

# **Project TIER:**

## **Teaching Transparency in Empirical Research**

Richard Ball  
Associate Professor of Economics  
Haverford College  
[rball@haverford.edu](mailto:rball@haverford.edu)

[www.haverford.edu/TIER](http://www.haverford.edu/TIER)

@Project\_TIER

Presented at the INET/YSI Workshop on  
Replication and Transparency in Economic Research  
San Francisco  
Day One: January 6, 2016

Project TIER is supported by a grant from the Alfred P. Sloan Foundation.

Thanks to Jan Hoeffler for the outstanding work he has done  
to conceive of and organize this workshop.

And thanks to the INET Young Scholars Initiative  
for providing support that made it possible.

## A prefatory comment on terminology:

I will use the terms *replicability* and *replication* somewhat loosely.

In some subcultures of the research transparency universe, people would say that what I am talking about should be called *reproducibility*, or more precisely *computational reproducibility*.

## Many Dimensions of Research Transparency

Videos of the complete set of lectures for an entire semester-long graduate-level course on research transparency in the social sciences, by Ted Miguel (Economics, UC Berkeley), are available at [https://www.youtube.com/watch?v=O3GBoVwQYwY&list=PL-XXv-cvA\\_iBN9JZND3CF91aouSHH9ksB&index=1](https://www.youtube.com/watch?v=O3GBoVwQYwY&list=PL-XXv-cvA_iBN9JZND3CF91aouSHH9ksB&index=1).

The issues subsumed within the concept of transparency include:

- Various notions of replication and computational reproducibility

- Pre-analysis plans and pre-registration of hypotheses and methods

- The “file drawer problem”

- P-hacking

## Pre-analysis plans and pre-registration

American Economic Association Randomized Control Trial Registry:  
<https://www.socialscienceregistry.org/>

Olken, Benjamin A (2015). "Promises and Perils of Pre-analysis Plans." *Journal of Economic Perspectives* 29(3): 61-80.

Coffman, Lucas C., and Muriel Niederle (2015). "Pre-analysis Plans Have Limited Upside, Especially Where Replications Are Feasible." *Journal of Economic Perspectives* 29(3): 81-98.

## The “file drawer problem”

Rosenthal, R. (1979). “The file drawer problem and tolerance for null results.” *Psychological Bulletin* 86: 638–641.

Franco, Annie, Neil Malhotra and Gabor Simonovits (2014). “Publication bias in the social sciences: Unlocking the file drawer.” *Science* 345(6203): 1502-1505.

PsychFileDrawer: <http://psychfiledrawer.org/>

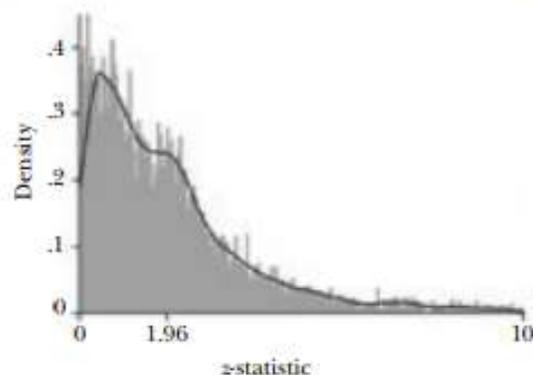
# P-hacking

Simmons, Joseph P., Leif D. Nelson, and Uri Simonsohn (2011). “False-Positive Psychology: Undisclosed Flexibility in Data Collection and Analysis Allows Presenting Anything as Significant.” *Psychological Science* 22(11): 1359–66.

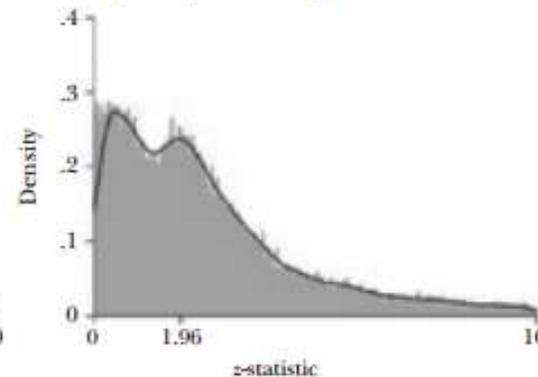
Figure 1

## Evidence of *p*-hacking

A: Laboratory experiments or randomized control trials data



B: Other [nonexperimental] data



Source: Figures 6e and f from Brodeur, L , Sangnier, and Zylberberg (forthcoming).

Notes: Displays distribution of z-statistics reported in all papers appearing in either the *American Economic Review*, *Journal of Political Economy*, or *Quarterly Journal of Economics* between 2005 and 2011. Experiments, both lab and field, are in the left panel; all other papers in the right panel.

The above figure is reproduced from: Coffman, Lucas C. and Muriel Niederle (2015). “Pre-Analysis Plans Have Limited Upside, Especially Where Replications Are Feasible.” *Journal of Economic Perspectives* 29(3): 81-98.

## Initiatives to Promote Transparency in Research Practice

Many initiatives were launched between 2012 and 2014, including:

Berkeley Initiative for Transparency in the Social Sciences (BITSS)

[www.bitss.org](http://www.bitss.org)

@UCBITSS

Hosts conferences and training institutes

Offers research grants and prizes

Has developed a manual of best-practices:

<https://github.com/garretchristensen/BestPracticesManual/blob/master/Manual.pdf>

## Center for Open Science (COS)

Based in Charlottesville, VA

[www.osf.io](http://www.osf.io)

@OSFramework

Developers of the Open Science Framework (OSF) platform  
for managing and sharing research documents

Two major projects on replicability of experimental research:

In psychology:

[https://osf.io/ezcuj/wiki/home/?\\_ga=1.169643940.48610305.1442958193](https://osf.io/ezcuj/wiki/home/?_ga=1.169643940.48610305.1442958193)

In cancer research:

[https://osf.io/e81xl/wiki/home/?\\_ga=1.208235415.48610305.1442958193](https://osf.io/e81xl/wiki/home/?_ga=1.208235415.48610305.1442958193)

## The Replication Wiki

<http://replication.uni-goettingen.de>

## Political Science Replication:

<http://projects.iq.harvard.edu/psreplication/home>

## Cambridge Replication Workshop:

<http://schreiberin.de/teaching/replication.html>

## Project TIER

Based at Haverford College

[www.haverford.edu/TIER](http://www.haverford.edu/TIER)

@Project\_TIER

Promotes the integration transparency and replicability in the research training of undergraduate and graduate students in the social sciences, with a focus on computational reproducibility.

## **Some Historical Context on the Issue of Computational Reproducibility**

Concern about the computational reproducibility of published economic research was sparked by a 1986 study known as the “*Journal of Money, Credit and Banking (JMCB) Project*.”

Dewald, William G., Jerry G. Thursby, and Richard G. Anderson (1986). “Replication in Empirical Economics: The *Journal of Money, Credit and Banking Project*.” *American Economic Review* 76(4):587-603.

## The *JMCB* Project

Editors of the *JMCB* attempted to reproduce the statistical results reported in all the empirical papers published in that journal in the preceding five years.

Requests for replication data and code were sent to authors of 154 papers.

In 37 cases (24%), the authors did not reply to the request.

In 24 cases (16%), the authors replied, but either refused to send data and code, or said they would but never did.

In 3 cases (2%), the authors said they could not provide the data because it was proprietary or confidential.

In the remaining 90 cases (58%), the authors sent *some information* in response to the request.

## **The *JMCB* Project (continued)**

Out of the 90 submissions received, the first 54 were investigated for completeness and accuracy.

Out of the 54 submissions that were investigated, the documentation provided by the authors of the papers successfully replicated the results of their papers in only 8 (15%) of the cases.

The remaining 46 (85%) of the papers could not be replicated because the information the authors submitted was insufficiently complete or precise.

## Conclusions of the *JMCB* Project

The authors of the *JMCB* study concluded:

“Our findings suggest that inadvertent errors in published empirical articles are a commonplace rather than a rare occurrence.”

and

“...we recommend that journals require the submission of programs and data at the time empirical papers are submitted. The description of sources, data transformations, and econometric estimators should be so exact that another researcher could replicate the study and, it goes without saying, obtain the same results.”

## Impact of the *JMCB* Project

It seems to me that the publication of the *JMCB* study should have hit the economics profession like a tsunami. But it did not.

It did, however, lead the *American Economic Review*, the flagship journal of the American Economic Association, to adopt a “data availability” policy. In the same issue of the *AER* in which the *JMCB* Project paper appeared, an editorial announcement stated (in part):

It is the policy of the *American Economic Review* to publish papers only where the data used in the analysis are clearly and precisely documented, are readily available to any researcher for purposes of replication, and where details of the computations sufficient to permit replication are provided.

This original *AER* data availability policy was weaker than the policy suggested in the *JMCB* Project paper.

The suggestion in the *JMCB* paper was that journals should require submission of replication documentation should be “at the time empirical papers are submitted.”

The original *AER* policy stated only that documentation should be “readily available.” It did not require authors to submit any data or documentation to the journal.

Another study documenting non-replicability of economic research came out in 2003:

McCullough, Bruce D., and H.D. Vinod (2003). “Verifying the Solution from a Nonlinear Solver: A Case Study,” *American Economic Review* 93(3): 873-892.

In this case, the articles found not to be replicable had appeared in the *American Economic Review*.

In response, the *AER* strengthened its data availability policy. It began requiring authors of papers accepted for publication in the journal to submit replication data and code, which would then be posted on the journals’ website.

The data availability policy currently in effect at the *AER* states, in part:

... Authors of accepted papers... must provide to the *Review*...the data, programs, and other details of the computations sufficient to permit replication. These will be posted on the *AER* Web site.

...the minimum requirement should include the data set(s) and programs used to run the final models, plus a description of how previous intermediate data sets and programs were employed to create the final data set(s). Authors are invited to submit these intermediate data files and programs as an option; if they are not provided, authors must fully cooperate with investigators seeking to conduct a replication who request them... Authors must provide a Readme PDF file listing all included files and documenting the purpose and format of each file provided, as well as instructing a user on how replication can be conducted.

I will discuss the strengths and weaknesses of the *AER* data availability policy later.

For now, I will just note that:

Most of the top journals in economics have adopted the *AER* data availability policy, verbatim or nearly so.

Nonetheless, evidence continues to accumulate that many published articles cannot be replicated. For a few examples, see:

McCullough, Bruce D., Kerry Anne McGeary, and Teresa D. Harrison (2006). “Lessons from the *JMCB* Archive,” *Journal of Money, Credit and Banking* 38(4): 1093-1107.

McCullough, Bruce D., Kerry Anne McGeary, and Teresa D. Harrison (2008). “Do Economics Journal Archives Promote Replicable Research?” *Canadian Journal of Economics* 41(4): 1406-1420.

Chang, Andrew C., and Phillip Li (2015). “Is Economics Research Replicable? Sixty Published Papers from Thirteen Journals Say “Usually Not.”” Finance and Economics Discussion Series 2015-083. Washington: Board of Governors of the Federal Reserve System, <http://dx.doi.org/10.17016/FEDS.2015.083>.

# **Replicability of empirical research in other social sciences: political science, psychology and sociology**

## POLITICAL SCIENCE

Style of research and nature of concerns about replicability similar to economics.

One recent high-profile case:

LaCour, Michael J., and Donald P. Green (2014). “When Contact Changes Minds: An Experiment on Transmission of Support for Gay Equality,” *Science* 346(6215): 1366-1369.

McNutt, Marcia (2015). “Expression of Concern on LaCour and Green, *Science* 346 (6215) 1366-1369.” *Science* 348(6239): 1100.

## POLITICAL SCIENCE (continued)

I don't know whether systematic assessments of replicability have been undertaken in political science. But concern among statistically-oriented political scientists has been high—much higher than in economics.

The Data Access-Research Transparency (DA-RT) initiative has formulated a set of principles for documenting statistical research, and many journals have or planned to adopt policies that embody these principles. See <http://www.dartstatement.org/>.

The *American Journal of Political Science* (<http://ajps.org/>) has adopted an exceptionally strong and carefully enforced policy (<http://ajps.org/ajps-replication-policy/>).

But there has been some push-back, largely from political scientists whose research does not involve analysis of statistical data. See <http://dialogueondart.org/2015/11/13/delay-petition-signed-by-1173-sent-to-jets-editors/>.

## PSYCHOLOGY

Nature of concerns somewhat different from economics because of the preponderance of experimental work in psychology.

A distinction is made between “*reproducing*” *computational results* (with the same data that were used for the original study) and “*replicating*” *experiments* (to see whether the new data lead to the same conclusions as the data collected when the experiment was originally run).

In a few minutes I will talk about an important study that just came out on the replicability of experimental research in psychology.

## SOCIOLOGY

I checked the “instructions to authors” for seven of the top-ranked journals in sociology, and in none of them did I find any statement about a requirement to submit replication data and code for papers that included statistical analyses.

In a couple of papers (though not in most of those I examined) there were notes that gave links to websites where the authors had (apparently voluntarily) made replication documentation available.

I am not aware of any systematic assessments of the replicability of statistical research in sociology, or of any initiatives to develop data availability policies for sociology journals.

## SOCIOLOGY (continued)

It happens, however, that one of the most important advocates of transparent and replicable statistical research in the social sciences is a sociologist, namely Scott Long of Indiana University (<http://www.indiana.edu/~jslsoc/index.htm>).

In particular, his book titled *The Workflow of Data Analysis using Stata*, is the definitive guide to reproducible research methods (at least for Stata users—but most of the principles apply to research conducted with any programmable software). See <http://www.stata.com/bookstore/workflow-data-analysis-stata/>.

One way of approaching the issue of replication can be called *ex post replication*.

This is what the Replication Wiki is about, as well as proposals for publishing replications (even creating a new journal to do just that).

*Ex post replication* is also used a pedagogical device in econometrics classes and in courses or seminars focused on training students in empirical research methods.

In this presentation I'll take a different approach to the issue of replication. I will call this approach *ex ante documentation*.

*Replication documentation*, prepared by the author of a paper and made available along with the printed (or pdf) paper, should make *ex post* replication easy.

So this talk is about:

What constitutes good replication documentation

Some practical ideas about efficient ways for researchers to construct documentation for their empirical papers

How to teach incoming cohorts of social scientists principles and practices of good documentation so that it becomes a matter of routine for them throughout their careers

I am speaking to you as:

An incoming cohort of social scientists

People who will be training future cohorts of social scientists

And (I hope!) future collaborators in developing and promoting transparent and replicable methods of empirical research

The methods and tools I will talk about are low-tech. Much of what I will suggest is largely a matter of common sense, and almost all of the ideas have been advocated by others.

These methods and tools are not intrinsically interesting from the point of view of computational or data science. (They can be used, however, to document sophisticated, computationally complex research.)

The interest and value of what I will present has to do with the fact that, as simple and low-tech as it is, lots of people don't pay attention to it, and so lots of research cannot be replicated. And when people do pay attention to a few basic principles, a lot of good things follow—both for the social enterprise of research and for the individuals who conduct it responsibly.

A little background on how I got involved in this work helps explain all the above.

Let's roll up our sleeves:

<https://osf.io/he87a/>