Introduction to the Selection Problem

- Aliens come and orbit earth, see sick people in hospitals and conclude “these ‘hospitals’ are hurting people”
- Motivated by anger and compassion, they kill the doctors to save the patients
- Sounds stupid, but earthlings do this too - all the time

#1: Correlation and causality are very different concepts

- Causal question:

  "If I hospitalize (D) my child, will her health (Y) improve?"

- Correlation question:

$$\frac{1}{n}\frac{Cov(D,Y)}{\sqrt{Var_D}\sqrt{Var_Y}}$$

- These are not the same thing

#2: Coming first may not mean causality!

- Every morning the rooster crows and then the sun rises
- Did the rooster cause the sun to rise? Or did the sun cause the rooster to crow?
- *Post hoc ergo propter hoc*: "after this, therefore, because of this"

#3: No correlation does not mean no causality!

- A sailor sails her sailboat across a lake
- Wind blows, and she perfectly counters by turning the rudder
- The same aliens observe from space and say "Look at the way she's moving that rudder back and forth but going in a straight line. That rudder is broken." So they send her a new rudder
- They're wrong but why are they wrong? There is, after all, no correlation

Introduction to potential outcomes model

- Let the treatment be a binary variable:

$$D_{i,t} = \begin{cases} 1 \text{ if hospitalized at time } t \\ 0 \text{ if not hospitalized at time } t \end{cases}$$

  where i indexes an individual observation, such as a person
- Potential outcomes:

$$Y^j_{i,t} = \begin{cases} 1 \text{ health if hospitalized at time } t \\ 0 \text{ health if not hospitalized at time } t \end{cases}$$

  where j indexes a counterfactual state of the world

Moving between worlds

- I'll drop t subscript, but note – these are potential outcomes for the same person at the exact same moment in time
- A potential outcome Y^1 is not the historical outcome Y either conceptually or notationally
- Potential outcomes are hypothetical states of the world but historical outcomes are ex post realizations
- Major philosophical move here: go from the potential worlds to the actual (historical) world based on your treatment assignment

Important definitions

Definition 1: Individual treatment effect

The individual treatment effect, δ_i, equals $Y_i^1 - Y_i^0$

Definition 3: Switching equation

An individual's observed health outcomes, Y, is determined by treatment assignment, D_i, and corresponding potential outcomes:

$$Y_i = D_i Y_i^1 + (1 - D_i) Y_i^0$$

$$Y_i = \begin{cases} Y_i^1 \text{ if } D_i = 1 \\ Y_i^0 \text{ if } D_i = 0 \end{cases}$$

Definition 2: Average treatment effect (ATE)

The average treatment effect is the population average of all i individual treatment effects

$$\begin{aligned} E[\delta_i] &= E[Y_i^1 - Y_i^0] \\ &= E[Y_i^1] - E[Y_i^0] \end{aligned}$$

So what's the problem?

Definition 4: Fundamental problem of causal inference

It is impossible to observe both Y_i^1 and Y_i^0 for the same individual and so individual causal effects, δ_i, are *unknowable*.

Conditional Average Treatment Effects

Definition 5: Average Treatment Effect on the Treated (ATT)

The average treatment effect on the treatment group is equal to the average treatment effect conditional on being a treatment group member:

$$\begin{aligned} E[\delta|D=1] &= E[Y^1 - Y^0|D=1] \\ &= E[Y^1|D=1] - E[Y^0|D=1] \end{aligned}$$

Definition 6: Average Treatment Effect on the Untreated (ATU)

The average treatment effect on the untreated group is equal to the average treatment effect conditional on being untreated:

$$\begin{aligned} E[\delta|D=0] &= E[Y^1 - Y^0|D=0] \\ &= E[Y^1|D=0] - E[Y^0|D=0] \end{aligned}$$

Causality and comparisons

- Comparisons are at the heart of the causal problem, but not all comparisons are equal because of the selection problem
- Does the hospital make me sick? Or am I sick, and that's why I went to the hospital?
- Why can't I just compare my health (Scott) with someone who isn't in the hospital (Nathan)? Aren't we supposed to have a "control group"?
- What are we actually measuring if we compare average health outcomes for the hospitalized with the non-hospitalized?

Definition 7: Simple difference in mean outcomes (SDO)

A simple difference in mean outcomes (SDO) is the difference between the population average outcome for the treatment and control groups, and can be approximated by the sample averages:

$$\begin{aligned} SDO &= E[Y^1|D=1] - E[Y^0|D=0] \\ &= E_N[Y|D=1] - E_N[Y|D=0] \end{aligned}$$

in large samples.

SDO vs. ATE

Notice the subtle difference between the SDO and ATE notation:

$$E[Y|D=1]-E[Y|D=0] \quad \lesseqgtr \quad E[Y^1]-E[Y^0]$$

- The SDO is an *estimate*, whereas ATE is a *parameter*
- SDO is a crank that turns data into numbers
- ATE is a parameter that is unknowable because of the fundamental problem of causal inference
- SDO can line up with the ATE and also cannot line up with the ATE.

Biased simple difference in mean outcomes

Decomposition of the SDO

The simple difference in mean outcomes can be decomposed into three parts (ignoring sample average notation):

$$\begin{aligned} E[Y^1|D=1]-E[Y^0|D=0] &= ATE \\ &\quad +E[Y^0|D=1]-E[Y^0|D=0] \\ &\quad +(1-\pi)(ATT-ATU) \end{aligned}$$

Seeing is believing so let's work through this identity

Decomposition of SDO

ATE is equal to sum of conditional average expectations by LIE

$$\begin{aligned} \text{ATE} &= E[Y^1] - E[Y^0] \\ &= \{\pi E[Y^1|D=1] + (1-\pi)E[Y^1|D=0]\} \\ &\quad -\{\pi E[Y^0|D=1] + (1-\pi)E[Y^0|D=0]\} \end{aligned}$$

Use simplified notations

$$\begin{aligned} E[Y^1|D=1] &= a \\ E[Y^1|D=0] &= b \\ E[Y^0|D=1] &= c \\ E[Y^0|D=0] &= d \\ \text{ATE} &= e \end{aligned}$$

Rewrite ATE

$$\begin{aligned} e &= \{\pi a + (1-\pi)b\} \\ &\quad -\{\pi c + (1-\pi)d\} \end{aligned}$$

Move SDO terms to LHS

$$
\begin{aligned}
e &= \{\pi a + (1-\pi)b\} - \{\pi c + (1-\pi)d\} \\
e &= \pi a + b - \pi b - \pi c - d + \pi d \\
e &= \pi a + b - \pi b - \pi c - d + \pi d + (\mathbf{a} - \mathbf{a}) + (\mathbf{c} - \mathbf{c}) + (\mathbf{d} - \mathbf{d}) \\
0 &= e - \pi a - b + \pi b + \pi c + d - \pi d - \mathbf{a} + \mathbf{a} - \mathbf{c} + \mathbf{c} - \mathbf{d} + \mathbf{d} \\
\mathbf{a} - \mathbf{d} &= e - \pi a - b + \pi b + \pi c + d - \pi d + \mathbf{a} - \mathbf{c} + \mathbf{c} - \mathbf{d} \\
\mathbf{a} - \mathbf{d} &= e + (\mathbf{c} - \mathbf{d}) + \mathbf{a} - \pi a - b + \pi b - \mathbf{c} + \pi c + d - \pi d \\
\mathbf{a} - \mathbf{d} &= e + (\mathbf{c} - \mathbf{d}) + (1-\pi)a - (1-\pi)b + (1-\pi)d - (1-\pi)c \\
\mathbf{a} - \mathbf{d} &= e + (\mathbf{c} - \mathbf{d}) + (1-\pi)(a-c) - (1-\pi)(b-d)
\end{aligned}
$$

Substitute conditional means

$$
\begin{aligned}
E[Y^1|D=1] - E[Y^0|D=0] &= \text{ATE} \\
&\quad + (E[Y^0|D=1] - E[Y^0|D=0]) \\
&\quad + (1-\pi)(\{E[Y^1|D=1] - E[Y^0|D=1]\} \\
&\quad - (1-\pi)\{E[Y^1|D=0] - E[Y^0|D=0]\}) \\
E[Y^1|D=1] - E[Y^0|D=0] &= ATE \\
&\quad + (E[Y^0|D=1] - E[Y^0|D=0]) \\
&\quad + (1-\pi)(ATT - ATU)
\end{aligned}
$$

Decomposition of difference in means

$$\begin{aligned}
\underbrace{E_N[y_i|d_i=1]-E_N[y_i|d_i=0]}_{\text{SDO}} &= \underbrace{E[Y^1]-E[Y^0]}_{\text{Average Treatment Effect}} \\
&\quad + \underbrace{E[Y^0|D=1]-E[Y^0|D=0]}_{\text{Selection bias}} \\
&\quad + \underbrace{(1-\pi)(ATT-ATU)}_{\text{Heterogenous treatment effect bias}}
\end{aligned}$$

where $E_N[Y|D=1] \to E[Y^1|D=1]$, $E_N[Y|D=0] \to E[Y^0|D=0]$ and $(1-\pi)$ is the share of the population in the control group.

Independence assumption

Independence assumption

Treatment is independent of potential outcomes

$$(Y^0, Y^1) \perp\!\!\!\perp D$$

In words: Random assignment means that the treatment has been assigned to units independent of their potential outcomes. Thus, mean potential outcomes for the treatment group and control group are the same *for a given state of the world*

$$\begin{aligned} E[Y^0|D=1] &= E[Y^0|D=0] \\ E[Y^1|D=1] &= E[Y^1|D=0] \end{aligned}$$

Random Assignment Solves the Selection Problem

$$
\begin{aligned}
\underbrace{E_N[y_i|d_i=1]-E_N[y_i|d_i=0]}_{\text{SDO}} \quad = \quad & \underbrace{E[Y^1]-E[Y^0]}_{\text{Average Treatment Effect}} \\
& + \underbrace{E[Y^0|D=1]-E[Y^0|D=0]}_{\text{Selection bias}} \\
& + \underbrace{(1-\pi)(ATT-ATU)}_{\text{Heterogenous treatment effect bias}}
\end{aligned}
$$

- If treatment is independent of potential outcomes, then swap out equations and **selection bias** zeroes out:

$$
E[Y^0|D=1]-E[Y^0|D=0] \quad = \quad 0
$$

Random Assignment Solves the Heterogenous Treatment Effects

- How does randomization affect heterogeneity treatment effects bias from the third line? Rewrite definitions for ATT and ATU:

$$\text{ATT} = E[Y^1|D=1] - E[Y^0|D=1]$$
$$\text{ATU} = E[Y^1|D=0] - E[Y^0|D=0]$$

- Rewrite the third row bias after $1-\pi$:

$$\begin{aligned} ATT - ATU &= \mathbf{E[Y^1 \mid D{=}1]} - E[Y^0|D=1] \\ &\quad -\mathbf{E[Y^1 \mid D{=}0]} + E[Y^0|D=0] \\ &= 0 \end{aligned}$$

- If treatment is independent of potential outcomes, then:

$$\begin{aligned} E_N[y_i|d_i=1] - E_N[y_i|d_i=0] &= E[Y^1] - E[Y^0] \\ SDO &= ATE \end{aligned}$$

Careful with this notation

- Independence only implies that the the average values for a given potential outcome (i.e., Y^1 or Y^0) are the same for the groups who did receive the treatment as those who did not
- Independence does **not** does not imply

$$E[Y^1|D=1] = E[Y^0|D=0]$$

SUTVA

- Potential outcomes model places a limit on what we can measure: the "stable unit-treatment value assumption" . Horrible acronym.
    1. **S**: ***s**table*
    2. **U**: across all ***u**nits*, or the population
    3. **TV**: ***t**reatment-value* ("treatment effect", "causal effect")
    4. **A**: ***a**ssumption*
- SUTVA means that average treatment effects are parameters that assume (1) homogenous dosage, (2) potential outcomes are invariant to who else (and how many) is treated (e.g., externalities), and (3) partial equilibrium

SUTVA: Homogenous dose

- SUTVA constrains what the treatment can be.
- Individuals are receiving the same treatment – i.e., the "dose" of the treatment to each member of the treatment group is the same. That's the "stable unit" part.
- If we are estimating the effect of hospitalization on health status, we assume everyone is getting the same dose of the hospitalization treatment.
- Easy to imagine violations if hospital quality varies, though, across individuals. But, that just means we have to be careful what we are and are not defining as *the treatment*

SUTVA: No spillovers to other units

- What if hospitalizing Scott (hospitalized, $D = 1$) is actually about vaccinating Scott from small pox?
- If Scott is vaccinated for small pox, then Nathan's potential health status (without vaccination) may be higher than when he isn't vaccinated.
- In other words, Y^0_{Nathan}, may vary with what Scott does *regardless of whether he himself receives treatment*.
- SUTVA means that you don't have a problem like this.
- If there are no externalities from treatment, then δ_i is stable for each i unit regardless of whether someone else receives the treatment too.

SUTVA: Partial equilibrium only

Easier to imagine this with a different example.

- Scaling up can be a problem because of rising costs of production
- Let's say we estimate a causal effect of early childhood intervention in some state
- Now the President wants to adopt it for the whole United States – will it have the same effect as we found?
- What if expansion requires hiring lower quality teachers just to make classes?

Demand for Learning HIV Status

- Rebecca Thornton implemented an RCT in rural Malawi for her job market paper at Harvard in mid-2000s
- At the time, it was an article of faith that you could fight the HIV epidemic in Africa by encouraging people to get tested; but Thornton wanted to see if this was true
- She randomly assigned cash incentives to people to incentivize learning their HIV status
- Also examined whether learning changed sexual behavior.

Experimental design

- Respondents were offered a free door-to-door HIV test
- Treatment is randomized vouchers worth between zero and three dollars
- These vouchers were redeemable once they visited a nearby voluntary counseling and testing center (VCT)
- Estimates her models using OLS with controls

Why Include Control Variables?

- To evaluate experimental data, one may want to add additional controls in the multivariate regression model. So, instead of estimating the prior equation, we might estimate:

$$Y_i = \alpha + \delta D_i + \gamma X_i + \eta_i$$

- There are 2 main reasons for including additional controls in the regression models:
  1. Conditional random assignment. Sometimes randomization is done *conditional* on some observable (e.g., gender, school, districts)
  2. Exogenous controls increase precision. Although control variables X_i are uncorrelated with D_i, they may have substantial explanatory power for Y_i. Including controls thus reduces variance in the residuals which lowers the standard errors of the regression estimates.

Table: Impact of Monetary Incentives and Distance on Learning HIV Results

	1	2	3	4	5
Any incentive	0.431***	0.309***	0.219***	0.220***	0.219 ***
	(0.023)	(0.026)	(0.029)	(0.029)	(0.029)
Amount of incentive		0.091***	0.274***	0.274***	0.273***
		(0.012)	(0.036)	(0.035)	(0.036)
Amount of incentive^2			−0.063***	−0.063***	−0.063***
			(0.011)	(0.011)	(0.011)
HIV	−0.055*	−0.052	−0.05	−0.058*	−0.055*
	(0.031)	(0.032)	(0.032)	(0.031)	(0.031)
Distance (km)				−0.076***	
				(0.027)	
Distance^2				0.010**	
				(0.005)	
Controls	Yes	Yes	Yes	Yes	Yes
Sample size	2,812	2,812	2,812	2,812	2,812
Average attendance	0.69	0.69	0.69	0.69	0.69

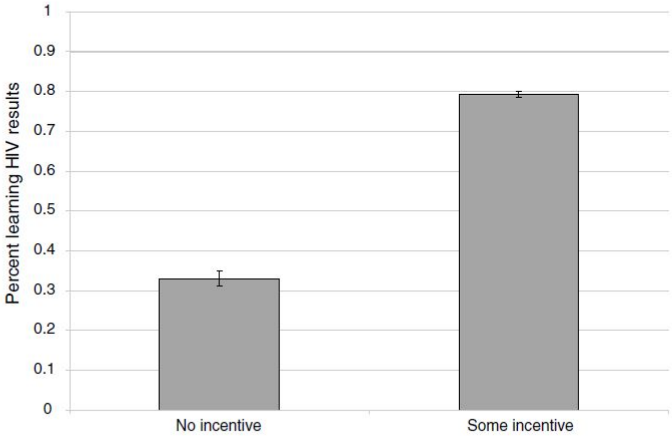

Figure: Visual representation of cash transfers on learning HIV test results.

Results

- Even small incentives were effective
- Any incentive increases learning HIV status by 43% compared to the control (mean 34%)
- Next she looks at the effect that learning HIV status has on risky sexual behavior

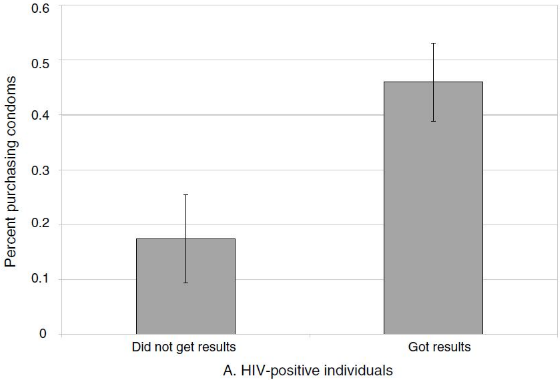

Figure: Visual representation of cash transfers on condom purchases for HIV positive individuals.

Table: Reactions to Learning HIV Results among Sexually Active at Baseline

Dependent variables:	Bought condoms		Number of condoms bought	
	OLS	IV	OLS	IV
Got results	−0.022	−0.069	−0.193	−0.303
	(0.025)	(0.062)	(0.148)	(0.285)
Got results × HIV	0.418***	0.248	1.778***	1.689**
	(0.143)	(0.169)	(0.564)	(0.784)
HIV	−0.175**	−0.073	−0.873	−0.831
	(0.085)	(0.123)	(0.275)	(0.375)
Controls	Yes	Yes	Yes	Yes
Sample size	1,008	1,008	1,008	1,008
Mean	0.26	0.26	0.95	0.95

Results

- For those who were HIV+ and got their test results, 42% more likely to buy condoms (but shrinks and becomes insignificant at conventional levels with IV).
- Number of condoms bought – very small. HIV+ respondents who learned their status bought 2 more condoms