MY457/MY557: Causal Inference for Observational and Experimental Studies

Week 1: Causal Frameworks

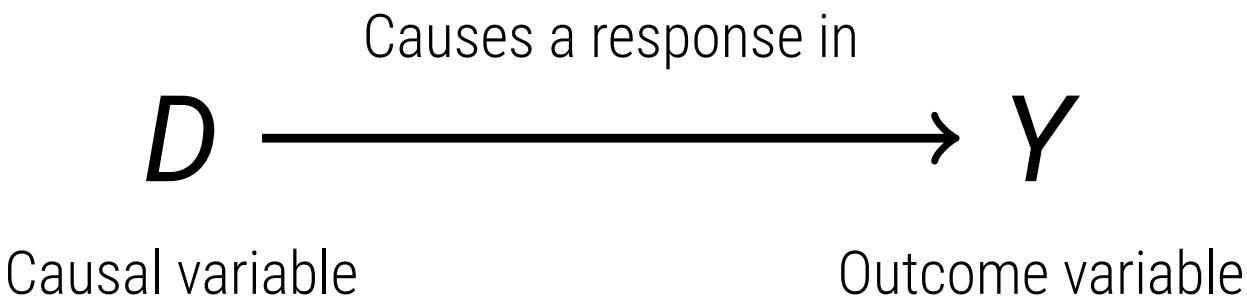
Daniel de Kadt

Department of Methodology
LSE

Winter Term 2026

Lecture Map

A (Very) Simple Causal Model



Effects of Causes

Causes and their effects have two properties: they are successive and can be reasoned about in counterfactual terms:

> *[...] We may define a cause to be an object followed by another, [...] where, if the first object had not been, the second never had existed.*
>
> *— Hume, 1748*

> *[...] would not have died if he had not eaten of it, people would be apt to say that eating of that dish was the cause of his death.*
>
> *— Mill, 1843*

One important implication is that causal variables must be manipulable:

> *No causation without manipulation.*
>
> *— Holland, 1986*

Good Causal Questions

Manipulability means we must think very carefully about causal questions...

- (Largely) immutable characteristics?
  - Judges' sex assigned at birth $\rightarrow$ decision making
  - Race and ethnicity $\rightarrow$ employment outcomes
  - Country of origin $\rightarrow$ political beliefs
- Major global events?
  - Russian revolution $\rightarrow$ Karl Marx's intellectual popularity
  - 9/11 $\rightarrow$ Arab Spring
- Non-successive chains?
  - Monthly expenditure $\rightarrow$ monthly savings
  - Holocaust $\rightarrow$ modern AFD election support

Concepts: Treatment, Outcomes, and Potential Outcomes

Treatment:

D_i: Indicator of treatment intake for *unit* i

$$D_i = \begin{cases} 1 & \text{if unit } i \text{ received the treatment} \\ 0 & \text{otherwise.} \end{cases}$$

Observed Outcome:

Y_i: Observed outcome variable of interest for unit i

Potential Outcome:

Y_{0i} and Y_{1i}: Potential outcomes for unit i:

$$\begin{aligned} Y_{1i} &\quad \text{Outcome for unit } i \text{ when } D_i = 1 \\ Y_{0i} &\quad \text{Outcome for unit } i \text{ when } D_i = 0 \end{aligned}$$

(Alternative notation: $Y_i(d)$, Y_i^d, etc.)

Concepts: Treatment, Outcomes, and Potential Outcomes

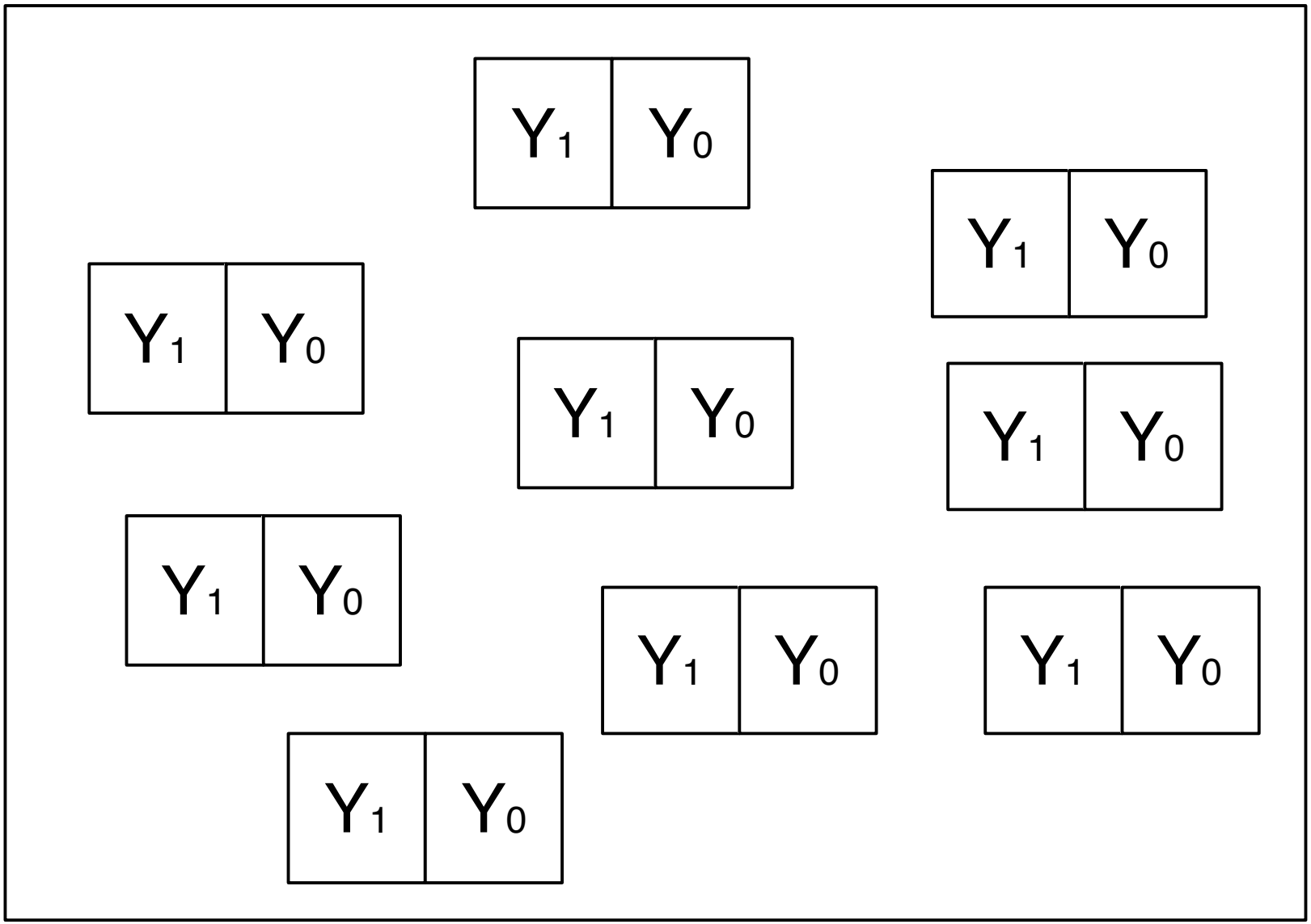
Under further assumptions ('SUTVA', more later), we can connect these three concepts mathematically:

$$
\begin{aligned}
Y_i &= D_i \cdot Y_{1i} + (1 - D_i) \cdot Y_{0i} \\
\text{i.e.} \quad Y_i &= \begin{cases} Y_{1i} & \text{if } D_i = 1 \\ Y_{0i} & \text{if } D_i = 0 \end{cases}
\end{aligned}
$$

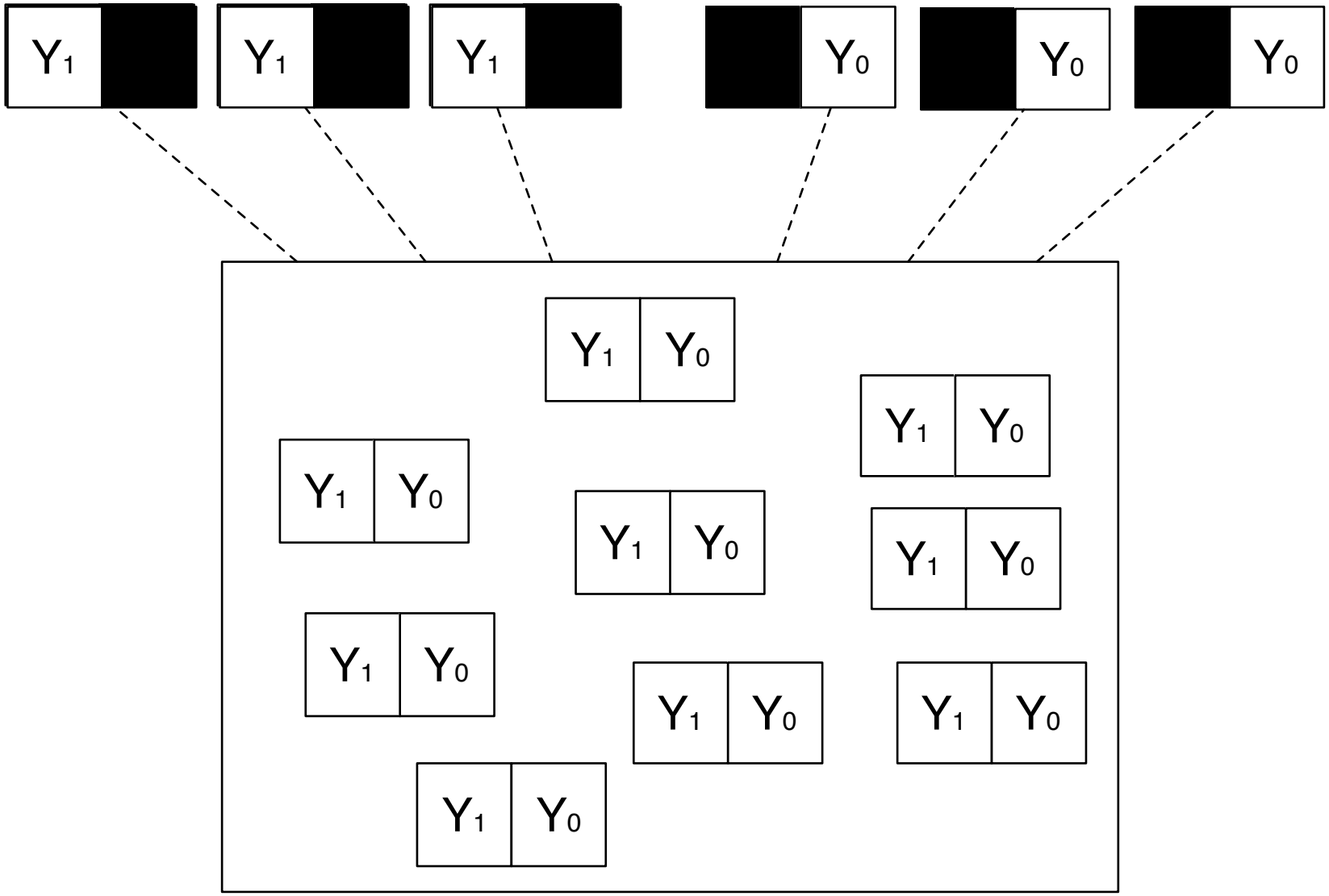
The punchline:

- *A priori* each potential outcome could be observed (manipulability!)
- After treatment assignment, one is observed, the other is counterfactual

Neyman Urn Model



Neyman Urn Model



Causal Inference as a Missing Data Problem

Imagine a population with 4 units:

i	D_i	Y_i	Y_{1i}	Y_{0i}
1	1	3	3	0
2	1	1	1	1
3	0	0	1	0
4	0	1	1	1

We take the values of both Y_{1i} and Y_{0i} to be real and fixed for all i

But we can only observe one of them for any i ...

This is known as the fundamental problem of causal inference (FPCI)

Causal Inference as a Missing Data Problem

... because of the FPCI we see only this:

i	D_i	Y_i	Y_{1i}	Y_{0i}
1	1	3	3	?
2	1	1	1	?
3	0	0	?	0
4	0	1	?	1

The goal:

1. Define causal estimands in terms of potential outcomes (previous table)
2. Estimate them using observable data on this slide (present table)

Essentially: fill in the missing counterfactuals as best as possible!

Esti-what?

Estimand:
$\rightsquigarrow$ Unobserved population parameter or function.

Estimator:
$\rightsquigarrow$ A function that can be applied to observed data.

Estimate:
$\rightsquigarrow$ A specific output of said function.

Unit-Level Causal Estimands

Individual Treatment Effect:

$$\tau_i = Y_{1i} - Y_{0i}$$

Read: Effect of the causal variable on outcome for unit i, defined by the comparison of two unit-level potential outcomes.

This cannot be observed, and is also very hard to estimate:

- We cannot observe both potential outcomes Y_{1i} and Y_{0i} for the same unit i.
- Hard to reliably fill in the missing potential outcome for any one unit i.

Group-Level Causal Estimands

Consider a fixed group (population*) of units $i = 1, \ldots, N$

Values of the potential outcomes for this population can be represented as two only partially observed vectors:

$$\mathbf{Y}_1 = (Y_{11}, Y_{12}, \ldots, Y_{1N})$$
$$\mathbf{Y}_0 = (Y_{01}, Y_{02}, \ldots, Y_{0N})$$

A group-level causal estimand is a comparison of $\mathbf{Y}_1$ and $\mathbf{Y}_0$

A common choice is a difference of their expected values.

*Note: The use of population here is possibly not what you are used to! We actually mean the particular realized sample. We can talk more about this if you want, but the idea is that there are two distinct sampling processes going on here – sampling from a super-population that generates the study sample, and sampling from potential outcomes of the realized sample. We are primarily interested in the latter, not the former.

Causal Estimand: The ATE

Average Treatment Effect:

$$\tau_{ATE} = \frac{1}{N}\sum_{i=1}^{N}(Y_{1i} - Y_{0i})$$

or equivalently

$$\tau_{ATE} = \mathbb{E}[Y_{1i} - Y_{0i}]$$

Read: Effect of the causal variable on outcome on average for the entire group, defined by the comparison of two group-level averages of potential outcomes.

Note: We represent the estimand as a greek letter (in this case τ, but could be anything). We typically represent an estimator for that estimand as a greek letter with something on top (e.g. $\tilde{\tau}$ or $\hat{\tau}$). An estimate will be a realised number (interval, etc.).

Causal Estimand: The ATT

Average Treatment Effect on the Treated:

$$\tau_{ATT} = \frac{1}{N_1} \sum_{i=1}^{N} D_i(Y_{1i} - Y_{0i}) \quad \text{where} \quad N_1 = \sum_{i=1}^{N} D_i$$

or equivalently

$$\tau_{ATT} = \mathbb{E}[Y_{1i} - Y_{0i} | D_i = 1]$$

(Note: The mathematical symbol | means "conditional on".)

Read: Effect of the causal variable on outcome on average for those units in the treated group, defined by the comparison of two group-level averages of potential outcomes.

What's going on here?

- When would $\tau_{ATT} \neq \tau_{ATE}$? When D_i and Y_{di} are associated.
- Exercise: Define τ_{ATU}, the average treatment effect on the untreated (control) units:

  $\tau_{ATU} = \mathbb{E}[Y_{1i} - Y_{0i} | D_i = 0]$

Causal Estimand: The CATE

Conditional Average Treatment Effect:

$$\tau_{CATE}(x) = \mathbb{E}[Y_{1i} - Y_{0i} | X_i = x]$$

where X_i is a pre-treatment covariate for unit i (more soon)

Read: Effect of the causal variable on outcome on average for those units where $X_i = x$, defined by the comparison of two subgroup averages of potential outcomes.

This estimand sometimes goes by other names (e.g. local average treatment effect or LATE, which we will see in a few weeks).

Generally, this is an increasingly important area for causal inference (e.g. optimal policy targeting, tailored medical therapies, algorithmic advertising).

Illustration: Average Treatment Effect

Let's return to our population of 4 units:

i	D_i	Y_i
1	1	3
2	1	1
3	0	0
4	0	1
$\mathbb{E}[Y_i \mid D_i = 1]$		2
$\mathbb{E}[Y_i \mid D_i = 0]$		0.5
$\mathbb{E}[Y_i \mid D_i = 1] - \mathbb{E}[Y_i \mid D_i = 0]$		1.5

What is $\tau_{ATE} = \mathbb{E}[Y_{1i} - Y_{0i}]$?

Naïve estimator:

$$\begin{aligned} \tilde{\tau}_{ATE} &= \mathbb{E}[Y_i \mid D_i = 1] - \mathbb{E}[Y_i \mid D_i = 0] \quad \text{(observed difference in means)} \\ &= \frac{3+1}{2} - \frac{0+1}{2} = 1.5 \quad \text{Could this be wrong?} \end{aligned}$$

Illustration: Average Treatment Effect

Let's return to our population of 4 units:

i	D_i	Y_i	Y_{1i}	Y_{0i}	τ_i
1	1	3	3	0	3
2	1	1	1	1	0
3	0	0	1	0	1
4	0	1	1	1	0
$\mathbb{E}[Y_{1i}]$			1.5		
$\mathbb{E}[Y_{0i}]$				0.5	
$\mathbb{E}[Y_{1i} - Y_{0i}]$					1

$$\tau_{ATE} = \mathbb{E}[Y_{1i} - Y_{0i}] = \mathbb{E}[\tau_i] = \frac{3+0+1+0}{4} = 1.$$

But recall that $\tilde{\tau}_{ATE} = 1.5$

Why does $\tau_{ATE} \neq \tilde{\tau}$? When would they be equal?

Illustration: Average Treatment Effect on the Treated

Again suppose we observe a population of 4 units:

i	D_i	Y_i	Y_{1i}	Y_{0i}	τ_i
1	1	3	3	0	3
2	1	1	1	1	0
3	0	0	1	0	1
4	0	1	1	1	0
$\mathbb{E}[Y_{1i} \mid D_i = 1]$			2		
$\mathbb{E}[Y_{0i} \mid D_i = 1]$				0.5	
$\mathbb{E}[Y_{1i} - Y_{0i} \mid D_i = 1]$					1.5

$$\tau_{ATT} = \mathbb{E}[Y_{1i} - Y_{0i} \mid D_i = 1] = \mathbb{E}[\tau_i \mid D_i = 1] = \frac{3+0}{2} = 1.5.$$

Average Treatment Effect on the Treated

In this example, why does $\tau_{ATT} \neq \tau_{ATE}$?

Because $\mathbb{E}[Y_{1i}] \neq \mathbb{E}[Y_{1i}|D_i = 1]$ (and likewise for $\mathbb{E}[Y_{0i}]$)

That is, D_i and Y_{di} are associated

Stable Unit Treatment Value Assumption (SUTVA)

Recall: $Y_i = Y_{D_i}$, or equivalently $Y_i = D_i Y_{1i} + (1 - D_i) Y_{0i}$

This notation implicitly makes an assumption:

SUTVA:

$$Y_{(D_1, D_2, \ldots, D_N)i} = Y_{(D'_1, D'_2, \ldots, D'_N)i} \quad \text{if} \quad D_i = D'_i$$

Read: For each i given a fixed value of D_i, potential outcomes under all possible randomization vectors are equal.

SUTVA comprises two sub-assumptions:

1. No interference between units: Potential outcomes for a unit not affected by treatment status of other units

   $\rightarrow$ Violations: spill-over effects, contagion, dilution

2. No different versions of treatment (stability, consistency): Nominally identical treatments are in fact identical

   $\rightarrow$ Violations: variable levels of treatment, technical errors

Causal Inference Without SUTVA

Let $\mathbf{D} = (D_1, D_2)$ be a vector of binary treatments for $N = 2$.

How many different values can $\mathbf{D}$ possibly take?

$$(D_1, D_2) = (0,0) \text{ or } (1,0) \text{ or } (0,1) \text{ or } (1,1)$$

How many potential outcomes for unit 1?

$$Y_{(0,0)1},\ Y_{(1,0)1},\ Y_{(0,1)1}\ Y_{(1,1)1}.$$

How many ITEs are defined for unit 1?

$$\begin{aligned} &Y_{(1,1)1} - Y_{(0,0)1}, \quad Y_{(1,1)1} - Y_{(0,1)1}, \\ &Y_{(1,0)1} - Y_{(0,0)1}, \quad Y_{(1,0)1} - Y_{(0,1)1}, \\ &Y_{(1,1)1} - Y_{(1,0)1}, \quad Y_{(0,1)1} - Y_{(0,0)1}. \end{aligned}$$

How many observed outcomes for unit 1? Only one: $Y_1 = Y_{(D_1,D_2)1}$

Without SUTVA, causal inference is exponentially more difficult as $n \uparrow$.

The Identification Problem for Causal Inference

Identification:
In statistics, an estimand (parameter) is said to be identified if its value can, asymptotically, be uniquely mapped to observed data and unidentified otherwise.

In the language of research design, if an estimand is not identified, there are alternative explanations (mappings) connecting the observed data (the estimate) and the estimand of interest. (We will dive into this intuition in a moment).

Recall that in causal inference, estimands are typically population causal effects but the FPCI tells us that at least half of the potential outcomes are always missing.

Identification Strategy:
A combination of data and assumptions which allows us to identify a causal estimand by estimating ("filling in") the missing potential outcomes (usually at a group level) in expectation.

Selection Bias

Let's see this idea in practice.

Consider again the naïve difference in observed means in the treatment groups:

$$\underbrace{E(Y_i|D_i=1) - E(Y_i|D_i=0)}_{\text{Observed difference in average outcome measures}} = E(Y_{1i}|D_i=1) - E(Y_{0i}|D_i=0)$$

$$= \underbrace{E(Y_{1i}|D_i=1) - E(Y_{0i}|D_i=1)}_{\text{ATT}} + \underbrace{E(Y_{0i}|D_i=1) - E(Y_{0i}|D_i=0)}_{\text{Selection bias}}$$

Read: The same observed mean difference could be due to different combinations of the ATT (estimand!) and selection bias terms.

Thus ATT is unidentified from the naïve observed mean difference: it is not uniquely mapped from the observed data. We need more assumptions.

Correlation [association, observed difference] is not necessarily causation.

Selection Bias

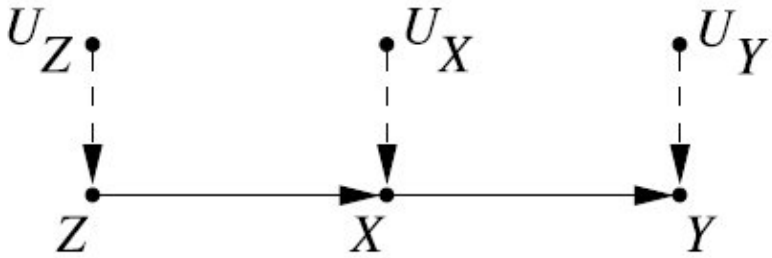
$$
\begin{aligned}
& E(Y_i|D_i = 1) - E(Y_i|D_i = 0) \\
&= \underbrace{E(Y_{1i}|D_i = 1) - E(Y_{0i}|D_i = 1)}_{\text{ATT}} + \underbrace{E(Y_{0i}|D_i = 1) - E(Y_{0i}|D_i = 0)}_{\text{Selection bias}}
\end{aligned}
$$

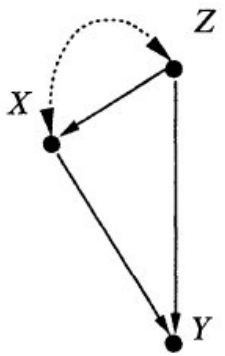
$E(Y_{0i}|D_i = 1) - E(Y_{0i}|D_i = 0)$ is called selection bias because, if it is not zero, treatment and control groups are systematically different in Y_{0i}. If non-zero, we might say the causal effect of D on Y is confounded.

Canonical example: Job training program

- Participants are self-selected from those in difficult labor situations
- Perhaps better resourced or more motivated individuals decide to take part
- Even in the absence of the program, post-training earnings for those people who participated would then have been higher than those for those who did not opt in: $E[Y_0|D = 1] - E[Y_0|D = 0] > 0$

Causal Diagrams





So far we have reasoned about causal effects using potential outcomes. An alternative (but intimately connected) framework is the graphical approach.

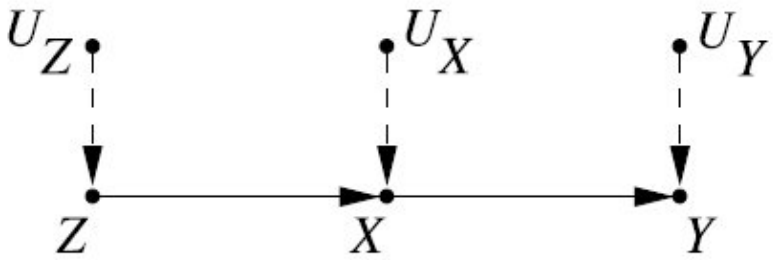
This uses causal diagrams, tools that allow us to:

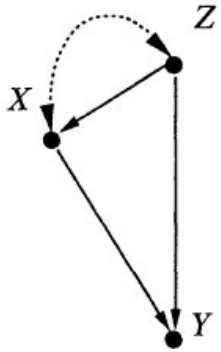
1. Specify the variables (observed and unobserved) we care about
2. Specify how those variables are connected
3. See what we can learn about causal effects, and with what assumptions.

This can help us to:

1. Study how conditioning affects our research designs
2. Create new research designs and methodologies.

Causal Diagrams as Directed Acyclic Graphs





Components of a causal diagram as a Directed Acyclic Graphs (DAG):

- Nodes: Representing “variables" (also called vertices)
- Directed Edges: Encoding one-way (causal) relationships

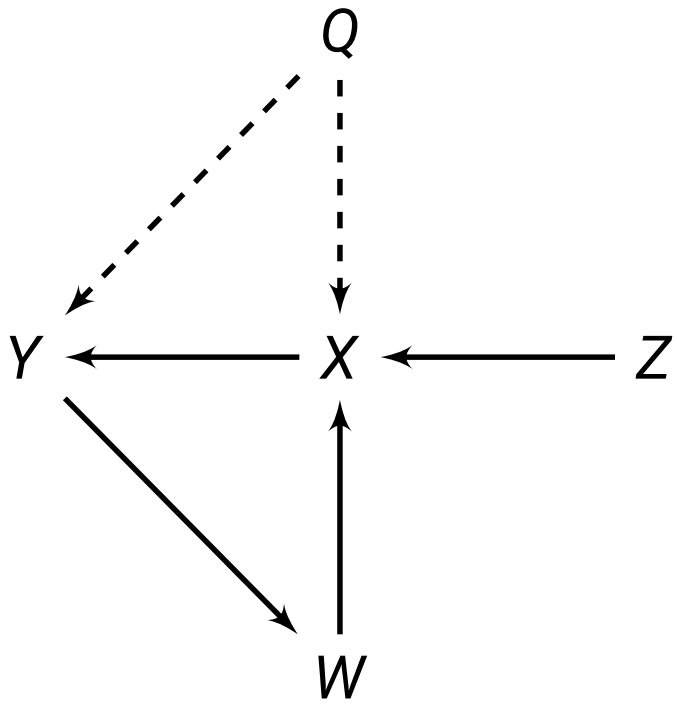
  $\rightarrow$ This implies nodes are ordered (each pair: tail $\rightarrow$ head)

  $\rightarrow$ These connections can be observed (solid) or unobserved (dashed)

Features of a DAG:

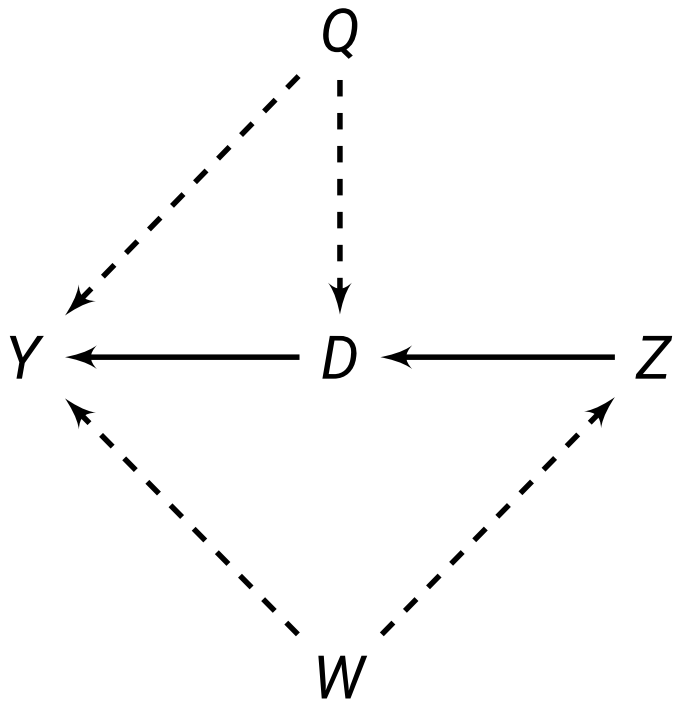
- Acyclic: No directed cycles (e.g. A does not terminate A)
- Non-connections: The absence of relationships between variables

Directed Acyclic Graphs: Example



"That's no DAG, it's a space station"

Directed Acyclic Graphs: Example

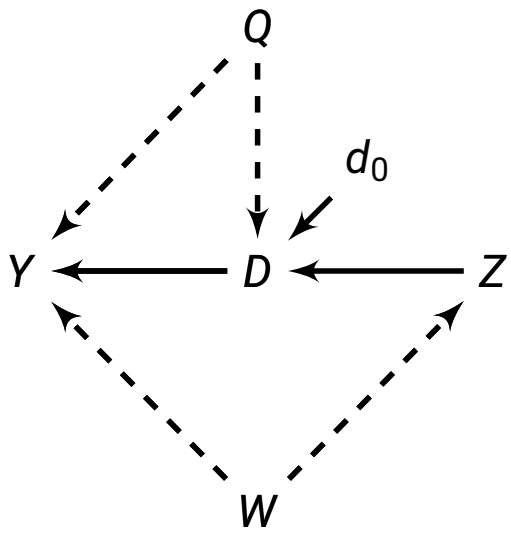


That's better! What can we learn from this DAG?

- $Z \rightarrow Y$ is confounded by W
- $D \rightarrow Y$ is confounded by Q
- $Z \rightarrow D$ is identified

But only if our DAG is correct!

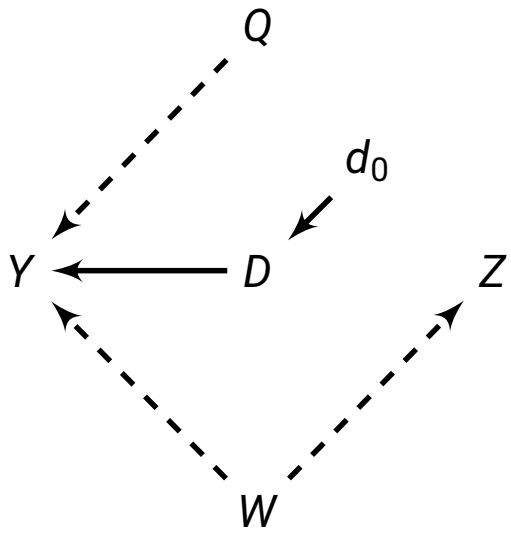
Representing Interventions



Treatments (interventions) are represented by the ***do***() operator.

For example, $do(d_0)$ holds $D = d_0$ exogenously.

Identification



ATE of D on Y defined as the average difference in Y between two interventions:

$$\mathbb{E}[Y \mid do(d_1)] - \mathbb{E}[Y \mid do(d_0)]$$

Problem: Can this be identified without an explicit intervention?

Insight: If the DAG is equivalent with and without $do()$, yes.

Generally: We can identify the effect of D on Y if all back-door paths are blocked.

Assignment Mechanism

Assignment Mechanism:
The procedure that determines the treatment status of each unit.

Identification in most causal inference methods relies on restricting the (assumed) assignment mechanism

For example, if we are willing to assume that treatment assignment is independent of potential outcomes under no treatment, then:

$$E(Y_{0i}|D_i = 1) - E(Y_{0i}|D_i = 0) = 0$$

Read: Selection bias is zero and the observed mean difference is (in expectation) equal to ATT (and also ATE in this case)

Different Assignment Mechanisms

Imbens and Rubin (2015, Ch. 3) present three assumptions about assignment mechanisms (for each unit) that provide the grounds for identification:

1. Individualistic: Assignment does not depend on the covariates or potential outcomes for other units.
2. Probabilistic: There is a nonzero probability of each treatment value, for every unit.
3. Unconfounded: Assignment does not depend on potential outcomes.

Assuming the above, we can distinguish:

- (Randomized) Experiments: The assignment mechanism is both known and controlled by the researcher (usually with some flavour of randomization).
- Observational Studies: The assignment mechanism is not known to, or not under the control of, the researcher.

Our Key Assignment Mechanisms

Randomised Experiments: These come in many flavours, many of which we won't even discuss, for example...

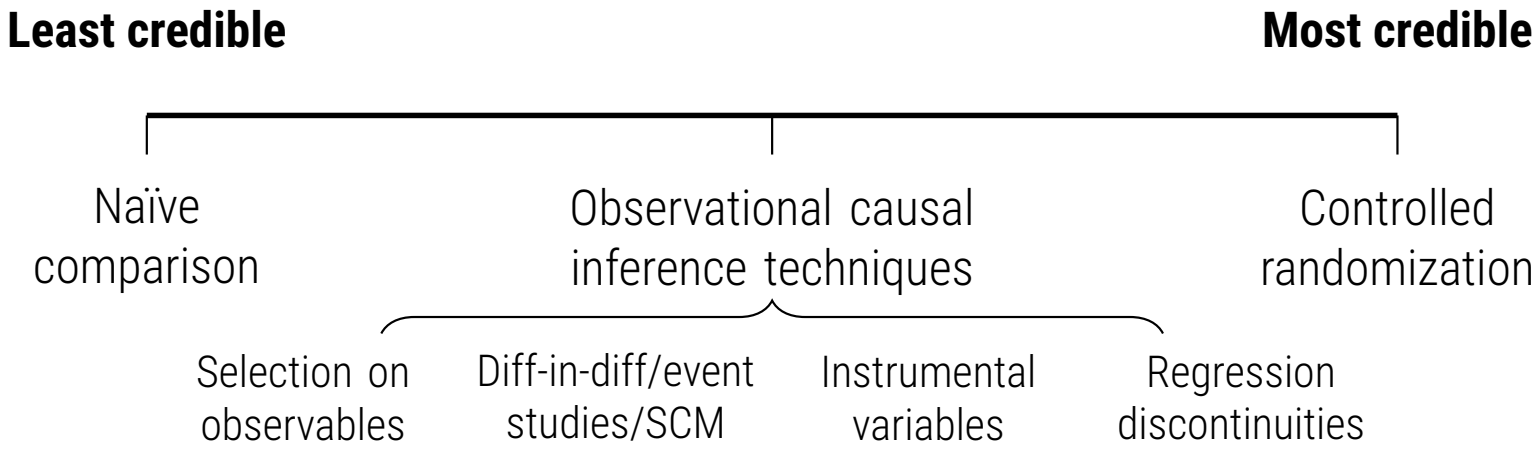
- Within designs, between designs
- Unit-randomized, cluster-randomized, dynamic randomization
- Crossover designs, stepped-wedge designs, etc. etc. etc.

Observational Studies:

- Selection on observables using regression, matching, etc.
- Instrumental variables, shift-share designs, etc.
- Sharp and fuzzy regression discontinuity designs, etc.
- Temporal: Diff-in-diff, event studies, synthetic control methods

Note: The first three observational designs are sometimes referred to as "design-based" methods. We'll talk about this distinction later.

The Continuum of Credibility™



Key point: The art (and science) of applied causal inference is making defensible assumptions. There is no 'magic' solution to the fundamental problem of causal inference, only assumptions all the way down!

"It's assumptions all the way down"

Key Ideas so Far

- Think about causal effects in terms of potential outcomes, not realized (observed) outcomes
- Observed association is neither necessary nor sufficient for causality – focused on one big problem, selection bias
- The graphical approach is an alternative framework for thinking about causal models
- Learning about causal effects should start from understanding the assignment mechanism for treatment
- Evaluate the plausibility of your assumptions to understand the credibility of your conclusions