# ECON42720 Causal Inference and Policy Evaluation

## 3 Potential Outcomes and Randomised Experiments

Ben Elsner (UCD)

# About this Lecture

This lecture introduces the **potential outcomes framework**

- This framework allows us to think about **causal effects in a structured way**
- We learn why **counterfactuals are important** but also **difficult to observe/construct**
- We learn what biases can occur when we **do not observe all confounders**

One **solution for causal inference: randomised experiments**

- Under what conditions do they allow for the **identification of a causal effect?**
- How can a hypothetical experiment serve as a **benchmark for other methods?**

# Lingo and Notation

The **lingo of causal inference** is borrowed from **medical trials**

- **treatment** is the intervention/variable whose effect we are interested in
- **treatment group** is the group of units that receives the treatment
- **control group** is the group of units that does not receive the treatment or receives a placebo
- **outcome** is the variable that is potentially affected by the treatment

**We are after the causal effect**: **treatment** $(D)$ $\rightarrow$ **outcome** $(Y)$

# Resources

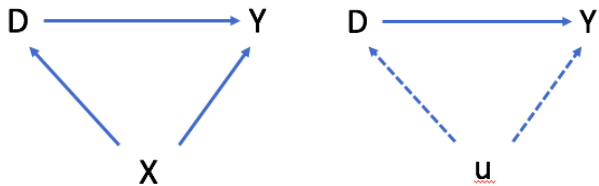
**This lecture is based on**

- Cunningham (2020), Chapter 4
- Angrist & Pischke (2009), Chapter 2

Find more about the course on the **course page**

# Causality

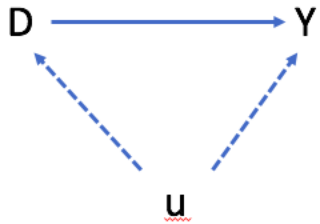
A common challenge in applied econometrics is to **separate a causal effect** from the **influence of third factors**




We often have a **good (theoretical) idea** why $D \to Y$

- but $X$ **is a confounding factor**
- often the problem: we also have an idea what $X$ could be...
- but **cannot observe it** (notation: $u$ for "unobservable")

# Causality




Common challenge: **selection into treatment**

In microeconomics we learn

- **people** make **rational choices**...
- ...as do **firms**
- ...as do **governments**

Problem: we do **not observe all the determinants** of these choices (i.e. $u$)

# Causality

Examples for **selection into treatment**:

**Going to the gym makes you healthier**

- good reason to believe so
- but people who go to the gym are different from those who don't
- observed correlation $\neq$ causation

**Exporting boosts firm profitability**

- good reason to believe so
- but exporters are different in many ways from non-exporters
- observed correlation $\neq$ causation

# Causality

Problem: it is **not easy to account for all confounding factors**

- because we know what they are but can't observe them
- or because we don't know what they are (e.g. "common shocks' ')

The **burden of proof is on the researcher**

- If you make a causal statement, you need to make a convincing case
- but causality cannot be proven without assumptions...
- ...and these assumptions need to be believed

# The Potential Outcomes Framework

We will now focus on **binary treatments**

Some **notation from experimental studies**

- $i$ is an index for the units in the population under study.

- $D_i$ is the **treatment status**:
- $D_i = 1$ if unit $i$ has been exposed to treatment,
- $D_i = 0$ if unit $i$ has not been exposed to treatment.

- $Y_i(D_i)$ indicates the **potential outcome according to treatment**:
- $Y_i(1) \equiv Y_{i1}$ is the outcome in case of treatment,
- $Y_i(0) \equiv Y_{i0}$ is the outcome in case of no treatment.

# The Potential Outcomes Framework (Rubin, 1974)

The **observed outcome** is then

$$Y_i = \begin{cases} Y_{i1} & \text{if } D_i = 1 \\ Y_{i0} & \text{if } D_i = 0 \end{cases}$$

**Switching equation**

$$= D_i Y_{i1} + (1 - D_i) Y_{i0}$$

$$= Y_{i0} + \underbrace{(Y_{i1} - Y_{i0})}_{\text{treatment effect}} D_i$$

We are interested in the (individual) **treatment effect** $\Delta_i = Y_{i1} - Y_{i0}$.

# The Fundamental Problem of Causal Inference

**We cannot identify** the individual treatment effect $\Delta_i = Y_{i1} - Y_{i0}$

This is **logically impossible**

- we either observe that a unit was treated or not
- but never both treatment statuses at the same time

# Wanted: The Counterfactual

A **critical ingredient** to establish causality: the **counterfactual**

**What would have happened** to a unit if

- the treatment status $D \in \{0, 1\}$ was different?
- or, if $D$ is continuous, the treatment intensity was different?

For units that were treated, we want to know $Y_{i0} \mid D_i = 1$. For units that were untreated, we want to know $Y_{i1} \mid D_i = 0$.

**No counterfactual, no causal claim!**

# Finding a Counterfactual: The Average Treatment Effect (ATE)

Problem: **we cannot observe the counterfactual**

- at the **individual level**, the counterfactual is **entirely hypothetical**

**Statistical solution**: **Average Treatment Effect** (**ATE**) for a random unit

$$E(\Delta_i) = E(Y_{i1} - Y_{i0}) = E(Y_{i1}) - E(Y_{i0})$$

If a **random unit was treated**, the **expected difference in their outcome** is the ATE

# The Average Treatment Effect on the Treated (**ATT**)

The **Average Treatment Effect on the Treated** (**ATT**)

$$\begin{aligned} ATT = E(\Delta_i | D_i = 1) &= E(Y_{i1} - Y_{i0} | D_i = 1) \\ &= E(Y_{i1} | D_i = 1) - E(Y_{i0} | D_i = 1). \end{aligned}$$

**Interpretation**: Average difference in potential outcomes for those who were treated

The ATT is **useful for policy evaluation**: take units who were treated and think what their outcomes would be had they not been treated

# The Average Treatment Effect on the Untreated (ATU)

Similarly, we can define the **Average Treatment Effect on the Untreated** (**ATU**) as:

$$\begin{aligned} ATU = E(\Delta_i | D_i = 0) &= E(Y_{i1} - Y_{i0} | D_i = 0) \\ &= E(Y_{i1} | D_i = 0) - E(Y_{i0} | D_i = 0). \end{aligned}$$

**Interpretation**: the **average difference in potential outcomes** for those who were **not treated**

The ATU can (sometimes) be **useful for policy evaluation**: take units who were untreated and think what their outcomes would be had they been treated

# Treatment Effects: Numerical Example

| Person | D | PO Treated $Y^1$ | PO Untreated $Y^0$ | Causal Effect |
|---|---|---|---|---|
| 1 | 1 | 80 | 60 | 20 |
| 2 | 1 | 75 | 70 | 5 |
| 3 | 0 | 70 | 60 | 10 |
| 4 | 1 | 85 | 80 | 5 |
| 5 | 0 | 75 | 70 | 5 |
| 6 | 0 | 80 | 80 | 0 |
| 7 | 0 | 90 | 100 | -10 |
| 8 | 1 | 85 | 80 | 5 |

Here we observe the **potential outcomes** for each person. Some people are treated ($D = 1$), some are not ($D = 0$).

# The Average Treatment Effect

| Person | D | PO Treated $Y^1$ | PO Untreated $Y^0$ | Causal Effect |
|---|---|---|---|---|
| 1 | 1 | 80 | 60 | 20 |
| 2 | 1 | 75 | 70 | 5 |
| 3 | 0 | 70 | 60 | 10 |
| 4 | 1 | 85 | 80 | 5 |
| 5 | 0 | 75 | 70 | 5 |
| 6 | 0 | 80 | 80 | 0 |
| 7 | 0 | 90 | 100 | -10 |
| 8 | 1 | 85 | 80 | 5 |

The ATE is the average of the last column: $ATE = 5$

# The Average Treatment Effect on the Treated (ATT)

| Person | D | PO Treated $Y^1$ | PO Untreated $Y^0$ | Causal Effect |
|---|---|---|---|---|
| 1 | 1 | 80 | 60 | 20 |
| 2 | 1 | 75 | 70 | 5 |
| 3 | 0 | 70 | 60 | 10 |
| 4 | 1 | 85 | 80 | 5 |
| 5 | 0 | 75 | 70 | 5 |
| 6 | 0 | 80 | 80 | 0 |
| 7 | 0 | 90 | 100 | -10 |
| 8 | 1 | 85 | 80 | 5 |

The ATT is the average causal effect of those who were treated (green cells):
$ATT = 8.75$

# The Average Treatment Effect on the Untreated (ATU)

| Person | D | PO Treated $Y^1$ | PO Untreated $Y^0$ | Causal Effect |
|---|---|---|---|---|
| 1 | 1 | 80 | 60 | 20 |
| 2 | 1 | 75 | 70 | 5 |
| 3 | 0 | 70 | 60 | 10 |
| 4 | 1 | 85 | 80 | 5 |
| 5 | 0 | 75 | 70 | 5 |
| 6 | 0 | 80 | 80 | 0 |
| 7 | 0 | 90 | 100 | -10 |
| 8 | 1 | 85 | 80 | 5 |

The ATU is the average causal effect of the units who were not treated (green cells):
$ATU = 1.25$

Here we have $ATT > ATE > ATU$, which may indicate **selection into treatment**: those who benefit most from the treatment are most likely to take it

# ATE/ATT/ATU: Which Parameter is most Relevant?

There is **no clear answer to this question**

**ATE is the most general parameter**

- what if we give the average person/firm/unit a treatment
- This is interesting for medical trials and for many policy questions

**ATT is often interesting for policy evaluation**

- take those who took up the policy
- what would have happened to them if they had not taken up the policy

**ATU is sometimes interesting for policy evaluation**

- We may be concerned about the people who did not take up the policy
- How would they be affected if they took up the policy

But: We **need all three to interpret typical estimation results**

# Comparison by Treatment Status

In many applications, we want to **estimate the ATE or the ATT**

But we **only observe**

- Whether a unit was treated or not
- The actual outcome of the unit

Solution (?): **comparison of means/simple difference in outcomes**, which can be estimated from two samples of data (treatment and control group)

$$
\begin{aligned}
SDO &= E(Y_i|D_i = 1) - E(Y_i|D_i = 0) \\
&= \underbrace{\frac{1}{N_T}\sum_i (y_i \mid d_i = 1)}_{\text{Treated units}} - \underbrace{\frac{1}{N_C}\sum_i (y_i \mid d_i = 0)}_{\text{"Control" units (untreated)}}
\end{aligned}
$$

# Simple Difference in Outcomes (SDO)

| Person | D | PO Treated $Y^1$ | PO Untreated $Y^0$ | Causal Effect |
|---|---|---|---|---|
| 1 | 1 | 80 | 60 | 20 |
| 2 | 1 | 75 | 70 | 5 |
| 3 | 0 | 70 | 60 | 10 |
| 4 | 1 | 85 | 80 | 5 |
| 5 | 0 | 75 | 70 | 5 |
| 6 | 0 | 80 | 80 | 0 |
| 7 | 0 | 90 | 100 | -10 |
| 8 | 1 | 85 | 80 | 5 |

We compare here the **observed outcomes** of the treated to the **observed outcomes** of the untreated

$SDO = 81.25 - 77.5 = 3.75$

# Comparison by Treatment Status

Problem with SDO: **treated and untreated units are not comparable**

$$
\begin{aligned}
E(Y_i|D_i=1) &- E(Y_i|D_i=0) \\
&= E(Y_{i1}|D_i=1) - E(Y_{i0}|D_i=0) \\
&= \underbrace{E(Y_{i1}|D_i=1) - E(Y_{i0}|D_i=1)}_{ATT} + \underbrace{E(Y_{i0}|D_i=1) - E(Y_{i0}|D_i=0)}_{\text{Selection bias}}
\end{aligned}
$$

**Selection bias**: the potential outcomes would differ even if both groups were untreated

Note how we add and subtract $E(Y_{i0}|D_i=1)$ here

# Now Suppose We Want to Estimate the ATE

Note that we can write the ATE as: $ATE = \pi ATT + (1 - \pi)ATU$

The **simple difference in outcomes yields**:

$$\begin{aligned} E(Y_i|D_i = 1) - E(Y_i|D_i = 0) &= ATE \\ &+ \underbrace{E(Y_{i0}|D_i = 1) - E(Y_{i0}|D_i = 0)}_{\text{Selection bias}} \\ &+ \underbrace{(1 - \pi)\,[ATT - ATU]}_{\text{HTE Bias}} \end{aligned}$$

Our estimate of the **ATE is contaminated by two biases**:

- **selection bias**
- **heterogeneous treatment effects (HTE) bias**

# Selection Bias: Genuine Difference between Treated and Untreated Units

| Person | D | PO Treated $Y^1$ | PO Untreated $Y^0$ | Causal Effect |
|---|---|---|---|---|
| 1 | 1 | 80 | 60 | 20 |
| 2 | 1 | 75 | 70 | 5 |
| 3 | 0 | 70 | 60 | 10 |
| 4 | 1 | 85 | 80 | 5 |
| 5 | 0 | 75 | 70 | 5 |
| 6 | 0 | 80 | 80 | 0 |
| 7 | 0 | 90 | 100 | -10 |
| 8 | 1 | 85 | 80 | 5 |

The **selection bias** is the **difference between the green and red cells**.

In this case, $E(Y_{i0}|D_i = 1) - E(Y_{i0}|D_i = 0) = 72.5 - 77.5 = -5$

So in absence of the treatment, the untreated have more favorable outcomes than the treated

# Selection Bias

**Prince Charles**

Male
Born in 1948
Raised in the UK
Married Twice
Lives in a castle
Wealthy and Famous

**Ozzy Osbourne**

Male
Born in 1948
Raised in the UK
Married Twice
Lives in a castle
Wealthy and Famous

Source: somewhere on X, before 2023

# Selection Bias: Genuine Difference between Treated and Untreated Units

The **selection bias** is $E(Y_{i0}|D_i = 1) - E(Y_{i0}|D_i = 0)$

This means that, even **without the treatment**, the **two groups would have different outcomes**

Selection bias is very common! Whenever there is a **confounder, we have selection bias**

# Heterogeneous Treatment Effects

$$\text{HTE bias} = (1 - \pi)\,[ATT - ATU]$$

**Treatment effect may vary across units**

- People who take the treatment may be different from those who do not
- People with **stronger treatment effects** may be **more likely** to take the treatment

**Heterogeneous treatment effects bias** can even occur in a randomised experiment

- We can **hardly ever enforce that people take the treatment**
- Unlike in physics or chemistry, in the social sciences we can only **encourage people to take the treatment**
- We speak here of **imperfect compliance**

# Heterogeneous Treatment Effects Bias

Source: Dall-E, OpenAI

# Heterogeneous Treatment Effects Bias

| Person | D | PO Treated $Y^1$ | PO Untreated $Y^0$ | Causal Effect |
|---|---|---|---|---|
| 1 | 1 | 80 | 60 | 20 |
| 2 | 1 | 75 | 70 | 5 |
| 3 | 0 | 70 | 60 | 10 |
| 4 | 1 | 85 | 80 | 5 |
| 5 | 0 | 75 | 70 | 5 |
| 6 | 0 | 80 | 80 | 0 |
| 7 | 0 | 90 | 100 | -10 |
| 8 | 1 | 85 | 80 | 5 |

The **heterogeneous treatment effects bias** is the difference between the green and red cells weighted by the **probability of treatment**

$$\text{HTE bias} = (1-\pi)\left[ATT - ATU\right] = 0.5 \times [8.75 - 1.25] = 3.75$$

# SDO vs. ATT

In our example, $SDO = 3.75$

$$\begin{aligned} SDO &= ATT + \text{Selection Bias} \\ &= 8.75 - 5 = 3.75 \end{aligned}$$

In our case, the **SDO under-estimates the ATT** by 5 because of selection bias. Without the treatment, treated units have less favourable outcomes than untreated units.

# SDO vs ATE

$$
\begin{array}{rcrcrcl}
SDO & = & ATE & + & \text{Selection Bias} & + & \text{HTE Bias} \\
 & = & 5 & - & 5 & + & 3.75 \\
 & = & 3.75 & & & &
\end{array}
$$

In our case, the SDO also under-estimates the ATE. This is the result of two biases going in opposite directions. The overall bias is negative.

# Putting it all together

The **previous slides** have shown that

- a **simple (a.k.a. naive) comparison of treated and untreated** units yields a **biased estimate** of the ATE and ATT
- It is **not random** whether a unit is treated or not or **chooses to get treated or not**
- This choice leads to **selection bias** and **heterogeneous treatment effects bias**
- A **randomised experiment** can **eliminate selection bias** but **not necessarily the HTE bias**

# So when Is the SDO Unbiased?

The SDO is unbiased ($SDO = ATE$) under the **Independence Assumption**

$$(Y_1, Y_0) \perp D \tag{1}$$

**The potential outcomes are independent of the treatment**

- Treated and untreated individuals have the same average potential outcomes $E(Y_0)$ and $E(Y_1)$
- I.e. they are statistically similar
- And no one selects into treatment because of their treatment effect

# So when Is the SDO Unbiased?

We can show that the **selection bias and HTE bias are zero under the independence assumption**

The **independence assumption implies**

$$E[Y_1 \mid D = 1] - E[Y_1 \mid D = 0] = 0$$
$$\underbrace{E[Y_0 \mid D = 1] - E[Y_0 \mid D = 0]}_{\text{Selection bias}} = 0$$

**Independence also implies that** $ATU = ATT$

$$\begin{aligned} ATT - ATU &= E[Y_1 \mid D = 1] - E[Y_0 \mid D = 1] \\ &\quad - E[Y_1 \mid D = 0] + E[Y_0 \mid D = 0] \\ &= 0 \end{aligned}$$

# Implication for research practice:

We need to **understand the sources of bias** in our context

We **cannot test directly for the presence of bias**

- no matter how much big data enthusiasts make us believe that we can

We need to be humble about **what parameters we can identify with our data** - Example: ATT does not suffer from HTE bias, but is it relevant?

# Comparison by Treatment Status

In most cases, **we run regressions** rather than comparing means

$$\begin{array}{ccccccc} Y_i = & \alpha & + & \beta & D_i & + & \varepsilon_i \\ & = & & = & & & = \\ & E(Y_{i0}) & & (Y_{i1} - Y_{i0}) & & & Y_{i0} - E(Y_{i0}) \end{array}$$

Take expectations conditional on $D_i$:

$$\begin{aligned} E(Y_i|D_i = 1) &= \alpha + \beta + E(\varepsilon_i|D_i = 1) \\ E(Y_i|D_i = 0) &= \alpha + E(\varepsilon_i|D_i = 0) \end{aligned}$$

# Comparison by Treatment Status

**Your regression will estimate a coefficient $\hat{\beta}$ that is a combination of a treatment effect (ATE or ATT) plus a bias**

The OLS estimate of $\beta$ is

$$\begin{aligned} E(Y_i|D_i=1) &- E(Y_i|D_i=0) \\ &= \underbrace{\beta}_{\text{Treatment Effect}} + \underbrace{E(\varepsilon_i|D_i=1) - E(\varepsilon_i|D_i=0)}_{\text{bias}}. \end{aligned}$$

The **RHS is either**

- ATT + Selection Bias
- ATE + Selection Bias + HTE Bias
- Both are equivalent

We should **always assume that the bias** $\neq 0$

# What if We Control for $X$

| **Person** | **D** | Female (**X**) | **PO Treated** $Y^1$ | **PO Untreated** $Y^0$ | **Causal Effect** |
|---|---|---|---|---|---|
| 1 | 1 | 1 | 80 | 60 | 20 |
| 2 | 1 | 1 | 75 | 70 | 5 |
| 3 | 0 | 1 | 70 | 60 | 10 |
| 4 | 1 | 0 | 85 | 80 | 5 |
| 5 | 0 | 1 | 75 | 70 | 5 |
| 6 | 0 | 0 | 80 | 80 | 0 |
| 7 | 0 | 0 | 90 | 100 | -10 |
| 8 | 1 | 0 | 85 | 80 | 5 |

# Controlling for $X$

If $X$ is binary, we **compare the outcomes conditional on $X$**

This means that

- We **compare the outcomes** of treated and untreated units **within each group** defined by $X$
- Our estimator is the **weighted average of these differences**

Consider $p = Pr(x = 1)$ and $q = Pr(x = 0)$, with $p + q = 1$

$$
\begin{aligned}
E(Y_i|D_i = 1, X) \quad &- \quad E(Y_i|D_i = 0, X) \\
&= \quad p\left[E(Y_{i1} \mid D_i = 1, x = 1) - E(Y_{i1} \mid D_i = 0, x = 1)\right] \\
&+ \quad q\left[E(Y_{i1} \mid D_i = 1, x = 0) - E(Y_{i1} \mid D_i = 0, x = 0)\right] \\
&= \quad p\left[ATT_{x=1} + E(Y_{i0} \mid D_i = 1, x = 1) - E(Y_{i0} \mid D_i = 0, x = 1)\right] \\
&+ \quad q\left[ATT_{x=0} + E(Y_{i0} \mid D_i = 1, x = 0) - E(Y_{i0} \mid D_i = 0, x = 0)\right]
\end{aligned}
$$

# What if We Control for $X$

Suppose $X$ is binary, for example male and female

| Person | D | Female ($\mathbf{X}$) | PO Treated $Y^1$ | PO Untreated $Y^0$ | Causal Effect |
|---|---|---|---|---|---|
| 1 | 1 | 1 | 80 | 60 | 20 |
| 2 | 1 | 1 | 75 | 70 | 5 |
| 3 | 0 | 1 | 70 | 60 | 10 |
| 4 | 1 | 0 | 85 | 80 | 5 |
| 5 | 0 | 1 | 75 | 70 | 5 |
| 6 | 0 | 0 | 80 | 80 | 0 |
| 7 | 0 | 0 | 90 | 100 | -10 |
| 8 | 1 | 0 | 85 | 80 | 5 |

- $SDO_{X=1} = 77.5 - 65 = 12.5$
- $SDO_{X=0} = 85 - 90 = -5$
- $SDO = \frac{1}{2} \times 12.5 + \frac{1}{2} \times (-5) = 3.75$

# Controlling for $X$

In our example, $E(D \mid X) = E(D)$, i.e. $D$ **is independent of** $X$

- This means that controlling for $X$ does not affect the estimate of the SDO

But often, $E(D \mid X) \neq E(D)$ because $X$ **is a confounder**

- That's exactly when we want to control for $X$
- If $X$ is the only confounder, **controlling for $X$ eliminates the selection bias**

We can identify the ATE under the **Conditional Independence Assumption**

$$(Y_1, Y_0) \perp D \mid X$$

# Randomised Experiments

**Experiments** provide a **clean way to eliminate selection bias**

Even if in most cases it is **not possible to run experiments**, a **clean experiment** serves as the **benchmark** for all the methods in this course

**Idea**: the closer we get to the **ideal experimental setting**, the closer we get to estimating a **causal effect**

# Randomised Experiments

**Basic idea**: **randomly assign treatment** across units

**Two conditions** have to be fulfilled:

- **assignment** to treatment and control group is **random**
- there is **full compliance** with the treatment (plus: no attrition)

In that case, the **treatment and control group** are **statistically identical**

# Randomised Experiments

**Formally** (T: treatment group, C: control group)

$$E\{Y_{i0}|i \in C\} = E\{Y_{i0}|i \in T\}$$

and

$$E\{Y_{i1}|i \in C\} = E\{Y_{i1}|i \in T\}.$$

- ... both groups have the **same potential outcomes** in expectation

# Randomised Experiments

Therefore, we can obtain an **unbiased and consistent estimate of the ATE**

$$E(\Delta_i) = E(Y_{i1} - Y_{i0}) = E\{Y_{i1}|i \in T\} - E\{Y_{i0}|i \in C\}.$$

$\Rightarrow$ **comparison of means** is meaningful (causal effect in expectation)

$\Rightarrow$ Randomisation solves the **fundamental problem of causal inference**

# Randomised Experiments

But **what if not everyone complies** with the treatment?

Examples

- Not every job seeker who is offered training takes part
- Not every person eligible for a medical card/social benefits/etc applies

$$E\{Y_{i1}|i \in T\} - E\{Y_{i0}|i \in C\} \neq ATE$$

As we will learn later in the course, **all is not lost**

- The resulting **treatment effect can still be meaningful**
- . . . as long as **treatment is random**

# Randomised Experiments

An **important assumption** underlying causal inference in experiments

**SUTVA**: **S**table **U**nit **T**reatment **V**alue **A**ssumption

In **plain English**

- treatment of one unit must not affect the potential outcomes of another
- i.e. no spillovers, general equilibrium effects, etc

# Randomised Experiments

**SUTVA** is an **untestable assumption**

Examples for **violations**:

- information leaks from treatment to control group
- large-scale job training programs $\Rightarrow$ GE effects
- health interventions (capacity constraints in medical facilities)

**Implications:**

- One reason why small interventions don't scale up
- Important to choose the right control group

# Cookbook for Analyzing a Randomised Experiment

**1) Explain experimental design** in detail, in particular

- How was the randomisation carried out
- Discuss compliance (or likely non-compliance)
- Argue why SUTVA holds

**2) Show balancing tests** based on **pre-treatment characteristics**

- Do treatment and control group differ before the experiment?
- Use pre-treatment outcomes if possible
- Use other pre-treatment characteristics that may predict selection into treatment
- If you use regression analysis in the paper, regress the pre-treatment x on the treatment
- Never use post-treatment outcomes (NEVER EVER!)

$$x_i = \delta_1 + \delta_2 D_i + \eta_i$$

# Cookbook for Analyzing a Randomised Experiment

**3) Show and discuss results**

- **Compare means** across treatment and control groups
- Add **pre-treatment characteristics** $X_i$ as controls

$$y_i = \alpha + \beta D_i + \boldsymbol{X_i}\boldsymbol{\gamma} + u_i$$

**Why add controls?**

- Additional test if randomisation worked ($\beta$ should not change)
- Estimates become more precise (less noise in the model $\Rightarrow$ lower SE)

# Example: Tennessee STAR Experiment

**Research Question**: **does class size matter** to student learning? **(Krueger, 1999)**

Implemented on cohort of kindergartners (i.e. senior infants) in 1985/86 in Tennessee.

Lasted 4 years, then everyone went back to regular size class.

**Three treatments**

1. Small class 13-17
2. Regular class 22-25
3. Regular class with Aide

# Example: Tennessee STAR Experiment

Schools had to have **at least 3 classes** to participate.

Entering cohort **randomly assigned to class type**. Teachers also randomly assigned.

I.e. randomisation was carried out **within schools**

Test scores measured towards **end of each school year** in March.

# Example: Tennessee STAR Experiment

**Balancing tests**: do treatment and control groups have the **same pre-treatment characteristics?**

TABLE 2.2.1
Comparison of treatment and control characteristics in the Tennessee STAR experiment

| Variable | Class Size | | | *P*-value for equality across groups |
|---|---|---|---|---|
| | Small | Regular | Regular/Aide | |
| Free lunch | .47 | .48 | .50 | .09 |
| White/Asian | .68 | .67 | .66 | .26 |
| Age in 1985 | 5.44 | 5.43 | 5.42 | .32 |
| Attrition rate | .49 | .52 | .53 | .02 |
| Class size in kindergarten | 15.10 | 22.40 | 22.80 | .00 |
| Percentile score in kindergarten | 54.70 | 48.90 | 50.00 | .00 |

*Notes*: Adapted from Krueger (1999), table I. The table shows means of variables by treatment status for the sample of students who entered STAR in kindergarten. The *P*-value in the last column is for the *F*-test of equality of variable means across all three groups. The free lunch variable is the fraction receiving a free lunch. The percentile score is the average percentile score on three Stanford Achievement Tests. The attrition rate is the proportion lost to follow-up before completing third grade.

- They are similar in age, race and SES
- Attrition rates are similar
- Treatment and outcomes differ

# Example: Tennessee STAR Experiment

Krueger (1999) uses a **regression to analyse the experiment**

The **most comprehensive specification** includes **pre-treatment characteristics** and **school fixed effects**

$$y_{is} = \alpha + \beta D_{is} + \boldsymbol{X}_{is}'\boldsymbol{\gamma} + \delta_s + u_{is}$$

**Why include fixed effects?**

- **Randomisation** was carried out **within schools**
- Children in different school types may **react differently to the treatment**
- The FE estimator compares children within the same school

# Example: Tennessee STAR Experiment

TABLE 2.2.2
Experimental estimates of the effect of class size on test scores

| Explanatory Variable | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| Small class | 4.82 | 5.37 | 5.36 | 5.37 |
| | (2.19) | (1.26) | (1.21) | (1.19) |
| Regular/aide class | .12 | .29 | .53 | .31 |
| | (2.23) | (1.13) | (1.09) | (1.07) |
| White/Asian | — | — | 8.35 | 8.44 |
| | | | (1.35) | (1.36) |
| Girl | — | — | 4.48 | 4.39 |
| | | | (.63) | (.63) |
| Free lunch | — | — | −13.15 | −13.07 |
| | | | (.77) | (.77) |
| White teacher | — | — | — | −.57 |
| | | | | (2.10) |
| Teacher experience | — | — | — | .26 |
| | | | | (.10) |
| Teacher Master's degree | — | — | — | −0.51 |
| | | | | (1.06) |
| School fixed effects | No | Yes | Yes | Yes |
| $R^2$ | .01 | .25 | .31 | .31 |

*Notes*: Adapted from Krueger (1999), table V. The dependent variable is the Stanford Achievement Test percentile score. Robust standard errors allowing for correlated residuals within classes are shown in parentheses. The sample size is 5,681.

# Example: Tennessee STAR Experiment

Result: **smaller classes are more effective**

**Experimental design appears valid**

- Balancing tables point to clean randomisation
- Estimates not affected by controls for pre-treatment characteristics

The **same cookbook approach** can be used for **non-experimental studies**

# Discrete vs. continuous treatment

So far, we discussed experiments with a **discrete treatment**

The **same assumptions apply** to experiments with a **continuous treatment**

- The **treatment intensity varies** across units
- treatment intensity is **randomly assigned**

Possibilites for researchers:

- **Discretise** and compare mean outcomes (e.g. above/below median)
- Estimate **marginal effect in a regression**

# Randomised Experiments as Templates

**Randomised experiments are often difficult or impossible** to conduct

But they serve as a **template for estimating causal effects** in non-experimental settings

Any study that **claims to estimate a causal effect** should

- explain what the treatment is
- and under what conditions the assignment of the treatment (intensity) is as good as random
- ...even if the world is not perfect, there is no harm thinking about the perfect

**My advice** (after 100+ referee reports, several published papers and many rejections): **ignore the experimental template at your own peril**

# References


Angrist, Joshua, & Pischke, Jörn-Steffen. 2009. *Mostly Harmless Econometrics - An Empiricist's Companion*. Princeton University Press.

Cunningham, Scott. 2020. *Causal Inference: The Mixtape*. Yale University Press.

Krueger, Alan B. 1999. Experimental Estimates of Education Production Functions. *The Quarterly Journal of Economics*, **114**(2), 497–532.

Rubin, Donald. 1974. Estimating Causal Effects of Treatments in Randomized and Nonrandomized Studies. *Journal of Educational Psychology*, **66**(5).

# Appendix

# Group Work I

Suppose you want to study the impact of medical cards on health outcomes in Ireland. Medical cards are means-tested and given to people with low income. Card holders can access medical care for free. You have access to administrative data covering all adults in Ireland, including information on whether they ever held a medical card and their health outcomes.

Tasks:

1. Explain: What is the treatment in this case?
2. Intuitively, explain what the ATE, ATU, ATT are. Which of these parameters is most relevant in a policy evaluation in this context?
3. What are potential sources of bias here?

# Group Work II

Consider the independence assumption

$$(Y_1, Y_0) \perp D$$

Evaluate the following statements (i.e. are they implied by the independence assumption)?

1. $E[Y_1 \mid D = 1] - E[Y_0 \mid D = 0] = 0$
2. $E[Y_1 \mid D = 1] - E[Y_0 \mid D = 1] = 0$
3. $E[Y_1 \mid D = 1] - E[Y_1 \mid D = 0] = 0$

# Group Work III

Suppose you want to study the impact of local police presence on crime in Dublin.

Questions:

- What would be the ideal experiment to run here?
- Why may this experiment not be feasible?
- If you can't run an experiment, what else could you do to get at the causal effect?

# Derivation of SDO Decomposition

$$
\begin{aligned}
&E(Y^1 \mid D=1) - E(Y^0 \mid D=0) \\
&\quad = E(Y^1 \mid D=1) - E(Y^0 \mid D=1) + E(Y^0 \mid D=1) - E(Y^0 \mid D=0) \\
&\quad = \underbrace{E(Y^1 \mid D=1) - E(Y^0 \mid D=1)}_{\text{ATT}} + \underbrace{E(Y^0 \mid D=1) - E(Y^0 \mid D=0)}_{\text{Selection Bias}}
\end{aligned}
$$

# Derivation of SDO Decomposition

Now consider $ATE = \pi ATT + (1 - \pi)ATU$

$$\begin{aligned} ATE &= \pi ATT + (1 - \pi)ATU \\ &= \pi ATT + ATT - ATT + (1 - \pi)ATU \end{aligned}$$

Re-arranging yields

$$\begin{aligned} ATT &= ATE + (1 - \pi)ATT - (1 - \pi)ATU \\ &= ATE + (1 - \pi)[ATT - ATU] \end{aligned}$$

# Derivation of SDO Decomposition

We can plug this into the SDO decomposition and get

$$
\begin{aligned}
E(Y^1 \mid D=1) - E(Y^0 \mid D=0) & \\
&= ATE \\
&+ \underbrace{E(Y^0 \mid D=1) - E(Y^0 \mid D=0)}_{\text{Selection Bias}} \\
&+ \underbrace{(1-\pi)[ATT - ATU]}_{\text{HTE Bias}}
\end{aligned}
$$

benjamin.elsner@ucd.ie

www.benjaminelsner.com

Sign up for office hours

YouTube Channel

@ben_elsner

LinkedIn

# Contact

**Prof. Benjamin Elsner**
University College Dublin
School of Economics
Newman Building, Office G206
benjamin.elsner@ucd.ie

Office hours: book on Calendly