Potential Outcomes and Causal Inference

Gov 51: Data Analysis and Politics

Gov 51 Lecture

Harvard University

Week 11
April 7, 2026

Two Experiments, One Lesson

April 26, 1954: 1.8 million children lined up to stop polio

The Salk Polio Vaccine Trial

- 1.8 million "Polio Pioneers" enrolled
- Randomized: vaccine vs. placebo
- Double-blind: children, parents, doctors all unaware
- Result: 71% reduction in paralytic polio

Largest field trial in history

Polio cases

Placebo group: 57 per 100k
Vaccine group: 16 per 100k

$p < 0.0001$

Lanarkshire, 1930: 20,000 children, four months, no answer

The Lanarkshire Milk Experiment

- 20,000 Scottish schoolchildren
- Question: does supplemental milk improve growth?
- Treatment: free daily milk for 4 months
- Selection: teachers chose who got milk

Teachers gave milk to the **neediest** children — thinner, malnourished, poorer

The problem

Treated group: worse baseline
Control group: healthier baseline

Comparison is contaminated

"Student" (W.S. Gosset) showed why Lanarkshire was useless

What the data showed:

- Milk group gained *less* weight than control
- Did milk hurt children? No.
- Milk group started *shorter and lighter*
- Selection of the treated group poisoned every comparison

$$E[Y_i^0 \mid D_i = 1] \neq E[Y_i^0 \mid D_i = 0]$$

The treated would have had *worse* outcomes even without milk

Design matters more than sample size

What Is a Cause?

Hume (1748): a cause is something whose absence prevents the effect

David Hume, *An Enquiry Concerning Human Understanding* (1748):

- An object followed by another such that all similar objects are followed by similar objects
- *Counterfactual form*: if the first object had not been, the second had never existed

"Had it been absent, its effects would have been absent as well"

The counterfactual is the cause

Mill (1843): change one thing, hold everything else fixed

John Stuart Mill, *A System of Logic* (1843):

- **Method of Difference**: compare two situations identical in every respect except the cause under investigation
- The only varying circumstance is the cause
- This is the logic of every controlled experiment

Unit A: treatment, outcome $= y_1$

Unit B: no treatment, outcome $= y_0$

All else identical

$\Rightarrow$ difference

$= y_1 - y_0$ is the effect

Lewis (1973): Y_i^0 is the outcome in the nearest possible world without treatment

David Lewis, *Counterfactuals* (Harvard University Press, 1973):

- Event c causes event e iff: in the nearest possible world where c doesn't occur, e doesn't occur
- **Possible worlds** formalize the counterfactual
- Causation = dependence across possible worlds

Y_i^0 = outcome in the closest counterfactual world where $D_i = 0$

The Potential Outcomes Framework

Every unit has two potential outcomes: one for each treatment state

Notation

- i indexes units (people, countries, households)
- $D_i \in \{0, 1\}$: treatment received
- Y_i^1: outcome *if* treated
- Y_i^0: outcome *if* not treated

Potential outcomes exist *before* treatment is assigned

Y_i^1: outcome if treated

Y_i^0: outcome if not treated

The individual treatment effect is the difference — and it is never observed

Individual Treatment Effect

- $\delta_i = Y_i^1 - Y_i^0$
- How much did treatment change *this person's* outcome?
- Requires observing the same person under both states
- **Impossible**: a person either receives treatment or they don't

$$\delta_i = Y_i^1 - Y_i^0$$

The Fundamental Problem of Causal Inference: we observe at most one potential outcome per unit

What we observe is determined by treatment received

$$Y_i^{\text{obs}} = D_i \cdot Y_i^1 + (1 - D_i) \cdot Y_i^0$$

If $D_i = 1$ (treated):
$Y_i^{\text{obs}} = Y_i^1$

Y_i^0 is the **counterfactual** — unobserved

If $D_i = 0$ (control):
$Y_i^{\text{obs}} = Y_i^0$

Y_i^1 is the **counterfactual** — unobserved

The counterfactual for every unit is missing data

The ATE averages the treatment effects across all units

Average Treatment Effect (ATE)

- Average δ_i across the full population
- What effect would we expect for a randomly chosen unit?
- Answerable by experiment even though δ_i is not

$$\text{ATE} = E[Y_i^1 - Y_i^0]$$

Other estimands:

- ATT $= E[\delta_i \mid D_i = 1]$ (effect on the treated)
- ATU $= E[\delta_i \mid D_i = 0]$ (effect on the untreated)

Why Naive Comparisons Fail

Do hospitals make people sicker?

People who go to hospitals are already sick — that's the selection

Naive comparison:

- Hospital patients have worse outcomes than non-patients
- Hospital mortality rate > community mortality rate
- **Conclusion?** Hospitals kill people

What's wrong:

- Hospital patients were already sicker
- Their Y_i^0 (without hospital) is already worse
- The two groups are not comparable

Treated and untreated groups differ in ways *correlated* with potential outcomes

A numerical example: job training with self-selection

Worker	Y_i^1	Y_i^0	δ_i	D_i	Y_i^{obs}
Anna	\$40K	\$30K	\$10K	1	\$40K
Bob	\$50K	\$40K	\$10K	1	\$50K
Carlos	\$20K	\$10K	\$10K	0	\$10K
Diana	\$30K	\$20K	\$10K	0	\$20K

True ATE: $\delta_i = \$10K$ for everyone

Naive SDO:
$\frac{40+50}{2} - \frac{10+20}{2} = \$45K - \$15K = \$30K$

The simple difference in outcomes decomposes into ATE plus selection bias

$$\text{SDO} = \underbrace{E[Y_i^1 - Y_i^0]}_{\text{ATE}} + \underbrace{E[Y_i^0 \mid D_i = 1] - E[Y_i^0 \mid D_i = 0]}_{\text{Selection Bias}}$$

In the example:
ATE $= \$10K$
Selection bias $= (35 - 15) = \$20K$
SDO $= 10 + 20 = \$30K$ ✓

Selection bias is the gap in baseline outcomes — how much better-off the treated were even *before* treatment

Selection bias: the treated would have differed even without treatment

What selection bias measures:

- $E[Y_i^0 \mid D_i = 1]$: what the treated *would have* earned without training
- $E[Y_i^0 \mid D_i = 0]$: what the untreated actually earned
- If these differ, the comparison is contaminated

Positive selection bias

$E[Y_i^0 \mid D_i = 1] > E[Y_i^0 \mid D_i = 0]$

SDO $>$ ATE

Negative selection bias

$E[Y_i^0 \mid D_i = 1] < E[Y_i^0 \mid D_i = 0]$

SDO $<$ ATE

Randomization as the Solution

Randomization makes potential outcomes independent of treatment

Random assignment means:

- Who gets treated is determined by a coin flip
- Not by earnings potential, health, or anything else
- Potential outcomes $\{Y_i^0, Y_i^1\}$ are *independent* of D_i

Consequence:

- $E[Y_i^0 \mid D_i = 1] = E[Y_i^0 \mid D_i = 0] = E[Y_i^0]$
- Selection bias $= 0$

$$\{Y_i^0, Y_i^1\} \perp D_i$$

"Independence": treatment assignment carries no information about potential outcomes

Under independence, the SDO is an unbiased estimator of the ATE

The algebra:

$$\begin{aligned}
\text{SDO} &= E[Y_i^1 \mid D_i = 1] - E[Y_i^0 \mid D_i = 0] \\
&= E[Y_i^1] - E[Y_i^0] \quad \text{(by independence)} \\
&= E[Y_i^1 - Y_i^0] \\
&= \text{ATE}
\end{aligned}$$

Under randomization:

$$\text{SDO} = \text{ATE}$$

The naive comparison *is* the causal effect

Randomization doesn't solve the Fundamental Problem — it averages around it

We still cannot observe $\delta_i = Y_i^1 - Y_i^0$

- Each person is still only seen in one state
- Individual treatment effects remain unidentified
- What randomization gives us: **average** effects

What randomization actually does:

- Creates two groups with identical distributions of Y^0 (in expectation)
- Any difference in outcomes is therefore due to treatment

Randomization creates *exchangeable* groups—neither group is special

From Theory to Data

Step 1: load a randomized experiment and check balance

```r
library(tidyverse)

# Load the resume audit study (Bertrand & Mullainathan 2004)
resume <- read_csv("resume.csv")

# Check covariate balance: sex composition by treatment
resume |>
  group_by(race) |>
  summarize(
    pct_female = mean(sex == "female"),
    n          = n()
  )
```

If randomization worked, `pct_female` should be similar across groups

Step 2: estimate the ATE as a difference in means

```r
# Difference in callback rates by race
resume |>
  group_by(race) |>
  summarize(callback_rate = mean(call))

# Black-sounding names:  3.8%
# White-sounding names:  9.7%

# ATE estimate:
9.7 - 3.8    # 5.9 percentage points
```

$$\widehat{\text{ATE}} = \bar{Y}^{\text{obs}}_{D=1} - \bar{Y}^{\text{obs}}_{D=0} = 0.097 - 0.038 = 0.059$$

Step 3: OLS with a binary treatment gives the same estimate

```r
# Create binary treatment: 1 = white-sounding name
resume <- resume |>
  mutate(white = as.integer(race == "white"))

# OLS regression
fit <- lm(call ~ white, data = resume)
coef(fit)
## (Intercept)       white
##  0.03448276  0.03203285

# Intercept = E[Y | white=0] = 3.4%
# Coefficient = difference in means = 3.2 pp
```

The OLS coefficient on a binary treatment equals the difference in means

$$\hat{\beta}_1 \quad = \quad \bar{Y}_{D=1} \; - \; \bar{Y}_{D=0}$$

Why this is true:

- With one binary regressor, OLS minimizes SSR by fitting two group means
- Slope = difference in group means
- Intercept = mean of the $D = 0$ group

OLS is not a special method for RCTs — it is the natural estimator. Adding covariates increases precision, never changes what it estimates under randomization.

Where Randomization Began

Fisher arrived at Rothamsted in 1919 to analyze 70 years of wheat and barley data

Rothamsted Experimental Station (Hertfordshire, UK)

- Founded 1843 — world's oldest agricultural research station
- Continuous crop trials since 1843 (Broadbalk wheat plot, still running)
- Fisher hired 1919 to make sense of decades of messy field data
- Problem: plots differ in soil, drainage, sunlight

Question: which fertilizer actually works?

Fisher's insight

Randomize which plots get which treatment

$\Rightarrow$ confounding averages out

$\Rightarrow$ inference is valid

Neyman's 1923 dissertation: potential outcomes to compare crop varieties

Jerzy Neyman (1923), written in Polish: *"On the Application of Probability Theory to Agricultural Experiments"*

- Formal framework for comparing crop varieties on randomized plots
- Introduced the notation $Y_i(1)$, $Y_i(0)$ for potential yields
- Derived unbiasedness of the difference in means estimator
- Published in English *67 years later* in 1990

Both Fisher and Neyman were answering one question:

Which seed grows more food?

The Method That Fed the World

Fisher's randomization methods spread from Rothamsted to every field station on Earth

The institutional chain:

1. Fisher develops randomized block designs at Rothamsted (1919–1933)
2. *Statistical Methods for Research Workers* (1925) spreads methods globally
3. Rockefeller Foundation funds Borlaug in Mexico (1944)
4. CIMMYT, IRRI run systematized randomized field trials
5. High-yield varieties tested, selected, deployed at scale

CIMMYT's breeding trials today still use alpha-lattice designs replicated two to three times — Fisher's principles, unbroken

Borlaug's varieties doubled wheat yields — twice, in two growing seasons

Norman Borlaug, Nobel Peace Prize 1970

- Semi-dwarf wheat varieties: shorter stem, more grain
- Tested through randomized variety trials across Mexico, India, Pakistan
- "Shuttle breeding": two growing seasons per year in different climates

Country	1965	1970
India	12M tonnes	20M tonnes
Pakistan	4.6M tonnes	8.4M tonnes

India doubled wheat output in two growing seasons (1966–68)

In 1965, India was importing millions of tonnes of grain. By 1974, it was self-sufficient.

World grain output rose 160% between 1950 and 1984

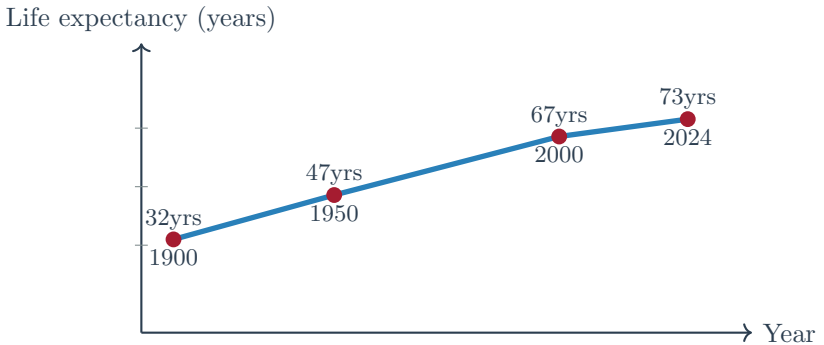
The Green Revolution in numbers:

- Global grain output +160% (1950–1984)
- Without it: global caloric availability would be 11–13% lower today
- 32–42 million preschool children with improved nutritional status

Borlaug's Nobel lecture (1970):
"We can't build world peace on empty stomachs and human misery"

Estimated lives saved
Gregg Easterbrook (1997):
∼1 billion
from Borlaug's methods alone

Life expectancy: 32 years in 1900, 73 years today



Key fact (Our World in Data): Today's global average exceeds what *any country* achieved in 1950

Extreme poverty: 85% of humanity in 1800, 10% today

Share of humanity in extreme poverty:

Year	Share (%)
1800	≈ 85
1910	≈ 66
1950	≈ 55
2000	≈ 29
2025	≈ 10

World population grew $8\times$ since 1800.

The poverty share fell $8\times$.

The absolute number of poor people shrank.

Child mortality: 1 in 3 died before age 5 in 1900, 1 in 27 today

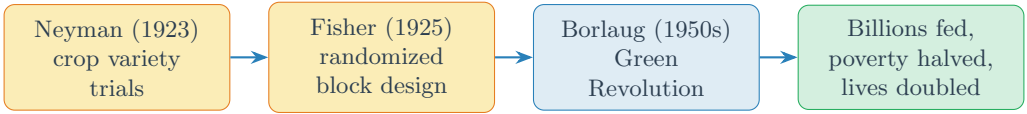
Under-5 mortality rate:

Era	Rate	Framing
Pre-modern	400–500/1k	1 in 2
1900	≈350/1k	1 in 3
1950	≈225/1k	1 in 4
1990	93/1k	1 in 11
2023	37/1k	1 in 27

Under-5 mortality fell **59%** between 1990 and 2023 alone

Source: UN IGME / Our World in Data

The logic of randomization traces a line from crop fields to a billion lives



Both founders of statistical randomization were trying to answer the same question: *which seed grows more food?*

Randomization, born to answer which seed grows more food, may be the statistical idea that has done the most for humanity