Causality, potential outcomes and research design

Prottoy A. Akbar

Principles of Empirical Analysis (ECON-A3000)
Lecture 4

Course outline and learning objectives

- Data, descriptive statistics and causality
  1. introduction, data
  2. samples and descriptive statistics
  3. conditional descriptive statistics
  4. **today: causality and research design**
  5. statistical inference and randomization
  6. revealed preferences in observed data
- Quasi-experimental methods

Course outline and learning objectives

- Data, descriptive statistics and causality
  1. introduction, data
  2. samples and descriptive statistics
  3. conditional descriptive statistics
  4. **today: causality and research design**
  5. statistical inference and randomization
  6. revealed preferences in observed data
- Quasi-experimental methods

Todays' learning objectives:

- Good understanding of:
  1. causality
  2. counterfactual
  3. potential outcomes
  4. treatment effect
  5. selection bias
- Good understanding of why randomization eliminates selection bias

Causal questions

- Thus far, we have focused on **descriptive** questions (lectures 1–3)
  - aim: measure the actual state of the world
  - "what is joint distribution of X and Y?"

Causal questions

- Thus far, we have focused on **descriptive** questions (lectures 1–3)
  - aim: measure the actual state of the world
  - "what is joint distribution of X and Y?"
- We would often need to evaluate the **impact** of X on Y, e.g.
  - education on earnings
  - marketing campaign on sales
  - carbon tax on emissions
  - R&D subsidy on innovation
  - fiscal stimulus on unemployment

Causal questions

- Thus far, we have focused on **descriptive** questions (lectures 1–3)
  - aim: measure the actual state of the world
  - "what is joint distribution of X and Y?"
- We would often need to evaluate the **impact** of X on Y, e.g.
  - education on earnings
  - marketing campaign on sales
  - carbon tax on emissions
  - R&D subsidy on innovation
  - fiscal stimulus on unemployment
- These are **causal** questions
  - aim: compare *counterfactual* states of the world
  - "how would Y change if we changed X?"
    - we typically refer to Y as "outcome" and to X as "treatment"

- Are almost impossible to observe for any single individual/entity!
  - where everything else is the same except the treatment ("ceteris paribus")
  - sometimes possible with highly controlled lab experiments in the natural sciences
  - but harder when studying people.

Source: Mastering Econometrics
from pre-class assignment 2

Counterfactual states of the world

- Are almost impossible to observe for any single individual/entity!
  - where everything else is the same except the treatment ("ceteris paribus")
  - sometimes possible with highly controlled lab experiments in the natural sciences
  - but harder when studying people.
- But we can often find counterfactuals for the average person in a sample
  - important to consider issues of sampling errors and representativeness (of wider population)

Source: Mastering Econometrics
from pre-class assignment 2

Identifying causal relationships via experiments

- Today's lecture will focus on answering causal questions using experimental designs
  - how to design comparisons to test for causality in observed relationships
  - the simplest context for learning relevant statistical concepts
- It is often helpful to ask:
  what would be the **ideal experiment** for answering this question?
  - Even when we can't run an experiment

Identifying causal relationships via experiments

- Today's lecture will focus on answering causal questions using experimental designs
  - how to design comparisons to test for causality in observed relationships
  - the simplest context for learning relevant statistical concepts
- It is often helpful to ask:
  what would be the **ideal experiment** for answering this question?
  - Even when we can't run an experiment
- helpful benchmark for "naturally occurring" or "quasi" experiments
  - we'll discuss examples of "natural experiment" involving actual randomization already next week
  - you'll see other types of quasi-experimental approaches in lectures 7–12

Elements of causal questions

1. **Treatment**
   - impact *of* [...]
2. **Counterfactual**
   - impact *in comparison to* [...]
3. **Outcome**
   - impact *on* [...]
4. **Population**
   - impact *for* [...]

Elements of causal questions

1. **Treatment**
   - impact *of* [...]
2. **Counterfactual**
   - impact *in comparison to* [...]
3. **Outcome**
   - impact *on* [...]
4. **Population**
   - impact *for* [...]

- Think of a causal question!
  - Hold it in your head or write it down. We will revisit.

Elements of causal questions

1. **Treatment**
   - impact *of* [...]
2. **Counterfactual**
   - impact *in comparison to* [...]
3. **Outcome**
   - impact *on* [...]
4. **Population**
   - impact *for* [...]

- Think of a causal question!
  - Hold it in your head or write it down. We will revisit.
- What is the causal question in the paper you read for the pre-class assignment?

Elements of causal questions

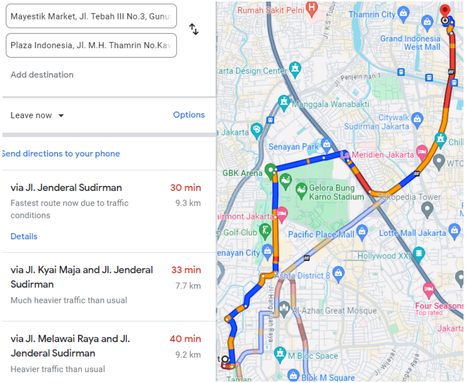
1. **Treatment**
   - impact *of* [...]
2. **Counterfactual**
   - impact *in comparison to* [...]
3. **Outcome**
   - impact *on* [...]
4. **Population**
   - impact *for* [...]

- Think of a causal question!
  - Hold it in your head or write it down. We will revisit.
- What is the causal question in the paper you read for the pre-class assignment?
  - my quick take: what is the impact of **Jakarta's high-occupancy vehicle restriction** in comparison to **unrestricted road travel** on **drivers'** **travel times**?

The causal question in Hanna et al. (2017)

What is the impact of **Jakarta's high-occupancy vehicle restriction** in comparison to **unrestricted road travel** on **drivers'** **travel times**?

- Treatment: Lifting of the HOV restriction
- Counterfactuals:
  1. State of the world just before the treatment
  2. Google's prediction under "typical traffic conditions"

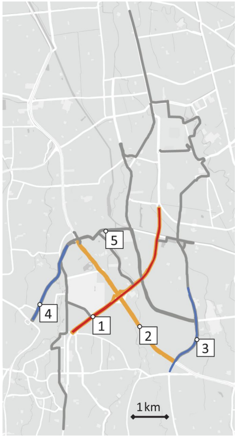


Source: Google Maps

The causal question in Hanna et al. (2017)

What is the impact of **Jakarta's high-occupancy vehicle restriction** in comparison to **unrestricted road travel** on **drivers'** **travel times**?

- Treatment: Lifting of the HOV restriction
- Counterfactuals:
  1. State of the world just before the treatment
  2. Google's prediction under "typical traffic conditions"
- Population: **drivers on trips that might use these routes**

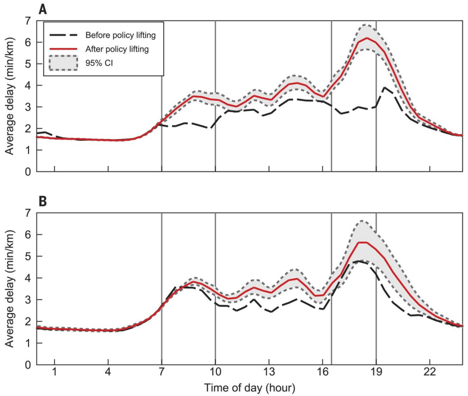


Source: Rema Hanna et al., Citywide effects of high-occupancy vehicle restrictions: Evidence from "three-in-one" in Jakarta. *Science* 357, 89-93 (2017).

The causal question in Hanna et al. (2017)

What is the impact of **Jakarta's high-occupancy vehicle restriction** in comparison to **unrestricted road travel** on **drivers' travel times**?

- Treatment: Lifting of the HOV restriction
- Counterfactuals:
  1. State of the world just before the treatment
  2. Google's prediction under "typical traffic conditions"
- Population: **drivers on trips that might use these routes**
- Outcome: the delay per km travelled

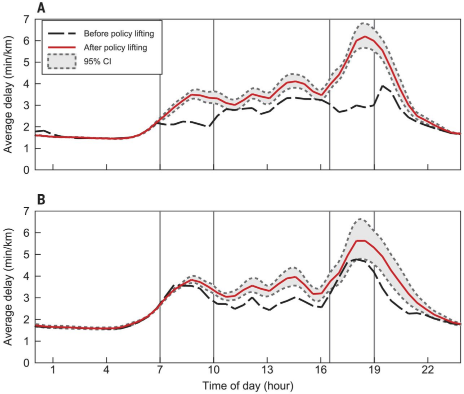


Source: Rema Hanna et al., Citywide effects of high-occupancy vehicle restrictions: Evidence from "three-in-one" in Jakarta. *Science* 357, 89-93 (2017).

The causal question in Hanna et al. (2017)

What is the impact of **Jakarta's high-occupancy vehicle restriction** in comparison to **unrestricted road travel** on **drivers'** **travel times**?

- Treatment: Lifting of the HOV restriction
- Counterfactuals:
  1. State of the world just before the treatment
  2. Google's prediction under "typical traffic conditions"
- Population: **drivers on trips that might use these routes**
- Outcome: the delay per km travelled
- Just one example of a well-defined question,
  - there are many others (even in this specific context)
- Next: formal definitions using the potential outcomes framework



Source: Rema Hanna et al., Citywide effects of high-occupancy vehicle restrictions: Evidence from "three-in-one" in Jakarta. *Science* 357, 89-93 (2017).

Potential Outcomes

- We focus on binary (0/1) treatments and denote **treatment status** of individual i as

$$D_i = \begin{cases} 1 & \text{if she receives the treament} \\ 0 & \text{if she doesn't} \end{cases}$$

Potential Outcomes

- We focus on binary (0/1) treatments and denote **treatment status** of individual i as

$$D_i = \begin{cases} 1 & \text{if she receives the treament} \\ 0 & \text{if she doesn't} \end{cases}$$

- We denote **outcomes** by y and define

$$\text{potential outcome} = \begin{cases} y_{1i} & \text{if } D_i = 1 \\ y_{0i} & \text{if } D_i = 0 \end{cases}$$

Potential Outcomes

- We focus on binary (0/1) treatments and denote **treatment status** of individual i as

$$D_i = \begin{cases} 1 & \text{if she receives the treament} \\ 0 & \text{if she doesn't} \end{cases}$$

- We denote **outcomes** by y and define

$$\text{potential outcome} = \begin{cases} y_{1i} & \text{if } D_i = 1 \\ y_{0i} & \text{if } D_i = 0 \end{cases}$$

in words: y_{1i} is the outcome of individual i in the state of the world where she is treated and y_{0i} is her outcome in the state of the world where she was *not* treated

Only one potential outcome can occur

Two roads diverged in a yellow wood,
And sorry I could not travel both
And be one traveler, long I stood
And looked down one as far as I could
To where it bent in the undergrowth;

...

I shall be telling this with a sigh
Somewhere ages and ages hence:
Two roads diverged in a wood, and I—
I took the one less traveled by,
And that has made all the difference.

Robert Frost (1915): The Road Not Taken

Robert Lee Frost (1874–1963) was an American poet, who frequently wrote about settings from rural life, using them to examine complex social and philosophical themes. Source: Wikipedia

Treatment effect

- The treatment effect for individual i is:

$$y_{1i} - y_{0i}$$

in words: difference in the potential outcomes with and without the treatment

Treatment effect

- The treatment effect for individual i is:

$$y_{1i} - y_{0i}$$

  in words: difference in the potential outcomes with and without the treatment

- The fundamental challenge of causal inference is that we cannot observe both y_{1i} and y_{0i} for the same individual. Instead, we observe

$$y_i = \begin{cases} y_{1i} & \text{if } D_i = 1 \\ y_{0i} & \text{if } D_i = 0 \end{cases}$$

Average treatment effect

- We can never identify the treatment effect for an individual person, but sometimes we can estimate average treatment effects:

$$\text{Average treatment effect (ATE)} \quad = \quad \mathbb{E}\left[y_{1i} - y_{0i}\right]$$

Average treatment effect

- We can never identify the treatment effect for an individual person, but sometimes we can estimate average treatment effects:

$$\begin{aligned} \text{Average treatment effect (ATE)} &= \mathbb{E}\left[y_{1i} - y_{0i}\right] \\ \text{ATE for the treated (ATT)} &= \mathbb{E}\left[y_{1i} - y_{0i} | D_i = 1\right] \end{aligned}$$

where $\mathbb{E}\left[a|b\right]$ is the expectation of a conditional on b

Average treatment effect

- We can never identify the treatment effect for an individual person, but sometimes we can estimate average treatment effects:

$$\begin{aligned}\text{Average treatment effect (ATE)} &= \mathbb{E}\left[y_{1i} - y_{0i}\right] \\ \text{ATE for the treated (ATT)} &= \mathbb{E}\left[y_{1i} - y_{0i} \middle| D_i = 1\right]\end{aligned}$$

  where $\mathbb{E}\left[a|b\right]$ is the expectation of a conditional on b

- Why ATE *and* ATT?
  - treatment effect may be different for those getting the treatment than it would be for those not getting it
  - internal validity: do we learn the true effect for the treated population?
  - external validity: can we extrapolate to other populations?

Research designs and control groups

- We use a comparison or **control group** to approximate what would have happened to the treated in the absence of the treatment
  - that is, we *estimate* the counterfactual $\mathbb{E}[y_{0i}|D_i = 1]$

Research designs and control groups

- We use a comparison or **control group** to approximate what would have happened to the treated in the absence of the treatment
  - that is, we *estimate* the counterfactual $\mathbb{E}[y_{0i}|D_i = 1]$
- In economics parlance, this approach is know as "**design-based**" or "**experimental**" approach
  - the alternative often is a "structural" approach, where we use quantitative economic models to simulate counterfactual states of the world

Research designs and control groups

- We use a comparison or **control group** to approximate what would have happened to the treated in the absence of the treatment
  - that is, we *estimate* the counterfactual $\mathbb{E}[y_{0i}|D_i = 1]$
- In economics parlance, this approach is know as "**design-based**" or "**experimental**" approach
  - the alternative often is a "structural" approach, where we use quantitative economic models to simulate counterfactual states of the world
- Invalid control group leads to **selection bias**
  - whether the control group provides a good counterfactual or not is the key question of all design-based causal inference

How to find a control group in Hanna et al. (2017)?

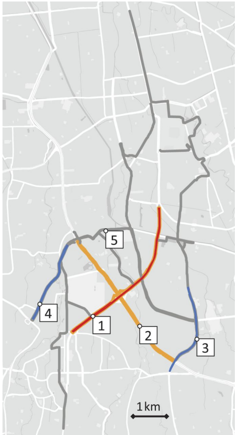


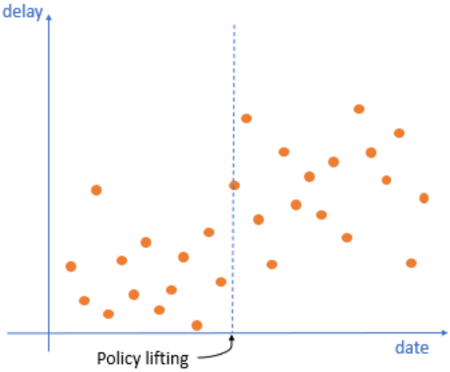
Fig. 1 Routes included in the analysis.

(1) Former three-in-one road (Jalan Sudirman, red and orange). (2) Former three-in-one road (Jalan Gatot Subroto, orange). (3) Unrestricted alternate road (Jalan Rasuna Said, blue). (4) Unrestricted alternate road (Jalan Tentara Pelajar, blue). (5) Eight unrestricted alternate routes from the Jakarta Department of Transport (gray). Routes from the first phase of data collection are drawn with thin lines: 1 (red), 3 (blue), and 4 (blue). [Map data from Google, 2017]

Regression estimation

$$delay_{idh} = \alpha + \beta \cdot post_d + \gamma \cdot north_i + \varepsilon_{idh}$$

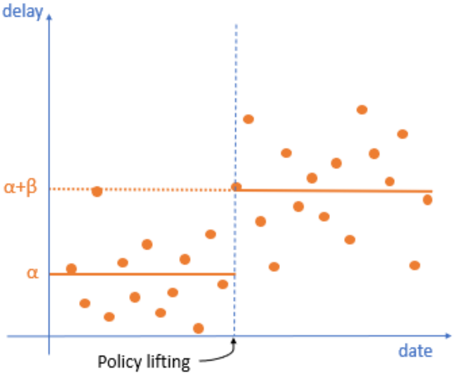
- Dependent/outcome variable: travel delay on segment i, on date d and departure hour h
- Independent/explanatory variable: indicator for whether date d is after the policy lifting
  - $post_d = 0$ before policy lifting ("control" group)
  - $post_d = 1$ after policy lifting ("treatment" group)
- Conditional on direction of travel



Regression estimation

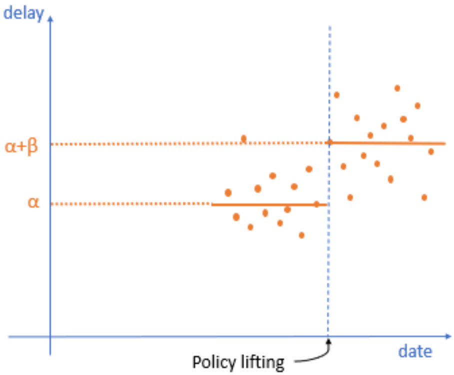
$$delay_{idh} = \alpha + \beta \cdot post_d + \gamma \cdot north_i + \varepsilon_{idh}$$

- Dependent/outcome variable: travel delay on segment i, on date d and departure hour h
- Independent/explanatory variable: indicator for whether date d is after the policy lifting
  - $post_d = 0$ before policy lifting ("control" group)
  - $post_d = 1$ after policy lifting ("treatment" group)
- Conditional on direction of travel



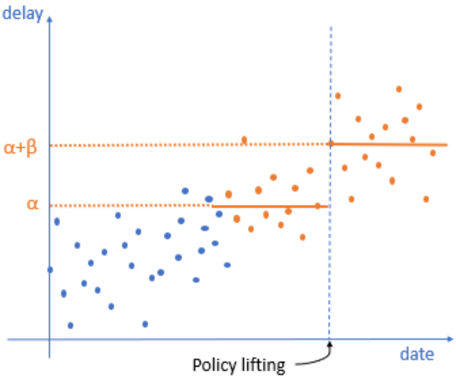
How good is the counterfactual?

- What if the timing of event is intended to coincide with the changes in outcomes?
  - as opposed to changes being caused by the treatment?
  - E.g. the delay would have occurred anyway (even in the absence of the policy lifting)
- What would outcomes have looked like in the absence of the policy?
  - Would the average delay have stayed at α?
  - Key assumption: Treated observations would resemble control observations in the absence of the treatment



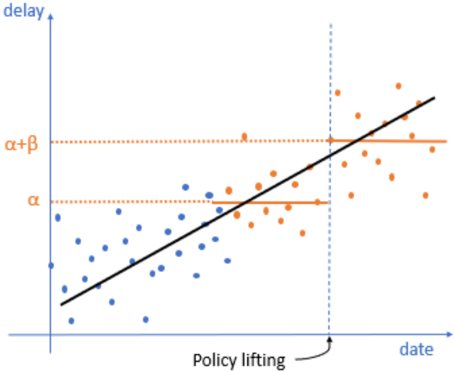
How good is the counterfactual?

- What if the timing of event is intended to coincide with the changes in outcomes?
  - as opposed to changes being caused by the treatment?
  - E.g. the delay would have occurred anyway (even in the absence of the policy lifting)
- What would outcomes have looked like in the absence of the policy?
  - Would the average delay have stayed at α?
  - Key assumption: Treated observations would resemble control observations in the absence of the treatment



How good is the counterfactual?

- What if the timing of event is intended to coincide with the changes in outcomes?
  - as opposed to changes being caused by the treatment?
  - E.g. the delay would have occurred anyway (even in the absence of the policy lifting)
- What would outcomes have looked like in the absence of the policy?
  - Would the average delay have stayed at α?
  - Key assumption: Treated observations would resemble control observations in the absence of the treatment



Selection Bias

- As the amount of data increases, the sample averages approach the population average (expectations)

$$\underbrace{Avg[y_i|D=1]}_{\text{treatment group}} - \underbrace{Avg[y_i|D=0]}_{\text{control group}} \rightarrow \mathbb{E}[y_i|D=1] - \mathbb{E}[y_i|D=0]$$

Selection Bias

- As the amount of data increases, the sample averages approach the population average (expectations)

$$\underbrace{Avg[y_i|D=1]}_{\text{treatment group}} - \underbrace{Avg[y_i|D=0]}_{\text{control group}} \rightarrow \mathbb{E}[y_i|D=1] - \mathbb{E}[y_i|D=0]$$

$$= \mathbb{E}[y_{1i}|D=1] - \mathbb{E}[y_{0i}|D=0]$$

Selection Bias

- As the amount of data increases, the sample averages approach the population average (expectations)

$$\underbrace{Avg[y_i|D=1]}_{\text{treatment group}} - \underbrace{Avg[y_i|D=0]}_{\text{control group}} \rightarrow \mathbb{E}[y_i|D=1] - \mathbb{E}[y_i|D=0]$$

$$= \mathbb{E}[y_{1i}|D=1] - \mathbb{E}[y_{0i}|D=0]$$

- Where the second row emphasizes that we observe y_{0i} only for the control group, while our objective is to estimate ATT, i.e.

$$\mathbb{E}\left[y_{1i} - y_{0i}|D_i = 1\right] = \mathbb{E}[y_{1i}|D=1] - \underbrace{\mathbb{E}[y_{0i}|D=1]}_{\text{never observed}}$$

Selection Bias

- As the amount of data increases, the sample averages approach the population average (expectations)

$$\underbrace{Avg[y_i|D=1]}_{\text{treatment group}} - \underbrace{Avg[y_i|D=0]}_{\text{control group}} \rightarrow \mathbb{E}[y_i|D=1] - \mathbb{E}[y_i|D=0]$$

$$= \mathbb{E}[y_{1i}|D=1] - \mathbb{E}[y_{0i}|D=0]$$

- Where the second row emphasizes that we observe y_{0i} only for the control group, while our objective is to estimate ATT, i.e.

$$\mathbb{E}\left[y_{1i} - y_{0i}|D_i = 1\right] = \mathbb{E}[y_{1i}|D=1] - \underbrace{\mathbb{E}[y_{0i}|D=1]}_{\text{never observed}}$$

- Selection bias arises when a control group leads to an incorrect estimate of the counterfactual, i.e. $\mathbb{E}[y_{0i}|D=0] \neq \mathbb{E}[y_{0i}|D=1]$

Selection Bias

- A particularly informative way to illustrate selection bias is:

$$\mathbb{E}[y_i|D=1] - \mathbb{E}[y_i|D=0] \quad = \quad \mathbb{E}[y_{1i}|D=1] - \mathbb{E}[y_{0i}|D=0]$$

Selection Bias

- A particularly informative way to illustrate selection bias is:

$$\begin{aligned}\mathbb{E}[y_i|D=1] - \mathbb{E}[y_i|D=0] &= \mathbb{E}[y_{1i}|D=1] - \mathbb{E}[y_{0i}|D=0] \\ &= \underbrace{\mathbb{E}[y_{1i}|D=1] - \mathbb{E}[y_{0i}|D=1]}_{\text{ATT}} + \\ &\quad \underbrace{\mathbb{E}[y_{0i}|D=1] - \mathbb{E}[y_{0i}|D=0]}_{\text{Selection bias}}\end{aligned}$$

Selection Bias

- A particularly informative way to illustrate selection bias is:

$$\begin{aligned}\mathbb{E}[y_i|D=1]-\mathbb{E}[y_i|D=0] &= \mathbb{E}[y_{1i}|D=1]-\mathbb{E}[y_{0i}|D=0] \\ &= \underbrace{\mathbb{E}[y_{1i}|D=1]-\mathbb{E}[y_{0i}|D=1]}_{\text{ATT}} + \\ &\quad \underbrace{\mathbb{E}[y_{0i}|D=1]-\mathbb{E}[y_{0i}|D=0]}_{\text{Selection bias}}\end{aligned}$$

where the first step is from the previous slide and the second step is taken by simply adding and substracting $\mathbb{E}[y_{0i}|D=1]$

  - i.e. $\mathbb{E}[y_{0i}|D=1]-\mathbb{E}[y_{0i}|D=1]=0$, so including it does not change the result, but allows us to rewrite the equation as ATT+SB

Selection Bias

- A particularly informative way to illustrate selection bias is:

$$\begin{aligned}\mathbb{E}[y_i|D=1]-\mathbb{E}[y_i|D=0] &= \mathbb{E}[y_{1i}|D=1]-\mathbb{E}[y_{0i}|D=0] \\ &= \underbrace{\mathbb{E}[y_{1i}|D=1]-\mathbb{E}[y_{0i}|D=1]}_{\text{ATT}} + \\ &\quad \underbrace{\mathbb{E}[y_{0i}|D=1]-\mathbb{E}[y_{0i}|D=0]}_{\text{Selection bias}}\end{aligned}$$

where the first step is from the previous slide and the second step is taken by simply adding and substracting $\mathbb{E}[y_{0i}|D=1]$

  - i.e. $\mathbb{E}[y_{0i}|D=1]-\mathbb{E}[y_{0i}|D=1]=0$, so including it does not change the result, but allows us to rewrite the equation as ATT+SB
  - in words: differences in the average outcomes between treatment and control groups include the treatment effect *and* the selection bias (the difference between the two groups if neither had been treated)

Randomization eliminates selection bias

- Random assignment into treatment/control ensures that the control groups is *comparable* to the treatment group

Randomization eliminates selection bias

- Random assignment into treatment/control ensures that the control groups is *comparable* to the treatment group
- Formally: their potential outcomes are in expectation the same, i.e.

$$\mathbb{E}[y_{1i}|D = 1] = \mathbb{E}[y_{1i}|D = 0]$$
$$\mathbb{E}[y_{0i}|D = 1] = \mathbb{E}[y_{0i}|D = 0]$$

Randomization eliminates selection bias

- Random assignment into treatment/control ensures that the control groups is *comparable* to the treatment group
- Formally: their potential outcomes are in expectation the same, i.e.

$$\mathbb{E}[y_{1i}|D=1] = \mathbb{E}[y_{1i}|D=0]$$
$$\mathbb{E}[y_{0i}|D=1] = \mathbb{E}[y_{0i}|D=0]$$

- Thus $\mathbb{E}[y_{0i}|D=1] - \mathbb{E}[y_{0i}|D=0] = 0$, i.e. no selection bias
  - in words: the control group tells us what would have happened to the treatment group in the absence of the treatment

Summary

- **Causality**: how one thing *affects* another thing
  - requires comparing counterfactual states of the world to each other ("how would Y change if we changed X?")
  - at most, one of them is observed
- **Control group** in an experimental research design
  - the outcomes of the control group are used to infer what would have happened to the treatment group in the absence of the treatment
- **Selection bias** occurs when the control group is not comparable to the treatment group, i.e. $\mathbb{E}[y_{0i}|D = 0] \neq \mathbb{E}[y_{0i}|D = 1]$
  - $=$ potential outcomes differ between the treatment and control groups
- **Randomization** eliminates selection bias
  - on expectation, the only difference between the groups is that the treatment group gets the treatment and the control group does not
  - $\rightarrow$ differences in average outcomes must be due to the treatment

Upcoming

- **Pre-class assignment 4**
  - Read and summarize an article
- **Homework 2**
  - Deadline: Jan 24 at 13:00
  - Don't wait till the last minute!
  - An important skill when working with data is to learn to troubleshoot efficiently.
  - This learning often involves spending time being "stuck".
- **Exercise Session 2** tomorrow!
- Use the course **Slack** channel to seek help and help others in the class
  - Quicker than waiting for private responses from the TA or me
  - Recall extra incentive: bonus points for active participation