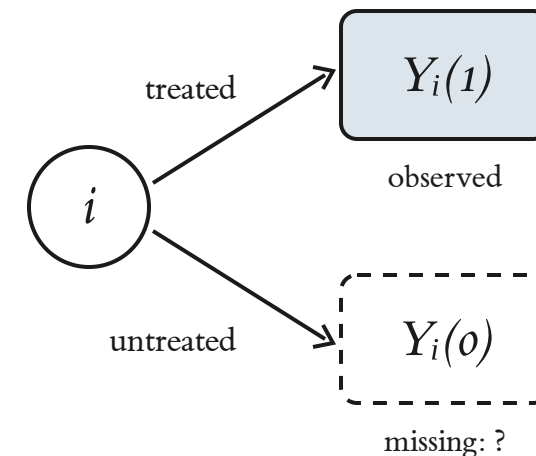


# Causal inference

*Which comparison answers the question “what if we intervene?”*

1 Effects and selection → 2 Designs → 3 Adjustment



Lecture 1

Probability theory

*What can happen, and how does information change it?*

Lecture 2

Data, variables & moments

*How do measurements vary?*

Lecture 3

Statistics

*What can one sample tell us about a population?*

Lecture 4

Time series

*What changes when observations depend on their past?*

Lecture 5

**Causal inference**

*Which comparison identifies an effect?*

## OPENING QUESTION

# The treated score 16.67 points lower. Did the treatment hurt them?

Seven people, outcomes in invented points. People 1, 3 and 4 were treated.

| Person   | 1   | 2  | 3   | 4   | 5  | 6  | 7  |                               |
|----------|-----|----|-----|-----|----|----|----|-------------------------------|
| Treated? | yes | no | yes | yes | no | no | no |                               |
| Outcome  | 30  | 30 | 46  | 54  | 60 | 70 | 80 | treated mean 43.33, others 60 |

## YOUR ANSWER

Vote now. Part 1 answers it.

A

Yes, by 16.67

B

No

C

Cannot tell yet

## WHY START HERE

Programme evaluations, university premiums and pay gaps all start from a comparison like this one. The question is what else differs between the groups.

## LECTURE 5 · THE ROUTE

# Define the effect, then find a comparison that identifies it

## PART 1

## The effect we want and the gap we see

*What is a causal effect, and why does a difference in group averages not measure it?*

potential outcomes

ATE

ATT

selection bias

Notes 5.1–5.2



## PART 2

## Designs that remove selection

*How can the way treatment is assigned make selection bias zero, and for whom is the answer then valid?*

random assignment

ITT

compliers

LATE

external validity

Notes 5.3, 5.4, 5.7



## PART 3

## Adjusting for covariates

*Without randomisation, when does comparing like with like recover the effect, and when does “controlling for” make things worse?*

CIA

overlap

target weights

bad control

Notes 5.5–5.6

## BY THE END OF THIS LECTURE YOU CAN

- ✓ Define potential outcomes, ATE and ATT, and explain why one potential outcome per person is always missing.
- ✓ Derive  $\Delta = ATT + B$  and argue the likely sign of selection bias in an economic comparison.
- ✓ Explain what random assignment identifies, and compute ITT and the Wald ratio for compliers.
- ✓ Judge an adjusted comparison: CIA, overlap, the right weights, and why occupation is a bad control for pay.

## PART 1 OF 3

# The effect we want and the gap we see

*What is a causal effect, and why does a difference in group averages not measure it?*

## KEY TERMS

potential outcomes

ATE

ATT

selection bias

[Notes 5.1–5.2 ↗](#)

## WHY IT MATTERS

People choose programmes, universities and jobs partly on their prospects, so participants differ from non-participants before any treatment. A naive comparison could close a helpful programme, or credit a university with earnings its students would have had anyway.

## IN THIS PART

- 5.1 Potential outcomes: one person, two outcomes, one observed
- 5.1 ATE and ATT: whose effect?
- 5.2 Every observed gap is ATT plus selection bias
- ▶ Animation: how selection bias can hide a positive effect
- 5.2 Selection can push a comparison either way

## 5.1 | TWO WORLDS FOR ONE PERSON

# Each person has two potential outcomes; we only ever see one

## DEFINITION

## Potential outcomes

For unit  $i$  and treatment  $D_i \in \{0, 1\}$ :  $Y_i(1)$  is the outcome under treatment,  $Y_i(0)$  without it. The individual effect is  $\tau_i = Y_i(1) - Y_i(0)$ . The outcome not realised is the **counterfactual**.

## KEY FORMULA IDENTITY

(1)

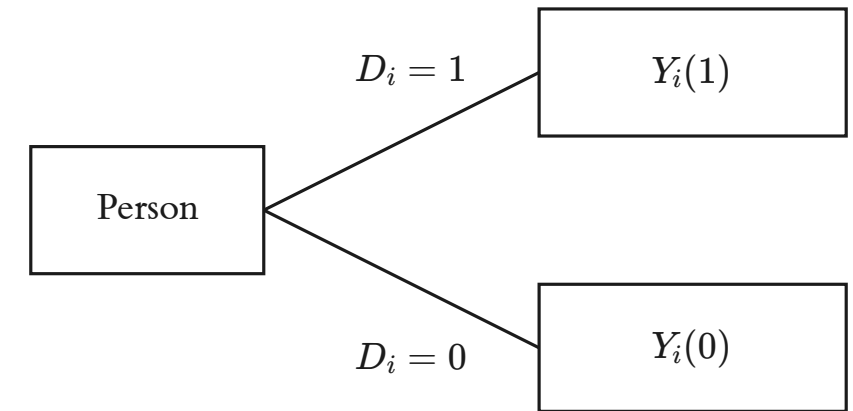
## The observed outcome

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

**IN WORDS** we observe  $Y_i(1)$  if treated and  $Y_i(0)$  if not, never both.

**NEEDS** a well-defined treatment, and no effect of one person's treatment on another's outcome.

One person, two possible outcomes, one observation



*So  $\tau_i$  is never observed for anyone: causal inference is a missing-data problem.*

## 5.1 | AVERAGE EFFECTS

## ATE and ATT average the same effects over different people

## KEY FORMULA

## DEFINITION

(2)

## Average treatment effects

$$ATE = E[Y(1) - Y(0)],$$

$$ATT = E[Y(1) - Y(0) \mid D = 1]$$

**IN WORDS** the average effect over everyone, and over the treated. Both terms of an effect average over the same people.

**ATE** and **ATT** differ when effects vary and the treated are not typical:  $ATE = P(D = 1) ATT + P(D = 0) ATU$ .

The notes' toy population (invented points); bold = observed

| Person | $D$ | $Y(0)$    | $Y(1)$    | $\tau$ |
|--------|-----|-----------|-----------|--------|
| 1      | 1   | 20        | <b>30</b> | 10     |
| 2      | 0   | <b>30</b> | 32        | 2      |
| 3      | 1   | 40        | <b>46</b> | 6      |
| 4      | 1   | 50        | <b>54</b> | 4      |
| 5      | 0   | <b>60</b> | 64        | 4      |
| 6      | 0   | <b>70</b> | 80        | 10     |
| 7      | 0   | <b>80</b> | 78        | -2     |

$$ATE = 34/7 \approx 4.86; ATT = (10 + 6 + 4)/3 \approx 6.67.$$

**TAKEAWAY** ATT subtracts the treated people's own no-treatment outcomes, not the control group's observed mean.

## 5.2 | AN IDENTITY, NOT AN ASSUMPTION

# Every observed gap is the effect plus selection bias

**KEY FORMULA** IDENTITY

(3)

Observed difference = effect + selection

$$\begin{aligned}\Delta &= E[Y \mid D = 1] - E[Y \mid D = 0] \\ &= \text{ATT} + B\end{aligned}$$

**IN WORDS** add and subtract  $E[Y(0) \mid D = 1]$ , the treated group's unseen no-treatment mean.

**DEFINITION**

## Selection bias

$B = E[Y(0) \mid D = 1] - E[Y(0) \mid D = 0]$ : how the treated would have fared untreated, compared with the untreated.

**WORKED EXAMPLE** *the toy population*

The treated would have averaged 110/3 untreated; the controls average 60.

$$B = 110/3 - 60 = -70/3 \approx -23.33$$

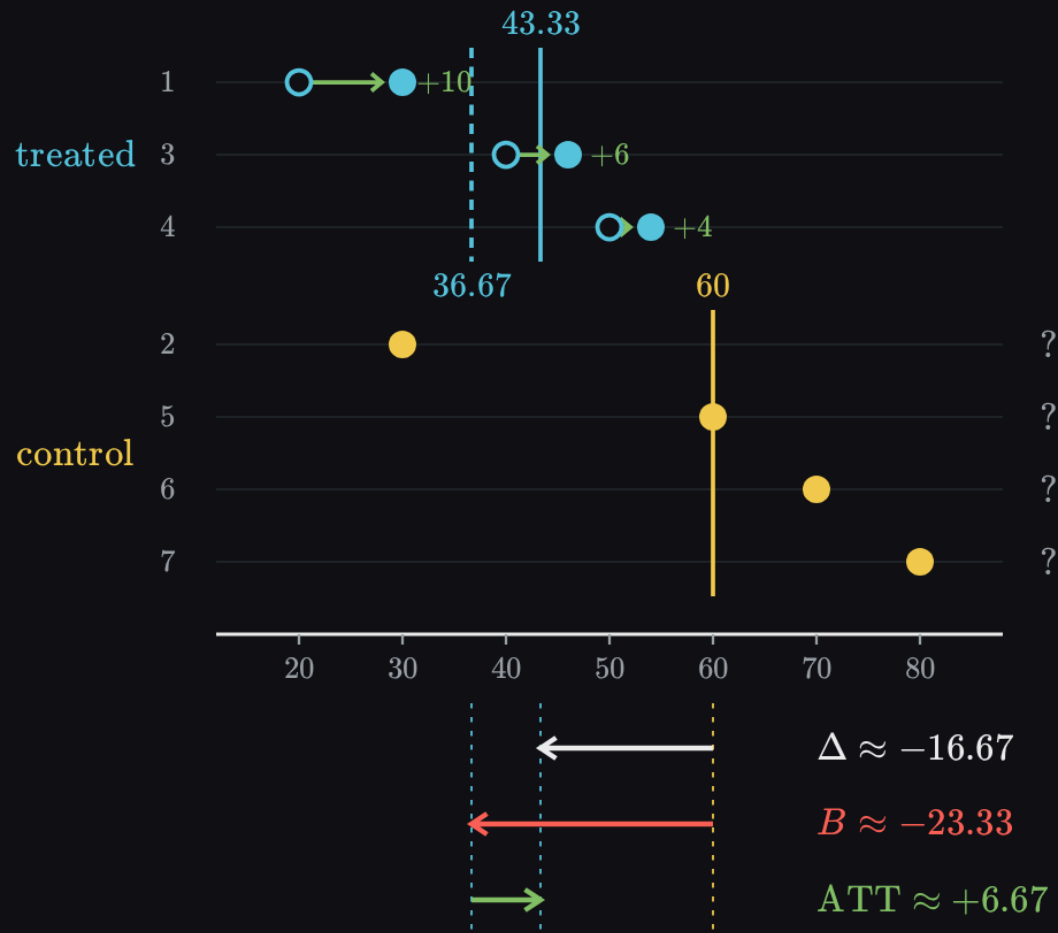
$$\Delta = 20/3 - 70/3 = -50/3 \approx -16.67$$

**COMMON MISTAKE**

Selection bias is not sampling bias. It arises even when the whole population is observed.

**TAKEAWAY** Identification is whatever justifies learning the unseen  $E[Y(0) \mid D = 1]$ .

The treated score lower. Did treatment hurt them?



observed: treated  $130/3 \approx 43.33$ , control 60

$$\Delta = 130/3 - 60 = -50/3 \approx -16.67$$

revealed:  $Y(0) = 20, 40, 50$  for the treated

$$ATT = (10 + 6 + 4)/3 = 20/3 \approx 6.67$$

$$\Delta = ATT + B$$

$$-16.67 = 6.67 + (-23.33)$$

$B = 110/3 - 60$ : lower baselines, not harm

A negative comparison can hide a positive effect;  
recovering ATT needs evidence about  $B$ .

## 5.2 | READING THE SIGN OF B

# Selection can push a comparison either way

**WORKED EXAMPLE** *a university premium (illustration)*

Graduates of university O earn £4,000 more than those of X:  $\Delta = \text{ATT} + B$ . If O's students would have earned more even at X, then  $B > 0$  and  $\text{ATT} < £4,000$ ;  $B = £3,000$  would leave £1,000.

**WORKED EXAMPLE** *a training programme (illustration)*

Participants earn £8.80 a week less:  $\Delta = -£8.80$ . If participant selection contributes  $-£12$ , then

$$\text{ATT} = -8.80 - (-12) = +£3.20 \text{ a week.}$$

**WHY ECONOMISTS CARE**

Closing a programme or justifying higher fees needs ATT, not  $\Delta$ . Economists expect selection wherever people choose treatment on their prospects.

**CHECK**

$\Delta = 12$ . Decompose it with  $\text{ATT} = 5$ , and with  $\text{ATT} = 15$ .

**ANSWER**

$B = 7$  and  $B = -3$ . Both satisfy the identity; the data alone cannot choose.

## PART 2 OF 3

# Designs that remove selection

*How can the way treatment is assigned make selection bias zero, and for whom is the answer then valid?*

## KEY TERMS

random assignment

ITT

compliers

LATE

external validity

[Notes 5.3, 5.4, 5.7 ↗](#)

SO FAR  $\Delta = ATT + B$  always holds, and the data alone cannot say how large  $B$  is.

## WHY IT MATTERS

Randomised trials, offers and eligibility rules are the economist's main evaluation tools. But people decline offers, effects spill over at scale and studies are short; each changes what is identified, and whether it carries over to a national policy.

## IN THIS PART

- 5.3 Random assignment:  $B = 0$  on average
- 5.3, 5.7 Compare by the offer: ITT, then LATE for compliers
- ▶ Animation: the Wald ratio recovers the effect for compliers
- 5.4 A credible trial need not transport
- 5.4 Natural experiments and a threshold rule

## 5.3 | RANDOM ASSIGNMENT

# Random assignment sets selection bias to zero on average, not in every draw

## KEY FORMULA ASSUMPTION

(4)

## Random assignment

$$D \perp\!\!\!\perp \{Y(0), Y(1)\}$$

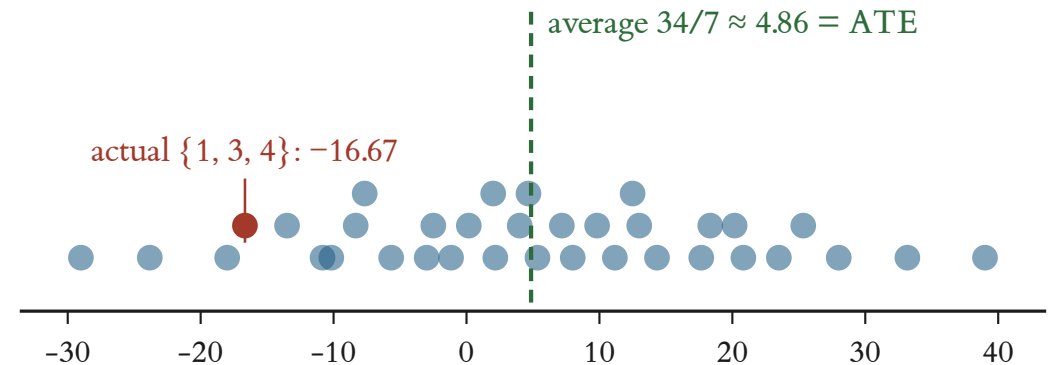
**IN WORDS** assignment is independent of the pair of potential outcomes; then  $B = 0$  and  $\Delta = \text{ATT} = \text{ATE}$ .

**NEEDS** the design assigns treatment with the same positive probability to every unit.

## COMMON MISTAKE

It does not make the observed  $Y$  independent of  $D$ : a treatment effect is exactly how assignment changes  $Y$ .

All 35 ways to treat three of the seven people:



Unbiased over allocations; one draw can still be far off. Inference (Lecture 3) measures that chance variation.

## 5.3, 5.7 | OFFERS AND TAKE-UP

# People refuse offers: compare by the offer, then rescale for compliers

## KEY FORMULA DEFINITION

(5)

## Intention to treat

$$\text{ITT} = E[Y \mid Z = 1] - E[Y \mid Z = 0]$$

**IN WORDS**  $Z$  is the randomised offer,  $D$  the treatment received. ITT is the effect of offering, take-up included.

Types by receipt without and with the offer,  $(D(0), D(1))$

never-taker (0, 0)      **complier** (0, 1)

always-taker (1, 1)      defier (1, 0): ruled out

## COMMON MISTAKE

Comparing takers with non-takers: take-up selects on motivation, so it forfeits the randomisation.

## KEY FORMULA RESULT

(6)

## The Wald ratio: LATE

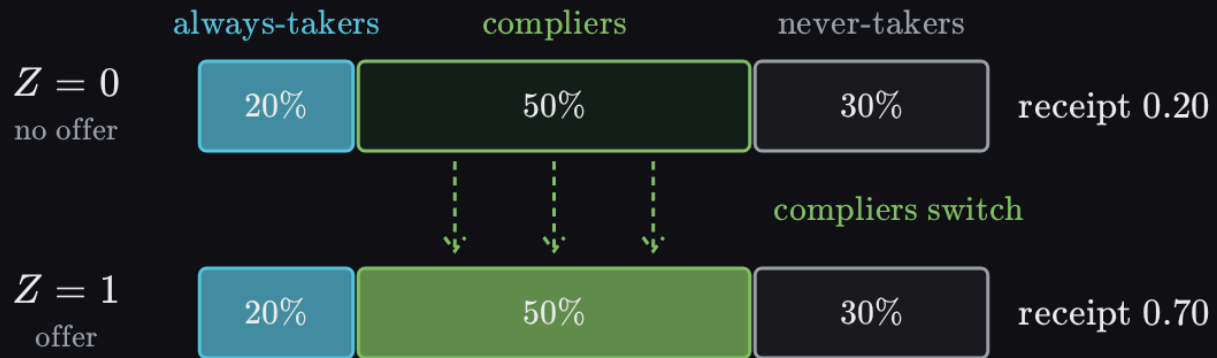
$$\text{LATE} = \frac{E[Y \mid Z = 1] - E[Y \mid Z = 0]}{E[D \mid Z = 1] - E[D \mid Z = 0]}$$

**IN WORDS** the offer's effect on outcomes over its effect on take-up: the average effect for compliers.

**NEEDS** a random offer that affects  $Y$  only through  $D$  (exclusion), changes take-up, and creates no defiers.

*LATE is local: a different offer moves different people, and its compliers may gain more or less than everyone else.*

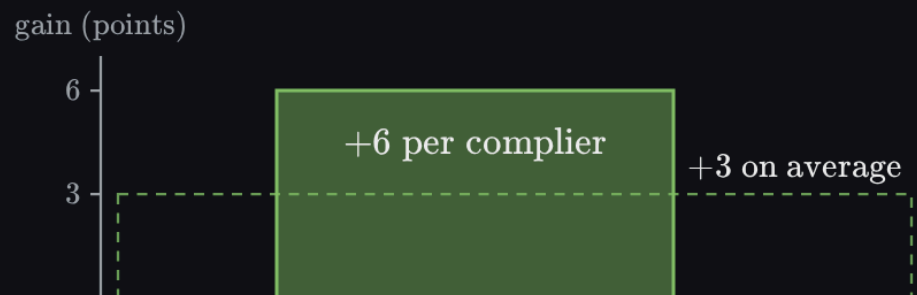
An offer raises outcomes by 3. What does treatment do?



random offer: the same mix in both groups

receipt 0.20  $\rightarrow$  0.70

$\Delta_D = 0.50$ , the complier share



$\Delta_Y = 3$  points, over everyone

only compliers switch:  $0.50 \times \text{LATE} = 3$

$$\text{LATE} = \Delta_Y / \Delta_D = 3 / 0.50 = 6$$

the effect for compliers, not for everyone

## 5.4 | INTERNAL AND EXTERNAL VALIDITY

# A credible trial need not transport to another setting

## DEFINITION

## Internal validity

Does the study credibly identify its stated effect for the units and treatment actually studied? Check assignment, take-up, measurement, missing outcomes, interference.

## DEFINITION

## External validity

Does the result transport to another population, setting, implementation or time horizon? Check people, institutions, scale, follow-up.

## WHY A GOOD TRIAL MAY NOT CARRY OVER

- **Spillovers.** If treating one person affects others, outcomes are  $Y_i(d_i, \mathbf{d}_{-i})$ : a pilot and a national policy are different interventions.
- **Surrogates.** An exam score next term is not earnings in ten years.
- **Being observed.** Hawthorne and placebo effects: define the comparison; blind where possible.

*Good random assignment supports internal validity without establishing external validity.*

## 5.4 | NATURAL EXPERIMENTS

## A natural experiment is an argument, not a label

## DEFINITION

## Natural and quasi-experiments

Variation from outside the researcher's control that resembles random allocation in the relevant respect. Name the source of variation, the comparison, the competing pathways and the local target.

## COMMON MISTAKE

“The weather was random.” A storm's exposure varies with location and housing, and it can hit income directly, not only through moving house.

WORKED EXAMPLE *a class-size threshold rule*

With classes capped at 40, enrolment  $n$  predicts class size

$$q(n) = \frac{n}{\lceil n/40 \rceil}$$

$$n = 40 \quad 40 \quad n = 80 \quad 40$$

$$n = 41 \quad 20.5 \quad n = 81 \quad 27$$

One more pupil drops the predicted size by 13. Compare schools just either side, if nothing else jumps at 40 and schools cannot sort.

## PART 3 OF 3

# Adjusting for covariates

*Without randomisation, when does comparing like with like recover the effect, and when does “controlling for” make things worse?*

## KEY TERMS

CIA

overlap

target weights

bad control

[Notes 5.5–5.6 ↗](#)

SO FAR randomisation sets  $B = 0$  by design; a random offer identifies ITT and, under conditions, LATE for compliers.

## WHY IT MATTERS

Most economic data are observational, and adding controls is routine. Whether the gender pay gap “disappears within occupations” turns on exactly this question.

## IN THIS PART

- 5.5 CIA: as good as random within covariate cells
- 5.5 Same cell effects, different targets
- 5.5 No comparison without overlap
- 5.6 Composition and the pay gap
- 5.6 Control for confounders, not for outcomes of treatment
- Animation: equal pay within a profession can hide a negative effect

## 5.5 | SELECTION ON OBSERVABLES

# Conditional independence: within covariate cells, assignment is as good as random

## KEY FORMULA ASSUMPTION

(7)

## Conditional independence (CIA)

$$D \perp\!\!\!\perp \{Y(0), Y(1)\} \mid X$$

**IN WORDS** within each group of pre-treatment covariates  $X$ , assignment is independent of the potential outcomes.

**NEEDS** an argument about how treatment was assigned. Then  $\tau(x) = m_1(x) - m_0(x)$ , the within-cell difference of observed means.

## WHAT CIA DOES NOT SAY

- CIA does not say potential outcomes are constant within a cell.
- No number of controls makes it true, and balance on observed variables does not test it.

WORKED EXAMPLE *military service and earnings*

Compare applicants who enlisted with applicants who did not, holding recorded screening characteristics fixed.  
Threat: persistence or reliable attendance may drive both enlistment and later earnings.

## 5.5 | AVERAGING CELL EFFECTS

## The same cell effects give different answers for different targets

## KEY FORMULA RESULT

(8)

## Target weights

$$ATE = \sum_x \tau(x) P(X = x)$$

$$ATT = \sum_x \tau(x) P(X = x \mid D = 1)$$

**IN WORDS** ATE weights cells by the population; ATT by the treated.

## CHECK

Cell effects 3 and 9; population shares 0.7 and 0.3; treated shares 0.2 and 0.8. ATE and ATT?

## ANSWER

ATE = 4.8, ATT = 7.8.

WORKED EXAMPLE *two cells, half the population each*

Treatment probability 0.2 in cell  $X = 0$ , 0.8 in cell  $X = 1$ .

| Cell    | Population | Treated | $\tau(x)$ |
|---------|------------|---------|-----------|
| $X = 0$ | 0.5        | 0.2     | 2         |
| $X = 1$ | 0.5        | 0.8     | 6         |

$$ATE = 0.5(2) + 0.5(6) = 4$$

$$ATT = 0.2(2) + 0.8(6) = 5.2$$

The raw gap,  $23.2 - 12 = 11.2$ , adds baseline selection of 6.

## 5.5 | OVERLAP AND SPARSE CELLS

# No comparison without overlap, and fine cells empty fast

**KEY FORMULA** ASSUMPTION

(9)

**Overlap (common support)**

$$0 < P(D = 1 \mid X = x) < 1$$

**IN WORDS** at every covariate value in the target, some units are treated and some are not.

**NEEDS** for ATT, untreated comparisons wherever treated people are. It can be checked in the data.

**COMMON MISTAKE**

If everyone with  $X = 1$  is treated,  $\tau(1)$  is not identified: a fitted model fills the cell only by extrapolation, and trimming changes the estimand.

**WORKED EXAMPLE** 100 points spread evenly over  $k$  variables

Expected count in a cell of side 0.2 of each range:

| $k$   | 1  | 2 | 3   | 4    | 5     |
|-------|----|---|-----|------|-------|
| count | 20 | 4 | 0.8 | 0.16 | 0.032 |

To expect 5 points with  $k = 5$ , each side must be  $(0.05)^{1/5} \approx 0.55$  of its range.

## 5.6 | COMPOSITION

# Composition alone can make the overall gap exceed every within-group gap

PUBLISHED STATISTIC ONS 2021, provisional, UK medians

Median hourly-pay gaps: 7.9% full-time, −2.7% part-time, 15.4% overall. They compare pay across jobs, not people in the same job.

## COMMON MISTAKE

Medians cannot be averaged across groups: the law of iterated expectations is about means.

WORKED EXAMPLE *a hypothetical mean model*

|           | Men   | Women | Men's share | Women's share |
|-----------|-------|-------|-------------|---------------|
| Full-time | 14.81 | 13.64 | 0.85        | 0.58          |
| Part-time | 9.07  | 9.31  | 0.15        | 0.42          |

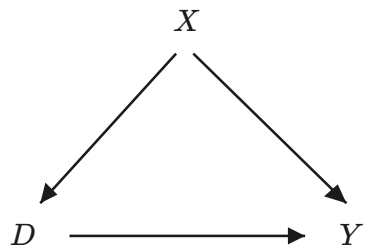
Means 13.949 and 11.8214: a gap of 15.25%, above both within-stratum gaps (7.90% and −2.65%), because more women work in the lower-paid stratum.

## 5.6 | BAD CONTROLS

# Control for confounders, not for outcomes of the treatment

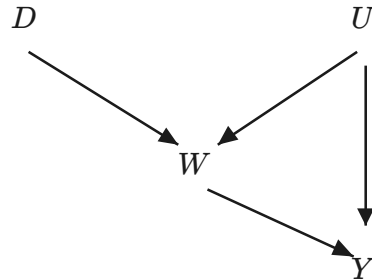
**A confounder:**  $X$  causes both  $D$  and  $Y$ . Conditioning on it can help.

A pre-treatment common cause



**A bad control:**  $W$  is caused by  $D$ . Holding it fixed changes who is compared.

Selection through an outcome



KEY FORMULA

IDENTITY

(10)

Comparing within an occupation

$$\Delta_W = \tau_W + B_W$$

**IN WORDS** the gap among professionals = an effect for those who would be professional if treated + a difference in who becomes professional.

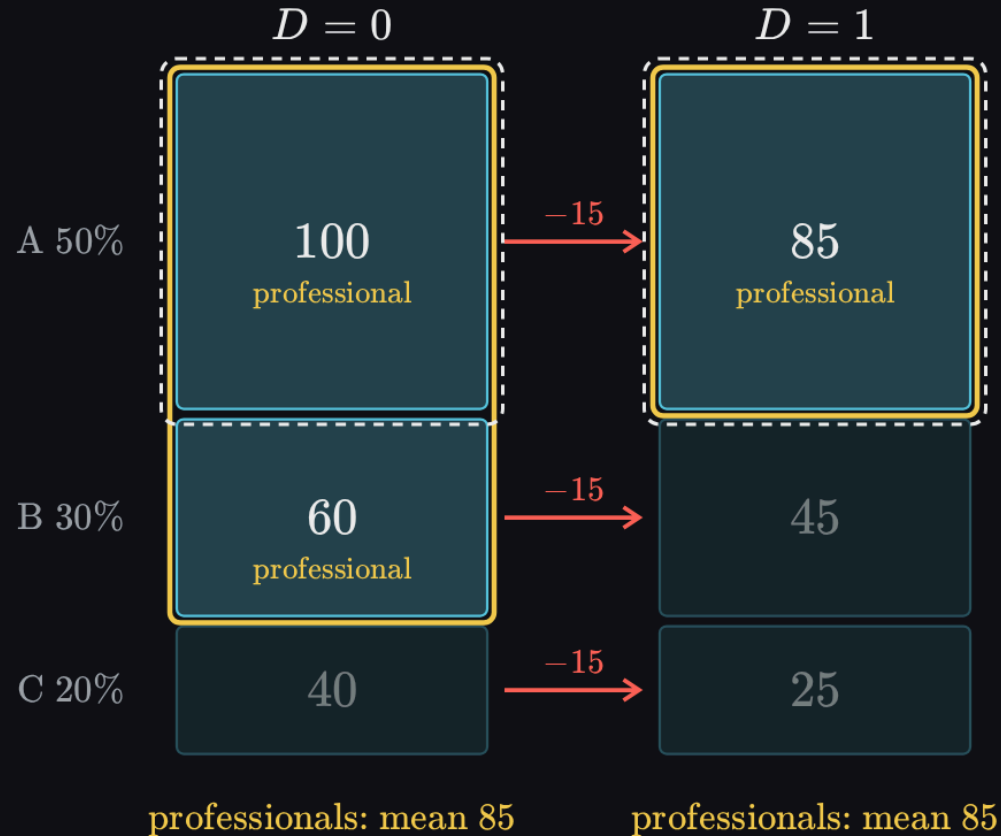
**NEEDS**  $D$  independent of all potential earnings and occupations jointly.

DEFINITION

## Bad control

A variable caused by the treatment, such as occupation for pay. Conditioning on it can hide part of the effect or create selection.

## Equal pay within a profession: no effect?



every type:  $Y(1) - Y(0) = -15$

professional under  $D = 0$ : types A and B

professional under  $D = 1$ : type A only

$$D = 0: (0.5 \cdot 100 + 0.3 \cdot 60) / 0.8 = 85$$

$$D = 1: 85; \text{ gap } 85 - 85 = 0$$

$$\tau_W = 85 - 100 = -15 \text{ (type A in both states)}$$

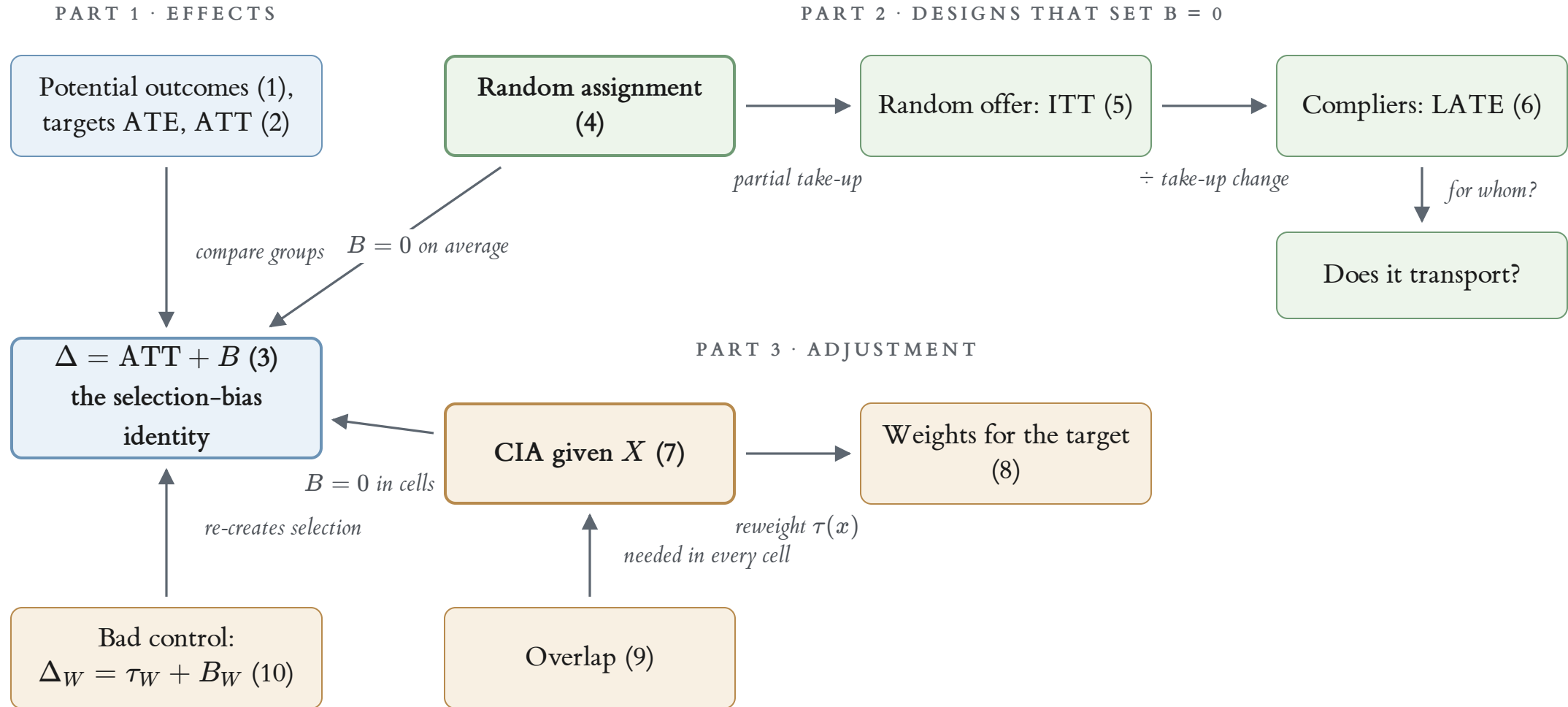
$$B_W = 100 - 85 = +15 \text{ (} D = 0 \text{ adds type B)}$$

$$0 = -15 + 15$$

Conditioning on a job changes who is compared.

## REVIEW · HOW THE IDEAS CONNECT

## One identity, three ways to deal with selection



## REVIEW · SUMMARY SHEET

## What to take away: the terms and the ten formulas

## KEY TERMS

## 1 EFFECTS AND SELECTION

**Potential outcomes**  $Y_i(1), Y_i(0)$ : one unit's outcomes with and without treatment; one is always missing.

**ATE** the average effect over everyone.

**ATT** the average effect among the treated.

**Selection bias** the groups' difference in no-treatment outcomes.

## 2 DESIGNS

**Random assignment** assignment independent of the potential outcomes.

**Complier** takes the treatment if and only if offered.

**LATE** the average effect for the compliers of one instrument.

**Internal and external validity** credible for the units studied; carries over to other settings.

## 3 ADJUSTMENT

**CIA** within covariate cells, assignment independent of potential outcomes.

**Overlap** treated and untreated units at every target covariate value.

**Bad control** a variable the treatment changes; conditioning on it alters who is compared.

## KEY FORMULAS

## 1 EFFECTS AND SELECTION

(1) **The observed outcome** IDENTITY

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

binary treatment, consistency, no interference

(2) **Average treatment effects** DEFINITION

$$\text{ATE} = E[\tau], \text{ATT} = E[\tau \mid D = 1]$$

each average over one stated population

(3) **Observed difference = effect + selection**

IDENTITY

$$\Delta = \text{ATT} + B$$

always true when both groups exist

## 2 DESIGNS

(4) **Random assignment** ASSUMPTION

$$D \perp\!\!\!\perp \{Y(0), Y(1)\}$$

gives  $B = 0$ ,  $\Delta = \text{ATT} = \text{ATE}$  (population model)

(5) **Intention to treat** DEFINITION

$$\text{ITT} = E[Y \mid Z = 1] - E[Y \mid Z = 0]$$

identified when the offer  $Z$  is randomised

(6) **The Wald ratio: LATE** RESULT

$$\text{LATE} = \Delta_Y / \Delta_D = E[Y(1) - Y(0) \mid C]$$

random offer, exclusion, nonzero take-up change, no defiers

## 3 ADJUSTMENT

(7) **Conditional independence (CIA)** ASSUMPTION

$$D \perp\!\!\!\perp \{Y(0), Y(1)\} \mid X$$

then  $\tau(x) = m_1(x) - m_0(x)$  where both groups exist

(8) **Target weights** RESULT

$$\text{ATE} = \sum_x \tau(x) P(X = x)$$

$$\text{ATT} = \sum_x \tau(x) P(X = x \mid D = 1)$$

CIA and overlap in every cell; discrete  $X$

(9) **Overlap (common support)** ASSUMPTION

$$0 < P(D = 1 \mid X = x) < 1$$

checkable; for every target covariate value

(10) **Comparing within an occupation** IDENTITY

$$\Delta_W = \tau_W + B_W$$

given joint independence of  $D$

## REVIEW · BEFORE YOU REPORT A CAUSAL ESTIMATE

# Name the target, defend the assumption, then measure the uncertainty

## BEFORE YOU REPORT A CAUSAL ESTIMATE

1. The intervention, population, outcome and time horizon.
2. How treatment was assigned.
3. The identifying condition: randomisation, exclusion, or CIA with overlap.
4. Only then, inference suited to the design. A confidence interval does not cover a failed assumption.

RESULT *the opening question*

From the data alone: cannot tell. In the model, every treated person gained ( $ATT \approx +6.67$ ); their lower baselines ( $B \approx -23.33$ ) produced the gap.

## THE COURSE IN ONE LINE

Probability defines uncertainty; distributions and moments describe measurements; statistics studies sampling variation; time series accounts for dependence over time; causal inference decides which comparison answers an intervention question.

## PRACTISE

Notes 5.1–5.7 with worked solutions; the five interactive figures (potential outcomes, selection, the 35 allocations, target weights, pay mixtures).