



BERKELEY INITIATIVE FOR TRANSPARENCY  
IN THE SOCIAL SCIENCES

Research  
Transparency  
in the Social  
Sciences

Introduction

Ethical  
Research

Publication  
Bias —  
Registrations  
Publication Bias  
Registrations

P-Hacking —  
Pre-Analysis  
Plans  
P-Hacking  
Pre-Analysis Plan

Replications  
— Guidelines  
& Protocols  
Project Protocol,  
Reporting Standards  
Workflow  
Version Control  
Data Sharing

Research

# Research Transparency in the Social Sciences

Garret Christensen<sup>1,2</sup>  
Fernando Hoces de la Guardia<sup>1</sup>

<sup>1</sup>UC Berkeley:  
Berkeley Initiative for Transparency in the Social Sciences  
<sup>2</sup>Berkeley Institute for Data Sciences

CEGA, November 2017

Slides available online at <https://goo.gl/RtfxqX>

# Outline

Research Transparency in the Social Sciences

Introduction

Ethical Research

Publication Bias — Registrations

Publication Bias Registrations

P-Hacking — Pre-Analysis Plans

P-Hacking

Pre-Analysis Plan

Replications — Guidelines & Protocols

Project Protocol, Reporting Standards

Workflow

Version Control

Data Sharing

Research

## 1 Introduction

## 2 Ethical Research

## 3 Publication Bias — Registrations

## 4 P-Hacking — Pre-Analysis Plans

## 5 Replications — Guidelines & Protocols

## 6 Research Transparency for Policy

## 7 Conclusion



# Outline

Research  
Transparency  
in the Social  
Sciences

Introduction

Ethical  
Research

Publication  
Bias —  
Registrations  
Publication Bias  
Registrations

P-Hacking —  
Pre-Analysis  
Plans

P-Hacking

Pre-Analysis Plan

Replications  
— Guidelines  
& Protocols

Project Protocol,  
Reporting Standards

Workflow

Version Control

Data Sharing

Research

## 1 Introduction

## 2 Ethical Research

## 3 Publication Bias — Registrations

## 4 P-Hacking — Pre-Analysis Plans

## 5 Replications — Guidelines & Protocols

## 6 Research Transparency for Policy

## 7 Conclusion



BERKELEY INITIATIVE FOR TRANSPARENCY  
IN THE SOCIAL SCIENCES



ICPSR



# Outline

Research Transparency  
in the Social Sciences

Introduction

Ethical Research

Publication Bias — Registrations  
Publication Bias Registrations

P-Hacking — Pre-Analysis Plans  
P-Hacking  
Pre-Analysis Plan

Replications — Guidelines & Protocols  
Project Protocol, Reporting Standards  
Workflow  
Version Control  
Data Sharing

Research

## 1 Introduction

## 2 Ethical Research

## 3 Publication Bias — Registrations

## 4 P-Hacking — Pre-Analysis Plans

## 5 Replications — Guidelines & Protocols

## 6 Research Transparency for Policy

## 7 Conclusion

# Ethical Research

Research Transparency in the Social Sciences

Introduction

Ethical Research

Publication Bias — Registrations

Publication Bias Registrations

P-Hacking — Pre-Analysis Plans

P-Hacking Pre-Analysis Plan

Replications — Guidelines & Protocols

Project Protocol, Reporting Standards  
Workflow  
Version Control  
Data Sharing

- Transparency is part of being an ethical researcher.
- Scientific values espoused by Robert Merton (Merton 1942):
  - Universalism: anyone can make a claim regardless of status.
  - Communism: open sharing of knowledge.
  - Disinterestedness: truth as motivation, not financial gains (COI).
  - Organized skepticism: peer review, replication.



BERKELEY INITIATIVE FOR TRANSPARENCY  
IN THE SOCIAL SCIENCES

## Research Transparency in the Social Sciences

Introduction

Ethical Research

Publication Bias — Registrations

Publication Bias Registrations

P-Hacking — Pre-Analysis Plans

P-Hacking Pre-Analysis Plan

Replications — Guidelines & Protocols

Project Protocol, Reporting Standards

Workflow

Version Control

Data Sharing

# Ethical Research

- Fraud exists (Simonsohn 2013), but mostly we should admit that we're human, subject to bias and motivated reasoning, transparency can help with this (Nosek, Spies, Motyl 2012).
- Those of us who run experiments or use data with personal identifying information should take IRBs seriously as part of transparency (Ch. 11–13 Morton & Williams 2010, Desposato 2014).



# Ethical Research

Research  
Transparency  
in the Social  
Sciences

Introduction

Ethical  
Research

Publication  
Bias —  
Registrations

Publication Bias  
Registrations

P-Hacking —  
Pre-Analysis  
Plans

P-Hacking  
Pre-Analysis Plan

Replications  
— Guidelines  
& Protocols

Project Protocol,  
Reporting Standards  
Workflow  
Version Control  
Data Sharing

## Why we worry:

- (Anderson, Martinson, De Vries 2007)
- (John, Loewenstein, Prelec 2011)

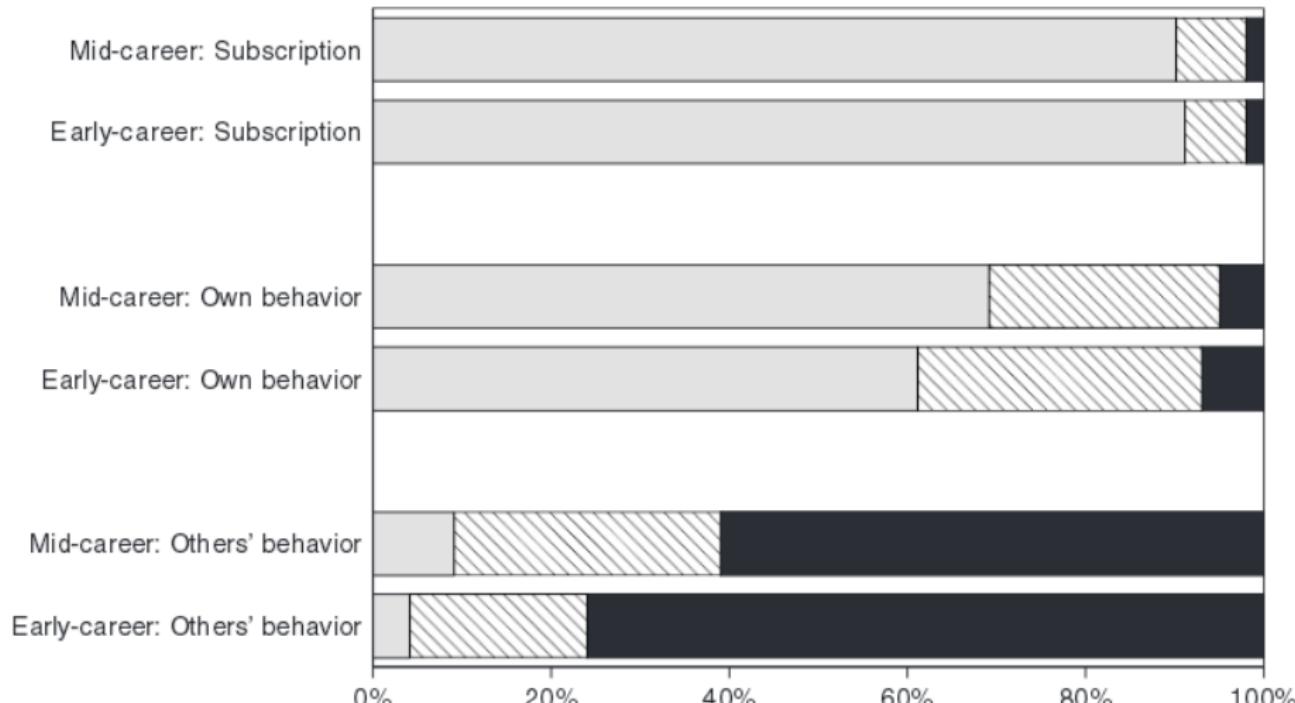
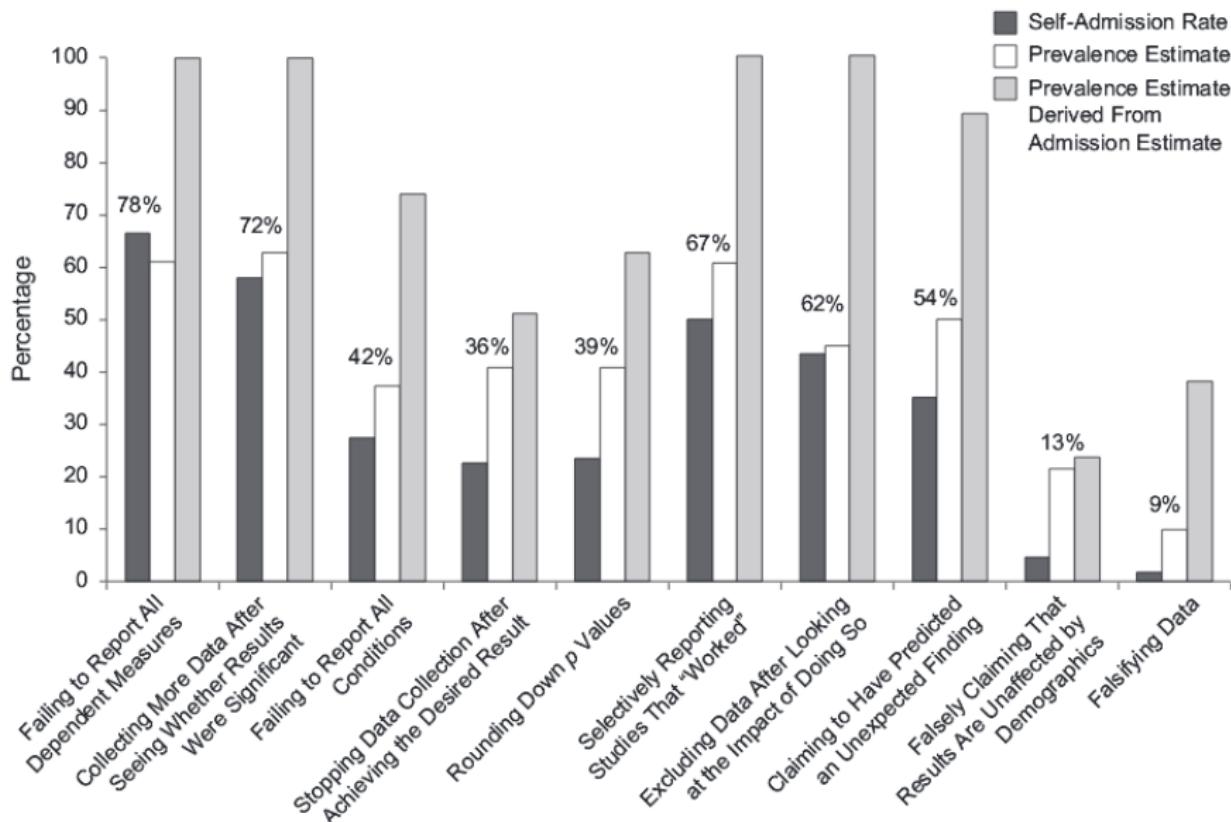


FIG. 3. Norm versus Counternorm Scores: Percent with Norm > Counternorm (dotted), Norm = Counternorm (striped), Norm < Counternorm (solid).



**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.



# Outline

## 1 Introduction

## 2 Ethical Research

## 3 Publication Bias — Registrations

## 4 P-Hacking — Pre-Analysis Plans

## 5 Replications — Guidelines & Protocols

## 6 Research Transparency for Policy

## 7 Conclusion



BERKELEY INITIATIVE FOR TRANSPARENCY

IN THE SOCIAL SCIENCES

Research  
Transparency  
in the Social  
Sciences

Introduction

Ethical  
Research

Publication  
Bias —  
Registrations

Publication Bias  
Registrations

P-Hacking —  
Pre-Analysis  
Plans

P-Hacking  
Pre-Analysis Plan

Replications  
— Guidelines  
& Protocols

Project Protocol,  
Reporting Standards

Workflow

Version Control

Data Sharing

# Publication Bias

## Existence of the problem:

- Effect sizes diminish with sample size (Gerber, Green, Nickerson 2001)
- There is a higher fraction of rejected hypothesis tests in social compared to hard sciences (Fanelli 2010).
- Published null results are disappearing over time, in all disciplines (Fanelli 2011).
- Data on the complete set of experiments run shows strong results are 40pp more likely to be published, and 60pp more likely to be written up. The file drawer problem is large. (Franco, Malhotra, Simonovits 2014)

# All Fields

Research Transparency in the Social Sciences

Introduction

Ethical Research

Publication Bias — Registrations

Publication Bias Registrations

P-Hacking — Pre-Analysis Plans

P-Hacking Pre-Analysis Plan

Replications — Guidelines & Protocols

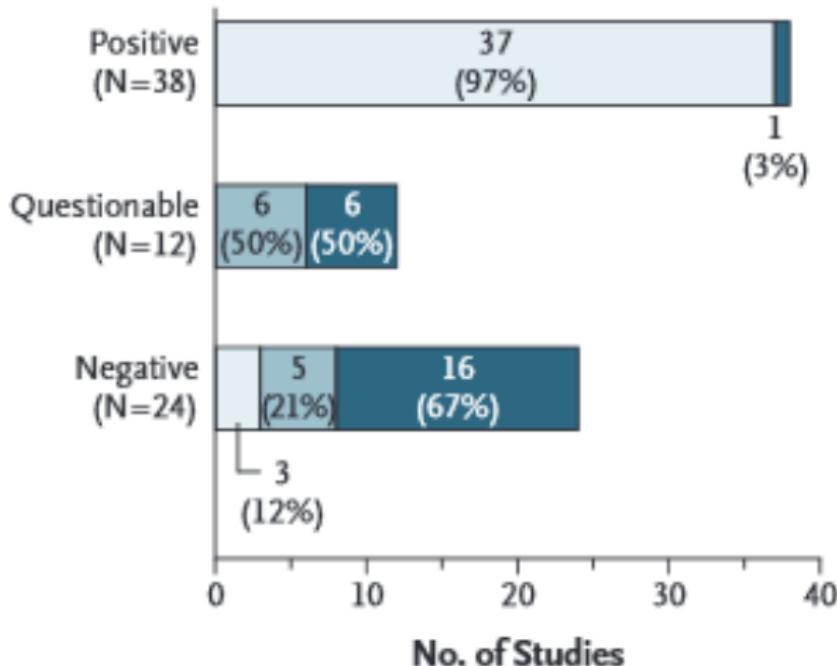
Project Protocol, Reporting Standards  
Workflow  
Version Control  
Data Sharing

- 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)

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

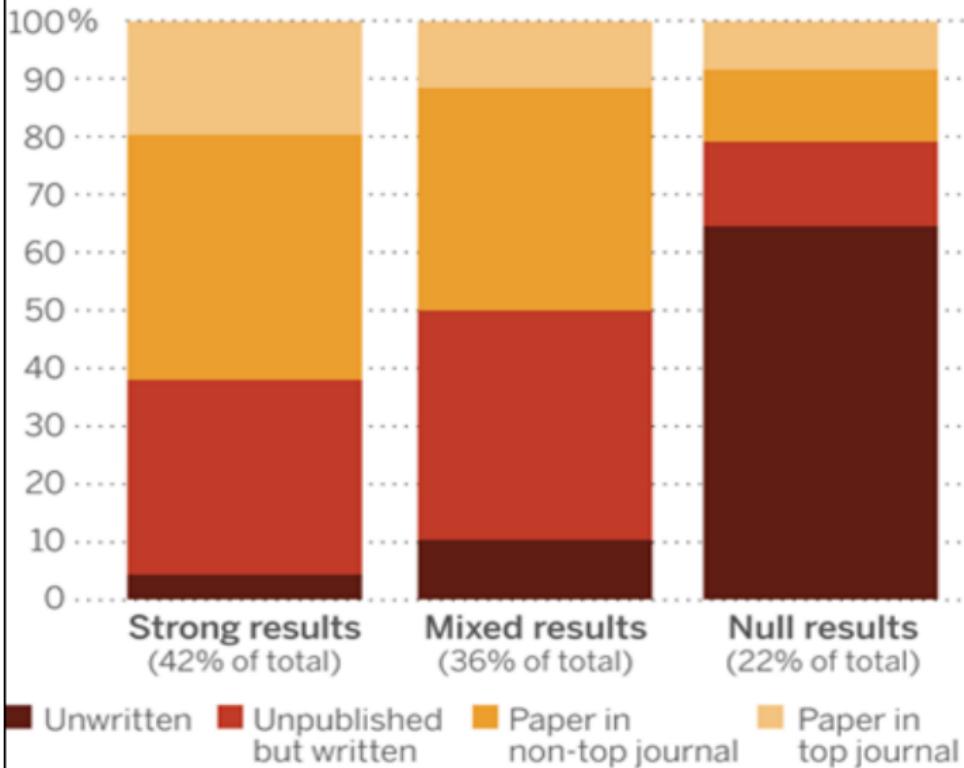
## A Studies (N=74)

### FDA Decision



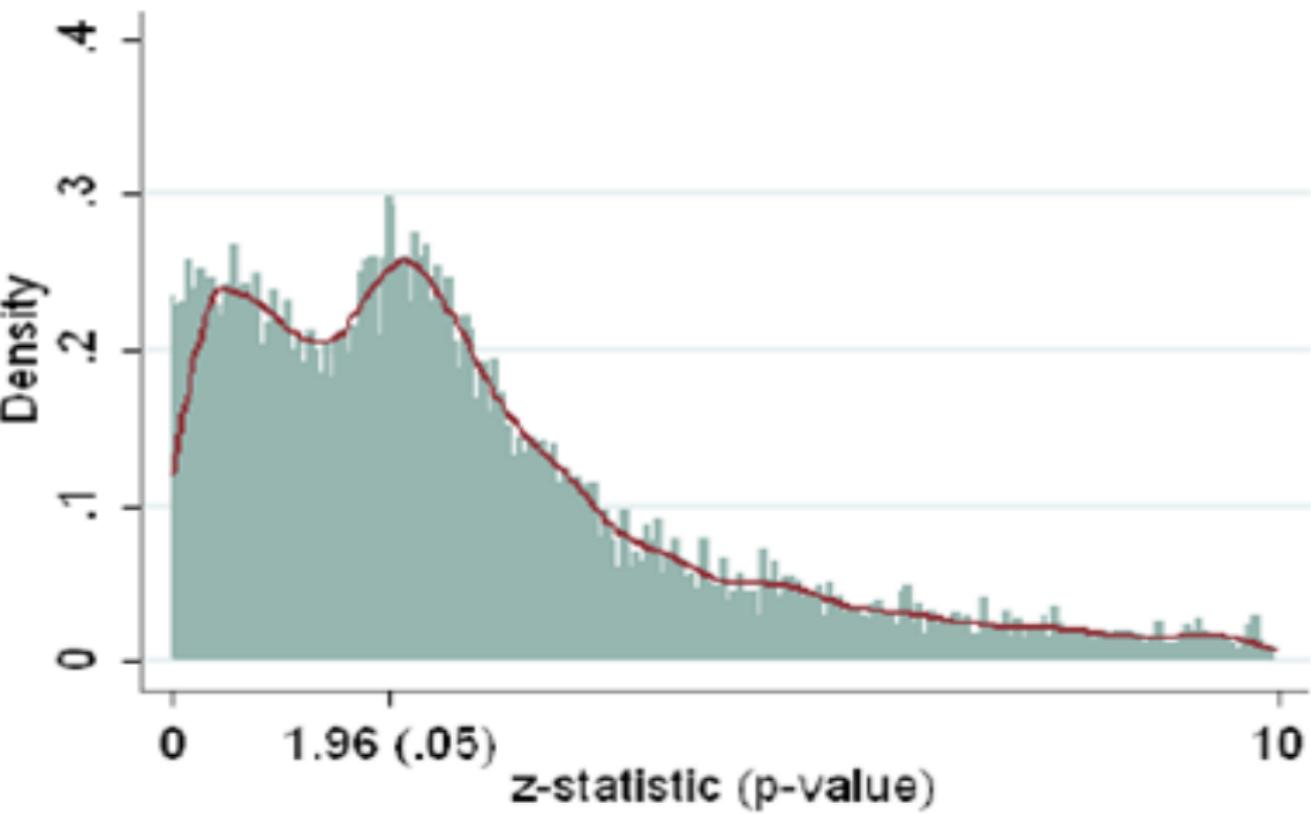
# Most null results are never written up

The fate of 221 social science experiments



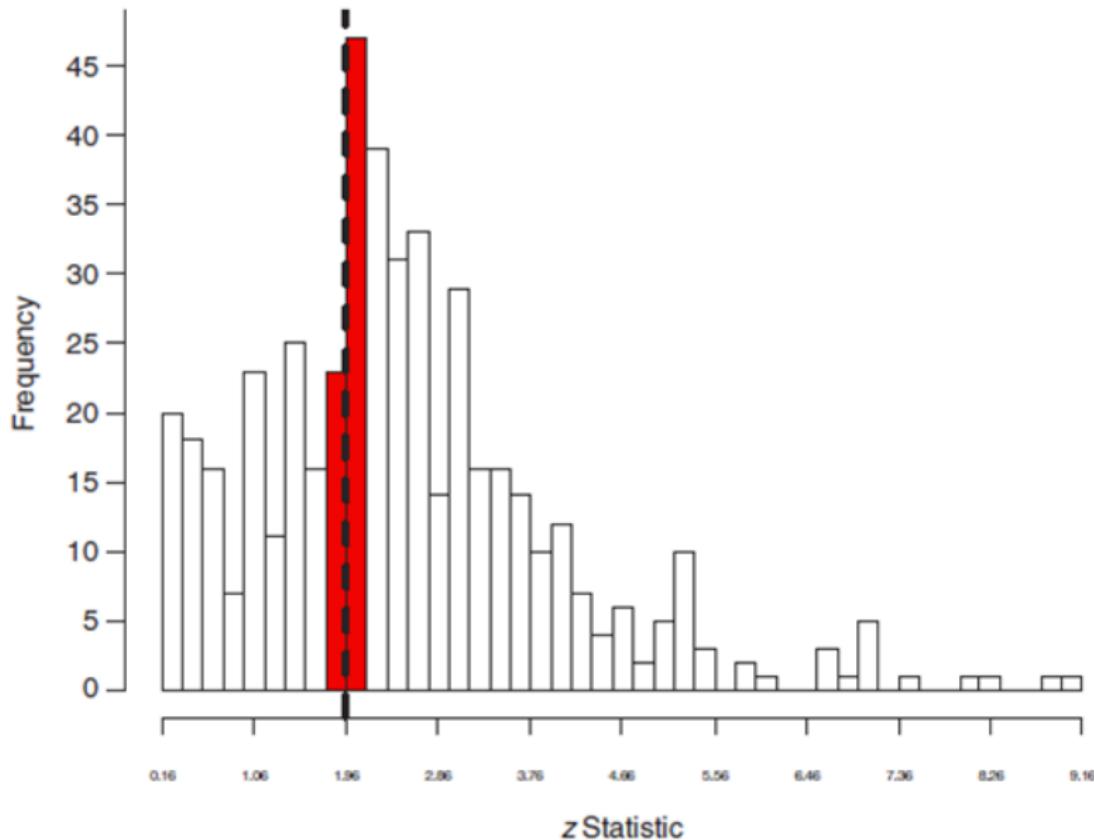
Source: A. Franco et al., *Science* (28 August)

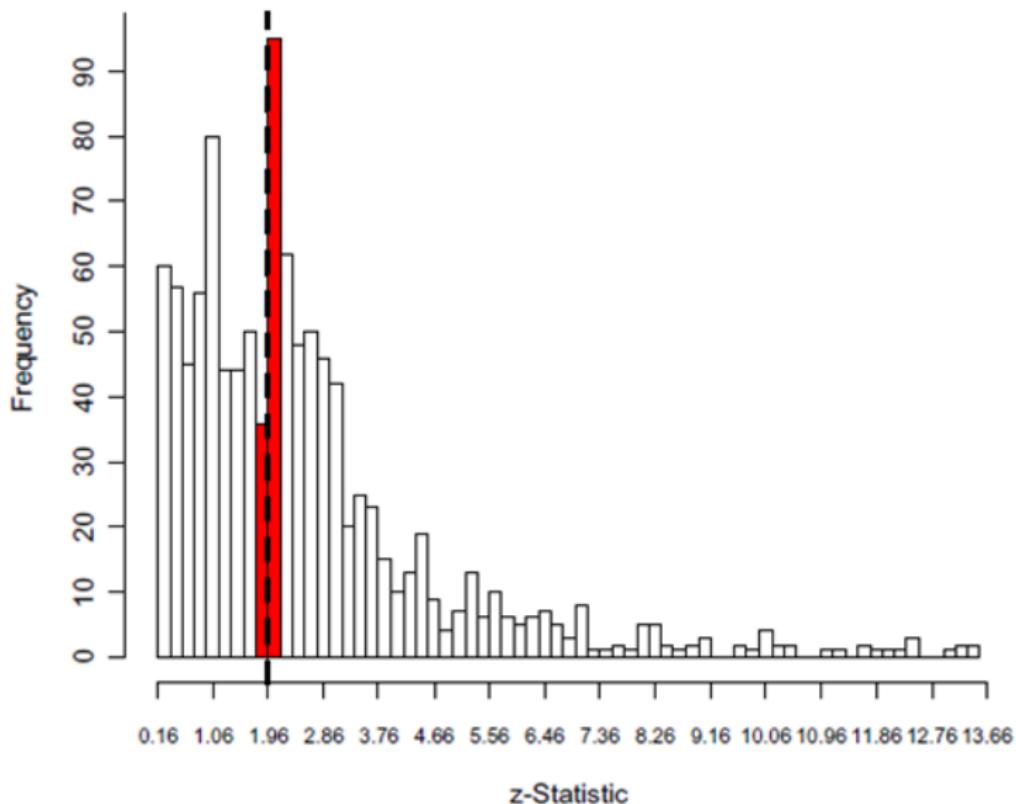
(b) Unrounded distribution of z-statistics.



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.



# Publication Bias

Research  
Transparency  
in the Social  
Sciences

Introduction

Ethical  
Research

Publication  
Bias —  
Registrations  
Publication Bias  
Registrations

P-Hacking —  
Pre-Analysis  
Plans

P-Hacking

Pre-Analysis Plan

Replications  
— Guidelines  
& Protocols

Project Protocol,  
Reporting Standards  
Workflow  
Version Control  
Data Sharing

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.



BERKELEY INITIATIVE FOR TRANSPARENCY  
IN THE SOCIAL SCIENCES

Research  
Transparency  
in the Social  
Sciences

Introduction

Ethical  
Research

Publication  
Bias —  
Registrations

Publication Bias  
Registrations

P-Hacking —  
Pre-Analysis  
Plans

P-Hacking  
Pre-Analysis Plan

Replications  
— Guidelines  
& Protocols

Project Protocol,  
Reporting Standards

Workflow

Version Control

Data Sharing

# 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 (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

Research Transparency in the Social Sciences

Introduction

Ethical Research

Publication Bias — Registrations

Publication Bias Registrations

P-Hacking — Pre-Analysis Plans

P-Hacking

Pre-Analysis Plan

Replications — Guidelines & Protocols

Project Protocol, Reporting Standards

Workflow

Version Control

Data Sharing

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

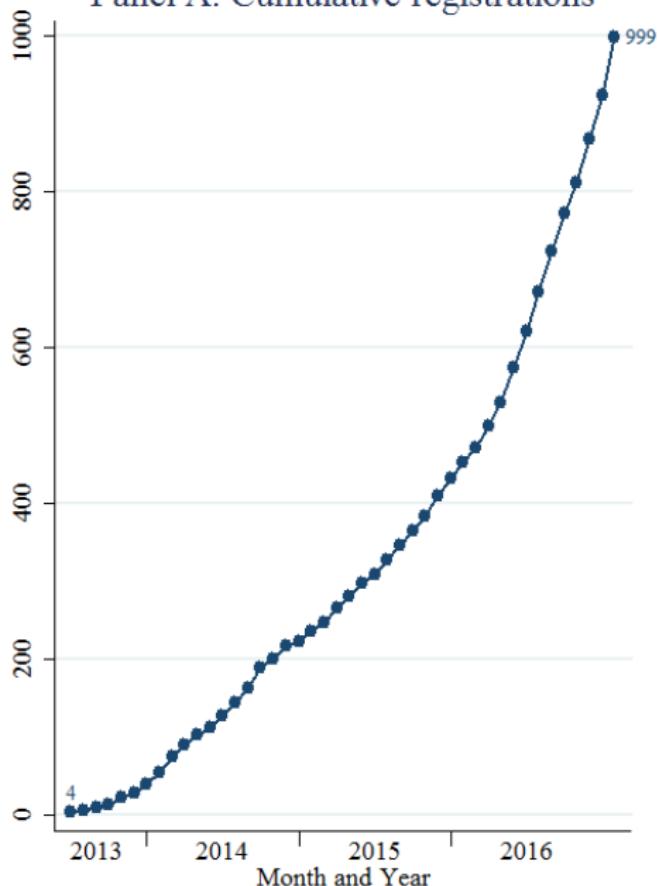
<http://osf.io>

- Open format

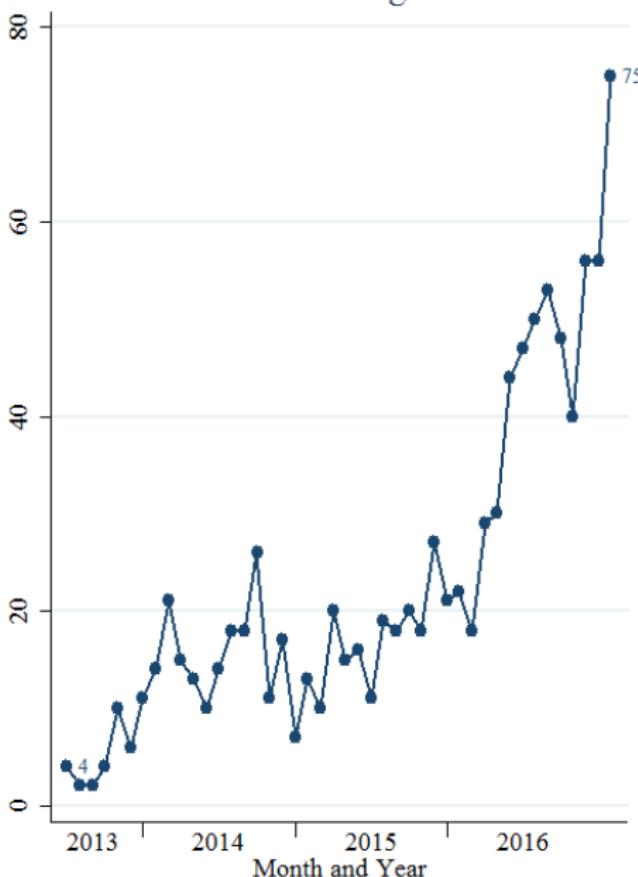
- Will soon sync with above

- Simple: <http://aspredicted.org>

Panel A: Cumulative registrations



Panel B: New registrations





BERKELEY INITIATIVE FOR TRANSPARENCY  
IN THE SOCIAL SCIENCES

Research  
Transparency  
in the Social  
Sciences

Introduction

Ethical  
Research

Publication  
Bias —  
Registrations  
Publication Bias  
Registrations

P-Hacking —  
Pre-Analysis  
Plans  
P-Hacking  
Pre-Analysis Plan

Replications  
— Guidelines  
& Protocols  
Project Protocol,  
Reporting Standards  
Workflow  
Version Control  
Data Sharing

Research

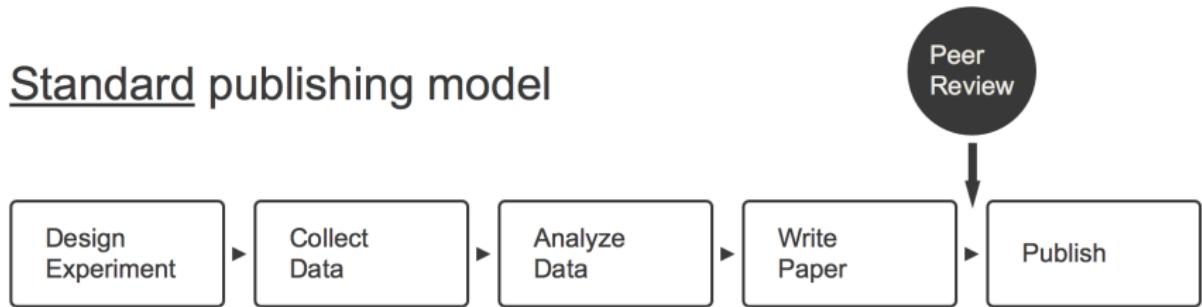
# Design-Based Publication

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

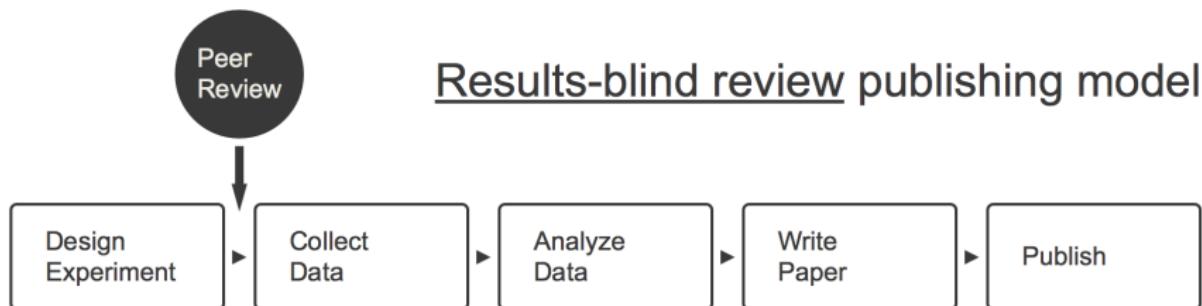
- 1 Design a project
- 2 Submit
- 3 Reviewed based on importance of question and quality of design
- 4 Get in-principle acceptance
- 5 Follow through, and nulls get published

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

## Standard publishing model



## Results-blind review publishing model





# Meta-Analysis

Research  
Transparency  
in the Social  
Sciences

Introduction

Ethical  
Research

Publication  
Bias —  
Registrations

Publication Bias  
Registrations

P-Hacking —  
Pre-Analysis  
Plans

P-Hacking  
Pre-Analysis Plan

Replications  
— Guidelines  
& Protocols

Project Protocol,  
Reporting Standards  
Workflow  
Version Control  
Data Sharing

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

Research  
Transparency  
in the Social  
Sciences

Introduction

Ethical  
Research

Publication  
Bias —  
Registrations

Publication Bias  
Registrations

P-Hacking —  
Pre-Analysis  
Plans

P-Hacking  
Pre-Analysis Plan

Replications  
— Guidelines  
& Protocols

Project Protocol,  
Reporting Standards  
Workflow  
Version Control  
Data Sharing

## Tools:

- Funnel plots of sample size vs. effect size or precision (Card & Krueger 1995)
- Funnel Asymmetry Test (Stanley & Doucouliagos 2012)
- P-curve (Simonsohn et al. 2014)
  - ▶ Online App
  - One for all P-checker
    - ▶ Shiny App



# Outline

Research Transparency  
in the Social Sciences

Introduction

Ethical Research

Publication Bias — Registrations  
Publication Bias Registrations

P-Hacking — Pre-Analysis Plans  
P-Hacking  
Pre-Analysis Plan

Replications — Guidelines & Protocols  
Project Protocol, Reporting Standards  
Workflow  
Version Control  
Data Sharing

Research

- 1 Introduction**
- 2 Ethical Research**
- 3 Publication Bias — Registrations**
- 4 P-Hacking — Pre-Analysis Plans**
- 5 Replications — Guidelines & Protocols**
- 6 Research Transparency for Policy**
- 7 Conclusion**

# P-Hacking

Research Transparency in the Social Sciences

Introduction

Ethical Research

Publication Bias — Registrations

Publication Bias Registrations

P-Hacking — Pre-Analysis Plans

P-Hacking

Pre-Analysis Plan

Replications — Guidelines & Protocols

Project Protocol, Reporting Standards

Workflow

Version Control

Data Sharing

Research

## 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).



BERKELEY INITIATIVE FOR TRANSPARENCY  
IN THE SOCIAL SCIENCES

## Research Transparency in the Social Sciences

Introduction

Ethical Research

Publication Bias —  
Registrations  
Publication Bias Registrations

P-Hacking —  
Pre-Analysis Plans

P-Hacking

Pre-Analysis Plan

Replications — Guidelines & Protocols

Project Protocol,  
Reporting Standards  
Workflow  
Version Control  
Data Sharing

# P-Hacking is fun!

- “Science isn’t Broken” —538 journalism piece with interactive demo [Link](#)
- Train your p-hacking skills R/Shiny App. [Link](#)
- An Exact Fishy Test [Link](#)

# 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\*

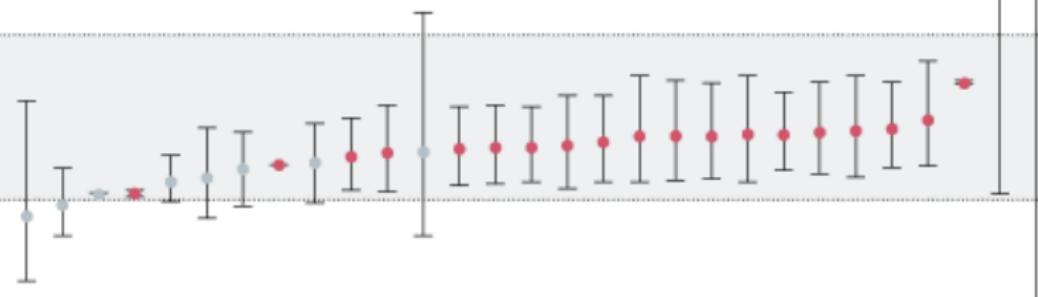
115\*

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.



# 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.”



BERKELEY INITIATIVE FOR TRANSPARENCY  
IN THE SOCIAL SCIENCES

Research  
Transparency  
in the Social  
Sciences

Introduction

Ethical  
Research

Publication  
Bias —  
Registrations

Publication Bias  
Registrations

P-Hacking —  
Pre-Analysis  
Plans

P-Hacking  
Pre-Analysis Plan

Replications  
— Guidelines  
& Protocols

Project Protocol,  
Reporting Standards

Workflow

Version Control

Data Sharing

Research

# Origin: FDA's Guidance for Industry

## “E9 Statistical Principles for Clinical Trials” (1998) Link

### §V Data Analysis Considerations

- 1 Prespecification of the Analysis
- 2 Analysis Sets
- 3 Missing Values and Outliers
- 4 Data Transformation
- 5 Estimation, Confidence Intervals, and Hypothesis Testing
- 6 Adjustment of Significance and Confidence Levels
- 7 Subgroups, Interactions, and Covariates
- 8 Integrity of Data and Computer Software Validity

# Glennerster, Takavarasha Suggestions

## *Running Randomized Evaluations*

- 1 the main outcome measures,
- 2 which outcome measures are primary and which are secondary,
- 3 the precise composition of any families that will be used for mean effects analysis,
  - Explain mean effects, FWER, FDR using Anderson (JASA 2008).
- 4 the subgroups that will be analyzed,
- 5 the direction of expected impact if we want to use a one-sided test, and
- 6 the primary specification to be used for the analysis.



BERKELEY INITIATIVE FOR TRANSPARENCY  
IN THE SOCIAL SCIENCES

Research  
Transparency  
in the Social  
Sciences

Introduction

Ethical  
Research

Publication  
Bias —  
Registrations

Publication Bias  
Registrations

P-Hacking —  
Pre-Analysis  
Plans

P-Hacking  
Pre-Analysis Plan

Replications  
— Guidelines  
& Protocols

Project Protocol,  
Reporting Standards  
Workflow

Version Control  
Data Sharing

Research

# 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

# Examples

Research Transparency in the Social Sciences

Introduction

Ethical Research

Publication Bias — Registrations

Publication Bias Registrations

P-Hacking — Pre-Analysis Plans

P-Hacking  
Pre-Analysis Plan

Replications — Guidelines & Protocols

Project Protocol, Reporting Standards  
Workflow  
Version Control  
Data Sharing

- J-PAL Hypothesis Registry (11), see <http://www.povertyactionlab.org/Hypothesis-Registry>  
6 published papers:
  - Sierra Leone CDD, Oregon Medicare, Turkey Job Training, El Salvador TOMS, two in Indonesia (Olken et al.)
  - Psychology: Hawkins, Fitzgerald, Nosek—Conception Risk and Prejudice

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
<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*



## THE PREREGISTRATION CHALLENGE

Learn more at [cos.io/prereg](http://cos.io/prereg)

# PAP—Observational Studies

Research Transparency in the Social Sciences

Introduction

Ethical Research

Publication Bias — Registrations

Publication Bias Registrations

P-Hacking — Pre-Analysis Plans

P-Hacking

Pre-Analysis Plan

Replications — Guidelines & Protocols

Project Protocol, Reporting Standards

Workflow

Version Control

Data Sharing

- 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.



# Outline

Research Transparency  
in the Social Sciences

Introduction

Ethical Research

Publication Bias — Registrations  
Publication Bias Registrations

P-Hacking — Pre-Analysis Plans

P-Hacking

Pre-Analysis Plan

Replications — Guidelines & Protocols

Project Protocol, Reporting Standards

Workflow

Version Control

Data Sharing

Research

- 1 Introduction**
- 2 Ethical Research**
- 3 Publication Bias — Registrations**
- 4 P-Hacking — Pre-Analysis Plans**
- 5 Replications — Guidelines & Protocols**
- 6 Research Transparency for Policy**
- 7 Conclusion**



# Replication

BERKELEY INITIATIVE FOR TRANSPARENCY  
IN THE SOCIAL SCIENCES

Research  
Transparency  
in the Social  
Sciences

Introduction

Ethical  
Research

Publication  
Bias —  
Registrations

Publication Bias  
Registrations

P-Hacking —  
Pre-Analysis  
Plans

P-Hacking

Pre-Analysis Plan

Replications  
— Guidelines  
& Protocols

Project Protocol,  
Reporting Standards

Workflow

Version Control

Data Sharing

- 1 The Problem (JMCC Project)**
- 2 Project Protocol, Reporting Standards**
- 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.*



BERKELEY INITIATIVE FOR TRANSPARENCY  
IN THE SOCIAL SCIENCES

## Research Transparency in the Social Sciences

Introduction

Ethical Research

Publication Bias —  
Registrations

Publication Bias  
Registrations

P-Hacking —  
Pre-Analysis Plans

P-Hacking  
Pre-Analysis Plan

Replications — Guidelines  
& Protocols

Project Protocol,  
Reporting Standards

Workflow

Version Control

Data Sharing

Research

# Project Protocol, Reporting Standards

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\*, 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

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



# Workflow

BERKELEY INITIATIVE FOR TRANSPARENCY  
IN THE SOCIAL SCIENCES

Research  
Transparency  
in the Social  
Sciences

Introduction

Ethical  
Research

Publication  
Bias —  
Registrations

Publication Bias  
Registrations

P-Hacking —  
Pre-Analysis  
Plans

P-Hacking  
Pre-Analysis Plan

Replications  
— Guidelines  
& Protocols

Project Protocol,  
Reporting Standards

Workflow

Version Control

Data Sharing

Research

“Reproducibility is just collaboration with people you don’t know, including yourself next week”

—Philip Stark, UC Berkeley Statistics

# Workflow

Research Transparency in the Social Sciences

Introduction

Ethical Research

Publication Bias — Registrations

Publication Bias Registrations

P-Hacking — Pre-Analysis Plans

P-Hacking Pre-Analysis Plan

Replications — Guidelines & Protocols

Project Protocol, Reporting Standards

Workflow

Version Control

Data Sharing

Research

- Practical coding and organizational suggestions
  - Making any changes to a file that has been posted/shared means it gets a new name.
  - Use version commands to ensure others get same results.
  - Long (2008) *The Workflow of Data Analysis Using Stata*
- Literate programming (extensive commenting, making the aim of code reading by a human)
- Version Control
- Dynamic Documents

# Version Control

Research Transparency in the Social Sciences

Introduction

Ethical Research

Publication Bias — Registrations

Publication Bias Registrations

P-Hacking — Pre-Analysis Plans

P-Hacking Pre-Analysis Plan

Replications — Guidelines & Protocols

Project Protocol, Reporting Standards

Workflow

Version Control

Data Sharing

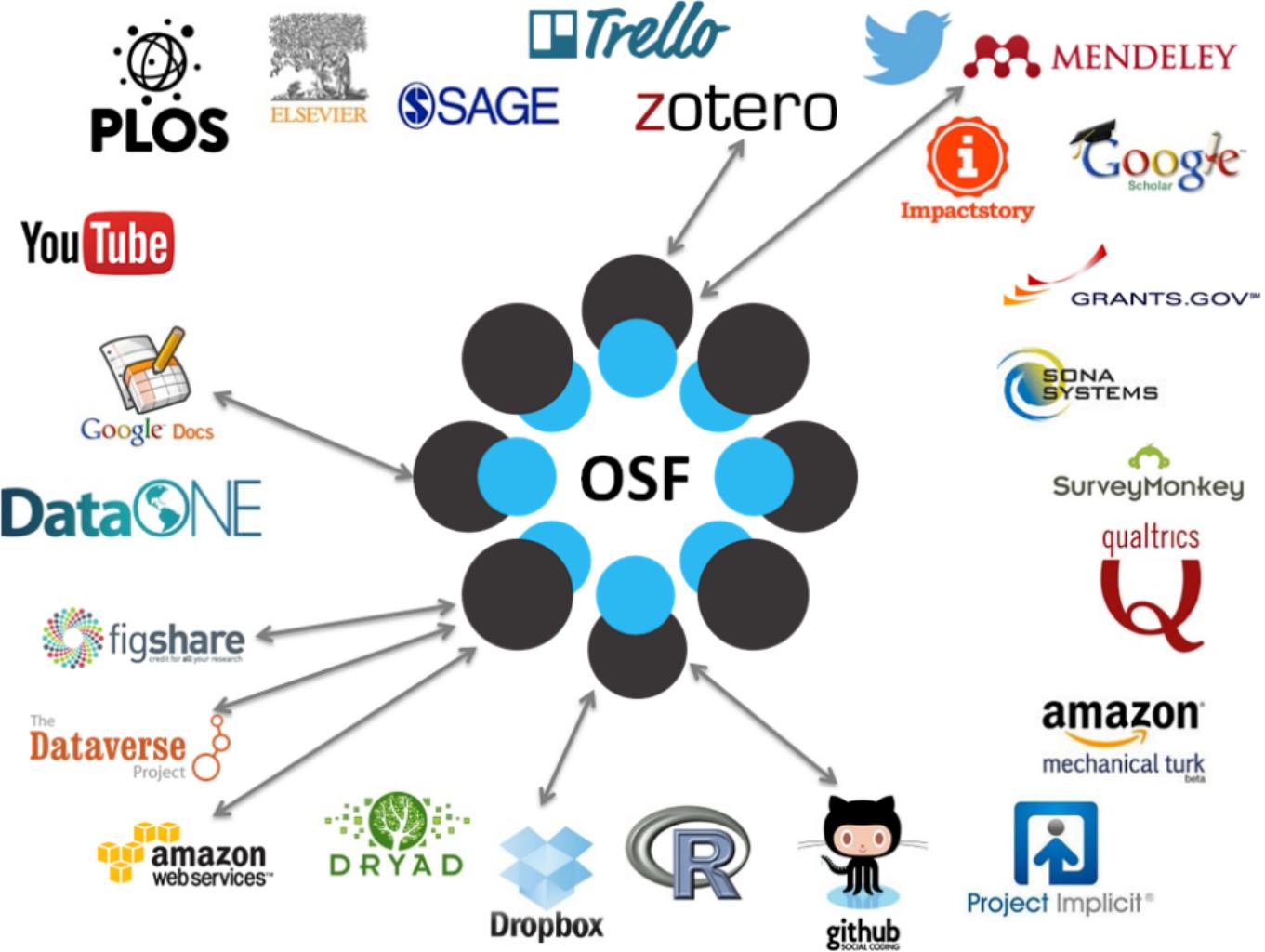
Research

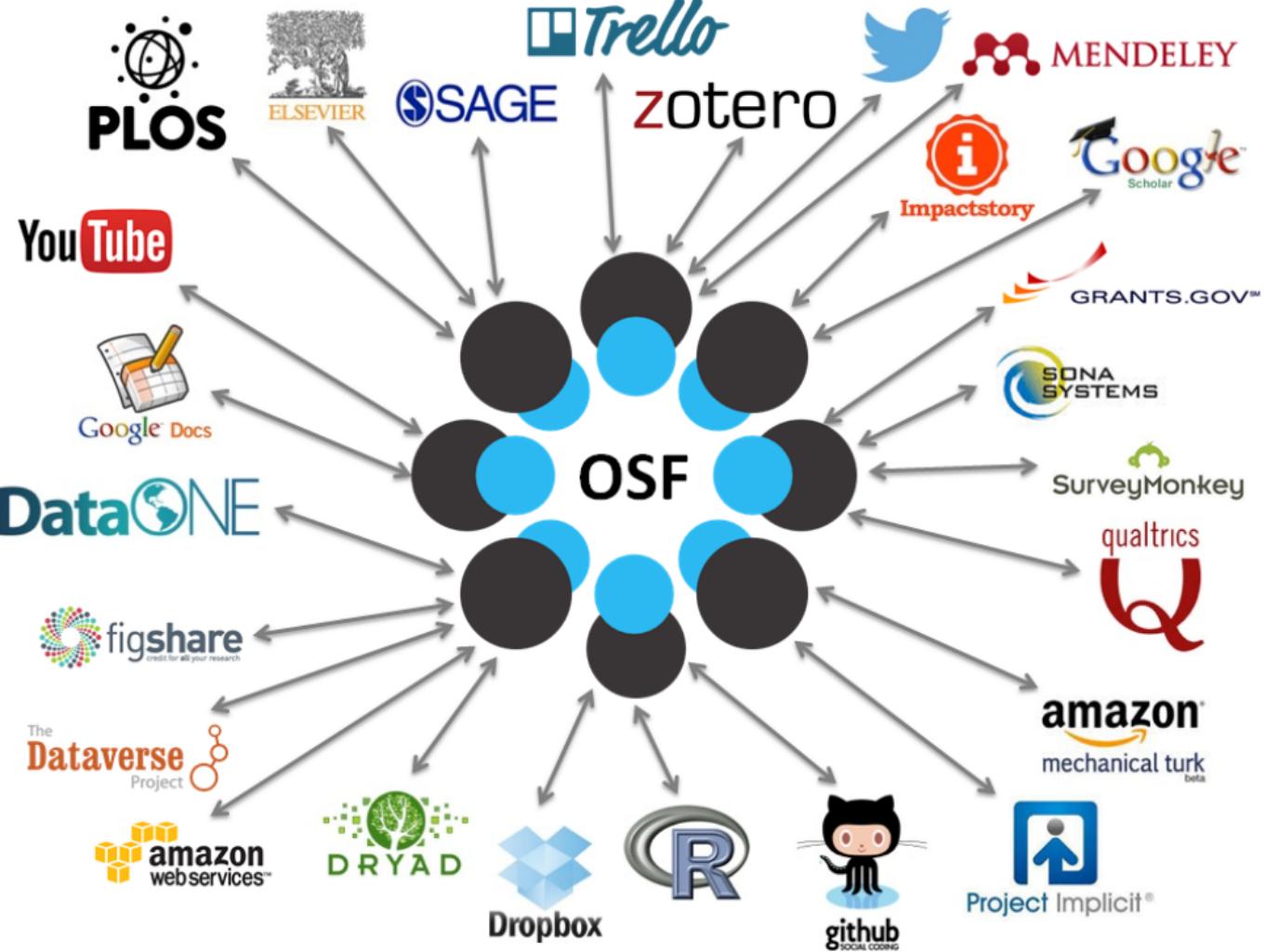
- Using version control (AKA revision control) can help to make your work more reproducible.
- What is version control?

*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

Research  
Transparency  
in the Social  
Sciences

Introduction

Ethical  
Research

Publication  
Bias —  
Registrations

Publication Bias  
Registrations

P-Hacking —  
Pre-Analysis  
Plans

P-Hacking  
Pre-Analysis Plan

Replications  
— Guidelines  
& Protocols

Project Protocol,  
Reporting Standards  
Workflow  
Version Control  
Data Sharing

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.

The logo for Jupyter consists of the word "jupyter" in a bold, black, sans-serif font. It is centered within a white circle. Above and below the circle are two thick, orange, curved bands that resemble the top and bottom edges of a smile. Small dark gray circles are positioned at the intersections of these curves with the circle.

jupyter



# Data Sharing

Research Transparency in the Social Sciences

Introduction

Ethical Research

Publication Bias — Registrations

Publication Bias Registrations

P-Hacking — Pre-Analysis Plans

P-Hacking  
Pre-Analysis Plan

Replications — Guidelines & Protocols

Project Protocol, Reporting Standards

Workflow

Version Control

Data Sharing

Research

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.



# Outline

Research Transparency  
in the Social Sciences

Introduction

Ethical Research

Publication Bias — Registrations

Publication Bias Registrations

P-Hacking — Pre-Analysis Plans

P-Hacking

Pre-Analysis Plan

Replications — Guidelines & Protocols

Project Protocol, Reporting Standards

Workflow

Version Control

Data Sharing

Research

## 1 Introduction

## 2 Ethical Research

## 3 Publication Bias — Registrations

## 4 P-Hacking — Pre-Analysis Plans

## 5 Replications — Guidelines & Protocols

## 6 Research Transparency for Policy

## 7 Conclusion



# Research Transparency for Policy

Research  
Transparency  
in the Social  
Sciences

Introduction

Ethical  
Research

Publication  
Bias —  
Registrations

Publication Bias  
Registrations

P-Hacking —  
Pre-Analysis  
Plans

P-Hacking

Pre-Analysis Plan

Replications  
— Guidelines  
& Protocols

Project Protocol,  
Reporting Standards

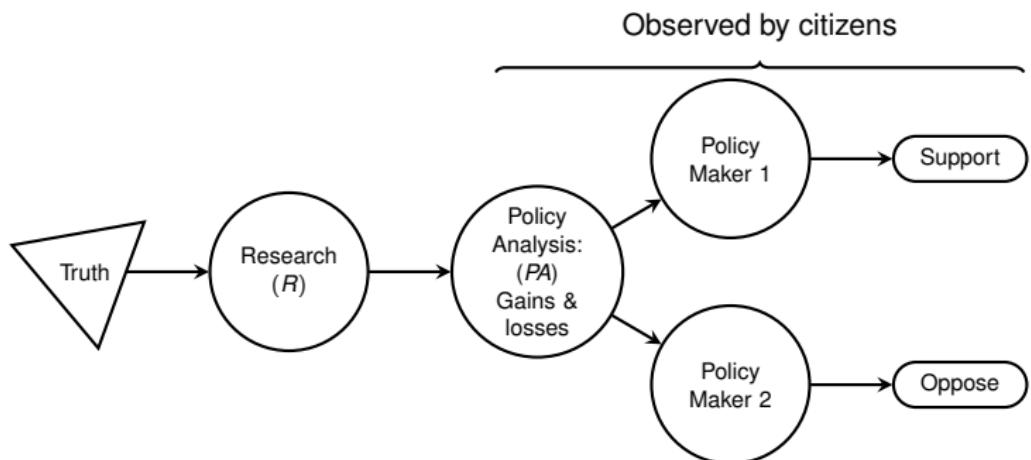
Workflow

Version Control

Data Sharing

- Policy-evidence link with **high** transparency.
- Policy-evidence link with **low** transparency.
- BITSS proposal.

# One link between evidence and public policy



**Figure: Simplified Model of Connection Between Evidence and Policy**

# Policy-Evidence Link With Low Transparency & Reproducibility

Research Transparency in the Social Sciences

Introduction

Ethical Research

Publication Bias — Registrations

Publication Bias Registrations

P-Hacking — Pre-Analysis Plans

P-Hacking  
Pre-Analysis Plan

Replications — Guidelines & Protocols

Project Protocol, Reporting Standards

Workflow

Version Control

Data Sharing

Research

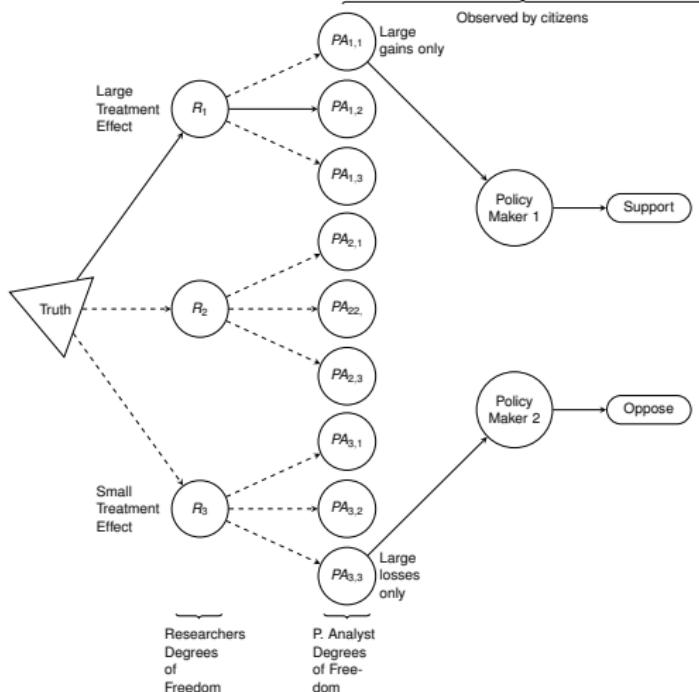


Figure: Policy-making with low TR in research and policy analysis

# Our Proposal Reproducibility and Transparency for Policy

Research Transparency in the Social Sciences

Introduction

Ethical Research

Publication Bias — Registrations

Publication Bias Registrations

P-Hacking — Pre-Analysis Plans

P-Hacking  
Pre-Analysis Plan

Replications — Guidelines & Protocols

Project Protocol, Reporting Standards  
Workflow  
Version Control  
Data Sharing

Research

- 1 Publish guidelines on transparency and reproducibility for policy analysis (similar to the TOP, described before).
- 2 Partner with agencies/think tanks interested in implementing this ideas. [Example here.](#)
- 3 Iterate.



# Outline

Research Transparency  
in the Social Sciences

Introduction

Ethical Research

Publication Bias — Registrations

Publication Bias Registrations

P-Hacking — Pre-Analysis Plans

P-Hacking

Pre-Analysis Plan

Replications — Guidelines & Protocols

Project Protocol, Reporting Standards

Workflow

Version Control

Data Sharing

Research

## 1 Introduction

## 2 Ethical Research

## 3 Publication Bias — Registrations

## 4 P-Hacking — Pre-Analysis Plans

## 5 Replications — Guidelines & Protocols

## 6 Research Transparency for Policy

## 7 Conclusion

# Conclusion

Research Transparency in the Social Sciences

Introduction

Ethical Research

Publication Bias — Registrations

Publication Bias Registrations

P-Hacking — Pre-Analysis Plans

P-Hacking  
Pre-Analysis Plan

Replications — Guidelines & Protocols

Project Protocol, Reporting Standards  
Workflow  
Version Control  
Data Sharing

Research

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

- Take our MOOC (E. Miguel). [▶ Link](#)
- Subscribe to the BITSS blog & E-mail list [▶ Link](#)
- Apply for our RT2 Summer Institute. [▶ Link](#)
- Become a Catalyst!
- Apply for our SSMART Grants (extra funding for developing country researchers). [▶ Link](#)
- Check out our resource website [▶ Link](#)
- Interested in incorporating some of these practices into policy analysis? Email me! [fhoce@berkeley.edu](mailto:fhoce@berkeley.edu)

# Summer Institute

Research  
Transparency  
in the Social  
Sciences

Introduction

Ethical  
Research

Publication  
Bias —  
Registrations

Publication Bias  
Registrations

P-Hacking —  
Pre-Analysis  
Plans

P-Hacking  
Pre-Analysis Plan

Replications  
— Guidelines  
& Protocols

Project Protocol,  
Reporting Standards  
Workflow  
Version Control  
Data Sharing

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



# SSMART Grant

Research Transparency in the Social Sciences

Introduction

Ethical Research

Publication Bias — Registrations

Publication Bias Registrations

P-Hacking — Pre-Analysis Plans

P-Hacking Pre-Analysis Plan

Replications — Guidelines & Protocols

Project Protocol, Reporting Standards Workflow Version Control Data Sharing

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

## Research Transparency in the Social Sciences

Introduction

Ethical Research

Publication Bias — Registrations

Publication Bias Registrations

P-Hacking — Pre-Analysis Plans

P-Hacking Pre-Analysis Plan

Replications — Guidelines & Protocols

Project Protocol, Reporting Standards

Workflow

Version Control

Data Sharing

Research

# Questions?

# Thank you!