BITSS

BERKELEY INITIATIVE FOR TRANSPARENCY
IN THE SOCIAL SCIENCES

Data Sharing
and
Replication

Christensen

Introduction

Project
Protocol,
Reporting
Standards

Data Sharing

Replication

Conclusion

Data Sharing and Replication
Enabling Reproducible Research

Garret Christensen[1]

[1] UC Berkeley: Berkeley Initiative for Transparency in the Social Sciences
Berkeley Institute for Data Science

New Delhi 2017

- What are problems associated with reproducibility?
- What are solutions to these problems?
- What are practical tools to implement these solutions?

Science advances by building on the work of others.

If I have seen further, it is by standing on the shoulders of giants

–Sir Isaac Newton, 1676

What prevents us from building on others' work?

- Data not shared
- Analysis not shared
- Methods/protocol not shared

What enables us to build on others' work?

- Data shared in trusted public repository
- Code/Analysis shared in trusted public repository
- Methods/protocol follow appropriate reporting standard
- Also: findings/scholarly publications available (open access)

Make sure you report everything another researcher would need to replicate your research, including the exact methods.

What to report (following medicine):

- Find the appropriate reporting standard for your field and follow it.

- Enhancing the QUAlity and Transparency Of health Research (EQUATOR Network)

- The most widely-adopted standard: Consolidated Standards of Reporting Trials (CONSORT).

- Standard Protocol Items: Recommendations for Interventional Trials (SPIRIT Statement).

- Transparency and Openness Promotion (TOP) Guidelines: http://cos.io/top

Project Protocol, Reporting Standards

Data Sharing and Replication

Christensen

Introduction

Project Protocol, Reporting Standards

Data Sharing

Replication

Conclusion

Make sure you report everything another researcher would need to replicate your research, including the exact methods.

What to report (following medicine):

- Find the appropriate reporting standard for your field and follow it.

- Enhancing the QUAlity and Transparency Of health Research (EQUATOR Network)

- The most widely-adopted standard: Consolidated Standards of Reporting Trials (CONSORT).

- Standard Protocol Items: Recommendations for Interventional Trials (SPIRIT Statement).

- Transparency and Openness Promotion (TOP) Guidelines: `http://cos.io/top`

Science

MAAAS

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

Taking up TOP

Marcia McNutt

+ Author Affiliations

Article   Figures & Data   Info & Metrics   eLetters   PDF

Nearly 1 year ago, a group of researchers boldly suggested that the standards for research quality, transparency, and trustworthiness could be improved if journals banded together to adopt eight standards called TOP (Transparency and Openness Promotion).* Since that time, more than 500 journals have been working toward their implementation of TOP. The editors at *Science* have held additional retreats

PHOTO: STACEY PENTLAND PHOTOGRAPHY

Where to report:
If not in the methods section of the article (of limited length),
supplementary online appendix linked with article or in
trusted digital repository.

- To build on the work of others, data must be shared.
- Data sharing is associated with more citations (causality unclear). Piwowar et al. 2007

History in Economics:

- Journal of Money Credit and Banking Project: Dewald, Thursby, Anderson *AER* 1986.
  - Low response rate to requests to share data.
  - Attempted to reproduce 9 papers, problems with all (some minor) even with help of original authors.

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.

TABLE 1 — RESPONSES TO REQUESTS FOR DATA FROM AUTHORS OF EMPIRICAL PAPERS[a]

	Published before Data Requested	Accepted before Data Requested	Under Review when Data Requested
Requests	62	27	65
Responses	42	26	49
Response Rate (Percent)	66	96	75
Mean Response Time (Days)	217	125	130
Not Submitted:			
Confidential Data	2	1[b]	0
Lost or Destroyed Data	14	2	1
Data Available, But Not Sent[c]	4	2	1
Nonrespondents	20	1	16
Total Not Submitted	40	6	18
Nonsubmission Rate (Percent)	66	22	28

[a] Includes all requests made through December 1984, and excludes authors whose papers were rejected.

[b] Two data sets were partially confidential.

[c] This category includes authors who (i) stated that their data were available from published sources, but did not send their data; and (ii) authors who claimed to have their data but were unwilling to sort through their papers to find the data.

History in Economics:

- Journal of Money Credit and Banking Project: Dewald, Thursby, Anderson *AER* 1986.
  - Low response rate to requests to share data.
  - Attempted to reproduce 9 papers, problems with all (some minor) even with help of original authors.
- A Decade After JMCB: Anderson and Dewald, St Louis Fed 1994.
  - Repeated similar experiment
  - Similar bleak results
- Verifying the Solution from a Nonlinear Solver, McCullough and Vinod, *AER* 2003.
  - Different software programs get you different answers.
 - But finally change—*AER* institutes data sharing requirement. Policy

History in Economics:

- Journal of Money Credit and Banking Project: Dewald, Thursby, Anderson *AER* 1986.
    - Low response rate to requests to share data.
    - Attempted to reproduce 9 papers, problems with all (some minor) even with help of original authors.
- A Decade After JMCB: Anderson and Dewald, St Louis Fed 1994.
    - Repeated similar experiment
    - Similar bleak results
- Verifying the Solution from a Nonlinear Solver, McCullough and Vinod, *AER* 2003.
    - Different software programs get you different answers.
    - But finally change—*AER* institutes data sharing requirement. Policy

How is econ doing as a discipline?

- *AER* internal review generally positive (Glandon 2010)
- Many, including McCullough, still skeptical of the ability to reproduce (Econ Journal Watch, 2007)
- Though *AER*, all AEA, and other top 5 journals have mandatory data policies, and shared data is often only the "analysis" data instead of raw data. Exemptions for proprietary/restricted data increasing.
- A study by the Replication Network shows that fewer than 27 journals regularly publish data, only 10 explicitly state they publish replications. (Duvendack et al 2015)

Year
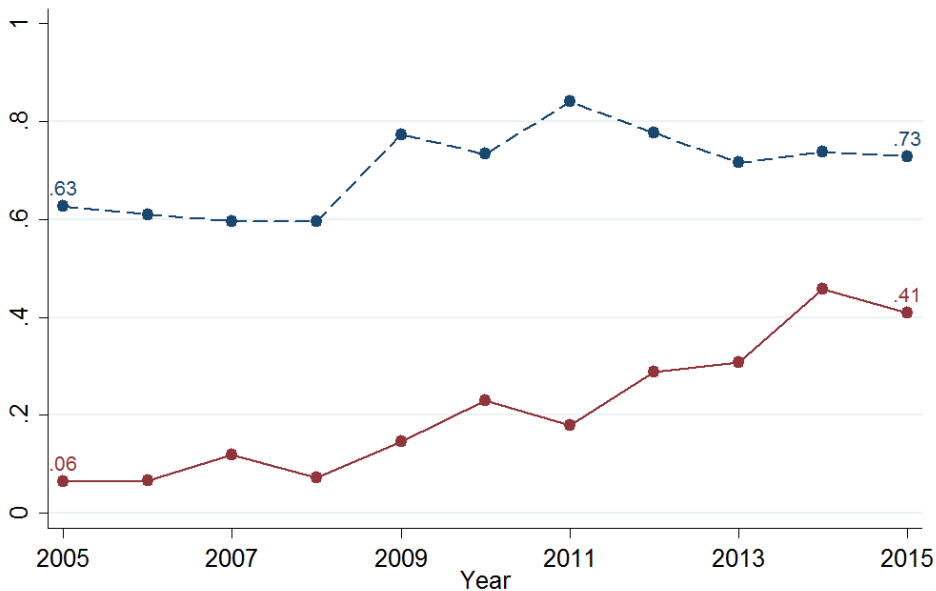
- Fraction papers with data
- Fraction data papers exempted

Why share 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.
- Repositories manage meta-data better.
- Repositories create digital citable identifiers (DOI).

Data Sharing

Examples of Trusted Repositories:

- Harvard's Dataverse
- Data Dryad
- figshare
- Open Science Framework
- OpenICPSR
- Check the journal–they may use one of these
  - *REStat*'s Dataverse

- With data available, we can begin to replicate studies.
- We should be very careful about what we mean by "replication."
- "The Meaning of Failed Replications" Michael Clemens, CGD Working Paper 399.

Table 1: A PROPOSED DEFINITION TO DISTINGUISH REPLICATION AND ROBUSTNESS TESTS

| | Sampling distribution for parameter estimates | Sufficient conditions for discrepancy | Types | Methods in follow-up study versus methods *reported* in original: | | | Examples |
				Same specification	Same population	Same sample	
Replication	*Same*	*Random chance, error, or fraud*	Verification	Yes	Yes	Yes	*Fix faulty measurement, code, dataset*
			Reproduction	Yes	Yes	No	*Remedy sampling error, low power*
Robustness	*Different*	*Sampling distribution has changed*	Reanalysis	No	Yes	Yes/No	*Alter specification, recode variables*
			Extension	Yes	No	No	*Alter place or time; drop outliers*

The "same" specification, population, or sample means the same as *reported* in the original paper, not necessarily what was contained in the code and data used by the original paper. Thus for example if code used in the original paper contains an error such that it does not run exactly the regressions that the original paper said it does, new code that fixes the error is nevertheless using the "same" specifications (as described in the paper).

Why Replicate? Motivation and suggestions from Nicole Janz of Political Science Replication and Cambridge University

- For science in general:
  - Uncover misconduct and sloppy science
  - Confirm previous findings and generalizability
  - Point to misuse of statistical methods
- For you as researchers:
  - Learn statistics
  - Jump to research frontier
  - Publish
  - Make your own research routinely reproducible
  - Fun

Why Replicate? Motivation and suggestions from Nicole Janz of Political Science Replication and Cambridge University

- For science in general:
  - Uncover misconduct and sloppy science
  - Confirm previous findings and generalizability
  - Point to misuse of statistical methods
- For you as researchers:
  - Learn statistics
  - Jump to research frontier
  - Publish
  - Make your own research routinely reproducible
  - Fun

Replication

Data Sharing
and
Replication

Christensen

Introduction

Project
Protocol,
Reporting
Standards

Data Sharing

Replication

Conclusion

Why Replicate? Motivation and suggestions from Nicole Janz of Political Science Replication and Cambridge University

- For science in general:
    - Uncover misconduct and sloppy science
    - Confirm previous findings and generalizability
    - Point to misuse of statistical methods
- For you as researchers:
    - Learn statistics
    - Jump to research frontier
    - Publish
    - Make your own research routinely reproducible
    - Fun

Which study should you pick to replicate? (Janz 2015)

- Don't select a study with methods that you don't know or can't learn within a reasonable time.
- Pick a recent study (<5 yo) from a good journal.
- Data (and code) should be publicly available.
- The journal that published the original study has published replications before.

Which journals publish replications?

- List from The Replication Network study, Duvendack et al.
- Sadly fairly limtied in economics (10).
- Selected journals from Janz (2015)

TABLE 2. Journals whose websites explicitly mention that they publish replications

1)	*Econ Journal Watch*
2)	*Economic Development and Cultural Change*
3)	*Economics of Education Review*
4)	*Empirical Economics*
5)	*Experimental Economics*
6)	*Explorations in Economic History*
7)	*International Journal of Forecasting*
8)	*Jahrbücher für Nationalökonomie und Statistik/Journal of Economics and Statistics*
9)	*Journal of Applied Econometrics*
10)	*Review of International Organizations*

Journals Open to Replication (selection)
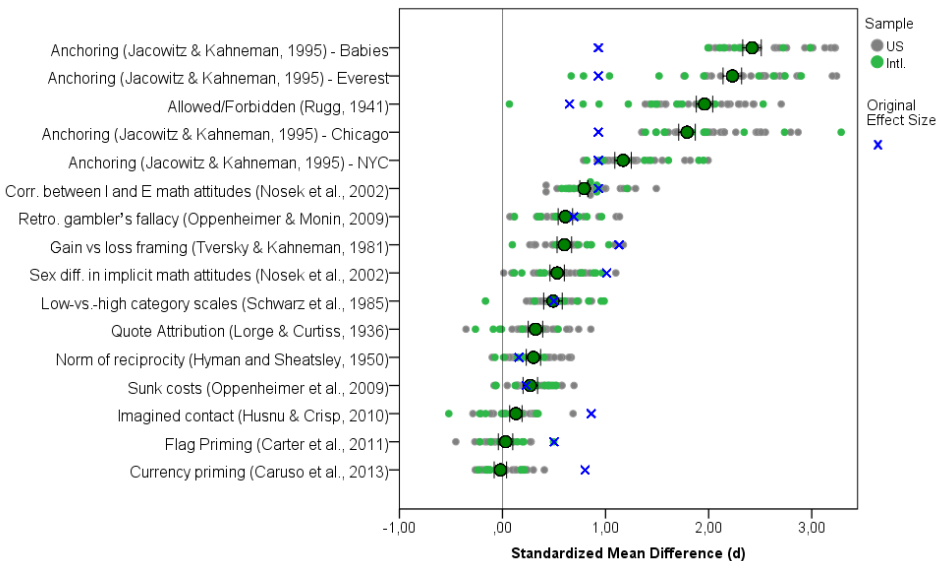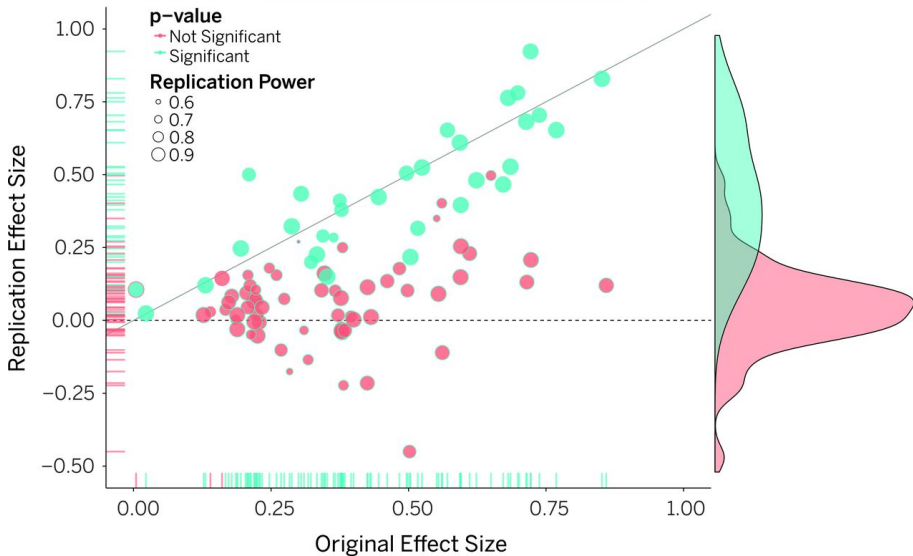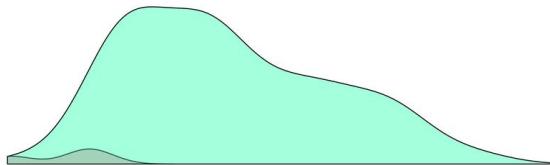

Political Science

Psychology

Economics

*original study was published in the same journal
+ home of the original 'Many Labs' project
# special issue dedicated to replications (March 2015)
^this journal invites replication studies

How exactly to replicate?

- Be systematic: write a pre-analysis plan.
- Don't just go on a fishing expedition. We all know that if you dig hard enough, you can find a specification that makes results appear weaker. Don't selectively report those specifications.
- Be courteous and professional.
- Take an entirely systematic approach:
    - Many Labs Project
    - Crowdsource your analysis

	Sample
	US (grey)
	Intl. (green)
	Original Effect Size (blue ×)

Anchoring (Jacowitz & Kahneman, 1995) - Babies
Anchoring (Jacowitz & Kahneman, 1995) - Everest
Allowed/Forbidden (Rugg, 1941)
Anchoring (Jacowitz & Kahneman, 1995) - Chicago
Anchoring (Jacowitz & Kahneman, 1995) - NYC
Corr. between I and E math attitudes (Nosek et al., 2002)
Retro. gambler's fallacy (Oppenheimer & Monin, 2009)
Gain vs loss framing (Tversky & Kahneman, 1981)
Sex diff. in implicit math attitudes (Nosek et al., 2002)
Low-vs.-high category scales (Schwarz et al., 1985)
Quote Attribution (Lorge & Curtiss, 1936)
Norm of reciprocity (Hyman and Sheatsley, 1950)
Sunk costs (Oppenheimer et al., 2009)
Imagined contact (Husnu & Crisp, 2010)
Flag Priming (Carter et al., 2011)
Currency priming (Caruso et al., 2013)

Standardized Mean Difference (d)

-1,00   ,00   1,00   2,00   3,00

Team	Analytic Approach	OR
12	Zero-inflated Poisson regression	0.89
17	Bayesian logistic regression	0.96
15	Hierarchical log-linear modeling	1.02
10	Multilevel regression and logistic regression	1.03
18	Hierarchical Bayes model	1.10
31	Logistic regression	1.12
1	Ordinary least squares with robust standard errors, logistic regression	1.18
4	Spearman correlation	1.21
14	Weighted least squares regression with referee fixed-effects and clustered standard errors	1.21
11	Multiple linear regression	1.25
30	Clustered robust binomial logistic regression	1.28
6	Linear Probability Model	1.28
26	Three-level hierarchical generalized linear modeling with Poisson sampling	1.30
3	Multilevel Binomial Logistic Regression using bayesian inference	1.31
23	Mixed model logistic regression	1.31
16	Hierarchical Poisson Regression	1.32
2	Linear probability model, logistic regression	1.34
5	Generalized linear mixed models	1.38
24	Multilevel logistic regression	1.38
28	Mixed effects logistic regression	1.38
32	Generalized linear models for binary data	1.39
8	Negative binomial regression with a log link analysis	1.39
20	Cross-classified multilevel negative binomial model	1.40
13	Poisson Multi-level modeling	1.41
25	Multilevel logistic binomial regression	1.42
9	Generalized linear mixed effects models with a logit link function	1.48
7	Dirichlet process Bayesian clustering	1.71
21	Tobit regression	2.88
27	Poisson regression	2.93
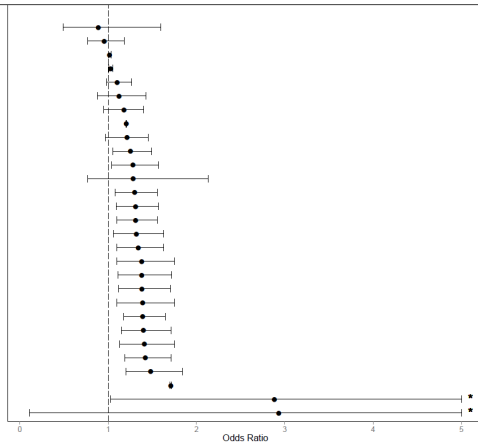
ONE DATA SET, MANY ANALYSTS

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*
11.5*

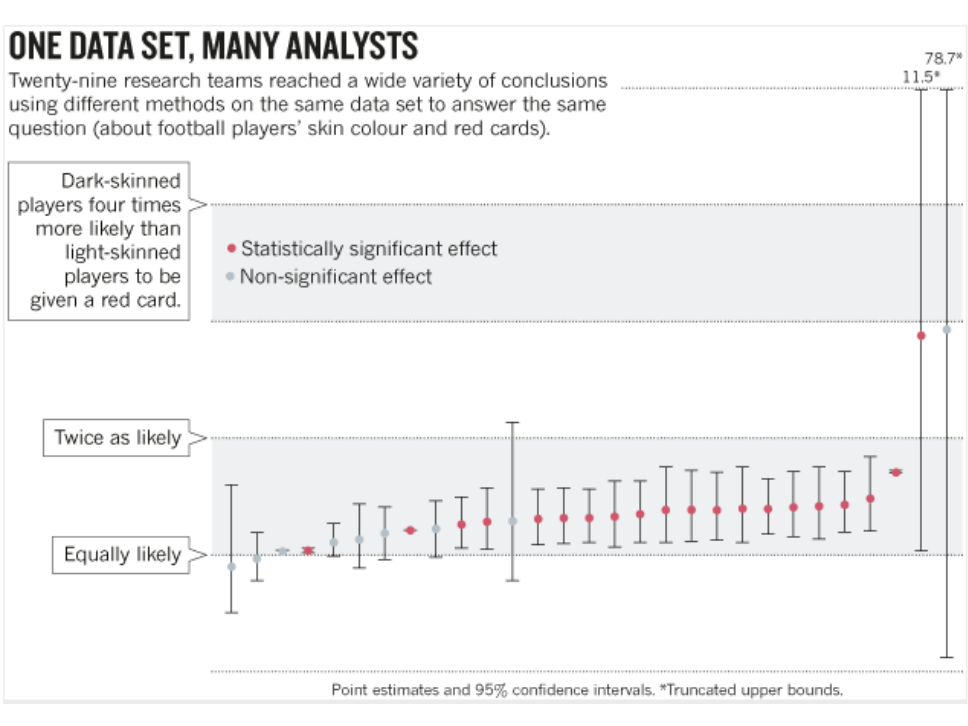Dark-skinned players four times more likely than light-skinned players to be given a red card.

● Statistically significant effect
● Non-significant effect

Twice as likely

Equally likely

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

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

- Science builds on previous work
- To do that, work must be public
- Share your data and code publicly
- Replicate the work of others