



BERKELEY INITIATIVE FOR TRANSPARENCY  
IN THE SOCIAL SCIENCES

Transparencia  
en  
Investigación  
en las  
Ciencias  
Sociales  
  
Hoces

Introducción

Ética en la  
Investigación  
Científica

Registros

Sesgo de  
Publicación

Registros

Pre-Analysis  
Plans

P-Hacking

Planes de  
Pre-Análisis

Replicaciones

Protocolo del  
Proyecto, Estándares  
de Información

Flujo de trabajo

Control de Versiones

Compartiendo los  
Datos

# Transparencia en Investigación en las Ciencias Sociales

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

Pontificia Universidad Católica, Octubre 2017  
Slides disponibles en

<https://goo.gl/wU81GA>



BERKELEY INITIATIVE FOR TRANSPARENCY  
IN THE SOCIAL SCIENCES

Transparencia  
en  
Investigación  
en las  
Ciencias  
Sociales

Hoces

Introducción

Ética en la  
Investigación  
Científica

Registros

Sesgo de  
Publicación

Registros

Pre-Analysis  
Plans

P-Hacking

Planes de  
Pre-Análisis

Replicaciones

Protocolo del  
Proyecto, Estándares  
de Información

Flujo de trabajo

Control de Versiones  
Compartiendo los  
Datos

Primero, un favor: ¿Podrían completar esta breve  
encuesta?

[goo.gl/dxfJRb](http://goo.gl/dxfJRb)



BERKELEY INITIATIVE FOR TRANSPARENCY  
IN THE SOCIAL SCIENCES

# Estructura de la Presentación

Transparencia  
en  
Investigación  
en las  
Ciencias  
Sociales

Hojas

Introducción

Ética en la  
Investigación  
Científica

Registros

Sesgo de  
Publicación

Registros

Pre-Analysis  
Plans

P-Hacking

Planes de  
Pre-Análisis

Replicaciones

Protocolo del  
Proyecto, Estándares  
de Información

Flujo de trabajo

Control de Versiones

Compartiendo los  
Datos

## 1 Introducción

## 2 Ética en la Investigación Científica

## 3 Registros

## 4 Pre-Analysis Plans

## 5 Replicaciones

## 6 Conclusión



BERKELEY INITIATIVE FOR TRANSPARENCY  
IN THE SOCIAL SCIENCES



ICPSR

# Ética en la Investigación Científica

Transparencia  
en  
Investigación  
en las  
Ciencias  
Sociales  
  
Hojas

Introducción

Ética en la  
Investigación  
Científica

Registros

Sesgo de  
Publicación

Registros

Pre-Analysis  
Plans

P-Hacking

Planes de  
Pre-Análisis

Replicaciones

Protocolo del  
Proyecto, Estándares  
de Información

Flujo de trabajo

Control de Versiones

Compartiendo los  
Datos

- Transparencia es un elemento central de la ética del investigador.
- Valores científicos acuñados por Robert Merton (Merton 1942):
  - **Universalismo:** cualquier persona puede presentar un argumento, independiente de su estatus.
  - **Comunismo:** el conocimiento es compartido de manera abierta.
  - **Desinterés:** la verdad como motivación, y no los beneficios monetarios.
  - **Escepticismo Organizado:** revisión a través de pares (peer review), replicación.

# Ética en la Investigación Científica

Transparencia  
en  
Investigación  
en las  
Ciencias  
Sociales

Hojas

Introducción

Ética en la  
Investigación  
Científica

Registros

Sesgo de  
Publicación  
Registros

Pre-Analysis  
Plans

P-Hacking

Planes de  
Pre-Análisis

Replicaciones

Protocolo del  
Proyecto, Estándares  
de Información

Flujo de trabajo

Control de Versiones

Compartiendo los  
Datos

- Casos de fraude existen (Simonsohn 2013), pero más importante como investigadores tenemos que admitir nuestra condición humana, sujetos a sesgos y razonamiento motivado, transparencia puede ayudar con esto (Nosek, Spies, Motyl 2012).
- Quienes llevamos a cabo experimentos o usamos datos con información identificable a nivel individual, tenemos que tomar con seriedad los Comités de Ética Institucionales (IRBs) (Ch. 11–13 Morton & Williams 2010, Desposato 2014).

# Ética en la Investigación Científica

Transparencia  
en  
Investigación  
en las  
Ciencias  
Sociales

Hoces

Introducción

Ética en la  
Investigación  
Científica

Registros

Sesgo de  
Publicación

Registros

Pre-Analysis  
Plans

P-Hacking

Planes de  
Pre-Análisis

Replicaciones

Protocolo del  
Proyecto, Estándares  
de Información

Flujo de trabajo

Control de Versiones  
Compartiendo los  
Datos

Por que nos preocupamos:

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

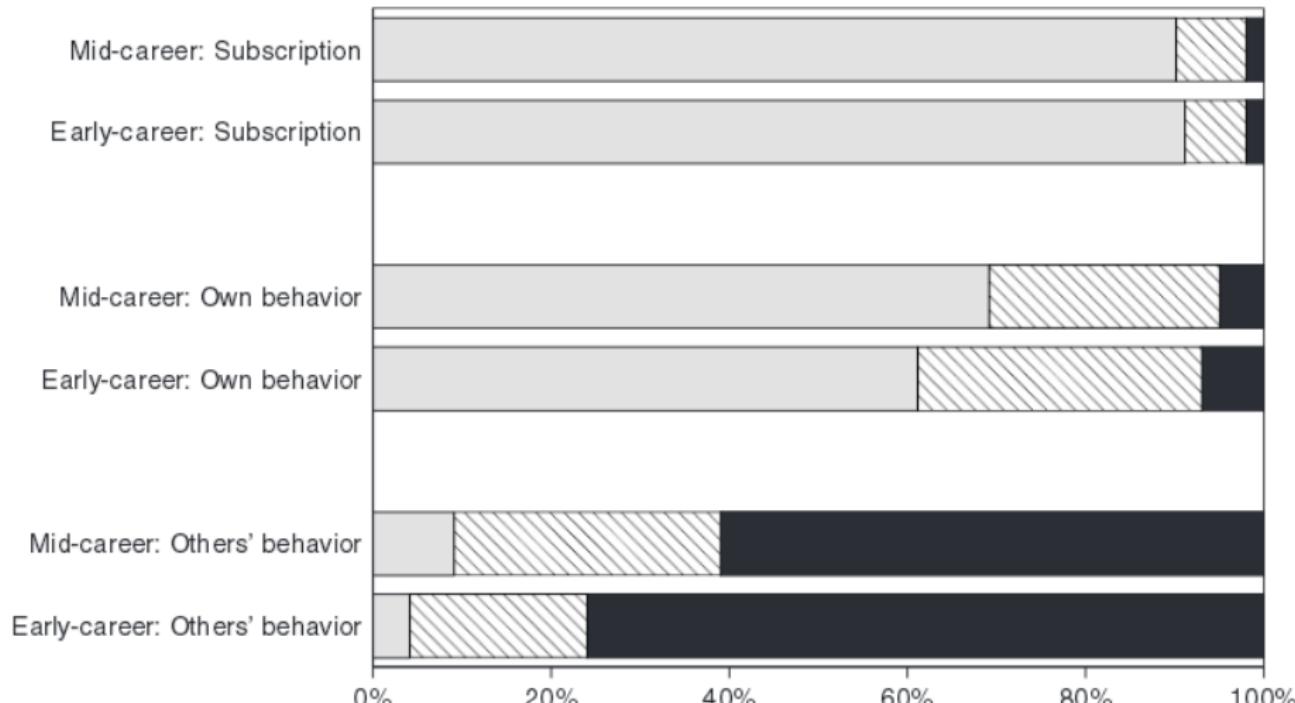
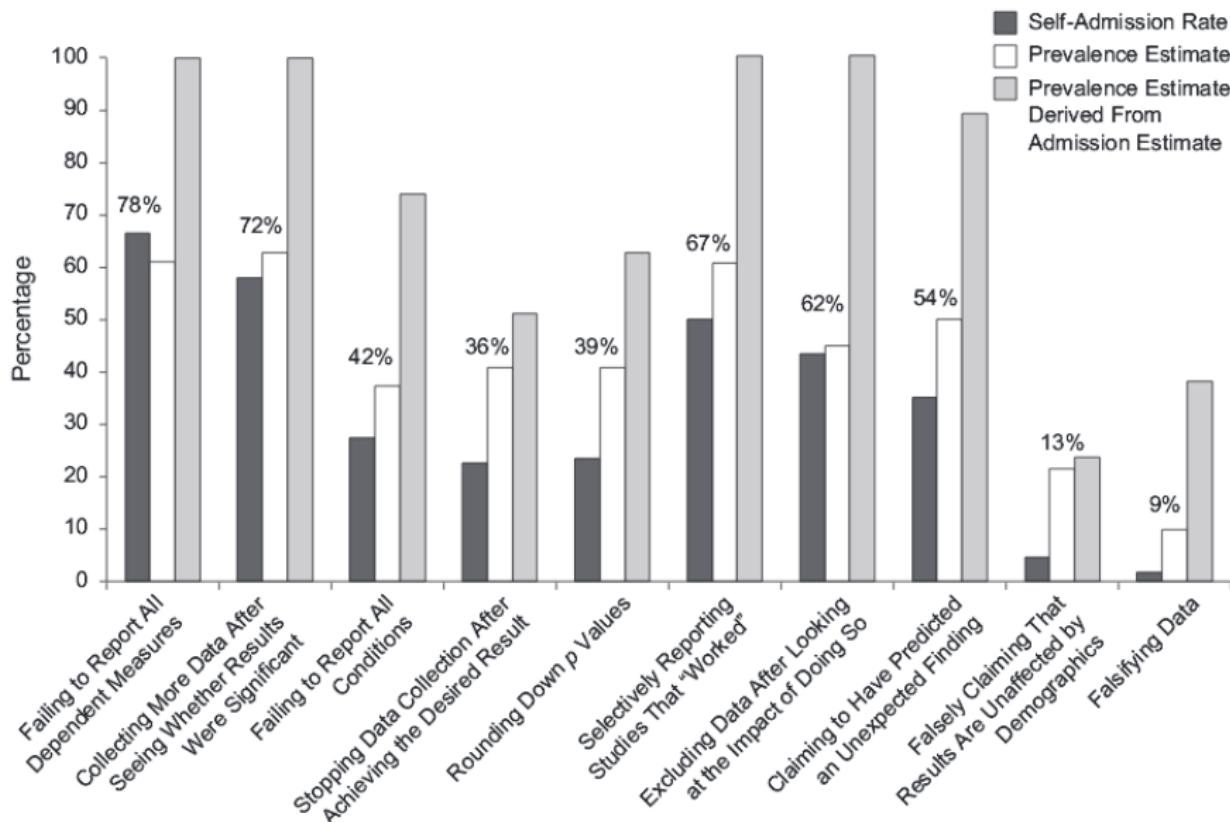


FIG. 3. Norm versus Counternorm Scores: Percent with Norm > Counternorm (dotted), Norm = Counternorm (striped), Norm < Counternorm (solid).



**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.



BERKELEY INITIATIVE FOR TRANSPARENCY  
IN THE SOCIAL SCIENCES

# Sesgo de Publicación

Transparencia en Investigación en las Ciencias Sociales Hoces

Introducción  
Ética en la Investigación Científica

Registros  
Sesgo de Publicación  
Registros

Pre-Analysis Plans  
P-Hacking  
Planes de Pre-Análisis

Replicaciones  
Protocolo del Proyecto, Estándares de Información  
Flujo de trabajo  
Control de Versiones  
Compartiendo los Datos

## Existencia del problema:

- El tamaño de los efectos disminuye con el tamaño muestral (Gerber, Green, Nickerson 2001)
- Las Ciencias Sociales muestran una tasa de rechazo de la hipótesis nula mayor que las ciencias duras (Fanelli 2010).
- La publicación de efectos nulos esta desapareciendo en el tiempo, en todas las disciplinas. (Fanelli 2011).
- Estudio que siguió a experimentos completados muestra que aquellos experimentos con fuertes resultados son 40pp más probable de ser publicados, y 60pp más probable de ser escritos. Alto “file drawer problem”. (Franco, Malhotra, Simonovits 2014)



BERKELEY INITIATIVE FOR TRANSPARENCY  
IN THE SOCIAL SCIENCES

# Problema en todas las disciplinas

Transparencia  
en  
Investigación  
en las  
Ciencias  
Sociales

Hojas

Introducción

Ética en la  
Investigación  
Científica

Registros

Sesgo de  
Publicación

Registros

Pre-Analysis  
Plans

P-Hacking

Planes de  
Pre-Análisis

Replicaciones

Protocolo del  
Proyecto, Estándares  
de Información

Flujo de trabajo

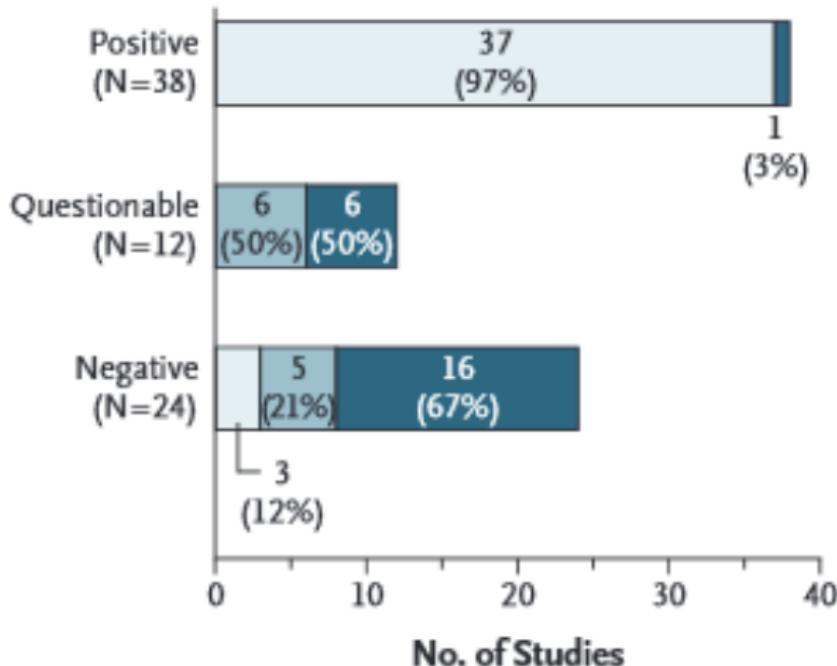
Control de Versiones  
Compartiendo los  
Datos

- Medicina: (Turner et al. 2008)
- Ciencias Sociales: (Franco, Malhotra, Simonovits 2014)
- Economía: (Brodeur et al. 2016)
- Sociología: (Gerber and Malhotra 2008)
- Ciencias Políticas: (Gerber and Malhotra 2008)

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

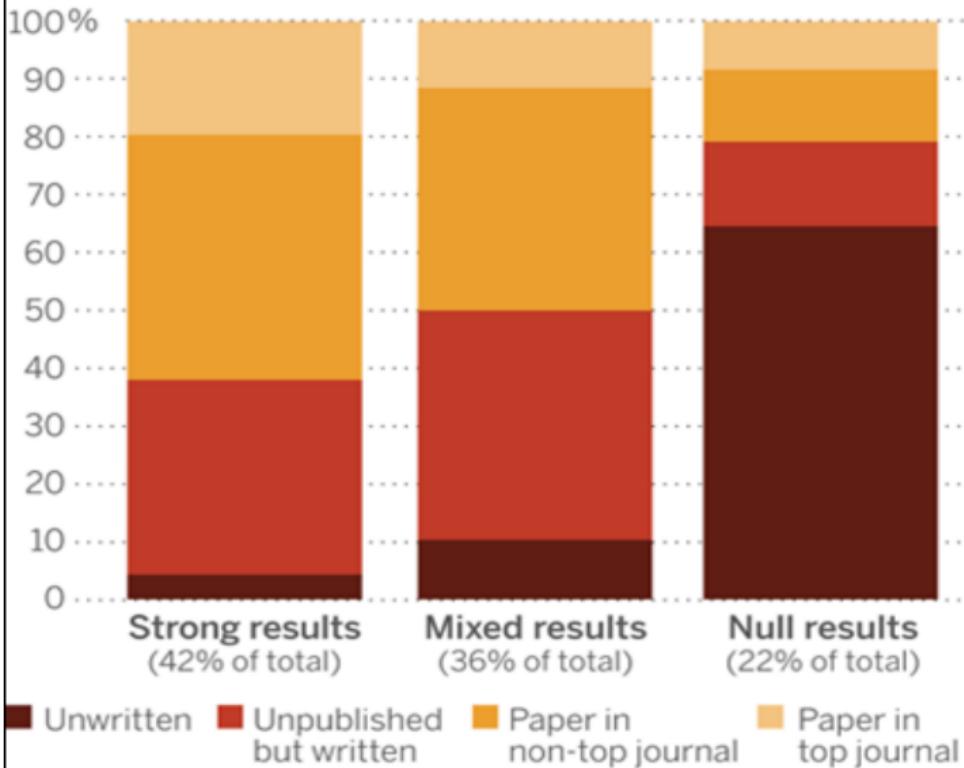
## A Studies (N=74)

### FDA Decision



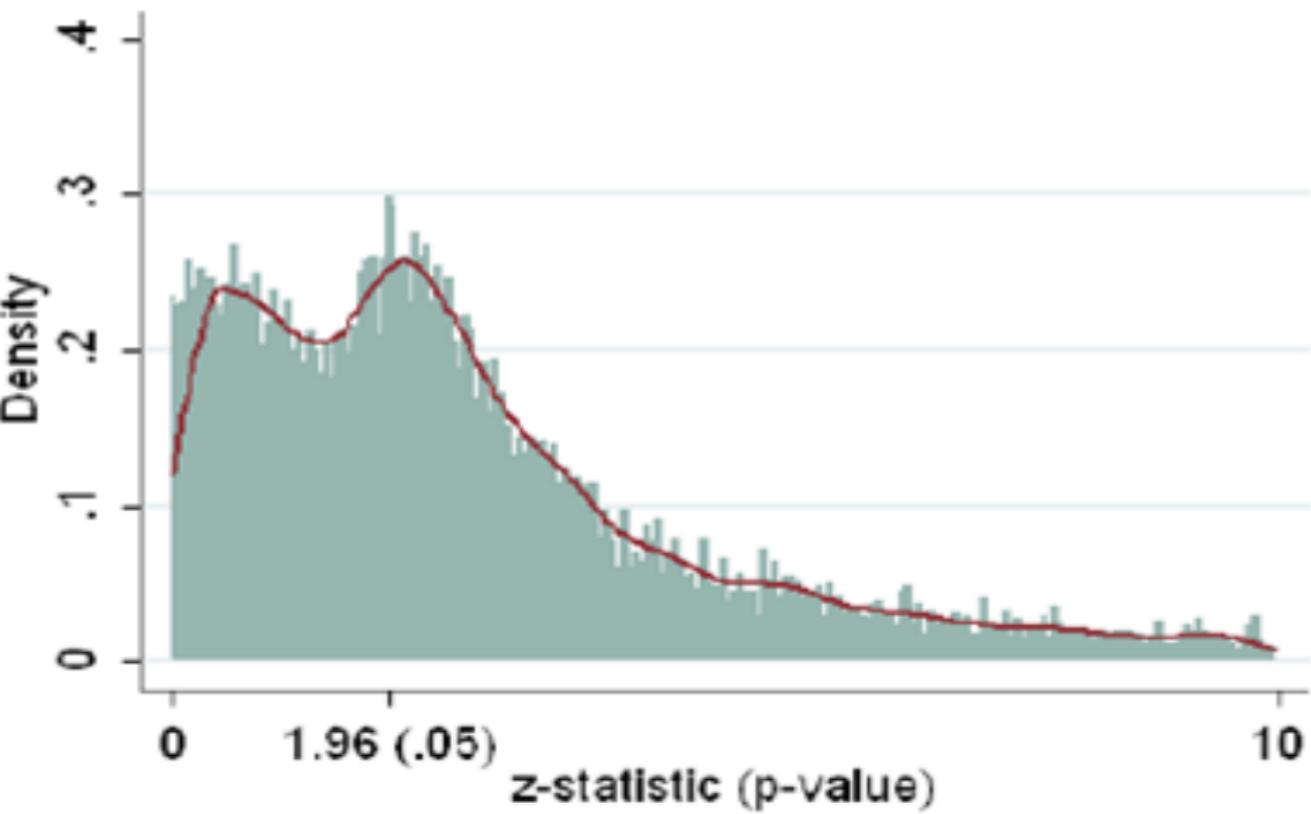
# Most null results are never written up

The fate of 221 social science experiments



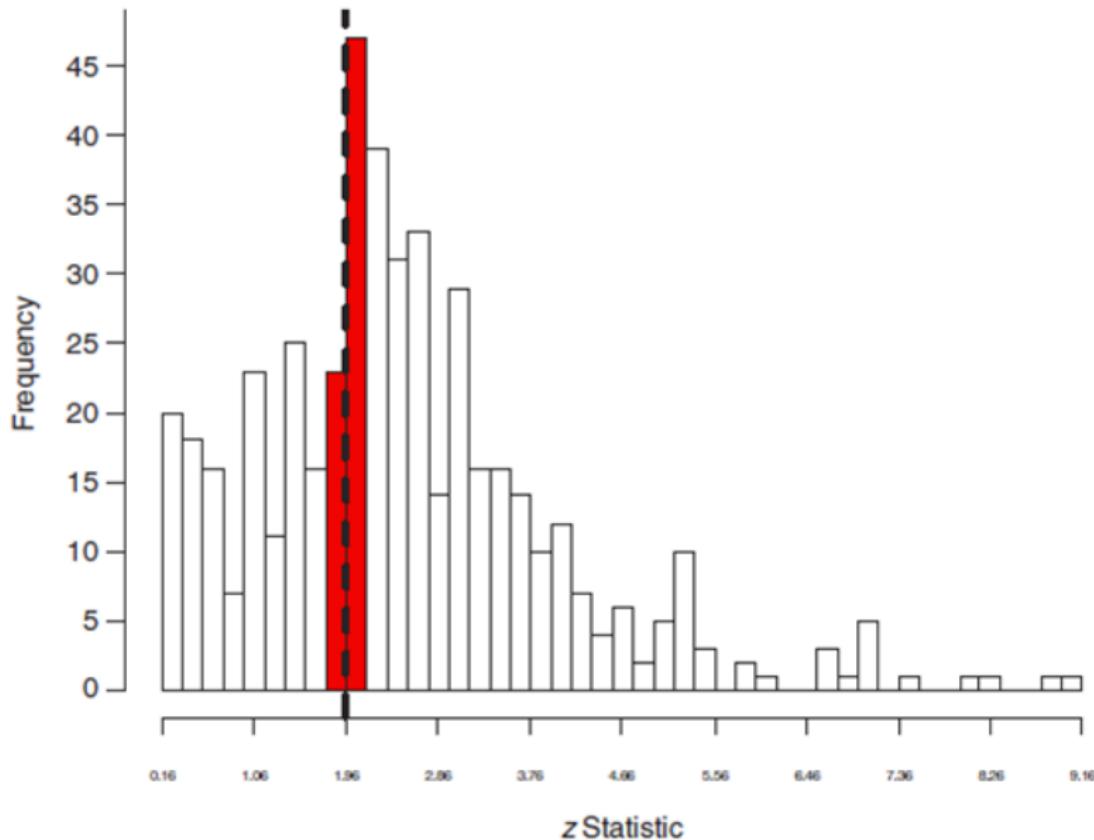
Source: A. Franco et al., *Science* (28 August)

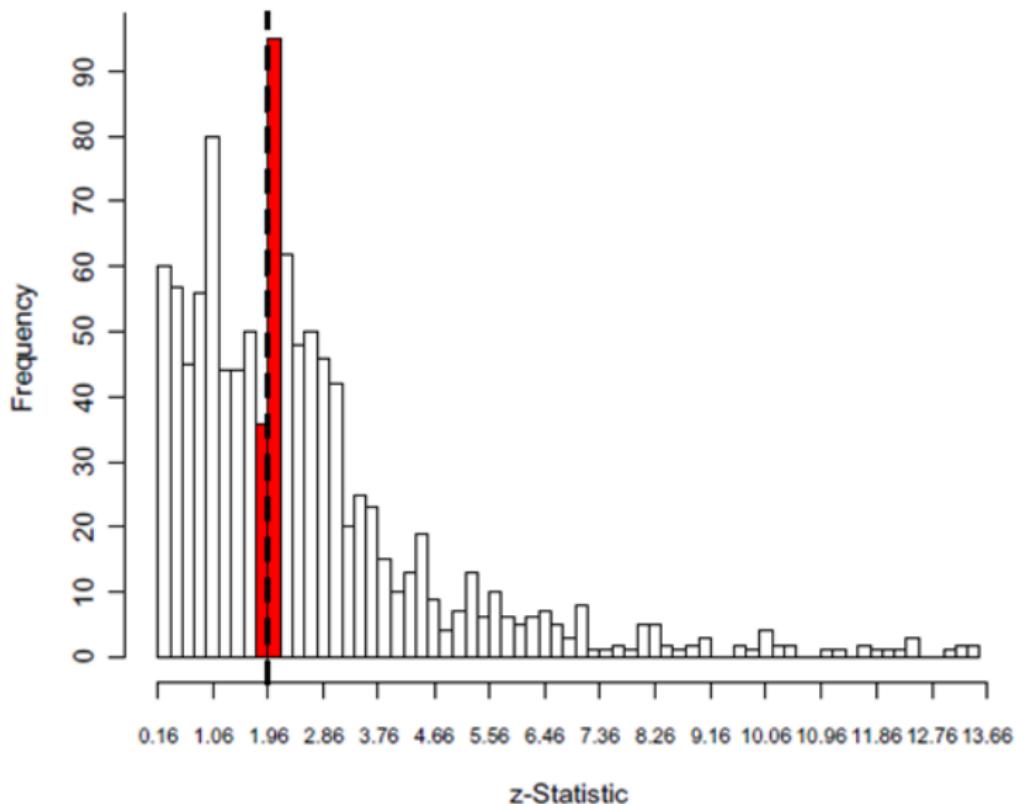
(b) Unrounded distribution of z-statistics.



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.

# Sesgo de Publicación

Transparencia  
en  
Investigación  
en las  
Ciencias  
Sociales

Hojas

Introducción

Ética en la  
Investigación  
Científica

Registros  
Sesgo de  
Publicación  
Registros

Pre-Analysis  
Plans

P-Hacking  
Planes de  
Pre-Análisis

Replicaciones  
Protocolo del  
Proyecto, Estándares  
de Información  
Flujo de trabajo  
Control de Versiones  
Compartiendo los  
Datos

Si solo escribimos/publicamos resultados significativos, y no dejamos registro de los no significativos, no tenemos forma de distinguir si nuestros resultados “significativos” son reales, o si son el 5% que deberíamos esperar debido a error estadístico.



BERKELEY INITIATIVE FOR TRANSPARENCY  
IN THE SOCIAL SCIENCES

# Registros

Transparencia  
en  
Investigación  
en las  
Ciencias  
Sociales

Hojas

Introducción

Ética en la  
Investigación  
Científica

Registros

Sesgo de  
Publicación

Registros

Pre-Analysis  
Plans

P-Hacking

Planes de  
Pre-Análisis

Replicaciones

Protocolo del  
Proyecto, Estándares  
de Información

Flujo de trabajo

Control de Versiones

Compartiendo los  
Datos

Pre-Registros como una solución al sesgo de publicación:

- Hacer pública la investigación a ejecutar, publicando por adelantado las hipótesis a testear.
- Adopción casi universal en RCTs en medicina.  
Journals top (ICMJE) no publican estudios si no están registrados. <http://clinicaltrials.gov>

# Registros

Transparencia en Investigación en las Ciencias Sociales

Hojas

Introducción

Ética en la Investigación Científica

Registros

Sesgo de Publicación

Registros

Pre-Analysis Plans

P-Hacking

Planes de Pre-Análisis

Replicaciones

Protocolo del Proyecto, Estándares de Información

Flujo de trabajo

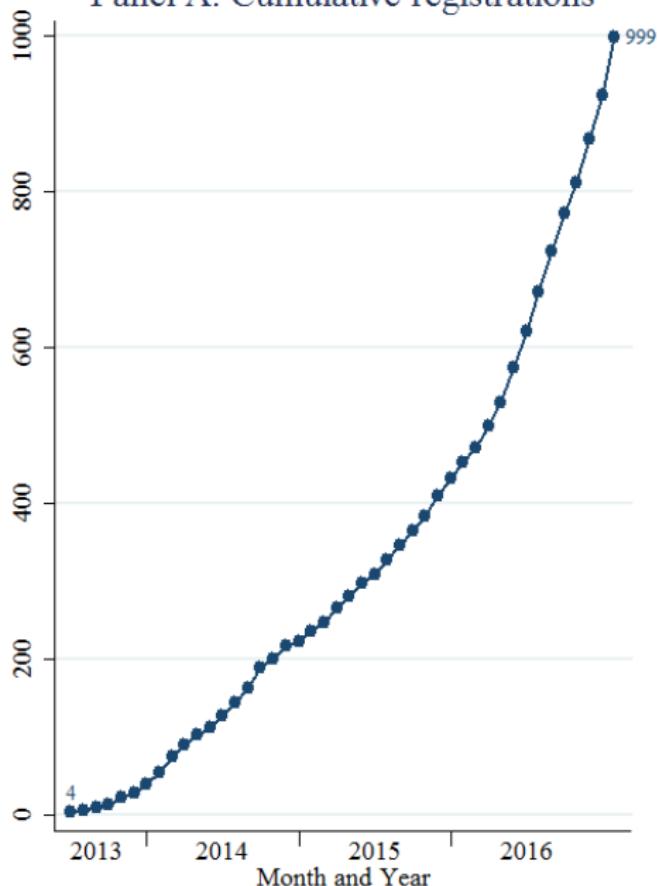
Control de Versiones

Compartiendo los Datos

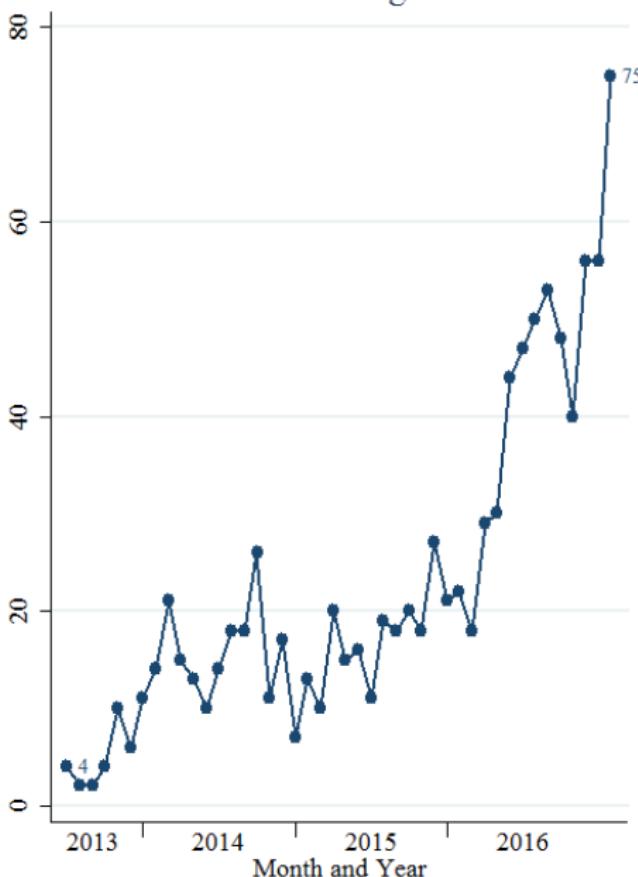
Nuevos en ciencias sociales, pero:

- Registro de AEA , actualmente solo para RCTs.  
<http://socialscienceregistry.org>
- Registro de EGAP  
<http://egap.org/design-registration>
- Registro de 3ie, para evaluaciones en países en desarrollo. <http://ridie.3ieimpact.org>
- Open Science Framework
  - Formato abierto
  - Pronto va a estar sincronizado con los de más arriba
- Simple: <http://aspredicted.org>

Panel A: Cumulative registrations



Panel B: New registrations



# Publicaciones Basadas en Diseño de la Investigación

Transparencia en Investigación en las Ciencias Sociales Hoces

Introducción Ética en la Investigación Científica

Registros Sesgo de Publicación Registros

Pre-Analysis Plans

P-Hacking

Planes de Pre-Análisis

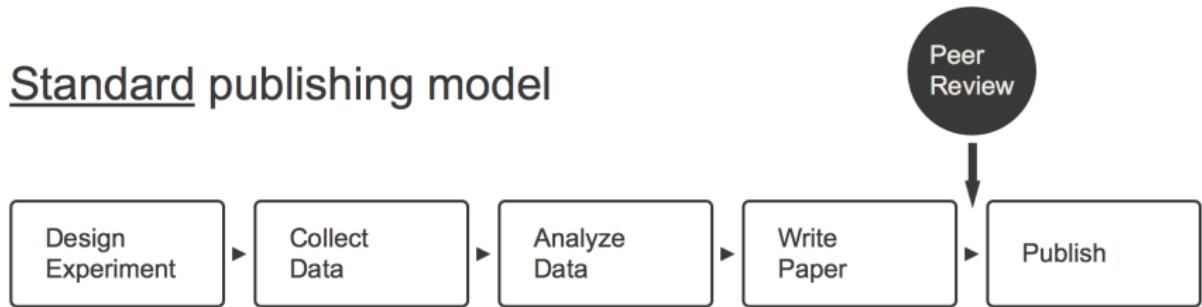
Replicaciones Protocolo del Proyecto, Estándares de Información Flujo de trabajo Control de Versiones Compartiendo los Datos

Alias Reportes Registrados, cambia momento de peer review hacia antes del la recolección de datos, análisis y resultados.

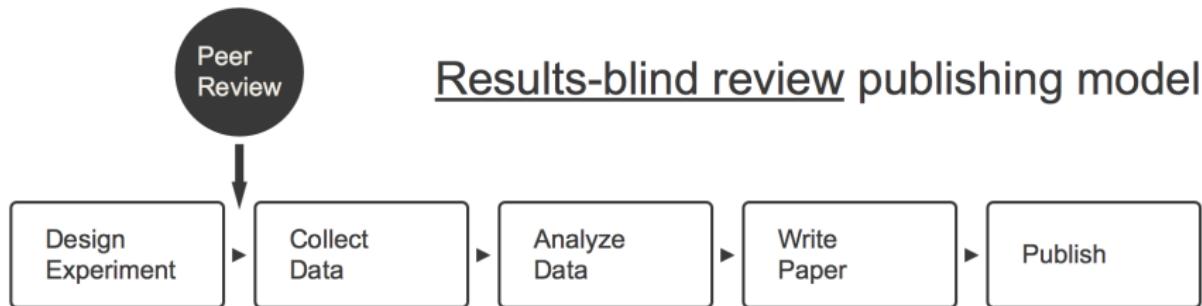
- 1 Diseñar un estudio
- 2 Enviar a un journal
- 3 Revisión basada en la importancia de la pregunta y calidad del diseño
- 4 Obtener aceptación en principio
- 5 Ejecutar el estudio, y publicar incluso con resultados nulos

20 Journals, 5 más con ediciones especiales [▶ Link](#)

## Standard publishing model



## Results-blind review publishing model



# Meta-Análisis

Transparencia  
en  
Investigación  
en las  
Ciencias  
Sociales  
  
Hoces

Introducción  
  
Ética en la  
Investigación  
Científica  
  
Registros  
  
Sesgo de  
Publicación  
Registros  
  
Pre-Analysis  
Plans  
  
P-Hacking  
Planes de  
Pre-Análisis  
  
Replicaciones  
  
Protocolo del  
Proyecto, Estándares  
de Información  
  
Flujo de trabajo  
  
Control de Versiones  
Compartiendo los  
Datos

## Síntesis sistemático de los resultados

### Organizaciones:

- Cochrane Collaboration (Medicina)
- Campbell Collaboration (Políticas Públicas)
- What Works Clearinghouse (Educación, Gob. US)
- CLEAR (Empleo, Gob. US)
- MAER-NET (Economía)



BERKELEY INITIATIVE FOR TRANSPARENCY  
IN THE SOCIAL SCIENCES

# Meta-Análisis

Transparencia  
en  
Investigación  
en las  
Ciencias  
Sociales

Hojas

Introducción

Ética en la  
Investigación  
Científica

Registros

Sesgo de  
Publicación

Registros

Pre-Analysis  
Plans

P-Hacking

Planes de  
Pre-Análisis

Replicaciones

Protocolo del  
Proyecto, Estándares  
de Información

Flujo de trabajo

Control de Versiones

Compartiendo los  
Datos

## Herramientas:

- Funnel plots del tamaño de la muestra vs. tamaño del efecto (Card & Krueger 1995)
- Funnel Asymmetry Test (Stanley & Doucouliagos 2012)
- P-curve (Simonsohn et al. 2014)
  - ▶ Online App
  - Un p-checker para todos
    - ▶ Shiny App

# P-Hacking

Transparencia  
en  
Investigación  
en las  
Ciencias  
Sociales

Hojas

Introducción

Ética en la  
Investigación  
Científica

Registros

Sesgo de  
Publicación  
Registros

Pre-Analysis  
Plans

P-Hacking

Planes de  
Pre-Análisis

Replicaciones

Protocolo del  
Proyecto, Estándares  
de Información

Flujo de trabajo

Control de Versiones

Compartiendo los  
Datos

## Definición del problema:

- Otros nombres: “fishing”, grados de libertad del investigador, o “data-mining”.
- Definición: flexibilidad en el análisis de datos permite presentar *casi cualquier resultado* bajo un umbral arbitrario; significancia estadística pierde sentido.
- No es algo único de investigadores con malas intenciones. Es subconsciente, o simplemente una práctica estándar del análisis estadístico (Gelman, Loken 2013).



BERKELEY INITIATIVE FOR TRANSPARENCY  
IN THE SOCIAL SCIENCES

# P-Hacking: hágalo usted mismo!

Transparencia  
en  
Investigación  
en las  
Ciencias  
Sociales

Hoces

Introducción

Ética en la  
Investigación  
Científica

Registros

Sesgo de  
Publicación

Registros

Pre-Analysis  
Plans

P-Hacking

Planes de  
Pre-Análisis

Replicaciones

Protocolo del  
Proyecto, Estándares  
de Información

Flujo de trabajo

Control de Versiones

Compartiendo los  
Datos

- “Science isn’t Broken” —538 reportaje periodístico con modulo interactivo [Link](#)
- Practique p-hacking en R/Shiny App. [Link](#)
- El 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).

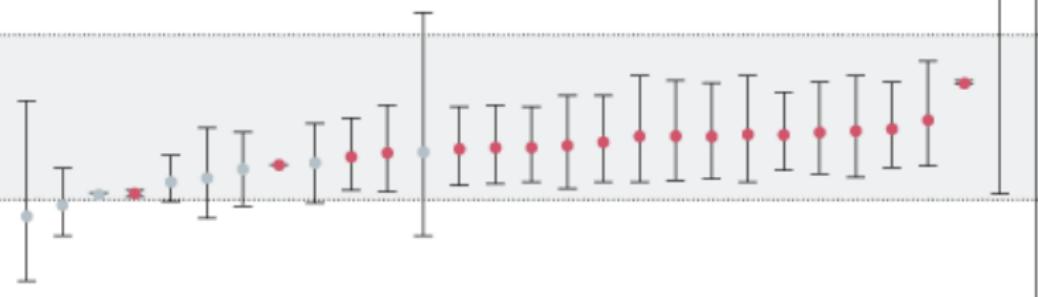
78.7\*  
115\*

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.

# Planes de Pre-Análisis

Transparencia en Investigación en las Ciencias Sociales

Hojas

Introducción

Ética en la Investigación Científica

Registros

Sesgo de Publicación

Registros

Pre-Analysis

Plans

P-Hacking

Planes de Pre-Análisis

Replicaciones

Protocolo del Proyecto, Estándares de Información

Flujo de trabajo

Control de Versiones

Compartiendo los Datos

## Explicación de la solución, de 3ie:

“Un plan de pre-análisis (PAPs) es una descripción detallada de los análisis a ser conducidos. Esta descripción es escrita antes de los datos que contienen los impactos del programa bajo evaluación. Puede especificar las hipótesis a testear, como se construirán la variables, ecuaciones a estimar, controles a utilizar, y otros aspectos del análisis. Una función clave de los PAPs es incrementar la transparencia en investigación. Al especificar los detalles por adelantado antes de ver los resultados, el plan es un resguardo contra data mining y búsqueda de especificaciones. Se les recomienda a los investigadores que desarrollen y suban dichos planes junto con el registro de sus estudios. Sin embargo esto no es requisito para el registro de un estudio.” study registration, but it is not required for registration.”

# Orígenes: Regulación Farmacéutica en USA

Transparencia  
en  
Investigación  
en las  
Ciencias  
Sociales

Hojas

Introducción

Ética en la  
Investigación  
Científica

Registros

Sesgo de  
Publicación

Registros

Pre-Analysis  
Plans

P-Hacking

Planes de  
Pre-Análisis

Replicaciones

Protocolo del  
Proyecto, Estándares  
de Información

Flujo de trabajo

Control de Versiones

Compartiendo los  
Datos

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

## §V Consideraciones en el Análisis de Datos

- 1 Pre-especificación del análisis
- 2 Grupos de análisis
- 3 Tratamiento de NAs y valores extremos
- 4 Transformaciones a los datos
- 5 Estimación, intervalos de confianza, y testeo de hipótesis
- 6 Ajuste de significancia y niveles de confianza
- 7 Subgrupos, interacciones, y variables de control
- 8 Integridad de los datos y validez del software computacional

# Sugerencias de Glennerster, Takavarasha

## *Running Randomized Evaluations*

- 1 como se van a medir los principales variables de resultado,
- 2 Dentro de las variables de resultado: cuales son primarias, cuales secundarias?,
- 3 composición exacta de toda la familia de tests utilizados en análisis de efectos promedios
  - Explicación de efectos promedio , FWER, FDR en Anderson (JASA 2008).
- 4 los subgrupos a ser analizados,
- 5 la dirección esperada de los impactos si queremos usar test de una cola, y
- 6 la especificación primaria a ser utilizada en el análisis.



BERKELEY INITIATIVE FOR TRANSPARENCY  
IN THE SOCIAL SCIENCES

Transparencia en  
Investigación en las  
Ciencias Sociales

Hojas

Introducción

Ética en la  
Investigación  
Científica

Registros

Sesgo de  
Publicación

Registros

Pre-Analysis  
Plans

P-Hacking

Planes de  
Pre-Análisis

Replicaciones

Protocolo del  
Proyecto, Estándares  
de Información

Flujo de trabajo

Control de Versiones

Compartiendo los  
Datos

# Sugerencias de McKenzie

## World Bank Development Impact Blog

- 1 Descripción de la muestra a ser utilizada en el estudio
- 2 Fuentes de dato fundamentales
- 3 Hipótesis a testear a través de la cadena causal
- 4 Especificar como se van a construir las variables
- 5 Especificar la ecuación donde se estimará el efecto del tratamiento
- 6 ¿Cuál es la estrategia para analizar múltiples variables de resultados y múltiples test de hipótesis?
- 7 Procedimiento a utilizar para enfrentar atrición en la muestra
- 8 ¿Qué hará el estudio en caso de limitada variación en las variables de resultados?
- 9 Si hay un modelo a testear, este debe ser incluido
- 10 No olvidar archivarlo (con una fecha estampada digitalmente)



BERKELEY INITIATIVE FOR TRANSPARENCY  
IN THE SOCIAL SCIENCES

# Ejemplos

Transparencia en Investigación en las Ciencias Sociales Hoces

Introducción Ética en la Investigación Científica

Registros Sesgo de Publicación Registros

Pre-Analysis Plans P-Hacking Planes de Pre-Análisis

Replicaciones Protocolo del Proyecto, Estándares de Información Flujo de trabajo Control de Versiones Compartiendo los Datos

- J-PAL Registro de Hipótesis , ver <http://www.povertyactionlab.org/Hypothesis-Registry>  
6 ejemplos de papers publicados:
  - Sierra Leone CDD, Oregon Medicare, Turkey Job Training, El Salvador TOMS, dos en Indonesia (Olken et al.)
  - Psicología: Hawkins, Fitzgerald, Nosek—Riesgos de Concepción y Prejuicios

Amplio rango de que es lo que se debe escribir exactamente y cuánto detalle incorporar en el PAP. En un nivel extremo el código estaría listo para ejecutar en cuanto lleguen los datos. .

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)



BERKELEY INITIATIVE FOR TRANSPARENCY  
IN THE SOCIAL SCIENCES

# PAP-Estudios Observacionales

Transparencia  
en  
Investigación  
en las  
Ciencias  
Sociales

Hojas

Introducción

Ética en la  
Investigación  
Científica

Registros

Sesgo de  
Publicación

Registros

Pre-Analysis  
Plans

P-Hacking

Planes de  
Pre-Análisis

Replicaciones

Protocolo del  
Proyecto, Estándares  
de Información

Flujo de trabajo

Control de Versiones

Compartiendo los  
Datos

- Actual debate en salud publica/epidemiología.
- Pre-especificar de manera verificable: difícil, pero no imposible.
- Ejemplo: Datos generados periódicamente por el gobierno.
- Ejemplo: Salario Mínimo (Neumark 2001)

# 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.



BERKELEY INITIATIVE FOR TRANSPARENCY  
IN THE SOCIAL SCIENCES

# Replicaciones

Transparencia  
en  
Investigación  
en las  
Ciencias  
Sociales

Hoces

Introducción

Ética en la  
Investigación  
Científica

Registros

Sesgo de  
Publicación  
Registros

Pre-Analysis  
Plans

P-Hacking

Planes de  
Pre-Análisis

Replicaciones

Protocolo del  
Proyecto, Estándares  
de Información  
Flujo de trabajo  
Control de Versiones  
Compartiendo los  
Datos

- 1 El problema (JMBC Project)**
- 2 Protocolo del proyecto, Estándares de publicación**
- 3 Organización del flujo de trabajo**
- 4 Compartir código y datos**

## **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.*

# Protocolo del Proyecto, Estándares de Información

Transparencia en Investigación en las Ciencias Sociales Hoces

Introducción Ética en la Investigación Científica

Registros Sesgo de Publicación Registros

Pre-Analysis Plans P-Hacking Planes de Pre-Análisis

Replicaciones Protocolo del Proyecto, Estándares de Información Flujo de trabajo Control de Versiones Compartiendo los Datos

Asegurar de que todo disponibilidad lo necesario para que otro investigador replique su investigación.

- Encuentre el estándar de información apropiado para su disciplina y sígalo:  
<http://www.equator-network.org>
- Informe todos los detalles sobre la implementación de un proyecto en un protocolo detallado:  
<http://www.spirit-statement.org>
- Pautas Transparency and Openness Promotion (TOP):  
<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)

# Flujo de trabajo

Transparencia  
en  
Investigación  
en las  
Ciencias  
Sociales

Hoces

Introducción

Ética en la  
Investigación  
Científica

Registros

Sesgo de  
Publicación

Registros

Pre-Analysis  
Plans

P-Hacking

Planes de  
Pre-Análisis

Replicaciones

Protocolo del  
Proyecto, Estándares  
de Información

**Flujo de trabajo**

Control de Versiones  
Compartiendo los  
Datos

“Reproducibilidad es simplemente una colaboración con personas que aun no conoces, incluido tu mismo en una semana más’ —Philip Stark, UC Berkeley Statistics



# Flujo de trabajo

Transparencia en Investigación en las Ciencias Sociales

Hojas

Introducción

Ética en la Investigación Científica

Registros

Sesgo de Publicación

Registros

Pre-Analysis Plans

P-Hacking

Planes de Pre-Análisis

Replicaciones

Protocolo del Proyecto, Estándares de Información

Flujo de trabajo

Control de Versiones  
Compartiendo los Datos

- Sugerencias prácticas sobre organización y programación
  - Al realizar cambios a archivos que han sido compartidos/publicados, implica que debe renombrar el archivo.
  - Haga referencia a la versión necesaria para ejecutar funciones.
  - Long (2008) *The Workflow of Data Analysis Using Stata*
- Programación Literaria: comentario extensivo en el código, apuntando a que sea legible por humanos
- Control de Versiones
- Documentos Dinámicos

# Control de Versiones

Transparencia en Investigación en las Ciencias Sociales Hoces

Introducción Ética en la Investigación Científica

Registros Sesgo de Publicación Registros

Pre-Analysis Plans P-Hacking Planes de Pre-Análisis

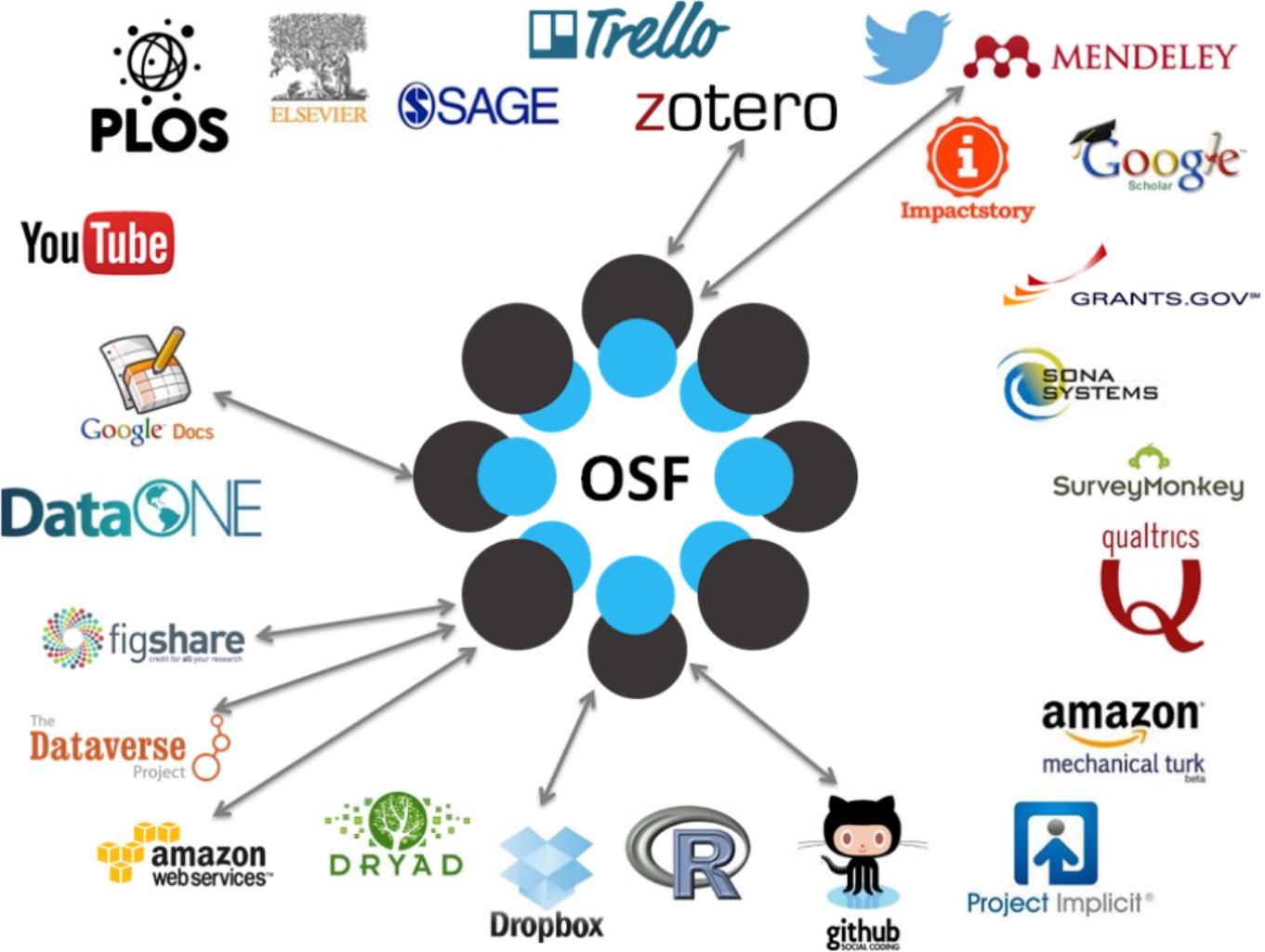
Replicaciones Protocolo del Proyecto, Estándares de Información Flujo de trabajo Control de Versiones Compartiendo los Datos

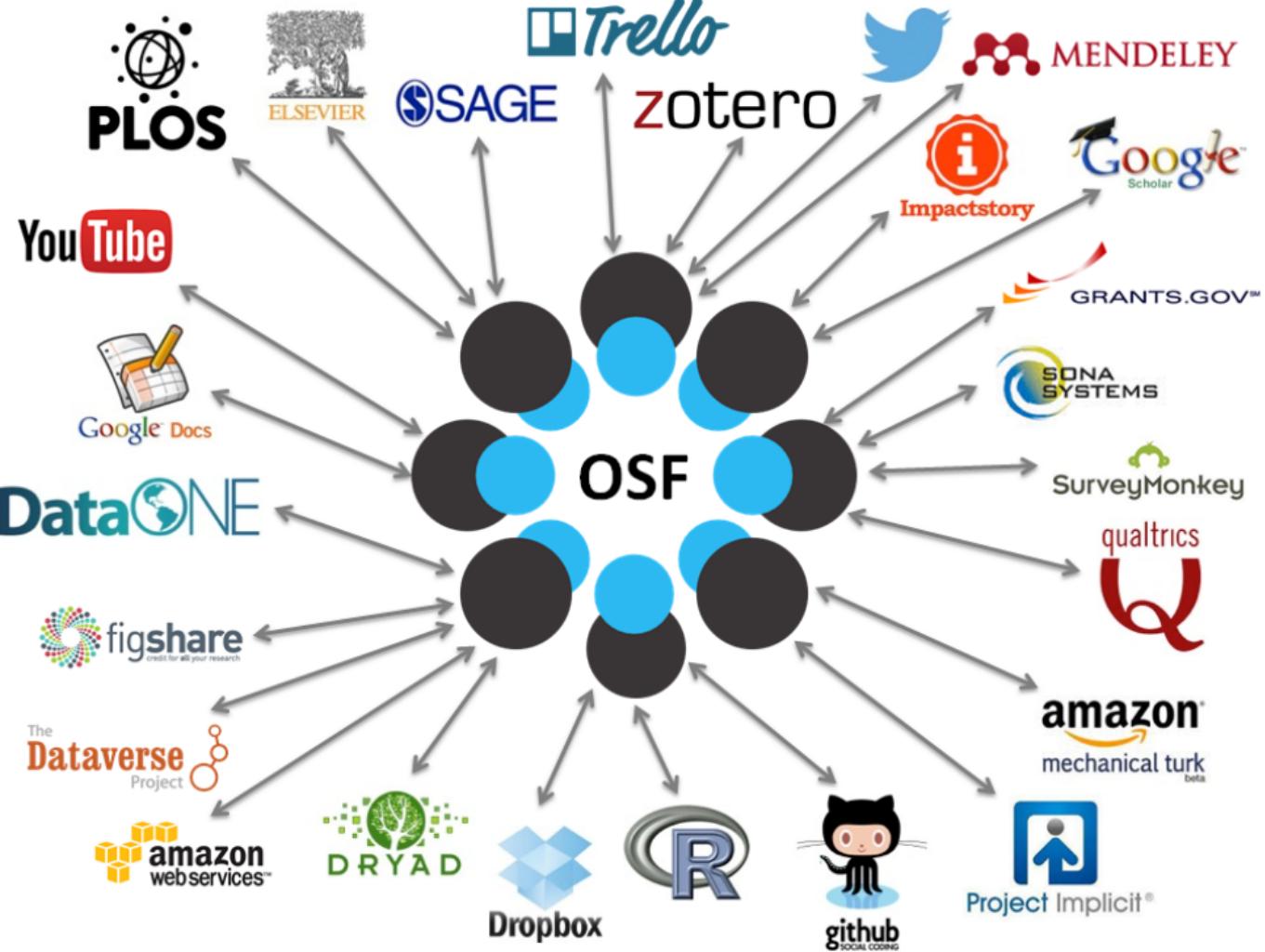
- Utilizar control de versiones puede ayudar a hacer su trabajo más reproducible.
- Qué es Control de Versiones?

*Control de versiones es un sistema que registra cambios a uno o varios archivos en el tiempo, de forma tal que versiones específicas pueden ser recuperados más adelante.*

## –Git, About Version Control









BERKELEY INITIATIVE FOR TRANSPARENCY  
IN THE SOCIAL SCIENCES

# Documentos Dinámicos

Transparencia en  
Investigación en las  
Ciencias Sociales  
Hoces

Introducción

Ética en la  
Investigación  
Científica

Registros

Sesgo de  
Publicación

Registros

Pre-Analysis  
Plans

P-Hacking

Planes de  
Pre-Análisis

Replicaciones

Protocolo del  
Proyecto, Estándares  
de Información

Flujo de trabajo

Control de Versiones

Compartiendo los  
Datos

**Meta aspiracional:** Escriba su código y su paper en el mismo archivo. De esta forma se minimiza la perdida de información y se eliminan errores de copiar/pegar.

Establecido en R/Python y comenzando en Stata.

- Incluye tablas dentro del archivo, en lugar de copiar-pegar-formatear elementos estáticos
- Todas las cifras dentro del texto también son calculadas de manera dinámica, eliminando la necesidad de tipeo de cifras manualmente.
- Cifras y tablas se actualizan automáticamente.
- El paper completo es (re)producido con uno o dos clicks

The logo for Jupyter consists of the word "jupyter" in a bold, black, sans-serif font, centered within a white circle. The circle is partially overlaid by two thick, orange, curved bands that meet at the top and bottom, resembling a smile. Small dark gray circles are positioned at the intersections of these curves with the circle's edge.

jupyter





BERKELEY INITIATIVE FOR TRANSPARENCY  
IN THE SOCIAL SCIENCES

# Compartiendo lo Datos

Transparencia  
en  
Investigación  
en las  
Ciencias  
Sociales  
  
Hojas

Introducción  
  
Ética en la  
Investigación  
Científica

Registros  
  
Sesgo de  
Publicación  
Registros

Pre-Analysis  
Plans  
  
P-Hacking  
Planes de  
Pre-Análisis

Replicaciones  
  
Protocolo del  
Proyecto, Estándares  
de Información  
  
Flujo de trabajo  
  
Control de Versiones  
  
Compartiendo lo  
Datos

Publique su código y sus datos en un repositorio público de confianza.

- Encuentre el repositorio apropiado:  
<http://www.re3data.org/>
- Los repositorios están diseñados para durar más que su website.
- También están diseñados para facilitar la búsqueda por parte de otros investigadores.
- Los repositorios guardan sus datos en formatos públicos que evitan obsolescencia en el tiempo.

# Conclusión

Transparencia en Investigación en las Ciencias Sociales Hoces

Introducción Ética en la Investigación Científica

Registros Sesgo de Publicación Registros

Pre-Analysis Plans P-Hacking Planes de Pre-Análisis

Replicaciones Protocolo del Proyecto, Estandares de Información Flujo de trabajo Control de Versiones Compartiendo los Datos

OK, estoy convencido. Como puedo implementar esto en mi propia investigación?

- Tomen nuestro MOOC (E. Miguel). [▶ Link](#)
- Suscribanse al blog & E-mail de BITSS [▶ Link](#)
- Apliquen a nuestro summer institute: RT2. [▶ Link](#)
- Apliquen a nuestros fondos de investigación SSMART (fondos adicionales para investigadores de países en desarrollo). [▶ Link](#)
- Apliquen al premio Leamer-Rosenthal. [▶ Link](#)
- Revisen nuestra página de recursos. [▶ Link](#)
- Interesado en incluir parte de estas prácticas en el análisis de políticas públicas? Hablen conmigo! [fhoces@berkeley.edu](mailto:fhoces@berkeley.edu)

Transparencia  
en  
Investigación  
en las  
Ciencias  
Sociales

Hoces

Introducción

Ética en la  
Investigación  
Científica

Registros

Sesgo de  
Publicación  
Registros

Pre-Analysis  
Plans

P-Hacking  
Planes de  
Pre-Análisis

Replicaciones

Protocolo del  
Proyecto, Estándares  
de Información

Flujo de trabajo

Control de Versiones

Compartiendo los  
Datos

Tres días de entrenamiento. Evento anual que alterna entre Berkeley y fuera de USA (Londres fue hace dos semanas)



Transparencia en Investigación en las Ciencias Sociales Hoces

Introducción

Ética en la Investigación Científica

Registros

Sesgo de Publicación

Registros

Pre-Analysis Plans

P-Hacking

Planes de Pre-Análisis

Replicaciones

Protocolo del Proyecto, Estándares de Información

Flujo de trabajo

Control de Versiones

Compartiendo lo Dato

# Fondos SSMART

Fondos de hasta \$30,000 para financiar investigación en transparencia:

- Desarrollo de nuevas metodologías
- Nuevas herramientas y usos de meta-análisis
- Investigación sobre comportamiento de investigadores y adopción de nuevas normas

Financiamiento extra para investigadores de países en desarrollo.



# Premio Leamer-Rosenthal

Transparencia  
en  
Investigación  
en las  
Ciencias  
Sociales  
  
Hoces

Introducción  
  
Ética en la  
Investigación  
Científica

Registros  
  
Sesgo de  
Publicación  
  
Registros

Pre-Analysis  
Plans  
  
P-Hacking  
  
Planes de  
Pre-Análisis

Replicaciones  
  
Protocolo del  
Proyecto, Estándares  
de Información  
  
Flujo de trabajo  
  
Control de Versiones  
  
Compartiendo los  
Datos

Reconocimiento a investigadores que han completado estudios en/sobre transparencia y reproducibilidad en ciencias sociales. Ejemplos:

- Psicología
- Ciencias Políticas
- Economía



Edward Leamer



Robert Rosenthal



BERKELEY INITIATIVE FOR TRANSPARENCY  
IN THE SOCIAL SCIENCES

Transparencia  
en  
Investigación  
en las  
Ciencias  
Sociales

Hoces

Introducción

Ética en la  
Investigación  
Científica

Registros

Sesgo de  
Publicación  
Registros

Pre-Analysis  
Plans

P-Hacking

Planes de  
Pre-Análisis

Replicaciones

Protocolo del  
Proyecto, Estándares  
de Información

Flujo de trabajo

Control de Versiones

Compartiendo los  
Datos

## Preguntas?

# Muchas Gracias