

# Causal Inference

Vanessa Didelez and Robin Evans

BIPS, University of Bremen (Germany), and University of Oxford (UK)

September 2024

APTS — Oxford

# Overview of Course



---

**Part 1:** Basic Causal Concepts

**Part 2:** Causal Diagrams

Directed Acyclic Graphs – DAGs; and  
Single World Intervention Graphs – SWIGs

**Part 3:** Estimating a Causal Effect (basics)

**Part 4:** Estimating a Causal Effect (advanced)

time-dependent / multiple treatments  
double-robust and debiased estimation

**Part 5:** Causal Discovery

**Appendix:** Causal mediation / IV (if time)

# Aims of Course



- Introduce basic concepts of causal learning (reasoning, modelling & inference)
- ... to enable you to read more advanced ‘causal’ papers
- Focus on:
  - formulating causal (research) questions
  - understanding sources of (avoidable and unavoidable) bias
  - some basic methods: g-methods, propensity score, IVs, causal discovery
- Principles / examples & a some maths

**ASK if you have QUESTIONS / comments etc. — ANYTIME!!!**

# Who are You?

---



- Statistics
- Mathematics
- Comp. Science
- Medical / biol / epidemiology
- Econometrics
- Others

# Who are You?

---



- What is a randomised controlled trial?
- Why do we randomise?
- What is a DAG?
- What is confounding?
- What is Berkson / collider bias?
- What is a propensity score?

- 
- Causality / causal inference very broad topic!
  - Has developed and evolved quite separately in different fields: philosophy, sociology, epidemiology, econometrics, computer science, (statistics), mathematics ...
  - Different terminology, approaches, accepted assumptions, designs / types of data sources
  - Last few (only!) years: some convergence has emerged across fields
  - Causality very fundamental to many research questions in many fields of data science!

## **Part 1**

# **Basic Causal Concepts**

- 
- Causation / causality: philosophical, moral and other usages of the term — not what we are concerned with here
  - *This course:* particular (narrow) view of causality most relevant for scientific enquiries: **causality we can implement**
  - “Causal effect” a difference in outcomes, or their distribution, between (hypothetical) experiments we might do,  
i.e. effect of **(hypothetical) interventions**



# Not a Statistical Problem



---

**Example:** We have data on treatment  $A \in \{0, 1\}$  and an outcome  $Y$  (positive is better). A further covariate  $Z \in \{0, 1\}$  has been measured.

- We find that for the treated, the average of  $Y$  is considerably *larger* than for the untreated
- However, within each level of  $Z$ , we find that the average of  $Y$  is considerably *smaller* for the treated than for the untreated

⇒ Do you recommend treatment or not?

---

To obtain a causal answer, **start with a causal question!**

Describe the decision problem you would like to solve,  
or the ideal (hypothetical) experiment with which you could  
investigate your research question

⇒ **Target Trial &**

⇒ **formal ‘language’!**

---

## **Descriptive / predictive:**

“Is this patient at high risk of developing complications during surgery?”

## **Causal:**

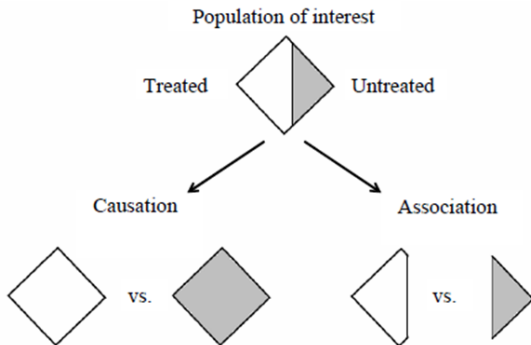
(A) “Which type of anaesthetic should this patient receive to minimise the risk of complications during surgery?”

(A’) “How does the amount of anaesthetic affect the risk of complications during surgery?”

(B) “What can be done to reduce the risk of complications during surgery for an average / a particular type of patient?”

# Causation versus Association

*(Hernan & Robins, 2020:book)*



**(Total) causal effect:** contrast of outcome if ‘everyone was treated’ versus if ‘no-one was treated’

# Target Trial and its Emulation



## Principle:

(Hernan & Robins, 2016:AJE)

- Start by formulating the **ideal** trial (experiment / ...) that would answer your **desired** research question
- Stick to good practice of trial design (PICOT):
  - eligibility criteria / relevant population
  - interventions / treatment strategies to be compared (controls?)
  - outcome (over what follow-up time)
  - other aspects: randomised? blinded? ...?
- Important: **time-zero** alignment of eligibility check, treatment assignment, start of follow-up
  - to avoid immortal-time bias
  - or prevalent-user bias

# Target Trial and its Emulation

---



## Principle ctd:

- Note: must not violate laws of physics (e.g. cannot turn back time); should not deliberately kill patients etc.
  - Then: **emulate target trial** as closely as possible by analysis & with available (obs.) data!
    - use sequence of trials (at all eligible times) for efficiency
    - use ‘cloning’ to avoid immortal-time bias
- ⇒ Systematic approach ensures meaningful research question & minimises design-based sources of bias

# Target Trial and its Emulation

---



- Actual RCTs describe ‘efficacy’:  
does the new drug have an effect at all?
- Analyse real-world (i.e. observational ) data:  
to describe ‘effectiveness’ in real population

---

Here: all models **probabilistic!**

## **Causal model:**

describes situation (distribution) under **(hypothetical) interventions** / manipulations / changes

... needs to be related to:

**observational** (no intervention / 'natural' / 'idle') situation (distribution) generating our data

## **Identifiability (informally):**

aspects of the interventional situation equal certain unique functions of the observational situation



# Basic Concepts

## Conditional (In)dependence



$P(Y = y)$ ,  $p(y)$  etc. probability / density / prob.mass function

### Conditional independence:

$A$  and  $Y$  are conditionally independent given  $Z$ ,  
write  $Y \perp\!\!\!\perp A \mid Z$ , if

$$P(Y = y, A = a \mid Z = z) = P(Y = y \mid Z = z)P(A = a \mid Z = z)$$

for all  $a, y, z$  s.t.  $p(z) > 0$ . Or, equivalently if:

$$P(Y = y \mid A = a, Z = z) = P(Y = y \mid Z = z)$$

or  $p(y|a, z) = p(y|z)$  — relate this to regression models!

# Basic Concepts

## Conditional Independence



In words: if we already know (observed) the value of  $Z$  then knowing the value  $A$  is not informative with respect to the distribution (prediction) of  $Y$

### Example:

- while knowing (only) that someone has tar-stained fingers is informative to predict if they will develop lung-cancer...
- ... once we also know that they are a smoker, the information on their tar-stained fingers becomes irrelevant

lung-cancer  $\perp\!\!\!\perp$  tar-fingers | smoking-status

**Formalisms** to make interventions explicit:

do-notation / causal DAGs / decision theory

Potential outcomes / counterfactuals

Structural equations / structural causal models:

*not much time to cover these...*

# do–Notation

*(Pearl, 2000/9:book)*



**Judea Pearl** introduced intuitive notation to distinguish association and causation: ‘do’ and ‘see’

$$p(y \mid \text{intervene to set } A = a) = p(y \mid \text{do}(A = a))$$

and

$$p(y \mid \text{observe } A = a) = p(y \mid \text{see}(A = a))$$

⇒ **do–calculus** / **axioms** / directed acyclic graphs (DAGs).

Usually  $p(y \mid \text{see}(A = a)) = p(y \mid a)$

$p(y \mid \text{do}(A = a))$  denotes point-intervention in wider system.

Consider:  $Y, A, X_1, X_2$  such that *observationally* ('see'):

$$p(y, a, x_1, x_2) = p(y|a, x_1, x_2)p(a|x_1, x_2)p(x_2|x_1)p(x_1)$$

May have reasons to believe that under intervention on  $A$ :

$$p(y, x_1, x_2 \mid \text{do}(A = \tilde{a})) = p(y|\tilde{a}, x_1, x_2)p(x_2|x_1)p(x_1).$$

**DAGs** help to **structure the factorisation** so as to represent prior-causal knowledge

---

Will see that under **three structural assumptions** we have for suitable set  $X$  of covariates:

$$p(y \mid \text{do}(A = a)) = \sum_x p(y \mid a, x)p(x)$$

left: interventional distribution;

right: observational distrib.

⇒ **non-parametrically identified**, i.e. without parametric assumptions like linearity, Gaussianity etc.

# Potential Outcomes (POs)

(Rubin, 1974; many others)



Consider binary ‘treatment’  $A^i \in \{0, 1\}$ , individual  $i$

$Y^i(0)$  = response under intervention setting  $A^i = 0$

$Y^i(1)$  = response under intervention setting  $A^i = 1$  for **same** subject (at the **same** time)

$\Rightarrow \{Y^i(0), Y^i(1)\}$  can *never be observed together*

$\Rightarrow$  **potential** outcomes.

## **Note:**

POs only well defined if way of manipulating  $A$  well defined!

More generally, for arbitrary treatment type  $A \in \mathcal{A}$

$Y^i(a)$  = response if we *set*  $A^i = a$



---

Once a treatment has been realised, say  $A^i = 1$ ,  
then  $Y^i(1)$  can be **observed**  
and  $Y^i(0)$  becomes **counterfactual** (and vice versa).

Approaches relying on assumptions / properties of the joint distribution of  $(Y(0), Y(1))$  can be called **counterfactual** as these assumptions are never empirically verifiable.

**Missing data?** Causal inference sometimes seen as missing data problem — counterfactual outcomes always missing!

---

Many approaches, in fact, do not rely on *joint* distribution of  $(Y(0), Y(1))$ , and could equivalently be expressed using  $\text{do}(\cdot)$ -notation.

(but POs strong tradition in biomedical / econometric literature.)

Can regard  $p(Y(a)) = p(y \mid \text{do}(A = a))$

But joint distribution of  $(Y(0), Y(1))$  has **no counterpart** in  $\text{do}$ -notation.

⇒ Can express more (also more dubious) concepts with POs.  
(for critique see e.g. Dawid, 2000)

# Structural Equations Models (SEMs)

## aka Structural Causal Models (SCMs)



What makes them *structural*? (Peters, Janzig, Schölkopf, 2018:book)

$$\text{output} \leftarrow f(\text{input})$$

function  $f(\cdot)$  is **invariant** to how the ‘input’ is chosen / generated, e.g. observed or manipulated.

**Caveat:** strong modelling assumption — system considered essentially a ‘machine’ with some random noise.

⇒ allows ‘**cross-world**’ assumptions (like counterfactuals)

⇒ see **single world intervention graphs SWIGs** as alternative  
(Richardson & Robins, 2013:TechRep)

Let's use the above causal languages to express our target of inference.

**Note:** no such thing as *'the'* causal effect

— always need to choose what to contrast with what and how

---

Typically formulated as contrasts of some aspect of

$$p(y \mid \text{do}(A = a)) \quad \text{versus} \quad p(y \mid \text{do}(A = a'))$$

or of  $p(Y(a))$  versus  $p(Y(a'))$ ,

possibly conditional on further variables

For simplicity:  $A$  binary, but with obvious generalisations.

# Average Causal Effect (ACE)

## (Population) Total / Average Treatment Effect (ATE)

---



$$ACE = E(Y \mid \text{do}(A = 1)) - E(Y \mid \text{do}(A = 0))$$

or with POs

$$ACE = E(Y(1)) - E(Y(0))$$

‘Population’: in expectation over whole underlying population  
(effect may change with population)

‘Total’: can be indirect, via multiple causal paths, or  
combination of direct / indirect effects.

Note: can consider ratio, odds-ratio etc. if preferred

Can now define:

$A$  is a **cause** of  $Y$  (and  $Y$  is an effect of  $A$ ) if for some  $a \neq a'$

$$p(y \mid \text{do}(A = a)) \neq p(y \mid \text{do}(A = a'))$$

or  $p(Y(a)) \neq p(Y(a'))$

i.e. if (hypothetically) intervening in  $A$  setting it to different values changes some aspect of the distribution of  $Y$

**Note:** this corresponds to how we check causation in a basic randomised experiment

# Other Causal Effects

## Conditional Causal / Treatment Effect (CATE)



... or **subgroup** causal effect

Let  $Z = z$  characterise subset of population, e.g. age group

Conditional causal effect of  $A$  on  $Y$  given  $Z = z$ :

$$E(Y|Z = z; \text{do}(A = 1)) - E(Y|Z = z; \text{do}(A = 0))$$

or, with POs

$$E(Y(1)|Z = z) - E(Y(0)|Z = z)$$

**Note:**  $Z$  must **not** itself be causally affected by  $A$ , i.e. must be pre-treatment



# Other Causal Effects

## Joint Causal Effect

---



Consider two (possibly sequential) exposures  $A_1, A_2$ .

The joint (total) causal effect of  $A_1$  and  $A_2$  on  $Y$  is

$$E(Y|\text{do}(A_1 = a_1, A_2 = a_2)) - E(Y|\text{do}(A_1 = a'_1, A_2 = a'_2))$$

**Note:** potential issue here: 'time-dependent' confounding  
→ Part 4

## Other Causal Effects

### Controlled Direct Effect (CDE)

---



Consider again two sequential exposures  $A_1, A_2$

Controlled direct effect of  $A_1$  while controlling  $A_2$  means:  
hold **fixed do**( $A_2 = 0$ ) and contrast different values for  $A_1$ , e.g.

$$CDE = E(Y|\text{do}(A_1 = a, A_2 = 0)) - E(Y|\text{do}(A_1 = a', A_2 = 0))$$

**Note:** ‘direct’ means this effect is not possibly mediated by  $A_2$   
(but other mediators allowed)

---

“Individual Causal Effect”: requires counterfactual concepts

“Population intervention effect”

“Effect of treatment on the treated (ETT)”

various versions of “(in) direct causal effects” (natural, interventional, separable...)

Other interventions:

- dynamic / adaptive: e.g. adapt dosage to patient history
- shift / random: add a constant or noise to the ‘treatment’

“Principal Stratum Effect” (or local average treatment effect): requires counterfactual concepts

- 
- For causal answers, start with an explicit causal question: use formal notation ('do' or PO) or describe target trial
  - Different causal parameters correspond to different research questions
  - Key: establish identifiability of causal parameter from observable data
    - so far: 'g-formula' / standardisation to adjust for confounding
  - Structural assumptions: causal consistency, positivity & conditional exchangeability ('no unmod. confounding').

# Thank You!

[www.leibniz-bips.de/en](http://www.leibniz-bips.de/en)

## Contact

Vanessa Didelez

Leibniz Institute for Prevention Research  
and Epidemiology – BIPS

Achterstraße 30  
D-28359 Bremen

[didelez@leibniz-bips.de](mailto:didelez@leibniz-bips.de)

