Class 5. Research designs

Nic Rivers

Advanced quantitative research methods, API6319
Fall 2019

Establishing causal relationships

Hurdles along the route to establishing causal relationships

- Most of our interesting research questions involve trying to understand whether changes in one variable (X, the *independent variable*) *cause* changes in another variable (Y, the *dependent* variable). ($X \rightarrow Y$):
    - Does increased development aid *cause* increase in growth rates? (How much?)
    - Does more education *cause* higher incomes? (How much?)
    - Do smaller class sizes in primary school *cause* improved grades? (How much?)
- But establishing causality is difficult.

Establishing causal relationships is difficult

Probabilistic vs. deterministic relationships

- We are used to thinking about cause and effect relationships in the physical sciences, which are *deterministic* (at least at scales that are relevant to most humans).
  - Turn on a switch $\rightarrow$ Light goes on (every time).
  - Drop a ball $\rightarrow$ It accelerates at 9.8m/s^2 (every time).
  - Feed arsenic to a frog $\rightarrow$ It dies (every time).

Establishing causal relationships is difficult

Probabilistic vs. deterministic relationships

- We are used to thinking about cause and effect relationships in the physical sciences, which are *deterministic* (at least at scales that are relevant to most humans).
  - Turn on a switch $\rightarrow$ Light goes on (every time).
  - Drop a ball $\rightarrow$ It accelerates at $9.8\text{m}/\text{s}^2$ (every time).
  - Feed arsenic to a frog $\rightarrow$ It dies (every time).
- In contrast, cause and effect relationships in the social sciences are *probabilistic*.
  - Increase in gasoline prices $\rightarrow$ Some people purchase more fuel efficient vehicles. (Average fuel efficiency improves.)
  - Increase in micro-lending $\rightarrow$ Income of some women goes up. (Average income of women increases.)
  - Increase in development aid $\rightarrow$ Growth rate of some countries increases. (On average, growth rate increases.)
- We need to learn to think in terms of averages rather than anecdotes.

Establishing causal relationships is difficult

Multiple things affect outcome

- Almost all of our theories in the social sciences involve a *bivariate* relationship: one (independent) variable *causes* changes in another (dependent) variable
  - For example: The availability of micro-lending helps to raise the incomes of poor women
- In reality, there are many things that affect the dependent variable
  - For example: incomes of poor women depend on culture, education, government policies, etc.
- It is often difficult to isolate the causal relationship of interest from the web of relationships

Establishing causal relationships is difficult

But our brains trick us into thinking we understand the world better than we do

- On the day of Saddam Hussein's capture in Iraq, Bloomberg headlines:

  "U.S. treasuries rise; Hussein's capture may not curb terrorism."

  Half an hour later

  "U.S. treasuries fall; Hussein's capture boosts the allure of risky assets."

Establishing causal relationships is difficult

But our brains trick us into thinking we understand the world better than we do

- On the day of Saddam Hussein's capture in Iraq, Bloomberg headlines:
  "U.S. treasuries rise; Hussein's capture may not curb terrorism."

  Half an hour later
  "U.S. treasuries fall; Hussein's capture boosts the allure of risky assets."
- Our brains are wired for jumping to conclusions. We "see" associations where there are only coincidences. We look for evidence to support our chosen interpretation of events (confirmation bias).
- This is the result of what Daniel Khaneman calls our "System 1" brain (bushes rattle - it must be a tiger).

Establishing causal relationships is difficult

But our brains trick us into thinking we understand the world better than we do

- On the day of Saddam Hussein's capture in Iraq, Bloomberg headlines:
  "U.S. treasuries rise; Hussein's capture may not curb terrorism."

  Half an hour later
  "U.S. treasuries fall; Hussein's capture boosts the allure of risky assets."
- Our brains are wired for jumping to conclusions. We "see" associations where there are only coincidences. We look for evidence to support our chosen interpretation of events (confirmation bias).
- This is the result of what Daniel Khaneman calls our "System 1" brain (bushes rattle - it must be a tiger).
  - How many animals of each kind did Moses take on to the ark?

Establishing causal relationships is difficult

But our brains trick us into thinking we understand the world better than we do

- On the day of Saddam Hussein's capture in Iraq, Bloomberg headlines:

  "U.S. treasuries rise; Hussein's capture may not curb terrorism."

  Half an hour later

  "U.S. treasuries fall; Hussein's capture boosts the allure of risky assets."

- Our brains are wired for jumping to conclusions. We "see" associations where there are only coincidences. We look for evidence to support our chosen interpretation of events (confirmation bias).
- This is the result of what Daniel Khaneman calls our "System 1" brain (bushes rattle - it must be a tiger).
  - How many animals of each kind did Moses take on to the ark?
  - Fred's parents arrived late. It was a sunny day. Fred was angry.

Establishing causal relationships is difficult

But our brains trick us into thinking we understand the world better than we do

- On the day of Saddam Hussein's capture in Iraq, Bloomberg headlines:

  "U.S. treasuries rise; Hussein's capture may not curb terrorism."

  Half an hour later

  "U.S. treasuries fall; Hussein's capture boosts the allure of risky assets."

- Our brains are wired for jumping to conclusions. We "see" associations where there are only coincidences. We look for evidence to support our chosen interpretation of events (confirmation bias).
- This is the result of what Daniel Khaneman calls our "System 1" brain (bushes rattle - it must be a tiger).
  - How many animals of each kind did Moses take on to the ark?
  - Fred's parents arrived late. It was a sunny day. Fred was angry.
- System 1 brain is "a machine for jumping to conclusions"
- We need to train our System 2 brain to carefully scrutinize System 1 conclusions, and avoid confirmation bias. Social scientists look for reasons to refute claims, not support them.

Hurdles to establishing causality

- We typically want to establish whether some variable X *causes* changes in some other variable Y.
- To do so, we need to carefully consider the answers to the following questions:
  1. Is there a credible causal mechanism that connects X to Y?
  2. Can we rule out the possibility that Y could cause X?
  3. Is there covariation between X and Y?
  4. Are there *confounding factors* X that might be responsible for the perceived covariation between X and Y?

Is there a credible causal mechanism that links X to Y?

- This logically about whether X could cause Y
- It is important to have some theory about why changes in X could cause changes in Y.
- It can be helpful to think through the *mechanism* via which the relationship takes place
- We want to avoid "shots in the dark" where we are mining data to find random relationships between variables (this is the domain of *machine learning* and not part of the *scientific method*)

Can we rule out the possibility that Y could cause X?

- In evaluating the claim that X *causes* Y, we will look at data to see if the two variables are related.
- One reason the two variables could be related is if X causes Y
- Another reason the two variables could be related is if Y causes X
- We need to distinguish between these two possibilities. Typically, there is no statistical test for claim, and we need to think it through logically.
    - Example: Does low income cause conflict? Data on conflict and GDP from all countries for 2015

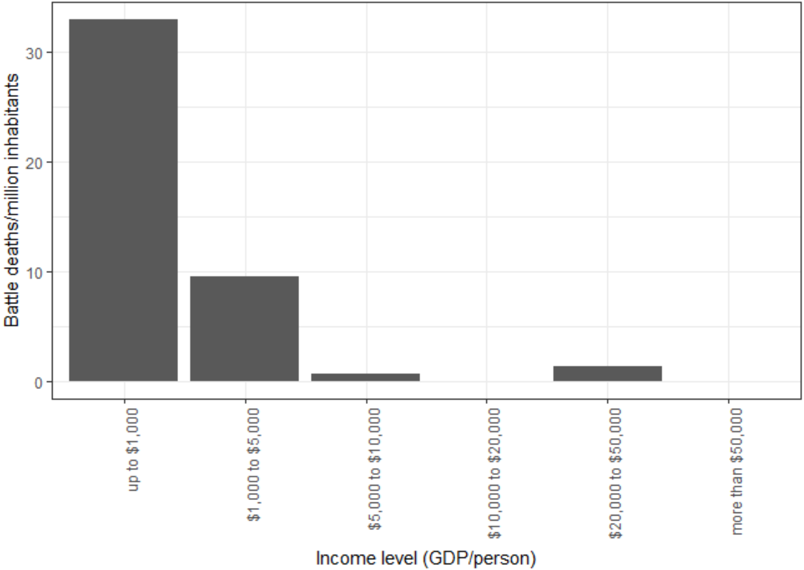
Can we rule out the possibility that Y could cause X?

- In evaluating the claim that X *causes* Y, we will look at data to see if the two variables are related.
- One reason the two variables could be related is if X causes Y
- Another reason the two variables could be related is if Y causes X
- We need to distinguish between these two possibilities. Typically, there is no statistical test for claim, and we need to think it through logically.
  - Example: Does low income cause conflict? Data on conflict and GDP from all countries for 2015
  - Example: Do older people have more conservative views about the role of government? Data from a survey

Can we rule out the possibility that Y could cause X?

- In evaluating the claim that X *causes* Y, we will look at data to see if the two variables are related.
- One reason the two variables could be related is if X causes Y
- Another reason the two variables could be related is if Y causes X
- We need to distinguish between these two possibilities. Typically, there is no statistical test for claim, and we need to think it through logically.
    - Example: Does low income cause conflict? Data on conflict and GDP from all countries for 2015
    - Example: Do older people have more conservative views about the role of government? Data from a survey
    - Example: Does paid parental leave improve infant health? Data on children's health from before and after parental leave policy

Deaths and incomes



Is there covariation between X and Y?

- In order to validate our theory about the causal effect of X on Y, we will collect data on these two variables
- We will learn how to implement statistical tests to see if these variables move together

  positive relationship More of X is associated with more of Y

  negative relationship More of X is associated with less of Y
- It is important to note that "correlation does not imply causation." However if there is a causal relationship between X and Y, the two variables will also be correlated.
- Sometimes another variable changes at the same time as X (call it Z). We may only observe correlation between X and Y once we "control for" Z.

Are there *confounding factors* X that might be responsible for the perceived covariation between X and Y?

- Our theories are typically *bivariate* - that is they relate a single independent variable (X) to a dependent variable (Y)
- In the real world, many factors can affect Y
- When we are looking for a relationship between X and Y, we therefore need to take account of these other factors. If we don't, our results will probably be misleading.
  - Example: we are interested in finding out if there is a relationship between education (X) and income (Y). We gather data on individuals from the Census, and find out the two variables are related. Is this evidence that education *causes* higher incomes?
- We will learn techniques for "controlling for" other variables in looking at a relationship between X and Y.

Establishing causality

- If one of the hurdles isn't passed, we need to be wary of claims that X causes Y.
- You need to think critically about research design, in your own work and when looking at that of others.
- We will be going through different *research designs* that can help us clear some of the hurdles.

Example 1: Do hospitals make people healthier?

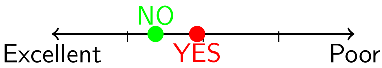
- Suppose we are interested in a causal question: "Do hospitals make people healthier?"
- We might think to answer this question by gathering data on groups of individuals that gone to hospital, and comparing to other individuals that have been in hospital.
- The 1991 wave of the General Social Survey asks two relevant questions:
    1. Did you spend any nights as a patient in a hospital nursing home or convalescent home during the last 12 months?
    2. Compared to other people your age, how would you describe your state of health? [1=excellent;5=poor]

Example 1 (con't)

- Taken at face value, the data in the table below suggest that spending time in hospital causes a deterioration in health outcomes.
- Is this plausible? What else might be going on? Think about four hurdles.

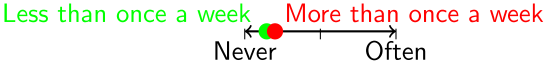
Time spent in hospital in last 12 months?	Mean self-reported health
YES	2.92
NO	2.37

Table: 1991 General Social Survey. [1=excellent;5=poor]



Example 2: Video games and violence

- Suppose we are interested in a causal question: "Does playing video games cause violent behaviour?"
- We might try to answer this question by comparing rates of fighting in kids that play video games with kids that don't play video games.
- The 1998-9 National Longitudinal Survey of Children and Youth asks two relevant questions (10-15 year olds):
  1. How often do you play video games?
  2. How often do you get into fights?
- Is this evidence that video games cause violent behaviour? (Think of four hurdles)



Examples

- Recall the four hurdles to establishing causality:
  1. Is there a credible causal mechanism that connects X to Y?
  2. Can we rule out the possibility that Y could cause X?
  3. Is there covariation between X and Y?
  4. Are there *confounding factors* X that might be responsible for the perceived covariation between X and Y?
- How would you assess the following comparisons in this light? (based on real headlines) [? ? ? ?]
  1. Communities with more police officers don't seem to experience lower crime rates.

Examples

- Recall the four hurdles to establishing causality:
  1. Is there a credible causal mechanism that connects X to Y?
  2. Can we rule out the possibility that Y could cause X?
  3. Is there covariation between X and Y?
  4. Are there *confounding factors* X that might be responsible for the perceived covariation between X and Y?
- How would you assess the following comparisons in this light? (based on real headlines) [? ? ? ?]
  1. Communities with more police officers don't seem to experience lower crime rates.
  2. Bilingual seniors have lower rates of dementia. Scientists find that there are "cognitive advantages" to bringing up children in a bilingual home.

Examples

- Recall the four hurdles to establishing causality:
  1. Is there a credible causal mechanism that connects X to Y?
  2. Can we rule out the possibility that Y could cause X?
  3. Is there covariation between X and Y?
  4. Are there *confounding factors* X that might be responsible for the perceived covariation between X and Y?
- How would you assess the following comparisons in this light? (based on real headlines) [? ? ? ?]
  1. Communities with more police officers don't seem to experience lower crime rates.
  2. Bilingual seniors have lower rates of dementia. Scientists find that there are "cognitive advantages" to bringing up children in a bilingual home.
  3. Recent college grads had a lower unemployment rate than their counterparts with only high school diplomas

Examples

- Recall the four hurdles to establishing causality:
  1. Is there a credible causal mechanism that connects X to Y?
  2. Can we rule out the possibility that Y could cause X?
  3. Is there covariation between X and Y?
  4. Are there *confounding factors* X that might be responsible for the perceived covariation between X and Y?
- How would you assess the following comparisons in this light? (based on real headlines) [? ? ? ?]
  1. Communities with more police officers don't seem to experience lower crime rates.
  2. Bilingual seniors have lower rates of dementia. Scientists find that there are "cognitive advantages" to bringing up children in a bilingual home.
  3. Recent college grads had a lower unemployment rate than their counterparts with only high school diplomas
  4. Evidence mounts that eating as a family can protect children from all sorts of harm, experts say

Police

Figure: Data on crime rates and police officers in 250 US cities.

Research design

- Our aim in social science research is to determine the causal effect of some variable X on some other variable Y. For example:

Research design

- Our aim in social science research is to determine the causal effect of some variable X on some other variable Y. For example:
  - Does stable employment reduce the propensity for violence in young men in Uganda?

Research design

- Our aim in social science research is to determine the causal effect of some variable X on some other variable Y. For example:
  - Does stable employment reduce the propensity for violence in young men in Uganda?
  - Does autocratic governance result in reduced social capital (trust, etc.)?

Research design

- Our aim in social science research is to determine the causal effect of some variable X on some other variable Y. For example:
  - Does stable employment reduce the propensity for violence in young men in Uganda?
  - Does autocratic governance result in reduced social capital (trust, etc.)?
- A key problem is that we know that there are typically many variables that affect Y, not just X. In addition, Y may affect X. How can we test our theory effectively to know if X affects Y?

Research design

- Our aim in social science research is to determine the causal effect of some variable X on some other variable Y. For example:
  - Does stable employment reduce the propensity for violence in young men in Uganda?
  - Does autocratic governance result in reduced social capital (trust, etc.)?
- A key problem is that we know that there are typically many variables that affect Y, not just X. In addition, Y may affect X. How can we test our theory effectively to know if X affects Y?
- Fundamental problem of causal inference: we can never observe the same person/country/etc. at the same time with different 'treatments'

Research design

- Our aim in social science research is to determine the causal effect of some variable X on some other variable Y. For example:
  - Does stable employment reduce the propensity for violence in young men in Uganda?
  - Does autocratic governance result in reduced social capital (trust, etc.)?
- A key problem is that we know that there are typically many variables that affect Y, not just X. In addition, Y may affect X. How can we test our theory effectively to know if X affects Y?
- Fundamental problem of causal inference: we can never observe the same person/country/etc. at the same time with different 'treatments'
- We need to design our research effectively. Effective research design rules out these competing explanations.

Research design

What is a *research design*?

- A **research design** is the overall strategy that is used to test a research hypothesis using data

Research design

What is a *research design*?

- A **research design** is the overall strategy that is used to test a research hypothesis using data
- It refers to the manner in which data is collected, measured, and analyzed

Research design

What is a *research design*?

- A **research design** is the overall strategy that is used to test a research hypothesis using data
- It refers to the manner in which data is collected, measured, and analyzed
- Since we will draw conclusions based on *comparisons*, most importantly a *research design* describes how we should construct the comparisons (i.e., what should we compare to what?)

Research design

What is a *research design*?

- A **research design** is the overall strategy that is used to test a research hypothesis using data
- It refers to the manner in which data is collected, measured, and analyzed
- Since we will draw conclusions based on *comparisons*, most importantly a *research design* describes how we should construct the comparisons (i.e., what should we compare to what?)
- Effective research designs aim to satisfy the four hurdles that make it difficult to draw causal inferences.

Establish causal relationships by comparison

- We seek to understand whether some X *causes* Y.
- The obvious way to find out is to compare some individual with X to some individual without X and see how Y differs between the two individuals.
- Comparison is the basis for establishing our empirical understanding of causal relationships.
- Yet a simple comparison may be misleading, since it does not account for competing explanations.
    - Think of causal hurdles/fundamental problem of causal inference.
- Designing research effectively involves carefully specifying what type of comparison would be most useful.

Example: a simple comparison

- The data shows that higher access to electricity seems to be associated with better nutritional outcomes. Can we say that access to electricity causes better nutritional outcomes?

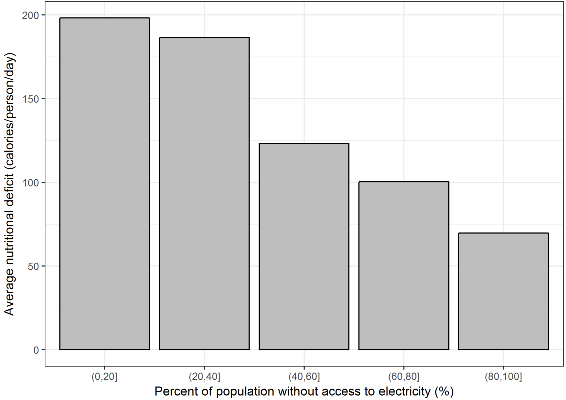


Figure: Based on data from World Bank for year 2010.

Problems with the simple comparison

- There seems to be a relationship between access to electricity (X) and nutritional outcomes (Y). Have we proven that this a causal relationship?

Problems with the simple comparison

- There seems to be a relationship between access to electricity (X) and nutritional outcomes (Y). Have we proven that this a causal relationship?
  - Is there a logical reason to think that X causes Y?

Problems with the simple comparison

- There seems to be a relationship between access to electricity (X) and nutritional outcomes (Y). Have we proven that this a causal relationship?
  - Is there a logical reason to think that X causes Y?
  - What if some other factor (Z) is causing the outcome (Y)?

Problems with the simple comparison

- There seems to be a relationship between access to electricity (X) and nutritional outcomes (Y). Have we proven that this a causal relationship?
  - Is there a logical reason to think that X causes Y?
  - What if some other factor (Z) is causing the outcome (Y)?
  - What if the causality runs in the other direction (Y to X)?

Problems with the simple comparison

- There seems to be a relationship between access to electricity (X) and nutritional outcomes (Y). Have we proven that this a causal relationship?
  - Is there a logical reason to think that X causes Y?
  - What if some other factor (Z) is causing the outcome (Y)?
  - What if the causality runs in the other direction (Y to X)?
- These concerns are the reason for the saying, "correlation is not causation."

Problems with the simple comparison

- There seems to be a relationship between access to electricity (X) and nutritional outcomes (Y). Have we proven that this a causal relationship?
  - Is there a logical reason to think that X causes Y?
  - What if some other factor (Z) is causing the outcome (Y)?
  - What if the causality runs in the other direction (Y to X)?
- These concerns are the reason for the saying, "correlation is not causation."
- Given these challenges, how can we make causal inferences?

Problems with the simple comparison

- There seems to be a relationship between access to electricity (X) and nutritional outcomes (Y). Have we proven that this a causal relationship?
  - Is there a logical reason to think that X causes Y?
  - What if some other factor (Z) is causing the outcome (Y)?
  - What if the causality runs in the other direction (Y to X)?
- These concerns are the reason for the saying, "correlation is not causation."
- Given these challenges, how can we make causal inferences?
- We need a research design that carefully considers these threats to validity

Experiments - Randomized controlled trials (RCTs) - can show causality

- An experiment is a particular type of *research design*. In an experiment, the researcher:

Experiments - Randomized controlled trials (RCTs) - can show causality

- An experiment is a particular type of *research design*. In an experiment, the researcher:
  1. **controls** the treatment (X) for each individual.

Experiments - Randomized controlled trials (RCTs) - can show causality

- An experiment is a particular type of *research design*. In an experiment, the researcher:
  1. **controls** the treatment (X) for each individual.
     - Since the researcher controls X for each individual, we can be certain that individuals themselves do not decide on X. We can also be sure that nature does not choose the level of X experienced by each individual.

Experiments - Randomized controlled trials (RCTs) - can show causality

- An experiment is a particular type of *research design*. In an experiment, the researcher:
  1. **controls** the treatment (X) for each individual.
     - Since the researcher controls X for each individual, we can be certain that individuals themselves do not decide on X. We can also be sure that nature does not choose the level of X experienced by each individual.
  2. applies the treatment by **randomly assigning** which individuals are treated and which are not

Experiments - Randomized controlled trials (RCTs) - can show causality

- An experiment is a particular type of *research design*. In an experiment, the researcher:
  1. **controls** the treatment (X) for each individual.
     - Since the researcher controls X for each individual, we can be certain that individuals themselves do not decide on X. We can also be sure that nature does not choose the level of X experienced by each individual.
  2. applies the treatment by **randomly assigning** which individuals are treated and which are not
     - Since the treatment is randomly assigned (e.g., roll of the dice), we can be sure that there is no systematic relationship between other variables (Z) and the treatment (X).

Experiments - Randomized controlled trials (RCTs) - can show causality

- An experiment is a particular type of *research design*. In an experiment, the researcher:
  1. **controls** the treatment (X) for each individual.
     - Since the researcher controls X for each individual, we can be certain that individuals themselves do not decide on X. We can also be sure that nature does not choose the level of X experienced by each individual.
  2. applies the treatment by **randomly assigning** which individuals are treated and which are not
     - Since the treatment is randomly assigned (e.g., roll of the dice), we can be sure that there is no systematic relationship between other variables (Z) and the treatment (X).
     - We can also be sure that Y doesn't cause X, since the treatment is applied randomly

Experiments - Randomized controlled trials (RCTs) - can show causality

- An experiment is a particular type of *research design*. In an experiment, the researcher:
  1. **controls** the treatment (X) for each individual.
     - Since the researcher controls X for each individual, we can be certain that individuals themselves do not decide on X. We can also be sure that nature does not choose the level of X experienced by each individual.
  2. applies the treatment by **randomly assigning** which individuals are treated and which are not
     - Since the treatment is randomly assigned (e.g., roll of the dice), we can be sure that there is no systematic relationship between other variables (Z) and the treatment (X).
     - We can also be sure that Y doesn't cause X, since the treatment is applied randomly
  3. observes Y in the control and treatment groups. The difference is the causal effect of X.

Random assignment vs. random sampling

- Experiments work as a result of *random assignment* of subjects to treatment and control groups (i.e., different values of X)

Random assignment vs. random sampling

- Experiments work as a result of *random assignment* of subjects to treatment and control groups (i.e., different values of X)
  - When subjects are randomly assigned to treatment, we can credibly infer the causal effect of treatment. Studies with random assignment are said to have *internal validity*

Random assignment vs. random sampling

- Experiments work as a result of *random assignment* of subjects to treatment and control groups (i.e., different values of X)
    - When subjects are randomly assigned to treatment, we can credibly infer the causal effect of treatment. Studies with random assignment are said to have *internal validity*
- In contrast, *random sampling* relates to how researchers decide which subjects to monitor.

Random assignment vs. random sampling

- Experiments work as a result of *random assignment* of subjects to treatment and control groups (i.e., different values of X)
  - When subjects are randomly assigned to treatment, we can credibly infer the causal effect of treatment. Studies with random assignment are said to have *internal validity*
- In contrast, *random sampling* relates to how researchers decide which subjects to monitor.
  - Randomly sampling people for inclusion ensures that the results of the study will be generalizable to the population at large. This is sometimes called *external validity*

Random assignment vs. random sampling

- Experiments work as a result of *random assignment* of subjects to treatment and control groups (i.e., different values of X)
    - When subjects are randomly assigned to treatment, we can credibly infer the causal effect of treatment. Studies with random assignment are said to have *internal validity*
- In contrast, *random sampling* relates to how researchers decide which subjects to monitor.
    - Randomly sampling people for inclusion ensures that the results of the study will be generalizable to the population at large. This is sometimes called *external validity*

Random assignment vs. random sampling

- Experiments work as a result of *random assignment* of subjects to treatment and control groups (i.e., different values of X)
  - When subjects are randomly assigned to treatment, we can credibly infer the causal effect of treatment. Studies with random assignment are said to have *internal validity*
- In contrast, *random sampling* relates to how researchers decide which subjects to monitor.
  - Randomly sampling people for inclusion ensures that the results of the study will be generalizable to the population at large. This is sometimes called *external validity*

	Random assignment	Non-random assignment
Random sampling	Externally valid and internally valid	Externally valid but not internally valid
Non-random sampling	Internally valid but not externally valid	Not internally valid or externally valid

Experiments in medicine

- The conventional way to test new drugs and treatments is with a randomized controlled trial (RCT). Research question - is some new drug X effective at reducing some disease Y?
    1. Participants are recruited
    2. Participants are **randomly assigned** into two groups - a treatment group and a control group
    3. The treatment group is given the new treatment. The control group is given a placebo
    4. The two groups are compared (e.g., disease rates).
    5. The difference in outcomes (Y) is the effect of the treatment (X).
- The analyst controls who is assigned to treatment, and assigns treatment randomly.

Experiments in science

- The conventional way to test new scientific theories is with an experiment. Research question - does air pollution (X) lead to aggression (Y)?
  1. Get some rats.
  2. Rats are divided **randomly** into two groups - a treatment group and a control group.
  3. The treatment group is put in a room with elevated levels of pollution; the control group is put into another room that is otherwise similar.
  4. The behaviour of the two sets of rats is compared.
  5. The difference in outcomes (Y) is the effect of the treatment (X).
- The analyst controls who is assigned to treatment, and assigns treatment randomly.

Can we do experiments in social science?

- People are not rats.

Can we do experiments in social science?

- People are not rats.
- Countries are not rats.

Can we do experiments in social science?

- People are not rats.
- Countries are not rats.
- Governments are not rats.

Can we do experiments in social science?

- People are not rats.
- Countries are not rats.
- Governments are not rats.
- Can we do experiments on important social science topics?

Can we do experiments in social science?

- People are not rats.
- Countries are not rats.
- Governments are not rats.
- Can we do experiments on important social science topics?

  Lab experiments: Social scientists have done lab experiments for several decades.

Can we do experiments in social science?

- People are not rats.
- Countries are not rats.
- Governments are not rats.
- Can we do experiments on important social science topics?

  Lab experiments: Social scientists have done lab experiments for several decades.

  - Lots of game theory research (e.g., *ultimatum game*)

Can we do experiments in social science?

- People are not rats.
- Countries are not rats.
- Governments are not rats.
- Can we do experiments on important social science topics?

  Lab experiments: Social scientists have done lab experiments for several decades.

  - Lots of game theory research (e.g., *ultimatum game*)
  - Lots of psychology research

Can we do experiments in social science?

- People are not rats.
- Countries are not rats.
- Governments are not rats.
- Can we do experiments on important social science topics?

  Lab experiments: Social scientists have done lab experiments for several decades.
  - Lots of game theory research (e.g., *ultimatum game*)
  - Lots of psychology research

  Survey experiments: Social scientists use hypothetical experiments (choice experiments)

Can we do experiments in social science?

- People are not rats.
- Countries are not rats.
- Governments are not rats.
- Can we do experiments on important social science topics?

  Lab experiments: Social scientists have done lab experiments for several decades.
  - Lots of game theory research (e.g., *ultimatum game*)
  - Lots of psychology research

  Survey experiments: Social scientists use hypothetical experiments (choice experiments)

  Field experiments: Increasingly, social scientists use field experiments to test theories *in the field*

Can we do experiments in social science?

- People are not rats.
- Countries are not rats.
- Governments are not rats.
- Can we do experiments on important social science topics?

Lab experiments
: Social scientists have done lab experiments for several decades.
  - Lots of game theory research (e.g., *ultimatum game*)
  - Lots of psychology research

Survey experiments
: Social scientists use hypothetical experiments (choice experiments)

Field experiments
: Increasingly, social scientists use field experiments to test theories *in the field*
  - Effectiveness of micro-lending

Can we do experiments in social science?

- People are not rats.
- Countries are not rats.
- Governments are not rats.
- Can we do experiments on important social science topics?

Lab experiments
: Social scientists have done lab experiments for several decades.
    - Lots of game theory research (e.g., *ultimatum game*)
    - Lots of psychology research

Survey experiments
: Social scientists use hypothetical experiments (choice experiments)

Field experiments
: Increasingly, social scientists use field experiments to test theories *in the field*
    - Effectiveness of micro-lending
    - How can we reduce school absenteeism?

Can we do experiments in social science?

- People are not rats.
- Countries are not rats.
- Governments are not rats.
- Can we do experiments on important social science topics?

Lab experiments Social scientists have done lab experiments for several decades.

- Lots of game theory research (e.g., *ultimatum game*)
- Lots of psychology research

Survey experiments Social scientists use hypothetical experiments (choice experiments)

Field experiments Increasingly, social scientists use field experiments to test theories *in the field*

- Effectiveness of micro-lending
- How can we reduce school absenteeism?
- Do sweatshops reduce poverty?

Can we do experiments in social science?

- People are not rats.
- Countries are not rats.
- Governments are not rats.
- Can we do experiments on important social science topics?

Lab experiments Social scientists have done lab experiments for several decades.
- Lots of game theory research (e.g., *ultimatum game*)
- Lots of psychology research

Survey experiments Social scientists use hypothetical experiments (choice experiments)

Field experiments Increasingly, social scientists use field experiments to test theories *in the field*
- Effectiveness of micro-lending
- How can we reduce school absenteeism?
- Do sweatshops reduce poverty?
- Can cash transfers eliminate poverty

Field experiments in social science: examples

Does micro-lending reduce poverty?

- Microcredit has generated considerable enthusiasm and hope for fast poverty alleviation.[1]
- In 2006, Mohammad Yunus (founder of Grameen Bank) and the Grameen Bank (micro-credit institution) were awarded the Nobel Prize for Peace, for their contribution to the reduction in world poverty.
- Yet there is substantial doubt as well: concerns about large profits made from the poor. Concerns about lenders pushing their clients to accept debt.
- How might we provide evidence to know whether micro-lending improves lives (e.g., reduces poverty)?

[1] See this study for more background.

Example: Does job micro-lending reduce poverty in Indian women?

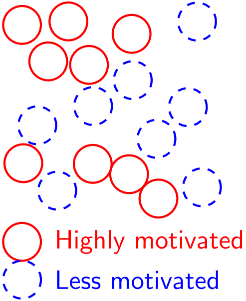
- There is interest in knowing whether micro-lending can help reduce poverty
- Imagine that we collect data on two groups of women in India - some that have received micro-loans and others that have not. We compare poverty rates in these two groups. Does this comparison tell us about the causal effects of micro-loans on poverty?

Example: Does job micro-lending reduce poverty in Indian women?

- There is interest in knowing whether micro-lending can help reduce poverty
- Imagine that we collect data on two groups of women in India - some that have received micro-loans and others that have not. We compare poverty rates in these two groups. Does this comparison tell us about the causal effects of micro-loans on poverty?
    1. Is there a credible causal mechanism that connects X to Y?
    2. Can we rule out the possibility that Y could cause X?
    3. Is there covariation between X and Y?
    4. Are there *confounding factors* X that might be responsible for the perceived covariation between X and Y?
- This comparison is problematic especially because of confounding factors: there are likely to be other factors Z that influence poverty rates.
- One example is motivation (e.g., more motivated women more likely to apply for loan; less likely to be in poverty).

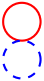
Randomized controlled trials vs. observational studies

Does job micro-lending reduce poverty in Indian women?



Randomized controlled trials vs. observational studies

Does job micro-lending reduce poverty in Indian women?

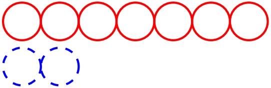


Highly motivated

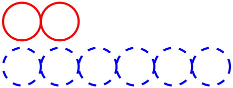
Less motivated

Observational study

Treatment group (micro-loans)



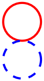
Control group (no micro-loans)



Problem: Highly motivated women select into micro-lending programs.

Randomized controlled trials vs. observational studies

Does job micro-lending reduce poverty in Indian women?

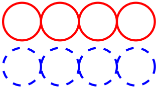


Highly motivated

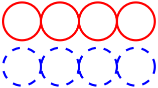
Less motivated

Experimental study

Treatment group (micro-loans)



Control group (no micro-loans)



Solution: Micro-loans are randomized, so highly motivated women are just as likely to receive a micro-loan as not.

What did the experiment do?

The experiment was conducted as follows. In 2005, 52 of 104 poor neighborhoods in Hyderabad were randomly selected for opening of an MFI branch by one of the fastest-growing MFIs in the area, Spandana, while the remainder were not. Hyderabad is the fifth largest city in India, and the capital of Andhra Pradesh, the Indian state were microcredit has expanded the fastest. Fifteen to 18 months after the introduction of microfinance in each area, a comprehensive household survey was conducted in an average of 65 households in each neighborhood, for a total of about 6,850 households.

- After 12 months, we see no difference in monthly per capita consumption and monthly non-durable consumption. We do see significant positive impacts on the purchase of durables ... households have reduced expenditures on what that they themselves describe as "temptation goods."
- Fifteen to 18 months after gaining access, households are no more likely to be entrepreneurs
- For all but the largest businesses, there is no difference in the profits of the businesses... contrary to most people's belief, to the extent microcredit helps businesses, it may help the larger businesses more
- We do not find any effect on any of the women's empowerment or human development outcomes either after 18 or 36 months

Experiments as a solution to fundamental problem of causal inference

- Experiments help to solve the fundamental problem of causal inference
- Because the experimenter assigns the treatment (X) to the subjects (randomly), it is clear that Y did not cause X. If there is a relationship between X and Y, it must be because X causes Y.
- Because the experimenter assigns the treatment randomly, there is no systematic relationship between other variables (Z) and X.
- Z can be anything — we might not even know what Z is. A well-conducted experiment will still balance Z between treatment and control groups.
- Therefore, in a well-conducted experiment, comparing the dependent variable Y for subjects with different levels of X yields the causal impact of X on Y.

An example of an experiment

Blattman, C. and Dercon, S. More sweatshops for Africa?

- Sweatshops (low-wage manufacturing) are extremely controversial:

An example of an experiment

Blattman, C. and Dercon, S. More sweatshops for Africa?

- Sweatshops (low-wage manufacturing) are extremely controversial:
  - Industrial sector growth is seen as an important step in broad-scale economic development and policy reduction

An example of an experiment

Blattman, C. and Dercon, S. More sweatshops for Africa?

- Sweatshops (low-wage manufacturing) are extremely controversial:
  - Industrial sector growth is seen as an important step in broad-scale economic development and policy reduction
  - Factory jobs can be hazardous, are low-wage, and may not be in the long-term interest of workers

An example of an experiment

Blattman, C. and Dercon, S. More sweatshops for Africa?

- Sweatshops (low-wage manufacturing) are extremely controversial:
    - Industrial sector growth is seen as an important step in broad-scale economic development and policy reduction
    - Factory jobs can be hazardous, are low-wage, and may not be in the long-term interest of workers
- Question: Does sweatshop employment lead to improved long-term economic outcomes (employment, wages)?

An example of an experiment

Blattman, C. and Dercon, S. More sweatshops for Africa?

- Sweatshops (low-wage manufacturing) are extremely controversial:
  - Industrial sector growth is seen as an important step in broad-scale economic development and policy reduction
  - Factory jobs can be hazardous, are low-wage, and may not be in the long-term interest of workers
- Question: Does sweatshop employment lead to improved long-term economic outcomes (employment, wages)?
- How would an observational study in this context proceed? What would be potential problems?

An example of an experiment

Blattman, C. and Dercon, S. More sweatshops for Africa?

- Sweatshops (low-wage manufacturing) are extremely controversial:
  - Industrial sector growth is seen as an important step in broad-scale economic development and policy reduction
  - Factory jobs can be hazardous, are low-wage, and may not be in the long-term interest of workers
- Question: Does sweatshop employment lead to improved long-term economic outcomes (employment, wages)?
- How would an observational study in this context proceed? What would be potential problems?
- Approach: Randomly hire applicants to sweathshops. Compare hired group with non-hired group.

An example of an experiment

Blattman, C. and Dercon, S. More sweatshops for Africa?

- Sweatshops (low-wage manufacturing) are extremely controversial:
  - Industrial sector growth is seen as an important step in broad-scale economic development and policy reduction
  - Factory jobs can be hazardous, are low-wage, and may not be in the long-term interest of workers
- Question: Does sweatshop employment lead to improved long-term economic outcomes (employment, wages)?
- How would an observational study in this context proceed? What would be potential problems?
- Approach: Randomly hire applicants to sweathshops. Compare hired group with non-hired group.
- Result: Sweatshop jobs did not lead to improved wages, or increased likelihood of long-term employment.

Pitfalls of experimental research designs

If experiments are so great, why don't we use them all the time?

- Experiments can be expensive

Pitfalls of experimental research designs

If experiments are so great, why don't we use them all the time?

- Experiments can be expensive
- Experiments can be unethical (e.g., what is the effect of violence on long-run health? Does bombing a village reduce or increase insurgency?)

Pitfalls of experimental research designs

If experiments are so great, why don't we use them all the time?

- Experiments can be expensive
- Experiments can be unethical (e.g., what is the effect of violence on long-run health? Does bombing a village reduce or increase insurgency?)
- Experiments can be impractical or impossible (e.g., what is the effect of gender on voting behaviour? What is the effect of autocratic governance on growth rate? What is the effect of being landlocked on economic outcomes?)

Pitfalls of experimental research designs

If experiments are so great, why don't we use them all the time?

- Experiments can be expensive
- Experiments can be unethical (e.g., what is the effect of violence on long-run health? Does bombing a village reduce or increase insurgency?)
- Experiments can be impractical or impossible (e.g., what is the effect of gender on voting behaviour? What is the effect of autocratic governance on growth rate? What is the effect of being landlocked on economic outcomes?)
- Experiments may have low **external validity**. Low external validity means that the results may not extend beyond the sample population. This may be an issue if the population selects into an experiment. (random sampling vs. random assignment)

Pitfalls of experimental research designs

If experiments are so great, why don't we use them all the time?

- Experiments can be expensive
- Experiments can be unethical (e.g., what is the effect of violence on long-run health? Does bombing a village reduce or increase insurgency?)
- Experiments can be impractical or impossible (e.g., what is the effect of gender on voting behaviour? What is the effect of autocratic governance on growth rate? What is the effect of being landlocked on economic outcomes?)
- Experiments may have low **external validity**. Low external validity means that the results may not extend beyond the sample population. This may be an issue if the population selects into an experiment. (random sampling vs. random assignment)
- If people know they are being experimented on, they may behave differently (Hawthorne effect).

Pitfalls of experimental research designs

If experiments are so great, why don't we use them all the time?

- Experiments can be expensive
- Experiments can be unethical (e.g., what is the effect of violence on long-run health? Does bombing a village reduce or increase insurgency?)
- Experiments can be impractical or impossible (e.g., what is the effect of gender on voting behaviour? What is the effect of autocratic governance on growth rate? What is the effect of being landlocked on economic outcomes?)
- Experiments may have low **external validity**. Low external validity means that the results may not extend beyond the sample population. This may be an issue if the population selects into an experiment. (random sampling vs. random assignment)
- If people know they are being experimented on, they may behave differently (Hawthorne effect).
- Experiments often cannot tell us about *general equilibrium* effects (i.e., what would happen if we scaled up program to apply to everyone?)

Esther Duflo - fighting poverty with experiments

Video

Nature News on experiments and global poverty: article

Observational (non-experimental) studies

- Experimental research designs are limited to those in which the analyst can choose the treatment (X)
- This leaves out a large number of topics of interest in the social sciences
- For these other topics, we have to rely on observational studies to understand the effect of X on Y. An **observational study** is one in which the researcher does not manipulate the treatment (X), but instead just passively observes outcomes (Y) for subjects with different treatments (X).

Observational studies

- The basic idea of an observational study is the same as in an experimental study: the researcher *compares* the outcomes (Y) of individuals with different treatments (X). But now we have to be aware that:
  1. It is possible that even if there is a relationship between X and Y, it is not due to the effect of X on Y, but rather due to some other variable Z.
  2. It is possible that even if there is a relationship between X and Y, it is not because X causes Y but instead because Y causes X.
- For these reasons, we have to be cautious in conducting and interpreting results of observational studies.
- However, if we are careful about research design, observational studies can provide high-quality information about causal effects.

Observational studies can yield credible estimates of the causal effect of X on Y if done carefully

- Careful research design aims to rule out the possibility that Y causes X and aims to rule out the possibility that Y is caused by some other factor Z
- All studies are based on comparison of Y for different values of X
- It is necessary to design research in such a way that the comparison is fruitful in telling us about the causal effect of X on Y
- Researchers have developed a number of techniques for doing this
- All are focused on the *fundamental problem of causal inference*: it is not possible to observe the same subject at the same time in both treated and untreated states.

Some advanced observational research designs

A basic observational research design compares Y for individuals with different levels of X. This may lead to problems with interpretation (four hurdles). More advanced research designs try to get over these hurdles.

- Control for other variables/Multiple regression
- Natural experiment
- Regression discontinuity design
- Instrumental variable

Multiple regression designs

- One our primary concerns in observational studies is that may be other factors (Z) that muddy the relationship between X and Y.

Multiple regression designs

- One our primary concerns in observational studies is that may be other factors (Z) that muddy the relationship between X and Y.
- When we see a relationship between X and Y, it can be difficult to know if it is because of a the causal effect of X on Y or because some other factor Z influences both X and Y.

Multiple regression designs

- One our primary concerns in observational studies is that may be other factors (Z) that muddy the relationship between X and Y.
- When we see a relationship between X and Y, it can be difficult to know if it is because of a the causal effect of X on Y or because some other factor Z influences both X and Y.
- In this latter case, we call the observed relationship between X and Y a ‘spurious relationship’ and we call Z ‘confounding variables’

Multiple regression designs

- One our primary concerns in observational studies is that may be other factors (Z) that muddy the relationship between X and Y.
- When we see a relationship between X and Y, it can be difficult to know if it is because of a the causal effect of X on Y or because some other factor Z influences both X and Y.
- In this latter case, we call the observed relationship between X and Y a 'spurious relationship' and we call Z 'confounding variables'
- One approach is to 'control for' confounding variables. 'Controlling for' confounding variables means taking them into account when we are studying the relationship between X and Y.

Multiple regression designs

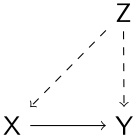
- One our primary concerns in observational studies is that may be other factors (Z) that muddy the relationship between X and Y.
- When we see a relationship between X and Y, it can be difficult to know if it is because of a the causal effect of X on Y or because some other factor Z influences both X and Y.
- In this latter case, we call the observed relationship between X and Y a 'spurious relationship' and we call Z 'confounding variables'
- One approach is to 'control for' confounding variables. 'Controlling for' confounding variables means taking them into account when we are studying the relationship between X and Y.
- When we learn *multiple regression* we will learn how to 'control for' confounding variables statistically.

Multiple regression designs

- One our primary concerns in observational studies is that may be other factors (Z) that muddy the relationship between X and Y.
- When we see a relationship between X and Y, it can be difficult to know if it is because of a the causal effect of X on Y or because some other factor Z influences both X and Y.
- In this latter case, we call the observed relationship between X and Y a 'spurious relationship' and we call Z 'confounding variables'
- One approach is to 'control for' confounding variables. 'Controlling for' confounding variables means taking them into account when we are studying the relationship between X and Y.
- When we learn *multiple regression* we will learn how to 'control for' confounding variables statistically.
- This is like looking for a relationship between X and Y in groups with similar Z

Controlling for other variables

- Suppose we suspect that the relationship we find between X and Y is actually due to a third variable Z



Controlling for other variables

- Suppose we suspect that the relationship we find between X and Y is actually due to a third variable Z

$$\begin{tikzcd}
& Z \arrow[dl, dashed] \arrow[d, dashed] \\
X \arrow[r] & Y
\end{tikzcd}$$

- We can 'control for' Z. For example, we could look at the relationship between X and Y for different subsets of Z

Only with $Z = Z_1$

$$X \longrightarrow Y$$

Only with $Z = Z_2$

$$X \longrightarrow Y$$

Example: Flossing and oral health

Feeling Guilty About Not Flossing? Maybe There's No Need

- NY Times article claims that there is no (or only very weak) evidence that flossing improves oral health. As a result, US gov has stopped advising Americans to floss.

Example: Flossing and oral health

Feeling Guilty About Not Flossing? Maybe There's No Need

- NY Times article claims that there is no (or only very weak) evidence that flossing improves oral health. As a result, US gov has stopped advising Americans to floss.
- Why not do an experiment on flossing?

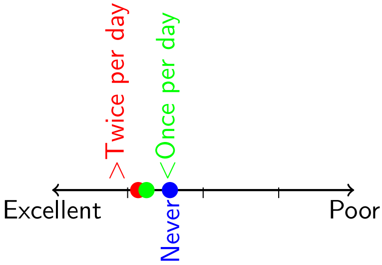
Example: Flossing and oral health

Feeling Guilty About Not Flossing? Maybe There's No Need

- NY Times article claims that there is no (or only very weak) evidence that flossing improves oral health. As a result, US gov has stopped advising Americans to floss.
- Why not do an experiment on flossing?
- It is difficult to do randomized controlled trials on flossing—it is a long-term behavioural change; people may not stick with experimental protocol.

Example: Flossing and oral health

- But studies exist that link flossing to oral health. For example, in a recent Canadian survey[2] Canadians were asked about the state of their oral health and their frequency of flossing



- Is this evidence that flossing causes improved oral health? Why? Why not?

[2] Canadian Community Health Survey, 2009, Healthy Aging.

Is X correlated with other factors Z that cause Y?

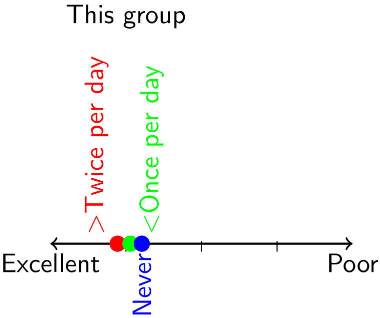
- What could some of these factors be?

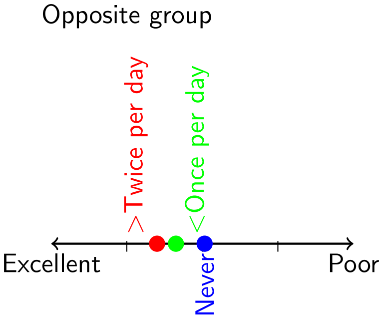
Is X correlated with other factors Z that cause Y?

- What could some of these factors be?
- We can repeat the comparison for a more homogeneous group - such as a group that (1) has seen a dentist in the past year, (2) has dental insurance, (3) brushes twice daily, (4) has enough income to meet basic expenses, (5) earns at least $80,000 (household) per year.

Is X correlated with other factors Z that cause Y?

- What could some of these factors be?
- We can repeat the comparison for a more homogeneous group - such as a group that (1) has seen a dentist in the past year, (2) has dental insurance, (3) brushes twice daily, (4) has enough income to meet basic expenses, (5) earns at least $80,000 (household) per year.



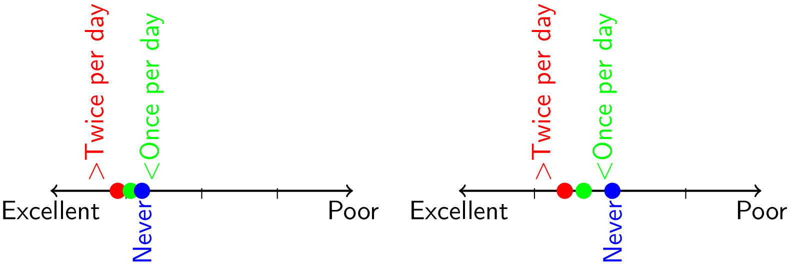


- Have we eliminated all factors Z that could cause X? Can we?

Is X correlated with other factors Z that cause Y?

- What could some of these factors be?
- We can repeat the comparison for a more homogeneous group - such as a group that (1) has seen a dentist in the past year, (2) has dental insurance, (3) brushes twice daily, (4) has enough income to meet basic expenses, (5) earns at least $80,000 (household) per year.

This group | Opposite group



- Have we eliminated all factors Z that could cause X? Can we?
- We can only 'control for' factors that we can measure. There may be other factors that we can't measure that confound the relationship between X and Y.

Does Y cause X?

Does oral health cause flossing?

- We are interested in testing whether more flossing causes improved oral health
- We do so by looking for a relationship in the data between oral health and flossing
- But the presence of a relationship could imply that differences in oral health are responsible for differences in flossing, not the reverse
- Our approach that controls for Z does not shed any light on the direction of causality. (in this case)

Multiple regression

- Later in the course, we will cover multiple regression.
- This is a formal statistical tool for "controlling" for other variables when drawing a conclusion about the link between X and Y
- Similar in concept to stratifying the sample, as described above.
- Same strengths and weaknesses as above:
  - We can control for other variables that we think might confound the relationship between X and Y if we can measure them.
  - We can't control for things that we can't measure or can't observe
  - We still need to be concerned about reverse causality.

Other observational research designs

- Regression discontinuity
- Matching
- Synthetic control
- Natural experiments
- Instrumental variables

Two good resources:

- Scott Cunningham. Causal inference: The mixtape. https://www.scunning.com/mixtape.html
- Angrist and Pischke. Mastering Metrics.

Wrapping up

- Any time we evaluate a study (or propose our own), we need to develop a *research design* that is capable of providing an estimate of the causal effect we are interested in
- Four causal hurdles
- Experiment - gold standard
- But all is not lost if we can't do an experiment
- Just need to think carefully about how our observational study may be compromised
- In any case, an experiment is a useful benchmark - think like an experiment.
- Beware country-level studies. Why?