Research Transparency in the Social Sciences

Garret Christensen[1,2]

[1]Berkeley Initiative for Transparency in the Social Sciences
UC Berkeley

[2]Berkeley Institute for Data Science

APHRC & CARTA, March 2016

Find these slides online at `http://osf.io/qdv5h`

Outline

BERKELEY INITIATIVE FOR TRANSPARENCY
IN THE SOCIAL SCIENCES


The
Dataverse
Project
ljaf
BIDS
CDL
COS
CENTER FOR
OPEN SCIENCE
CEGA
Lab
egap
experiments in governance and politics
ALFRED P. SLOAN FOUNDATION
1934
RR
REPLICATION
Science
AAAS
DA RT
PROJECT TIER
TEACHING INTEGRITY
IN EMPIRICAL RESEARCH
ICPSR

Ethical Research

- 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.
  - Communality: open sharing of knowledge.
  - Disinterestedness: truth as motivation, not financial gains (COI).
  - Organized skepticism: peer review, replication.

Ethical Research

- 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.
  - Communality: open sharing of knowledge.
  - Disinterestedness: truth as motivation, not financial gains (COI).
  - Organized skepticism: peer review, replication.

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

- 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).

Brian Nosek I don't know if it will cheer you up, but a journalist this week asked me if my biases may have contributed to how we did and discussed RPP. Here is my response to her. I personally find the solutions, external to relying on my intentions, soothing for my confidence in science:

"Yes, it is absolutely the case that my research practices and interpretation are likely to be influenced by my preconceptions, assumptions, and biases. That must be so because I am human. Psychology has demonstrated how we humans are likely to use motivated reasoning in order to shape evidence to conform to the conclusions that we want rather than the conclusions that are correct.

Science offers some good tools to try to mitigate these biases. For example, one tool is transparency. If others can observe how I made my claims, then there is more opportunity to identify potential biases in the methodology, reasoning, and conclusions. That is why we made the entire project public right from the start - all protocols, methods, data, analysis scripts, etc. are available for review and critique on the Open Science Framework. Another tool is preregistration. Even if I desire not to be biased, once I observe the data, if I have multiple ways that I could analyze, and choices about what I should report, then I am more likely to use motivated reasoning to justify - even unintentionally - reporting the analyses and outcomes that support my point of view, rather than those that counter it. So, what we did in the Reproducibility Project is seek advice of original authors to maximize the quality of the designs before running then, and then preregistering the design and analysis plan in advance. In that way, we put constraint on ourselves to follow the plan we pre-specified and removed the opportunity for flexibility in how we interpreted the data.

TheUpshot

A STUDY IN CONTROVERSY

Professors' Research Project Stirs Political Outrage in Montana

Derek Willis @derekwillis OCT. 28, 2014

The only thing that three political scientists wanted to do was send mailers to thousands of Montana voters as part of a study of nonpartisan elections. What could possibly go wrong?

Monkey Cage

Campaign experiment found to be in violation of Montana law

By Jeremy Johnson May 13, 2015

Ethical Research

Why we worry:

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

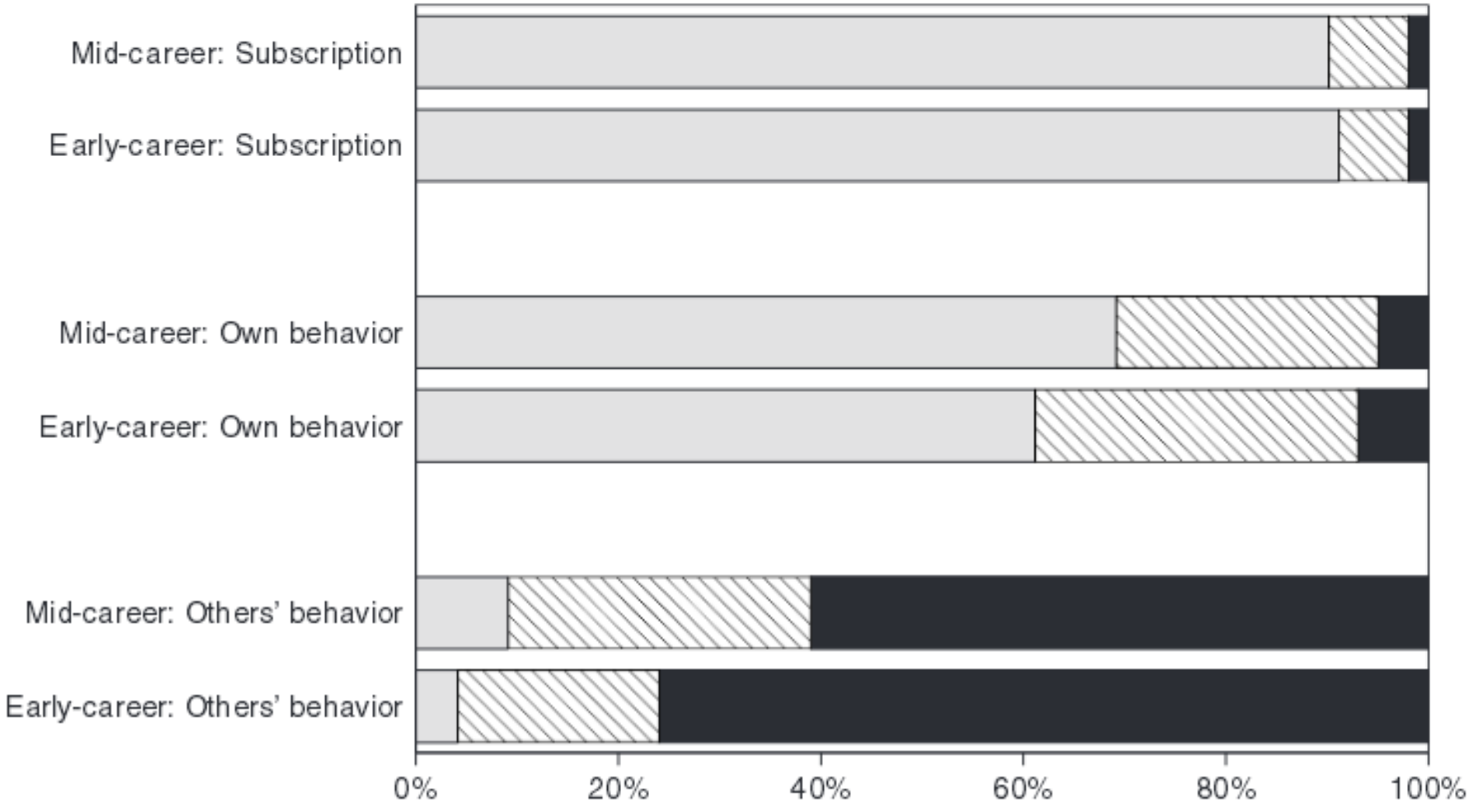


FIG. 3. Norm versus Counternorm Scores: Percent with Norm > Counternorm (dotted), Norm = Counternorm (striped), Norm < Counternorm (solid).

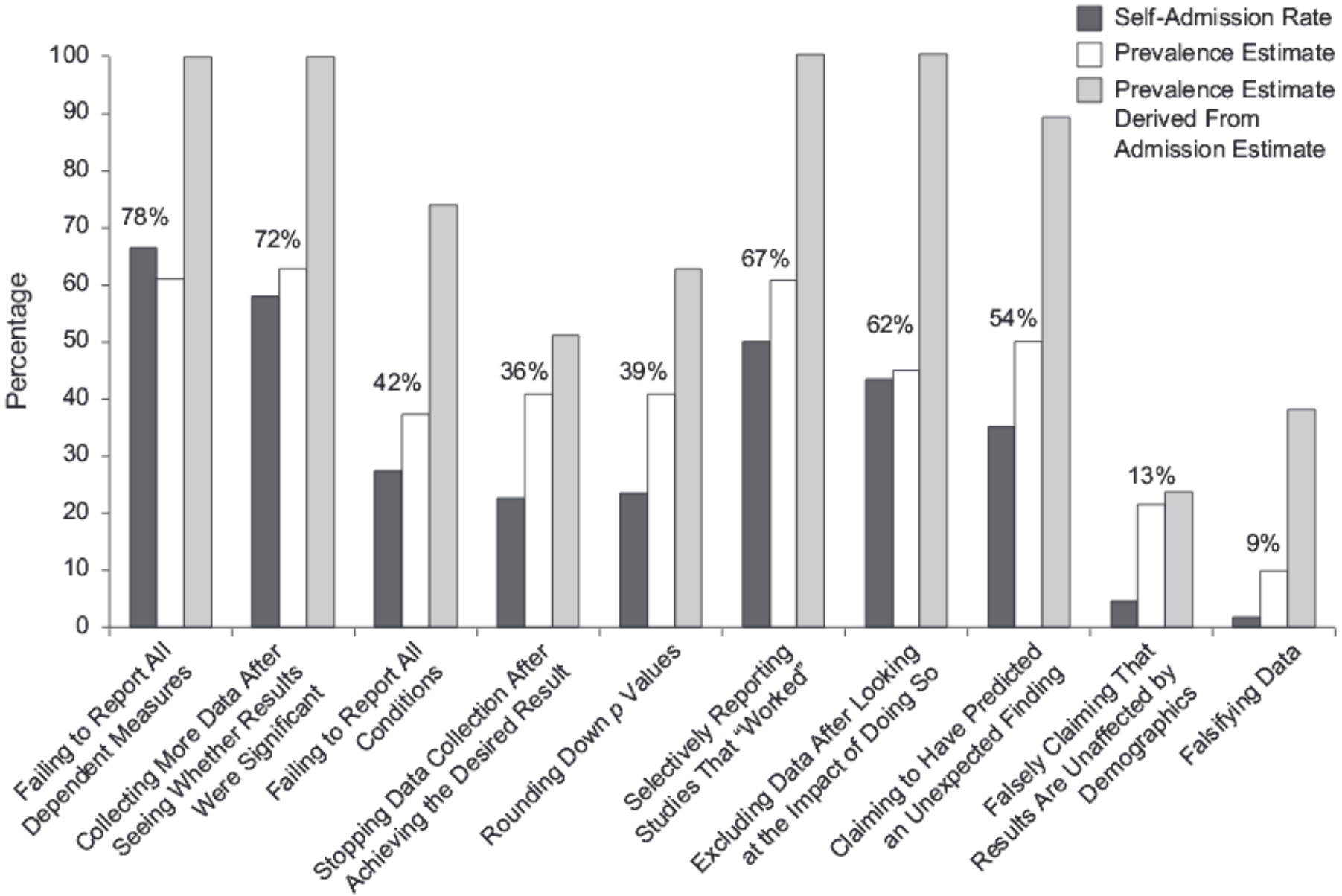


Fig. 1. Results of the Bayesian-truth-serum condition in the main study. For each of the 10 items, the graph shows the self-admission rate, prevalence estimate, prevalence estimate derived from the admission estimate (i.e., self-admission rate/admission estimate), and geometric mean of these three percentages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

Study Design and Power

- Adequately power trials to help prevent spurious significant results.
- Practical suggestions:
  - Collaborate with other labs to mutually run each others' experiments (Open Science Collaboration 2014, 2015).
    - Replication Project: Psychology
    - Many Labs 1, 2, 3
    - Crowdsourcing Analysis (Silberzahn and Uhlmann 2015)
    - Experimental economics replications (Camerer et al. 2016)
  - Maximize power subject to budget constraint by adjusting expensive treatment arm (relative) size (Duflo, Glennerster, Kremer 2007).

Study Design and Power

- Adequately power trials to help prevent spurious significant results.
- Practical suggestions:
  - Collaborate with other labs to mutually run each others' experiments (Open Science Collaboration 2014, 2015).
    - Replication Project: Psychology
    - Many Labs 1, 2, 3
    - Crowdsourcing Analysis (Silberzahn and Uhlmann 2016)
    - Experimental economics replications (Camerer et al. 2016)
  - Maximize power subject to budget constraint by adjusting expensive treatment arm (relative) size (Duflo, Glennerster, Kremer 2007).

Study Design and Power

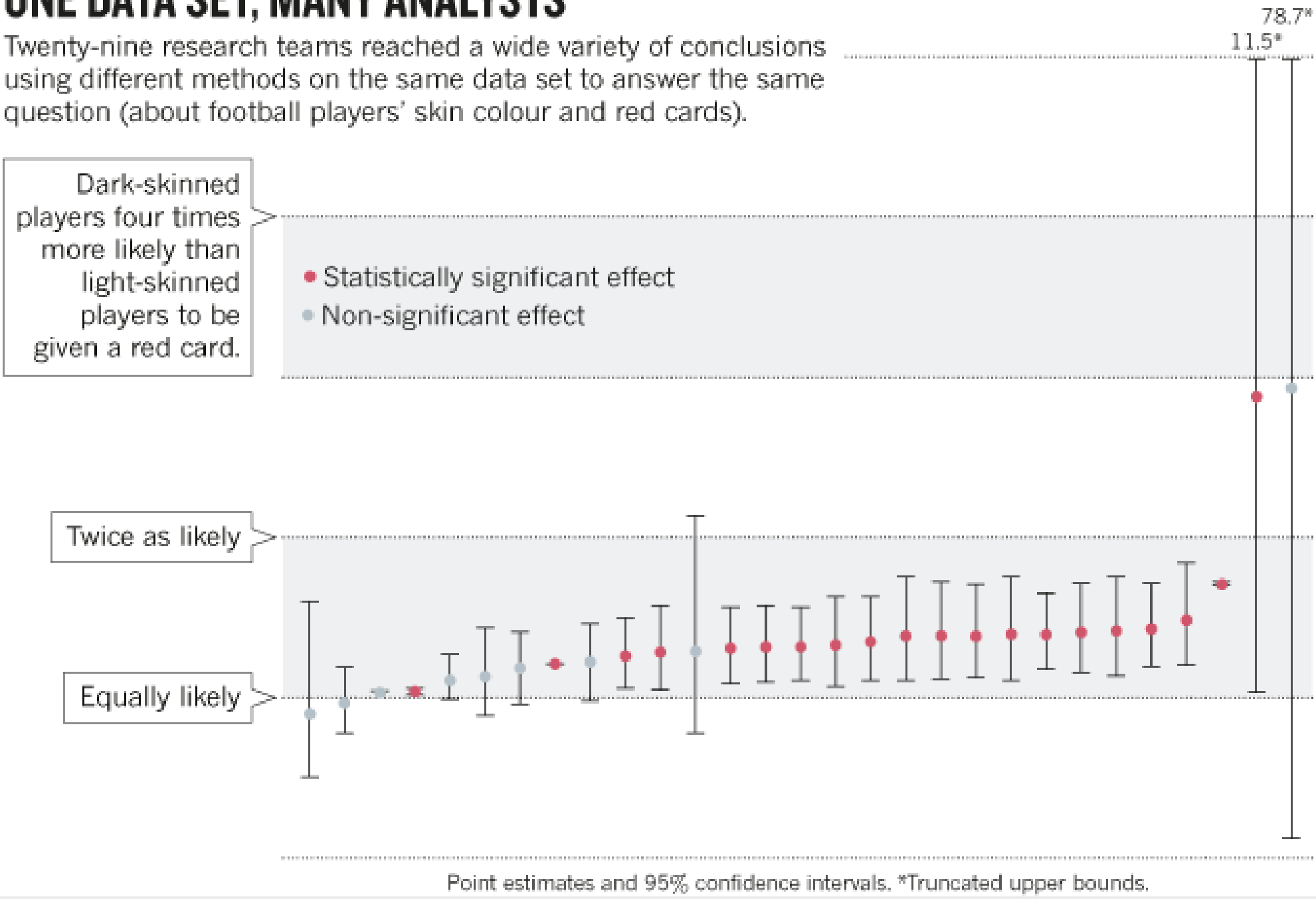
- Adequately power trials to help prevent spurious significant results.
- Practical suggestions:
  - Collaborate with other labs to mutually run each others' experiments (Open Science Collaboration 2014, 2015).
    - Replication Project: Psychology
    - Many Labs 1, 2, 3
    - Crowdsourcing Analysis (Silberzahn and Uhlmann 2016)
    - Experimental economics replications (Camerer et al. 2016)
  - Maximize power subject to budget constraint by adjusting expensive treatment arm (relative) size (Duflo, Glennerster, Kremer 2007).

Study Design and Power

- Adequately power trials to help prevent spurious significant results.
- Practical suggestions:
  - Collaborate with other labs to mutually run each others' experiments (Open Science Collaboration 2014, 2015).
    - Replication Project: Psychology
    - Many Labs 1, 2, 3
    - Crowdsourcing Analysis (Silberzahn and Uhlmann 2016)
    - Experimental economics replications (Camerer et al. 2016)
  - Maximize power subject to budget constraint by adjusting expensive treatment arm (relative) size (Duflo, Glennerster, Kremer 2007).

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).



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

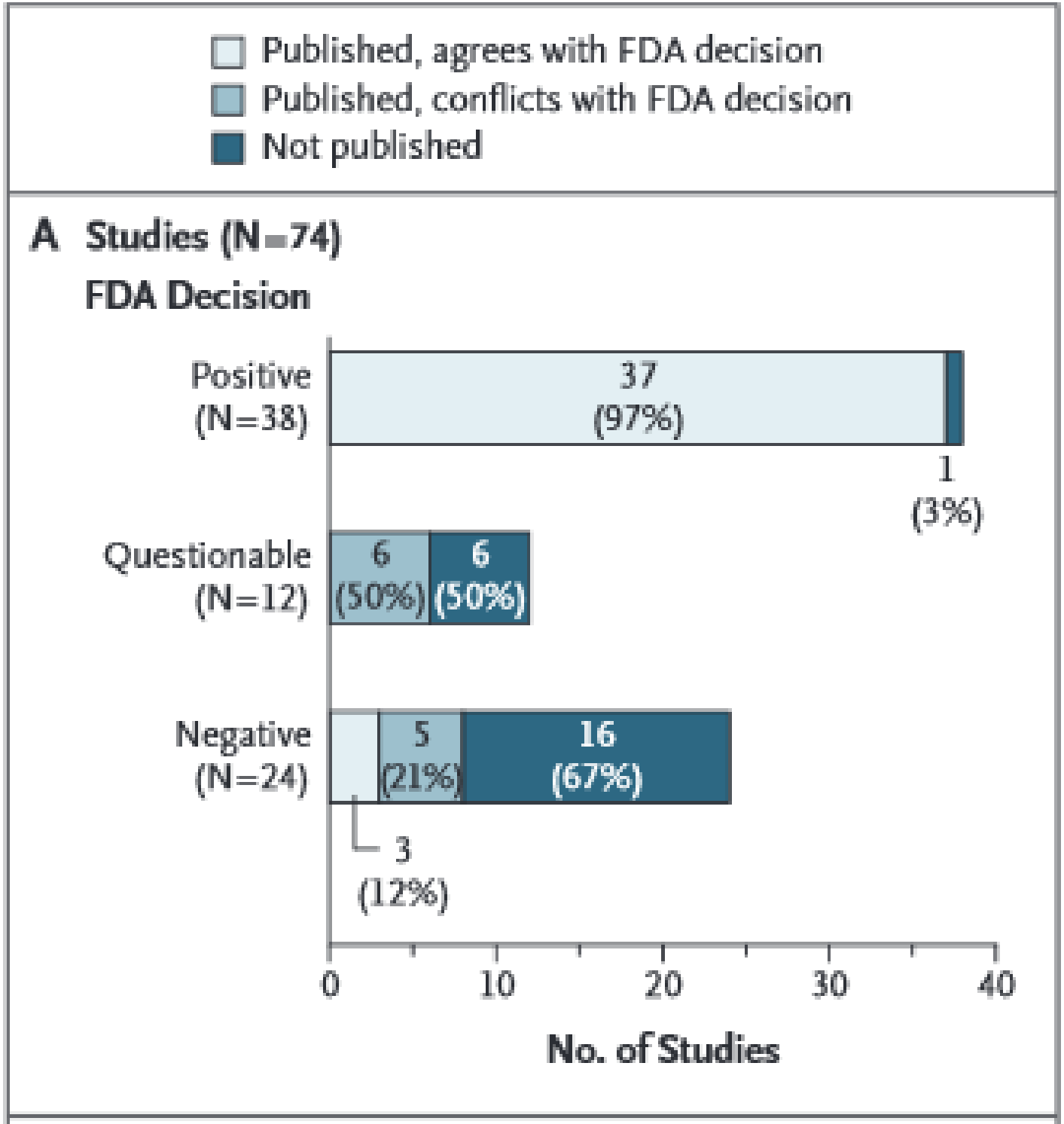
Publication Bias

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)

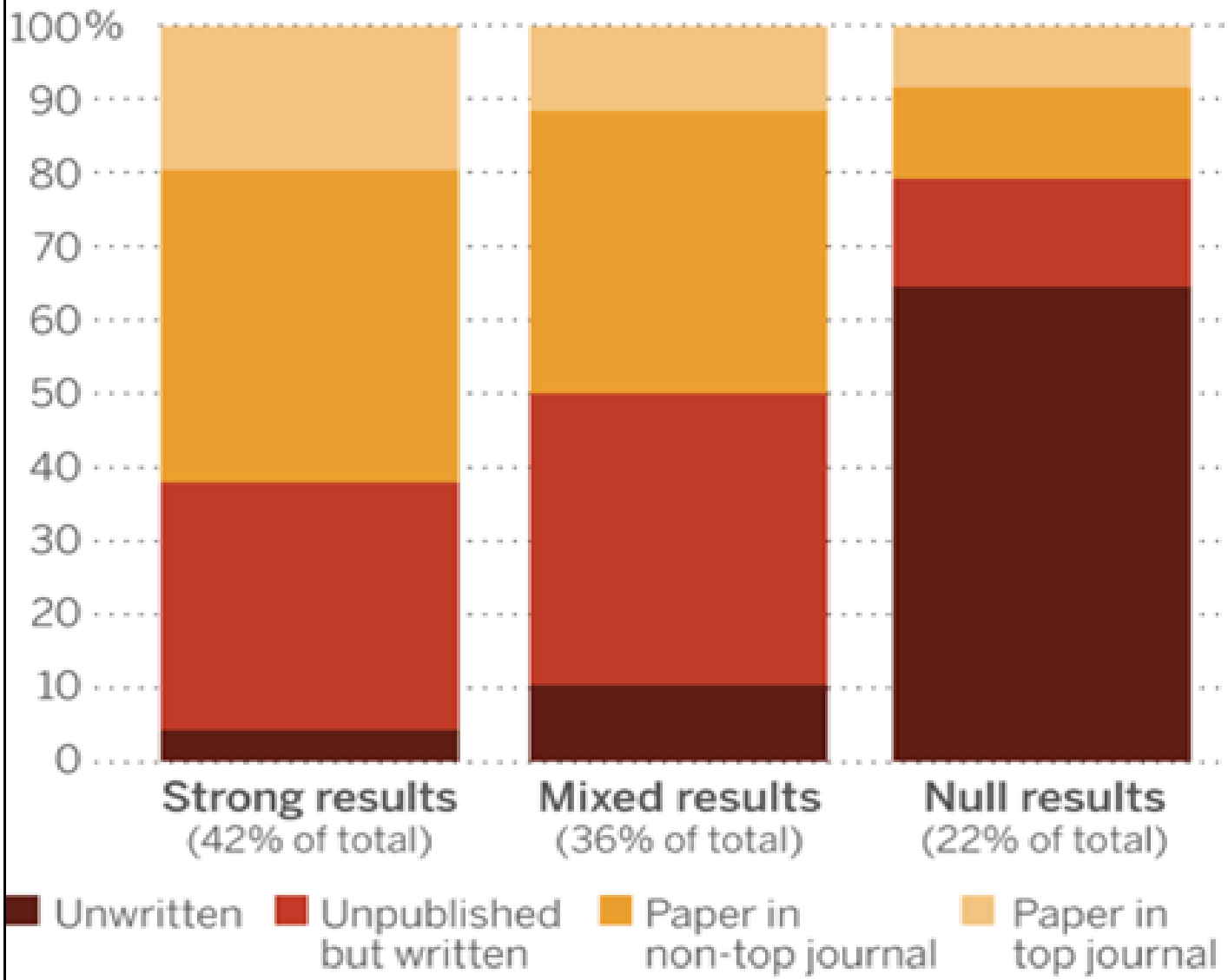
All Fields

- 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
Positive (N=38)
37 (97%)
1 (3%)
Questionable (N=12)
6 (50%)
6 (50%)
Negative (N=24)
5 (21%)
16 (67%)
3 (12%)
0
10
20
30
40
No. of Studies

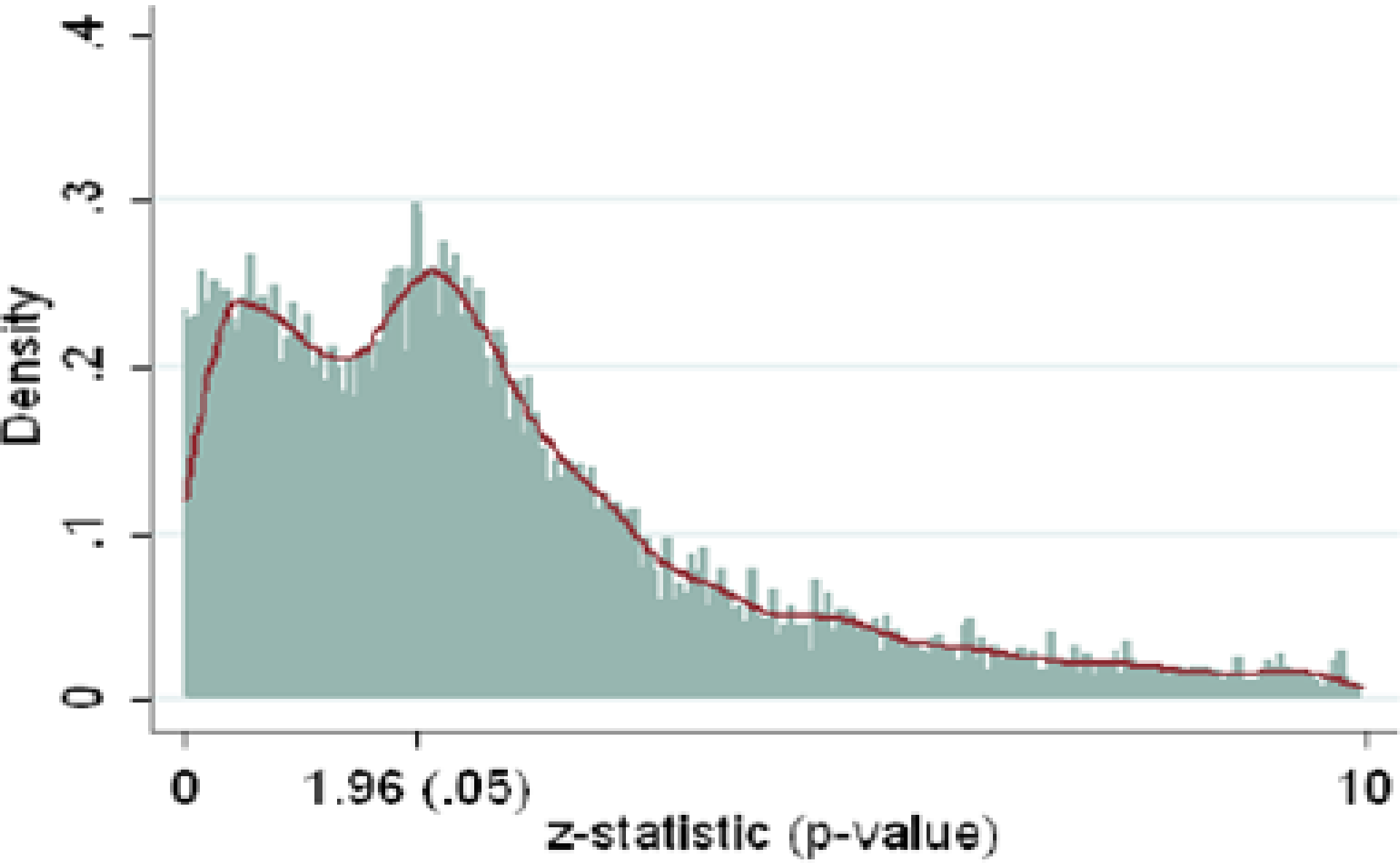
Most null results are never written up

The fate of 221 social science experiments

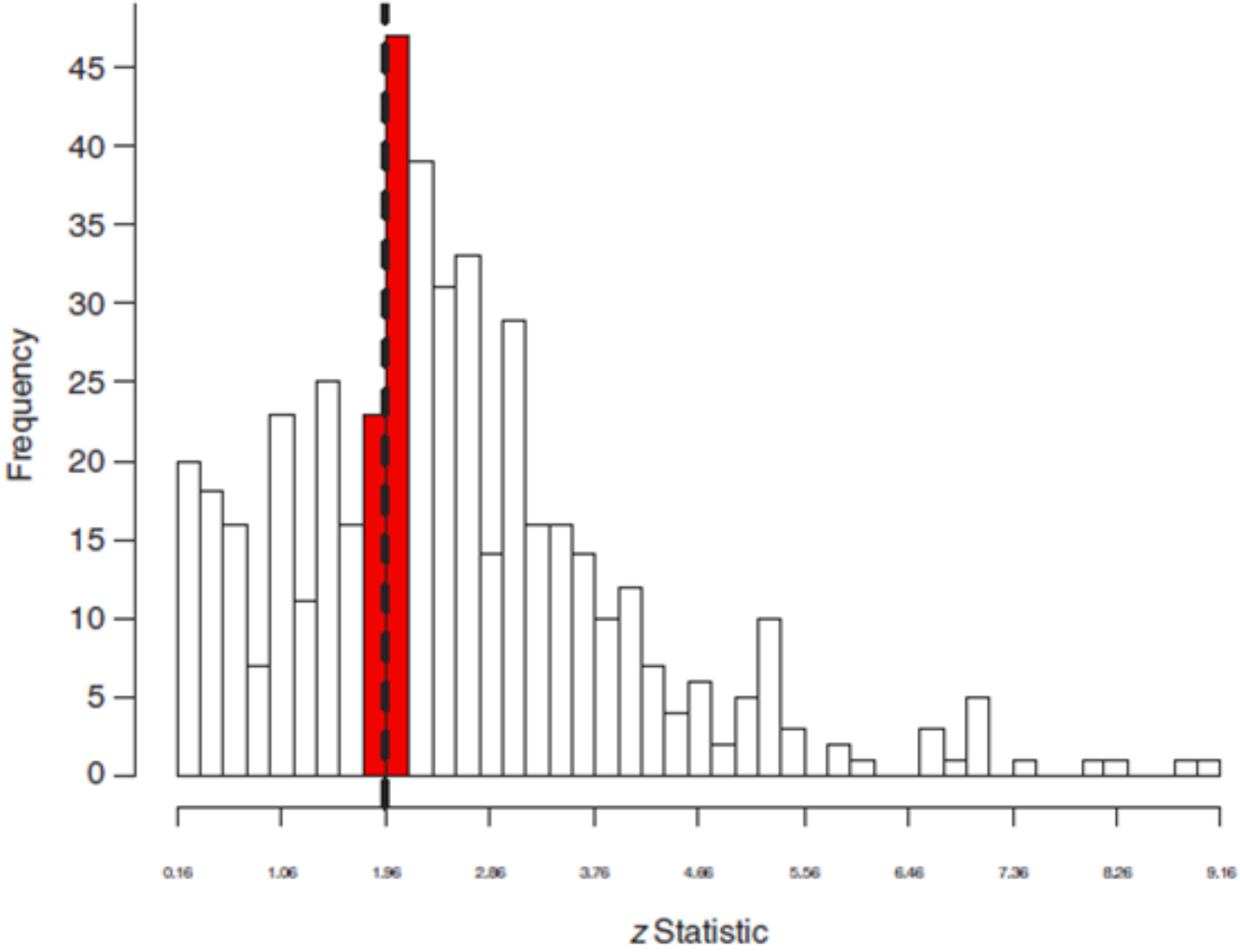


Source: A. Franco *et al.*, *Science* (28 August)

(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)



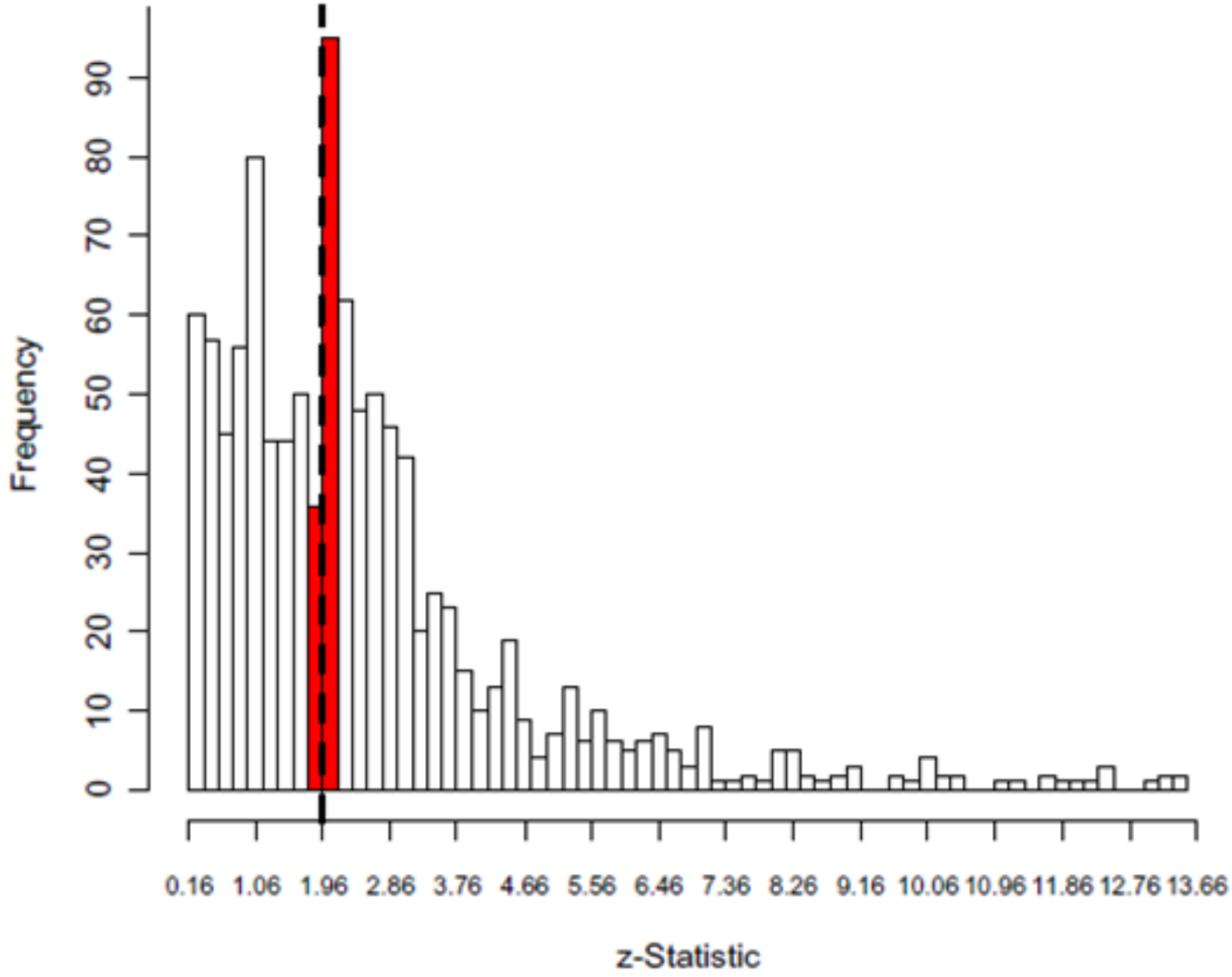


Figure 1(a). Histogram of z-statistics, *APSR & AJPS* (Two-Tailed). Width of bars (0.20) approximately represents 10% caliper. Dotted line represents critical z-statistic (1.96) associated with $p = 0.05$ significance level for one-tailed tests.

Publication Bias

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

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

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.

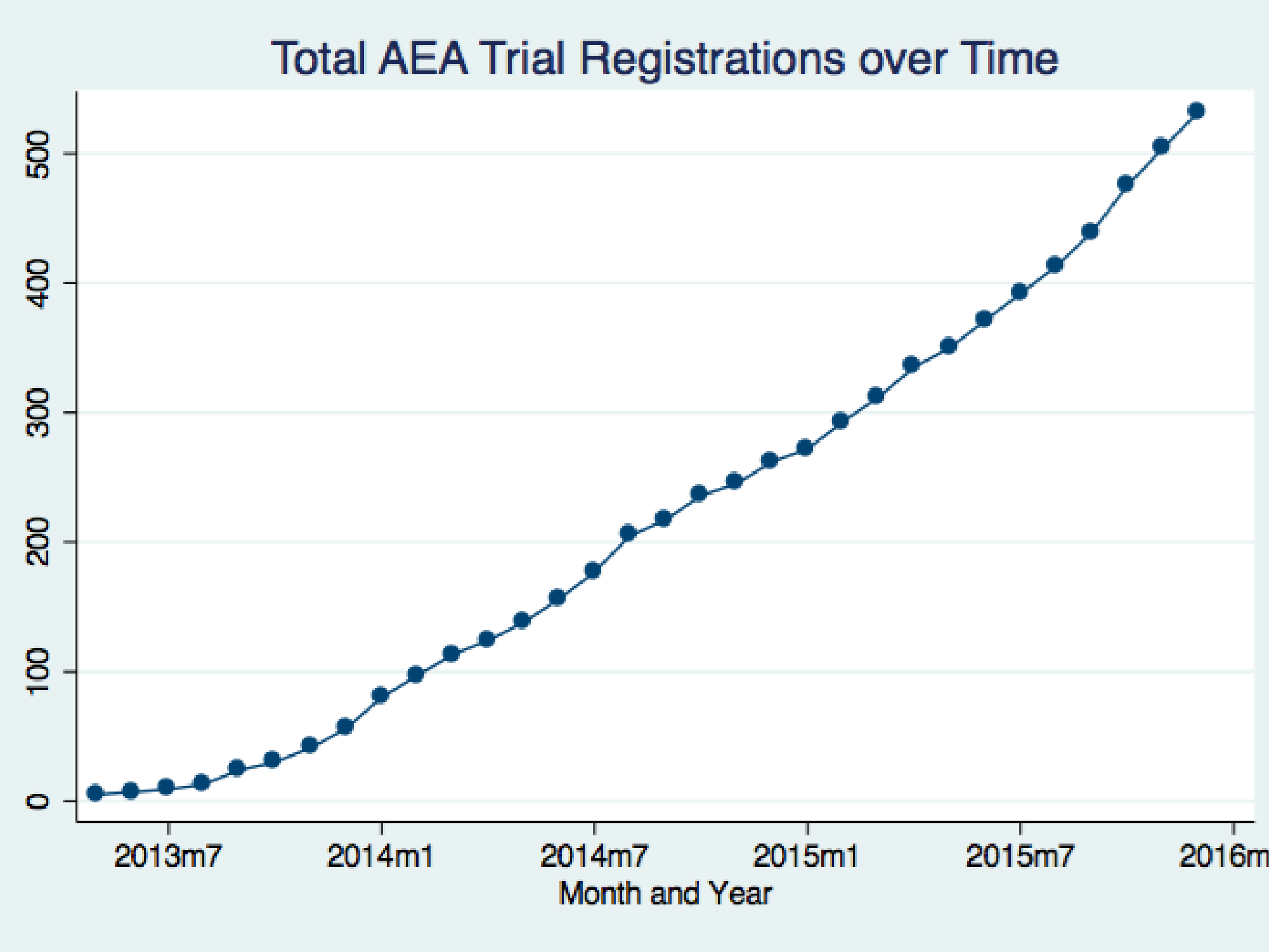
Registration

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

- 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


Total AEA Trial Registrations over Time
0
100
200
300
400
500
2013m7
2014m1
2014m7
2015m1
2015m7
2016m
Month and Year

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

Meta-Analysis

Synthesize results systematically

Organizations:

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

Meta-Analysis

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

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 threshhold; significance loses its meaning.
- Not something only evil people do. It's subconcious, 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.”

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.

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

- 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.

Replication

1. The Problem (JMCB 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.

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`

Science AAAS

Home News Journals Topics Careers

Science Science Advances Science Immunology Science Robotics Science Signaling Science Translational Medicine

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

Science 26 Jun 2015:
Vol. 348, Issue 6242, pp. 1422-1425
DOI: 10.1126/science.aab2374

Workflow

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

Workflow

- 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

Workflow

- 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

Workflow

- 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

Workflow

- 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

- 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 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

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

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

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

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

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

SSMART Grant

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

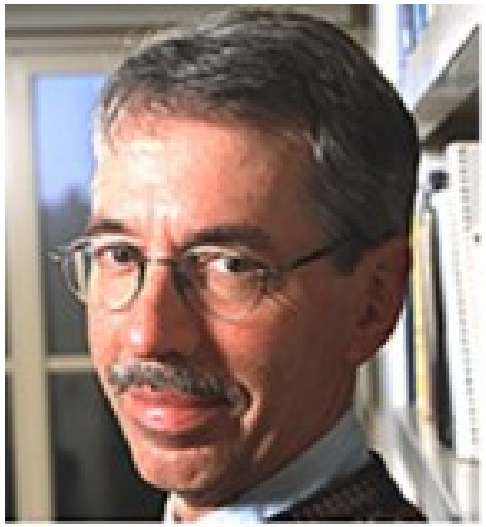
- Develop new methodology
- New tools and approaches for meta-analysis
- Research on researchers and adoption of new norms

Extra funding source for researchers from developing countries.

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

Questions?

Thank you!