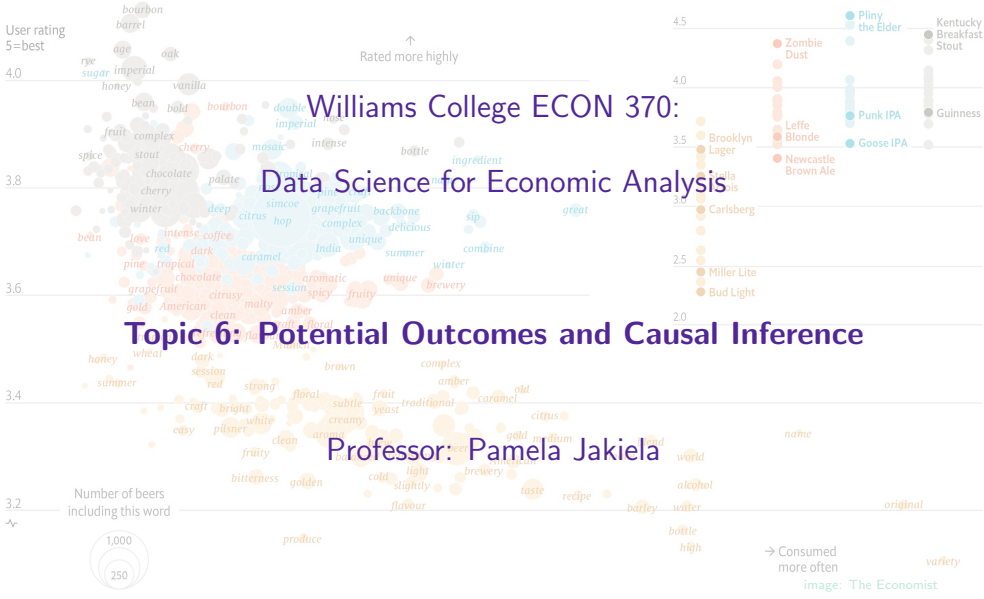


Williams College ECON 370:

Data Science for Economic Analysis

Topic 6: Potential Outcomes and Causal Inference

Professor: Pamela Jakiela

Outline

- Potential Outcomes, Selection Bias, and Causal Inference
- Randomization
- Statistical Power in Randomized Trials

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	3.93	-0.72^{***}
	(0.014)	(0.003)	
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 (eg. self-assessed health status)

Each individual is either treated or not:

- D_i = is 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

Potential Outcomes: Example

	Yes Hospital	**No Hospital**
Alejandro	$Y_{1,a}$	$Y_{0,a}$
Benicio	$Y_{1,b}$	$Y_{0,b}$

The Fundamental Problem of Causal Inference

The fundamental problem of causal inference:

We never observe both potential outcomes for the same individual

$\Rightarrow$ Creates a missing data problem whenever we try to compare treated to untreated

For any individual, we can only observe one potential outcome:

$$Y_i = \begin{cases} Y_{0,i} & \text{if } D_i = 0 \\ Y_{1,i} & \text{if } D_i = 1 \end{cases}$$

Potential outcomes without treatment (i.e. values of $Y_{0,i}$) may differ between those who choose to take-up treatment (Alejandro with a broken leg) and those who do not (healthy Benicio)

Selection Bias

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
- Difference in potential outcomes without treatment leads to **selection bias**
- The difference in outcome means, $\bar{Y}_T - \bar{Y}_C$, is a biased estimator of program impacts
- $\bar{Y}_T - \bar{Y}_C$ 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

Notation: Mathematical Expectations

The **expected value** or mathematical expectation of Y_i, $E[Y_i]$:

- Equivalent to population mean or sample average in an infinite population
  - Example: probability coin flipped lands heads
  - Equivalent to fraction heads after a (very) large number of flips

Law of Large Numbers:

- In small samples, realized average of Y_i might be far from the true mean of Y_i
- Average of Y_i gets very close to $E[Y_i]$ as number of observations gets large

Notation: Conditional Expectations

Conditional expectation:

$E[Y_i|X_i = x]$

Conditional expectation of Y_i given $X_i = x$ is average of Y_i in infinite population where $X_i = x$

Example:
Let Y_i be height, and let $X_i \in \{0, 1\}$ be an "economics professor dummy"

- $E[Y_i|X_i = 1]$ is the average height among (infinitely many) economics professors
- $E[Y_i|X_i = 0]$ is the population mean of height among everybody else

Notation: 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 (usually) across treated population
- Fundamental problem of causal inference: we never observe $Y_{0,i}$ for treatment group
  - $\bar{Y}_T$ is an unbiased estimator of $E[Y_i|D_i = 1] = E[Y_{1,i}|D_i = 1]$
  - We need an unbiased estimator of $E[Y_{0,i}|D_i = 1]$

Notation: Selection Bias

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

$$E[\bar{Y}_T - \bar{Y}_C] = E[Y_i|D_i = 1] - E[Y_i|D_i = 0]$$

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

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

Difference in group means

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

Selection Bias: Summary

We would like to calculate average treatment effect by comparing potential outcomes for i both with and without treatment, but for each i we can only observe one potential outcome

- Can't observe the **counterfactual** (i.e. the other potential outcome)

To estimate causal impacts on the set of people who choose to take up treatment, we must identify a comparison group that is similar to the treatment group in the absence of treatment

- This is hard – typically impossible in observational data
- An **identification strategy** is a research design specifying treatment, comparison groups
- A good identification strategy: variation in treatment status that is good-as-random

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)

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

Expected outcomes are equal in the absence of the program

Random Assignment Eliminates Selection Bias

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

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

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

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

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

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

Random Assignment Eliminates Selection Bias: Assumptions

Excellent news: random assignment eliminates selection bias*

*Some restrictions apply

Random assignment is not (quite) magic:

- Relies on Law of Large Numbers, which only makes sense for large(ish) samples
- Stable Unit Treatment Value Assumption (SUTVA): individual outcomes depend on one's own treatment status, but not on anyone else's treatment status (i.e. no spillovers)

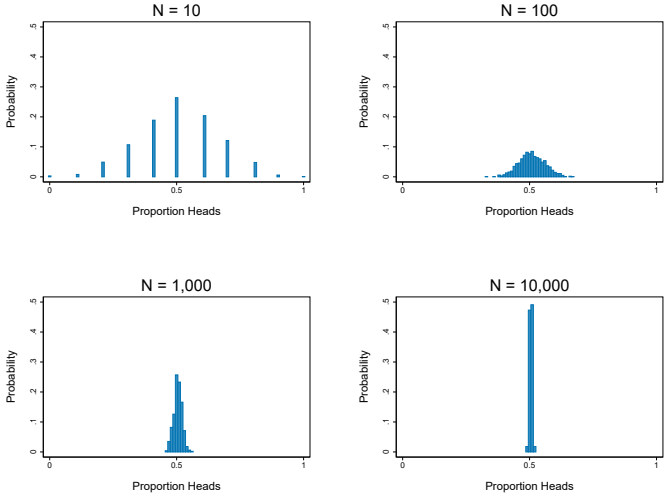
Sample Size Matters: Example

Example: imagine that I want to evaluate the impact of fancy new software Stata 138, so I randomly choose which of my two research assistants (below) should receive a copy

They're different! Omitted variables likely to matter — by chance — in small samples

> *"Randomization works not by eliminating individual difference but rather by ensuring that the mix of individuals being compared is the same. Think of this as comparing barrels that include equal proportions of apples and oranges."*
> – Angrist and Pischke (2009)

The Law of Large Numbers in Practice



Stable Unit Treatment Value Assumption (SUTVA)

The **Stable Unit Treatment Value Assumption (SUTVA)**:

- Imbens and Rubin (2015):
  "potential outcomes for any unit do not vary with the treatments assigned to other units"
- Remember: binary treatment, two potential outcomes is only a model

When is SUTVA likely to be violated?

- When there are spillovers (so i's treatment impacts j)
- Examples: vaccination/health, network externalities, equilibrium effects
  - This is why we have "cluster-randomized" trials

Randomization Eliminates Selection Bias: Summary

When treatment is randomly assigned (at an appropriate level), difference in outcomes between treatment and control groups provides an unbiased estimate of the causal impact of treatment

Randomly assigning treatment status eliminates selection bias (at least in expectation) because treatment, control groups are random samples of same underlying population

A Very Short History of Randomized Trials in the Social Sciences

Regression Analysis of RCTs

Treatment Effects Under Random Assignment

Expected value of control group mean:

$$E[\bar{Y}_C] = E[Y_i|D_i = 0] = E[Y_{0,i}|D_i = 0] = E[Y_{0,i}]$$

equal to population mean because control group is a random sample

Treatment Effects Under Random Assignment

Expected value of control group mean:

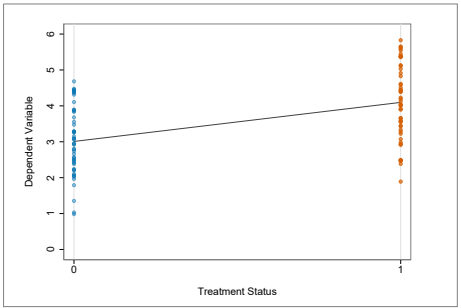
$$E[\bar{Y}_C] = E[Y_i|D_i = 0] = E[Y_{0,i}|D_i = 0] = E[Y_{0,i}]$$

Expected value of treatment group mean:

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

OLS Regression on a Binary Independent Variable

$$Y = \alpha + \beta D$$



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)

RCT Regression Specification with Controls

More typical regression specification:

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

We typically include these controls:

- Dummies for randomization strata (κ_{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)

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

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

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

Subsidizing Malaria Treatment in Kenya

TABLE 2—IMPACT OF ACT SUBSIDY ON TREATMENT SEEKING AND ACT ACCESS

	Took ACT (1)	Took ACT from drug shop (2)	Took ACT from health center (3)	Visited drug shop (4)	Visited health center (5)	Sought no care (6)	Took malaria test (7)	Took antibiotic (8)
Panel A. Pooled impact								
Any ACT subsidy	0.187***	0.222***	−0.038	0.167***	−0.079*	−0.096***	−0.014	−0.072**
	(0.038)	(0.031)	(0.030)	(0.046)	(0.042)	(0.036)	(0.038)	(0.034)
Panel B. Impact by subsidy level								
B1. ACT subsidy = 92 percent	0.225***	0.249***	−0.024	0.159***	−0.055	−0.110***	−0.031	−0.046
	(0.053)	(0.046)	(0.037)	(0.058)	(0.053)	(0.042)	(0.048)	(0.043)
B2. ACT subsidy = 88 percent	0.161***	0.217***	−0.056	0.167***	−0.070	−0.097**	−0.042	−0.062
	(0.050)	(0.043)	(0.037)	(0.058)	(0.052)	(0.042)	(0.047)	(0.040)
B3. ACT subsidy = 80 percent	0.178***	0.206***	−0.035	0.173***	−0.106**	−0.085*	0.023	−0.100***
	(0.048)	(0.042)	(0.035)	(0.054)	(0.047)	(0.045)	(0.046)	(0.038)
p-value: B1 = B2 = B3 = 0	0.000***	0.000***	0.498	0.004***	0.164	0.048**	0.533	0.066
p-value: B1 = B2 = B3	0.531	0.723	0.660	0.968	0.535	0.846	0.362	0.304
DV mean (control group)	0.190	0.071	0.119	0.488	0.286	0.226	0.214	0.185
Observations	631	631	631	631	631	631	631	631

Notes: "Substandard" malaria treatment includes non-ACT antimalarials and antipyretics. Sample excludes all households selected for a surprise or subsidized RDT. The unit of observation is the first illness episode with at least one malaria-like symptom that the household experienced following the baseline. A few households have multiple observations if multiple household members were ill simultaneously. Robust standard errors clustered at the household level in parentheses. All regressions control for household head age and a full set of strata dummies.

***Significant at the 1 percent level.
**Significant at the 5 percent level.
*Significant at the 10 percent level.

Subsidizing Malaria Treatment in Kenya

POLICY LESSONS

PHOTO BY IPA KENYA

Understanding the Context

The data collected during this evaluation suggest that households in the study area:

- Tend to bypass the public health care system if they are poor, likely because they live far from health centers, making travel costs too high. Instead they rely on local drug shops that do not offer diagnostic services.
- Experience illnesses suspected to be malaria very often. These illness episodes are generally not formally diagnosed and are typically presumptively treated with less effective antimalarials procured from a drug shop.

Subsidizing ACTs provides measurable benefits, especially for vulnerable children and the poorest households. Many households effectively miss out on the existing free treatment at public facilities and either do not seek care for malaria at all or take less effective medicines. For these families, a retail-sector ACT subsidy substantially improves access to proper treatment.

A slightly lower subsidy can improve targeting without compromising access for children. Moving from the AMFm target subsidy level (roughly 92 percent) to a somewhat lower subsidy (80 percent) reduced overtreatment among adults, while keeping access constant for children. These results suggest that an ACT subsidy is clearly needed, but that a slightly lower subsidy may achieve similar benefits at a lower cost.

Rapid diagnostic tests may be a promising means to improve targeting. People were very willing to try out rapid diagnostic testing, including sharing the cost of the test. More than half of adults who suspected malaria but got a negative test result decided not to purchase the subsidized ACT. Imperfect compliance with malaria test results is also common among public health workers, and thus it may take some time for people with malaria to become familiar with and trust RDTs.

Lab #6: Takeaways

1. You should be able to open the Cohen, Dupas, and Schaner (2015) data set
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 comparison group, which can also be recovered from a t-test of the equality of means in the two groups
3. Same logic applies when we include separate dummies for multiple (randomly-assigned) treatments, with no interaction terms and no additional covariates (or strata dummies)
4. When the treatment dummy aggregates multiple distinct treatment intensities, each treated observation weighted equally in calculating the estimated treatment effect
5. In a bivariate OLS regression on a continuous independent variable, the estimated OLS coefficient is a linear combination of the observed values of the outcome variable; the "weights" on each value if Y_i are proportional to $X_i - \bar{X}$ (scaled by $1/\sum_i \left(X_i - \bar{X}\right)^2$)
6. Including an additional control/covariate is equivalent to regressing Y_i and X_i on that covariate, and the regressing the predicted residuals of Y on the predicted residuals of X