MY457/557: Causal Inference for Observational and Experimental Studies

Introduction

Daniel de Kadt

Department of Methodology
LSE

Winter Term 2025

nytimes.com/2024/02/17/well/eat/red-wine-heart-health.html

The New York Times

How Red Wine Lost Its Health Halo

For a glorious decade or two, the drink was lauded as good for the heart. What happened?

Listen to this article · 7:06 min Learn more

Share full article

1.8K

mcgill.ca/oss/article/covid-19-health/how-we-know-covid-19-vaccines-work

McGill | Office for Science and Society
Separating Sense from Nonsense

Search

Our Articles | Who We Are | Dr. Joe's Books | Media and Press | Events | Our History | Public Lectures | Past Newsletters

Home

How We Know That the COVID-19 Vaccines Work

Saying that the COVID-19 vaccines work is more than just cheerleading: there are many studies to back it up

Jonathan Jarry M.Sc. | 23 Jul 2021 COVID-19

X Post

We keep hearing that the COVID-19 vaccines work but some people have asked us, "How do we know that they work?"

Luckily, we have a growing trove of data that show that these vaccines are indeed effective against catching the virus, transmitting it to other people, and ending up in the hospital because of it.

Efficacy in clinical trials

First, there was the experimental work done as the vaccines were developed and tested in both animal models and in human participants. Large clinical trials were eventually conducted where

How the project works

Policy Experimentation is a new project launched by the Directorate-General for Research and Innovation in order to provide key results to further inform and embed policy experimentation into EU research and innovation policy-making. Lessons from it will be instrumental to guide and inform the inclusion of more experimentation actions funded by the framework programme currently and in the future.

To promote the use of policy experimentation in research and innovation policy, the project will follow a multifaceted approach

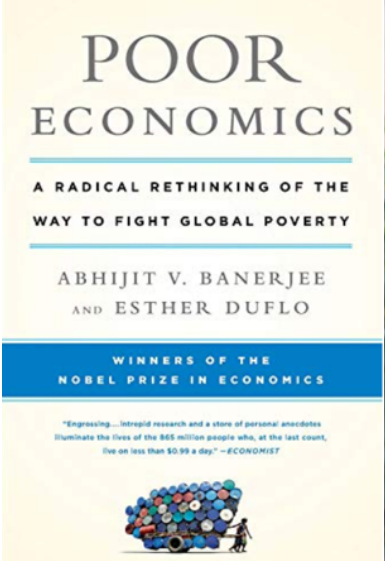
- **Lessons from experimentation for a more impactful future**, which will capitalise on evidence of enablers and barriers to experimentation.
- **Supporting policy with experimental evidence** through application of policy experimentation methods within the scope of EU Missions to shed light on how to best ensure goals implementation through citizen engagement actions. The experiments will be conducted in partnership with key stakeholders for 2 specific EU Missions:
  - 100 Climate-Neutral and Smart Cities by 2030
  - A Soil Deal for Europe: 100 living labs and lighthouses to lead the transition
- **Promotion of experimentation across Europe** by raising awareness and building capacity through information campaigns and a training module for EU Commission policy officers, EU countries and other stakeholders.

A consortium led by Nesta together with the Barcelona School of Economics (BSE) will carry out the project activities. The project will be led by two of their centres, the Innovation Growth Lab (IGL), at Nesta and BSE, and the Centre for Collective Intelligence Design (CCID) at Nesta.

What is policy experimentation?

The experimental approach allows testing new policies on a small scale, measuring the impact and learning what works best, to then scale up the most successful solutions.

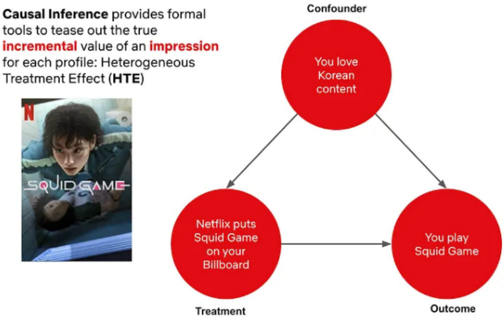
Through policy experimentation, policymakers are able to test new ideas while building credible evidence of what works, what does not work and why.


POOR
ECONOMICS
A RADICAL RETHINKING OF THE
WAY TO FIGHT GLOBAL POVERTY
ABHIJIT V. BANERJEE
AND ESTHER DUFLO
WINNERS OF THE
NOBEL PRIZE IN ECONOMICS
"Engrossing....intrepid research and a store of personal anecdotes illuminate the lives of the 865 million people who, at the last count, live on less than $0.99 a day." —ECONOMIST


오징어 게임
넷플릭스 시리즈
9월 17일 공개 | NETFLIX

netflixtechblog.com/a-survey-of-causal-inference-applications-at-netflix-b62d25175e6f

Search



Compared to machine learning, causal inference allows us to build a robust framework that controls for confounders in order to estimate the true incremental impact to members

At Netflix, many surfaces today are powered by recommendation models like the personalized rows you see on your homepage. We believe that many of these surfaces can benefit from additional algorithms that focus on making each recommendation as useful to our members as possible, beyond just identifying the title or feature someone is most likely to engage with. Adding this new model on top of existing systems can help improve recommendations to those that are right in the moment, helping find the exact title members are looking to stream now.

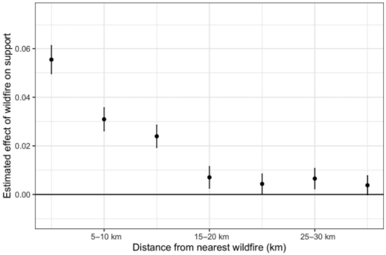
American Political Science Review (2020) 114, 4, 1359–1365
doi:10.1017/S0003055420000441

Letter
Wildfire Exposure Increases Pro-Environment Voting within Democratic but Not Republican Areas

CHAD HAZLETT *University of California, Los Angeles*
MATTO MILDENBERGER *University of California, Santa Barbara*

One political barrier to climate reforms is the temporal mismatch between short-term policy costs and long-term policy benefits. Will public support for climate reforms increase as climate-related disasters make the short-term costs of inaction more salient? Leveraging variation in the timing of Californian wildfires, we evaluate how exposure to a climate-related hazard influences political behavior rather than self-reported attitudes or behavioral intentions. We show that wildfires increased support for costly, climate-related ballot measures by 5 to 6 percentage points for those living within 5 kilometers of a recent wildfire, decaying to near zero beyond a distance of 15 kilometers. This effect is concentrated in Democratic-voting areas, and it is nearly zero in Republican-dominated areas. We conclude that experienced climate threats can enhance willingness-to-act but largely in places where voters are known to believe in climate change.

FIGURE 1. Estimated Effect of Wildfire Exposure on Pro-Environment Voting, by Distance



Note: Estimates compared with response at the median distance (35–40 kilometers). All estimates derived from a linear model with block-group and year fixed effects and controlling for Democratic vote share in Congressional elections four years prior. Error bars show 99% confidence intervals, using standard errors clustered on block group.

STRICT ID LAWS DON'T STOP VOTERS: EVIDENCE FROM A U.S. NATIONWIDE PANEL, 2008–2018*

ENRICO CANTONI AND VINCENT PONS

U.S. states increasingly require identification to vote—an ostensible attempt to deter fraud that prompts complaints of selective disenfranchisement. Using a difference-in-differences design on a panel data set with 1.6 billion observations, 2008–2018, we find that the laws have no negative effect on registration or turnout, overall or for any group defined by race, gender, age, or party affiliation. These results hold through a large number of specifications. Our most demanding specification controls for state, year, and voter fixed effects, along with state and voter time-varying controls. Based on this specification, we obtain point estimates of −0.1 percentage points for effects both on overall registration and turnout (with 95% confidence intervals of [−2.3; 2.1 percentage points] and [−3.0; 2.8 percentage points], respectively), and +1.4 percentage points for the effect on the turnout of nonwhite voters relative to whites (with a 95% confidence interval of [−0.5; 3.2 percentage points]). The lack of negative impact on voter turnout cannot be attributed to voters' reaction against the laws, measured by campaign contributions and self-reported political engagement. However, the likelihood that nonwhite voters were contacted by a campaign increases by 4.7 percentage points, suggesting that parties' mobilization might have offset modest effects of the laws on the participation of ethnic minorities. Finally, strict ID requirements have no effect on fraud, actual or perceived. Overall, our findings suggest that efforts to improve elections may be better directed at other reforms. *JEL Codes:* D72.

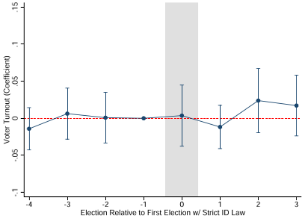
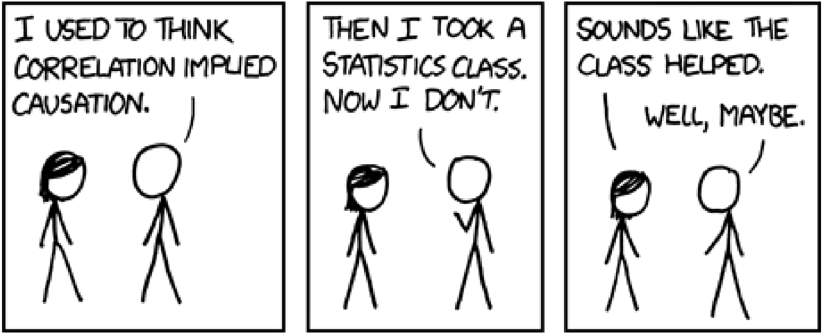


FIGURE I

Event Study Graph of the Turnout Effects of Strict ID Laws

The figure plots event study estimates and 95% confidence intervals from a regression (in the form of equation (2)) run on all registered and unregistered voters. The sample includes treated and control states. To avoid picking up variation from 2016 North Dakota, 2016 Texas, and 2018 Texas (which, unlike 2014 and 2018 North Dakota and 2014 Texas, did not enforce a strict law), we define $ID^{\tau=1}_{ND,2016} = ID^{\tau=1}_{TX,2016} = ID^{\tau=2}_{TX,2018} = 0$.

Goals of this Course



Goals of this Course

By the end of the class you should know the following about causal inference:

1. Concepts, definitions, and theoretical foundations
2. Canonical research designs and examples
3. Some of what is happening at the field's frontiers
4. Basic implementation of these designs using real data in `R`

Some limitations:

- Variables must be numerical (though not necessarily quantitative!)
- Measurement will only come up briefly, though it is very important
- 'Effects of causes' but not 'causes of effects' (*rerum cognoscere causas*)
- Minimal focus on:
  - Mechanisms (why a causal variable has the effect it does)
  - Generalization and meta-learning
  - Intersection between causal inference and machine learning
  - Programming/coding (but you will need these skills!)

Course Map

- **Week 1:** Causal frameworks
- **Week 2:** Randomization
- **Week 3:** Selection on observables I
- **Week 4:** Selection on observables II
- **Week 5:** Selection on observables III
- **Week 6:** *Reading week*
- **Week 7:** Instrumental variables I
- **Week 8:** Instrumental variables II
- **Week 9:** Regression discontinuity
- **Week 10:** Difference-in-differences I
- **Week 11:** Difference-in-differences II

Should You Take This Class?

Yes! (Probably)

Prerequisites:

1. Probability and linear regression, to the level of MY452 or equivalent.
2. Familiarity with research design, to the level of MY400 or equivalent.
3. The seminars and problem sets use `R`. Some prior experience is required, and we are not able to offer extensive support on programming as this is not a programming class. The Digital Skills Lab offers a 3 week introduction to `R` which you could take.

Auditors are welcome, so long as there is room (lectures and classes). See `https://www.lse.ac.uk/Methodology/Auditing/Auditing` to register.

Fair warning: The material is hard, and we move through it quite quickly (there is a lot to cover). But I'm confident you will succeed, and enjoy the journey.

Teaching

Ten two-hour lectures, Wed 12:00–14:00

- Taught by yours truly.
- Focus on intuition, theory, and real examples. Some practical content, but not much.

Five two-hour seminars (two groups), Thu 14:00–16:00 (Group 2) and Fri 12:00–14:00 (Group 1)

- Led jointly by me and Dr Michael Schultz, every second week, starting in week 2
- Each seminar will focus on a specific design, linked to the lecture material.
- 1 hour 15 minutes focused on discussing an applied paper you will pre-read.
- 45 minutes focused on implementing the design of interest in `R`.
- Please bring your own laptop to the seminars if possible, with `R` and RStudio/VScode installed.

See LSE timetables for room information.

Office hours for both me and Dr Schultz can be booked on studenthub.

Formative Assessment

Six Problem Sets:

- Mix of theory, simulation, and real-world data.
- Use the provided `Rmarkdown` template and submit on time through Moodle.
- You will receive feedback from our graduate teaching assistant, Pedro Torres-Lopez (PhD student in Social Policy).
- You are welcome to use Generative AIs to assist you with these problem sets, but do so responsibly and with the goal of improving your understanding of the material and coding.

While not mandatory, these problem sets are strongly encouraged and will greatly enhance your learning in this class. I recommend you do them.

Summative Assessment

MY457 (MSc): A reappraisal of a published social science paper (100%)

- Five papers will be chosen from leading journals in Economics, Political Science, and Sociology. These will be announced by the start of reading week (week 6).
- You will choose one paper for reproduction and reappraisal (sometimes called 'replication').
- You are welcome to use Generative AI to assist you, but be thoughtful, careful, and responsible in your use of these tools.
- Specific instructions will be provided before reading week, but anticipate producing a 5$\pm$ page report plus a comprehensive code appendix and github repo, due at 5pm on 23 May 2025.

MY557 (MRes/PhD): An original paper applying methods from the course to a substantive research question (100%)

- I am quite flexible about what this paper is, and it should ideally enhance and contribute to your PhD research.
- Please have a preliminary discussion with me about your proposed topic with me before reading week.
- Due at 5pm on 23 May 2025.

Online Learning Platforms

Moodle:

- Forum for questions (please use this!)
- Lecture recordings (after lectures)
- Formative homework assignments: Submission and feedback
- Summative submissions

Github (https://lse-my457.github.io/):

- Weekly reading lists (to be read before lecture)
- Lecture slides (updated ahead of the lectures)
- Seminar materials (paper to pre-read, guiding questions, example code)
- Formative homework assignments (instructions and data)
- MY457 summative assignment (instructions)

Core Books

We will use three core textbooks throughout the term:

- **MHE**: Angrist and Pischke, *Mostly Harmless Econometrics: An Empiricist's Companion*, 2009, Princeton University Press.
- **CIS**: Imbens and Rubin, *Causal Inference for Statistics, Social, and Biomedical Sciences*, 2015, Cambridge University Press.
- **TE**: Huntington-Klein, *The Effect: An Introduction to Research Design and Causality*, 2022, CRC Press.

For those very interested in the material, we also recommend:

- **CMRI**: Pearl, *Causality: Models Reasoning and Inference* (2nd Ed), 2009, Cambridge University Press.
- **CISAP**: Pearl, Glymour, and Jewell, *Causal Inference in Statistics: A Primer*, 2016, Wiley.
- **CIWI**: Hernan and Robins, *Causal Inference: What If* (2nd Ed), 2025, Routledge.

All are available as electronic copies.