Causality and study design

Gianluca Campanella

Contents

What does ‘*X* caused *Y*’ mean?

What is a cause?

'…we may define a cause to be an object, followed by another, and where all the objects similar to the first are followed by objects similar to the second.
Or in other words where, if the first object had not been, the second never had existed.'

— D Hume (1748)

What is a cause?

'We think of a cause as something that makes a difference, and the difference it makes must be a difference from what would have happened without it. Had it been absent, its effects — some of them, at least, and usually all — would have been absent as well.'

*— D Lewis (1973), J Phil **70**(17)*

Counterfactual theory of causation

Y is present

but

Y wouldn't have been present
if *X* wasn't present

Contribution, not attribution

- We'd like to figure out the effect of *X*, not the cause of *Y*
- *X* is not 'responsible' (i.e. the main or even the only reason) for *Y*
- Causes are not rivals: no point in 'apportioning' outcomes

Contribution, not attribution

- We'd like to figure out the effect of *X*, not the cause of *Y*
- *X* is not 'responsible' (i.e. the main or even the only reason) for *Y*
- Causes are not rivals: no point in 'apportioning' outcomes

✗

What is the cause of *Y*?

✓

How much does *X* affect *Y*?

Correlation is not causation

A correlation is a relation between factual outcomes,
not between factual and counterfactual outcomes

Correlation is not causation

A correlation is a relation between factual outcomes, not between factual and counterfactual outcomes

Example

Taking cough syrup:

- Is positively correlated with coughing
- Has a negative causal effect on coughing (hopefully)

Fundamental problem of causal inference

Causal effects are differences between:

- What happened
- What could have happened (but didn't)

Fundamental problem of causal inference

Causal effects are differences between:

- What happened
- What could have happened (but didn't)

…and so cannot be measured!

Potential outcomes

$Y_i(1)$

is the outcome for unit *i* that
was observed
under some condition

$Y_i(0)$

is the outcome for unit *i* that
would have been observed
under some *other* condition

(all other things being equal)

Potential outcomes

$Y_i(1)$

is the outcome for unit *i* that
was observed
under some condition

$Y_i(0)$

is the outcome for unit *i* that
would have been observed
under some *other* condition

(all other things being equal)

Causal effect for unit *i*

$$Y_i(1) - Y_i(0)$$

Average causal effects

We cannot conclude whether *X* caused *Y*
in any given case, but…

Average causal effects

We cannot conclude whether X caused Y in any given case, but…

We can still figure out if X causes Y on average

$$\mathbb{E}[Y_i(1) - Y_i(0)] = \mathbb{E}[Y_i(1)] - \mathbb{E}[Y_i(0)]$$

Average causal effects are not transitive

- *A* causes *B* and *B* causes *C* on average
- Does it follow that *A* causes *C*?

Average causal effects are not transitive

- *A* causes *B* and *B* causes *C* on average
- Does it follow that *A* causes *C*?

No!

Example

Imagine that…

- *A* causes *B* for men only (so *A* causes *B* on average)
- *B* causes *C* for women only (so *B* causes *C* on average)

…then there is no one for whom *A* causes *C* through *B*!

How can we decide if '*X* caused *Y*'?

There is no causation without manipulation…

…because we need to be able to observe what happens under different conditions all other things being equal

There is no causation without manipulation…

…because we need to be able to observe what happens under different conditions all other things being equal

Try it and find out!

Study design

Study designs

Observational

The researcher observes
but does not alter
what occurs

Experimental

The researcher
intervenes to change reality
then observes what happens

Observational studies

Selection based on…

Exposure X	Outcome Y	Neither
↓	↓	↓
Cohort	**Case-control**	**Cross-sectional**

Cohort studies

- Selection based on exposure X
- Prospective: exposure before outcome
- Usually very lengthy, leading to attrition

Example

- Select two groups:
  1. Smokers
  2. Non-smokers
- After 10 years, check who developed lung cancer

Case-control studies

- Sampling based on outcome Y
- Retrospective: outcome before exposure
- May be biased by imperfect recall

Example

- Select two groups:
  1. Lung cancer patients
  2. Cancer-free ‘controls’
- Ask them whether they’ve ever smoked

Cross-sectional

- All data are collected at the same time
- No distinction between exposure and outcome

Strengths

- Often population-based
- Less expensive

Weaknesses

- No direction of causality
- Over-representation of cases with longer durations

Confounding variables

- Associated with both exposure and outcome
- May explain correlations that have no direct causal connection

Confounding variables

- Associated with both exposure and outcome
- May explain correlations that have no direct causal connection

Example
‘Grey hair causes heart disease’

Experimental studies

Treatment randomly assigned?

Yes

↓

Randomised Controlled Trial

No

↓

Quasi-experiment
(a.k.a. **natural experiment**)

Randomised Controlled Trials (RCTs)

- Control for all main forms of bias
- Ethical concerns

Example

- Divide patients in two groups:
  1. Those who take the drug
  2. Those who take the placebo
- Evaluate influence of drug on disease course

Quasi-experiments

- More practical than RCTs
- Allocation bias

Example (1854 Broad Street cholera outbreak)

- Public water pumps supplied by:
  1. Southwark and Vauxhall Waterworks Company
  2. Lambeth Waterworks Company
- High disease rate in districts supplied by 1
- Water obtained downstream from sewage discharge