



BERKELEY INITIATIVE FOR TRANSPARENCY
IN THE SOCIAL SCIENCES

Short Paper
Title

Christensen

Introduction

Project
Protocol,
Reporting
Standards

Data Sharing

Replication

Data Sharing and Replication

Enabling Reproducible Research

Garret Christensen¹

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

APHRC, Summer 2015



Outline

Short Paper
Title

Christensen

Introduction

Project
Protocol,
Reporting
Standards

Data Sharing

Replication

1 Introduction

2 Project Protocol, Reporting Standards

3 Data Sharing

4 Replication



Reproducibility & Transparency

Short Paper
Title

Christensen

Introduction

Project
Protocol,
Reporting
Standards

Data Sharing

Replication

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



Introduction

Short Paper
Title

Christensen

Introduction

Project
Protocol,
Reporting
Standards

Data Sharing

Replication

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)



BITSS

BERKELEY INITIATIVE FOR TRANSPARENCY
IN THE SOCIAL SCIENCES

Project Protocol, Reporting Standards

Short Paper
Title

Christensen

Introduction

Project
Protocol,
Reporting
Standards

Data Sharing

Replication

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).

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).



BITSS

BERKELEY INITIATIVE FOR TRANSPARENCY
IN THE SOCIAL SCIENCES

Project Protocol, Reporting Standards

Short Paper
Title

Christensen

Introduction

Project
Protocol,
Reporting
Standards

Data Sharing

Replication

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.



Data Sharing

Short Paper
Title

Christensen

Introduction

Project
Protocol,
Reporting
Standards

Data Sharing

Replication

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.

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

^aIncludes all requests made through December 1984, and excludes authors whose papers were rejected.

^bTwo data sets were partially confidential.

^cThis 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.



- 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 are we 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 journals have a good policy, enforcement is limited, and shared data is often only the “analysis” data instead of raw data, and *QJE* has no policy whatsoever.
- 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)

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.



Data Sharing

Short Paper
Title

Christensen

Introduction

Project
Protocol,
Reporting
Standards

Data Sharing

Replication

Data repositories
Harvard's Dataverse
Check the journal



APHRC Repository

Short Paper
Title

Christensen

Introduction

Project
Protocol,
Reporting
Standards

Data Sharing

Replication

APHRC has its own repository. You should use it.



Replication

BERKELEY INITIATIVE FOR TRANS-PARACITY
IN THE SOCIAL SCIENCES

Short Paper
Title

Christensen

Introduction

Project
Protocol,
Reporting
Standards

Data Sharing

Replication

Define Replication Michael Clemens



Replication

BERKELEY INITIATIVE FOR TRANSPARENCY
IN THE SOCIAL SCIENCES

Short Paper
Title

Christensen

Introduction

Project
Protocol,
Reporting
Standards

Data Sharing

Replication

How to do a replication



Replication

BERKELEY INITIATIVE FOR TRANSPARENCY
IN THE SOCIAL SCIENCES

Short Paper
Title

Christensen

Introduction

Project
Protocol,
Reporting
Standards

Data Sharing

Replication

Encourage it, make it publishable



Conclusion

Short Paper
Title

Christensen

Introduction

Project
Protocol,
Reporting
Standards

Data Sharing

Replication

- Share your data
- Replicate