

Causal Learning for Data Science

Vanessa Didelez

with much help from Ronja Foraita and Christine W Bang

Leibniz Institute for Prevention Research & Epidemiology – BIPS

December 2021
Data Train — UBRA Bremen

Acknowledgements



Helpers have been and will be for the practicals:

Ronja Foraita



Christine W Bang



Many thanks to them!!

Overview of Course



Part 1: Causal Reasoning

— with directed acyclic graphs (DAGs)

Part 2: Estimation (Learning) of Causal Effects

Part 3: Causal Discovery

— finding (potential) causes

Note: all of this is a *subjective* selection of material based on what I **like** and **know** (though I try to cover a variety of topical material).

Aim of Course



- Introduce basic concepts of causal learning (reasoning, modelling & inference)
- ... to enable you to read more advanced 'causal' papers
- Focus on:
 - formulating causal (research) questions
 - some basic methods
 - understanding sources of (avoidable and unavoidable) bias
- Mix of mathematics & stories/examples

Questions?



If at any point anything is unclear, please feel free to interrupt and ask questions!

Causality in Data Science



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

Part 1: Causal Reasoning

Preamble



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

Preamble



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

Describe the ideal (hypothetical) experiment with which you could investigate your research question ⇒ Target Trial!

Or: describe the decision problem you would like to solve.

Research Questions



Descriptive / predictive:

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

Causal:

- (A) "Which type of anaesthetic should this patient receive to reduce the risk of complications during surgery?"
- (A') "How does the amount of anaesthetic affect the risk of complications during surgery?"
- (B) "What can be done to reduce the risk of complications during surgery for an average / a particular type of patient?"

Research Questions



Descriptive / predictive:

“Which type of client will buy which kind of product?”

Causal:

(A) “Should advert be at the top or bottom of website to increase the probability of viewing product?”

(A') “How does the size of advert affect the probability of viewing product?”

(B) “How can I get a client to buy my product?”

Research Questions



Descriptive / predictive:

“Who is most likely to become long-term unemployed?”

Causal:

- (A) “Will a minimum wage legislation increase the unemployment rate of a country?”
- (B) “What can be done to prevent someone from becoming unemployed?”

Research Questions



Type-A causal questions: **Causal Effects**

“what is the causal effect of a ‘treatment’*?”

“dose-response relation”

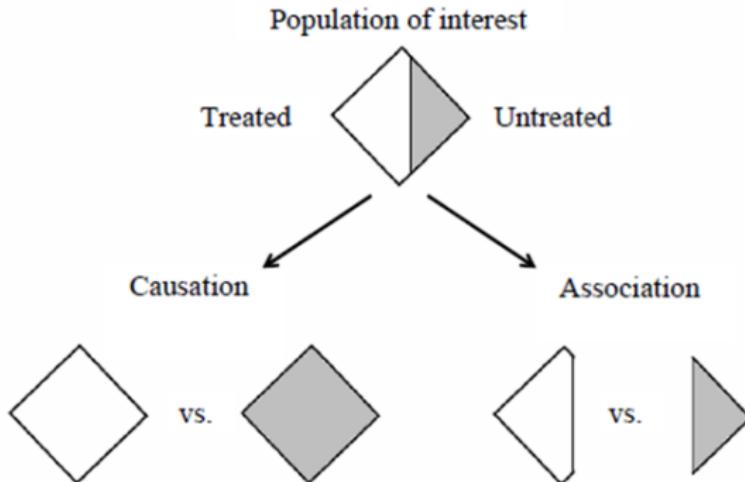
Type-B causal questions: **Causal Discovery**

“where can / should we possibly intervene?”

* Note: ‘treatment’ needs not be medical, could be: policy, teaching method, physical activity etc.

Causation versus Association

(Hernan & Robins, 2020 book)



Causal effect: contrast of outcome if 'everyone was treated' versus if 'no-one was treated'

Here: all models **probabilistic!**

Causal model:

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

... needs to be related to:

observational (no intervention / ‘natural’ / ‘idle’) situation
(distribution) generating our data

Identifiability:

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

Basic Concepts

Conditional Independence



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

Conditional independence:

X and Y are conditionally independent given Z ,
write $Y \perp\!\!\!\perp X | Z$, if

$$P(Y = y, X = x | Z = z) = P(Y = y | Z = z)P(X = x | Z = z)$$

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

$$P(Y = y | X = x, Z = z) = P(Y = y | Z = z)$$

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

Basic Concepts

Conditional Independence



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

Example:

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

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

Basic Causal Concepts



Formalisms to make interventions explicit:

do-notation / causal DAGs / decision theory

Potential Responses / counterfactuals

Structural equations / structural causal models: *not enough time to cover these*

do-Notation

(Pearl, 2000)



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

$$p(y \mid \text{intervene to set } X = x) = p(y \mid \mathbf{do}(X = x))$$

and

$$p(y \mid \text{observe } X = x) = p(y \mid \mathbf{see}(X = x))$$

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

Usually $p(y \mid \mathbf{see}(X = x)) = p(y \mid x)$

do-Intervention



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

Consider: Y, X, C_1, C_2 such that *observationally* ('see'):

$$p(y, x, c_1, c_2) = p(y|x, c_1, c_2)p(x|c_1, c_2)p(c_2|c_1)p(c_1)$$

May have reasons to believe that under intervention:

$$p(y, c_1, c_2 \mid \text{do}(X = \tilde{x})) = p(y|\tilde{x}, c_1, c_2)p(c_2|c_1)p(c_1).$$

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

Identifiability



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

$$p(y \mid \text{do}(X = x)) = \sum_c p(y \mid x, c)p(c)$$

left: interventional distribution; right: observational distrib.

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

Potential Responses (PRs)

(*Rubin, 1974; many others*)



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

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

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

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

\Rightarrow **potential** responses (or potential outcomes).

Potential Responses



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

$Y^i(x)$ = response if we **set** $X^i = x$

Counterfactuals



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

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

Potential Responses and ‘do’



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

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

Can regard $p(Y(x)) = p(y \mid \text{do}(X = x))$

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

⇒ Can express more (also more dubious) concepts with PRs

Causal Effects



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

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

Causal effects: typically formulated as contrasts of some aspect of

$$p(y \mid \text{do}(X = x)) \quad \text{versus} \quad p(y \mid \text{do}(X = x'))$$

or of $p(Y(x))$ versus $p(Y(x'))$

For instance: **Average Causal Effect**

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

Cause and Effect



Can now define:

X is a **cause** of Y and Y is an effect of X if for some $x \neq x'$

$$p(y \mid \text{do}(X = x)) \neq p(y \mid \text{do}(X = x'))$$

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

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

Counterfactual Prediction



Decision needed about treatment $X = 1$ or $X = 0$

Want to predict what happens with Y under either setting
 $X = 1$ or $X = 0$

⇒ Only one can be applied: counterfactual prediction.

Key Assumptions

for identifiability from observational data



Causal Consistency Assumption:

if we observe $X = x$ then $Y = Y(x)$

Positivity Assumption:

$$p(x \mid c) > 0 \text{ for all } x, c \quad (p(c) > 0)$$

Key Assumptions



Assumption of **no unmeasured confounding**:

(aka: random treatment assignment, or cond. exchangeability, ignorability, or ...)

Set C of observed (measured) **pre-treatment covariates** exists such that

$$Y(x) \perp\!\!\!\perp X \mid C$$

for all x to be considered as treatment values

Interpretation:

within values of C , can consider X like randomised wrt Y

Denote: C is **sufficient** to adjust (control) for confounding;

or ‘valid adjustment set’

No-Unmeasured-Conounding with $\text{do}(\cdot)$



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

$$p(y \mid c; \text{do}(X = x)) = p(y \mid c, x)$$

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

Identifiability Revisited



We consider $p(y \mid \text{do}(X = x))$ or equivalently $p(Y(x))$.

With the above assumptions:

$$p(Y(x)) \stackrel{(i)}{=} \sum_c p(Y(x)|c)p(c) \stackrel{(ii)}{=} \sum_c p(Y(x)|x,c)p(c)$$

$$\dots \stackrel{(iii)}{=} \sum_c p(y|x,c)p(c)$$

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

Adjustment / Standardisation



Consider the above result

$$p(y \mid \text{do}(X = x)) = \sum_c p(y \mid x, c)p(c)$$

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

Confounding



Above: confounding is present if

$$Y(x) \not\perp\!\!\!\perp X$$

or if $p(y \mid \text{do}(X = x)) \neq p(y \mid X = x)$

Usually:

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

⇒ Use causal DAGs to clarify!

Quiz



Possible break for quiz.

Causal Directed Acyclic Graphs (DAGs)

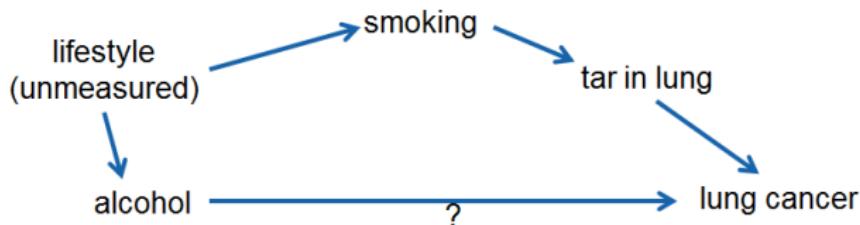
Graphs — Terminology

Graph $G = (V, E)$

V = vertices 0 nodes = variables / features

E = edges = possible (causal) dependence

Non-edge = known (conditional) independence



Note: nodes shown as ‘events’ represent binary indicator variables, e.g. ‘lung cancer’ $\in \{0, 1\}$ for ‘no’ / ‘yes’.

(Causal) Graphs

aka: (causal) DAGs / diagrams / Bayesian networks



A causal graph is a (probabilistic) model for a set of random variables imposing

- restrictions on conditional independencies within the **observational distribution**

and
- restrictions on conditional independencies within the distribution under **hypothetical interventions**
- ‘non-parametric’: graph contains no information on the functional shape of relations between variables (nor on strength / size of dependencies)

Why Graphs?



Make explicit: underlying assumptions & required background knowledge!

Graphs: one way to *represent* & *organise* assumptions / prior knowledge

Here: will focus on bias sources related to

- confounding
- selection

Graphs for Identifiability



- Can we identify causal effects from observational data?
- ... for what do we need to adjust?
- ... for what must we not adjust?

Nodes / Variables



The typical / traditional approach assumes one **already** has access to **variables which represent high-level semantic concepts**

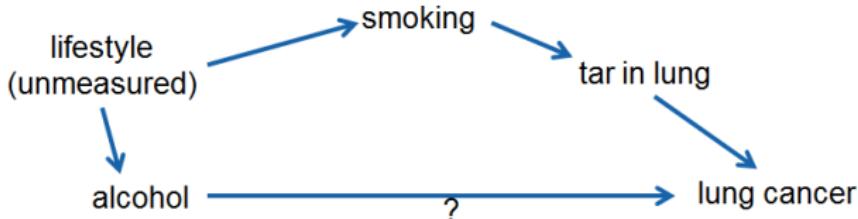
This may not be the case when learning from raw video or imaging data, for example

⇒ Formulating causal DAG for such situations: active research!

Graphs — Terminology

Graphical terms:

'parents', 'children', 'ancestors', '(non-)descendants' etc.
'(directed) paths', '(directed) cycles'



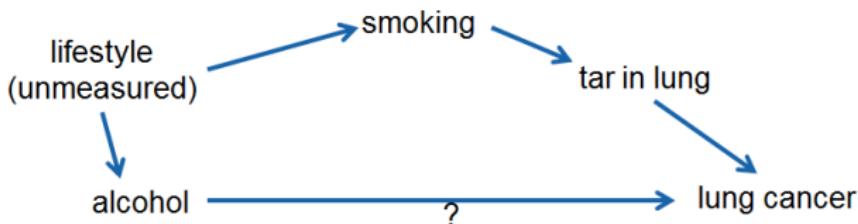
Graphs — Markov Property

with DAGitty



Observationally:

Absence of edges into outcome: if we know whether there is tar in the lungs and whether person drinks alcohol, then smoking status or any further information on lifestyle are non-informative for the probability of lungcancer.

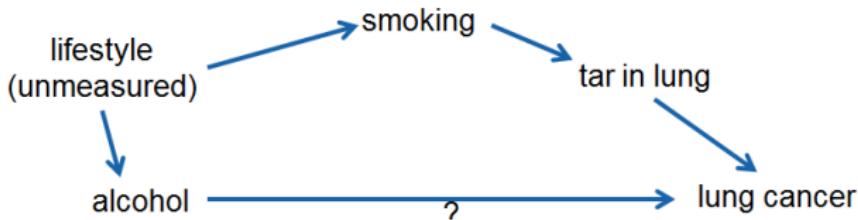


Check with DAGitty, software for querying DAGs (Textor et al, 2016)!

Causal Markov Condition

Causally:

Edge, e.g., if we control ‘tar’ by an intervention and vary ‘alcohol’ by an intervention then this will possibly change the probability for ‘lung cancer’
⇒ an edge represents a possible ‘controlled direct effect’



Notes:

- ‘direct effect’ relative to nodes included
- better: **absence** of edge guarantees **no direct effect**

Direct Causal Effect



Notion of **Controlled direct effect**:

For other parent node(s) Z of Y , what is the effect of intervening in X on Y while **fixing Z** by intervention?

Principle: block certain causal pathways by **fixing Z** ; then assess remaining effect of X on Y .

⇒ contrast of

$$p(y \mid \text{do}(X = x, \mathbf{Z} = \mathbf{z})) \quad \text{versus} \quad p(y \mid \text{do}(\mathbf{X} = \mathbf{x}', \mathbf{Z} = \mathbf{z}))$$

Causal Markov Condition



Axiom (Causal Markov Condition):

if neither X *direct* cause of Y nor vice versa

\Rightarrow there exists a set S s.t. $X \perp\!\!\!\perp Y | S$

('direct' relative to other nodes)

Graphical: every variable is cond. independent of its non-effects (descendants) given its direct causes (parents).

Factorisation for DAGs



Factorisation: a distribution P (with pdf/pmf p) factorises according to a DAG G and is called **G–Markov** iff

$$p(\mathbf{x}) = \prod_{i=1}^K p(x_i | \mathbf{x}_{\text{pa}(i)})$$

Note: the above factorisation is **equivalent to**

$$X_i \perp\!\!\!\perp \mathbf{X}_{\text{nd}(i) \setminus \text{pa}(i)} \mid \mathbf{X}_{\text{pa}(i)} \text{ for every } i \in V$$

Rule: read off cond. independencies using **d-separation** (later)

⇒ testable implications of DAG models!

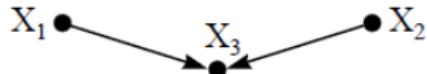
Selection Effect

“collider bias”



Important for the interpretation:

Conditioning on common child (**selection**) \Rightarrow dependence



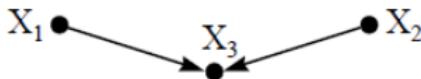
here: $X_1 \perp\!\!\!\perp X_2$ but $X_1 \not\perp\!\!\!\perp X_2 \mid X_3$

$$p(x_1, x_2, x_3) = p(x_1)p(x_2)p(x_3|x_1, x_2)$$

does not generally imply $X_1 \perp\!\!\!\perp X_2 \mid X_3$

Selection Effect

“collider bias”



Example: some school admission process is such that pupils are admitted (X_3) if they are either good at maths (X_1) or good at sports (X_2).

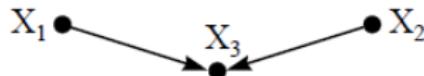
Assume in population X_1 and X_2 are independent(!)

If we randomly draw a pupil from this school, $X_3 = 1$, and find this pupil is no good at sports, $X_2 = 0$, then we know s/he must be good at maths, $X_1 = 1$!

In other words, given X_3 , X_2 becomes informative for X_1 .

Separation in DAGs

Motivated by selection effect: want general rule to describe
“separation”



Here: \emptyset separates X_1 and X_2

but X_3 does not separate X_1 and X_2 .

d-Separation in DAGs

(Pearl, 1988)

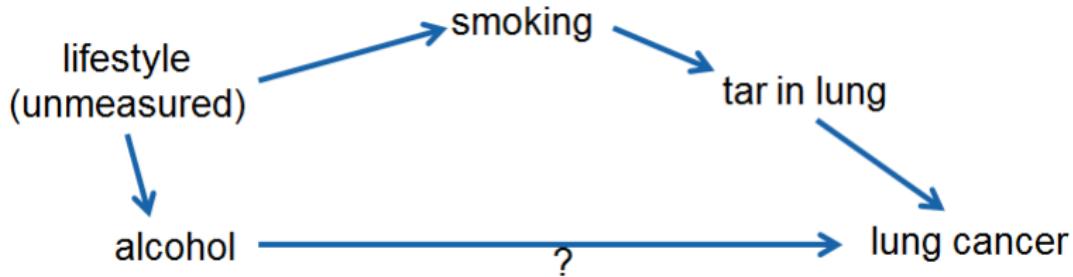


Given DAG $G = (V, E)$. A path between a and $b \in V$ is **blocked** by $S \subset V \setminus \{a, b\}$ if

- (i) it contains a non-collider $\leftarrow z \rightarrow$ or $\leftarrow z \leftarrow$ and $z \in S$ or
- (ii) it contains a collider $\rightarrow z \leftarrow$ and **neither z nor** any descendants of z are elements of S

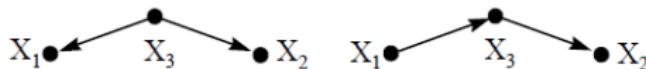
A and $B \subset V$ are **d-separated** by $S \subset V \setminus (A \cup B)$ if every path between A and B is blocked by S .

d-Separation — Quiz



Markov Equivalence of DAGs

Marginalizing w.r.t. common parent (**confounder**) or intermediate variables \Rightarrow dependence



Here: $X_1 \perp\!\!\!\perp X_2 \mid X_3$, but $X_1 \not\perp\!\!\!\perp X_2$

Markov equivalence:

different DAGs imply same conditional independencies!

Implication:

cannot distinguish between equivalent DAGs from data alone.

Causal DAG



So what makes a DAG into a **causal DAG**?

Additional **semantics** relating DAG to interventions:

- effects of **interventions follow direction** of edges, i.e. can affect all descendants, but cannot affect non-descendants
⇒ DAGitty depicts ‘causal paths’ and ‘non-causal’ paths inducing associations
- **intervention distribution** corresponds to DAG-model after **removing edges** into the intervened node.

Example 1

$$X \longrightarrow Y$$

This causal DAG expresses:

- an intervention in X can affect Y
- an intervention in Y *cannot* affect X

Note: The DAG expresses no (cond.) independencies.

Example 1 ctd.



$$\text{do}(X=x) \longrightarrow Y$$

Moreover:

- an intervention in X removes arrows into X (here: none)
- the intervention distribution is identical to the (observational) conditional distribution

$$p(y | \text{do}(X = x)) = p(y | x)$$

Note: the latter reflects that the DAG expresses the assumption of no common causes for X and Y .

This would be plausible if X was known to be randomised.

Example 1 ctd.



$$X \qquad \text{do}(Y=y)$$

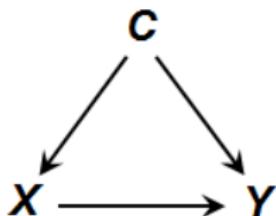
Finally:

- an intervention in Y removes arrows into Y
- the intervention distribution is identical to the (observational) marginal distribution

$$p(x \mid \text{do}(Y = y)) = p(x)$$

- i.e. X is independent of Y under an intervention in Y .

Example 2



This causal DAG expresses:

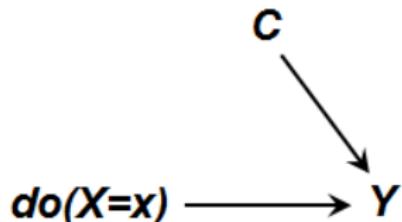
- an intervention in X can affect Y , *but not C*
- an intervention in C can affect X and Y
- an intervention in Y *cannot* affect X nor C .

Note: The DAG expresses no (cond.) independencies.

Example 2 ctd.

Moreover:

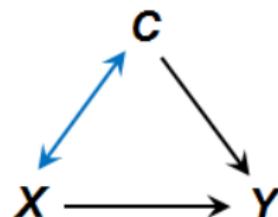
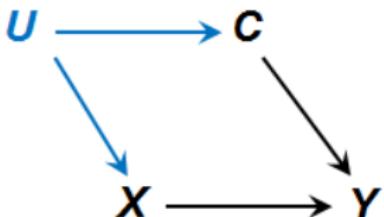
- an intervention in X removes arrows into X
- the intervention distribution is identical to the (observational) conditional distribution $p(y, c | \text{do}(X = x)) = p(y | c, x)p(c)$ and hence (standardisation again!)



$$p(y | \text{do}(X = x)) = \sum_c p(y | c, x)p(c)$$

Note: because of the **assumption of a common cause** C , the formula for $p(y | \text{do}(X = x))$ is now different than in Example 1.

Example 3



Assume U unobserved (often represented by bi-directed edge)

- X and C are not independent (due to common cause U)
- but intervention in X does not affect C and intervention in C does not affect X
- otherwise, regarding X, C, Y same as Example 2.

Example 4



This causal DAG expresses:

- an intervention in X can affect Z and Y
- an intervention in Z can affect Y , but not X
- an intervention in Y cannot affect X nor Z

Example 4 ctd.

$$X \quad do(Z=z) \longrightarrow Y$$

Moreover:

- an intervention in Z prevents an intervention in X having any effect on Y
- ⇒ relative to the considered set of variables:
 Z is a direct cause of Y , X is an indirect cause of Y
- ⇒ the direct effect of X on Y controlling for Z is null.

Causal DAG

(for the mathematically interested)



Definition:

DAG G , distribution P is G -Markov. Then, G causal wrt $B \subset V$ if for any $A \subset B$

$$p(\mathbf{x}_V \mid \text{do}(A = a)) = \prod_{i \in V \setminus A} p(x_i \mid \mathbf{x}_{\text{pa}(i)}) \Big|_{\mathbf{x}_A = a}$$

in words:

- P describes ‘behaviour’ under observation, factorises
- under intervention, $\text{do}(A = a)$, the variables in \mathbf{X}_A are simply **fixed to a** when appearing in $\mathbf{x}_{\text{pa}(i)}$
- and all **conditional specifications** on $V \setminus A$ **remain the same** (‘invariance’)

Use of causal DAGs



“Draw your assumptions before your conclusions!”

(Hernán)

- Make explicit your assumptions \Rightarrow draw DAG based on background knowledge (good thing in any case)
- Check if some implications of DAG can be verified empirically, e.g. implied conditional independencies
- Check if your desired target can be identified with the observable data
- Possibly: motivate sensitivity analysis by different competing causal DAGs reflecting uncertainty in subject matter knowledge.

Standardisation



Remember: identifying functional for the effect of X on Y

$$p(y \mid \text{do}(X = x)) = \sum_c p(y \mid x, c)p(c)$$

Requires assumption ‘no-unmeasured confounding given C ’.

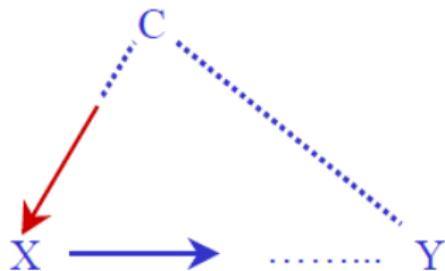
Graphical formulation:

C must ‘block all back-door paths’ from X to Y ...

Back-Door Path

Definition

A back-door path from X to Y starts with an edge $X \leftarrow \dots Y$.



Back-Door Criterion

(Pearl, 1995)



Theorem

Given a DAG G on V , causal wrt. $X \in V$. Then $C \subset V \setminus \{X, Y\}$ identifies causal effect of X on Y if

- (i) C is non-descendant of X and
- (ii) all 'back-door' paths from X to Y are blocked by C

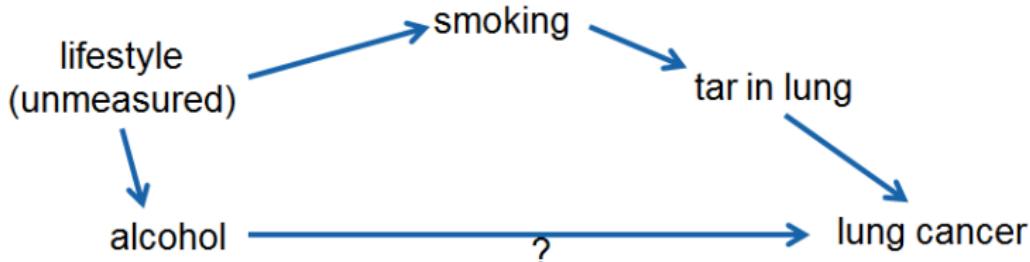
C is then *sufficient* adjustment set.

Note: C not unique; *minimal* C not unique.

Back-door Criterion — Exercise

with DAGitty

Note: lifestyle is *the* confounder (common cause), but unobserved!



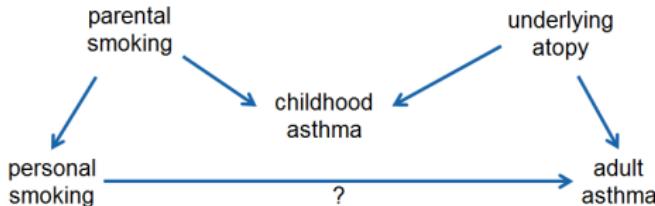
Sufficient set of covariates to identify the effect of X on Y ?

$X = \text{alcohol consumption}$, $Y = \text{lung cancer}$

M-Bias

(more generally: *collider-Bias*)

Example (simplified from Williamson et al., 2014): want effect of smoking on adult asthma; know that childhood asthma is associated with smoking and with adult asthma.
Is “childhood asthma” sufficient to adjust for confounding?



Note: it is impossible to define or empirically check for ‘confounding’ in terms of associations!

Always need prior structural knowledge.

Back-door Criterion



How can we use the Back-door Criterion in practice?

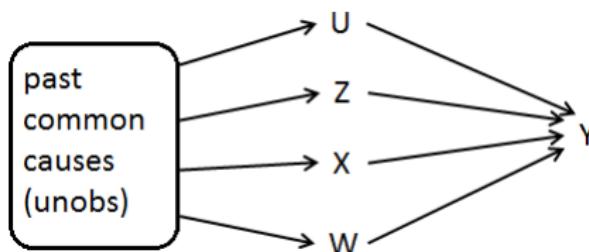
- Construct the DAG based on knowledge of
 - subject matter (basic biology etc.)
 - temporal ordering
 - study design
 - statistical evidence
 - justify all missing edges and absence of further hidden variables (i.e. include all common causes)

⇒ Causal DAG will typically **include unobservable** variables!

- check for which choice of C (if any) properties (i) and (ii) of Theorem hold → check for separations

Association due to Past

Common situation might be: associations between exposure X and other covariates are due to common past history, e.g. past life-style / disease process etc.



⇒ need all of U, Z, W to identify effect of X on Y .

Question:

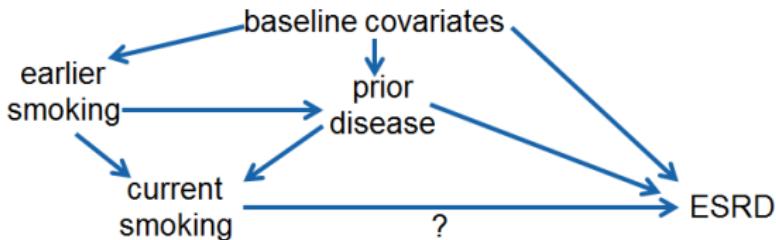
what happens if W and Y affected by unobserved factor?

Further Examples

do this with DAGitty !

Wanted: effect of current smoking on end-stage renal disease (ESRD) (Staplin et al., 2016)

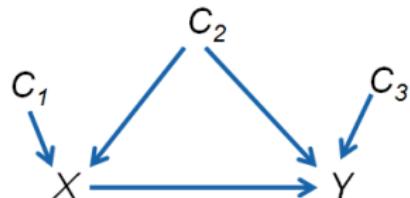
No data available on ‘earlier smoking’ – is this a problem?



Question: what if ‘prior disease’ and ESDR affected by further unobserved factors?

Further Examples

Some covariates are unnecessary:
here C_1, C_3 not required to
adjust for confounding,
 C_2 is sufficient.



But: while it can improve efficiency to include C_3 as additional predictor of outcome Y , it can be inefficient and even harmful to include C_1 ...

Bias amplification: can show that if there is some small residual unobserved confounding (e.g. C_2 measured with error), then including variables like C_1 will increase the bias.

Confounding

some misconceptions



- Confounding is a causal concept
- ...a definition of confounding in terms of associations is impossible (wrong in many textbooks)
- ‘associations’ cannot be confounded, only causal relations can be confounded
- notion of ‘confounder’ problematic — often better: ‘deconfounder’ = variables that are useful for reducing bias

Selection Effect (or Bias)



Traditional meaning

Potential to induce bias regarding causal inference through the way how the sample is selected.

Formally

Assume causal effect identified from marginal (observational) distribution of (X, Y, C) , then selection effect occurs if it is not necessarily identified from $(X, Y, C | Sel = 1)$ (i.e. given selection).

More general meaning

Some form of *collider-bias*: potential to induce bias regarding causal inference by *conditioning / stratifying* on covariates
≈ opposite of confounding.

Selection Effect — Graphically



Let DAG represent background knowledge on conditional independencies and *causal order wrt. X* .

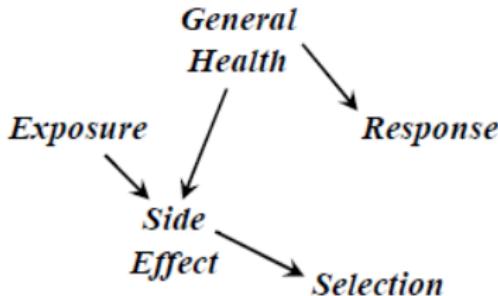
i.e. variables known not to be affected by an intervention in X must not be descendants of X .

Assume set of covariates sufficient to adjust for confounding.

Trick: draw graph under null-hypothesis of no causal effect
⇒ check if exposure $\perp\!\!\!\perp$ response | (selection, covariates)

If above check fails, then inference will typically be biased (even if there is a causal effect, i.e. not under null).

Graphical Check — Exercise



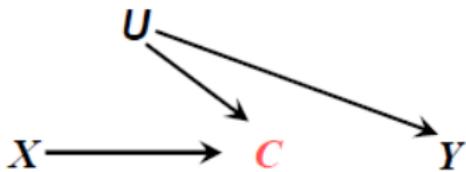
Let X = exposure, Y = response, E = side effect, S = selection (patients with bad side effects drop out of the study).

Exercise: Can we test the null-hypothesis of no causal effect from the patients remaining in the study?

Post-Treatment Covariates

If C is post-treatment covariate (e.g. liver function after treatment) we typically do not adjust for it as we may find $Y \perp\!\!\!\perp X|C$ even when X has a causal effect (but mediated by C). But often done to find the ‘direct effect’ of X on Y .

Less well known: This can lead to $Y \not\perp\!\!\!\perp X|C$ even when X has no causal effect (direct or indirect) on Y ! See DAG below...



Selection Bias in COVID Research?



nature paper found ‘protective effect’ of smoking on COVID-19 death

Article

Factors associated with COVID-19-related death using OpenSAFELY

<https://doi.org/10.1038/s41586-020-2521-4>

Received: 15 May 2020

Accepted: 1 July 2020

Published online: 8 July 2020

Check for updates

Elizabeth J. Williamson^{1,2}, Alex J. Walker^{2,3}, Krishnan Bhaskaran^{1,6}, Seb Bacon^{2,4}, Chris Bates^{2,5}, Caroline E. Morton², Helen J. Curtis⁵, Amir Mehrkar², David Evans², Peter Inglesby², Jonathan Cockburn², Helen I. McDonald^{1,4}, Brian MacKenna³, Laurie Tomlinson¹, Ian J. Douglas¹, Christopher T. Rentsch¹, Rohini Mathur¹, Angel Y. S. Wong¹, Richard Grieve¹, David Harrison¹, Harriet Forbes¹, Anna Schultze¹, Sam Harper², Rafael Perera², Stephen J. W. Evan

Coronavirus disease 2019 (COVID-19) has raised unprecedented urgency to understand what

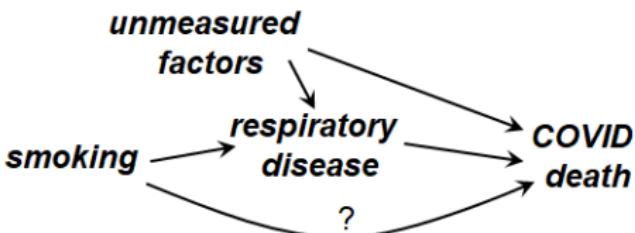


Table-2 Fallacy?



American Journal of Epidemiology
© The Author 2013. Published by Oxford University Press on behalf of the Johns Hopkins Bloomberg School of Public Health. All rights reserved. For permissions, please e-mail: journals.permissions@oup.com.

Vol. 177, No. 4
DOI: 10.1093/aje/kws412
Advance Access publication:
January 30, 2013

Commentary

The Table 2 Fallacy: Presenting and Interpreting Confounder and Modifier Coefficients

Daniel Westreich* and Sander Greenland

* Correspondence to Dr. Daniel Westreich, Department of Obstetrics and Gynecology, Duke Global Health Institute, Duke University, DUMC 3967, Durham, NC 27710 (e-mail: daniel.westreich@duke.edu).

Initially submitted January 13, 2012; accepted for publication October 11, 2012.

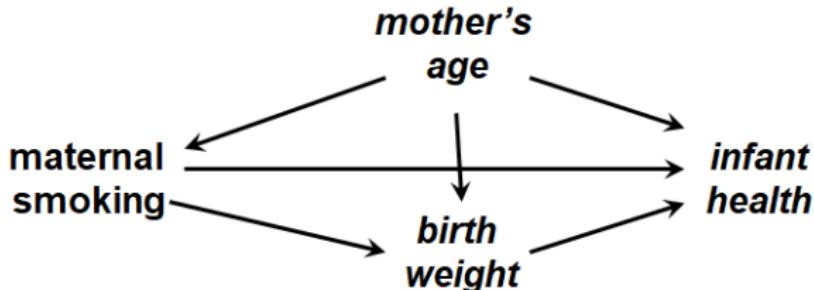
It is common to present multiple adjusted effect estimates from a single model in a single table. For example, a table might show odds ratios for one or more exposures and also for several confounders from a single logistic regression. This can lead to mistaken interpretations of these estimates. We use causal diagrams to display the sources of the problems. Presentation of exposure and confounder effect estimates from a single model may lead to several interpretative difficulties, inviting confusion of direct-effect estimates with total-effect estimates for covariates in the model. These effect estimates may also be confounded even though the effect estimate for the main exposure is not confounded. Interpretation of these effect estimates is further complicated by heterogeneity (variation, modification) of the exposure effect measure across covariate levels. We offer suggestions to limit potential misunderstandings when multiple effect estimates are presented, including precise distinction between total and direct effect measures from a single model, and use of multiple models tailored to yield total-effect estimates for covariates.

causal diagrams; causal inference; confounding; direct effects; epidemiologic methods; mediation analysis;

Prediction — Causation

Prediction of infant health: use *all* available information

Causal effect of maternal smoking on infant health: ignore birth-weight



Selection Effect in Longitudinal / Duration Studies



Problem

more potential for selection effect by inadvertently conditioning on information that occurs later in time.

Chance

time ordering is explicit and potential for selection effect easier to detect.

If time: simulated example

Causal DAG Construction?



- Domain knowledge (check literature etc.) — talk a lot with subject matter experts!
- Include relevant unmeasured nodes (common causes) & justify absence of further edges and further nodes
- Can empirically assess *some* cond. indep. implications but key assumption of no unmeasured confounding cannot be tested...
- **Can do sensitivity analyses with multiple DAGs if uncertain!**

Causal DAGs

Summary



- Graphs are helpful to organise your causal reasoning / structuring of a given causal question with data at hand.
 - Main purpose: can the causal effect be identified from the available data in the first place? Can we test for causal effect? Can we estimate the causal effect?
 - Confounding: which covariates do we have to take into account? ⇒ Back-door criterion.
 - Selection- / collider-bias: which covariates should we not condition on?
- ⇒ **Recommended:** always *draw your assumptions before your conclusions!* (Hernán)

Further Topics



- **Software:** DAGitty — R package or online.
Carries out queries on DAGs, e.g. find all minimal sufficient adjustment sets.
- Other identification criteria exist: e.g. Front-door criterion.
Complete identification algorithm due to Shpitser (2006)
available in software `ananke` (Python)
- Causal DAGs also used for:
 - decide transportability of inference across populations
 - identifiability with missing values
 - expert systems etc.

Further Topics

Appendix



- Workflow of causal analysis?
- Single world intervention graphs (SWIGs)
link between potential responses and graphs
- Alternative (niche): influence diagrams
- Structural equation models → impose most structure
- Other interventions: nudging / shifting / stochastic
interventions — active research