University of Basel
Department of
Mathematics and Computer Science

A gentle introduction to causal thinking
Basic concepts and definitions

Giusi Moffa
a_randomwalker giusi.moffa@unibas.ch

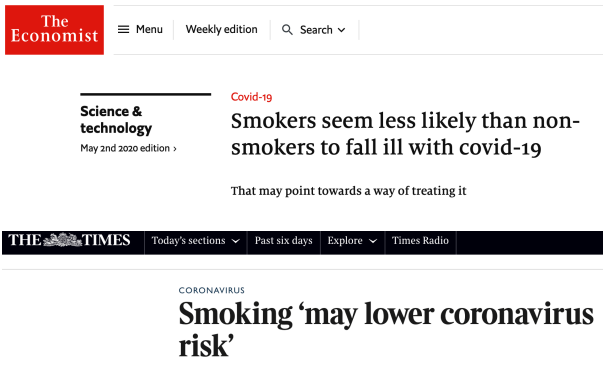
BBS Training Series 2021
Tuesday, February 2nd, 2021

Course plan and objectives

- The concept of causal questions
- Definition of causal effects in the framework of potential outcomes
- (Causal) Identifiability assumptions
- Causal diagrams: reasoning about causality with DAGs

How do we define *causal effects*?

...and why do we even care?



Causality related concepts are an integral part of human life

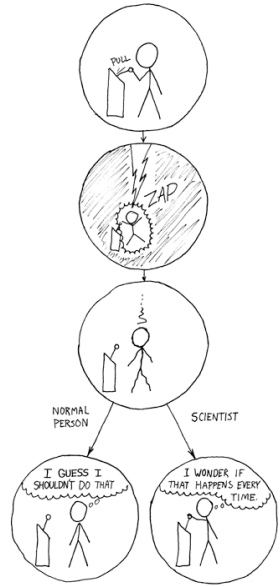
Our ability to

- learn about the world
- achieve scientific progress
- or even make (informed) personal decisions

relies on being able to distinguish

- the signal from the noise
- evidence from anecdotes

and establish *reproducible* cause effect relationships...



https://xkcd.com/242/

How do we formulate (valid) causal questions?

Are any of the following valid causal questions?

Not everyone agrees on the fine prints of course...

- Is going to university causing brain tumour?
- Is coffee good/bad for you?
- Is coffee decreasing the risk of early death?
- Is alcohol bad for you?
- Does alcohol accelerate cognitive decline?
- Does smoking increase the risk of cancer?
- Is eating red meat increasing the risk of cancer?
- Are people responsible for climate change?

Does Obesity Shorten Life? Or is it the Soda? On Non-manipulable Causes *Judea Pearl, 2018*

Does obesity shorten life? ...well-defined interventions ... *Hernán & Taubman, 2008*

What is the difference and how do we answer such questions once we agree on the defintion?

In Donald Rubin's words

If you are not talking about intervention, you can't talk about causality.

Pragmatic definition of *causal effect* as the effect of an intervention

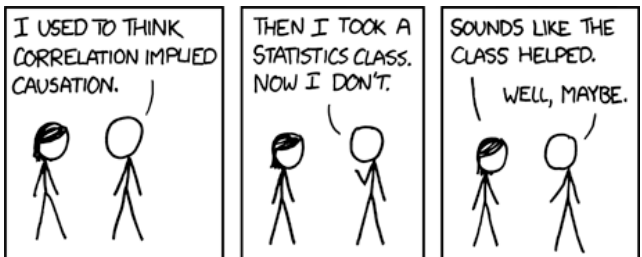
Intuitively if we have an outcome Y *(: lung cancer)* and an exposure (treatment, intervention) variable A *(: smoking)*, we may agree that A *causes/has a causal effect on* Y if **changing** the value of A **will change** the distribution of Y.

Does it mean that we can only define causal effects of manipulable causes?
Are causal effects of age or sex ill defined concepts?

And what sort of intervention? **Real? Hypothetical? Simulated?**

To find out what happens when you change something, it is necessary to change it
Statistics for Experimenters: Design, Innovation, and Discovery, 1st edition 1978
George E.P. Box, J.Stuart Hunter and William G. Hunter

If an exposure A has a causal effect on an outcome Y, we expect to find an association when observing Y and A



If we have taken a statistics class before, no doubt we have become accustomed with the **"correlation does not imply causation" mantra**, and we know that the reverse is not necessarily true: if we simply observe a correlation between A and Y we cannot conclude that a causal relationship exists - as tempting as it may be...

Everyday in the news - Anecdotes vs evidence

- genetics
- lifestyle
- never stopped working
- healthy diet
- lots of walking
- drinking tea

105 year old claims secret to her long life is 'full fat cream, butter and whisky'; Gorham woman's secret to living to 100 is no sweets; Cruises and keeping fit are 107-year-old's secret to a long life; Azerbaijan's secret to long life? Mountain air

Do they sound ridiculous? How about:

"I took an aspirin and my headache went away - the drug worked!"

If correlation does not imply causation, what does?

So the question is: what are we actually trying to estimate, and when does association imply causation?

How do we distinguish anecdotes from evidence?

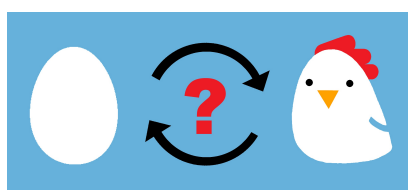
How do we evaluate the effect of a (possibly hypothetical) intervention?

What sort of study do we need?

Interventional? Controlled? Randomised?

Block what you can, randomise what you cannot. (Box?)

Effects of causes or causes of effects?



Potential outcomes

A framework to evaluate the effects of causes

Treatment, outcome and potential outcomes

Denote by A the action of interest (may be referred to as treatment, exposure or intervention, depending on the field of study) and by Y the outcome under study, with both A and Y random variables.

A: flu shot, surgery... Y: flu, death...

The entity (person, object, place or whatever) upon which the action operates *at a particular point in time*, is the **unit** (the same entity at a different point in time is treated as a different unit)

Potential outcomes - counterfactuals

At the individual level we can define the *potential outcomes* as the random variables corresponding to each of the outcome Y^a we would see for a given unit under each treatment option $A = a$.
Once treatment is chosen/happens only one of the potential outcomes can be observed, the other is counterfactual...

Potential outcomes - binary treatment and outcome

Specifically for a dichotomous treatment A (e.g. flu shot, surgery, ...) and a dichotomous outcome Y

$$A = \begin{cases} 1 & \text{treated} \\ 0 & \text{untreated} \end{cases} \qquad Y = \begin{cases} 1 & \text{flu (death)} \\ 0 & \text{no flu (survival)} \end{cases}$$

Formally

- $Y^{a=1}$ denotes the potential outcome under treatment $a = 1$
- $Y^{a=0}$ denotes the potential outcome under treatment $a = 0$.

Each unit has two potential outcomes $Y^{a=1}, Y^{a=0}$, and before treatment happens (*is chosen*) both have the potential to become the *actual* outcome

Individual level causal effect - formal definition

Unit	Potential outcomes $Y^{a=1}$	$Y^{a=0}$
Jupiter	0	1
Juno	1	1
Apollo	1	1
Diana	0	0
Neptune	1	0
Minerva	0	1
Mars	1	0
Venus	1	0
Mercury	0	1
Ceres	0	0
Vulcan	0	0
Vesta	1	1

Colours highlight categories of individuals based on their potential outcome patterns, but no theoretical reason for the types to be balanced

Gods' example adapted from Hernán and Robins (2020)

Individual level: treatment A has a causal effect on outcome Y if the potential outcomes for that individual differ

$$Y^{a=1} \neq Y^{a=0}$$

For each unit once the treatment is assigned only one potential outcome can happen - *whence* the terminology of *factual and counterfactual outcomes*

1. actual treatment $A = 1$ and observed outcome $Y^{a=1}$
2. actual treatment $A = 0$ and observed outcome $Y^{a=0}$

Counterfactual outcomes: those which would have been observed had the treatment been different from the actual treatment (*counter to the fact*)

Fundamental problem of causal inference

In practice we can only ever observe one potential outcome for each unit.

Based on the data we can observe, and regardless of the type[#] of study (observational or interventional), *without explicit assumptions on the structural model of the data generating mechanism:*

The individual level causal effect is out of reach (non identifiable)

If a unit gets the flu shot, the question of what would have happened to them if they did not get the shot is pretty much **hopeless**
The same is true for a unit who does not get the flu shot: we cannot know what would have happened if they did.

The population level causal effect may be identifiable under milder assumptions

The proportion of units who get the flu if everybody gets a flu shot, vs the proportion of units who get the flu if nobody gets a shot

[#] Xover experimental designs may enable individual causal effects identification under additional assumptions

Population level (average) causal effect

Unit	Potential outcomes $Y^{a=1}$	$Y^{a=0}$
Jupiter	0	1
Juno	1	1
Apollo	1	1
Diana	0	0
Neptune	1	0
Minerva	0	1
Mars	1	0
Venus	1	0
Mercury	0	1
Ceres	0	0
Vulcan	0	0
Vesta	1	1

At the population level: treatment A has a non-null average causal effect on outcome Y in the population of interest if

$$\mathbb{E}\left[Y^{a=1}\right] \neq \mathbb{E}\left[Y^{a=0}\right]$$

- $\mathbb{E}\left[Y^{a=1}\right]$ the expected outcome in the population if everybody is treated, from the column of potential outcomes $Y^{a=1}$
- $\mathbb{E}\left[Y^{a=0}\right]$ the expected outcome in the population if everybody is left untreated, from the column of potential outcomes $Y^{a=0}$

For binary outcomes: $P(Y^{a=1} = 1) \neq P(Y^{a=0} = 1)$

The sharp causal null hypothesis (no individual causal effects), implies the null hypothesis of no average causal effect - but the reverse does not hold.

Equivalent terminology: ATE (Average treatment effect), ACE (Average causal effect)

Is the population average causal effect identifiable (and estimable)?

If we were given the potential outcomes for all (N) sample units an estimate for the causal estimand at the population level would be

$$\frac{1}{N}\sum_{i=1}^{N}(Y_i^{a=1} - Y_i^{a=0})$$

Rather than comparing functions of individual level potential outcomes we compare functions of their distributions, in this case expectations.

For binary response it may also be risk ratios or odds ratio - (risk difference are the same as expectation for binary variables...)

Are we getting around the fundamental problem of causal inference? $$$

- can we actually identify (and estimate) population level causal effects from observed data?

Not yet, unless we make *causal assumptions*

What do the data we may observe look like?

In reality, we may only ever see one realization of the world

	Potential outcomes		Treatment	Outcome
Unit	$Y^{a=1}$	$Y^{a=0}$	A	Y
Diana	0	0	0	0
Ceres	0	0	1	0
Vulcan	0	0	1	0
Jupiter	0	1	1	0
Minerva	0	1	1	0
Mercury	0	1	1	0
Neptune	1	0	0	0
Mars	1	0	0	0
Venus	1	0	1	1
Juno	1	1	0	1
Apollo	1	1	0	1
Vesta	1	1	1	1

- Treatment: *actual* treatment assigned in a real study
- Outcome: observed outcome (given the actual treatment)

The *conditional* probability of getting the flu for gods who *happened* to be treated gods is

$$P(Y = 1 \mid A = 1) = 2/7$$

and for those who *happened* to be left untreated is

$$P(Y = 1 \mid A = 0) = 2/5$$

Is there a population level causal effect?

Potential outcomes: $P(Y^{a=1} = 1) = P(Y^{a=0} = 1) = 5/10 = .5$

Association versus causation

For the gods' example the conditional probabilities differ

$$P(Y = 1 \mid A = 1) = 2/7 < 2/5 = P(Y = 1 \mid A = 0)$$

but the potential outcomes do not $P(Y^{a=1} = 1) = 5/10 = .5 = 5/10 = P(Y^{a=0} = 1)$

	Potential outcomes		Treatment	Outcome
Unit	$Y^{a=1}$	$Y^{a=0}$	A	Y
Diana	0	0	0	0
Ceres	0	0	1	0
Vulcan	0	0	1	0
Jupiter	0	1	1	0
Minerva	0	1	1	0
Mercury	0	1	1	0
Neptune	1	0	0	0
Mars	1	0	0	0
Venus	1	0	1	1
Juno	1	1	0	1
Apollo	1	1	0	1
Vesta	1	1	1	1

Ignoring sampling variability or assuming we have millions of observations...

Different conclusions in terms of association and causal effects

- an association between treatment and outcome
- no causal effect of the treatment on the outcome

Flu shot disclaimer

The example is entirely synthetic and it is not meant to cast doubt on the effectiveness of flu vaccines

Seeing versus doing

Why is it that the difference between the expected values of the potential outcomes is not the same as the expected value of quantities we may observe?

Why do we have the inequalities?

- $P(Y^{a=1} = 1) \neq P(Y = 1 \mid A = 1)$
- $P(Y^{a=0} = 1) \neq P(Y = 1 \mid A = 0)$

Why are the potential outcome risks *in general* different than the risks we observe in treated and untreated?

or more generally

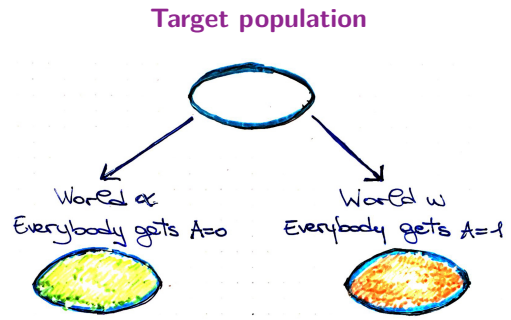
- $\mathbb{E}\left[Y^{a=1}\right] \neq \mathbb{E}\left[Y \mid A = 1\right]$
- $\mathbb{E}\left[Y^{a=0}\right] \neq \mathbb{E}\left[Y \mid A = 0\right]$

...or the expected values...

What is the difference between the quantities on the left hand side and those on the right hand side of these equations?

We observe the real or actual world, but causal effects are defined in terms of a hypothetical world...

Hypothetical worlds



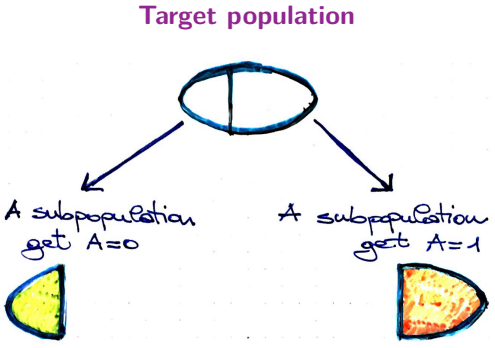
Causal effect: the effect we would see when comparing two hypothetical worlds with the same target population, but where in one world everybody in the population gets treated while in the other world one no-one gets treated.

Questions about causation are about asking **"what if"**:
what would be the risk if everybody had been treated? versus *what would be the risk if everybody had been left untreated?*

The difference between the mean in world α and the mean in world ω is the average causal effect

$$\mathbb{E}\left[Y^{a=1} - Y^{a=0}\right] = \mathbb{E}\left[Y^{a=1}\right] - \mathbb{E}\left[Y^{a=0}\right]$$

Illustration adapted from Hernán and Robins (2020)



Real world

It may be the best of all possible worlds if we trust **Gottfried Leibniz**; but it still complicates things for causal inference.

Inferences about associations are concerned with questions about the **actual world**: *what is the risk in the treated?* versus *what is the risk in the untreated?*

The difference in mean this time is comparing different groups of people, defined by the treatment they receive.

They may also differ in other ways which may lead to different outcome expectations. Therefore in general the difference observed in the real world is not the same as the causal effect, which should compare the **same** group of people treated in different ways

$$\mathbb{E}[Y \mid A = 1] - \mathbb{E}[Y \mid A = 0] \neq \mathbb{E}\left[Y^{a=1} - Y^{a=0}\right]$$

Illustration adapted from Hernán and Robins (2020)

Recap: Association vs causation, Conditioning vs setting

Association is defined by a different average response/outcome in *disjoint subsets* of the study population. For a binary treatment we have two sets, with membership defined by the actual treatment of each unit, treated ($A = 1$) and untreated ($A = 0$)

Causation is defined by a different average response/outcome in the **same population**, under different treatments (again with two possibilities for a binary treatment: $a = 1$ and $a = 0$)

The distinction between *setting* a variable to a given value by intervention with respect to *observing* it is absolutely essential in causal inference.

Conditioning (seeing) may be interpreted as referring to a *subpopulation* (those who happened to take the flu shot), which may not be representative of the entire target population.

When comparing the average response in those who *happened* to take the flu shot to that in those who *happened* not to, we are comparing different subpopulations, which may differ from each other in important and fundamental ways.

Measures of causal effect - causal estimands

Dichotomous outcomes

The causal null hypothesis may be expressed in terms of... as...

- Causal risk difference: $P(Y^{a=1} = 1) - P(Y^{a=0} = 1) = 0$, additive scale
- Causal risk ratio: $\frac{P(Y^{a=1}=1)}{P(Y^{a=0}=1)} = 1$, multiplicative scale
- Causal odds ratio: $\frac{P(Y^{a=1}=1)/P(Y^{a=1}=0)}{P(Y^{a=0}=1)/P(Y^{a=0}=0)} = 1$

Continous outcome

One way of expressing the causal null hypothesis is with respect to the outcome expected value

- Average causal effect: $\mathbb{E}\left[Y^{a=1}\right] - \mathbb{E}\left[Y^{a=0}\right] = 0$

In the presence of a causal effect causal parameters will differ from the values characterising the causal null hypothesis, in a measure which quantifies the strength of the causal effect on different scales (fit for different purposes).

Causal effects in different populations

More generally we may also define a causal effect in different sub-groups, for example within a sub-population with a certain level of a covariate $V = v$ as

$$\mathbb{E}\left[Y^{a=1} \mid V = v\right] - \mathbb{E}\left[Y^{a=0} \mid V = v\right]$$

or within the sub-population identified by the treated group

$$\mathbb{E}\left[Y^{a=1} \mid A = 1\right] - \mathbb{E}\left[Y^{a=0} \mid A = 1\right]$$

(Causal) Identifiability assumptions

Do real world data ever imply causal statements?

Are there any conditions under which we can use *real world data* to make causal inference?

Running a randomised experiment is perhaps the most well established method or gold standard...

Treatment randomisation ensures that the observed average difference is an **unbiased** estimate of the population average causal effect

Randomised study: $P(A = a \mid Z) = P(A = a) = p_a$ independently of any covariates Z.

For a binary treatment (1/0:Yes/No) typically $P(A = 1) = P(A = 0) = \frac{1}{2}$

but in general equality is not required for the identifiability of causal effects (though a convenient choice for statistical arguments, e.g. power)

Precise probabilistic statements are another important "side effect" of randomisation

On the additive scale (and for a continuous variable) for example, from a randomised study we can estimate the average treatment/causal effect as

$$\begin{aligned}
\mathbb{E}\left[Y^{a=1}\right] - \mathbb{E}\left[Y^{a=0}\right] &= \mathbb{E}\left[Y^{a=1} \mid A=1\right] - \mathbb{E}\left[Y^{a=0} \mid A=0\right] \\
&= \mathbb{E}\left[Y \mid A=1\right] - \mathbb{E}\left[Y \mid A=0\right] \\
&\simeq \frac{1}{N_1} \sum_{i \in S_1} Y_i - \frac{1}{N_0} \sum_{i \in S_0} Y_i
\end{aligned}$$

with S_1 the set of N_1 indices (randomly) assigned to treatment $A=1$, and S_0 the set of N_0 indices (randomly) assigned to $A=0$.

The 1st and 2nd equalities follow from ignorability and consistency respectively, with the first failing for observational studies *(more on identifiability assumptions in a moment)*.

Rubin's perfect doctor example

Unit	Potential outcomes $Y^{a=1}$	$Y^{a=0}$
Jupiter	14	13
Juno	0	6
Apollo	1	4
Diana	2	5
Neptune	3	6
Minerva	1	6
Mars	10	8
Venus	9	8
True average	5	7

Potential outcomes from a hypothetical study comparing two types of surgery with respect to post-operative life expectancy

- the outcome Y is the number of years lived since surgery
- $a = 1$: novel surgical treatment
- $a = 0$: traditional surgery

The true average causal effect is
$\mathbb{E}\left[Y^{a=1} = 1\right] - \mathbb{E}\left[Y^{a=0} = 1\right] = 5 - 7 = -2$

What happens if we have our dream doctor, who chooses the best treatment for each patient (the one under which the patient will live longer), and they flip a coin if there is no difference?

	Potential outcomes		Treatment	Outcome
Unit	$Y^{a=1}$	$Y^{a=0}$	A	Y
Jupiter	14	13	1	14
Juno	0	6	0	6
Apollo	1	4	0	4
Diana	2	5	0	5
Neptune	3	6	0	6
Minerva	1	6	0	6
Mars	10	8	1	10
Venus	9	8	1	9

The observed average survival times are

- $\mathbb{E}[Y \mid A = 1] = 11$
- $\mathbb{E}[Y \mid A = 0] = 5.4$

If we were to use the observed values to estimate the average causal effect we would get

$\mathbb{E}[Y \mid A = 1] - \mathbb{E}[Y \mid A = 0] = 5.6 \neq -2$ and in the opposite (!) direction.

If we were to draw conclusions from the outcomes observed following the **perfect doctor assignment** we would say that

- the new surgery on average increases life expectancy by 5 years
- the average life expectancy if everyone underwent the new surgery would be 11 years

But we *know* (hypothetically, based on the potential outcomes) that

- the new treatment on average shortens life by 2 years
- the average life expectancy if everyone received the new surgical treatment would be 5 years

Where is the mistake?

The treatment assignment depends on the potential outcome (**non ignorable** in Rubin's terms), in such a way that we observe the most extreme outcome, so much that we get an effect in the wrong direction, totally misleading...

Identifiability assumptions

With observational data we need additional **identifiability assumptions**

- SUTVA - stable unit-treatment value assumption $\sim$ consistency
- ignorability $\sim$ exchangeability $\sim$ no unmeasured confounders
- probabilistic (as opposed to deterministic) assignment $\sim$ positivity

about the observed data, usually: an outcome Y, a treatment A and a set of pre-treatment covariates Z

There are different assumptions for different approaches, but these are probably the most common ones. Identifiability means that we can use observed data Y, A, Z to estimate a causal parameter defined in terms of potential outcomes.

Assuming that the identifiability assumptions hold for a *non-randomised* study equates to conceptualise it as if the study were randomised within levels of the variables Z which guarantee *unconfoundedness* (i.e. a conditionally randomised experiment).

SUTVA

Stable Unit Treatment Value Assumption - two parts

a. There is only one form of treatment
   - for each exposure considered: e.g. one single form of active treatment, one single form of control

b. There is no interference among units
   - treatment assignment of one unit does not affect the outcome of another unit
   - no spillover or contagion

Some vaccine trials are a typical example where the no interference assumption does not hold.

SUTVA implies that we can write the potential outcome of each unit i only in terms of that unit treatment $Y_i^{A_1,...A_i,...A_N} = Y_i^{A_i}$

Consistency

For every study unit i the observed outcome coincides with the potential outcome corresponding to the actual treatment received

$$Y_i = Y_i^a \text{ if } A_i = a \ \forall a$$

Exchangeability

$$Y^a \perp\!\!\!\perp A \forall a$$

The potential outcome and the actual treatment are independent; the actual treatment does not predict the potential outcome and vice versa.

Conditional exchangeability: $Y^a \perp\!\!\!\perp A \mid Z \forall a$

The potential outcome and the actual treatment are independent within levels of the covariates Z.

Positivity

The probability of treatment assignment is non-zero for all treatments and for all levels of the covariates Z

$$P(A = a \mid Z = z) > 0 \forall a, \forall z : P(Z = z) > 0$$

Suitable adaptations needed for continuous Z, to consider ranges of Z with positive density...

It is important for the identification of causal effects that there is variability in treatment assignment, in that all "strata" of the target population have a non-zero probability of being assigned to any of the possible treatments.

Unlike exchangeability, positivity may sometimes be empirically verified.

See Westreich and Cole (2010) for a discussion of practical aspects of the positivity assumption.

Alternatively: Strong ignorability - no unmeasured confounders

For a binary treatment

$$(Y^1, Y^0) \perp\!\!\!\perp A \mid Z; 0 < P(A = 1 \mid Z = z) < 1 \forall z : P(Z = z) > 0$$

Ignorable and probabilistic treatment assignment (positivity).

Treatment assignment is ignorable if it is independent of all potential outcomes, conditional on a given a set of observed pre-treatment covariates Z characterizing treatment assignment.

Treatment assignment is not marginally independent of the potential outcomes, but independence holds within levels of Z.

Ignorability implies conditional *exchangeability* $(Y^a \perp\!\!\!\perp A \mid Z \forall a)$

Randomisation implies strong ignorability.

Under the given assumptions we can link the observed data distribution to the potential outcome distribution

The conditional expected value of the (observed) outcome Y given treatment $A = a$ and covariates $Z = z$ is

$$\mathbb{E}\left[Y \mid A = a, Z = z\right] = \mathbb{E}\left[Y^a \mid A = a, Z = z\right]$$

with the equality justified by the consistency assumption. Ignorability then implies

$$\mathbb{E}\left[Y^a \mid A = a, Z = z\right] = \mathbb{E}\left[Y^a \mid Z = z\right]$$

So that we can estimate the conditional expectation of potential outcomes from the observed data

$$\mathbb{E}\left[Y \mid A = a, Z = z\right] = \mathbb{E}\left[Y^a \mid Z = z\right]$$

Standardization: stratifying plus averaging

To obtain marginal causal effects we will need to average over Z

$$\mathbb{E}\left[Y^a\right] = \mathbb{E}_Z\left[\mathbb{E}\left[Y^a \mid Z = z\right]\right] = \sum_z \mathbb{E}\left[Y \mid A = a, Z = z\right] P(Z = z)$$

with the first equality following from the law of iterated expectations.

Finite populations and increasing dimension of Z may quickly result in (random) non-positivity, due to empty cells - raising the need for modelling.

Non-parametric estimation of effects becomes quickly unfeasible.

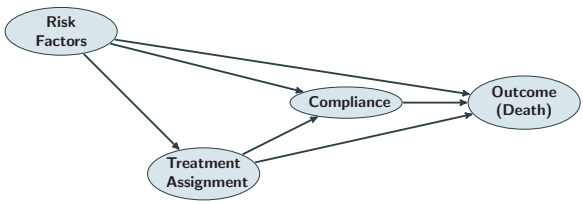
The sum will be an integral for continuous covariates

Causal diagrams

Reasoning about causality with Directed Acyclic Graphs

Graphical representation of causal relations

...in medias res...



Causal diagrams

Alternative (? rather complementary) framework to describe causal effects

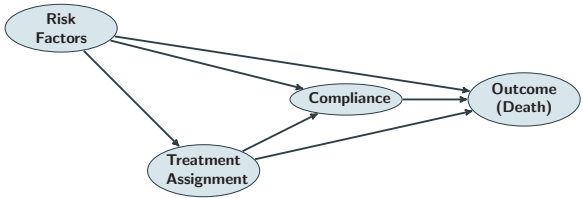
Directed edges represent direct causal effect - **changing** the value of a variable where an arrow originates will determine a change in the value of the variable where the arrow points

Directed paths represent *causal paths* (e.g. RF $\rightarrow$ Comp $\rightarrow$ Out)

Remarks

- DAGs represent real world variables and cannot explicitly represent potential outcomes
  That requires SWIGs (Single World Intervention Graphs), Richardson and Robins (2003)
- However, given a **valid** causal diagram - there are graphical criteria which inform us about which variables may be sufficient to identify the effect of any variable on another

Causal paths as water pipes



By simply looking at a causal diagram we can establish which variables are expected to be associated (marginally and/or conditionally) with each other, and which associations correspond to a causal path.

Causation only goes through directed paths

Association may flow through any path depending on whether it is open or closed - and we can think of every path as a water pipe with taps on the way

If we are interested in the *total effect* of treatment assignment on the outcome, we wish to make sure that we

- leave all causal paths open
- close all non-causal paths
- do not *accidentally* open any non-causal paths (!!!)

Which paths are open and which ones are closed?

Open paths in causal diagrams

(Marginal) association between A and Y may flow in any of the following configurations

direct link

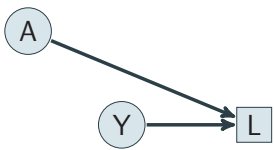
causal chain

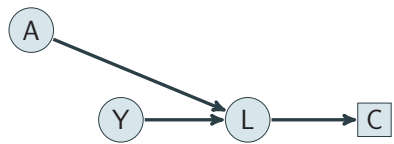
causal fork

(Conditional) association between A and Y may flow when they have a common effect L and we restrict the analysis to certain levels of L or of one of its descendants C.

causal collider

A square indicates conditioning on...

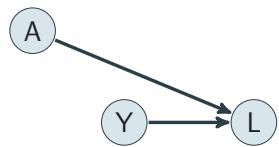




Closed paths in causal diagrams

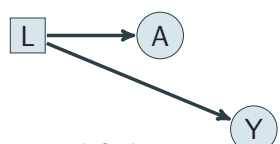
(Marginal) association between A and Y cannot flow on a path with colliding arrow heads

A collider is a (marginally) closed tap - no water flows until we open it.



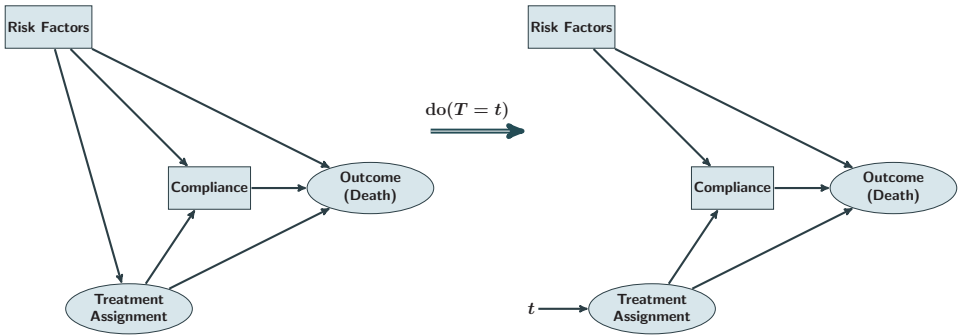
(Conditional) association between A and Y cannot flow through chains and forks when we restrict the analysis to certain levels of the middle node L

causal chain



causal fork

Randomisation: what it does and what it doesn't (do)



Randomisation takes care of the confounding due to the risk factors (common causes)

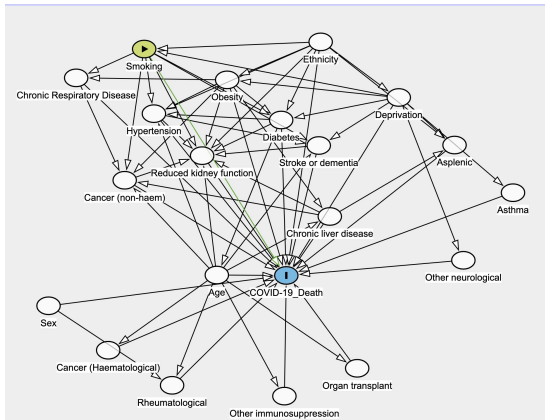
Randomisation does not address the potential bias for selecting on effects of treatment

Loss to follow-up, with unobserved outcome, would also be such an example - ITT effect?

Back to covid and smoking

Misunderstandings originating by the Table 2 fallacy through potential *overadjustment* in the presence of mediating variables and colliders.

Drawing a diagram has the potential to clearly highlight this type of problem and avoid flawed conclusions.



Article | Published: 08 July 2020

Factors associated with COVID-19-related death using OpenSAFELY

Elizabeth J. Williamson, Alex J. Walker, [...] Ben Goldacre ✉

Nature **584**, 430–436(2020) | Cite this article

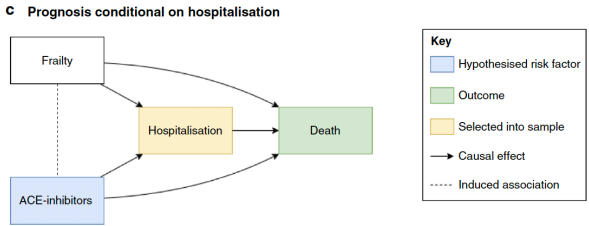
#tweetorial by @EpiEllie

Collider bias: why it's difficult to find risk factors or effective medications for COVID-19 infection and severity

ARTICLE

https://doi.org/10.1038/s41467-020-19478-2 OPEN

Collider bias undermines our understanding of COVID-19 disease risk and severity



Restricting the analysis to hospitalised patients is *closing* a causal path between an exposure of interest and the outcome - blocking the flow of association on that path.

Griffith et al. (2020)

Wrapping up: draw your assumptions before your conclusions

Thought provoking title of a free online course by Miguel Hernán.

Formulating valid causal questions is a cornerstone of causal inference (CI).

CI rests on expert knowledge and untestable assumptions about the causal network linking treatment, outcome and other covariates - especially in observational studies.

Randomisation is the gold standard to establish the effect of interventions, with the unique merit of eliminating common cause confounding (of measured *& unmeasured confounders*)...the best we have, but nobody is perfect and potential hurdles remain to deal with unplanned post-randomisation events, where we may still benefit from a CI mindset.

Being explicit about what we know and what we assume is a critical component of CI.

Causal diagrams represent *qualitative* a priori knowledge of a causal mechanism of interest, and they may

- help uncover/clarify conceptual misconceptions
- enhance communication among investigators
- increase transparency with respect to assumptions

References - suggested readings

Causal inference: what if
Miguel A Hernán, James M Robins, 2020

Causal Inference for Statistics, Social and Biomedical Sciences
Guido W Imbens and Donald B Rubin, 2015

A Conversation with Donald B. Rubin
Fan Li and Fabrizia Mealli
Statistical Science. 2014

The central role of the propensity score in observational studies for causal effects
Paul R Rosenbaum Donald B Rubin, Biometrika, 1983

Causal : Draw Your Assumptions Before Your Conclusions
A course by Miguel Hernán to *Learn simple graphical rules that allow you to use intuitive pictures to improve study design and data analysis for causal inference.*

Causal inference in statistics: a primer
Judea Pearl, Madelyin Glymour, Nicholas P Jewell, 2016

Collider bias undermines our understanding of COVID-19 disease risk and severity
Gareth J Griffith et al., Nature Communications. 2020

The Table 2 Fallacy: Presenting and Interpreting Confounder and Modifier Coefficients
Daniel Westreich and Sander Greenland, American Journal of Epidemiology. 2013

Use of directed acyclic graphs (DAGs) to identify confounders in applied health research: review and recommendations
Peter W G Tennant et al., International Journal of Epidemiology. 2020

Formulating causal questions and principled statistical answers
Els Goetghebeur et al., ArXiv preprint. 2020

Single World Intervention Graphs (SWIGs): A Unification of the Counterfactual and Graphical Approaches to Causality
Thomas S Richardson and James M Robins, 2013

Does obesity shorten life? The importance of well-defined interventions to answer causal questions
Miguel A Hernán & Sarah L Taubman, International Journal of Obesity, 2008

Does Obesity Shorten Life? Or is it the Soda? On Non-manipulable Causes
Judea Pearl, Journal of Causal Inference, 2018

Identifiability, exchangeability and confounding revisited
Sander Greenland and James M Robins, Epidemiologic Perspectives & Innovations, 2009

The role of exchangeability in causal inference
Olli Saarela, David A. Stephens and Erica E. M. Moodie, ArXiv preprint. 2020

Invited Commentary: Positivity in Practice
Daniel Westreich and Stephen R. Cole, American Journal of Epidemiology, 2010

Up next

Drug development, the ICH E9 addendum and causal inference

Björn Bornkamp, Novartis

Practical application with implementation details: Estimating the causal treatment effect in a subgroup defined by a post-baseline biomarker

Dominik Heinzmann, Roche, BBS board member

Thanks for your attention

Thanks to the working team: Björn Bornkamp, Dominik Heinzmann
...and to Amanda Ross, Marcel Wolbers and Heinz Schmidli for their helpful comments

a_randomwalker giusi.moffa@unibas.ch

Questions?