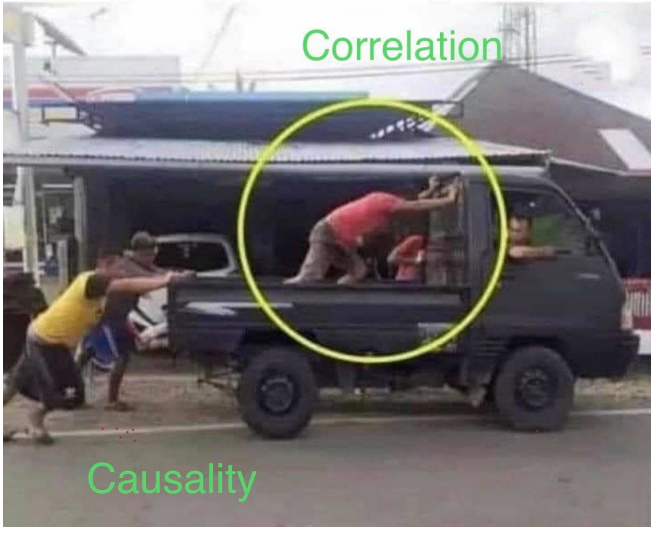
Part A: Regression and causality

A2: Potential outcomes and RCTs

Kirill Borusyak

ARE 213 Applied Econometrics

UC Berkeley, Fall 2025


Correlation
Causality

Outline

1. The basics: Potential outcomes, average causal effects, and RCTs
2. Other estimands
3. Conceptual issues with the potential outcome framework. Alternative frameworks
4. Complex RCTs
5. (Skipped) Causality or prediction?

(Neyman-)Rubin causal model

- Consider some population of units i
- Each unit is observed one of several treatment conditions $D_i \in \mathfrak{D}$
  - E.g. $\mathfrak{D} \in \{0, 1\}$: untreated and treated
- Suppose we can imagine each unit under all possible conditions (in the same period)
  - Causality always requires specifying alternatives
  - Corresponding **potential outcomes** are $\{Y_i(d) \colon d \in \mathfrak{D}\}$
    - e.g. $(Y_i(0), Y_i(1))$ (equivalently written as (Y_{0i}, Y_{1i}))
    - e.g. demand function $Q_i(p)$
  - **Causal effects** $Y_i(d') - Y_i(d)$ are defined by this abstraction
  - Writing $Y_i(d)$ encodes a possibility that D_i impacts Y_i

Rubin causal model (2)

- Unobserved vars: $Y_i(0), Y_i(1)$ (or $\{Y_i(d) : d \in \mathfrak{D}\}$ in general)
- Observed vars: **treatment status** D_i; **realized outcome** Y_i
- Restriction:

$$Y_i = \begin{cases} Y_i(0) & \text{if } D_i = 0 \\ Y_i(1) & \text{if } D_i = 1 \end{cases}$$

  - Can write it as $Y_i = Y_i(0) + (Y_i(1) - Y_i(0)) \cdot D_i \equiv Y_i(0) + \tau_i D_i$ with

$$\tau_i = Y_i(1) - Y_i(0) \qquad \text{(causal effect)}$$

  - Or as $Y_i = Y_i(D_i)$ in general
  - (More restrictions will come from experimental design, theory/contextual knowledge, or out of the blue)

Fundamental problem of causal inference

- We cannot learn the causal effect $Y_i(1) - Y_i(0)$ for any particular unit without strong assumptions
  - "Fundamental problem of causal inference": multiple potential outcomes are never observed at once

i	D	Y	$Y(0)$	$Y(1)$	$Y(1) - Y(0)$
1	1	Y_1	?	Y_1	?
2	0	Y_2	Y_2	?	?
3	0	Y_3	Y_3	?	?
...					
N	1	Y_N	?	Y_N	?

Good news

... But we can sometimes learn some averages

- E.g. **Average structural function** $d \mapsto \mathbb{E}[Y_i(d)]$
  - $\mathbb{E}[Y_i(0)]$: what if nobody was treated
  - $\mathbb{E}[Y_i(1)]$: what if everyone was treated
- Causal inference can be understood as a missing data problem:
  - $Y_i(1)$ is observed for treated units $\Rightarrow$ we know $\mathbb{E}[Y_i(1) \mid D_i = 1] = \mathbb{E}[Y_i \mid D_i = 1]$
  - ... but not for untreated units. We need to impute $\mathbb{E}[Y_i(1) \mid D_i = 0]$

Weakly causal estimands

- Averages of causal effects: $\mathbb{E}\left[\omega_i \cdot \left(Y_i(1) - Y_i(0)\right)\right] / \mathbb{E}\left[\omega_i\right]$ for observed or unobserved weights ω_i
  - A **weakly causal** estimand when $\omega_i \geq 0$ for all i (see Blandhol et al. (2025) for non-binary D_i)
  - This precludes **sign reversals**: when $Y_i(1) \geq Y_i(0)$ for all i but estimand < 0
  - (See pset for an estimator that can produce a non-convex average of effects)

Most common causal estimands

- Average treatment/causal effect: $ATE = \mathbb{E}[Y_i(1) - Y_i(0)] = \mathbb{E}[Y_i(1)] - \mathbb{E}[Y_i(0)]$
  - The effect of treating everyone relative to not treating anyone
  - ATE is the same regardless of how D_i is actually determined
- Average effect on the treated: $ATT = \mathbb{E}[Y_i(1) - Y_i(0) \mid D_i = 1]$, a.k.a. TOT, TT
  - ATT depends on how selection into treatment happened (e.g. RCT vs self-selection)
  - The effect of treating those who were treated relative to not treating anyone:

$$\mathbb{E}[Y_i - Y_i(0)] = \mathbb{E}[(Y_i(1) - Y_i(0)) \cdot D_i] = ATT \cdot P(D_i = 1)$$

- Average effect on the untreated: $ATU = \mathbb{E}[Y_i(1) - Y_i(0) \mid D_i = 0]$

When is "correlation" causal?

All these estimands follow from the distribution of $(Y(1), Y(0), D)$. But are they identified from observed (Y, D)?

$$\begin{aligned}
\beta_{OLS} &= \mathbb{E}[Y \mid D=1] - \mathbb{E}[Y \mid D=0] && \text{(Difference in means)} \\
&= \mathbb{E}[Y_1 \mid D=1] - \mathbb{E}[Y_0 \mid D=0] \\
&= \mathbb{E}[Y_1 - Y_0 \mid D=1] && \text{(ATT)} \\
&\quad + (\mathbb{E}[Y_0 \mid D=1] - \mathbb{E}[Y_0 \mid D=0]) && \text{(Selection bias)}
\end{aligned}$$

- Selection bias $= 0$ and $\beta_{OLS} = ATT$ iff Y_0 is mean-independent of D:

$$\mathbb{E}[Y_0 \mid D=1] = \mathbb{E}[Y_0 \mid D=0] = \mathbb{E}[Y_0]$$

- $ATE = ATT$ iff $(Y_1 - Y_0)$ is mean-independent of D:

$$\mathbb{E}[Y_1 - Y_0 \mid D=1] = \mathbb{E}[Y_1 - Y_0 \mid D=0] = \mathbb{E}[Y_1 - Y_0]$$

What an RCT does

- In a randomized control trial (**RCT**), $(Y_0, Y_1) \perp\!\!\!\perp D$ by **design**
  - Both mean-independence assumptions follow
- Consider two causal diagrams ("directed acyclic graphs", DAGs):

- RCT: $Y = y(D, \xi_Y)$ for $\xi_Y \perp\!\!\!\perp D$
  - Potential outcomes $Y(d) \equiv y(d, \xi_Y) \perp\!\!\!\perp D$
- Observational data: $D = d(U, \xi_D)$, $Y = y(D, U, \xi_Y)$
  - Potential outcomes $Y(d) \equiv y(d, U, \xi_Y) \not\perp\!\!\!\perp D$ because of **confounders** U

External vs. internal validity

- If you run an RCT on a selected subpopulation P, e.g. those who signed up to a clinical trial, you'd get
$$\mathbb{E}\left[Y_1 - Y_0 \mid P\right]$$
- This may not equal to $ATE = \mathbb{E}\left[Y_1 - Y_0\right]$ in the full population — **external validity** problem

- But selection bias in observational studies is worse

  - Even if this pill has no effect on anyone, can have $\mathbb{E}\left[Y_0 \mid D = 1\right] - \mathbb{E}\left[Y_0 \mid D = 0\right] \neq 0$ — **internal validity** problem

Connecting to linear models

- With a binary treatment,

$$Y_i = Y_{0i} + (Y_{1i} - Y_{0i})\, D_i = \beta_0 + \tau_i D_i + \varepsilon_i$$

  where $\beta_0 = \mathbb{E}[Y_0]$, $\tau_i = Y_{1i} - Y_{0i}$ and $\varepsilon_i = Y_{0i} - \mathbb{E}[Y_0]$
- With homogeneous effects, $Y_i = \beta_0 + \beta_1 D_i + \varepsilon_i$ becomes a causal *model* where $Y_{1i} - Y_{0i} \equiv \beta_1$
  - Regardless of whether $\varepsilon_i \perp\!\!\!\perp D_i$ (think IV). But if $D_i \perp\!\!\!\perp Y_{0i}$, OLS yields β_1
- With heterogeneous effects, can rederive that OLS identifies ATE in RCTs: if $(\varepsilon_i, \tau_i) \perp\!\!\!\perp D_i$ and denoting $\mu = \mathbb{E}[D_i]$,

$$\beta_{OLS} = \frac{\mathbb{E}[(D_i - \mu)\, Y_i]}{\mathrm{Var}[D_i]} = \frac{\mathrm{Cov}[D_i, \varepsilon_i]}{\mathrm{Var}[D_i]} + \frac{\mathbb{E}[D_i (D_i - \mu_i)\, \beta_{1i}]}{\mathrm{Var}[D_i]} = \mathbb{E}[\beta_i] \equiv ATE$$

RCT with ordered or continuous treatments

Consider a RCT where D takes more than two values (e.g. different dosages)

- $D \perp\!\!\!\perp \{Y(d)\}_{d \in \mathfrak{D}} \Longrightarrow \mathbb{E}[Y \mid D = d] = \mathbb{E}[Y(d) \mid D = d] = \mathbb{E}[Y(d)]$
- A saturated regression of Y on dummies for all values of D (or a nonparametric regression with continuous D) traces the average structural function $\mathbb{E}[Y(d)]$
- A simple regression of Y on D identifies a convexly-weighted average of $\partial \mathbb{E}[Y(d)] / \partial d$ (or its discrete version):

$$\beta_{OLS} = \int_{-\infty}^{\infty} \omega(\tilde{d}) \frac{\partial \mathbb{E}\left[Y(\tilde{d})\right]}{\partial \tilde{d}} \mathrm{d}\tilde{d}, \qquad \omega(\tilde{d}) = \frac{\mathrm{Cov}\left[\mathbf{1}\left[D \geq \tilde{d}\right], D\right]}{\mathrm{Var}[D]}$$

$$\text{or } \beta_{OLS} = \sum_{k=1}^{K} \omega_k \frac{\mathbb{E}[Y(d_k) - Y(d_{k-1})]}{d_k - d_{k-1}}, \quad \omega_k = \frac{(d_k - d_{k-1}) \mathrm{Cov}[\mathbf{1}[D \geq d_k], D]}{\mathrm{Var}[D]}$$

Outline

1. The basics: Potential outcomes, average causal effects, and RCTs
2. Other estimands
3. Conceptual issues with the potential outcome framework. Alternative frameworks
4. Complex RCTs
5. (Skipped) Causality or prediction?

Distribution of gains

Some other interesting parameters (cf. Heckman et al. (1999)):

1. How widely are the gains distributed?
    a. The proportion of people taking the program who benefit from it: $P(Y_1 > Y_0)$
    b. Median gains: $Med(Y_1 - Y_0)$
2. Does the program help the lower tail?
    a. Distribution of gains by untreated value: e.g. $\mathbb{E}[Y_1 - Y_0 \mid Y_0 = \bar{y}]$
    b. Increase in % above a threshold due to a policy: $P(Y_1 > \bar{y}) - P(Y_0 > \bar{y})$

Identification

Does an RCT identify these other parameters, e.g. median gains?

- Not without extra restrictions!
- E.g. imagine an RCT where Y takes values 0, 1, 2 with equal prob. in both treated and control groups
- This is consistent with (Y_0, Y_1) taking values $(0,0), (1,1), (2,2)$ with equal prob.
  - No casual effect for anyone. Median gain $= 0$
- Or with (Y_0, Y_1) taking values $(0,1), (1,2), (2,0)$ with equal prob.
  - Median gain $= 1$

Exception: $P(Y_1 > \bar{y}) - P(Y_0 > \bar{y})$ is identified — *how?*

Other Parameters: Individual causal effects

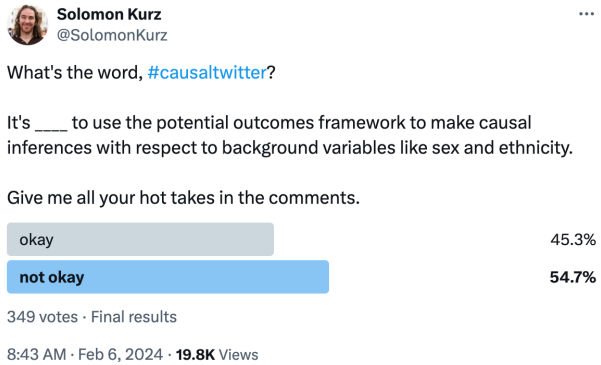
- What if we want to identify $Y_i(0)$, $Y_i(1)$, and τ_i for each i?
  - Possible under very strong effect homogeneity restrictions
  - E.g. if $\tau_i \equiv \beta$, we can learn β from an RCT (or relax to $\tau_i = \beta(X_i)$)
  - Then $Y_i(0) = Y_i - \tau_i$ for treated, $Y_i(1) = Y_i + \tau_i$ for untreated
- International trade: the effect of NAFTA on wages in Mexico
- Industrial organization: the effect of a merger between Northwest and Republic Airlines on fares from Minneapolis
- Berry and Haile (2021): *"Suppose a researcher is able to randomly assign price vectors. It may be natural to imagine that identification of demand would be trivial in this case. It is not. The observed variation in quantities will reveal certain averages of demand responses. But such averages are of very limited value, as they do not reveal any elasticity of demand — not at the observed prices or any other known point"* (abridged)

Outline

1. The basics: Potential outcomes, average causal effects, and RCTs
2. Other estimands
3. Conceptual issues with the potential outcome framework. Alternative frameworks
4. Complex RCTs
5. (Skipped) Causality or prediction?

What can be a cause/treatment?

Is it meaningful to say *"She did not get this position because she is a woman"* (example from Imbens (2020))?



Attributes as treatments

Imagining each unit under all possible conditions is non-trivial:

> *"No causation without* [imagining] *manipulation"* (Holland & Rubin)

- *"She did not get this position because she is a woman"* ✗
  - Gender is an **attribute**, not a cause; same for race
- *"She got an orchestra job because of a gender-blind audition"* ✓
  (cf. Goldin and Rouse (2000))
  - Resume studies (e.g., Bertrand and Mullainathan (2004)) similarly manipulate *perceived* gender

Poorly defined treatments

"Attending college increases average earnings by $X" — what does $Y_i(1)$ mean?

- How could you force someone to attend college?
  - Not realistic $\Rightarrow$ less useful or even (in the strict sense) ill-defined
  - Meaningful causal effects are tightly linked to policy relevance
- And which college?
  - A matter of defining $Y_i(1)$, not heterogeneous effects
  - *"Attending a 4-year public university"* or *"Attending college of the person's choice"* would be better

SUTVA (1): Poorly defined treatments

- This is a violation of **SUTVA** ("stable unit treatment value assumption")
  - D should summarize everything outcome-relevant about the intervention
  - Defining treatment variables is imposing a causal model. Don't take it lightly!
- Note: Perhaps it's easier to imagine forcing people out of the college they attended
  - So $Y_i(0)$ and ATT are more well-defined
- Note 2: current practice seems more permissive
  - E.g. Orefice et al. (2025) study causal effects of migrant diversity, measured as Herfindahl index of migrant origins
  - This measure can't be manipulated without changing anything else

SUTVA (2): Interference

- More common meaning of SUTVA: no unmodeled **interference**
  - I.e., treatment statuses of other units d_{-i} do not affect Y_i
  - Frequently violated: e.g. vaccines and infectious disease; information and technology adoption; equilibrium effects via prices
- Allowing for interference, we could start from a different model: $Y_i(d_1, \ldots, d_N)$ for the population of size N
  - We may be interested in own-treatment effects $Y_i(d_i', d_{-i}) - Y_i(d_i, d_{-i})$ and various spillover effects, e.g. $Y_i(d_i, 1, \ldots, 1) - Y_i(d_i, 0, \ldots, 0)$
  - No interference is an exclusion restriction: $Y_i(d_i, d_{-i}) = Y_i(d_i, d_{-i}') \equiv Y_i(d_i)$, $\forall d_i, d_{-i}, d_{-i}'$
- Intermediate model: $Y_i(\overrightarrow{d}_i)$ for **exposure mapping** $\overrightarrow{d}_i = (d_i, \sum_{k \in \text{Friends}(i)} d_k)$

Effects of causes vs. causes of effects

Statistical analysis focuses on effects of causes (treatments) rather than causes of effects (outcomes)

- Causes are not clearly defined

> For example, do bacteria cause disease? Well, yes . . . until we dig deeper and find that it is the toxins the bacteria produce that really cause the disease; and this is really not it either. Certain chemical reactions are the real causes . . . and so on, ad infinitum.
>
> Holland (1986, p.959)

- But *asking* questions about causes is helpful: observing an anomaly in the data motivates us to ask statistical questions about effects (Gelman and Imbens, 2013)

Criticisms by Heckman and Vytlacil (2007)

1. Estimated effects cannot be transferred to new environments (limited external validity) and to new programs never previously implemented
   - Interventions are black boxes, with little attempt to unbundle their components
   - Mechanisms are not possible to pin down
   - Knowledge does not cumulate across studies (contrast with estimates of a labor supply elasticity — a structural parameter)
     - ⋆ Counterpoint from Angrist and Pischke (2010): *"Empirical evidence on any given causal effect is always local, derived from a particular time, place, and research design. Invocation of a superficially general structural framework does not make the underlying variation more representative. Economic theory often suggests general principles, but extrapolation to new settings is always speculative. A constructive response to the specificity of a given research design is to look for more evidence, so that a more general picture begins to emerge."*

Criticisms by Heckman and Vytlacil (2007) (cont.)

2. Estimands need not be relevant even to analyze the observed policy
   - Informative on whether to throw out the program entirely (ATT) and whether to extend it forcing it on everyone not covered yet (ATU)
   - But not whether to extend/shrink it on the margin (Heckman et al. (1999), Sec.3.4)
   - Or a policy change that affects the assignment mechanism, e.g. available options
   - No analysis from the social planner's point of view, e.g. accounting for externalities
   - No analysis of causal parameters other than means, e.g. median gains

Optional exercise: read Heckman-Vytlacil's Sec. 4.4. Do you agree with everything?

Roy model

Alternative "structural" approach: to model self-selection explicitly

- Original Roy (1951) model: self-selection based on outcome comparison
  - D = choice of occupation (e.g. agriculture vs not) or education level
  - $Y(d)$ = earnings for a given occupation/education
  - People vary by occupational productivities/returns to education, known to them
  - They choose based on them: $D = \arg\max_{d \in \mathfrak{D}} Y_i(d)$, perhaps with homogeneous costs $C(d)$
- Extended Roy model: costs are heterogeneous but fully determined by observables
  - which may or may not affect the outcome at the same time
- Generalized Roy model: self-selection based on unobserved preferences
  - $D = \arg\max_{d \in \mathfrak{D}} R_i(d)$ where e.g. $R_i(d) = Y_i(d) - C_i(d)$ for costs $C_i(d)$

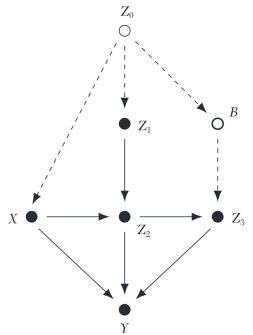
Roy model: Identification

What does this structure buy us?

- No free lunch: *"for general skill distributions* [i.e., without parametric restrictions], *the* [original Roy] *model is not identified* [from a single cross-section] *and has no empirical content"* (Heckman and Honore, 1990)
- But with more data and restrictions can identify the ATE and even the distribution of $(Y_0, Y_1, R_1 - R_0) \Longrightarrow$ distribution of gains
- Assumptions are often parametric: e.g. Heckman correction via normality of potential outcomes
  - Not living up to the goal of using economic theory for identification?
- Can do better with cost shifters that shift selection but not outcomes
  - Value over traditional IV methods is not so clear?

Another alternative: Directed acyclic graphs (DAGs)

Directed acyclic graphs of Judea Pearl represent causal relationships graphically: e.g.



X = soil treatment (fumigation)
Y = crop yield
Z_1 = eelworm population before the treatment
Z_2 = eelworm population after the treatment
Z_3 = eelworm population at the end of season

Z_0 = eelworm population last season (unobserved)
B = bird population (unobserved)

- "Do-calculus" allows to verify whether the average total effect of X on Y is identified from observing (X, Y, Z_1, Z_2, Z_3)
- Popular in epidemiology but not in economics. Why?

Some limitations of DAGs

Imbens (2020) lists some pitfalls of DAGs relative to potential outcomes:

1. Economists avoid complex models with many variables
2. Randomization and manipulability have no special value in DAGs
3. Too nonparametric:
    a. Not possible to incorporate additional assumptions, such as continuity (important in regression discontinuity designs)
    b. Too much focus on identification, relative to estimation and inference
4. Difficult to model interference
5. Clunky to model simultaneity, e.g. demand and supply

Outline

1. The basics: Potential outcomes, average causal effects, and RCTs
2. Other estimands
3. Conceptual issues with the potential outcome framework. Alternative frameworks
4. Complex RCTs
5. (Skipped) Causality or prediction?

What's special about RCTs

Simplest RCT with a binary treatment:

1. A large population of subjects
2. Treatment status is assigned by the researcher
3. Treatment status is assigned according to a known protocol
4. The protocol involves randomization
5. All units have the same probability of being assigned to the treatment group
6. Treatment status is assigned independently across units

Which of these is most important? Which can be relaxed?

Complex RCTs

Abdulkadiroglu et al. (2017) study the effects of charter school attendance D_i on grades Y_i in Denver (see also Borusyak and Hull (2021), Appx D.6)

- Centralized assignment mechanism run by the city
- Families submit preferences, they are processed by some algorithm that gives offers
- The algorithm involves lotteries to tie breaks for oversubscribed schools
- Lotteries can have knock on effects on other kids

Difference-in-means is no longer causal:

- Treatment and control groups are not balanced on geography and other inputs

Identifying causal effects in complex RCTs

- Suppose the researcher has access to the algorithm and its inputs
  - No self-selection is possible here
- Running the algorithm many times, can measure the **propensity score** of each kid:

$$p_i \equiv P(D_i = 1 \mid \text{preferences of all families}), \qquad \text{separately across } i$$

  - E.g. lower p_i in more competitive areas
  - Assume there is enough randomness so $0 < p_i < 1$
- Adjusting for p_i as a covariate is feasible and sufficient to get ATE and other causal estimands
  - Regression adjustment or reweighting *(coming soon)*
- Note: RCT not run by the researcher; p_i varies across i; D_i can be interdependent
  - Interdependence of D_i affects efficiency and SE but doesn't generate bias

What's special about RCTs

Simplest RCT with a binary treatment:

1. A large population of subjects
2. ~~Treatment status is assigned by the researcher~~
3. Treatment status is assigned according to a known protocol
4. The protocol involves randomization
5. ~~Unit are assigned to the treatment group with equal probabilities~~
6. ~~Treatment status is assigned independently across units~~

RCT with one subject

Can we learn something about causal effects if we have $N = 1$? Yes!

- Randomize treatment with $P(D_i = 1) = p$
- Estimate $\widehat{Y_i(1)} = \begin{cases} Y_i/p, & D_i = 1 \\ 0, & D_i = 0 \end{cases}$
- This estimate is unbiased:

$$\begin{aligned} \mathbb{E}\left[\widehat{Y_i(1)}\right] &= \mathbb{E}\left[\widehat{Y_i(1)} \mid D_i = 1\right] \cdot P(D_i = 1) + \mathbb{E}\left[\widehat{Y_i(1)} \mid D_i = 0\right] \cdot P(D_i = 0) \\ &= \frac{Y_i(1)}{p} \cdot p + 0 = Y_i(1) \end{aligned}$$

RCT with one subject (2)

- Similarly, $\widehat{Y_i(0)} = \begin{cases} 0, & D_i = 1 \\ Y_i/(1-p), & D_i = 0 \end{cases}$ is unbiased for $Y_i(0)$
- So $\hat{\tau}_i = \widehat{Y_i(1)} - \widehat{Y_i(0)} = \begin{cases} Y_i/p, & D_i = 1 \\ -Y_i/(1-p), & D_i = 0 \end{cases}$ is unbiased for the causal effect
- Of course, a single $\hat{\tau}_i$ is very noisy
  - A real RCT essentially combines $\hat{\tau}_i$ across $i = 1, \ldots, N$ to consistently estimate $ATE = \frac{1}{N}\sum_i \tau_i$
  - Different propensity scores are fine $\Rightarrow$ Horvitz-Thompson pscore adjustment

RCT in a small sample: Randomization inference

- In a small sample, **randomization inference** yields exact tests for the **sharp null** $\tau_i \equiv 0$ or any other constant effect, e.g. $\tau_i \equiv 3$
  - No asymptotic approximation needed
- Idea: D_i and Y_i are independent under the null but not under the alternative
  - Random permutation $\{D_i^*\}$ is independent from Y_i in both cases
  - Reject $\tau_i \equiv 0$ if $\mathrm{Cov}\left[D_i, Y_i\right]$ (or any other statistic of dependence) is in the tail of the distribution of $\mathrm{Cov}\left[D_i^*, Y_i\right]$ over many permutations $\{D_i^*\}$
- Collecting all effects that are not rejected (**"test inversion"**) gives a confidence interval for the effect
  - But only under constant effects

RCT without randomization?

- Bernoulli randomization: randomize treatment independently, e.g. with $p = 1/2$
- More efficient: "complete randomization," i.e. treat exactly $N/2$ units

  $\Rightarrow$ Better *not* to randomize the number of treated units

- Even more efficient: stratified randomization
  - For each discrete strata $X_i = x$, treat exactly half units — ensures better balance

  $\Rightarrow$ Better not to randomize the number of treated units per strata

- Kasy (2016): with a continuous X_i, best not to randomize at all!
  - Intuition: ordering units by X_i and treated every other one ensures better balance
  - It's essentially by chance whether i is even or odd on the list

References I

Abdulkadiroglu, A., J. D. Angrist, Y. Narita, and P. A. Pathak (2017): "Research Design Meets Market Design: Using Centralized Assignment for Impact Evaluation," *Econometrica*, 85, 1373–1432.

Angrist, J. D. and J.-S. Pischke (2010): "The Credibility Revolution in Empirical Economics: How Better Research Design is Taking the Con out of Econometrics," *Journal of Economic Perspectives*, 24, 3–30.

Berry, S. T. and P. A. Haile (2021): *Foundations of demand estimation*, vol. 4, Elsevier B.V., publication Title: Handbook of Industrial Organization Issue: 1 ISSN: 1573448X.

Bertrand, M. and S. Mullainathan (2004): "Are Emily and Greg more employable than Lakisha and Jamal? A field experiment on labor market discrimination," *The American Economic Review*, 94, 991–1013.

Blandhol, C., J. Bonney, M. Mogstad, and A. Torgovitsky (2025): "When is TSLS Actually LATE?" *Working Paper*.

Borusyak, K. and P. Hull (2021): "Non-Random Exposure to Exogenous Shocks: Theory and Applications," *Mimeo*.

References II

Gelman, A. and G. Imbens (2013): "Why ask Why? Forward Causal Inference and Reverse Causal Questions," Tech. Rep. w19614, National Bureau of Economic Research, Cambridge, MA.

Goldin, C. and C. Rouse (2000): "Orchestrating Impartiality: The Impact of "Blind" Auditions on Female Musicians," *American Economic Review*, 90, 715–741.

Heckman, J. J. and B. E. Honore (1990): "The Empirical Content of the Roy Model," .

Heckman, J. J., R. J. Lalonde, and J. A. Smith (1999): "The Economics and Econometrics of Active Labor Market Programs," in *Handbook of Labor Economics*, Elsevier, vol. 3, 1865–2097.

Heckman, J. J. and E. J. Vytlacil (2007): "Chapter 70 Econometric Evaluation of Social Programs, Part I: Causal Models, Structural Models and Econometric Policy Evaluation," in *Handbook of Econometrics*, Elsevier, vol. 6, 4779–4874.

Holland, P. (1986): "Statistics and causal inference," *Journal of the American Statistical Association*, 81, 945–960.

Imbens, G. W. (2020): "Potential Outcome and Directed Acyclic Graph Approaches to Causality: Relevance for Empirical Practice in Economics," *Journal of Economic Literature*, 58, 1129–1179.

References III

Kasy, M. (2016): "Why Experimenters Might Not Always Want to Randomize, and What They Could Do Instead," *Political Analysis*, 24, 324–339.

Orefice, G., H. Rapoport, and G. Santoni (2025): "How Do Immigrants Promote Exports?" *Journal of Development Economics*, 174, 103443.

Outline

1. The basics: Potential outcomes, average causal effects, and RCTs
2. Other estimands
3. Conceptual issues with the potential outcome framework. Alternative frameworks
4. Complex RCTs
5. (Skipped) Causality or prediction?

Causality vs. prediction

- Economists obsess with causality but sometimes prediction is the relevant goal
- The choice should be guided by the ultimate goal: decision making
- Two scenarios (see Kleinberg et al., 2015):

1. The action/policy $D \in \{0, 1\}$ affects the outcome Y, and the payoff (i.e., utility) π depends on Y
   - E.g. $D =$ rain dance in a drought, $Y =$ it rains

$$\pi(d) = aY(d) - bd \quad \Longrightarrow \quad \mathbb{E}\left[\pi(1) - \pi(0)\right] = a\mathbb{E}\left[Y(1) - Y(0)\right] - b$$

   - Optimal decision: $D = \mathbf{1}\left[\mathbb{E}\left[Y(1) - Y(0)\right] \geq b/a\right]$
   - This is a causal problem. Running an RCT is very helpful
   - Better knowledge of heterogeneous causal effects $\mathbb{E}\left[Y(1) - Y(0) \mid X\right]$ based on observed covariates X also yields better decisions $D(X)$

Causality vs. prediction (2)

2. Y is unaffected by D but the marginal payoff of actions, $\partial\pi/\partial D$, depends on Y
   - E.g. $D =$ take an umbrella, $Y =$ it rains

$$\pi(d) = aY \cdot d - bd \quad \Longrightarrow \quad \mathbb{E}\left[\pi(1) - \pi(0)\right] = a\mathbb{E}\left[Y\right] - b$$

   - Optimal decision: $D = \mathbf{1}\left[\mathbb{E}\left[Y\right] \geq b/a\right]$
   - This is a prediction problem. Running an RCT is not helpful
   - Better prediction $\mathbb{E}\left[Y \mid X\right]$ yields better decisions $D(X)$

- *Note:* This scenario can also be recast as a causal problem:
  - D affects $\tilde{Y}(D) =$ you get wet $= Y \cdot (1 - D)$
  - But we know potential outcome $\tilde{Y}(1) = 0$
  - And we have data on $\tilde{Y}(0) = Y$ to make a *prediction* of $\tilde{Y}(1) - \tilde{Y}(0)$

Policy-relevant prediction problems: Examples

1. Eliminating futile hip and knee replacement surgeries
   - Surgery has costs: monetary + painful recovery
   - Benefits depend on life expectancy
   - Kleinberg et al. (2015) show 10% (1%) of patients have *predictable* probability of dying within a year of 24% (44%) for reasons unrelated to this surgery
2. Improving admissions by predicting college success
   - Geiser and Santelices (2007) show that high-school GPA is a better predictor of performance at UC colleges than SAT
   - If UC had to reduce admissions, rejecting applicants with marginal GPAs would result in losing fewer good students than rejecting marginal SAT applicants
3. See Kleinberg et al. "Human Decisions and Machine Predictions" (2018) for a more subtle example on bail decisions by judges