An introduction to causal inference

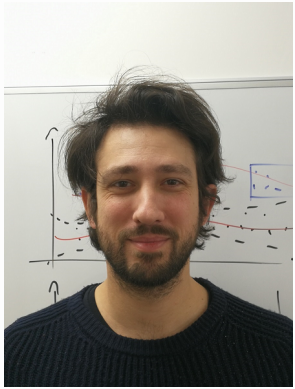
Federico Andreis
federico.andreis@gmail.com

University of Milan-Bicocca
MEBRIC - *May 2021*

Something about me

Dr Federico Andreis,

- Previously Lecturer / Faculty Statistician @ Health Sciences and Sport, University of Stirling
- PhD in Statistics @ University of Milan-Bicocca / Karolinska Institutet, Stockholm
- sampling theory and applications, statistical modelling of electronic health records, non-standard Bootstrap, item response theory



Federico_Andreis · Chicco_Stat · Federico Andreis

Plan for the lecture

The course material for these lectures include:

- recording
- lecture slides
- references for further reading.

Learning outcomes

By the end of these lectures, you should be able to:

- describe the difference between association and causation
- identify the main challenges in estimation of causal effects
- choose and apply basic causal inference methods.

A quick refresher: statistical inference

Statistical inference is the scientific process used to form judgments about the **parameters** of a **population**, typically based on a probabilistic **sample** and **statistics** obtained therefrom.

Population: a finite or infinite collection of statistical units of interest. *Example*: all individuals residing in Scotland.

Parameter: a quantifiable characteristic of the population. *Example*: the average income of all individuals residing in Scotland.

Sample: any subset of a population. *Example*: all individuals residing in Edinburgh.

Statistic: a synthesis of the sample information, typically used to make inference about a parameter. *Example*: the average income of the individuals in the sample.

The exposure and the outcome

Many epidemiological research questions are centered around a particular **exposure** and a particular **outcome**. A few examples:

- Does your diet (exposure) affect your risk for breast cancer (outcome)?
- Is the risk for autism (outcome) bigger in IVF pregnancies (exposure), as compared to 'ordinary' pregnancies?
- Do antibiotics (exposure) cause asthma (outcome)?
- Is the risk for sudden infant death (outcome) bigger if the baby sleeps on the front (exposure) than if the baby sleeps on the back?
- Is coffee drinking related to pancreatic cancer?
- Is coffee drinking during pregnancy related to Pre-Term Delivery?

Association

We often want to learn if there is an **association** between the exposure and the outcome.

- Do the exposure and the outcome tend to 'appear together' in the study population?
- Is asthma more common in children who frequently take on antibiotics, than in other children?
- ...

Statistics and association

Statistics is branch of science that deals with association. Using statistics, we can formally (i.e. using mathematical language) define and quantify association

Common statistical measures of association include:

- correlation coefficients
- regression coefficients
- risk ratios
- odds ratios
- hazard ratios.

Causation

However, the goal is often more ambitious. Ultimately, we often want to learn to what extent the exposure **causes** the outcome

- Does drinking coffee cause pancreatic cancer?
- Do antibiotics cause asthma?

Association is not the same as causation: in observational studies the exposure and outcome may be associated, even in the absence of a causal effect.

Causation

However, the goal is often more ambitious. Ultimately, we often want to learn to what extent the exposure **causes** the outcome

- Does drinking coffee cause pancreatic cancer?
- Do antibiotics cause asthma?

Association is not the same as causation: in observational studies the exposure and outcome may be associated, even in the absence of a causal effect.

Statistics and causation

For most of the 20th century, causation was largely ignored in statistics. In fact, causation cannot even be defined with 'traditional' statistics language.

For instance, the 'associational' risk ratio (RR)

$$RR = \frac{P(Y = 1|A = 1)}{P(Y = 1|A = 0)}$$

cannot in general be given a causal interpretation.

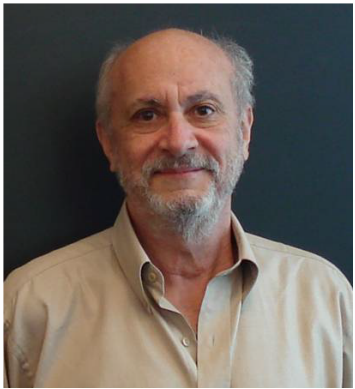
But... what does a 'causal' risk ratio look like?

Brief history of causal inference: the 70s

In the mid-70s, eminent American statistician Donald Rubin developed a formal definition of causation

- potential outcomes
- counterfactuals.

You can find a recent interview with Rubin here.



Brief history of causal inference: the 80s

American epidemiologist and biostatistician James M. Robins discovered - and solved - some important problems with longitudinal studies, from a causal inference perspective

- marginal structural models
- structural nested models.

Here's a recent interview.

Brief history of causal inference: the 90s

Israeli-American computer scientist and philosopher Judea Pearl developed Directed Acyclic Graphs (DAGs)

- simplify interpretation and communication in causal inference
- useful for covariate selection in observational studies.

Here you can find a recent interview.

Outline of topics

During these lectures, we will introduce and discuss the main ideas relating to:

- association vs causation
- estimation of causal effects
- Directed Acyclic Graphs
- causality and regression models

Intense research field over the last 20 years. We will be just scratching the surface!

Association - preliminaries

Suppose we are interested in the relation between an exposure A and an outcome Y.

- We assume for simplicity that both A and Y are binary
  - we use '0' for 'unexposed/no outcome', and '1' for 'exposed/outcome'
- we assume that population data are available (infinite sample size)
  - no need for p-values, confidence intervals etc
- these conditions are often unrealistic, but are useful for pedagogical purposes
  - we will relax them later on.

Association - preliminaries

Suppose we are interested in the relation between an exposure A and an outcome Y.

- We assume for simplicity that both A and Y are binary
  - we use '0' for 'unexposed/no outcome', and '1' for 'exposed/outcome'
- we assume that population data are available (infinite sample size)
  - no need for p-values, confidence intervals etc
- these conditions are often unrealistic, but are useful for pedagogical purposes
  - we will relax them later on.

Association - preliminaries

Suppose we are interested in the relation between an exposure A and an outcome Y.

- We assume for simplicity that both A and Y are binary
  - we use '0' for 'unexposed/no outcome', and '1' for 'exposed/outcome'
- we assume that population data are available (infinite sample size)
  - no need for p-values, confidence intervals etc
- these conditions are often unrealistic, but are useful for pedagogical purposes
  - we will relax them later on.

Association - preliminaries

Suppose we are interested in the relation between an exposure A and an outcome Y.

- We assume for simplicity that both A and Y are binary
  - we use '0' for 'unexposed/no outcome', and '1' for 'exposed/outcome'
- we assume that population data are available (infinite sample size)
  - no need for p-values, confidence intervals etc
- these conditions are often unrealistic, but are useful for pedagogical purposes
  - we will relax them later on.

Joint probability

Suppose that the population proportions of A and Y are given by

$A \setminus Y$	$Y = 0$	$Y = 1$	
$A = 0$	0.88	0.02	
$A = 1$	0.09	0.01	

- Among all subjects, 1% are both exposed and have the outcome
- we say that the **joint probability** of $(A = 1, Y = 1)$ is 0.01
- we denote this as $P(A = 1, Y = 1) = 0.01$.

Marginal probability

$A \setminus Y$	$Y = 0$	$Y = 1$	
$A = 0$	0.88	0.02	0.90
$A = 1$	0.09	0.01	0.10
	0.97	0.03	1

- Among all subjects, 3% have the outcome
- we say that the **marginal probability** of $Y = 1$ is 0.03, denoted as $P(Y = 1) = 0.03$
- we say that the **marginal probability** of $A = 1$ is 0.10, denoted as $P(A = 1) = 0.10$.

Conditional probability

$A \setminus Y$	$Y = 0$	$Y = 1$	
$A = 0$	0.88	0.02	
$\mathbf{A = 1}$	**0.09**	**0.01**	

- Among the exposed subjects, $0.01/(0.01 + 0.09) = 0.1$ have the outcome
- we say that the **conditional probability** of having the outcome, for exposed subjects, is 0.1
- we denote this as $P(Y = 1|A = 1) = 0.1$.

What about the unexposed?

Conditional probability

$A \setminus Y$	$Y = 0$	$Y = 1$	
$\mathbf{A = 0}$	**0.88**	**0.02**	
$A = 1$	0.09	0.01	

- Among the unexposed subjects, $0.02/(0.02 + 0.88) \approx 0.022$ have the outcome
- we say that the **conditional probability** of having the outcome, for unexposed subjects, is 0.022
- we denote this as $P(Y = 1|A = 0) = 0.022$.

Independence and association

We say that A and Y are **independent** if the risk of the outcome is the same for exposed and unexposed:

$$P(Y = 1|A = 1) = P(Y = 1|A = 0) = P(Y = 1)$$

a shorthand for which is $Y \perp\!\!\!\perp A$.

- We say that A and Y are **associated** if the risk of the outcome is different for exposed and unexposed:

$$P(Y = 1|A = 1) \neq P(Y = 1|A = 0) \neq P(Y = 1)$$

a shorthand for which is $Y \not\perp\!\!\!\perp A$.

Independence and association

We say that A and Y are **independent** if the risk of the outcome is the same for exposed and unexposed:

$$P(Y = 1|A = 1) = P(Y = 1|A = 0) = P(Y = 1)$$

a shorthand for which is $Y \perp\!\!\!\perp A$.

- We say that A and Y are **associated** if the risk of the outcome is different for exposed and unexposed:

$$P(Y = 1|A = 1) \neq P(Y = 1|A = 0) \neq P(Y = 1)$$

a shorthand for which is $Y \not\perp\!\!\!\perp A$.

A numerical example

Are A and Y independent or associated, given the table below?

$A \setminus Y$	$Y = 0$	$Y = 1$	
$A = 0$	0.88	0.02	
$A = 1$	0.09	0.01	

A numerical example - solution

$A \setminus Y$	$Y = 0$	$Y = 1$	
$A = 0$	0.88	0.02	
$A = 1$	0.09	0.01	

$$P(Y = 1|A = 1) = \frac{0.01}{0.01 + 0.09} = 0.1$$

$$P(Y = 1|A = 0) = \frac{0.02}{0.02 + 0.88} \approx 0.022$$

$$P(Y = 1) = 0.02 + 0.01 = 0.03$$

$P(Y = 1|A = 1) \neq P(Y = 1|A = 0) \neq P(Y = 1)$, so A and Y are associated.

An important remark

There may be several explanations to an observed association between A and Y. For example:

- A causes Y
- Y causes A ('reverse causation')
- A and Y have common causes ('confounding')

That A and Y appear associated only means that certain values of A and Y tend to 'appear together'. Why this happens is a different question.

Common association measures for binary variables

- The **risk difference (RD)**

$$RD = P(Y = 1|A = 1) - P(Y = 1|A = 0)$$

$Y \perp\!\!\!\perp A \iff RD = 0$

- the **risk ratio (RR)**

$$RR = \frac{P(Y = 1|A = 1)}{P(Y = 1|A = 0)}$$

$Y \perp\!\!\!\perp A \iff RR = 1$

- the **odds ratio (OR)**

$$OR = \frac{P(Y = 1|A = 1)}{P(Y = 0|A = 1)} \Big/ \frac{P(Y = 1|A = 0)}{P(Y = 0|A = 0)}$$

$Y \perp\!\!\!\perp A \iff OR = 1.$

A numerical example

Compute RD , RR, and OR for the table.

$A \setminus Y$	$Y = 0$	$Y = 1$	
$A = 0$	0.88	0.02	
$A = 1$	0.09	0.01	

A numerical example - solution

Since we know, from our previous example, that

$$P(Y = 1|A = 1) = 0.1, \quad P(Y = 1|A = 0) = 0.022$$

it follows

$$RD = 0.1 - 0.022 = 0.078$$

$$RR = \frac{0.1}{0.022} \approx 4.55$$

$$OR = \frac{0.1}{1 - 0.1} / \frac{0.022}{1 - 0.022} \approx 4.94.$$

Conditional association/independence

Sometimes we wish to stratify. Let L denote "male" (0) or "female" (1). Then

- $P(Y = 1|A = a; L = 1)$ is the conditional probability of having the outcome, for women with exposure level a
- $P(Y = 1|A = a; L = 0)$ is the conditional probability of having the outcome, for men with exposure level a

We will say that

- A and Y are **conditionally independent**, given L, if
$$P(Y = 1|A = 1; L) = P(Y = 1|A = 0; L) = P(Y = 1|L)$$
and write $Y \perp\!\!\!\perp A|L$
- A and Y are **conditionally associated**, given L, if
$$P(Y = 1|A = 1; L) \neq P(Y = 1|A = 0; L) \neq P(Y = 1|L)$$
and write $Y \not\perp\!\!\!\perp A|L$.

Conditional association/independence

Sometimes we wish to stratify. Let L denote "male" (0) or "female" (1). Then

- $P(Y = 1|A = a; L = 1)$ is the conditional probability of having the outcome, for women with exposure level a
- $P(Y = 1|A = a; L = 0)$ is the conditional probability of having the outcome, for men with exposure level a

We will say that

- A and Y are **conditionally independent**, given L, if
$$P(Y = 1|A = 1; L) = P(Y = 1|A = 0; L) = P(Y = 1|L)$$
and write $Y \perp\!\!\!\perp A|L$
- A and Y are **conditionally associated**, given L, if
$$P(Y = 1|A = 1; L) \neq P(Y = 1|A = 0; L) \neq P(Y = 1|L)$$
and write $Y \not\perp\!\!\!\perp A|L$.

Technical note

In principle, we could have that $P(Y = 1|A = 1; L) = P(Y = 1|A = 0; L)$ for some values of L, and $P(Y = 1|A = 1; L) \neq P(Y = 1|A = 0; L)$ for some others.

When we write $Y \perp\!\!\!\perp A|L$, we mean that

$$P(Y = 1|A = 1; L) = P(Y = 1|A = 0; L)$$

for **all** values of L.

When we write $Y \not\perp\!\!\!\perp A|L$, we mean that

$$P(Y = 1|A = 1; L) \neq P(Y = 1|A = 0; L)$$

for **at least one** value of L.

Technical note

In principle, we could have that $P(Y = 1|A = 1; L) = P(Y = 1|A = 0; L)$ for some values of L, and $P(Y = 1|A = 1; L) \neq P(Y = 1|A = 0; L)$ for some others.

When we write $Y \perp\!\!\!\perp A|L$, we mean that

$$P(Y = 1|A = 1; L) = P(Y = 1|A = 0; L)$$

for **all** values of L.

When we write $Y \not\perp\!\!\!\perp A|L$, we mean that

$$P(Y = 1|A = 1; L) \neq P(Y = 1|A = 0; L)$$

for **at least one** value of L.

Technical note

In principle, we could have that $P(Y = 1|A = 1; L) = P(Y = 1|A = 0; L)$ for some values of L, and $P(Y = 1|A = 1; L) \neq P(Y = 1|A = 0; L)$ for some others.

When we write $Y \perp\!\!\!\perp A|L$, we mean that

$$P(Y = 1|A = 1; L) = P(Y = 1|A = 0; L)$$

for **all** values of L.

When we write $Y \not\perp\!\!\!\perp A|L$, we mean that

$$P(Y = 1|A = 1; L) \neq P(Y = 1|A = 0; L)$$

for **at least one** value of L.

Measures of conditional association

- The **conditional risk difference**

$$RD|L = P(Y = 1|A = 1; L) - P(Y = 1|A = 0; L)$$

- the **conditional risk ratio**

$$RR|L = \frac{P(Y = 1|A = 1; L)}{P(Y = 1|A = 0; L)}$$

- the **conditional odds ratio**

$$OR|L = \frac{P(Y = 1|A = 1; L)}{P(Y = 0|A = 1; L)} \Big/ \frac{P(Y = 1|A = 0; L)}{P(Y = 0|A = 0; L)}.$$

Models to deal with causality

Many **causal models** have been proposed, that can be used to describe causal relationships. Notable examples include:

- the sufficient-component cause model (Rothman)
- potential outcomes, counterfactuals (Rubin, Robins)
- structural equations, causal diagrams (Pearl).

All common causal models are essentially (mathematically) equivalent: different languages, same content.

To define 'causation', we will mostly rely on potential outcomes, but borrow from the other models as well. For a specialised discussion on model differences, see here.

Models to deal with causality

Many **causal models** have been proposed, that can be used to describe causal relationships. Notable examples include:

- the sufficient-component cause model (Rothman)
- potential outcomes, counterfactuals (Rubin, Robins)
- structural equations, causal diagrams (Pearl).

All common causal models are essentially (mathematically) equivalent: different languages, same content.

To define 'causation', we will mostly rely on potential outcomes, but borrow from the other models as well. For a specialised discussion on model differences, see here.

Motivating example

Consider the following situation: August has been smoking 5 cigarettes a day since he was 15 years old. At the age of 60 he develops lung cancer.

Did the smoking cause the cancer?

Human reasoning about cause and effects

To assess causality, we typically mentally compare two scenarios:

- the outcome when the exposure is present
- the outcome when the exposure is absent.

while keeping everything else equal.

If the two outcomes differ, we say that the exposure has a causal (or preventative) effect.

Potential outcomes

Let Y_a be shorthand for the outcome that we would observe, for a given subject, if the subject **potentially** received exposure level $A = a$:

- Y_1 is the outcome under exposure
- Y_0 is the outcome under non-exposure.

Y_1 and Y_0 are referred to as **potential outcomes**.

The ideal dataset

Ideally we would observe both potential outcomes for every subject:

subject	Y_1	Y_0
August	1	0
Selma	0	0
Fjodor	1	1

In practice, this is **not realistic**. However, it is useful to introduce the framework.

Subject-specific causal effects

subject	Y_1	Y_0
August	1	0
Selma	0	0
Fjodor	1	1

We say that A has a **causal effect** on Y, **for a given subject**, if the potential outcomes Y_1 and Y_0 differ for that subject.

- For August, the exposure has an effect: $Y_1 \neq Y_0$
- For Selma and Fjodor, the exposure has no effect; $Y_1 = Y_0$.

A more realistic situation

subject	A	Y	Y_1	Y_0
August	1	1	1	?
Selma	0	0	?	0
Fjodor	0	1	?	1

August is exposed ($A = 1$). Thus, for August

- Y_1 is observed and equal to the factual outcome Y
- Y_0 is unobserved, or **counterfactual**.

Selma and Fjodor are unexposed ($A = 0$). Thus, for Selma and Fjodor

- Y_0 is observed and equal to the factual outcome Y
- Y_1 is unobserved, or counterfactual.

A more realistic situation

subject	A	Y	Y_1	Y_0
August	1	1	1	?
Selma	0	0	?	0
Fjodor	0	1	?	1

August is exposed ($A = 1$). Thus, for August

- Y_1 is observed and equal to the factual outcome Y
- Y_0 is unobserved, or **counterfactual**.

Selma and Fjodor are unexposed ($A = 0$). Thus, for Selma and Fjodor

- Y_0 is observed and equal to the factual outcome Y
- Y_1 is unobserved, or counterfactual.

A more realistic situation

subject	A	Y	Y_1	Y_0
August	1	1	1	?
Selma	0	0	?	0
Fjodor	0	1	?	1

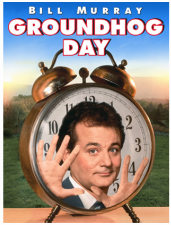
August is exposed ($A = 1$). Thus, for August

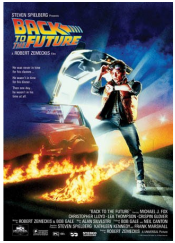
- Y_1 is observed and equal to the factual outcome Y
- Y_0 is unobserved, or **counterfactual**.

Selma and Fjodor are unexposed ($A = 0$). Thus, for Selma and Fjodor

- Y_0 is observed and equal to the factual outcome Y
- Y_1 is unobserved, or counterfactual.

Counterfactuals in popular culture





A fundamental problem of causation


DAD'S LANDMARK £1M LEGAL BATTLE WITH PHONE FIRM

I'm suing Nokia to prove mobile was the cause of brain tumour

More radiation as old phones found a signal

World is glued to case in UK

It is very difficult to say whether the exposure causes the outcome for a specific subject - because we cannot observe the same subject under two exposure levels simultaneously.

Population causal effects

Fortunately, it is much easier to make causal claims on population levels, such as

"*if everybody would quit smoking, the incidence of lung cancer would decrease by* 15%"

Denote with $P(Y_a = 1)$ is the proportion of subjects that would develop the outcome, if **everybody** would receive exposure level a, then

- A has a population causal effect on Y if

$$P(Y_1 = 1) \neq P(Y_0 = 1)$$

- A has no population causal effect on Y if

$$P(Y_1 = 1) = P(Y_0 = 1).$$

Population causal effects

Fortunately, it is much easier to make causal claims on population levels, such as

"*if everybody would quit smoking, the incidence of lung cancer would decrease by* 15%"

Denote with $P(Y_a = 1)$ is the proportion of subjects that would develop the outcome, if **everybody** would receive exposure level a, then

- A has a population causal effect on Y if

$$P(Y_1 = 1) \neq P(Y_0 = 1)$$

- A has no population causal effect on Y if

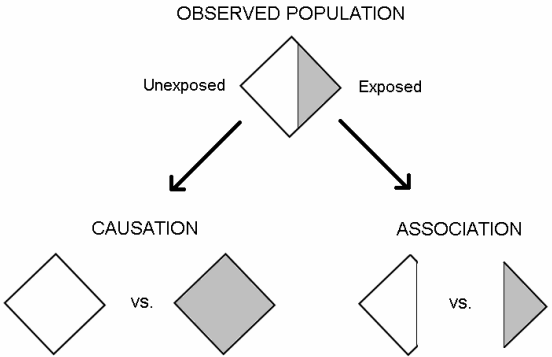
$$P(Y_1 = 1) = P(Y_0 = 1).$$

Notation

In statistics, we use

- upper case letters (e.g. A, Y) for random variables
- lower case letters (e.g. a, y) for fixed numbers
- when writing Y_a, we consider the exposure to be fixed to a (0 or 1)
- when writing $P(Y_a = 1)$, we consider a scenario where the exposure A is fixed to a for everybody.

Association vs Causation



$P[Y_1 = 1]$ vs $P[Y_0 = 1]$ $\quad$ $P[Y = 1|A = 1]$ vs $P[Y = 1|A = 0]$

Measures of causal effects for binary variables

- The **causal risk difference (CRD)**

$$CRD = P(Y_1 = 1) - P(Y_0 = 1)$$

  causal null hypothesis $\iff$ $CRD = 0$

- the **causal risk ratio (CRR)**

$$CRR = \frac{P(Y_1 = 1)}{P(Y_0 = 1)}$$

  causal null hypothesis $\iff$ $CRR = 1$

- the **causal odds ratio (COR)**

$$COR = \frac{P(Y_1 = 1)}{P(Y_1 = 0)} \Big/ \frac{P(Y_0 = 1)}{P(Y_0 = 0)}$$

  Causal null hypothesis $\iff$ $COR = 1$.

A numerical example

subject	Y_1	Y_0
1	0	0
2	1	0
3	0	0
4	1	1
5	0	0
6	1	1
7	1	1
8	1	1
9	0	0
10	1	0

Compute CRD, CRR, and COR.

Solution

subject	Y_1	Y_0
1	0	0
2	1	0
3	0	0
4	1	1
5	0	0
6	1	1
7	1	1
8	1	1
9	0	0
10	1	0

$$P(Y_1 = 1) = 6/10 = 0.6$$

$$P(Y_0 = 1) = 4/10 = 0.4$$

$$CRD = 0.6 - 0.4 = 0.2$$

$$CRR = 0.6/0.4 = 1.5$$

$$COR = \frac{0.6}{1-0.6} / \frac{0.4}{1-0.4} = 2.25.$$

Conditional causal effects

- Conditional causal risk difference

$$CRD|L = P(Y_1 = 1|L) - P(Y_0 = 1|L)$$

- Conditional causal risk ratio

$$CRR|L = \frac{P(Y_1 = 1|L)}{P(Y_0 = 1|L)}$$

- Conditional causal odds ratio

$$COR|L = \frac{P(Y_1 = 1|L)}{P(Y_1 = 0|L)} \Big/ \frac{P(Y_0 = 1|L)}{P(Y_0 = 0|L)}.$$

Some remarks

- Both association and causation can be quantified with risk differences, risk ratios, and odds ratios
- for convenience, we will mostly focus on risk ratios
- everything that we say holds for risk differences and odds ratios as well.

Association vs causation

The difference between association and causation is critical. Consider the following example:

- the CRR of 5-year mortality is 0.5 for aspirin vs. no aspirin
- the corresponding RR is 1.5 because individuals at high risk of cardiovascular death are preferentially prescribed aspirin.

Physician decides to withhold aspirin because those treated have greater risk of dying. The doctor will be sued for malpractice.

When is a counterfactual “well defined”?

We say that a counterfactual is “well defined” when we have a clear understanding of what the counterfactual represents ‘in real life’.

- Are all counterfactuals well defined?
- If some counterfactuals are not well defined, then causal effects based on these are not well defined either!

An example of poorly defined counterfactual

Let $A = 1$ if BMI > 30, and $A = 0$ if BMI < 30.

Certain diseases occur more frequently in obese than in non-obese, i.e.

$$P(Y = 1|A = 1) > P(Y = 1|A = 0)$$

Does 'obesity' have a causal effect on the risk for disease?

$$P(Y_1 = 1) \neq P(Y_0 = 1)$$

Quite a vague question

Translated into plain English, the counterfactual comparison reads:

"*what risk if everybody had BMI>30 compared to if everybody had BMI* < 30*?*"

But what does "if everybody had BMI > 30" really mean?

- everybody short or heavy?
- everybody fat or muscular?
- everybody belly fat or hips fat?

Outcome likely very different under these alternative scenarios. Unless we specify more precisely what scenario we refer to, the counterfactual outcome is not well defined.

Quite a vague question

Translated into plain English, the counterfactual comparison reads:

"*what risk if everybody had BMI>30 compared to if everybody had BMI* < 30*?*"

But what does "if everybody had BMI > 30" really mean?

- everybody short or heavy?
- everybody fat or muscular?
- everybody belly fat or hips fat?

Outcome likely very different under these alternative scenarios. Unless we specify more precisely what scenario we refer to, the counterfactual outcome is not well defined.

An important difference between association and causation

In order for the causal effect of A on Y to be well defined we require that:

- we can tell whether an observed subject has $A = 1$ or $A = 0$
- we agree on what it means that an observed subject with $A = 0$ **would have had** $A = 1$, and vice versa.

As concept of association is based on factual observations, rather than counterfactuals, first condition is enough for association between A and Y to be well defined.

A quote

"*Some counterfactuals are ill-defined, most are somewhat vague, but many are useful*". Lewis, 1973

Remarks

- Association is not equal to causation
- to define causation, we use **potential outcomes** and **counterfactuals**
- not all counterfactuals (and causal effects) are well defined.

Exercise - general use of counterfactual

Consider the problem of estimating the causal effect of coffee drinking on pancreatic cancer (PC).

- Expose subject to coffee drinking, follow him up, see if PC occurs
- On the same subject, turn back the clock and expose him to NOT coffee, then follow him up and see if PC occurs.

Possible PC responses: $Y = 1$, $Y = 0$

group	coffee ($a = 1$)	no coffee ($a = 0$)	number
1	$Y = 1$	$Y = 1$	Np_1
2	$Y = 1$	$Y = 0$	Np_2
3	$Y = 0$	$Y = 1$	Np_3
4	$Y = 0$	$Y = 0$	Np_4

Remarks

A few things to consider:

- repeating observation on same individual implies *identical conditions*
- coffee exposure must occur prior to incidence of pancreatic cancer
- factors determined before we set coffee exposure are held fixed.

Remarks

- Group 1 subjects are called **doomed**, as they will develop PC, no matter what
- group 2 subjects are called **responders** or **causative**, as they will develop PC if they drink coffee, but not if they do not drink
- group 3 subjects are called **preventative**
- group 4 subjects are called **immune**, as they will not develop PC, no matter what.

Remarks

The following relationships hold:

- $p_1 + p_2 + p_3 + p_4 = 1$
- $P(Y_1 = 1) = p_1 + p_2$
- $P(Y_0 = 1) = p_1 + p_3$
- $CRR = \frac{p_1+p_2}{p_1+p_3}$.

group	Y_1	Y_0	number
1	1	1	Np_1
2	1	0	Np_2
3	0	1	Np_3
4	0	0	Np_4

Remarks

Consider the null value of $CRR = 1$. This would mean, that the population risk of PC when everyone drinks coffee is the same as when everyone does not:

- $CRR = 1$ if and only if $p_2 = p_3$.

It does not mean that coffee has no causal effect on any individual:

- there may be individuals in group 2 or 3 where coffee drinking makes the difference as to whether they contract pancreatic cancer or not
- the number of group 2 individuals exactly matches the number of group 3 individuals so drinking coffee does not change the population risks of PC.

Ideal data - encore

Let Y_a be the outcome that we would observe, for a given subject, if the subject potentially received exposure level a.

- Y_1 is the outcome under exposure
- Y_0 is the outcome under non-exposure

Ideally (**unrealistically**) we would observe both potential outcomes for any given subject

subject	Y_1	Y_0
August	1	0
Selma	0	0
Fjodor	1	1

Subject-specific causal effects

A has a causal effect on Y, for a given subject, if the potential outcomes Y_1 and Y_0 differ for this subject

subject	Y_1	Y_0
August	1	0
Selma	0	0
Fjodor	1	1

- for August, the exposure has an effect, $Y_1 \neq Y_0$
- for Selma and Fjodor, the exposure has not effect, $Y_1 = Y_0$.

Observed data

August is exposed ($A = 1$). Thus, for August

- Y_1 is observed and equal to the factual outcome Y
- Y_0 is unobserved, or counterfactual.

Selma and Fjodor are unexposed ($A = 0$). Thus, for Selma and Fjodor

- Y_0 is observed and equal to the factual outcome Y
- Y_1 is unobserved, or counterfactual.

subject	A	Y	Y_1	Y_0
August	1	1	1	?
Selma	0	0	?	0
Fjodor	0	1	?	1

A fundamental problem of causation

It is very difficult to say whether the exposure causes the outcome for a specific subject, because we cannot observe the same subject under two exposure levels simultaneously.

Fortunately, easier to make causal claims on population levels, such as "*if everybody would quit smoking, then the incidence of liver cancer would decrease by* 15%".

Population causal effects

$P(Y_a = 1)$ is the proportion of subjects that would develop the outcome, if **everybody** would receive exposure level a

- A has a population causal effect on Y if

$$P(Y_1 = 1) \neq P(Y_0 = 1)$$

- A has no population causal effect on Y if

$$P(Y_1 = 1) = P(Y_0 = 1)$$

Population causal effects

- A population causal effect is assessed by comparing $P(Y_1 = 1)$ vs $P(Y_0 = 1)$.
- Direct computation requires comparing the whole population under exposure, with the whole population under no exposure.
- However, just like for any given subject, we cannot in general observe the whole population under two exposure levels.

How can we estimate population causal effects?

A numerical example - encore

Compute the CRR.

subject	Y_1	Y_0
1	0	0
2	1	0
3	0	0
4	1	1
5	0	0
6	1	1
7	1	1
8	1	1
9	0	0
10	1	0

A numerical example - encore

Compute the CRR.

subject	Y_1	Y_0
1	0	0
2	1	0
3	0	0
4	1	1
5	0	0
6	1	1
7	1	1
8	1	1
9	0	0
10	1	0

$$P(Y_1 = 1) = 6/10 = 0.6$$

$$P(Y_0 = 1) = 4/10 = 0.4$$

$$CRR = \frac{0.6}{0.4} = 1.5$$

Randomised trials

In a **randomised trial**, we have that

$$\underbrace{P(Y_1 = 1|A = 1)}_{=P(Y=1|A=1)} = P(Y_1 = 1)$$
$$\underbrace{P(Y_0 = 1|A = 0)}_{=P(Y=1|A=0)} = P(Y_0 = 1)$$

an immediate implication of which is that $RR = CRR$.

In randomised trials, **association is causation**! This is due to the randomisation of exposure. More on this later.

Exercise - randomised trial

Consider data obtained from a **randomised trial**. Compute the RR.

subject	A	Y	Y_1	Y_0
1	1	0	0	?
2	1	1	1	?
3	0	0	?	0
4	1	1	1	?
5	0	0	?	0
6	1	1	1	?
7	0	1	?	1
8	0	1	?	1
9	1	0	0	?
10	0	0	?	0

Exercise - randomised trial

Consider data obtained from a **randomised trial**. Compute the RR.

subject	A	Y	Y_1	Y_0
1	1	0	0	?
2	1	1	1	?
3	0	0	?	0
4	1	1	1	?
5	0	0	?	0
6	1	1	1	?
7	0	1	?	1
8	0	1	?	1
9	1	0	0	?
10	0	0	?	0

$$P(Y = 1|A = 1) = 3/5 = 0.6 = P(Y_1 = 1)$$

$$P(Y = 1|A = 0) = 2/5 = 0.4 = P(Y_0 = 1)$$

$$RR = \frac{0.6}{0.4} = 1.5 = CRR.$$

Exchangeability

In a **randomised trial**, we have that

$$\underbrace{P(Y_1 = 1|A = 1)}_{=P(Y=1|A=1)} = P(Y_1 = 1)$$
$$\underbrace{P(Y_0 = 1|A = 0)}_{=P(Y=1|A=0)} = P(Y_0 = 1).$$

Thanks to randomisation, Y_0 and Y_1 are independent of A:

$$(Y_0, Y_1) \perp\!\!\!\perp A.$$

We say that the exposed and unexposed are **exchangeable**. Under exchangeability, association = causation.

A numerical example

Compute the CRR, assuming exchangeability

subject	A	Y
1	1	0
2	1	1
3	0	0
4	1	1
5	0	0
6	1	1
7	0	1
8	0	1
9	1	0
10	0	0

A numerical example

Compute the CRR, assuming exchangeability

subject	A	Y
1	1	0
2	1	1
3	0	0
4	1	1
5	0	0
6	1	1
7	0	1
8	0	1
9	1	0
10	0	0

$$
\begin{aligned}
CRR &= \frac{P(Y_1 = 1)}{P(Y_0 = 1)} = \{(Y_0, Y_1) \perp\!\!\!\perp A\} \\
&= \frac{P(Y_1 = 1 | A = 1)}{P(Y_0 = 1 | A = 0)} \\
&= \frac{P(Y = 1 | A = 1)}{P(Y = 1 | A = 0)} \\
&= \frac{3/5}{2/5} = 1.5
\end{aligned}
$$

Why does randomisation work?

Under randomisation, all pre-exposure variables:

- are equally distributed across levels of A
- are independent of A.

The potential outcomes (Y_0, Y_1) are pre-exposure variables! They describe how the subject 'would react' to $A = 0$ and $A = 1$.

This reaction depends on numerous factors which are determined before the factual exposure level is received: genes, lifestyle, age, etc.

Thus, under randomization (Y_0, Y_1) are independent of A: $(Y_0, Y_1) \perp\!\!\!\perp A$.

This is amazing! Why then not always randomize?

Why does randomisation work?

Under randomisation, all pre-exposure variables:

- are equally distributed across levels of A
- are independent of A.

The potential outcomes (Y_0, Y_1) are pre-exposure variables! They describe how the subject 'would react' to $A = 0$ and $A = 1$.

This reaction depends on numerous factors which are determined before the factual exposure level is received: genes, lifestyle, age, etc.

Thus, under randomization (Y_0, Y_1) are independent of A: $(Y_0, Y_1) \perp\!\!\!\perp A$.

This is amazing! Why then not always randomize?

Example

Does heart transplant (A) increase 5-year survival (Y)?

1. Select a large population of potential recipients of a transplant
2. Get funding and ethical approval
3. Randomly allocate to either transplant ($A = 1$) or medical treatment ($A = 0$)
4. 5 years later, calculate the causal risk ratio.

Easy! Or... is it?

Example

Does heart transplant (A) increase 5-year survival (Y)?

1. Select a large population of potential recipients of a transplant
2. Get funding and ethical approval
3. Randomly allocate to either transplant ($A = 1$) or medical treatment ($A = 0$)
4. 5 years later, calculate the causal risk ratio.

Easy! Or... is it?

Non-ignorable drop out

Let $D = 1$ if a subject drops out of study before end of follow up, $D = 0$ otherwise.

- We can calculate $P(Y = 1|A, D = 0)$, but not $P(Y = 1|A)$
- **potential problem!** among those who remain in the study, exposed and unexposed may not be exchangeable:

$$(Y_0, Y_1) \not\perp\!\!\!\perp A|D = 0.$$

Unblinding

When the study subjects are aware of what treatment they receive, they may change their behavior accordingly:

- transplant receivers may change their diet to keep their new heart healthy
- the causal effect of A on Y combines effect of exposure and behaviour change
- even if treated and untreated behave similarly, pure knowledge of treatment received may affect the outcome.

Non-compliance

Some subjects who are assigned to the new treatment may take the old treatment, and vice versa.

Traditional analyses:

- Intention To Treat (ITT)
- As Treated (AT).

Both analyses are likely to be biased! Alternative 'causal inference methods' exist (beyond the scope of this course).

Conclusion - RCTs

Real randomized trials often suffer from several important problems

Observational studies are needed. In fact, most human knowledge comes from observations, e.g. evolution theory, smoking causes lung cancer etc.

Methods for causal inference from observational studies **very active field of research**!

A numerical example - encore

Compute the CRR.

subject	Y_1	Y_0
1	0	0
2	1	0
3	0	0
4	1	1
5	0	0
6	1	1
7	1	1
8	1	1
9	0	0
10	1	0

$$P(Y_1 = 1) = 6/10 = 0.6$$

$$P(Y_0 = 1) = 4/10 = 0.4$$

$$CRR = \frac{0.6}{0.4} = 1.5$$

Exercise - observational study

Consider data obtained from an **observational study**. Compute the RR.

subject	A	Y	Y_1	Y_0
1	0	0	?	0
2	1	1	1	?
3	0	0	?	0
4	1	1	1	?
5	1	0	0	?
6	1	1	1	?
7	0	1	?	1
8	1	1	1	?
9	0	0	?	0
10	0	0	?	0

Exercise - observational study

Consider data obtained from an **observational study**. Compute the RR.

subject	A	Y	Y_1	Y_0
1	0	0	?	0
2	1	1	1	?
3	0	0	?	0
4	1	1	1	?
5	1	0	0	?
6	1	1	1	?
7	0	1	?	1
8	1	1	1	?
9	0	0	?	0
10	0	0	?	0

$$P(Y = 1|A = 1) = 4/5 = 0.8 > P(Y_1 = 1)$$

$$P(Y = 1|A = 0) = 1/5 = 0.2 < P(Y_0 = 1)$$

$$RR = \frac{0.8}{0.2} = 4 > CRR.$$

Remarks

In an **observational study**, we have that

$$\underbrace{P(Y_1 = 1|A = 1)}_{=P(Y=1|A=1)} \neq P(Y_1 = 1)$$
$$\underbrace{P(Y_0 = 1|A = 0)}_{=P(Y=1|A=0)} \neq P(Y_0 = 1).$$

In other words, we **do not have exchangeability**, meaning Y_0 and Y_1 depend on A:

$$(Y_0, Y_1) \not\perp\!\!\!\perp A.$$

As a consequence, **association** $\neq$ **causation** and $RR \neq CRR$.

Three important questions

- What is the cause of non-exchangeability in observational studies?
- Can we identify non-exchangeability in a population/sample?
- How can we estimate causal effects in the presence of non-exchangeability?

What is the cause of non-exchangeability in observational studies?

Suppose that there is a covariate, L, which affects both A and Y. For example:

- $L =$ 'age'; older people have higher BMI (A) than young people, and are more likely to develop cancer (Y)

If so, we will find an association between A and Y, even if A has no causal effect on Y.

The association between A and Y suffers from **confounding** by L. More on confounding later, but remember: **confounding causes non-exchangeability**.

Can we identify non-exchangeability in a population/sample?

We have non-exchangeability if (Y_0, Y_1) and A are not independent. That is, if

$$P(Y_1 = 1|A = 1) \neq P(Y_1 = 1)$$

or

$$P(Y_0 = 1|A = 0) \neq P(Y_0 = 1)$$

However, Y_1 is not observed for the unexposed ($A = 0$), and Y_0 is not observed for the exposed ($A = 1$). Thus, **the observed data can never tell us whether we have exchangeability or not**, or whether we have unmeasured confounding.

To judge whether exchangeability is plausible, we must rely on subject matter knowledge.

How can we estimate causal effects in the presence of non-exchangeability?

There are *several* ways to 'adjust' the analysis for potential confounders, including:

- stratification
- matching
- standardisation
- propensity scores
- regression modelling
- inverse probability weighting.

Conditional exchangeability

Adjusting for a potential confounder L produces a causal effect **if L is sufficient for confounding control** (more on this later).

Technically, if we have conditional exchangeability, given L:

$$\underbrace{P(Y_1 = 1|A = 1, L)}_{=P(Y=1|A=1,L)} = P(Y_1 = 1|L)$$

$$\underbrace{P(Y_0 = 1|A = 0, L)}_{=P(Y=1|A=0,L)} = P(Y_0 = 1|L)$$

$$(Y_0, Y_1) \perp\!\!\!\perp A|L.$$

Conditional exchangeability cannot be tested, and must be judged by subject matter knowledge. Exchangeability can be achieved by adjustments, but can also be 'destroyed'.

Stratification

The conceptually simplest way to adjust for a potential confounder L is by **stratification**:

- the study population is partitioned into strata (groups), one for each level of L
- each stratum is analysed separately
- within strata, no variation in $L \implies$ no imbalance in L across exposure levels.

A numerical example

	$L=1$		$L=0$	
subject	$Y=1$	$Y=0$	$Y=1$	$Y=0$
$A=1$	1	3	6	3
$A=0$	2	3	2	1

Compute $CRR|L$ for $L=1$ and $L=0$, assuming conditional exchangeability, given L.

Solution

	$L=1$		$L=0$	
subject	$Y=1$	$Y=0$	$Y=1$	$Y=0$
$A=1$	1	3	6	3
$A=0$	2	3	2	1

$$
\begin{aligned}
CRR(1) &= \frac{P(Y_1=1|L=1)}{P(Y_0=1|L=1)} = \{(Y_0,Y_1) \perp\!\!\!\perp A|L\} \\
&= \frac{P(Y_1=1|A=1,L=1)}{P(Y_0=1|A=0,L=1)} = \frac{P(Y=1|A=1,L=1)}{P(Y=1|A=0,L=1)} \\
&= \frac{1/4}{2/5} = 0.625 \\
CRR(0) &= \frac{P(Y_1=1|L=0)}{P(Y_0=1|L=0)} = \ldots = \frac{6/9}{2/3} = 1.
\end{aligned}
$$

Regression model for the outcome

Another relatively simple approach is to use regression methods. For example,

$$\text{logit}P(Y=1|A,L) = \alpha + \beta A + \gamma L$$

$$\beta = \ln(OR|L)$$

Which is asymptotically equivalent to stratification by L, if the model is correct.

This approach is useful for finite samples and sparse data (more on this later). Note: if the model is incorrect, then it may not produce anything interpretable.

Conditional effects vs marginal effects

Stratification gives **conditional causal effects** (within subsets of the population). For example, stratification by 'sex' gives the causal effect for men and women separately.

We may want to calculate the causal effect for the whole study population - a **marginal causal effect**:

- easier to interpret **one** marginal effect than **several** conditional effects
- randomized trials give marginal effects, and we may want to make results from observational studies comparable
- we may want to consider future interventions to the whole population, rather than to subsets.

The standardisation formula

Under conditional exchangeability, $P(Y_1)$ can be computed via **standardisation**:

$$P(Y_a = 1) = \sum_L P(Y = 1|A = a, L)P(L).$$

For binary L:

$$\begin{aligned} P(Y_a = 1) = {} & P(Y = 1|A = a, L = 1)P(L = 1) \\ & + P(Y = 1|A = a, L = 0)P(L = 0) \end{aligned}$$

Proof

By the law of total probability

$$P(Y_a = 1) = \sum_L P(Y_a = 1|L)P(L)$$

by conditional exchangeability, given L

$$\sum_L P(Y_a = 1|L)P(L) = \sum_L P(Y_a = 1|A = a, L)P(L)$$

by definition of potential outcomes

$$\sum_L P(Y_a = 1|A = a, L)P(L) = \sum_L P(Y = 1|A = a, L)P(L)$$

A numerical example - encore

	$L=1$		$L=0$	
subject	$Y=1$	$Y=0$	$Y=1$	$Y=0$
$A=1$	1	3	6	3
$A=0$	2	3	2	1

Compute $CRR|L$ for $L=1$ and $L=0$, assuming conditional exchangeability, given L.

Solution

	$L=1$		$L=0$	
subject	$Y=1$	$Y=0$	$Y=1$	$Y=0$
$A=1$	1	3	6	3
$A=0$	2	3	2	1

$$
\begin{aligned}
CRR &= \frac{P(Y_1=1)}{P(Y_0=1)} = \{(Y_0, Y_1) \perp\!\!\!\perp A|L\} \\
&= \frac{\sum_L P(Y=1|A=1,L)P(L)}{\sum_L P(Y=1|A=0,L)P(L)} = \\
&= \frac{\overbrace{1/4}^{P(Y=1|A=1,L=1)} \times \overbrace{9/21}^{P(L=1)} + \overbrace{6/9}^{P(Y=1|A=1,L=0)} \times \overbrace{12/21}^{P(L=0)}}{\underbrace{2/5}_{P(Y=1|A=0,L=1)} \times \underbrace{9/21}_{P(L=1)} + \underbrace{2/3}_{P(Y=1|A=0,L=0)} \times \underbrace{12/21}_{P(L=0)}} \\
&= 0.884.
\end{aligned}
$$

Stratification-based standardisation

$$P(Y_a = 1) = \sum_L P(Y = 1|A = a, L)P(L)$$

Explicit use of the standardisation formula leads to a two step procedure:

1. $P(Y = 1|A = a, L)$ is calculated for all levels of L, by stratification on L
2. these probabilities are averaged over the population distribution of L.

We refer to this method of standardisation as **stratification-based**.

Weighting-based standardisation

$P(Y_a = 1)$ can also be calculated using a method called **Inverse Probability Weighting (IPW)**.

Under conditional exchangeability, IPW uses weighting instead of stratification. We refer to this method of standardisation as **weighting-based**.

Summary

- Under **exchangeability**, association is equal to causation
- exchangeability follows by **randomisation**
- we typically don't have exchangeability in observational studies
- causal effects can be estimated in observational studies **if we make sufficient confounder adjustments**
  - issue: whether the adjustments are sufficient or not is untestable
- **stratification** produces sub-population (conditional) effects
- **standardisation** produces population (marginal) effects.

Ideal randomised trials

Exposed and unexposed are exchangeable:

$$(Y_0, Y_1) \perp\!\!\!\perp A$$

Association $=$ causation:

$$RR = CRR.$$

Observational studies

Exchangeability is often implausible. We may consider exchangeability more plausible if we adjust for some set of covariates:

$$(Y_0, Y_1) \perp\!\!\!\perp A|L$$

$$RR|L = CRR|L$$

However, selecting an appropriate set of covariates to adjust for is a non-trivial task.

Motivating example

Consider an observational study to investigate whether smoking during pregnancy (exposure) causes malformations (outcome) in newborns.

- For a large number of pregnancies, we collect data on both exposure and outcome
- We record five additional covariates:
  - the mothers age at conception
  - the mothers socioeconomic status at conception
  - the mothers diet during pregnancy
  - indicator of whether there is a family history of birth defects
  - indicator of whether the baby was liveborn or stillborn.

Motivating example, cont'd

We observe an unadjusted inverse association between smoking and malformations ($RR = 0.8$). However, we suspect that there is confounding of the exposure and outcome!

- If so, exposed and unexposed are not exchangeable, and
- the observed risk ratio cannot be given a causal interpretation

To reduce bias due to confounding we want to adjust for observed covariates.

Covariate selection

One strategy would be to adjust for all observed covariates. However, this approach may

- increase non-exchangeability, if the covariates are not confounders
- increase statistical uncertainty (e.g. wider confidence intervals).

Therefore, it is desirable to select a subset of covariates to adjust for. But what selection strategy should we use?

Traditional covariate selection strategies

- Adjust for covariates that are selected in a stepwise regression procedure
- Adjust for covariates that change point estimate of interest more than, say, 10%
- Adjust for covariates that
  - are associated with the exposure, and
  - are conditionally associated with the outcome, given the exposure, and
  - are not in the causal pathway between exposure and outcome.

Problem with traditional strategies

Traditional methods rely on statistical analyses of observed data, rather than a priori knowledge about causal structures. This has two important implications:

- they require data already collected, and cannot not be used at design stage
- they may select non-confounders, which may increase non-exchangeability if adjusted for.

Directed Acyclic Graphs

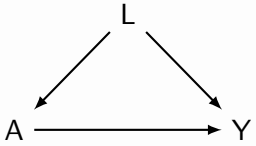
Directed Acyclic Graphs (DAG) can be used to overcome the problems with traditional covariate selection strategies.

A DAG is a graphical representation of underlying causal structures.

DAGs for covariate selection:

- encode our a priori causal knowledge/beliefs into a DAG
  - apply simple graphical rules to determine what covariates to adjust for.

A simple DAG



Each arrow represents a causal influence. The graph is

- **Directed**, since each connection between two variables consists of an arrow
- **Acyclic**, since the graph contains no directed cycles

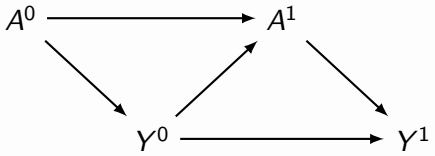
There is a formal connection to potential outcomes/counterfactuals through non-parametric structural equations (beyond the scope of this course).

A note on acyclicness

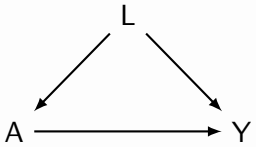
We impose acyclicness since a variable cannot cause itself: "*my BMI today has no effect on my BMI today*".

Observed variables are often snapshots of time varying processes: "*my BMI today certainly affects my BMI tomorrow*"

Time varying processes can be depicted by explicitly adding one 'realization' of each variable per time unit (more on this later).

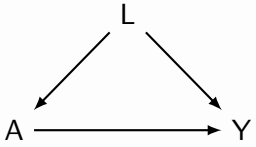


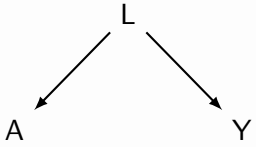
Underlying assumptions



Assumptions are encoded by the direction of arrows: the arrow from A to Y means that A may affect Y, but not the other way around.

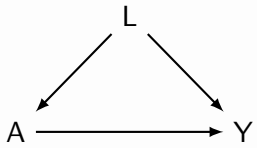
Underlying assumptions, cont'd

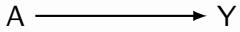




Assumptions are encoded by the absence of arrows: the presence of an arrow from A to Y means that A may or may not affect Y, while its absence means that A does not affect Y.

Underlying assumptions, cont'd

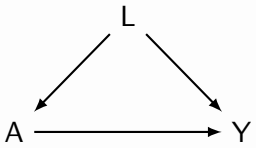




Assumptions about common causes:

- the presence of L means that A and Y may or may not have common causes
- the absence of L means that A and Y do not have any common causes.

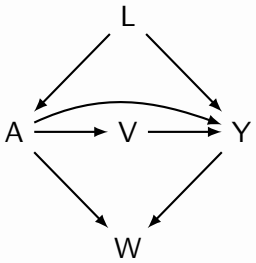
Ancestors and descendents



The **ancestors** of a variable are all other variables that affect it, either directly or indirectly: L is the single ancestor of A

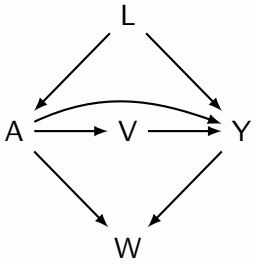
The **descendents** of a variable are all other variables that are affected by it, either directly or indirectly: Y is the single descendent of A.

Paths



A **path** is a route between two variables, not necessarily following the direction of arrows: which are the paths between A and Y?

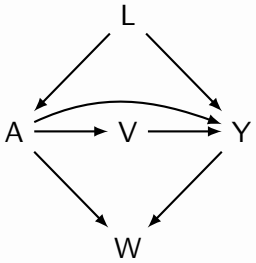
Solution



Four paths exist between A and Y:

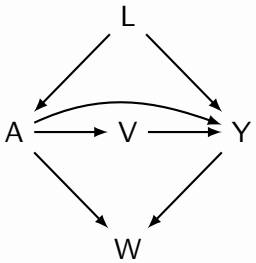
- $A \longrightarrow Y$
- $A \longrightarrow V \longrightarrow Y$
- $A \longleftarrow L \longrightarrow Y$
- $A \longrightarrow W \longleftarrow Y$.

Causal paths



A **causal path** is a route between two variables, **following the direction of arrows**. The causal paths from A to Y **mediate** the causal effect of A on Y, the non-causal paths do not. Which are the causal paths between A and Y?

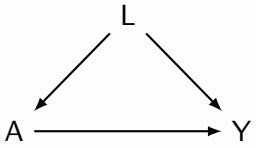
Solution



There are two causal paths from A to Y:

- $A \longrightarrow Y$
- $A \longrightarrow V \longrightarrow Y$.

Blocking of paths



Paths (both causal and non-causal) are either **open** or **blocked**, according to two rules.

Rule 1

A path is **blocked** if somewhere along the path there is a variable L that sits in a **chain**

$$\longrightarrow L \longrightarrow$$

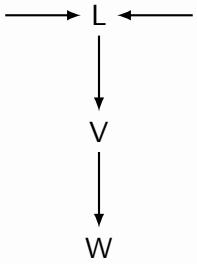
or in a **fork**

$$\longleftarrow L \longrightarrow$$

and we have adjusted for L.

Rule 2

A path is blocked if somewhere along the path there is a variable L that sits in an **inverted fork**



and we have **not** adjusted for L, or any of its descendents.

Once blocked, it stays blocked

$$A \longleftarrow V \longrightarrow W \longleftarrow Y$$

- Adjusting for V blocks the path from A to Y (rule 1)
- adjusting for W leaves the path open (rule 2)
- adjusting for both V and W blocks the path.

d-separation

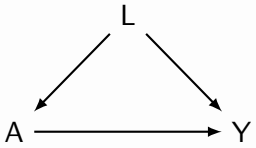
If all paths between A and Y are blocked by adjusting for L, then A and Y are conditionally independent, given L

$$Y \perp\!\!\!\perp A|L$$

In graph terminology, A and Y are **d-separated** by L.

Conversely, if at least one path is open, then A and Y are (most likely) conditionally associated, given L.

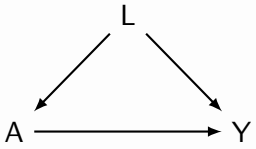
Example



Suppose the DAG above depicts the true causal structure. We want to test whether there is a causal effect of A on Y, i.e. does the causal path $A \longrightarrow Y$ exist?

- Concrete example?
- Adjust or not adjust for L?

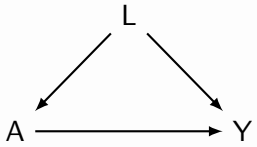
Heuristic argument



A smoking, Y malformations, L age.

- Young mothers smoke more often, but their babies have smaller risk for malformations, than old mothers.
- Hence, smokers are more likely to be young, and for this reason less likely to have babies with malformations, than old mothers.
- Thus, by not adjusting for age, we may observe an inverse association between smoking and malformations, even in the absence of a causal effect.

Formal solution based on d-separation

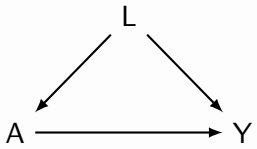


Suppose that we do not adjust for L, and that we observe an association between A and Y. There are two possibleexplanations for this association:

- the causal path $A \longrightarrow Y$
- the open non-causal path $A \longleftarrow L \longrightarrow Y$ (Rule 1)

An unadjusted association between A and Y does not prove that the causal path $A \longrightarrow Y$ exists!

Formal solution, cont'd

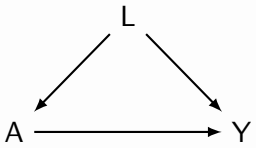


Suppose we do adjust for L: we block the non-causal path $A \longleftarrow L \longrightarrow Y$ (Rule 1).

Suppose that we observe an association between A and Y. This can now only be explained by the causal path $A \longrightarrow Y$.

Here, an adjusted association between A and Y proves that there is a causal effect of A on Y.

Conclusion



If the aim is to test for a causal effect of A on Y, then we should adjust for L.

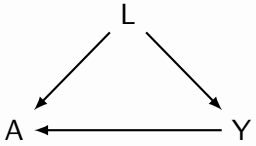
In terms of potential outcomes:

- exposed and unexposed are not marginally exchangeable
- exposed and unexposed are conditionally exchangeable, given L.

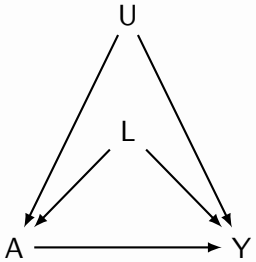
Remark

The argument assumes that the DAG is correct, in particular that

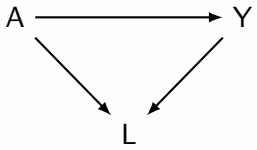
- Y does not cause A



- there is no additional common cause of A and Y



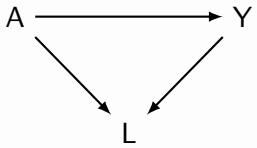
Example



Suppose that the DAG above depicts the true causal structure. We want to test whether there is a causal effect of A on Y, i.e. does the causal path $A \longrightarrow Y$ exist?

- Concrete example?
- Adjust or not adjust for L?

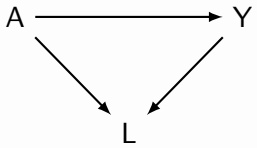
Heuristic argument



A smoking, Y malformations, L birth status (live/stillborn).

- Smoking and malformations increase the risk for stillbirth.
- Consider the group of woman who has stillbirths, but do not smoke: there must be another reason for the stillbirth, presumably that the baby had malformations.
- Thus, by adjusting for birth status, we may observe an inverse association between smoking and malformations, even in the absence of a causal effect.

Formal solution based on d-separation

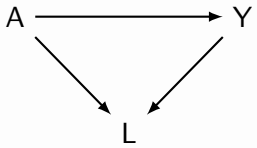


Suppose that we adjust for L, and that we observe an association between A and Y. There are two explanations for this association:

- the causal path $A \longrightarrow Y$
- the open non-causal path $A \longrightarrow L \longleftarrow Y$ (Rule 2).

Here, an adjusted association between A and Y does not prove that the causal path $A \longrightarrow Y$ exists.

Formal solution, cont'd

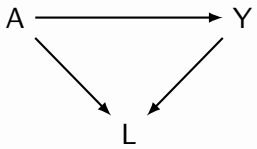


Suppose that we do not adjust for L. We block the non-causal path $A \longrightarrow L \longleftarrow Y$ (Rule 2).

Suppose that we observe an association between A and Y. This can now only be explained by the causal path $A \longrightarrow Y$.

Hence, here, an unadjusted association between A and Y proves that there is a causal effect of A on Y.

Conclusion



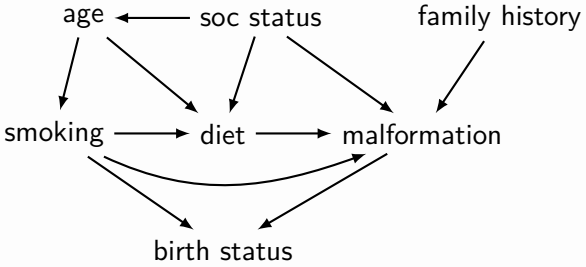
If the aim is to test for a causal effect of A on Y, then we should not adjust for L.

In terms of potential outcomes:

- exposed and unexposed are marginally exchangeable
- exposed and unexposed are not conditionally exchangeable, given L.

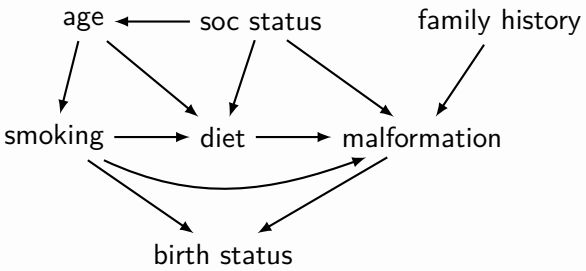
A possible DAG for the motivating example

Suppose we agree that the causal structures for our data can be described by the DAG below



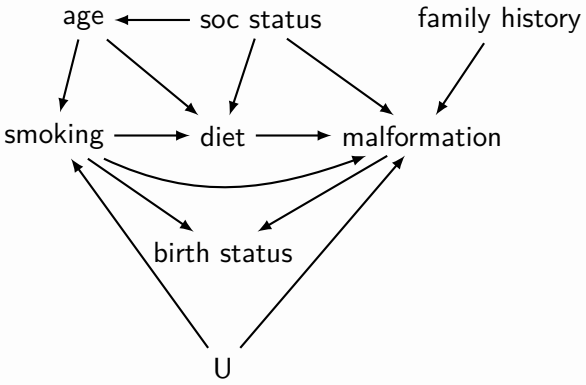
- Which assumptions are encoded in this DAG?
- Can these assumptions be tested?

Covariate selection



- Given the DAG, which covariates should we adjust for?
- Which covariates would be selected by the traditional strategies?

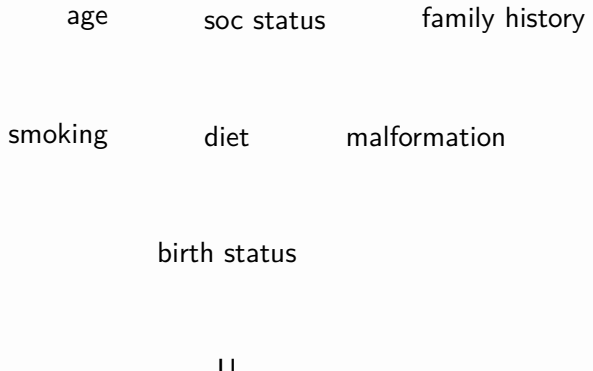
Unmeasured confounding



- Not a problem with DAGs, rather with observational studies
- reduce confounding bias by blocking as many non-causal paths as possible.

No a priori knowledge

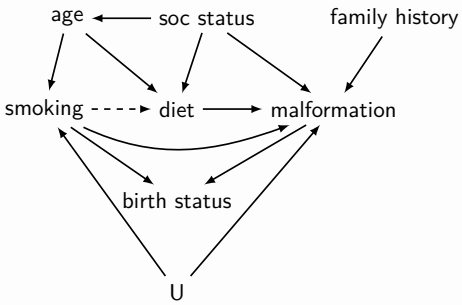
Without a priori knowledge we cannot construct a plausible DAG

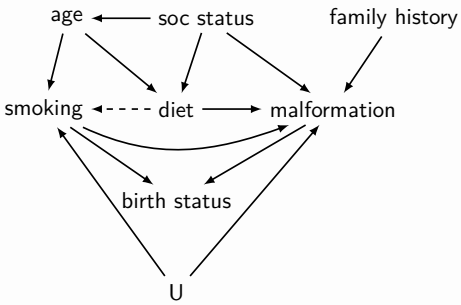


DAG-based covariate selection cannot be used, and we have to resort to traditional strategies. But be aware of the pitfalls.

Weak a priori knowledge

What if we cannot settle with **one** plausible DAG?

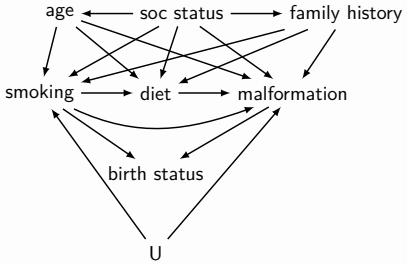




Present all plausible DAGs, and the implied analyses.

A complicated DAG

Consider a more complex situation where no/little covariate reduction is feasible.

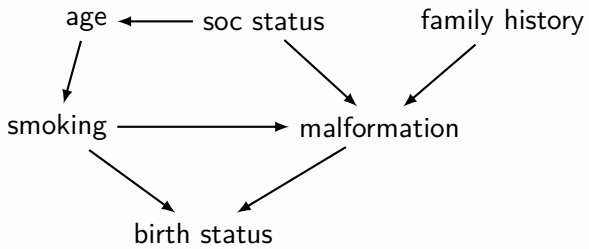


For every confounder that we adjust for, we may:

- decrease the bias due to confounding, but
- increase the statistical uncertainty (e.g. wider confidence intervals).

A reasonable trade off: exclude covariates with a relatively weak 'confounding effect'.

Confounding vs confounder



We have confound**ing** due to the non-causal path

$$\text{smoking} \longleftarrow \text{age} \longleftarrow \text{soc status} \longrightarrow \text{malformation}$$

Confound**ing** means that the exposure and the outcome have common causes.

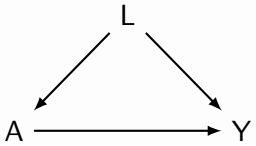
The path can be blocked by adjusting for either social status or age. A confound**er** is any variable that can be used to block a non-causal path, not necessarily the common cause.

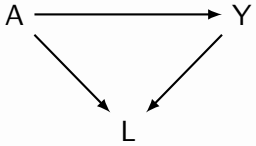
Testing vs estimation

We have learned how to use DAGs for covariates selection, given that we want to **test** for a causal effect, i.e. '*is there an effect, or is there no effect?*'

Often, our aim is on estimation: '*what is the magnitude of the effect?*'

Examples revisited





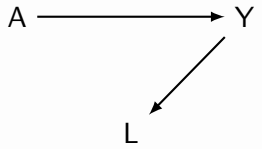
- In the left DAG, the conditional (given L) RD, RR, OR equal their **causal** conditional counterparts

$$RD|L = CRD|L, \quad RR|L = CRR|L, \quad OR|L = COR|L$$

- In the right DAG, the marginal RD, RR, OR equal their **causal** marginal counterparts

$$RD = CRD, \quad RR = CRR, \quad OR = COR.$$

A valid test does not imply a valid estimate



In this DAG, adjusting for L is not necessary.

- it does not 'invalidate' the test: only one path, $A \longrightarrow Y$, which is causal
- it does 'invalidate' the RD and the RR

$$RD|L \neq CRD|L, \quad RR|L \neq CRR|L$$

but not the OR

$$OR|L = COR|L.$$

Summary

Traditional covariate selection strategies do not use a priori knowledge about causal structures

- difficult to apply at the design stage
- may select non-confounders, which may increase non-exchangeability.

DAGs can be used for covariate selection

- encode our a priori causal knowledge/beliefs into a DAG
- apply simple graphical rules to determine what covariates to adjust for.

DAGs are not only tools for covariate selection: generally speaking, they are used to facilitate interpretation and communication in causal inference.

References

- Hernán MA and Robins J (2021). Causal Inference - What If. CRC Press url, book page.
- Hernán MA, Hernández-Díaz S, Werler MM and Mitchell AA (2002). Causal Knowledge as a Prerequisite for Confounding Evaluation: An Application to Birth Defects Epidemiology. American Journal of Epidemiology 155 (2): 176–84 url.
- Hèrnan MA (2017). Causal Diagrams: Draw Your Assumptions Before Your Conclusions. url.
- Hèrnan MA,Clayton D and Keiding N (2011). The Simpon's paradox unraveled. International Journal of Epidemiology 40 (3): 780–85.
- Munafò MR, Tilling K, Taylor AE, Evans DM and Smith GD (2018). Collider Scope: When Selection Bias Can Substantially Influence Observed Associations. International Journal of Epidemiology 47 (1):226–35 url.

References

- Murray E (2018). What Is a Cause? url.
- Neuhaus JM and Jewell NP (1993). A Geometric Approach to Assess Bias Due to Omitted Covariates in Generalized Linear Models. Biometrika 80 (4): 807–15.
- Pearl J (2018). The Book of Why: The New Science of Cause and Effect. New York, US: Allen Lane.
- Pearl J, Glymour M and Jewell NP (2016). Causal Inference in Statistics: A Primer. Chichester, UK: Wiley.
- Pearl J (1995). Causal Diagrams for Empirical Research. Biometrika 82 (4): 669–88 url.
- Rubin DB (2005). Causal Inference Using Potential Outcomes. Journal of the American Statistical Association, 100(469): 322–31 url.