

Outline

- Potential outcomes, selection bias, and causal identification
- Randomized trials
- The use of prediction in RCTs

Potential Outcomes

Do Hospitals Make People Healthier?

Your health status is: excellent, very good, good, fair, or poor?

	Hospital	No Hospital	Difference
Health status	3.21 (0.014)	3.93 (0.003)	-0.72***
Observations	7,774	90,049	

source: 2005 National Health Interview Survey (Angrist & Pischke 2009)

A comparison of means suggests hospitals make people worse off: those with a hospital stay in last 6 months are, on average, less healthy than those that were not admitted to the hospital

- What's wrong with this picture?

The Causal Impact of Treatment

We are interested in the relationship between some “**treatment**” (e.g. going to the hospital) and some outcome that may be impacted by the treatment (e.g. self-assessed health status)

Each individual is either treated or not:

- D_i = a **treatment dummy** equal to 1 if i is treated and 0 otherwise

Outcome of interest:

- Y = outcome we are interested in studying (e.g. health)
- Y_i = value of outcome of interest *for individual i*

Potential Outcomes

For each individual, there are two **potential outcomes**:

- $Y_{0,i}$ = i 's outcome if she **doesn't** receive treatment
- $Y_{1,i}$ = i 's outcome if she **does** receive treatment

The **causal impact** of treatment on individual i is: $Y_{1,i} - Y_{0,i}$

- How much does treatment change outcome of interest for i ?
- We are interested in **average treatment effect** – average of $Y_{1,i} - Y_{0,i}$ across people

Potential Outcomes: Example

Alejandro has a broken leg.

- $Y_{0,a}$ = If he doesn't go to the hospital, his leg won't heal properly
- $Y_{1,a}$ = If he goes to the hospital, his leg heals completely

Benicio doesn't have any broken bones. His health is fine.

- $Y_{0,b}$ = If he doesn't go to the hospital, his health is still fine
- $Y_{1,b}$ = If he goes to the hospital, his health is still fine

The Fundamental Problem of Causal Inference

The fundamental problem of causal inference:

We never observe both potential outcomes for the same individual

Comparing the mean outcome among program participants to the mean outcome among those who don't choose to participate doesn't normally provide an unbiased estimate of causal impact

- Treated, untreated likely different in absence of program:

$$E[Y_{0,i}|D_i = 1] \neq E[Y_{0,i}|D_i = 0]$$

- Difference in potential outcomes without treatment leads to **selection bias**
- $E[Y_i|D_i = 1] - E[Y_i|D_i = 0]$, the expected difference in outcomes between those who choose to participate in a program and those who do not is a biased estimator of impacts
- Could be biased up or down, relative to true average causal effect of treatment
- Bias does not disappear in large samples, even large numbers of controls may not help

The Average Treatment Effect (ATE)

The quantity of interest is the **average treatment effect** (ATE), or average causal effect, or conditional average treatment effect, or average impact, or treatment effect. . .

$$E[Y_{1,i} - Y_{0,i} | D_i = 1] = E[Y_{1,i} | D_i = 1] - E[Y_{0,i} | D_i = 1]$$

- ATE is average difference in potential outcomes across treated (or eligible) population
- Fundamental problem of causal inference: we never observe $Y_{0,i}$ for treatment group
 - ▶ $\bar{Y}_T$ is an unbiased estimator of $E[Y_{1,i} | D_i = 1]$
 - ▶ We need an unbiased estimator of $E[Y_{0,i} | D_i = 1]$
 - ▶ $\bar{Y}_C$ is **not** an unbiased estimator of $E[Y_{0,i} | D_i = 1]$

Selection Bias

When we compare (many) participants to (many) non-participants:

$$\begin{aligned} E[\bar{Y}_T - \bar{Y}_C] &= E[Y_i | D_i = 1] - E[Y_i | D_i = 0] \\ &= E[Y_{1,i} | D_i = 1] - E[Y_{0,i} | D_i = 0] \end{aligned}$$

Adding in $\underbrace{-E[Y_{0,i} | D_i = 1] + E[Y_{0,i} | D_i = 1]}_{=0}$, we get:

Difference in group means

$$= \underbrace{E[Y_{1,i} | D_i = 1] - E[Y_{0,i} | D_i = 1]}_{\text{average causal effect on participants}} + \underbrace{E[Y_{0,i} | D_i = 1] - E[Y_{0,i} | D_i = 0]}_{\text{selection bias}}$$

The Experimental Ideal

Random Assignment Eliminates Selection Bias*

Experimental approach:

- **Random assignment to treatment:** eligibility for program is determined at random, e.g. via pulling names out of a hat, or using a computer pseudo-random number generator

When treatment status is randomly assigned,
treatment, control groups are random samples of a single population
(e.g. the population of all eligible applicants for the program)

$$\Rightarrow E[Y_{0,i}|D_i = 1] = E[Y_{0,i}|D_i = 0] = E[Y_{0,i}]$$

$\Rightarrow$ Expected outcomes are equal in the absence of the program, eliminating selection bias

*Some restrictions apply (see ECON 379 and ECON 474)

Random Assignment Eliminates Selection Bias

$\bar{Y}_T - \bar{Y}_C$ provides an unbiased estimate of the (causal) average treatment effect (or ATE):

$$= E[Y_i | D_i = 1] - E[Y_i | D_i = 0]$$

$$= E[Y_{1,i} | D_i = 1] - E[Y_{0,i} | D_i = 0]$$

$$= E[Y_{1,i} | D_i = 1] - E[Y_{0,i} | D_i = 1] + E[Y_{0,i} | D_i = 1] - E[Y_{0,i} | D_i = 0]$$

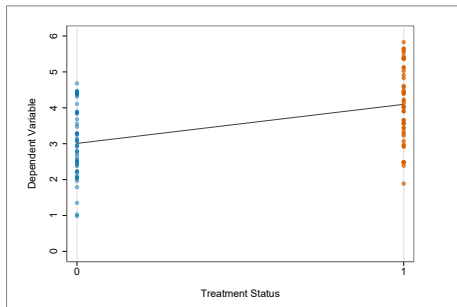
$$= \underbrace{E[Y_{1,i} | D_i = 1] - E[Y_{0,i} | D_i = 1]}_{\text{average treatment effect on participants}} + \underbrace{E[Y_{0,i} | D_i = 1] - E[Y_{0,i} | D_i = 0]}_{=0}$$

$$= \underbrace{E[Y_{1,i}] - E[Y_{0,i}]}_{\text{ATE}}$$

Regression Analysis of RCTs

Null hypothesis (H_0): average treatment effect is zero $\Leftrightarrow ATE = 0 \Leftrightarrow \bar{Y}_T = \bar{Y}_C$

Testing H_0 in a regression framework: $Y = \alpha + \beta D$



control	treatment
$D = 0$	$D = 1$
$\hat{\alpha}$	$\hat{\alpha} + \hat{\beta}$

$\hat{\alpha} = \bar{Y}_C$ (control group mean)

$\hat{\beta} = \bar{Y}_T - \bar{Y}_C$ (difference in means)

A Very Short History of Randomized Trials in the Social Sciences



A Very Short History of Randomized Trials in the Social Sciences



RCTs and the Replication Crisis

- Economists (& other social scientists) have conducted thousands of RCTs since the 1990s
- RCTs differ from other types of empirical economics because researchers typically collect their own (usually survey) data to measure outcomes of interest and important covariates
- Because researchers propose, and then fund, studies before data is collected, RCTs have played an important role in the evolution of practices around transparency and replicability
 - ▶ **Pre-analysis plans** are required by most top economics journals that publish RCTs
 - ▶ **Replication files** are now required by almost all economics journals

Regression Analysis of RCTs (in Practice)

RCTs and Machine Learning

Social scientists conducting RCTs use machine learning in two (main) ways:

- Researchers choose which baseline covariates to measure and include in analysis
- Researchers try to identify sub-groups most-impacted by treatment

Both uses reflect **prediction** problems:

- Covariate selection: identify baseline X variables that predict $Y_{0,i}$
- Treatment effect heterogeneity: identify baseline X variables that predict $Y_{1,i} - Y_{0,i}$

RCT Regression Specification with Controls

Typical regression specification:

$$Y_{t=1,i} = \alpha + \beta D_i + \gamma Y_{t=0,i} + \delta X_{t=0,i} + \eta_{strata} + \varepsilon_i$$

We typically include these controls:

- Dummies for randomization strata (η_{strata})
- Selected baseline covariates that are not balanced across treatments*
- Baseline covariates that predict the outcome
 - ▶ Baseline values of outcome variables are (sometimes) most important (ANCOVA)

We do not want to include:

- Controls that could be impacted by treatment (“bad controls” problem)

The Minimum Detectable Effect

The **minimum detectable effect** (or MDE) is the smallest effect size that we can detect with power of 0.8 (i.e. the probability of a Type II error, failing to reject a false null, is 0.2)

$$\begin{aligned} \text{MDE} &= (t_{\alpha/2} + t_{1-\kappa}) \sqrt{\frac{1}{P(1-P)}} \sqrt{\frac{\sigma^2}{N}} \\ &\approx 2.8 \sqrt{\frac{1}{P(1-P)}} \sqrt{\frac{\sigma^2}{N}} \end{aligned}$$

where:

- P is the proportion of the sample assigned to treatment
- N is the sample size
- σ^2 is the variance of the (residualized) outcome

The MDE Depends on Covariates

The **Frisch-Waugh-Lovell** Theorem tells us that the $\hat{\beta}$ that minimizes RSS in the regression

$$Y = \alpha + \beta D + \theta_1 X_1 + \dots + \theta_K X_K$$

is equal to the $\hat{\beta}$ that minimizes RSS in the regression

$$\tilde{Y} = \alpha + \beta \tilde{D}$$

where $\tilde{Y}$ and $\tilde{D}$ are the residuals from regressions of Y and D on $X_1, \dots, X_K$

⇒ Choosing the right covariates can increase statistical power (the covariate selection problem)

Treatment Effect Heterogeneity

	control	treatment
boys	$\bar{Y}_C^{boys}$	$\bar{Y}_T^{boys}$
girls	$\bar{Y}_C^{girls}$	$\bar{Y}_T^{girls}$

Interaction Terms and Predicted Values

$$Y = \alpha + \beta \textit{Treatment} + \lambda \textit{Girl} + \theta(\textit{Girl} \times \textit{Treatment})$$

$\bar{Y}_C^{boys}$	$\bar{Y}_T^{boys}$
$\bar{Y}_C^{girls}$	$\bar{Y}_T^{girls}$

Treatment Effect Heterogeneity with Sparse Data

Thought experiment: you test for treatment effect heterogeneity using 100 baseline variables

- How many variables do you expect to be statistically significant ($p\text{-values} < 0.05$)?

The “old-fashioned” approach to testing for treatment effect heterogeneity works very well (and has good statistical power) if you can identify key dimensions of heterogeneity in advance

- We often want a more flexible approach that allows us to search for heterogeneity
- But we also want to avoid **p-hacking** (don't test 100 variables and report 2 of them!)
- Machine learning allows us to identify X s that predict large differences in the mean of Y between the treatment and control groups (the treatment effect heterogeneity problem)

Optional Lab

Subsidizing Malaria Treatment in Kenya

American Economic Review 2015, 105(2): 609–645
<http://dx.doi.org/10.1257/aer.20130267>

Price Subsidies, Diagnostic Tests, and Targeting of Malaria Treatment: Evidence from a Randomized Controlled Trial[†]

By JESSICA COHEN, PASCALINE DUPAS, AND SIMONE SCHANER[✉]

Both under- and over-treatment of communicable diseases are public bads. But efforts to decrease one run the risk of increasing the other. Using rich experimental data on household treatment-seeking behavior in Kenya, we study the implications of this trade-off for subsidizing life-saving antimalarials sold over-the-counter at retail drug outlets. We show that a very high subsidy (such as the one under consideration by the international community) dramatically increases access, but nearly one-half of subsidized pills go to patients without malaria. We study two ways to better target subsidized drugs: reducing the subsidy level, and introducing rapid malaria tests over-the-counter. (JEL D12, D82, I12, O12, O15)

Subsidizing Malaria Treatment in Kenya

Comparison Group

No subsidy. Households received vouchers to purchase unsubsidized ACTs at the pre-AMFm retail price in Kenya: KSh 500 (approximately US\$6.25, using a 2009 exchange rate of KSh 80/ US\$1).

ACT Subsidy

Households were randomly selected to receive vouchers for ACTs at one of three subsidy levels:

- **92 percent** (US\$0.50 per adult dose, corresponds to the Kenyan government's target retail price of KSh 40 under the AMFm)
- **88 percent** (US\$0.75 per adult dose)
- **80 percent** (US\$1.25 per adult dose)

ACT & RDT Subsidy

Households received one of the three ACT subsidy levels above and were also randomly assigned to receive vouchers for rapid diagnostic tests (RDTs) either for free or at an 85 percent subsidy (US\$0.20).



PHOTO BY AUDE GUERRUCCI

Optional Lab #6

1. Open the Cohen, Dupas, and Schaner (2015) data set (E1-CohenEtAl-data.dta)
2. In a bivariate regression on a (single) dummy variable, the estimated OLS coefficient $\hat{\beta}$ is the difference in means between the treatment group and the control group. Regress the outcome `c_act` on the treatment dummy `act_any` to show this.
3. Baseline covariates are in columns that begin with `b_`. How many are there in the data set? Which of the baseline covariates predict the outcome `c_act`? How much does the R-squared increase when you include them in the regression from Question 2?
4. Define a `low_education` dummy equal to one if the household head had 0 or 1 years of schooling. Estimate the OLS specification that tests for treatment effect heterogeneity. How does the estimated impact differ between high and low education households?