

# Contenidos

1. **BITSS**
2. Transparencia en la Investigación Científica
3. Problemas y Soluciones

# Contenidos

1. BITSS
2. **Transparencia en la Investigación Científica**
3. Problemas y Soluciones

# Ética en la Investigación Científica

- 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/Comunalismo**: 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.

# Contenidos

1. BITSS
2. Transparencia en la Investigación Científica
3. **Problemas y Soluciones**

# Problema #1: Sesgo de Publicación

El sesgo de publicación ocurre cuando los estudios publicados en revistas científicas están sobrerrepresentados por estudios que obtienen un particular tipo de resultados (e.g., rechazan la hipótesis nula).

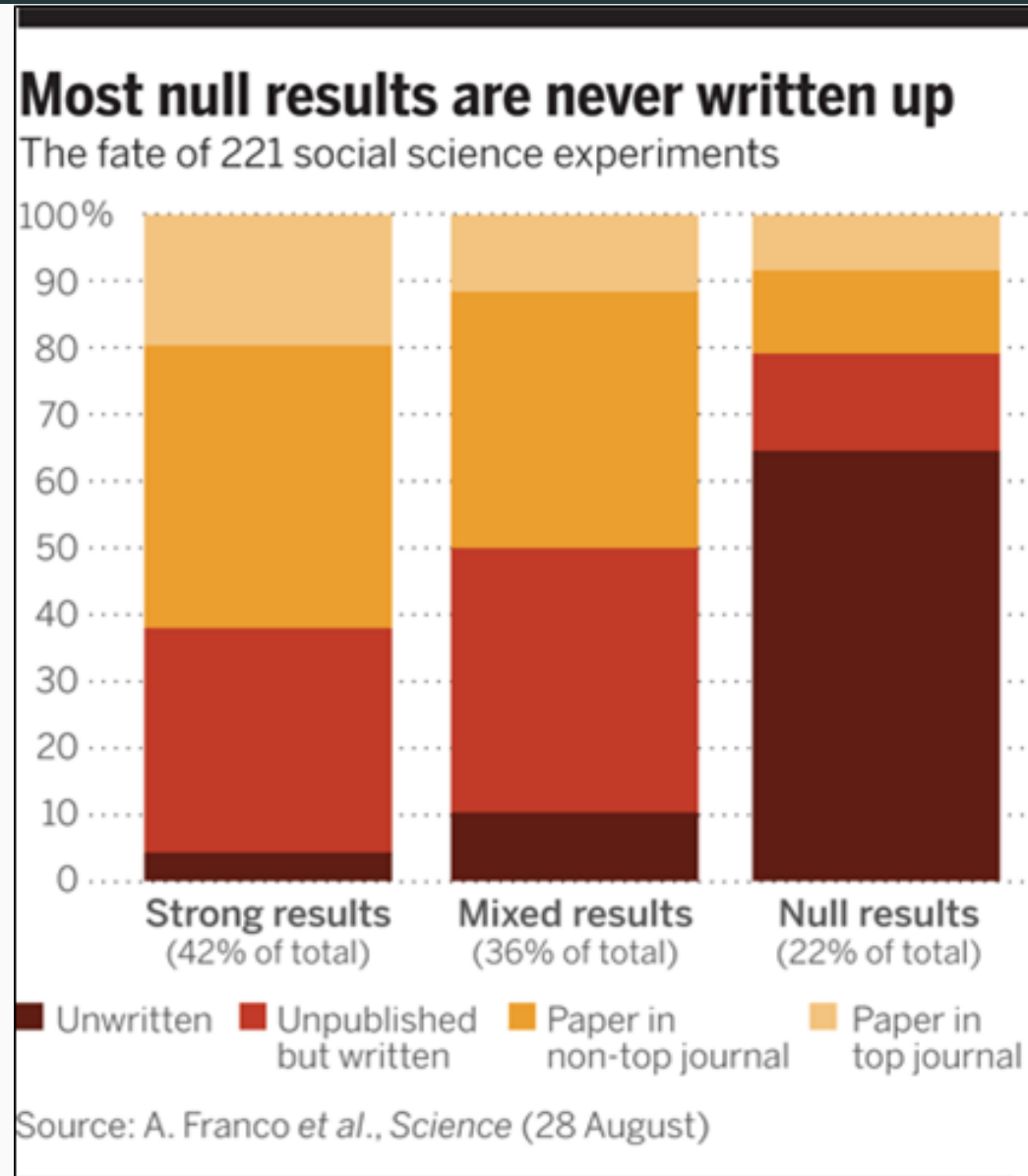
Evidencia que sugiere la existencia de sesgo de publicación:

- El tamaño de los efectos disminuye con el tamaño muestral (Gerber et al 2001).
- La publicación de efectos nulos está desapareciendo en el tiempo, en todas las disciplinas (Fanelli 2011).

Evidencia que mide la magnitud del sesgo de publicación:

- 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 et al 2014)
- En economía Andrews and Kasy (2019) estiman que, para algunas literaturas, los estudios que rechazan la nula son entre 3 y 30 (!) veces más probables de ser publicados en journals top.

# Sesgo de Publicacion en TESS/NSF (Franco et al 2014)

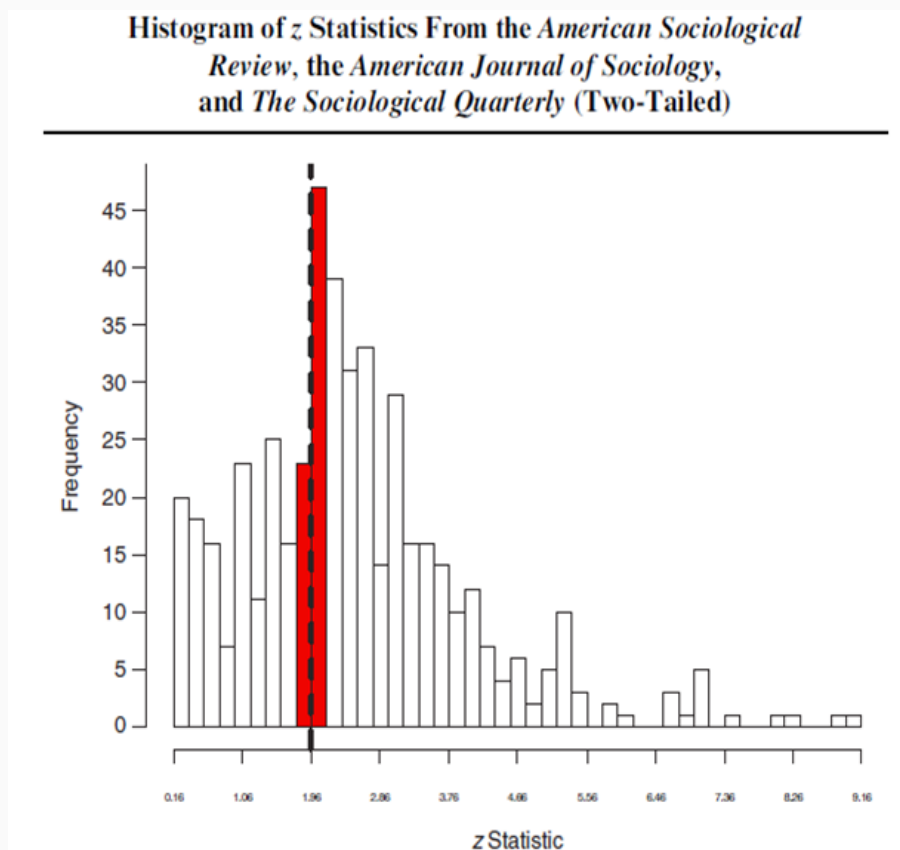


# Problema #2: P-Hacking

- Definición: flexibilidad en el análisis de datos permite presentar *casi cualquier resultado* bajo un umbral arbitrario; significancia estadística pierde sentido.
- Otros nombres: "specification searching" (Leamer 1983), "data-fishing", grados de libertad del investigador, o "data-mining".
- No implica intencionalidad. Puede ser subconsciente, o simplemente una practica estándar del análisis estadístico (Gelman and Loken 2013).
- Evidencia: comportamiento anomalo de test estadisticos entorno a umbrales arbitrarios.

# Evidencia: Sociología y Ciencias Políticas

Sociología (Gerber and Malhotra 2008a)



Ciencias Políticas (Gerber and Malhotra 2008b)

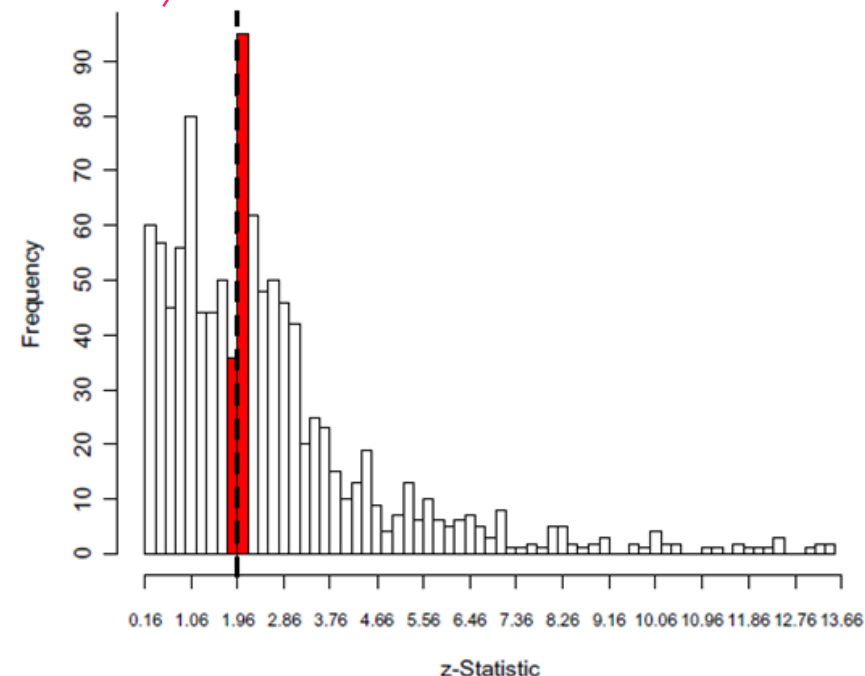


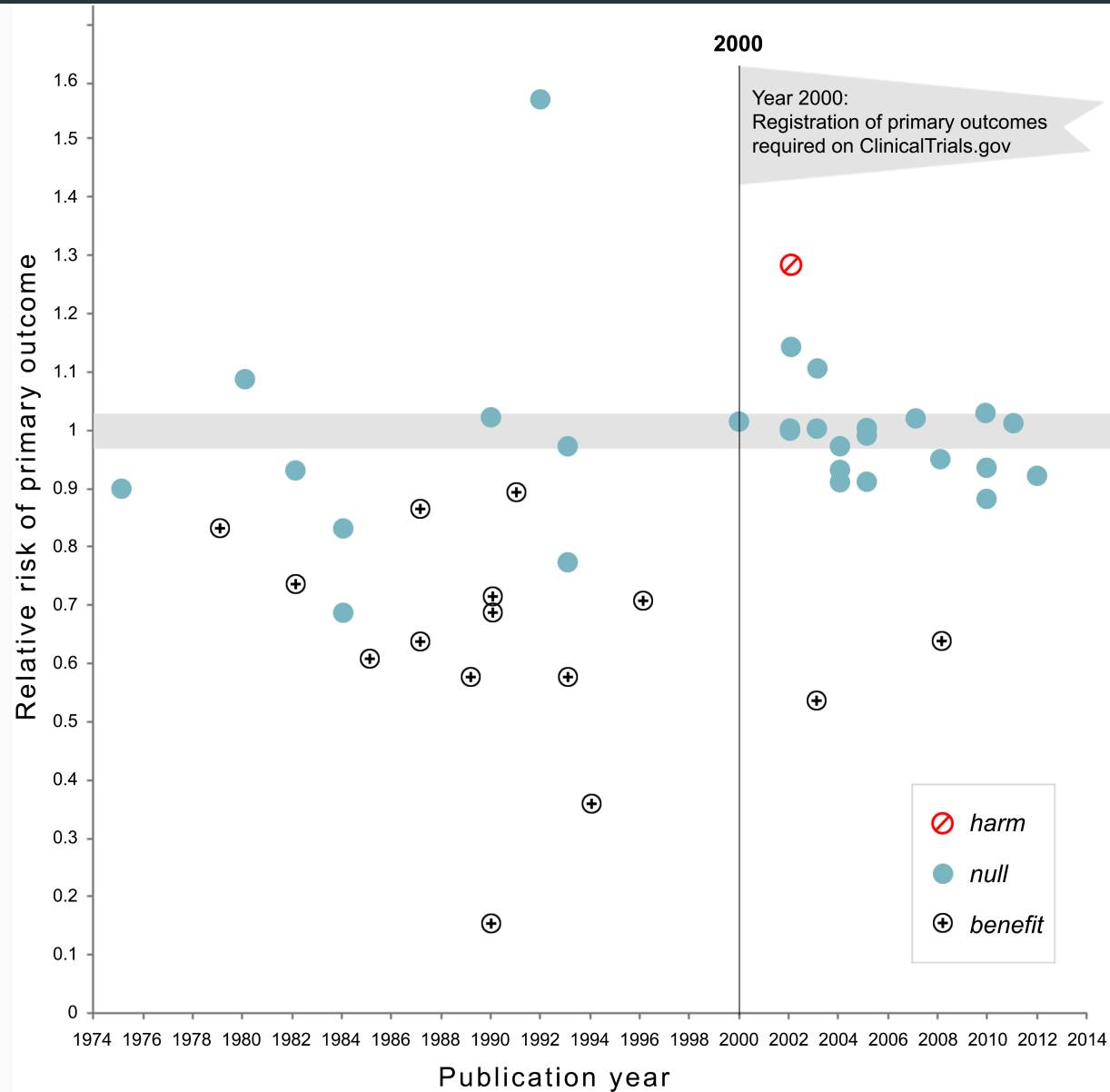
Figure 1(a). Histogram of  $z$ -statistics, *APSR* & *AJPS* (Two-Tailed). Width of bars (0.20) approximately represents 10% caliper. Dotted line represents critical  $z$ -statistic (1.96) associated with  $p = 0.05$  significance level for one-tailed tests.



# Soluciones Para Problemas 1 y 2

- Registros (o pre-registros)
- Planes de pre-analysis
- Reportes registrados

# Registros en Estudios Nutricionales Kaplan and Irvin 2015



# Importancia 1: Un Concepto mas Inclusivo del Objeto Académico

Principio Clarebout:

“Un artículo sobre ciencias computacionales en una publicación científica no es el objeto académico relevante, es simplemente publicidad sobre este objeto. El objeto académico relevante se comprende de todo el entorno de desarrollo de software y todo el conjunto de instrucciones necesarias para generar las figuras [y tablas del artículo]”

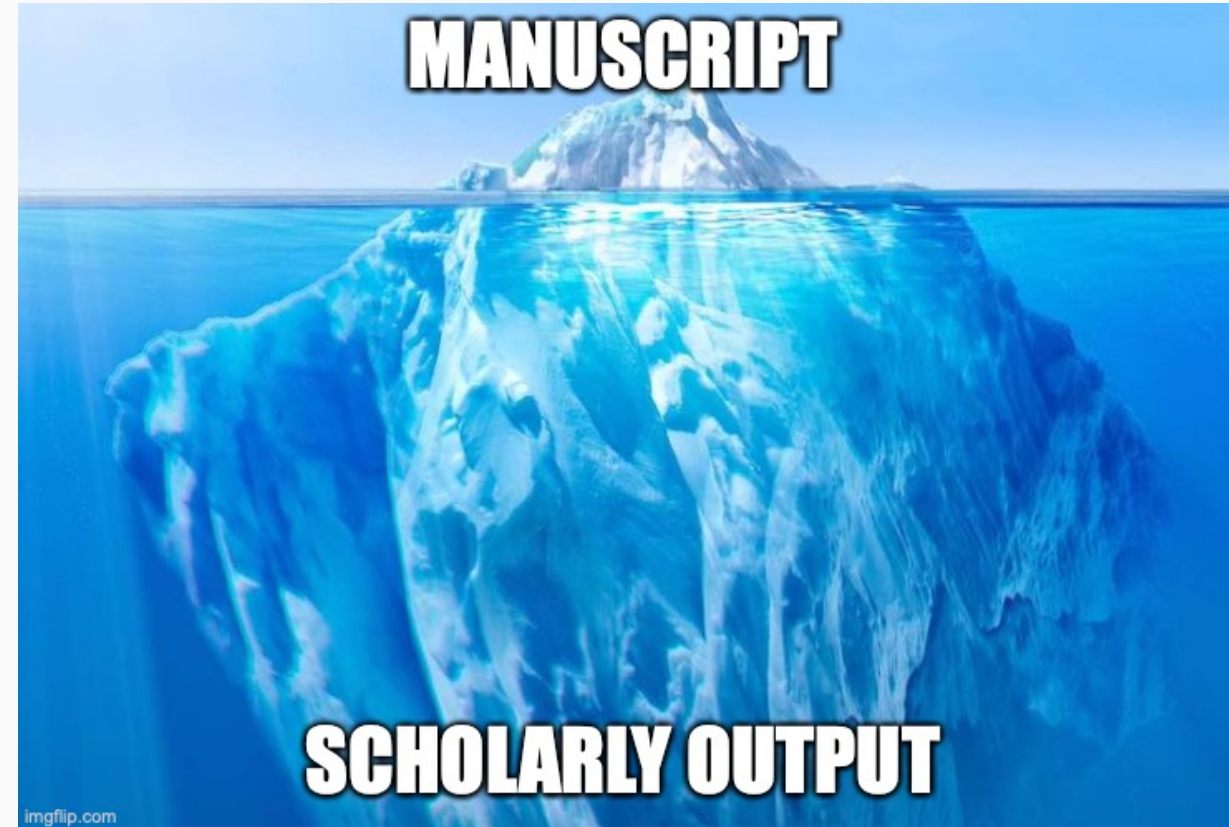
*Buckheit and D.L. Donoho (1995, 2009)*

# Importancia 1: Un Concepto mas Inclusivo del Objeto Académico

Principio Clarebout:

“Un **artículo** sobre ciencias computacionales en una publicación científica no es el objeto académico relevante, **es simplemente publicidad sobre este objeto**. El **objeto académico relevante se comprende de todo el entorno de desarrollo de software y todo el conjunto de instrucciones necesarias para generar las figuras** [y tablas del artículo]>”

*Buckheit and D.L. Donoho (1995, 2009)*



# Importancia 1: Un Concepto mas Inclusivo del Objeto Académico

Beneficios potenciales de adoptar el Principio Clarebout

El PC típicamente se alude para discutir la posibilidad de efectos positivos en:

- Pedagogia
- Generación incremental de conocimiento

Otra dimensión menos discutida:

- Posibles efectos positivos en diversidad, equidad e inclusión: reduce requerimiento de conecciones o habilidades de lenguaje ("apropiadamente polite") para obtener materiales

# Importancia 2: Prevenir la Pérdida de Conocimiento

Cada semestre, estudiantes de posgrado en todo **el mundo** toman cursos en empíricos o aplicados (e.g., Economía Laboral, Psicología Social). Una tarea común consiste en reproducir los resultados de un paper y, posiblemente, testear la robustez de sus resultados.

Etapa	Nuevo Conocimiento
Alcance (Selección y Verificación)	¿Existen datos y código?
Evaluación	Grado de reproducibilidad de una parte específica del artículo
Mejoras	Ej. Corregir directorios, librerías, añadir archivos faltantes, etc.
Testear robustez	Resultados son robustos a especificaciones adicionales

# Prevenir la Pérdida de Conocimiento

- En cada uno de los pasos anteriores el estudiante generó conocimiento nuevo y valioso.
- Este conocimiento no es usualmente diseminado
  - Reporte de fin de semestre, o presentación que no es publicada.
  - El lenguaje a través de estos ejercicios no es estándar (reproducción? replicación?).
  - Existen algunas iniciativas pero se acercan mucho más a un paper completo (ver [Replication Wiki](#)).
- Como resultado, cada año se pierde una cantidad importante de conocimiento.
  - Cada semestre nuevo, nuevos estudiantes repiten ejercicios similares y reinventan la rueda cada vez.
  - Oportunidad perdida de poder construir sobre la base de ejercicios anteriores.
  - Oportunidad perdida de poder agregar este nuevo conocimiento.

# Contexto para ACRe

- La American Economics Association (AEA) crea su primera política de datos en 2006.
  - Se debe publica algunos datos (excepciones disponibles)
- AEA actualiza su política en 2019.
  - Debe publicar todos los datos y código. Publicación es condicional a verificar reproducibilidad (si datos son confidenciales debe proveer documentación extensiva)
  - El 2020 sumó un nuevo requisito de publicar todos los códigos de limpieza de datos, incluso si los datos son confidenciales.
- Debemos esperar alto grado de reproducibilidad después del 2019 (en revistas AEA).
- Antes del 2019 no debemos esperar reproducibilidad del 100%. Pero podemos identificar brechas y cerrarlas.



# Accelerating Computational Reproducibility

## Más Allá de Juicios Binarios

Reproducciones pueden gravitar fácilmente hacia interacciones hostiles

- Investigadores junior tienen incentivo para enfatizar reproducciones fallidas.
- Autores están en una posición más senior donde puede desalentar reproducciones de investigadores junior.
- La prensa tiende a enfocarse en resultados produzcan titulares atractivos.

## Nuestro approach:

No queremos decir:

"Artículo X es (ir)reproducible"

Si queremos decir:

"Resultado Y en artículo X tienen un alto/bajo **grado** de reproducibilidad de acuerdo a **varios** ejercicios de reproducción. Más aún, **mejoras** fueron realizadas al paquete de reproducción original, **aumentando** su reproducibilidad a un nivel más alto"

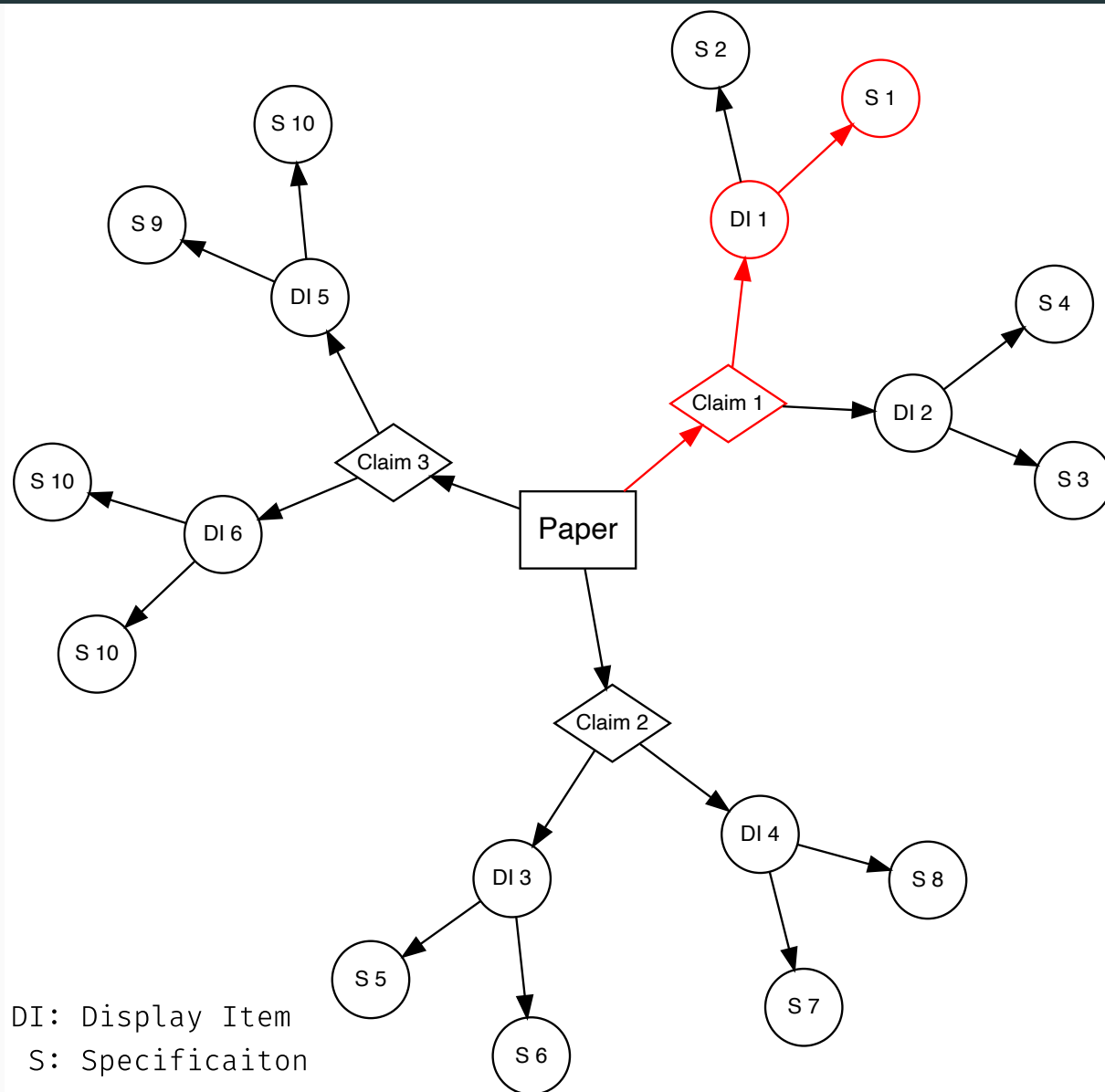
# ACRe: Marco Conceptual

Cada **ejercicio de reproducción** está centrado entorno a una **afirmación científica**

Un artículo puede contener varias afirmaciones.

Cada afirmación se basará en **objetos de resultados**: tablas, figuras y resultados en texto.

Cada ejercicio de reproducción es a nivel de afirmaciones, y los reproductores deben documentar sus **especificaciones** de interes



# Demo: socialsciencereproduction.org

[Home](#)[About](#)[Events](#)[Guide](#)[Forum](#)[Metrics](#)

Welcome to the

## Social Science Reproduction Platform

Accelerating computational  
reproducibility in the social sciences

GET STARTED!

Search reproductions by paper title, DOI, or author name



# Seleccionar un Artículo

## Article Information

### Abstract

How large are the benefits of transportation infrastructure projects, and what explains these benefits? This paper uses archival data from colonial India to investigate the impact of India's vast railroad network. Guided by four results from a general equilibrium trade model, I find that railroads: (1) decreased trade costs and interregional price gaps; (2) increased interregional and international trade; (3) increased real income levels; and (4) that a sufficient statistic for the effect of railroads on welfare in the model accounts well for the observed reduced-form impact of railroads on real income in the data.

### Citation

Donaldson, Dave. 2018. "Railroads of the Raj: Estimating the Impact of Transportation Infrastructure." *American Economic Review*. 108 (4-5): 899-934.

DOI: 10.1257/aer.20101199

Choose Format:

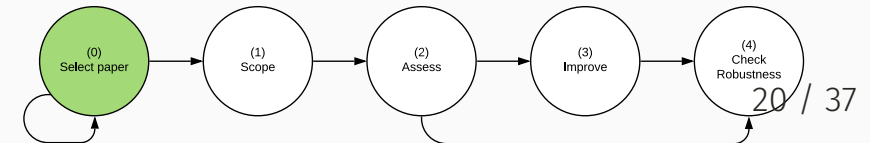


### Additional Materials

Data Set (17.80 MB)

Online Appendix (247.97 KB)

Author Disclosure Statement(s) (63.46 KB)



# Definir Alcance del Ejercicio

Go to page 19

## C. Results

Table 2 presents ordinary least squares (OLS) estimates of equation (12). In column 1 I estimate the effect of the lowest-cost route effective distance on trade when the relative costs of each mode ( $\alpha$ ) are set to observed historical relative freight rate estimates. I use the relative per unit distance freight rates described in Section IB (at their midpoints):  $\alpha^{road} = 4.5$ ,  $\alpha^{river} = 3.0$ , and  $\alpha^{coast} = 2.25$  relative to the freight rate of railroad transport, normalized to 1). Column 1 demonstrates that the elasticity of trade costs with respect to the lowest-cost route effective distance, calculated at observed freight rates, is 0.088, and this is statistically significant at the 5 percent level.

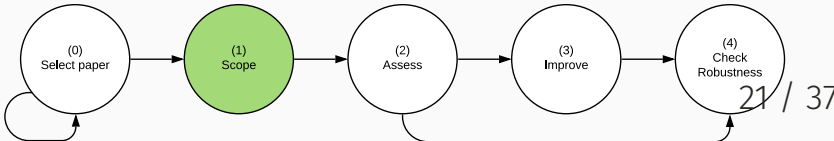
However, as argued in Section IB, it is possible that these observed relative freight rates do not capture the full benefits (such as increased certainty or savings) of railroad transport relative to alternative modes of transportation. For this reason the NLS specification in column 2 estimates the relative freight rates (i.e., the parameters  $\alpha$ ) that minimize the sum of squared residuals in equation (12).

Column 2 is my preferred specification. When the mode-wise distance costs (i.e.,  $\alpha$ ) are not restricted to be equal to the observed freight rates, the estimated elasticity of trade costs with respect to effective distance (i.e.,  $\delta$ ) rises to 0.169. Even when controlling for all unobserved, time-constant determinants of trade costs between all salt sources and destinations, as well as unrestricted shocks to the source price of each salt type, reductions in trade costs along lowest-cost routes (estimated from railroad-driven time variation in these routes alone) have a large effect on reducing salt price gaps over space.

TABLE 2—RAILROADS AND TRADE COSTS: STEP 1

Dependent variable: log salt price at destination	(1)	(2)
log effective distance to source, along lowest-cost route (at historical freight rates)	0.088 (0.028)	
log effective distance to source, along lowest-cost route (at estimated mode costs)		0.169 [0.062, 0.296]
Estimated mode costs per unit distance:		
Railroad (normalized to 1)		1 N/A
Road		2.375 [1.750, 10.000]
River		2.250 [1.500, 6.250]
Coast		6.188 [5.875, 10.000]
Observations	7,345	7,345
R <sup>2</sup>	0.946	0.946

Notes: Regressions estimating equation (12) using data on 6 types of salt (listed in online Appendix A), from 133 districts in Northern India, annually from 1861 to 1930. Column 1 and column 2 estimated by OLS and NLS respectively; both include salt type  $\times$  year and salt type  $\times$  destination fixed effects. "Effective distance to source."



# Evaluar

Analysis

descriptives

T1\_descriptives.do

income

income estimation.do

income\_es...\_temp.dta

income\_estimation.log

simulation

eq\_cond2.m

exp\_shares.csv

func\_solve2.m

land\_area.csv

secant\_price2.m

sim\_prep\_temp.dta

T5\_simulation.m

trade costs

haversine.m

matlab\_bg1\_4.0\_osx64

TC estimation.do

TC\_est\_prep.m

TC\_estima...n\_temp.dta

TC\_estimation.log

trade flows

gravity estimation.do

gravity\_estimation.log

income\_temp.dta

rainfall\_temp.dta

TF\_est\_prep.m

Data

crosswalks

district-bl...\_wdistid.dta

district-bl...\_relation.dta

income

income.dta

maps

bd\_ns\_boundary2.csv

coast\_simplepoint2.csv

RAIL\_dummies.dta

railways\_D...point2.csv

rivers\_simplepoint2.csv

Prices\_salt

od\_salt\_list.csv

prices\_salt.dta

rainfall

crop rainfall.dta

trade flows

freight class.dta

trade\_data.dta

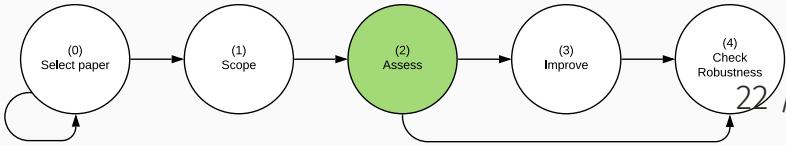
trade\_prices\_1870.dta

weight-value.dta

Readme for data files.pdf

Table 2

TC estimation.do	
LCRED_RoX_CoY_RiZ.csv	
TC_est_prep.m	
railways_Dissolve_Simplify2_point2.csv	
bd_ns_boundary2.csv	
rivers_simplepoint2.csv	
rivers_simplepoint2.csv	
coast_simplepoint2.csv	
od_salt_list.csv	
prices_salt.dta	
railways_Dissolve_Simplify2_point2.csv	
bd_ns_boundary2.csv	
rivers_simplepoint2.csv	
coast_simplepoint2.csv	
od_salt_list.csv	



# Evaluar

Analysis

descriptives

T1\_descriptives.do

income

income estimation.do

income\_es...\_temp.dta

income\_estimation.log

simulation

eq\_cond2.m

exp\_shares.csv

func\_solve2.m

land\_area.csv

secant\_price2.m

sim\_prep\_temp.dta

T5\_simulation.m

trade costs

haversine.m

matlab\_bg1\_4.0\_osxf

TC estimation.do

TC\_est\_prep.m

TC\_estima...n\_temp.dta

TC\_estimation.log

trade flows

gravity estimation.do

gravity\_estimation.lo

income\_temp.dta

rainfall\_temp.dta

TF\_est\_prep.m

Data

crosswalks

district-bl...wdistid.i

district-bl...relation.i

income

income.dta

maps

bd\_ns\_boundary2.csv

coast\_simplepoint2.c

RAIL\_dummies.dta

railways\_D...point2.c

rivers\_simplepoint2.c

Prices\_salt

od\_salt\_list.csv

prices\_salt.dta

rainfall

crop rainfall.dta

trade flows

freight class.dta

trade\_data.dta

trade\_prices\_1870.dta

weight-value.dta

Readme for data files.pdf

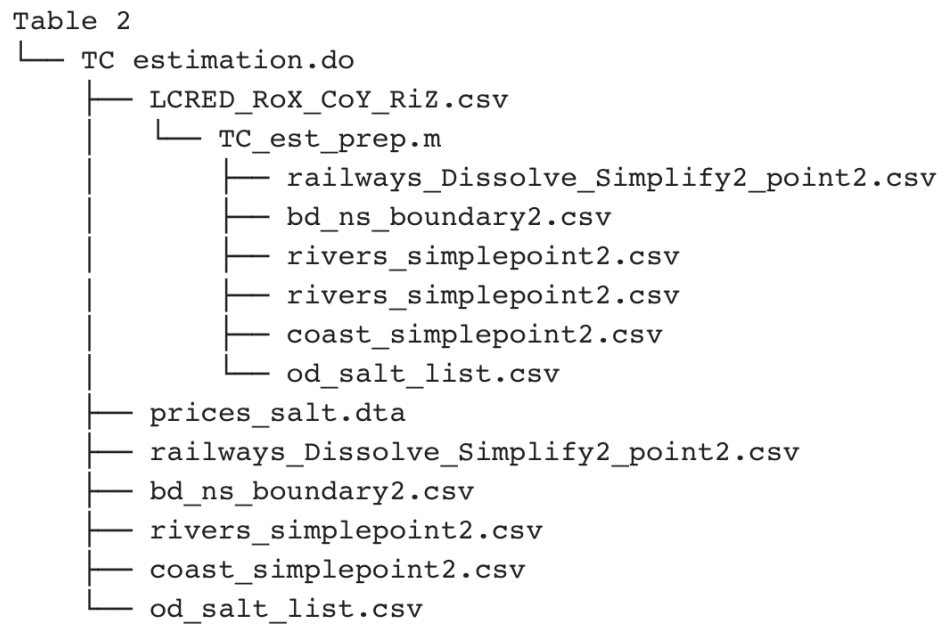


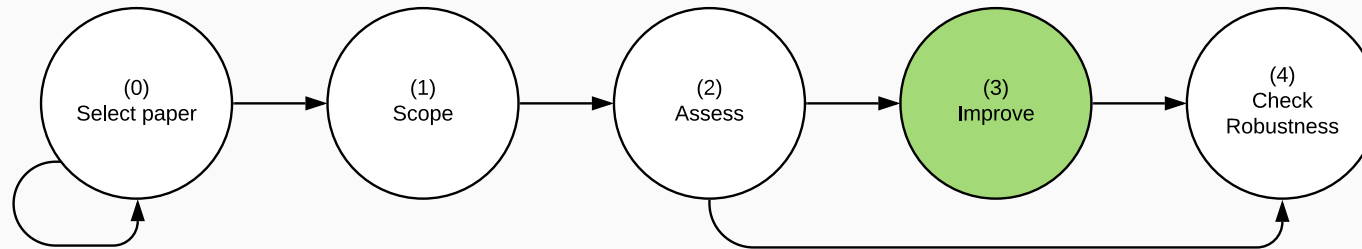
Table 3.5: Levels of Computational Reproducibility  
(P denotes “partial,” C denotes “complete”)

Availability of materials, and reproducibility										
	Analysis Code		Analysis Data		CRA	Cleaning Code		Raw Data		CRR
	P	C	P	C		P	C	P	C	
L1: No materials	–	–	–	–	–	–	–	–	–	–
L2: Only code	✓	✓	–	–	–	–	–	–	–	–
L3: Partial analysis data & code	✓	✓	✓	–	–	–	–	–	–	–
L4: All analysis data & code	✓	✓	✓	✓	–	–	–	–	–	–
L5: Reproducible from analysis	✓	✓	✓	✓	✓	–	–	–	–	–
L6: All cleaning code	✓	✓	✓	✓	–	✓	✓	–	–	–
L7: Some raw data	✓	✓	✓	✓	–	✓	✓	✓	–	–
L8: All raw data	✓	✓	✓	✓	–	✓	✓	✓	✓	–
L9: All raw data + CRA	✓	✓	✓	✓	✓	✓	✓	✓	✓	–
L10: Reproducible from raw data	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓

<sup>a</sup> **Computationally Reproducible from Analytic data (CRA):** The output can be reproduced with minimal effort starting from the *analytic* datasets.

<sup>b</sup> **Computationally Reproducible from Raw data (CRR):** The output can be reproduced with minimal effort from the *raw* datasets.

# Mejoras



## Tres tipos de mejoras

1. Mejoras a nivel de artículo
2. Mejoras a nivel de objeto de resultado
3. Sugerencias específicas de mejoras futuras



# Mejoras: A Nivel de Artículo

- Use version control software (Git/Github).
- Improve documentation: comments, indentations, object names, etc.
- Re-organize the reproduction package into a set of folders and sub-folders that follow **standardized best practices**, and add a master script that executes all the code in order, with no further modifications. **See AEA's reproduction template.**
- Literate programming environment (e.g., Jupyter notebooks, RMarkdown)
- Re-write code using a different statistical software (ideally open source, like R, Python, or Julia).
- Set up a computing capsule (e.g., **Binder** and **Code Ocean**).



# Mejoras: A Nivel de Objeto de Resultado

- Adding missing raw data: files or meta-data
  - Example: "Add raw temperature and relative humidity data"
- Adding missing analytic data files
  - Example: "Copy the row files from Data folder into new `Analysis\trade cost\Input`"
- Adding missing analysis or cleaning code
  - Example: "Replaced broken Wald bootstrap code with updated code/command"
- Debugging code
  - Example: "was counting each group 4 times in round 1, so fixed that"



# Mejoras: Sugerencias Especificas

We ask reproducer to leave concise and actionable tasks for other reproducers in the future.

Example 1:

"Revise the .aml and .bat code scripts to reflect reorganized structure"

Example 2:

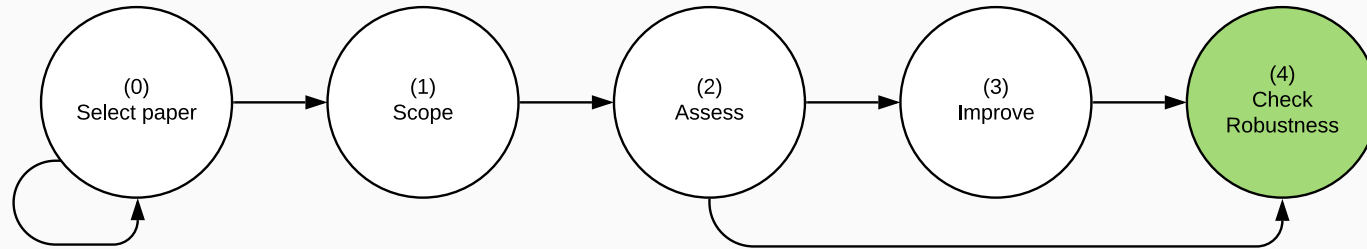
"Provide data and codes generating the other two figures in the paper, which are not given in the replication file."

Example 3:

"Table 3 can be reproduced identically from the [...] analytic data files. I was not able to reproduce the analytic data files due to lack of access to ArcGIS software, but the code scripts and raw data files [...] are included in the reproduction package."



# Chequeos de Robustez



## Dos partes de robustez:

1. Incrementar el numero de decisiones analiticas
2. Justificar un específico chequeo de robustez

# Robustez

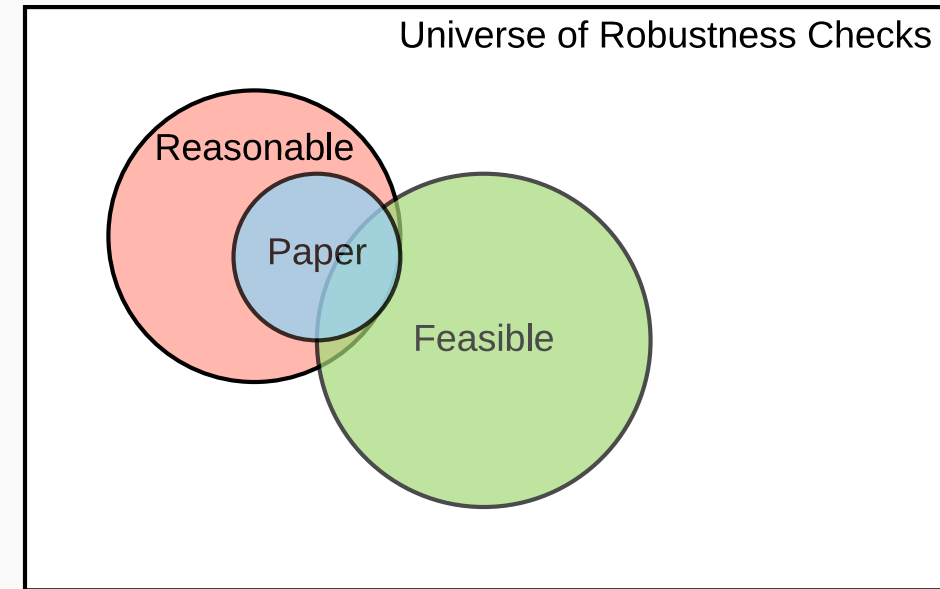
**Cheques de Robustez:** revisión de los resultados ante cualquier posible cambio en una decisión computacional, incluyendo análisis y limpieza de datos

**Especificaciones Razonables** (Simonsohn et. al., 2018):

1. Test sensatos de la pregunta de investigación subyacente
2. Estadísticamente válidos, y
3. No redundantes con otras especificaciones.

Reproductores son capaces de contribuir en dos dimensiones:

- Mapear el universo de todos los posibles chequeos de robustez.
- Proponer y justificar un test específico como razonable.



# Reproducción Completa: Ejemplo

Reproduction of: Railroads of the Raj: Estimating the Value of Transportation Infrastructure in American Colonial India  
<https://doi.org/10.1257/aer.20101199>

Reproducer: Emma Ng  
Date created: November 19 2020  
Date submitted: August 03 2021

Claims assessed: 1  
Display Items (DI) assessed: 1  
Improvements: 3

Robustness tests:

- Feasible choices added: 0
- Specifications justified: 0

Claims identified by reproducer:

1. "The paper estimated the value of the extent to which the costs of different modes of transportation are estimated to be reduced by the reduction of LCRED (lowest-cost route effective distance) trading cost. This estimate has a 95% confidence interval."

Reproducibility of Display Items:

1. Table 2 - Table 2 -- Railroads and Trade Costs: Step 1. ...

Original reproduction package: <https://www.aeaweb.org/articles?id=10.1257/aer.20101199>  
Revised reproduction package: <https://github.com/em-ng21/railroads-of-the-raj>  
Original authors' availability for further inquiries: Not sure

0

Select a paper

👁

Declare the paper that you will analyze in the remainder of the exercise and any other "candidate" papers for which you were unable to obtain access to the full text.

VIEW THIS SECTION

1

Scoping

👁

2

Assessment

👁

3

Improvement

👁

4

Robustness

👁

0

Select a paper

👁

Declare the paper that you will analyze in the remainder of the exercise and any other "candidate" papers for which you were unable to obtain access to the full text.

# Reproducción Completa: Ejemplo

Reproduction of: Railroads of the Raj: Estimating the Impact of Transportation Infrastructure on the Indian Economy  
<https://doi.org/10.1257/aer.20101199>

Reproducer: Emma Ng  
Date created: November 19 2020  
Date submitted: August 03 2021

Claims assessed: 1  
Display Items (DI) assessed: 1  
Improvements: 3

Robustness tests:

- Feasible choices added: 0
- Specifications justified: 0

Claims identified by reproducer:

1. "The paper estimated the value of the extent to which the costs of different modes of transportation are estimated to reduce the reduction of LCRED (lowest-cost route effective distance trading cost. This estimate has a 95% confidence interval."

Reproducibility of Display Items:

1. Table 2 - Table 2 -- Railroads and Trade Costs: Step 1: ...

Original reproduction package: <https://www.aeaweb.org/articles?id=10.1257/aer.20101199>  
Revised reproduction package: <https://github.com/em-ng21/railroads-of-the-raj>  
Original authors' availability for further inquiries: Not sure

0

Select a paper

Declare the paper that you will analyze in the remainder of the exercise and record other "candidate" papers for which you were unable to obtain access to a reproduction package. See detailed guidance [here](#).

VIEW THIS SECTION

1

Scoping

2

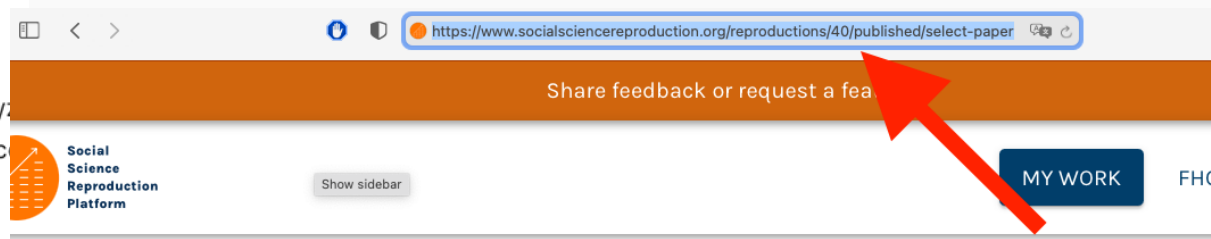
Assessment

3

Improvement

4

Robustness



Select a paper

Declare the paper that you will analyze in the remainder of the exercise and record other "candidate" papers for which you were unable to obtain access to a reproduction package. See detailed guidance [here](#).

RETURN TO STAGES OVERVIEW

## Basic information

- **Compartelo:** Con los autores, tus profesores, colegas, ponlo en to CV
- **Discutelo:** Foro en Discourse para hablar sobre reproducciones
- **Citalo:** Cada reproduccione completa tiene un DOI!

# Promoviendo un Intercambio Construtivo de Ideas

- 1 - Contacting the original author(s) when there is no reproduction package
  - 2 - Contacting the original author(s) to request specific missing items of a reproduction package
  - 3 - Asking for additional guidance when some materials have been shared
  - 4 - Response when the original author has refused to share due to *undisclosed reasons*
  - 5 - Response when the original author has refused to share due to legal or ethical restrictions of the data
  - 6 - Contacting the original author to share the results of your reproduction exercise
  - 7 - Responding to hostile responses from original authors
- Under development: sample responses from authors to reproducers



# Ejemplo 1: Cuando No Hay Materiales de Reproducción

**Subject:** Reproduction package for [“Title of the paper”]

Dear Dr. [Lastname of Corresponding Author],

I am contacting you to request a reproduction package for your paper titled [Title] which was published in [Reference]. A reproduction package may contain (raw and/or analytic) data, code, and other documentation that makes it possible to reproduce paper. Would you be able to share any of these items?

I am a [position] at [Institution], and I would like to reproduce the results, tables, and other figures using the reproduction materials mentioned above. I have chosen this paper because [add context ...]. **Unfortunately, I was not able to locate any of these materials on the journal website, Dataverse [or other data and code repositories], or in your website.**

**I will record the result of my reproduction attempt on ACRE [...]. With your permission, I will also record the materials you share with me, which would allow access for other reproducers and avoid repeated requests directed to you. Please let me know if there are any legal or ethical restrictions that apply to all or parts of the reproduction materials so that I can take that into consideration during this exercise.**

In addition to your response above, would you be available to respond to future (non-repetitive) inquiries from me or other reproducers conducting an ACRE exercise? **Though your cooperation with my and/or any future request would be extremely helpful, please note that you are *not required to comply*.**

Since I am required to complete this project by [date], I would appreciate your response by [deadline].

Let me know if you have any questions. Please also feel free to contact my supervisor/instructor [Name (email)] for further details on this exercise. Thank you in advance for your help!

Best regards,

[Reproducer]

# Ejemplo 2: Respondiendo ante la ausencia de algunos materiales

## Template email:

**Subject:** Clarification for reproduction materials for ["Title of the paper"]

Dear Dr. [Lastname of Corresponding Author],

Thank you for sharing the materials. They have been immensely helpful for my work.

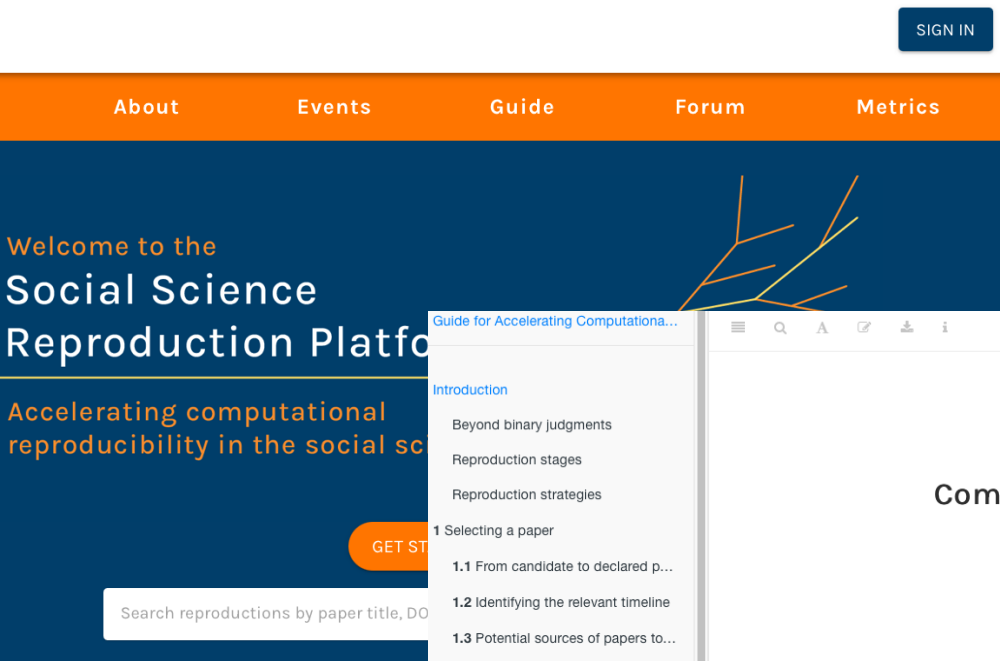
Unfortunately, I ran into a few issues as I delved into the reproduction exercise, and I think your guidance would be helpful in resolving them. **[Describe the issues and how you have tried to resolve them. Describe whatever files or parts of the data or code are missing. Refer to examples 1 and 2 below for more details]**.

Thank you in advance for your help.

Best regards,

[Reproducer]

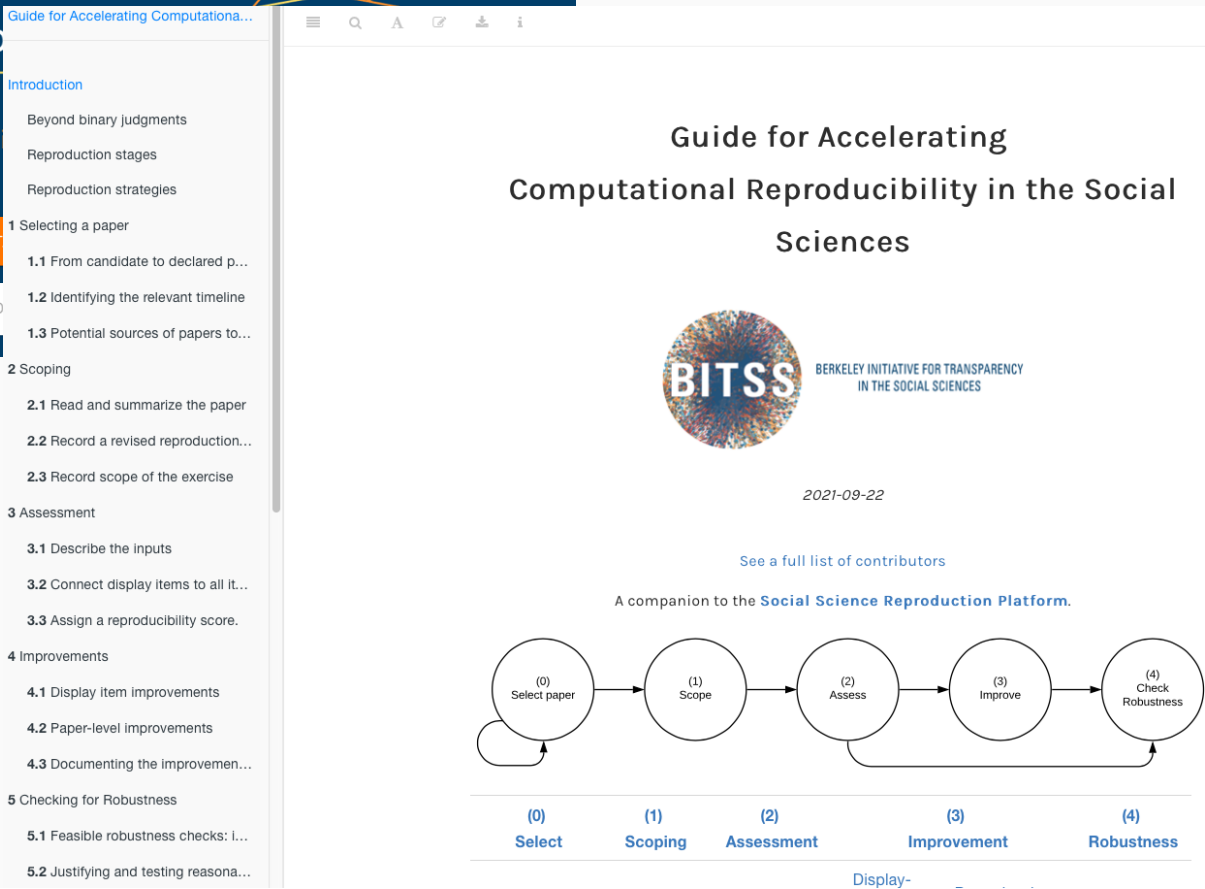
# Donde Empiezo?



Consulta  
nuestra  
guía

SIGN IN

Úsala en tu clase  
o proyecto independiente



# Donde Empiezo?

About

Events

Guide

Forum

Metrics

WELCOME TO THE

SOCIAL SCIENCE

REPRODUCTION PLATFORM

ACCELERATING COMPUTATIONAL

REPRODUCIBILITY IN THE SOCIAL SCIENCES

GET STARTED

Search reproductions by paper title, DOI, or keyword

Guide for Accelerating Computational Reproducibility in the Social Sciences

Introduction

Beyond binary judgments

Reproduction stages

Reproduction strategies

1 Selecting a paper

1.1 From candidate to declared paper

1.2 Identifying the relevant timeline

1.3 Potential sources of papers to consider

2 Scoping

2.1 Read and summarize the paper

2.2 Record a revised reproduction

2.3 Record scope of the exercise

3 Assessment

3.1 Describe the inputs

3.2 Connect display items to all items

3.3 Assign a reproducibility score

4 Improvements

4.1 Display item improvements

4.2 Paper-level improvements

4.3 Documenting the improvements

5 Checking for Robustness

5.1 Feasible robustness checks: I...

5.2 Justifying and testing reasona...

Úsala en tu clase  
o proyecto independiente

Consulta  
nuestra  
guía

SIGN IN

Guide for Accelerating  
Computational Reproducibility in the Social  
Sciences

BITSS

BERKELEY INITIATIVE FOR TRANSPARENCY  
IN THE SOCIAL SCIENCES

2021-09-22

See a full list of contributors

A companion to the [Social Science Reproduction Platform](#).

(0) Select paper

(1) Scope

(2) Assess

(3) Improve

(4) Check Robustness

(0) Select

(1) Scoping

(2) Assessment

(3) Improvement

(4) Robustness

Social Science Reproduction Platform

SEARCH

≡

F

all categories ▾

all tags ▾

Categories

Latest

New (3)

Unread (1)

Top

+ New Topic

Category

Topics

Latest

Getting Started

0

Version 1.0, developed in April 2021

Reproductions

5

1 unread

Discuss and share resources about ongoing or completed reproductions sorted by papers reproduced. To join the discussion about a particular paper, look up the paper's DOI, replacing "." with "-".

Tips for Reproducibility

1

Find and share tips and resources for reproducible workflows that aren't linked to a particular reproduction.

Help

3

3 new

Ask questions on how to use the SSRP.

For instructors

0

Connect with a community of

Welcome to Discourse

0

Nov '20

Restricted-use data

0

23h

Catalog of reproductions

0

23h

No replication package

1

23h

Feedback for my reproduction of Twenty-year economic impacts of deworming

1

Aug 11

How to use the SSRP

2

Jul 22

Pregunta en el foro

36 / 37

# Gracias

---

[fhoces@berkeley.edu](mailto:fhoces@berkeley.edu)