

How Research Transparency Can Improve Social Science: **Quality and Quantity**

Pre-doctoral Research in Economics (PRE) Workshop

Fernando Hoces de la Guardia, BITSS
June 2021 | [slides](#)

BITSS

The Berkeley Initiative for Transparency in the Social Sciences works to improve the credibility of science by advancing transparency, reproducibility, rigor, and ethics in research.

Core Team

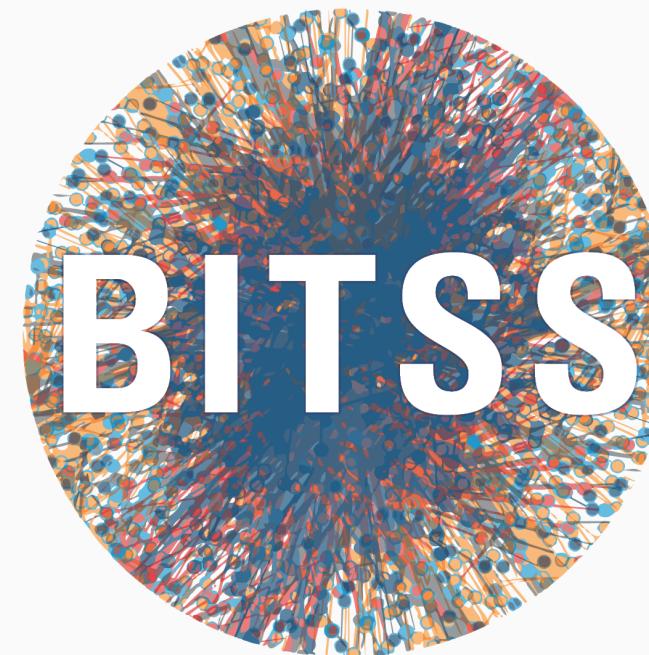
Aleks Bogdanoski

Fernando Hoces de la Guardia

Katie Hoeberling

Edward Miguel

We are part of the Center for Effective Global Action ([CEGA](#)).



BERKELEY INITIATIVE FOR TRANSPARENCY
IN THE SOCIAL SCIENCES

Many Others

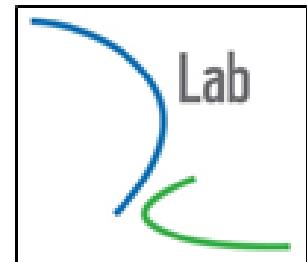
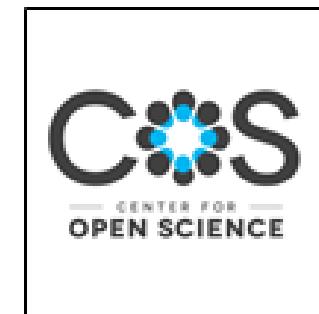
CEGA staff

Undergrad and Graduate RAs

Catalysts

Outside Collaborators
(Researchers, Programmers)

Part of the much larger Open Science Community



ICPSR

Table of Contents

1. **Quality: Problems and Solutions**
2. Quantity: How research transparency can make social science more inclusive

Researchers and Scientific Norms (Anderson et. al., 2007)

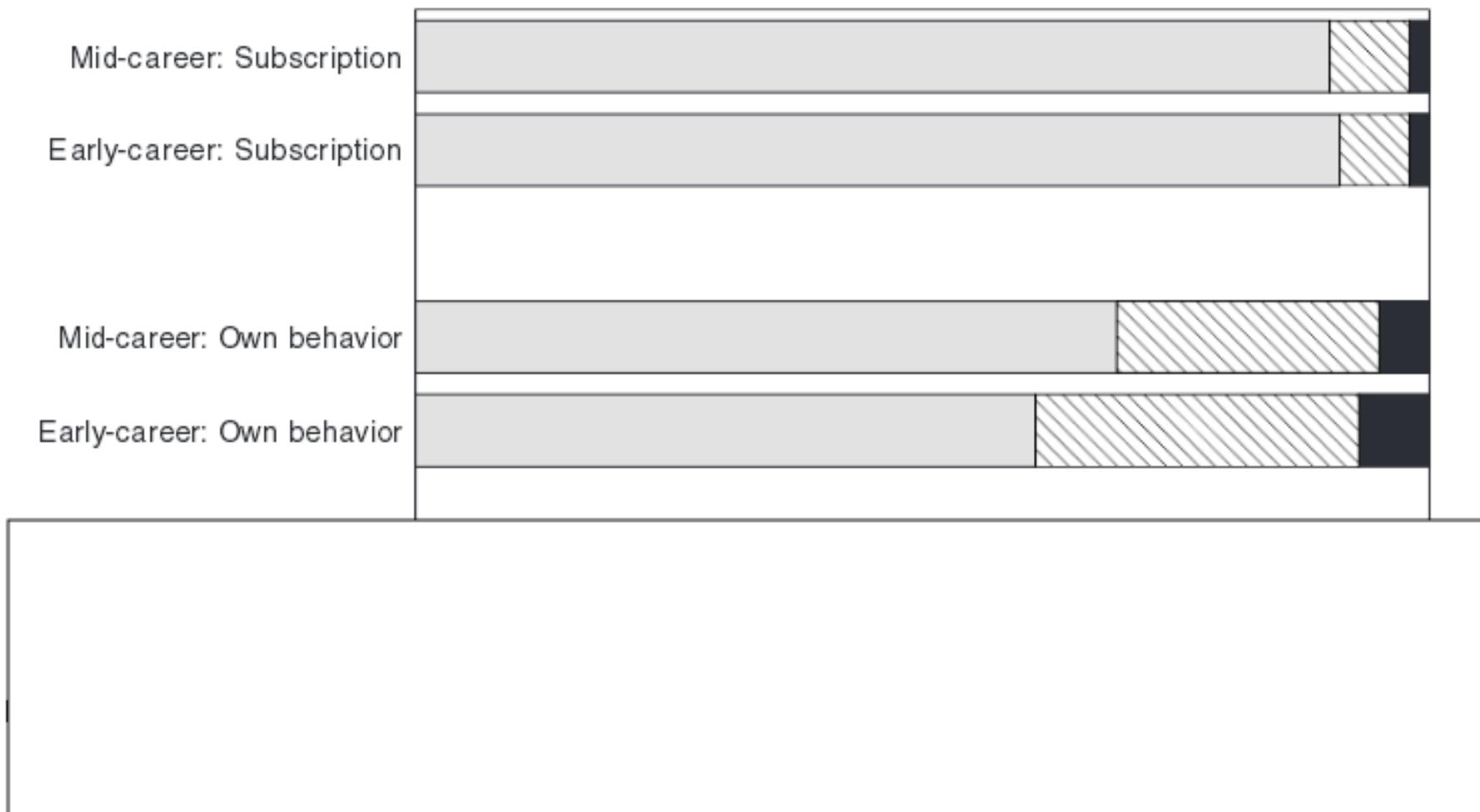


FIG. 3. Norm versus Counternorm Scores: Percent with Norm > Counternorm (dotted), Norm = Counternorm (striped), Norm < Counternorm (solid).

Researchers and Scientific Norms (Anderson et. al., 2007)

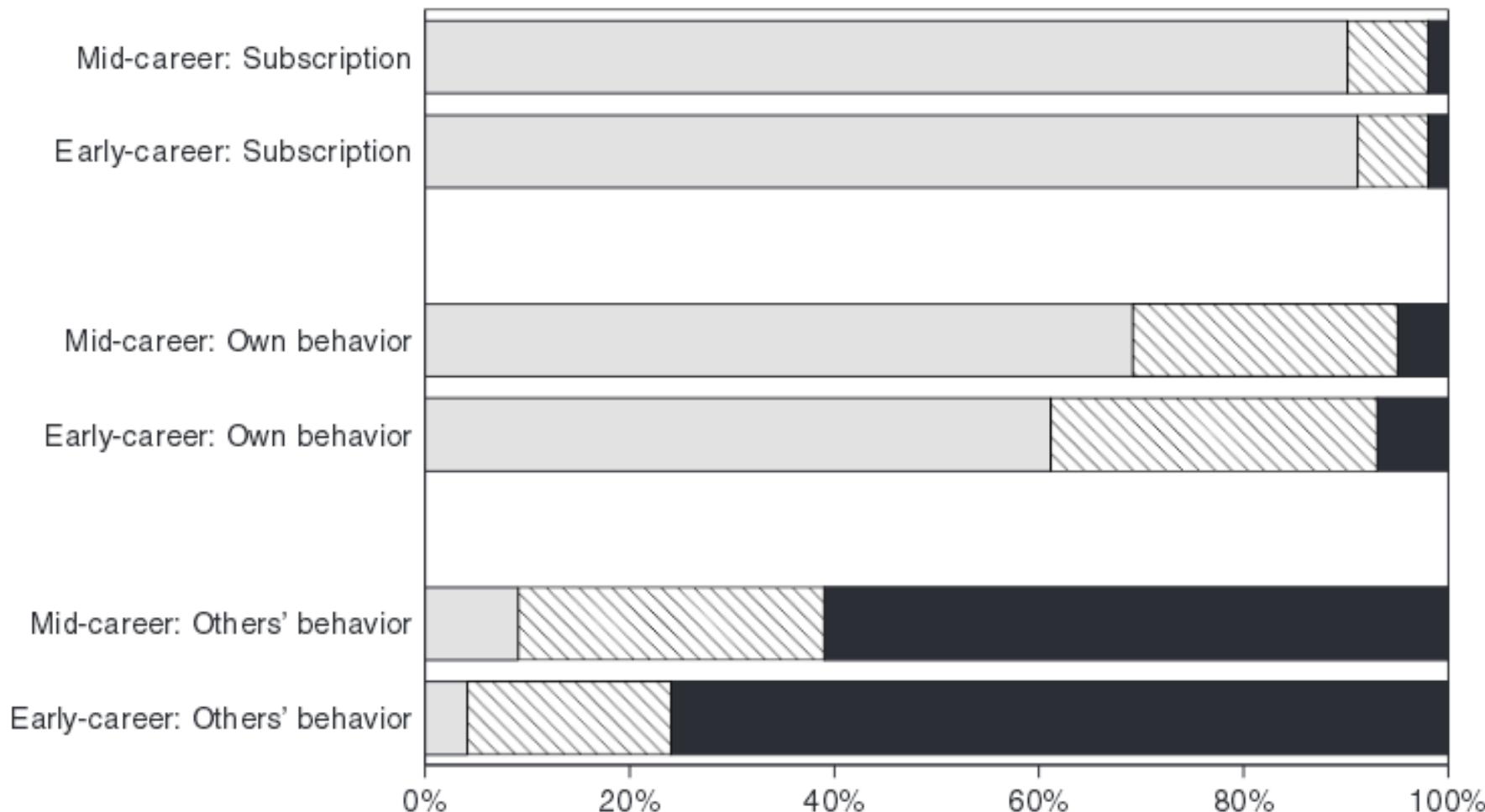
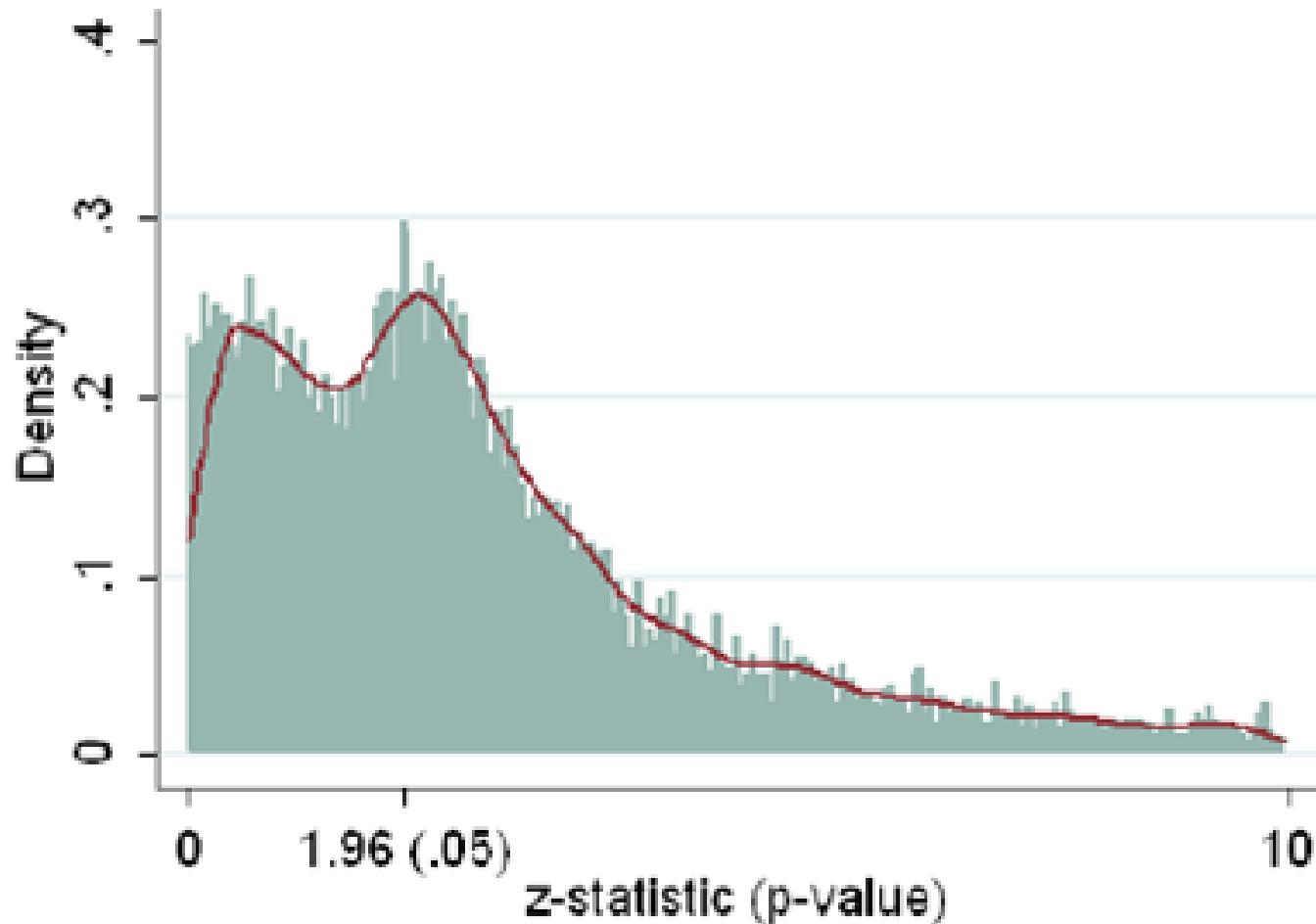


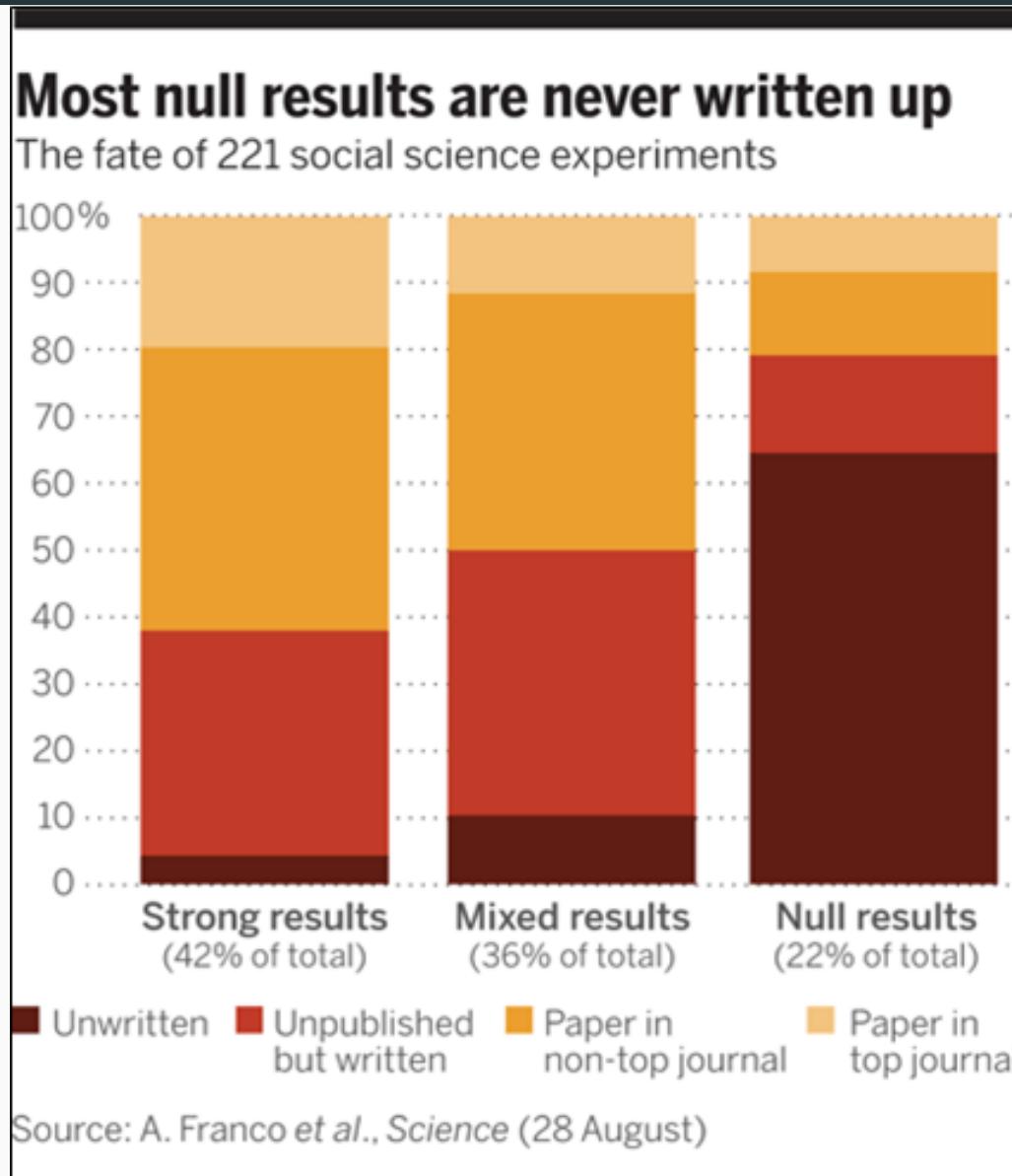
FIG. 3. Norm versus Counternorm Scores: Percent with Norm > Counternorm (dotted), Norm = Counternorm (striped), Norm < Counternorm (solid).

P-hacking (for Economics: Brodeur et. al 2016, 2020)

(b) Unrounded distribution of z-statistics.



Publication Bias (Franco et. al. 2014)



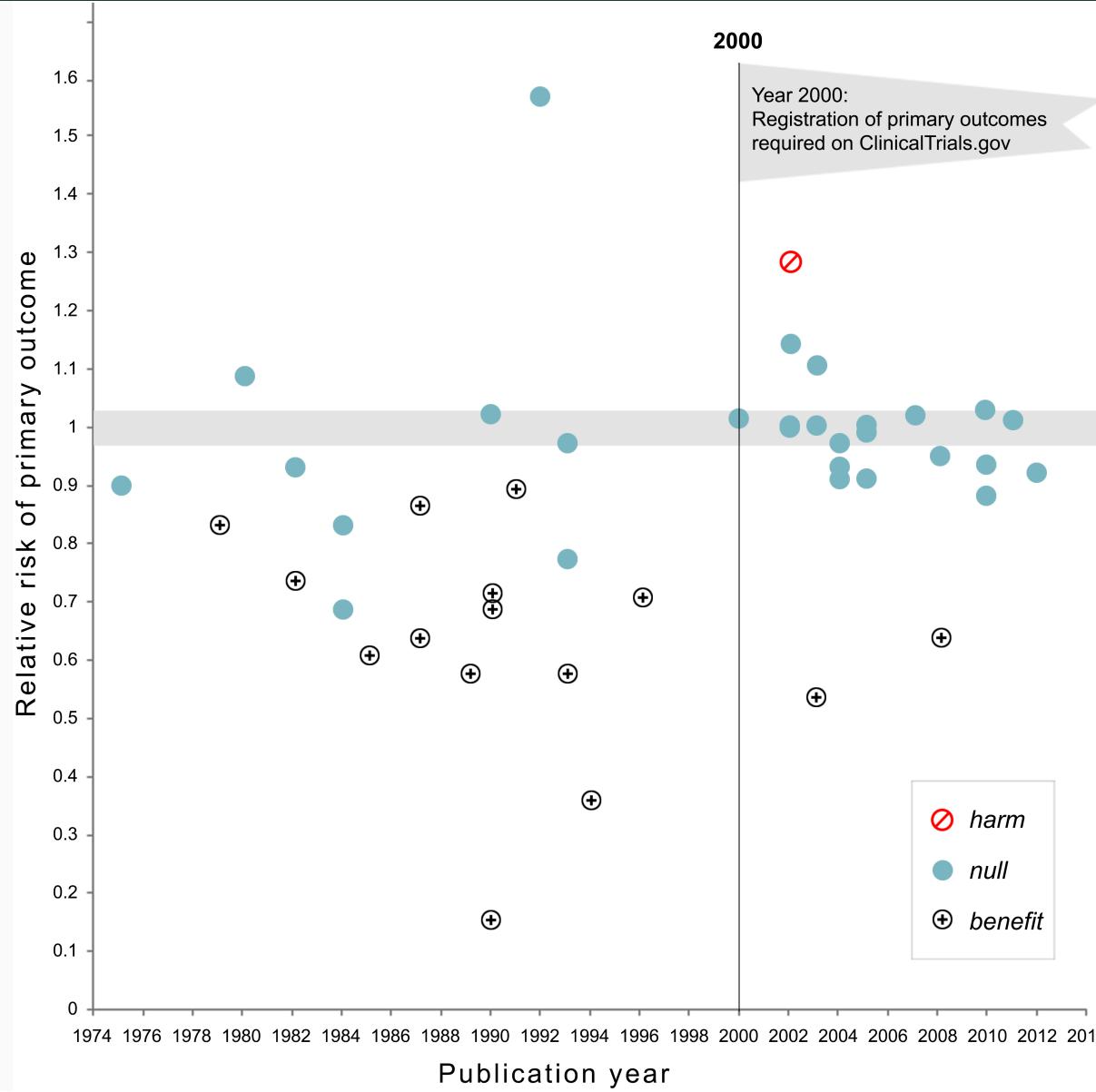
Low Replicability and Reproducibility ("Reproducibility Crisis")

Replication in Social Sciences (same method, different sample)	Reproduction in Economics (same data and methods)
OSC (2015): 30%-60%	Chang & Li (2015): 43%
Camerer et. al. (2016): ~60%	Gertler et. al. (2017): 14%
Nosek & Camerer et. al. (2018): ~60%	Kingi et. al. (2018): 43%
Klein et. al. (2018): 50%	Wood et. al. (2018): 25%

Registrations

- A registration is a record that contains minimal information about a study: title, authors, study country, status, keywords, abstract, start and end dates, outcomes, intervention information, basic research design, whether or not treatments are clustered (when performing an RCT), and Institutional Review Board (IRB) information.
- Preferably, it should be recorded before analyzing data
- **The main goal:** track the complete body of knowledge in a topic of research, regardless of the direction and magnitude of the results.

Why Register: Kaplan and Irvin (2015)



Pre-Analysis Plans

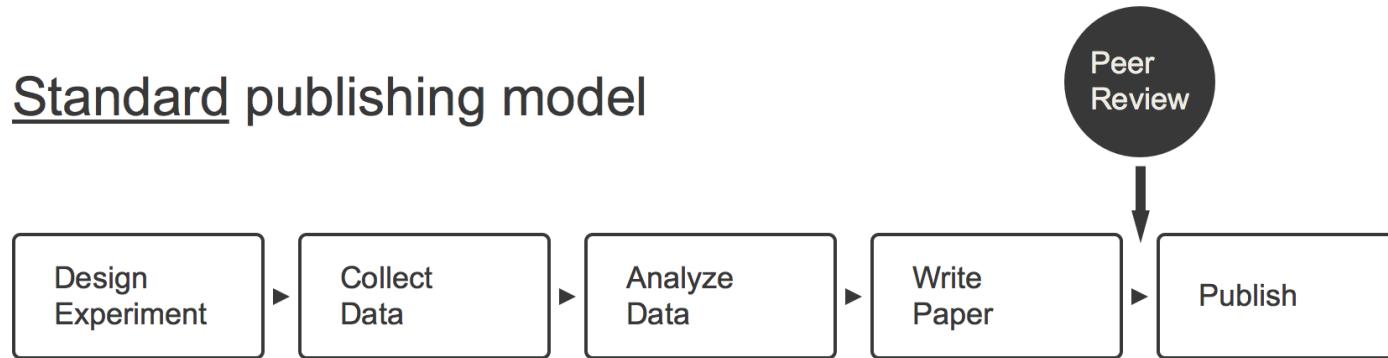
- PAPs are **extensive** methodological descriptions of the analysis to be performed before the endline data is collected
- Helps to prevent p-hacking
- Only way to guarantee accurate statistical testing
- Distinguishes confirmatory from exploratory analysis

Common Concerns About PAPs

Critique	Response
PAPs take too much time and are too difficult (Olken 2015)	A PAP changes the timing of the analytic component, not clear that it increases it
Scientific discovery often comes from surprises. PAPs stifle discovery (Olken 2015)	PAPs do not prevent researchers from doing exploratory work; they only require researchers to be clear about the objectives of their analyses (Ofosu and Posner 2020).
If replications are cheap they will rule out false positives, making PAPs irrelevant. (Coffman and Niederle 2015)	Very few experiments are inexpensive as to perform many replications. Moreover, most of the false positives have been identified where experiments are least expensive (lab experiments).

Another Reason: Register Reports

Standard publishing model



Results-blind review publishing model

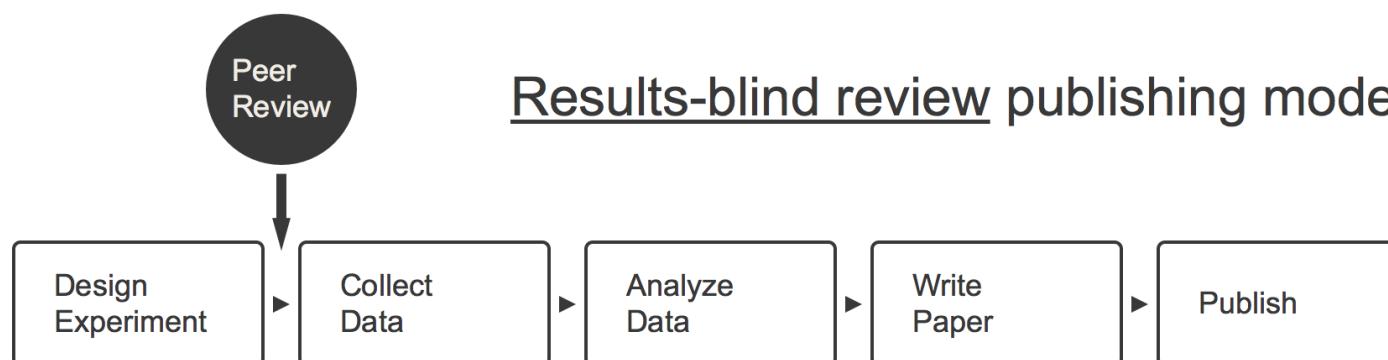


Table of Contents

1. Quality: Problems and Solutions
2. **Quantity: How research transparency can make social science more inclusive**

Historical Background

Economics Before Credibility Revolution (circa 1990, Angrist & Pieshke 2010):

- Largely driven by theory with unclear standards of how to judge empirical research
- Conjecture: opacity in standards made it more difficult for outsiders to publish

Economics After Credibility Revolution:

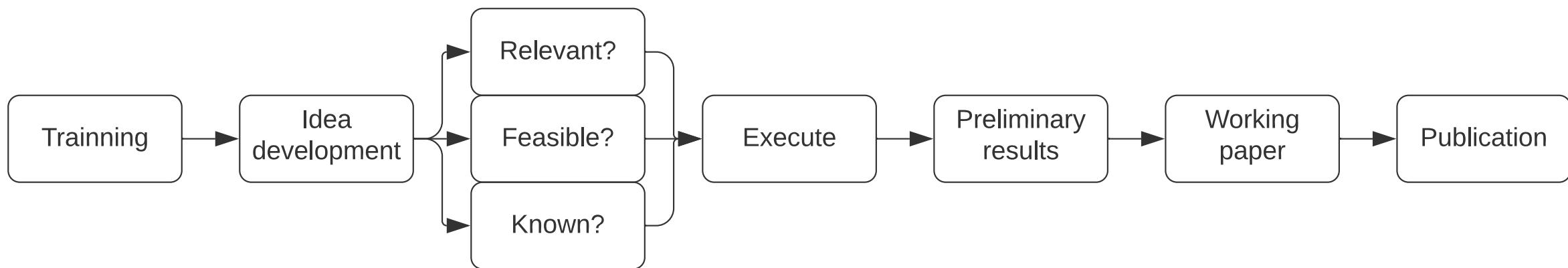
- Increase emphasis on well justified (credible) research design.
- More transparency on the definition of rigor may have open the gates to a more diverse body of researchers.

There are still several steps in the production of knowledge where more transparency can drastically increase scale of scientific production in economics. Examples of questions where opacity still favors elite universities:

- What defines an area of research as "hot"?
- What makes a paper publishable (after the analysis is completed)?
- How does a researcher obtain access to confidential/proprietary data?

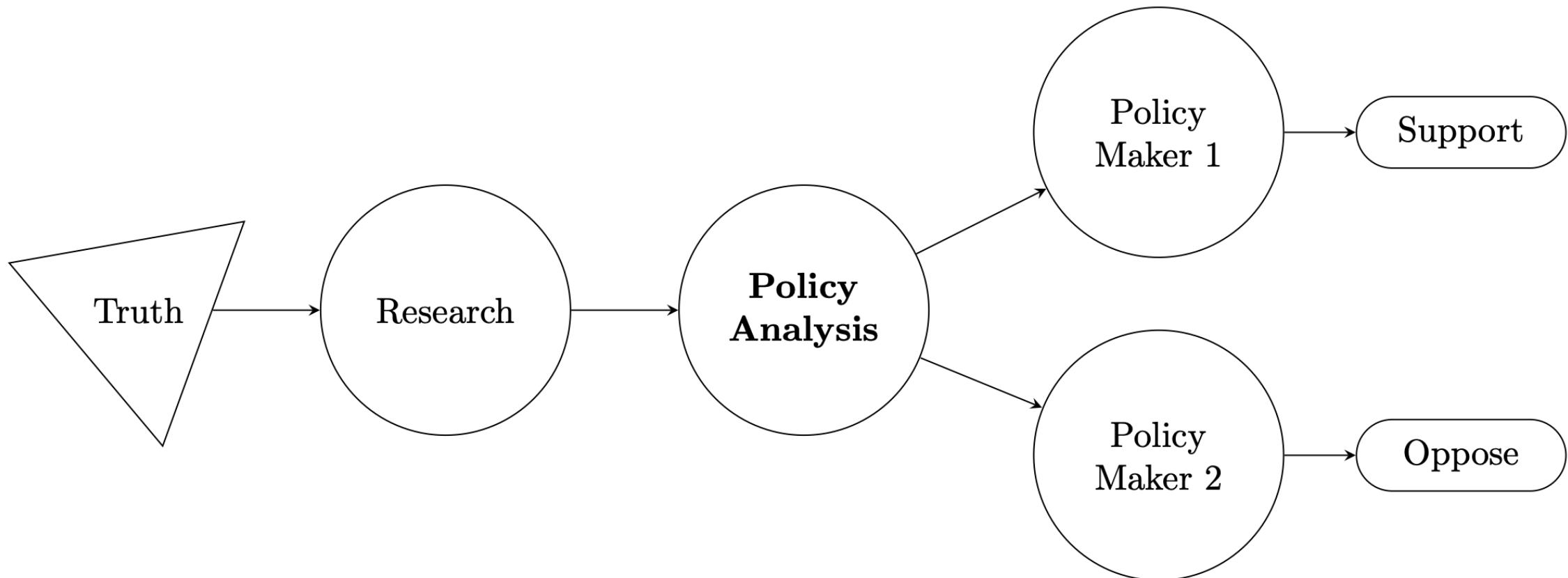
Sample Timeline for the Production of a Scientific

- If we only look at training and publications (what is observable on a CV!) we might attribute a disproportionate role to skill ("genius myth").

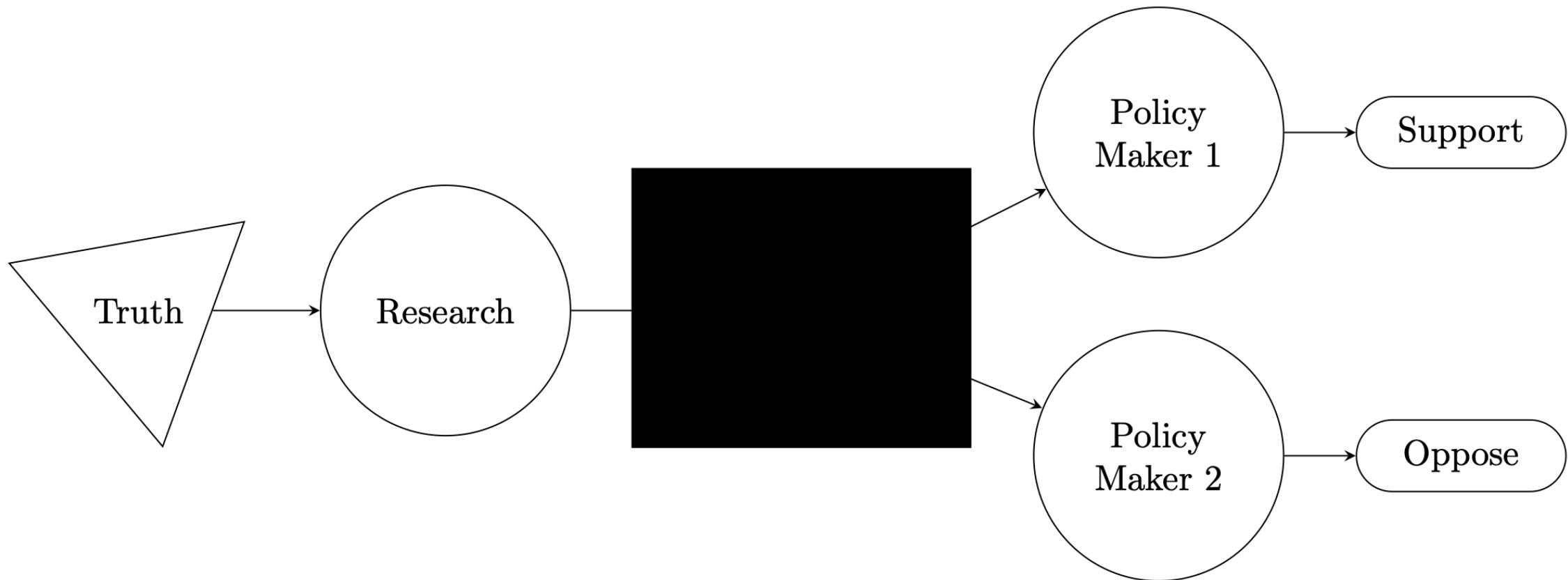


- Research transparency gives us tools to remove opacity out of every stage in this process

Why Open Policy Analysis?

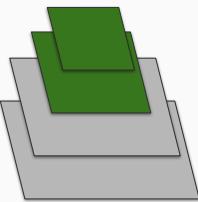


Why Open Policy Analysis?



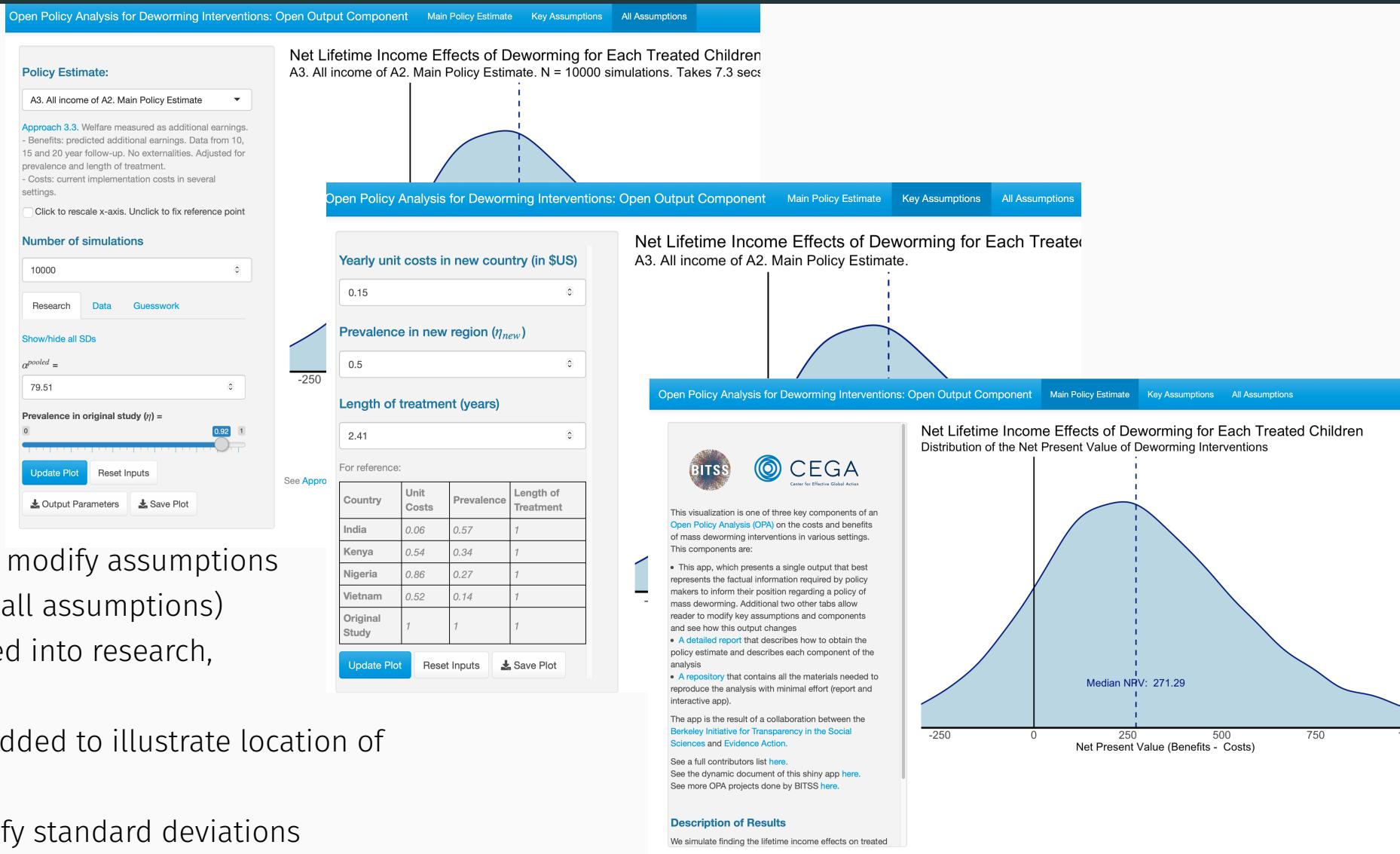
Open Output

Demo



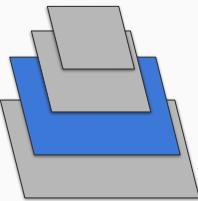
Main features

- One clear output previously agreed in consultation with policy partner
- Two additional tabs to modify assumptions (key assumptions and all assumptions)
- Each source is classified into research, data, or guesswork
- High level equations added to illustrate location of components
- Added feature to modify standard deviations
- Track values of each component



Open Analysis

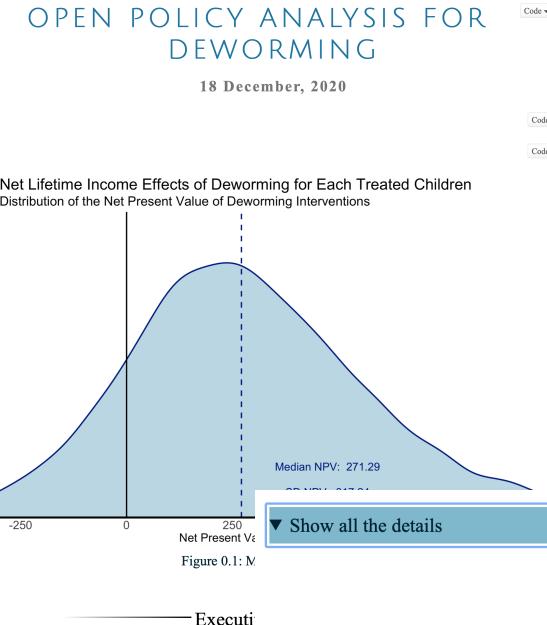
Demo



Main features

- Complete narrative description of the methodology
- Translation of each narrative step into an equation
- Implementation of each equation into code
- Combine all of the above into using a dynamic document (RMarkdown)
- Presentation of narrative, equations, and code in layered fashion to avoid overwhelming the reader
- Icon figure

BITSS CEGA
Open Policy Analysis
1 Introduction
2 Methodology
3 Main Results
References



$$B = \sum_{t=0}^{50} \left(\frac{1}{1+r} \right)^t E_t \quad (1)$$

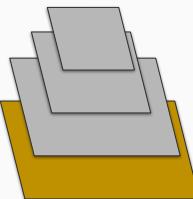
Where:

- E_t : earnings individuals are expected to generate at period t
- r : real interest rate as the discounting rate
- t : period t. Period 0 represents time of intervention. Individuals are assumed to enter the labor market 9 years after treatment.

```
# - inputs: stream earnings, discounting rate, number of periods
# - outputs: function that computes the present value of benefits
chunk_benefits <- function(){
#####
pv_benef_f <- function(
  earnings_var = earnings_in,
  interest_r_var = interest_in,
  periods_var = periods_so
) {
  index_t <- 0:periods_var
  res1 <- sum( ( 1 / (1 + interest_r_var) )^index_t * earnings_var )
  return(res1)
}
```

Open Materials

Demo



Main features

- One-click reproducible documentation and app
- Extensive readme files
- Clear folder structure
- Version controlled
- Open data
- Acknowledgment to all contributors

BITSS-OPA / [opa-deworming](#)

Code Issues Pull requests Actions Projects Wiki Security Insights Settings

master 21 branches 1 tag Go to file Add file Code

fhoces Change title of readmen file 53bb6f1 1 minute ago 728 commits

.binder update install.R 2 months ago

code Merge branch 'master' of <https://github.com/fhoces/opa-deworming> 1 hour ago

data

docs

rawdata

.gitignore

contributors.R

contributors.csv

opa-deworming.Rproj

readme.Rmd

readme.md

readme.md

Open Policy Analysis of Deworming

BITSS BERKELEY INITIATIVE FOR TRANSPARENCY IN THE SOCIAL SCIENCES

R version 4.0.0 (2020-04-24) -- "Arbor Day"
(Copyright (C) 2020 The R Foundation for Statistical Computing
Platform: x86_64-apple-darwin17.0 (64-bit)

R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.

Natural language support but running in an English locale

R is a collaborative project with many contributors.
(Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.)

Go to file Add file Code

Environment History Connections

Import Dataset Global Environment

Files Plots Packages Help Viewer

OPEN PO

BITSS CEGA

Open Policy Analysis

1 Introduction
2 Methodology
3 Main Results
References

Net Lifetime Income Effect
Distribution of the Net Present

22 / 23

