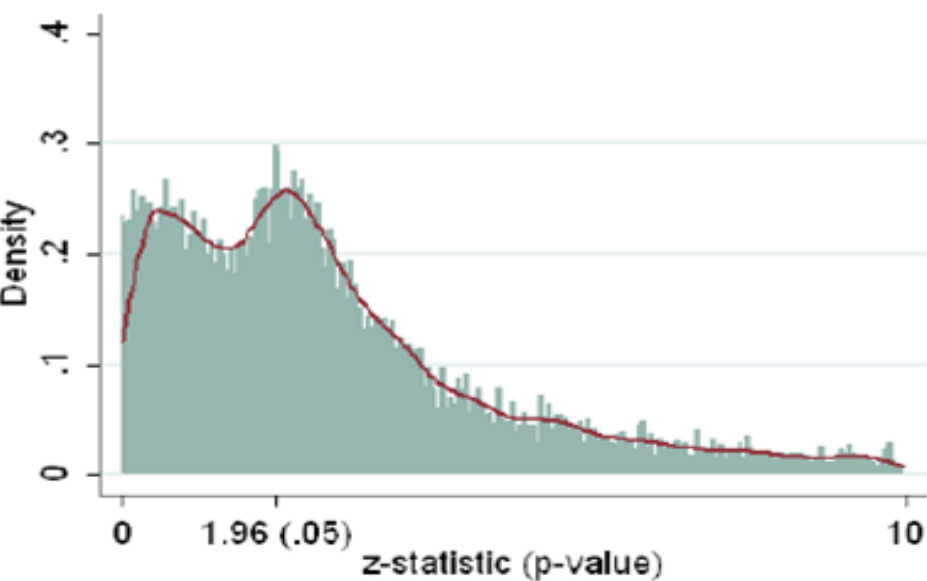
Most null results are never written up

The fate of 221 social science experiments



Source: A. Franco *et al.*, *Science* (28 August)

(b) Unrounded distribution of z-statistics.



Registration

Registration as Solution to Publication Bias:

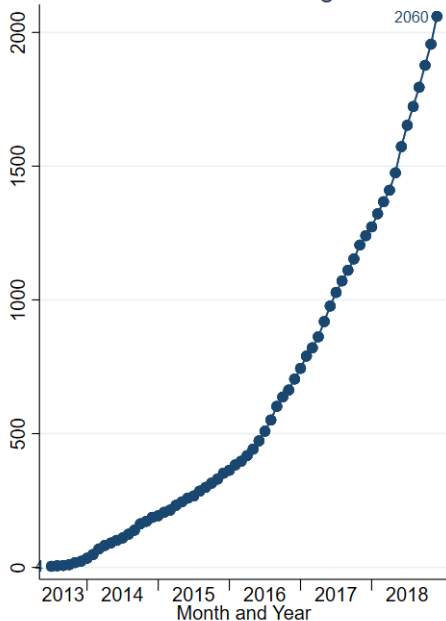
- ▶ Publicly stating all research you will do, what hypotheses you will test, prospectively.
- ▶ Near universal adoption in medical RCTs. Top journals won't publish if it's not registered.
<http://clinicaltrials.gov>
- ▶ Even better if registry requires outcomes from after study. (Some required by US National Institutes of Health.)

Registration

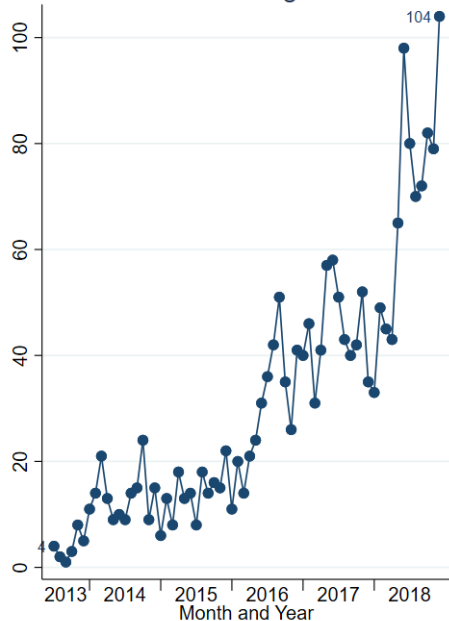
Newer to social sciences, but:

- ▶ American Economic Association (AEA), for randomized trials*. <http://socialscienceregistry.org>
- ▶ Evidence in Governance and Politics
<http://egap.org/design-registration>
- ▶ 3ie registry, for developing country evaluations.
<http://ridie.3ieimpact.org>
- ▶ Open Science Framework
<http://osf.io>
 - ▶ Open format
 - ▶ Will soon sync with above
- ▶ Simple: <http://aspredicted.org>

Panel A: Cumulative registrations



Panel B: New registrations



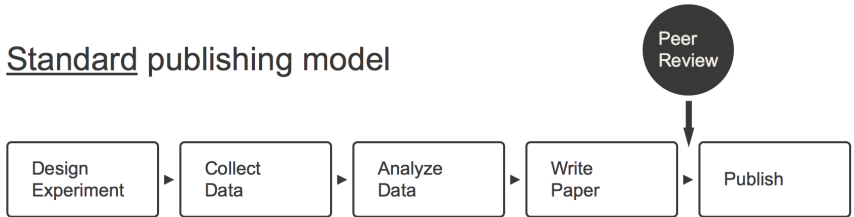
Design-Based Publication

AKA Registered Reports, moves peer review before data gathering, results, and analysis.

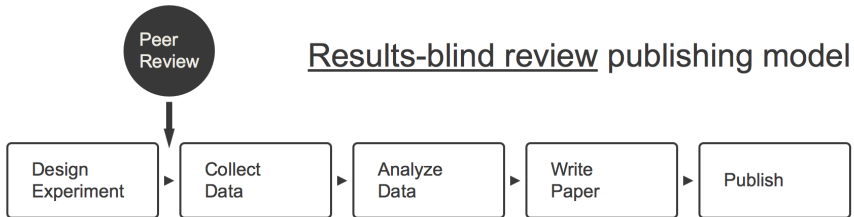
1. Design a project
2. Submit
3. Reviewed based on importance of question and quality of design
4. Get in-principle acceptance
5. Follow through, and nulls get published

75 Journals, 20 more with Special Issues [▶ Link](#)

Standard publishing model



Results-blind review publishing model



**Enter Manuscript Information**

Upload Files

Provide Additional Information

Review & Submit

[Guide for Authors](#)

Enter Manuscript Information

To begin, select an issue from the dropdown list.

Issue:

Select an article type from the dropdown list. Enter the required information into the fields that then appear.

Article Type:

- Full Length Article
- Registered Report Stage 2: Full Article
- Registered Report Stage 1: Proposal

P-Hacking

Define the problem:

- ▶ Also called fishing, researcher degrees of freedom, or data-mining.
- ▶ Definition: flexibility in data analysis allows portrayal of *anything* as below an arbitrary p-value threshold; significance loses its meaning.
- ▶ Not something only evil people do. It's subconscious, or simply built into statistics (Gelman, Loken 2013).

Pre-Analysis Plan

Explain the solution:

- ▶ From 3ie: “A pre-analysis plan is a detailed description of the analysis to be conducted that is written in advance of seeing the data on impacts of the program being evaluated. It may specify hypotheses to be tested, variable construction, equations to be estimated, controls to be used, and other aspects of the analysis. A key function of the pre-analysis plan is to increase transparency in the research. By setting out the details in advance of what will be done and before knowing the results, the plan guards against data mining and specification searching. Researchers are encouraged to develop and upload such a plan with their study registration, but it is not required for registration.”

Examples

- ▶ Sierra Leone Community-Driven Development: Casey et al. (2012)*
- ▶ Oregon Medicare: Finkelstein et al. (2012)
- ▶ J-PAL Hypothesis Registry (N=11): <http://www.povertyactionlab.org/Hypothesis-Registry>
- ▶ AEA Registry

Wide range of when exactly to write and how detailed to make the plan. At the extreme level of detail you would have your entire code already written before you got any data.

Outcome variable	(1) Mean for controls	(2) Treatment effect
Panel A: GoBifo “weakened” institutions		
Attended meeting to decide what to do with the tarp	0.81	−0.04 ⁺
Everybody had equal say in deciding how to use the tarp	0.51	−0.11 ⁺
Community used the tarp (verified by physical assessment)	0.90	−0.08 ⁺
Community can show research team the tarp	0.84	−0.12 [*]
Respondent would like to be a member of the VDC	0.36	−0.04 [*]
Respondent voted in the local government election (2008)	0.85	−0.04 [*]
Panel B: GoBifo “strengthened” institutions		
Community teachers have been trained	0.47	0.12 ⁺
Respondent is a member of a women’s group	0.24	0.06 ^{**}
Someone took minutes at the most recent community meeting	0.30	0.14 [*]
Building materials stored in a public place when not in use	0.13	0.25 [*]
Chiefdom official did not have the most influence over tarp use	0.54	0.06 [*]
Respondent agrees with “Responsible young people can be good leaders” and not “Only older people are mature enough to be leaders”	0.76	0.04 [*]
Correctly able to name the year of the next general elections	0.19	0.04 [*]

PAP—Observational Studies

- ▶ Debated in public health/epidemiology.
- ▶ Difficult, but not impossible, to verifiably pre-specify.
- ▶ Example: Government data releases
- ▶ Example: Minimum Wage (Neumark 2001*)
- ▶ Example: US, Italian Elections

The Employment Effects of Minimum Wages: Evidence from a Prespecified Research Design

DAVID NEUMARK*

This article presents evidence on the employment effects of recent minimum wage increases from a prespecified research design that entailed committing to a detailed set of statistical analyses prior to “going to” the data. The limited data to which the prespecified research design can be applied may preclude finding many significant effects. Nonetheless, the evidence is most consistent with disemployment effects of minimum wages for younger, less-skilled workers.

NOVEMBER 8, 2016. WHAT REALLY HAPPENED?

[REGISTER A DESIGN \(closed\)](#)[OFFICIAL RULES](#)[GET TWITTER UPDATES](#)

An opportunity for scholars studying elections:

- Preregister a research design for a study of the 2016 general election using ANES data *before* the data are publicly released. **UPDATE (3/31/17): ANES DATA RELEASED. THE REGISTRY IS CLOSED. THANK YOU FOR YOUR PARTICIPATION!**
1. **Submit an article including this design to a participating journal**, which will consider your submission *before* data are available.
 2. **Win a cash award for publishing your article.**

Replication

1. Lack of replicability in code:
 - ▶ 1986: JMCB Project*
 - ▶ 2018: Gertler, Galiani, Romero*
2. Also a lack of replicability in experiments: Camerer et al 2016*, Many Labs, Reproducibility Project: Psychology
3. Organizing Workflow
4. Code & Data Sharing

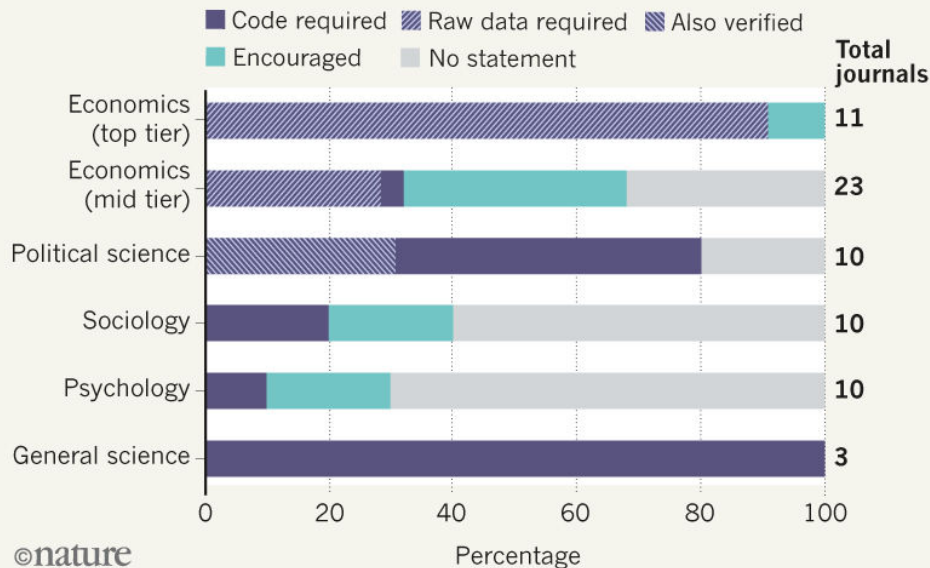
Replication in Empirical Economics: *The Journal of Money, Credit and Banking Project*

By WILLIAM G. DEWALD, JERRY G. THURSBY, AND RICHARD G. ANDERSON*

This paper examines the role of replication in empirical economic research. It presents the findings of a two-year study that collected programs and data from authors and attempted to replicate their published results. Our research provides new and important information about the extent and causes of failures to replicate published results in economics. Our findings suggest that inadvertent errors in published empirical articles are a commonplace rather than a rare occurrence.

DATA CHECKED?

In a survey of 67 journals, most of the political-science and top-tier economics titles required authors to submit software code and data to editors before publication. Journals in sociology and psychology rarely did so.

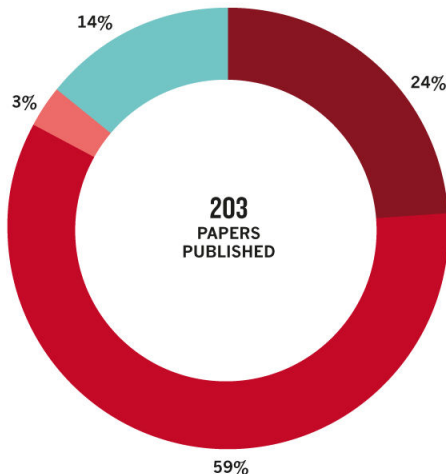


REPLICATION RARELY POSSIBLE

An analysis of 203 economics papers found that fewer than one in seven supplied the materials needed for replication.

ELEMENTS PROVIDED*:

- None
- One or more missing
- All, code doesn't run
- All, code runs



*The elements assessed were raw data, raw code, estimation data and estimation code.



An economics study featuring a performance by Robin Williams failed to replicate after the actor's death.

Bonnie Schiffman/Touchstone/The
Kobal Collection

About 40% of economics experiments fail replication survey

By **John Bohannon** | Mar. 3, 2016, 2:00 PM

Another social science looks at itself

Experimental economists have joined the reproducibility discussion by replicating selected published experiments from two top-tier journals in economics. Camerer *et al.* found that two-thirds of the 18 studies examined yielded replicable estimates of effect size and direction. This proportion is somewhat lower than unaffiliated experts were willing to bet in an associated prediction market, but roughly in line with expectations from sample sizes and P values.

Science, this issue p. 1433

Abstract

The replicability of some scientific findings has recently been called into question. To contribute data about replicability in economics, we replicated 18 studies published in the *American Economic Review* and the *Quarterly Journal of Economics* between 2011 and 2014. All of these replications followed predefined analysis plans that were made publicly available beforehand, and they all have a statistical power of at least 90% to detect the original effect size at the 5% significance level. We found a significant effect in the same direction as in the original study for 11 replications (61%); on average, the replicated effect size is 66% of the original. The replicability rate varies between 67% and 78% for four additional replicability indicators, including a prediction market measure of peer beliefs.

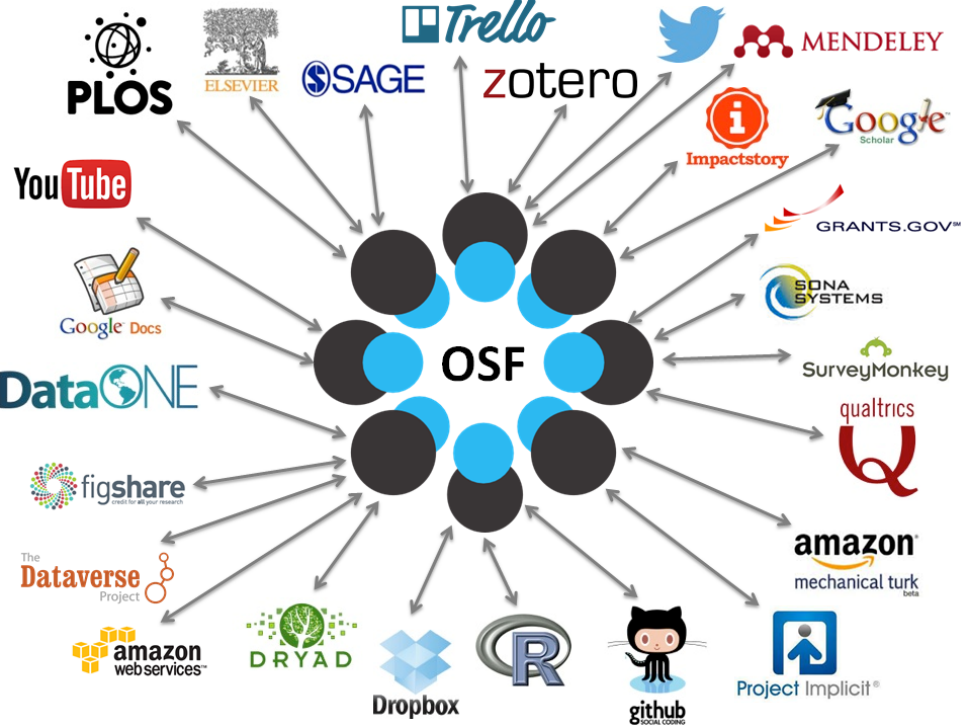
Workflow

- ▶ Literate programming (extensive commenting, making the aim of code reading by a human)
- ▶ Version Control
- ▶ Dynamic Documents

Version Control

- ▶ Using version control (AKA revision control) can help to make your work more reproducible.
- ▶ What is version control?
Version control is a system that records changes to a file or set of files over time so that you can recall specific versions later. For the examples in this book you will use software source code as the files being version controlled, though in reality you can do this with nearly any type of file on a computer.
—Git, About Version Control





Dynamic Documents

Write your code and your paper in the same file so you won't lose information or make copy and paste mistakes.

Possible in R and Stata.

- ▶ Include tables by linking to a file, instead of a static image.
- ▶ Include number by linking to a value calculated by an analysis file, instead of a static number typed manually.
- ▶ Automatically update tables and numbers.
- ▶ Produce entire paper with one or two clicks.

Data Sharing

Post your code and your data in a trusted public repository.

- ▶ Find the appropriate repository:

<http://www.re3data.org/>

- ▶ Repositories will last longer than your own website.
- ▶ Repositories are more easily searchable by other researchers.
- ▶ Repositories will store your data in a non-proprietary format that won't become obsolete.

Conclusion

OK, I'm convinced. How can I learn more?

- ▶ Christensen, Miguel (2018) J Econ Lit review. [▶ Link](#)
- ▶ Berkeley Initiative for Transparency in the Social Sciences (BITSS) [▶ Link](#)
- ▶ Center for Open Science (COS) [▶ Link](#)
- ▶ The Replication Network [▶ Link](#)
- ▶ Meta-Analysis in Economics Network (MAER-Net) [▶ Link](#)

Questions?

Thank you!