



BERKELEY INITIATIVE FOR TRANSPARENCY  
IN THE SOCIAL SCIENCES

Transparency,  
Reproducibil-  
ity, and the  
Credibility of  
Economics  
Research

Christensen

Introduction

Publication  
Bias

Registration

Specification  
Searching

Pre-Analysis Plans

Replicability

Questionable  
Research Practices

Project Protocol,  
Reporting Standards

Workflow

Data Sharing

Conclusion

# Transparency, Reproducibility, and the Credibility of Economics Research

Garret Christensen, Edward Miguel

UC Berkeley

YSI Plenary, October 2016  
Slides available online <http://www.github.com/BITSS/YSIPlenary2016>



# Outline

Transparency,  
Reproducibil-  
ity, and the  
Credibility of  
Economics  
Research

Christensen

Introduction

Publication  
Bias  
Registration

Specification  
Searching

Pre-Analysis Plans

Replicability

Questionable  
Research Practices

Project Protocol,  
Reporting Standards

Workflow

Data Sharing

Conclusion

## 1 Introduction

## 2 Publication Bias

### ■ Registration

## 3 Specification Searching

### ■ Pre-Analysis Plans

## 4 Replicability

### ■ Workflow

### ■ Data Sharing

## 5 Conclusion



BERKELEY INITIATIVE FOR TRANSPARENCY  
IN THE SOCIAL SCIENCES



# Publication Bias

Transparency,  
Reproducibil-  
ity, and the  
Credibility of  
Economics  
Research

Christensen

Introduction

Publication  
Bias

Registration

Specification  
Searching

Pre-Analysis Plans

Replicability

Questionable  
Research Practices

Project Protocol,  
Reporting Standards

Workflow

Data Sharing

Conclusion

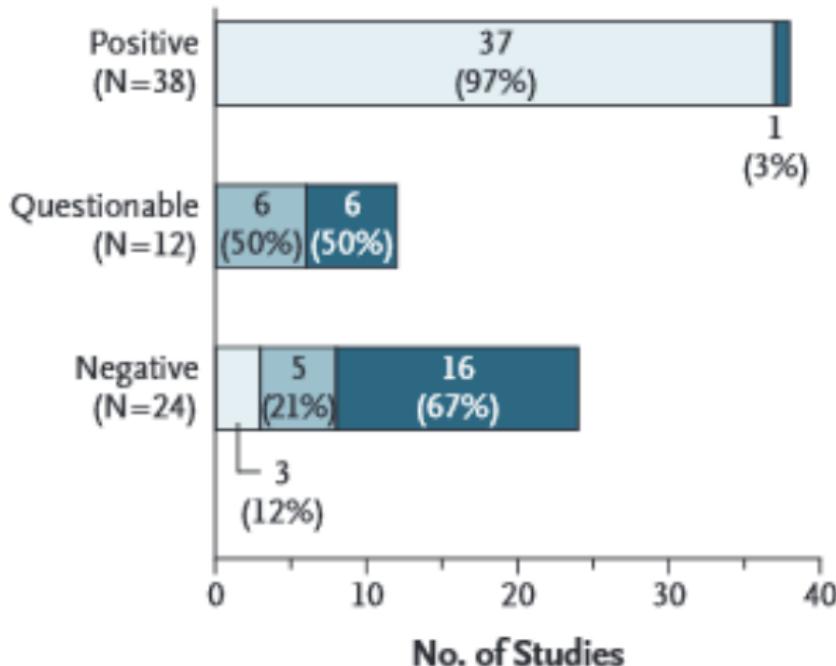
## All fields:

- Medicine: (Turner et al. 2008)
- Social Sciences: (Franco, Malhotra, Simonovits 2014)
- Economics: DeLong and Lang 1992, (Brodeur et al. 2016)
- Sociology: (Gerber and Malhotra 2008)
- Political Science: (Gerber and Malhotra 2008)

- Published, agrees with FDA decision
- Published, conflicts with FDA decision
- Not published

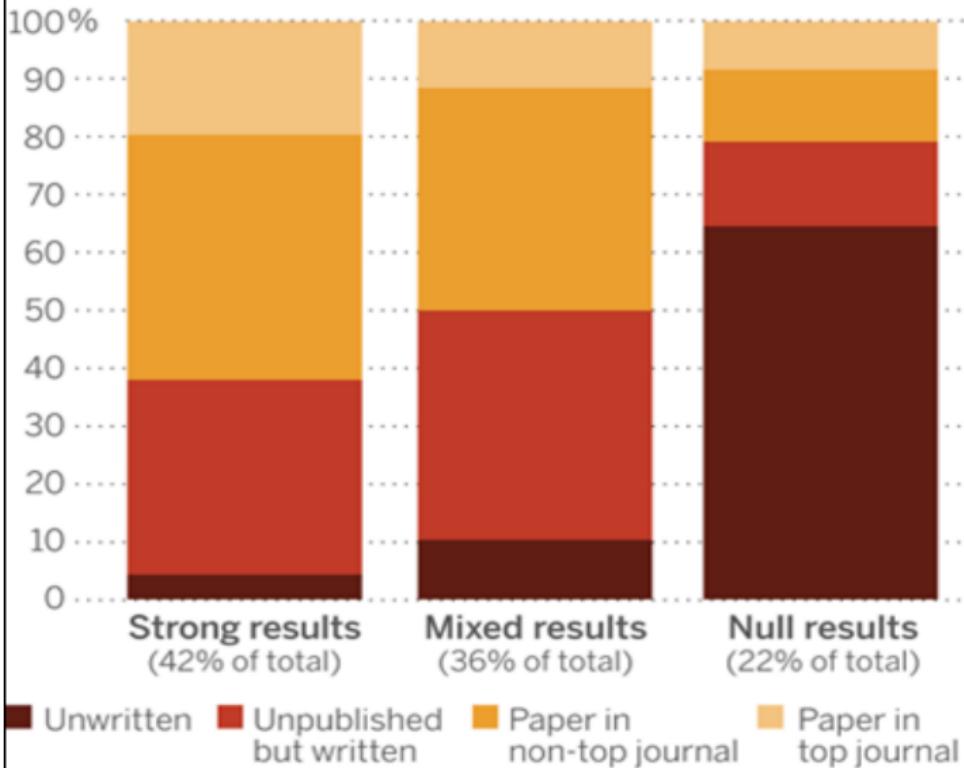
## A Studies (N=74)

### FDA Decision



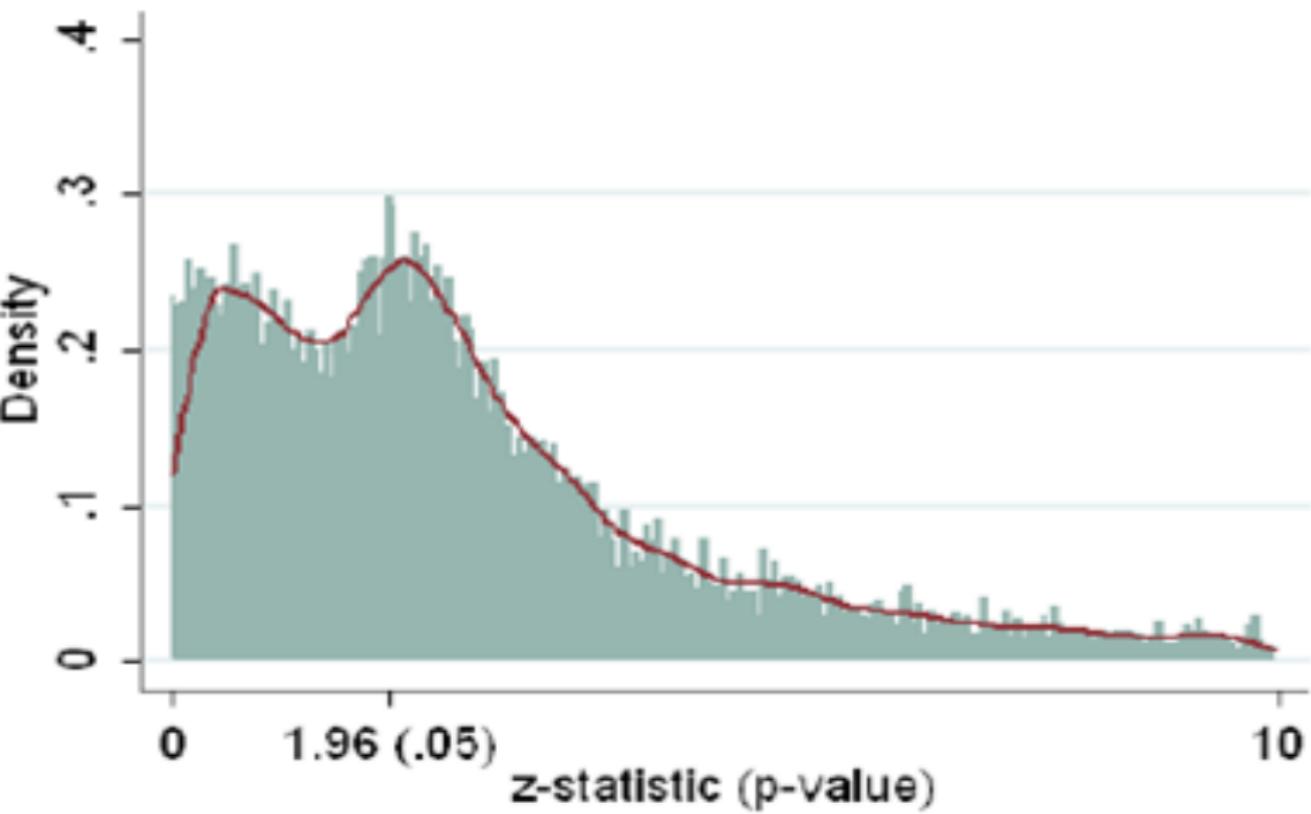
# Most null results are never written up

The fate of 221 social science experiments



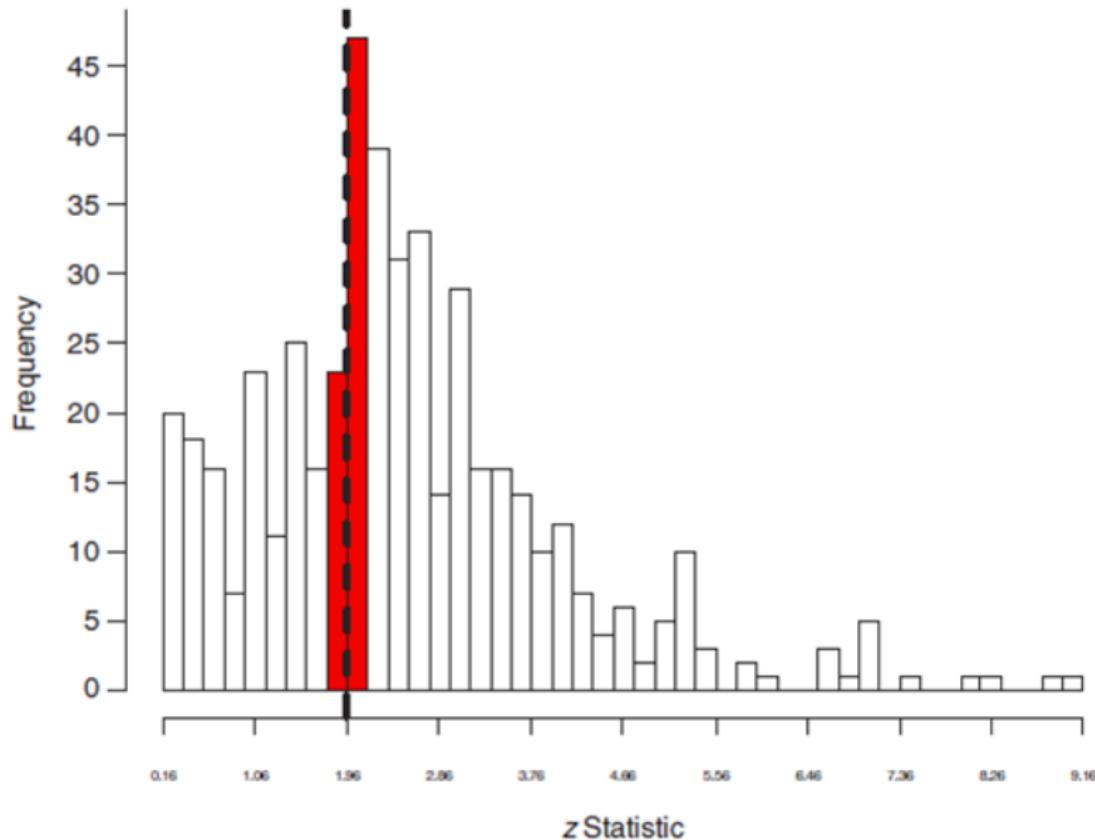
Source: A. Franco et al., *Science* (28 August)

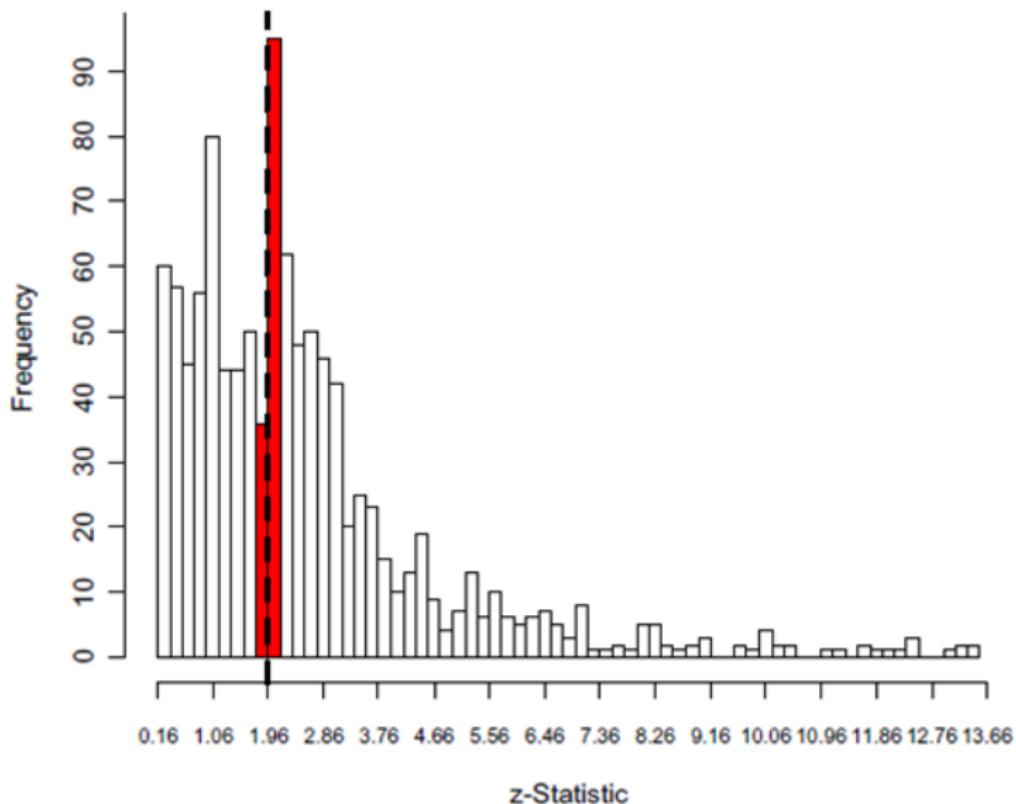
(b) Unrounded distribution of z-statistics.



Histogram of  $z$  Statistics From the *American Sociological Review*, the *American Journal of Sociology*, and *The Sociological Quarterly* (Two-Tailed)

---





**Figure 1(a).** Histogram of  $z$ -statistics, *APSR* & *AJPS* (Two-Tailed). Width of bars (0.20) approximately represents 10% caliper. Dotted line represents critical  $z$ -statistic (1.96) associated with  $p = 0.05$  significance level for one-tailed tests.



BERKELEY INITIATIVE FOR TRANSPARENCY  
IN THE SOCIAL SCIENCES

# Publication Bias

Transparency,  
Reproducibil-  
ity, and the  
Credibility of  
Economics  
Research

Christensen

Introduction

Publication  
Bias

Registration

Specification  
Searching

Pre-Analysis Plans

Replicability

Questionable  
Research Practices

Project Protocol,  
Reporting Standards

Workflow

Data Sharing

Conclusion

If we only write up/publish significant results, and we have no record of all the insignificant results, we have no way to tell if our ‘significant’ results are real, or if they’re the 5% we should expect due to noise.

# Registration

Transparency,  
Reproducibil-  
ity, and the  
Credibility of  
Economics  
Research

Christensen

Introduction

Publication  
Bias  
Registration

Specification  
Searching  
Pre-Analysis Plans

Replicability  
Questionable  
Research Practices  
Project Protocol  
Reporting Standards  
Workflow  
Data Sharing

Conclusion

## 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 (ICMJE) won't publish if it's not registered.  
<http://clinicaltrials.gov>
- Even better if registry requires outcomes from after study. Currently limited, but NIH is moving on this.



# Registration

Transparency,  
Reproducibil-  
ity, and the  
Credibility of  
Economics  
Research

Christensen

Introduction

Publication  
Bias  
Registration

Specification  
Searching  
Pre-Analysis Plans

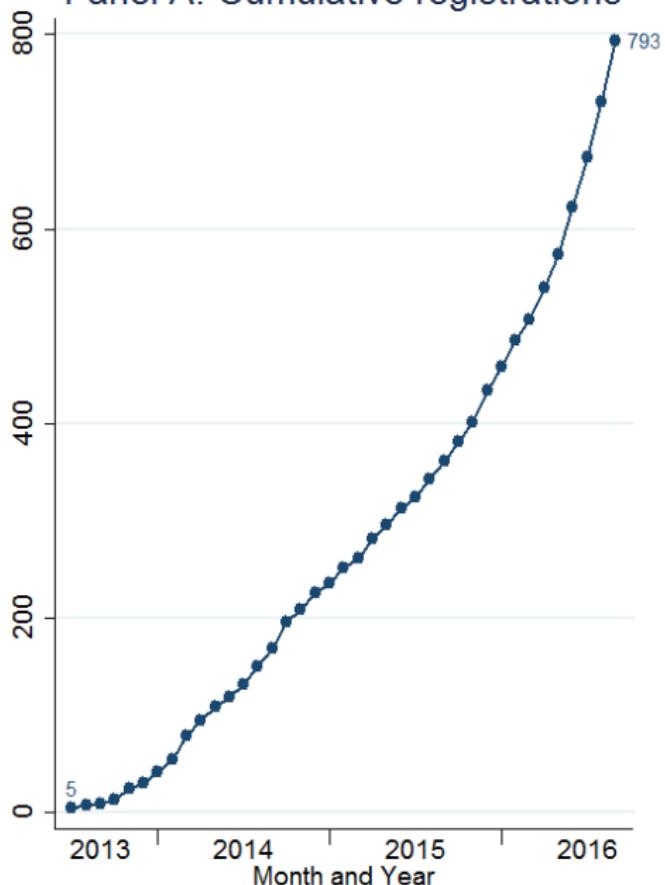
Replicability  
Questionable  
Research Practices  
Project Protocol,  
Reporting Standards  
Workflow  
Data Sharing

Conclusion

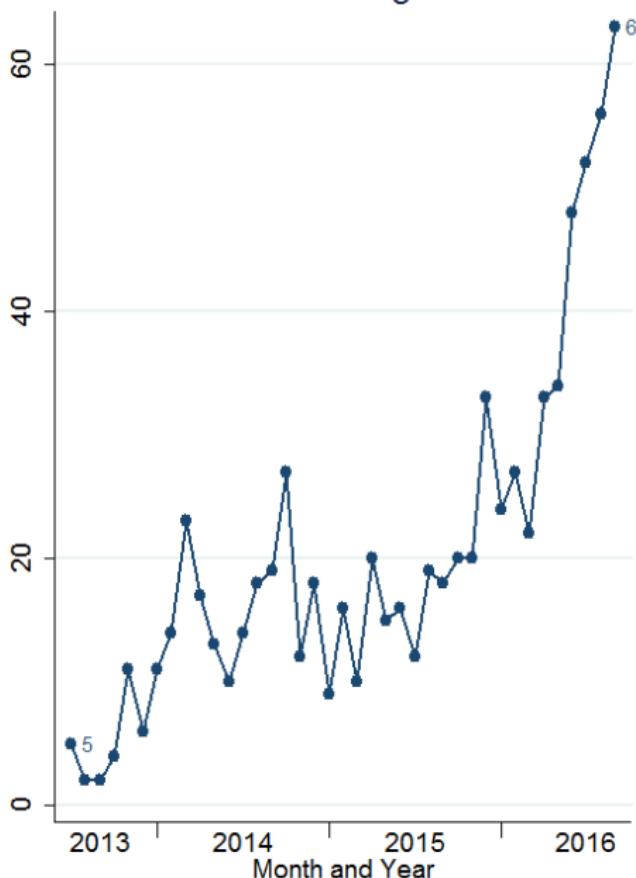
Newer to social sciences, but:

- **AEA registry, currently only for RCTs.**  
<http://socialscienceregistry.org>
- **EGAP registry**  
<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

Panel A: Cumulative registrations



Panel B: New registrations



# Design-Based Publication

Transparency,  
Reproducibil-  
ity, and the  
Credibility of  
Economics  
Research

Christensen

Introduction

Publication  
Bias  
Registration

Specification  
Searching  
Pre-Analysis Plans

Replicability

Questionable  
Research Practices  
Project Protocol,  
Reporting Standards  
Workflow  
Data Sharing

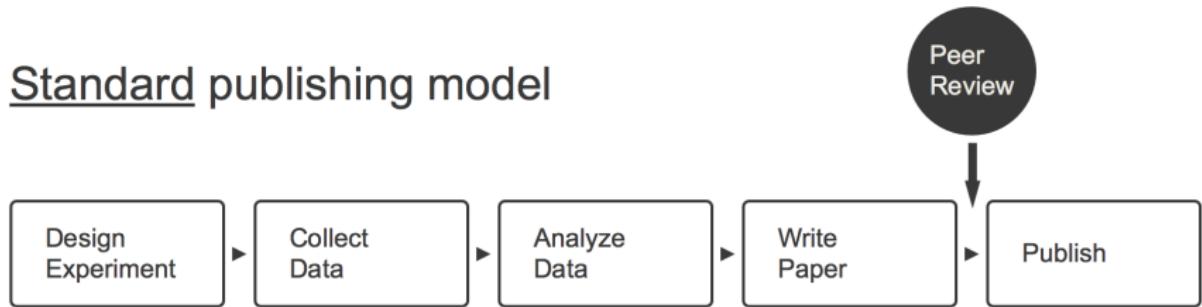
Conclusion

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

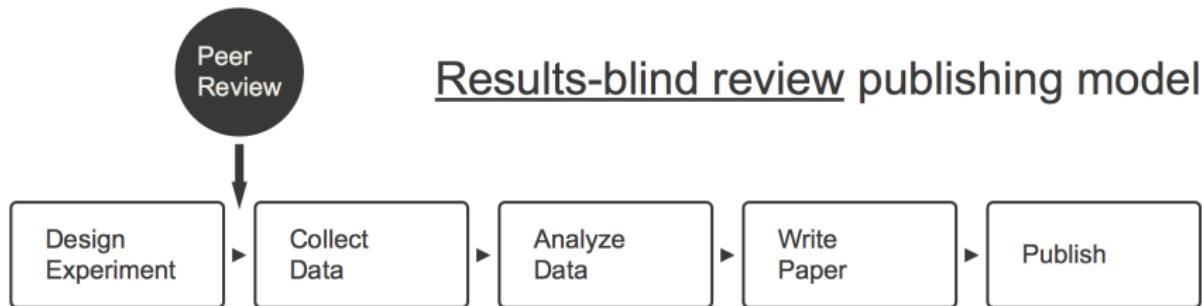
- 1 Design a project
- 2 Submit
- 3 Reviewed based on importance of question and quality of design
- 4 Get in-principle acceptance
- 5 Follow through, and nulls get published

20 Journals, 5 more with Special Issues [▶ Link](#)

## Standard publishing model



## Results-blind review publishing model





# Meta-Analysis

Transparency,  
Reproducibil-  
ity, and the  
Credibility of  
Economics  
Research  
Christensen

Introduction  
Publication Bias  
Registration  
Specification Searching  
Pre-Analysis Plans  
Replicability  
Questionable Research Practices  
Project Protocol, Reporting Standards  
Workflow  
Data Sharing  
Conclusion

Synthesize results systematically

Organizations:

- Cochrane Collaboration (Medicine)
- Campbell Collaboration (Policy)
- What Works Clearinghouse (US Gov't, Education)
- CLEAR (US Gov't, Labor)
- MAER-NET (Economics)

# Meta-Analysis

Transparency,  
Reproducibil-  
ity, and the  
Credibility of  
Economics  
Research

Christensen

Introduction

Publication  
Bias  
Registration

Specification  
Searching  
Pre-Analysis Plans

Replicability

Questionable  
Research Practices  
Project Protocol,  
Reporting Standards  
Workflow  
Data Sharing

Conclusion

## Tools:

- Funnel plots of sample size vs. effect size or precision (Card & Krueger 1995)
- Funnel Asymmetry Test (Stanley & Doucouliagos 2012)
  - Are All Economic Effects Greatly Exaggerated? (2013)  
Meta-meta of N=87
  - P-curve (Simonsohn et al. 2014) [▶ Online App](#)



# P-Hacking

Transparency,  
Reproducibil-  
ity, and the  
Credibility of  
Economics  
Research

Christensen

Introduction

Publication  
Bias  
Registration

Specification  
Searching  
Pre-Analysis Plans

Replicability

Questionable  
Research Practices  
Project Protocol,  
Reporting Standards  
Workflow  
Data Sharing

Conclusion

## 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.
- Discussed since at least Leamer (1983).
- Not something only evil people do. It's subconscious, or simply built into statistics (Gelman, Loken 2013).



BERKELEY INITIATIVE FOR TRANSPARENCY  
IN THE SOCIAL SCIENCES

# Pre-Analysis Plan

Transparency,  
Reproducibility,  
and the  
Credibility of  
Economics  
Research

Christensen

Introduction

Publication  
Bias  
Registration

Specification  
Searching

Pre-Analysis Plans

Replicability

Questionable  
Research Practices

Project Protocol,  
Reporting Standards

Workflow

Data Sharing

Conclusion

Fairly standard, if less detailed statistically, in FDA trials.

- 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.”



BERKELEY INITIATIVE FOR TRANSPARENCY  
IN THE SOCIAL SCIENCES

Transparency,  
Reproducibil-  
ity, and the  
Credibility of  
Economics  
Research

Christensen

Introduction

Publication  
Bias  
Registration

Specification  
Searching  
Pre-Analysis Plans

Replicability

Questionable  
Research Practices

Project Protocol,  
Reporting Standards

Workflow

Data Sharing

Conclusion

# McKenzie Suggestions

## World Bank Development Impact Blog

- 1 Description of the sample to be used in the study
- 2 Key data sources
- 3 Hypotheses to be tested throughout the causal chain
- 4 Specify how variables will be constructed
- 5 Specify the treatment effect equation to be estimated
- 6 What is the plan for how to deal with multiple outcomes and multiple hypothesis testing?
- 7 Procedures to be used for addressing survey attrition
- 8 How will the study deal with outcomes with limited variation?
- 9 If you are going to be testing a model, include the model
- 10 Remember to archive it

# RESHAPING INSTITUTIONS: EVIDENCE ON AID IMPACTS USING A PREANALYSIS PLAN\*

KATHERINE CASEY  
RACHEL GLENNERSTER  
EDWARD MIGUEL

Despite their importance, there is limited evidence on how institutions can be strengthened. Evaluating the effects of specific reforms is complicated by the lack of exogenous variation in institutions, the difficulty of measuring institutional performance, and the temptation to “cherry pick” estimates from among the large number of indicators required to capture this multifaceted subject. We evaluate one attempt to make local institutions more democratic and egalitarian by imposing participation requirements for marginalized groups (including women) and test for learning-by-doing effects. We exploit the random assignment of a governance program in Sierra Leone, develop innovative real-world outcome measures, and use a preanalysis plan (PAP) to bind our hands against data mining. The intervention studied is a “community-driven development” program, which has become a popular strategy for foreign aid donors. We find positive short-run effects on local public goods and economic outcomes, but no evidence for sustained impacts on collective action, decision making, or the involvement of marginalized groups, suggesting that the intervention

Outcome variable	(1) Mean for controls	(2) Treatment effect
<b>Panel A: GoBifo “weakened” institutions</b>		
Attended meeting to decide what to do with the tarp	0.81	-0.04 <sup>+</sup>
Everybody had equal say in deciding how to use the tarp	0.51	-0.11 <sup>+</sup>
Community used the tarp (verified by physical assessment)	0.90	-0.08 <sup>+</sup>
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*
<b>Panel B: GoBifo “strengthened” institutions</b>		
Community teachers have been trained	0.47	0.12 <sup>+</sup>
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

Transparency,  
Reproducibil-  
ity, and the  
Credibility of  
Economics  
Research

Christensen

Introduction

Publication  
Bias

Registration

Specification  
Searching

Pre-Analysis Plans

Replicability

Questionable  
Research Practices

Project Protocol,  
Reporting Standards

Workflow

Data Sharing

Conclusion

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

# 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.



BERKELEY INITIATIVE FOR TRANSPARENCY  
IN THE SOCIAL SCIENCES

# Replication (or Lack Thereof)

Transparency,  
Reproducibil-  
ity, and the  
Credibility of  
Economics  
Research

Christensen

Introduction

Publication  
Bias

Registration

Specification  
Searching

Pre-Analysis Plans

Replicability

Questionable  
Research Practices

Project Protocol,  
Reporting Standards

Workflow

Data Sharing

Conclusion

## 1 The Problem

- Questionable Research Practices
- Replication Projects: JMCB, Chang & Li, Camerer

## 2 Project Protocol, Reporting Standards

## 3 Organizing Workflow

## 4 Code & Data Sharing

# Questionable Research Practices

Transparency,  
Reproducibil-  
ity, and the  
Credibility of  
Economics  
Research

Christensen

Introduction

Publication  
Bias

Registration

Specification  
Searching

Pre-Analysis Plans

Replicability

Questionable  
Research Practices

Project Protocol,  
Reporting Standards

Workflow

Data Sharing

Conclusion

## Why we worry (a little):

- List (2001), Economists at AEA meetings
- Anderson, Martinson, De Vries (2007)
- John, Loewenstein, Prelec (2011)

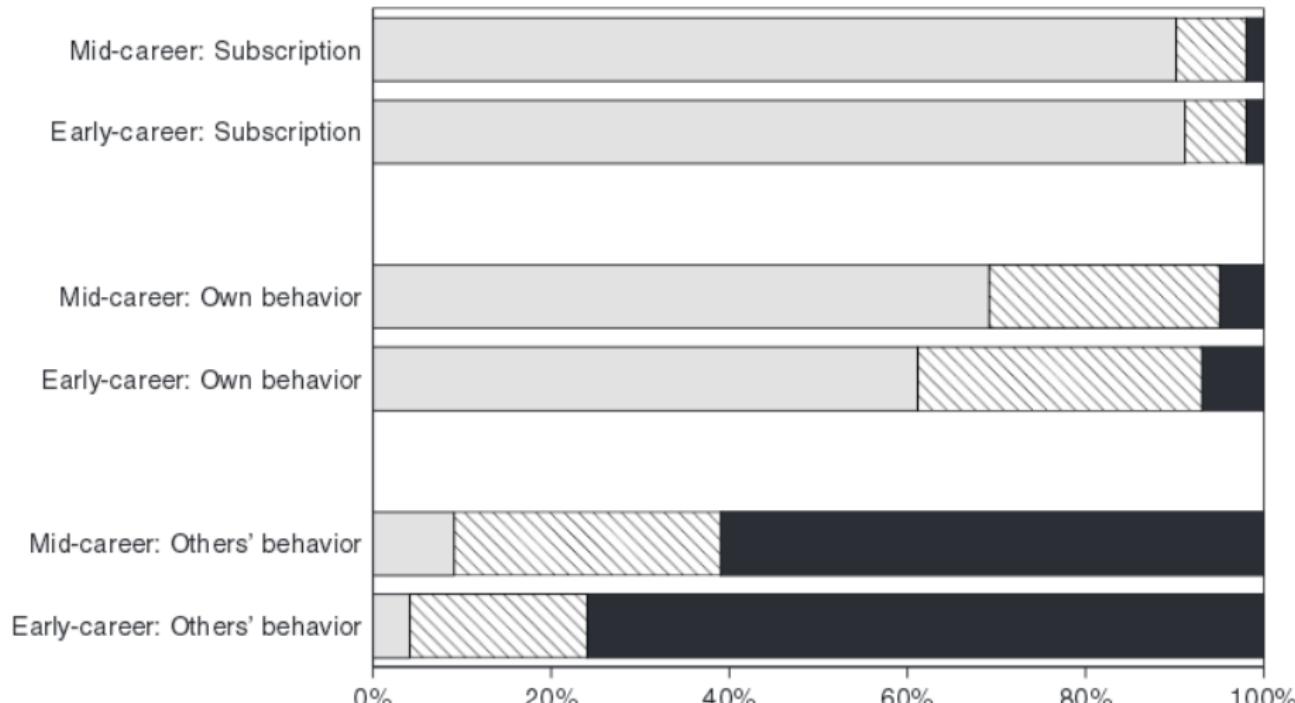
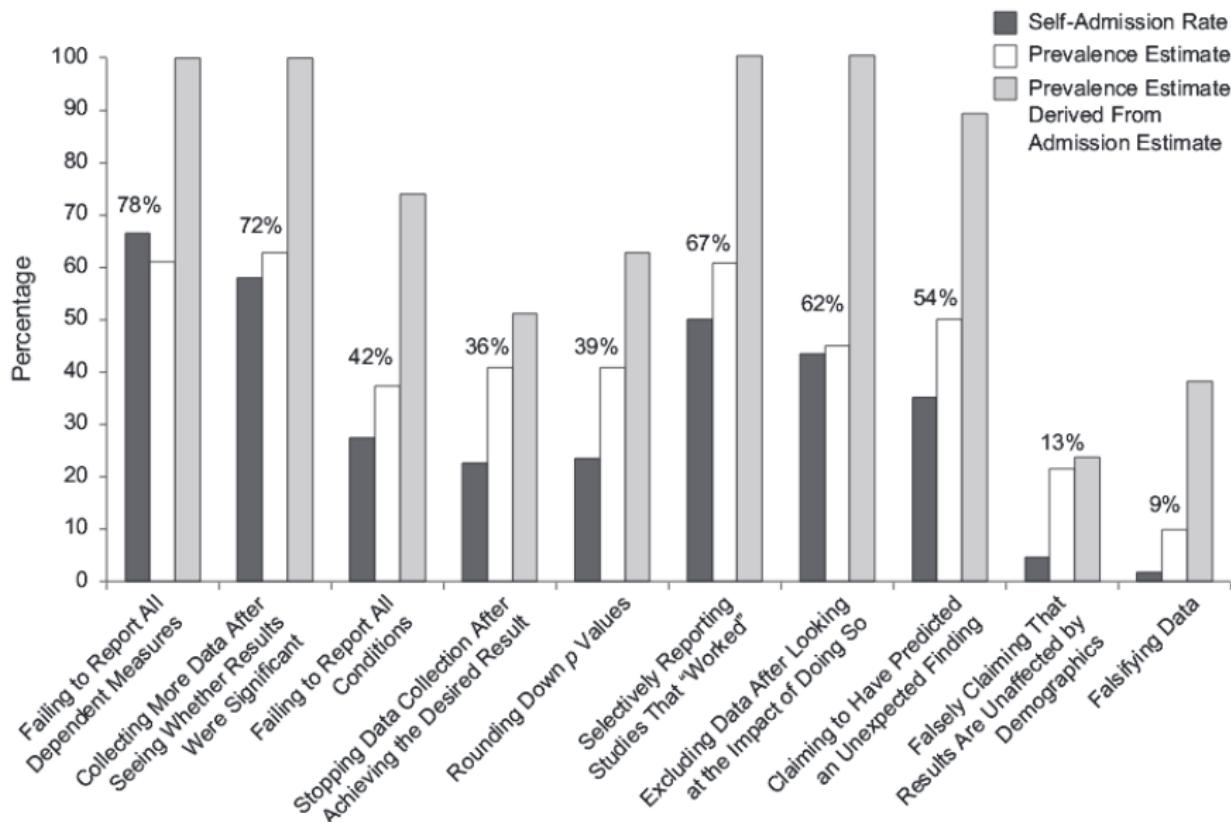


FIG. 3. Norm versus Counternorm Scores: Percent with Norm > Counternorm (dotted), Norm = Counternorm (striped), Norm < Counternorm (solid).



**Fig. 1.** Results of the Bayesian-truth-serum condition in the main study. For each of the 10 items, the graph shows the self-admission rate, prevalence estimate, prevalence estimate derived from the admission estimate (i.e., self-admission rate/admission estimate), and geometric mean of these three percentages (numbers above the bars). See Table I for the complete text of the items.

## **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.*

## 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.



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

# ONE DATA SET, MANY ANALYSTS

Twenty-nine research teams reached a wide variety of conclusions using different methods on the same data set to answer the same question (about football players' skin colour and red cards).

78.7\*

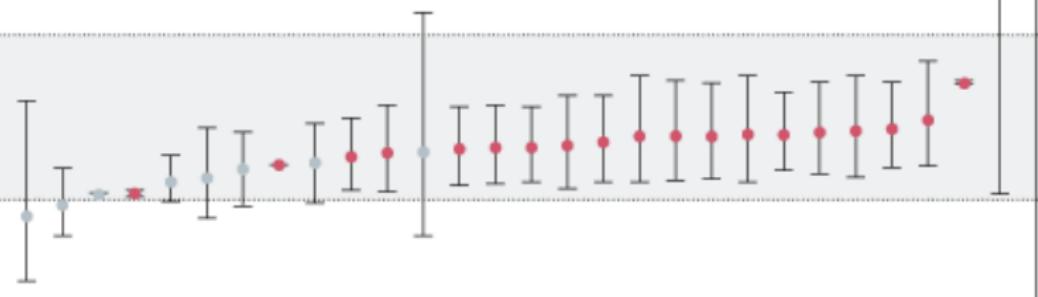
115\*

Dark-skinned  
players four times  
more likely than  
light-skinned  
players to be  
given a red card.

- Statistically significant effect
- Non-significant effect

Twice as likely

Equally likely



Point estimates and 95% confidence intervals. \*Truncated upper bounds.



# Project Protocol, Reporting Standards

Transparency,  
Reproducibil-  
ity, and the  
Credibility of  
Economics  
Research

Christensen

Introduction

Publication  
Bias  
Registration

Specification  
Searching  
Pre-Analysis Plans

Replicability

Questionable  
Research Practices

Project Protocol,  
Reporting Standards

Workflow

Data Sharing

Conclusion

Make sure you report everything another researcher would need to replicate your research.

- Find the appropriate reporting standard for your field and follow it: <http://www.equator-network.org>
- Report the nuts and bolts of the project implementation in a detailed protocol:  
<http://www.spirit-statement.org>
- Transparency and Openness Promotion (TOP) Guidelines: <http://cos.io/top>

**SHARE**[POLICY FORUM](#) | [SCIENTIFIC STANDARDS](#)

## Promoting an open research culture



B. A. Nosek<sup>\*</sup>, G. Alter, G. C. Banks, D. Borsboom, S. D. Bowman, S. J. Breckler, S. Buck, C. D. Chambers, G. Chin, G. Christensen, M. Contestabile, A. Dafoe, E. Eich, J. Freese, R. Glennerster, D. Goroff, D. P. Green, B. Hesse, M. Humphreys, J. Ishiyama, D. Karlan, A. Kraut, A. Lupia, P. Mabry, T. Madon, N. Malhotra, E. Mayo-Wilson, M. McNutt, E. Miguel, E. Levy Paluck, U. Simonsohn, C. Soderberg, B. A. Spellman, J. Turitto, G. VandenBos, S. Vazire, E. J. Wagenmakers, R. Wilson, T. Yarkoni

+ Author Affiliations

<sup>\*</sup>Corresponding author. E-mail: [nosek@virginia.edu](mailto:nosek@virginia.edu)



# Workflow

BERKELEY INITIATIVE FOR TRANSPARENCY  
IN THE SOCIAL SCIENCES

Transparency,  
Reproducibil-  
ity, and the  
Credibility of  
Economics  
Research

Christensen

Introduction

Publication  
Bias

Registration

Specification  
Searching

Pre-Analysis Plans

Replicability

Questionable  
Research Practices

Project Protocol,  
Reporting Standards

Workflow

Data Sharing

Conclusion

“Reproducibility is just collaboration with people you don’t know, including yourself next week”  
—Philip Stark, UC Berkeley Statistics



# Workflow

Transparency,  
Reproducibil-  
ity, and the  
Credibility of  
Economics  
Research

Christensen

Introduction

Publication  
Bias  
Registration

Specification  
Searching

Pre-Analysis Plans

Replicability

Questionable  
Research Practices

Project Protocol,  
Reporting Standards

Workflow

Data Sharing

Conclusion

New tools exist to make workflow more easily reproducible

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





# Data Sharing

Transparency,  
Reproducibil-  
ity, and the  
Credibility of  
Economics  
Research

Christensen

Introduction

Publication  
Bias  
Registration

Specification  
Searching

Pre-Analysis Plans

Replicability

Questionable  
Research Practices  
Project Protocol  
Reporting Standards

Workflow

Data Sharing

Conclusion

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

Transparency,  
Reproducibil-  
ity, and the  
Credibility of  
Economics  
Research

Christensen

Introduction

Publication  
Bias  
Registration

Specification  
Searching

Pre-Analysis Plans

Replicability

Questionable  
Research Practices

Project Protocol,  
Reporting Standards

Workflow

Data Sharing

Conclusion

OK, I'm convinced. How do I implement this in my own research?

- Read the manual I wrote. [▶ Link](#)
- Subscribe to the BITSS blog & E-mail list [▶ Link](#)
- Apply for our Summer Institute. [▶ Link](#)
- Apply for our SSMART Grants (extra funding for developing country researchers). [▶ Link](#)
- Apply for our Leamer-Rosenthal Prizes. [▶ Link](#)

# Summer Institute

Transparency,  
Reproducibil-  
ity, and the  
Credibility of  
Economics  
Research

Christensen

Introduction

Publication  
Bias  
Registration

Specification  
Searching  
Pre-Analysis Plans

Replicability

Questionable  
Research Practices

Project Protocol,  
Reporting Standards

Workflow

Data Sharing

Conclusion

Three days of training in June at UC Berkeley, or two days of training in July at the University of Michigan.



# SSMART Grant

Transparency,  
Reproducibil-  
ity, and the  
Credibility of  
Economics  
Research

Christensen

Introduction

Publication  
Bias  
Registration

Specification  
Searching

Pre-Analysis Plans

Replicability

Questionable  
Research Practices

Project Protocol,  
Reporting Standards

Workflow

Data Sharing

Conclusion

Up to \$30,000 grant for a research project on:

- Develop new methodology
- New tools and approaches for meta-analysis
- Research on researchers and adoption of new norms

Extra funding source for researchers from developing countries.





BERKELEY INITIATIVE FOR TRANSPARENCY  
IN THE SOCIAL SCIENCES

Transparency,  
Reproducibil-  
ity, and the  
Credibility of  
Economics  
Research

Christensen

Introduction

Publication  
Bias

Registration

Specification  
Searching

Pre-Analysis Plans

Replicability

Questionable  
Research Practices

Project Protocol,  
Reporting Standards

Workflow

Data Sharing

Conclusion

# Leamer-Rosenthal Prizes

Up to \$10,000 prize for completed transparent research in the social sciences, especially:

- Economics
- Political Science
- Psychology



Edward Leamer



Robert Rosenthal



BERKELEY INITIATIVE FOR TRANSPARENCY  
IN THE SOCIAL SCIENCES

## Transparency, Reproducibil- ity, and the Credibility of Economics Research

Christensen

Introduction

Publication  
Bias

Registration

Specification  
Searching

Pre-Analysis Plans

Replicability

Questionable  
Research Practices

Project Protocol,  
Reporting Standards

Workflow

Data Sharing

Conclusion

## Questions?

# Thank you!