Causal Inference

Vanessa Didelez and Robin Evans
BIPS, University of Bremen (Germany), and University of Oxford (UK)

August 2023
APTS — Glasgow

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 (Point Treatment)

Part 4: Multiple / Sequential Treatments and Causal Mediation

Part 5: Outlook: Instrumental Variables & Causal Discovery

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?

Causal Inference — History

- Causality / causal inference very broad topic!

Causal Inference — History

- Causality / causal inference very broad topic!
- Has developed and evolved quite separately in different fields: philosophy, sociology, epidemiology, econometrics, computer science, (statistics), mathematics ...

Causal Inference — History

- 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

Causal Inference — History

- 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

Causal Inference — History

- 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

Preamble

- Causation / causality: philosophical, moral and other usages of the term — not what we are concerned with here

Preamble

- 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**

Preamble

- 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.

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

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

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

$\Rightarrow$ Do you recommend treatment or not?

Causal Questions

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

Causal Questions

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’**!

Research Questions

Descriptive / predictive:

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

Research Questions

Descriptive / predictive:

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

Causal:

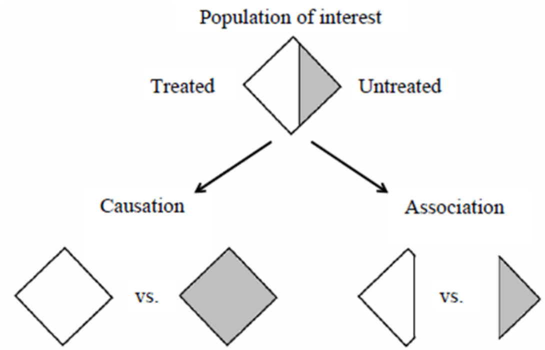
(A) "Which type of anaesthetic should this patient receive to minimise the risk of complications during surgery?"

(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

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? ...?

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.

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

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

Causal Models 1-0-1

Here: all models **probabilistic**!

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

Causal Models 1-0-1

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

Causal Models 1-0-1

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. $p(z) > 0$.

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 $p(y|a, z) = p(y|z)$

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

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 some-one 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

$$\text{lung-cancer} \perp\!\!\!\perp \text{tar-fingers} \mid \text{smoking-status}$$

Basic Causal Concepts

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 intuitve notation to distinguish association and causation: 'do' and 'see'

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

and

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

$\Rightarrow$ **do–calculus** / **axioms** / directed acyclic graphs (DAGs).

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

do–Intervention

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

do–Intervention

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

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

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

do–Intervention

$p(y \,|\, \text{do}(A = a))$ 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:

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

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

Identifiability

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

$$p(y \,|\, \mathsf{do}(A = a)) = \sum_x p(y \,|\, a, x) p(a)$$

left: interventional distribution; right: observational distrib.

Identifiability

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

$$p(y \,|\, \mathsf{do}(A = a)) = \sum_x p(y \,|\, a, x) p(a)$$

left: interventional distribution; right: observational distrib.

$\Rightarrow$ **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!

Potential Outcomes

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

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

Counterfactuals

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).

Counterfactuals

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.

Counterfactuals

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!

Potential Outcomes and ‘do’

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

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

Potential Outcomes and 'do'

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

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

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

Potential Outcomes and 'do'

Many approaches, in fact, do not rely on *joint* distribution of $(Y(0), Y(1))$, and could equivalently be expressed using 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 do–notation.

Potential Outcomes and ‘do’

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

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

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

But joint distribution of $(Y(0), Y(1))$ has no counterpart in do–notation.

$\Rightarrow$ 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.

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.

$\Rightarrow$ allows 'cross-world' assumptions (like counterfactuals)

$\Rightarrow$ see single world intervention graphs SWIGs as alternative (Richardson & Robins, 2013:TechRep)

Non-Parametric SEMs

(NPSEMs-IE) ***(Pearl, 2000/9:book)***

A = treatment / exposure, Y = response, C = covariate

Structural equation model (SEM) — ingredients:

- Directed acyclic graph (DAG) defines 'parents' $=$ inputs;

Non-Parametric SEMs

(NPSEMs-IE) ***(Pearl, 2000/9:book)***

A = treatment / exposure, Y = response, C = covariate

Structural equation model (SEM) — ingredients:

- Directed acyclic graph (DAG) defines 'parents' $=$ inputs;
- equations: $A := f_A(\text{pa}(A), U_A)$
  $$Y := f_Y(\text{pa}(Y), U_Y)$$
  $$C := f_C(\text{pa}(C), U_C)$$

  where f_A, f_Y, f_C describe 'stable' functional relations

Non-Parametric SEMs

(NPSEMs-IE) ***(Pearl, 2000/9:book)***

A = treatment / exposure, Y = response, C = covariate

Structural equation model (SEM) — ingredients:

- Directed acyclic graph (DAG) defines 'parents' $=$ inputs;
- equations: $A := f_A(\text{pa}(A), U_A)$

  $$Y := f_Y(\text{pa}(Y), U_Y)$$

  $$C := f_C(\text{pa}(C), U_C)$$

  where f_A, f_Y, f_C describe 'stable' functional relations
- probability distribution on (U_A, U_Y, U_C)

Non-Parametric SEMs

(NPSEMs-IE) *(Pearl, 2000/9:book)*

A = treatment / exposure, Y = response, C = covariate

Structural equation model (SEM) — ingredients:

- Directed acyclic graph (DAG) defines 'parents' $=$ inputs;
- equations: $$\begin{aligned} A &:= f_A(\text{pa}(A), U_A) \\ Y &:= f_Y(\text{pa}(Y), U_Y) \\ C &:= f_C(\text{pa}(C), U_C) \end{aligned}$$

  where f_A, f_Y, f_C describe 'stable' functional relations
- probability distribution on (U_A, U_Y, U_C)

$\Rightarrow$ induce probability distribution on (A, Y, C).

Often: (U_A, U_Y, U_C) mutually independent $\Rightarrow$ NPSEM-IE

NPSEMs and POs

With NPSEM-IE we have

$$Y(0) = f_Y(\text{pa}(Y)\backslash A, A = 0, U_Y)$$
$$Y(1) = f_Y(\text{pa}(Y)\backslash A, A = 1, U_Y)$$

with the same U_Y

NPSEMs and POs

With NPSEM-IE we have

$$Y(0) = f_Y(\text{pa}(Y)\backslash A, A = 0, U_Y)$$
$$Y(1) = f_Y(\text{pa}(Y)\backslash A, A = 1, U_Y)$$

with the same U_Y

$\Rightarrow$ distribution on (U_A, U_Y, U_C) also induces a probability distribution on $(Y(0), Y(1), A, Y, C)$

... in particular a joint distribution for $(Y(0), Y(1))$!

NPSEMs and POs

With NPSEM-IE we have

$$Y(0) = f_Y(\text{pa}(Y)\backslash A, A = 0, U_Y)$$
$$Y(1) = f_Y(\text{pa}(Y)\backslash A, A = 1, U_Y)$$

with the same U_Y

$\Rightarrow$ distribution on (U_A, U_Y, U_C) also induces a probability distribution on $(Y(0), Y(1), A, Y, C)$

... in particular a joint distribution for $(Y(0), Y(1))$!

Example: linear case $Y := \alpha + \beta x + U_Y$

$\Rightarrow Y^i(0) = \alpha + u_Y^i$ and $Y^i(1) = \alpha + \beta + u_Y^i$
$\Rightarrow$ individual causal effect: $Y^i(1) - Y^i(0) = \beta$

Known as **treatment–unit additivity** assumption.

Causal Frameworks?

1) do$(A = a)$ approach at distributional level: imposes least structure

2) FFRCISTG: uses POs but only allows 'single world'

3) PO's $Y(a)$: imposes more structure as it allows counterfactual variables and cross-worlds

4) NPSEM-IE: imposes most structure as it allows to construct joint distributions of all counterfactuals under 'multiple worlds'

Causal Effects

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

Causal Effects

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

Causal Effects

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

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

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)

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.

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

Cause and Effect

Can now define:

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

$$p(y \,|\, \mathrm{do}(A = a)) \neq p(y \,|\, \mathrm{do}(A = a'))$$

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

Cause and Effect

Can now define:

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

$$p(y \,|\, \mathrm{do}(A = a)) \neq p(y \,|\, \mathrm{do}(A = a'))$$

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

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

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
$\rightarrow$ 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)

Further Causal Effects

“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

Key Assumptions

identifiability of ACE

Causal Consistency Assumption:

if we observe $A = a$ then $Y = Y(a)$

Key Assumptions

identifiability of ACE

Causal Consistency Assumption:

if we observe $A = a$ then $Y = Y(a)$

Positivity Assumption:

$p(a \,|\, x) > 0$ for all $a, x \quad (p(x) > 0)$

where X is sufficient for adjustment as defined next

Key Assumptions

Assumption of **conditional exchangeability:**

(aka: random treatment assignment, or no unmeasured confounding / ignorability, ...)

Key Assumptions

Assumption of **conditional exchangeability:**
(aka: random treatment assignment, or no unmeasured confounding / ignorability, ...)

Set X of observed (measured) pre–treatment covariates exists such that

$$A \perp\!\!\!\perp Y(a) \mid X$$

for all a to be considered as treatment values

Key Assumptions

Assumption of **conditional exchangeability:**
(aka: random treatment assignment, or no unmeasured confounding / ignorability, ...)

Set X of observed (measured) pre–treatment covariates exists such that

$$A \perp\!\!\!\perp Y(a) \mid X$$

for all a to be considered as treatment values

Interpretation:
within values of X, can consider A like randomised wrt Y

Key Assumptions

Assumption of **conditional exchangeability:**
(aka: random treatment assignment, or no unmeasured confounding / ignorability, ...)

Set X of observed (measured) pre–treatment covariates exists such that

$$A \perp\!\!\!\perp Y(a) \mid X$$

for all a to be considered as treatment values

Interpretation:
within values of X, can consider A like randomised wrt Y

Denote: X is *sufficient* to adjust (control) for confounding;
or 'valid adjustment set'

No-Unmeasured-Confounding

with do$(\cdot)$

Assumption of **no unmeasured confounding** & **'consistency'** with do–notation:

$$p(y \mid x; \mathrm{do}(A = a)) = p(y \mid x, a)$$

Interpretation:
within values of X, whether $A = a$ obtained by intervention or observation makes no difference wrt. distribution of Y.

Note: graphical check by **back-door** criterion (Pearl, 1995:Btka)

Randomisation

Under full randomisation: $A \perp\!\!\!\perp$ of all (pre-)baseline variables.

$\Rightarrow$ Exchangeability / no-confounding satisfied for $X = \emptyset$ (or any pre-A set X).

In non-randomised studies:
expert judgement required to determine / justify X as sufficient;
very helpful to use causal DAGs.

Identifiability Revisited

We consider $p(y \,|\, \text{do}(A = a))$ or equivalently $p(Y(a))$.

With the above assumptions:

$$p(Y(a)) \overset{(i)}{=} \sum_x p(Y(a)|x)p(x) \overset{(ii)}{=} \sum_x p(Y(a)|a, x)p(x)$$

$$\ldots \overset{(iii)}{=} \sum_x p(y|a, x)p(x)$$

Identifiability Revisited

We consider $p(y \,|\, \text{do}(A = a))$ or equivalently $p(Y(a))$.

With the above assumptions:

$$p(Y(a)) \overset{(i)}{=} \sum_x p(Y(a)|x)p(x) \overset{(ii)}{=} \sum_x p(Y(a)|a, x)p(x)$$

$$\ldots \overset{(iii)}{=} \sum_x p(y|a, x)p(x)$$

(i) probability calculus

Identifiability Revisited

We consider $p(y \,|\, \text{do}(A = a))$ or equivalently $p(Y(a))$.

With the above assumptions:

$$p(Y(a)) \overset{(i)}{=} \sum_x p(Y(a)|x)p(x) \overset{(ii)}{=} \sum_x p(Y(a)|a,x)p(x)$$

$$\ldots \overset{(iii)}{=} \sum_x p(y|a,x)p(x)$$

(i) probability calculus
(ii) valid adjustment set

Identifiability Revisited

We consider $p(y \,|\, \text{do}(A = a))$ or equivalently $p(Y(a))$.

With the above assumptions:

$$p(Y(a)) \overset{(i)}{=} \sum_x p(Y(a)|x)p(x) \overset{(ii)}{=} \sum_x p(Y(a)|a,x)p(x)$$

$$\dots \overset{(iii)}{=} \sum_x p(y|a,x)p(x)$$

(i) probability calculus
(ii) valid adjustment set
(iii) causal consistency & positivity

Adjustment / Standardisation

Consider the above result

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

- left = causal quantity; right = observational quantity
  $\Rightarrow$ identified if covariates C measured
- right hand side = identifying functional (under the assumptions)
- know as adjustment formula, or standardisation (to the marginal distribution of X)
- also: simplest case of so-called 'g-formula' (Robins, 1986)

Confounding

Above: confounding is present if

$$Y(a) \not\perp\!\!\!\perp A$$

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

Confounding

Above: confounding is present if

$$Y(a) \not\perp\!\!\!\perp A$$

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

Usually:

Confounding = some *(unobserved) common cause* of A and Y

$\Rightarrow$ Use causal DAGs to clarify!

Summary

- 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 assumtpions: causal consistency, positivity & conditional exchangeability.

Thank You!

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