UiO : Universitetet i Oslo

Directed Acyclic Graphs

Cecilie Dahl
Researcher
Department of Community Medicine and Global Health
University of Oslo
cecilie.dahl@medisin.uio.no

Outline- 3 hour session

- Why do we need to use DAGs in epidemiology?
- Short- Need to know
- Analyzing DAGs (with exercises and discussion)
- Daggity- the «easy way» to draw and analyze a DAG

- Practice drawing your own DAGs, discussion

WHY DO WE NEED TO USE DAGS?

Why causal graphs (DAGs)?

- Estimate effect of exposure on disease (causal relation)
- Problem
  – Association measures are biased
- DAGs help in :
  – Understanding
    - Confounding, selection bias, mediation
  – Analysis
    - Adjust or not
  – Discussion
    - Precise statement of prior assumptions

Criteria for a variable to be a confounder:

1. Associated with the exposure in the source population

2. Associated with the outcome (also in the unexposed)

3. Not on the causal pathway between exposure and outcome

Criteria for a variable to be a confounde

1. Associated with the exposure in the ... pulation

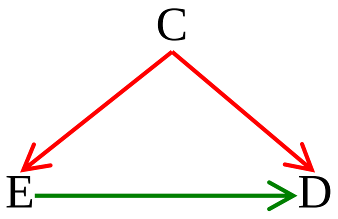
2. Associated with the outcom... ...e unexposed)

3. Not on the causal pa... ...tween exposure and outcome

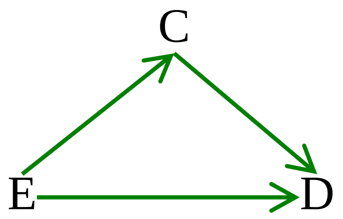
Not a sufficient definition

Adjust for C: cause versus association

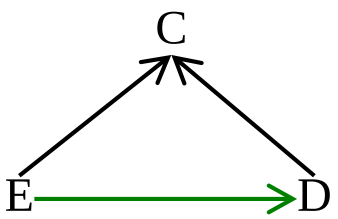
Cause:



Confounder

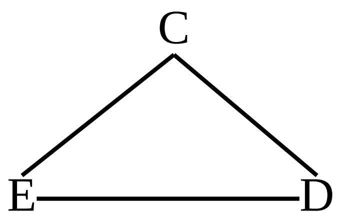


Mediator



Collider

Association:



Statistical criteria:

likelihood ratio, AIC, 10% change in estimate

cannot differentiate between

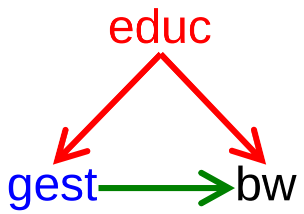
Confounder, Mediator or Collider

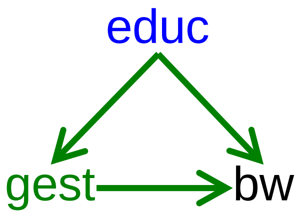
Need causal model to do a proper analysis

Table 2 fallacy, gestation age and birth weight

- Pre DAGs: report all covariate effects from one model
- Post DAGs:
  - report only exposure effect
  - separate models for other covariates

	Model 1	Model 2
Exposure:	gest	educ





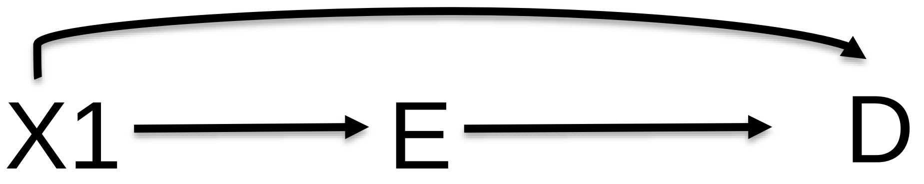
Variable	model 1	model 2
gest	202	
educ	~~113~~	278
	educ confounder adjust	gest mediator not adjust

(Westreich and Greenland 2013)

SHORT- NEED TO KNOW

Time

- Time flows from left to right, and thus X1 is temporally prior to E and D , and E is temporally prior to D

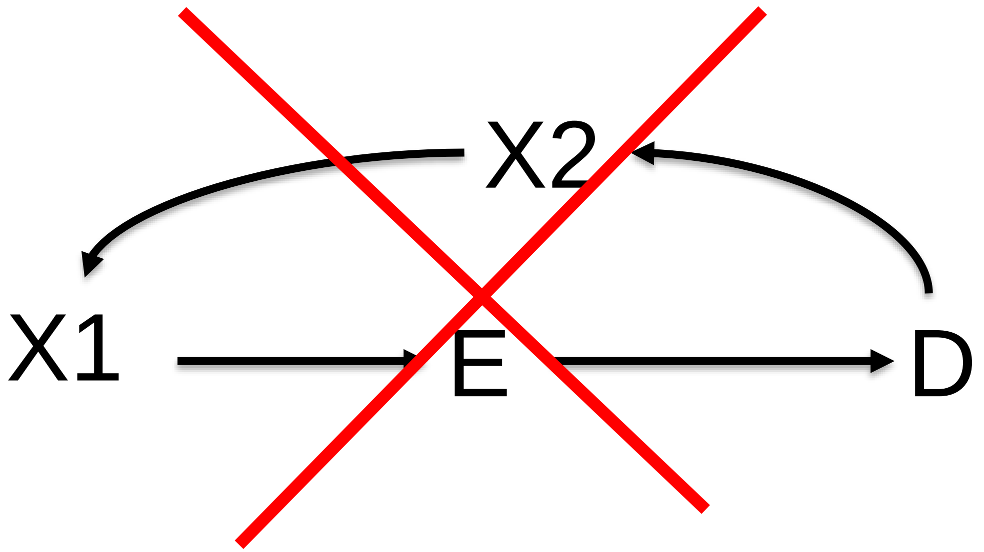


Parent → Child

Ancestor → Descendant

Acyclic

- Arrows in series cannot lead back to the same node (ACYCLIC)



Components

- Variables represented as «nodes» or «vertices»

- Arrows (directed edges) between variables represent causal effects
  - Depicting the *existence*, but *not the strenght* of causal relationships (nor whether it is positive or negative)
  - Causation vs. Association paths

- Omission of an arrow is a stronger claim than the inclusion of an arrow

Causation vs. Associtition paths

E and D are associated through X1

X1 is a cause of E and D

Causal DAG?

- Is the DAG a «complete view of the causal structure of reality»
  - – Did we include all unmeasured varibles (U)?
    - A DAG should also include these
  - – Did we include all common ancestors of two variables?
  - – Did we include all the arrows?
    - Absense of an arrow is a strong statement

ANALYZING A DAG

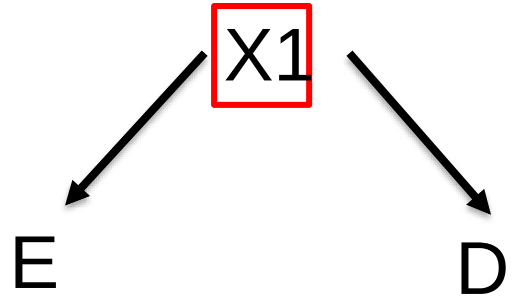
3 reasons why two variables may be associated

Non-comparability (bias of the total effect)

1. They share common causes (confounding)
2. They share a common effect and the analysis is restricted to certain level of that common effect (or of its descendants)
3. One causes the other (directly or indirectly)

1. Conditioning on a common cause

- Confounding



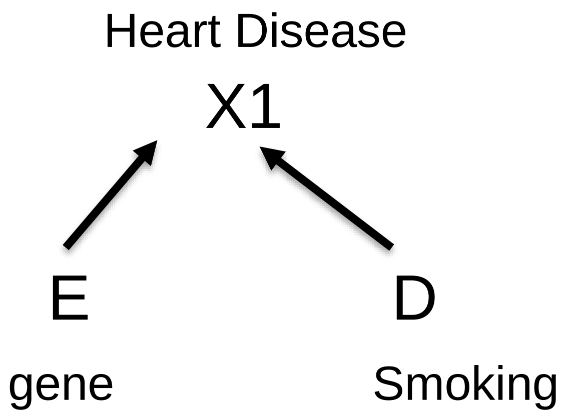
Unconditioned- Information on E could give information on D through X1

Conditioning- Restricting X1=1 blocks all variation in E, i.e. variation in E gives no information in D

Example: Carrying a lighter (E) is associated with lung cancer (D). Smoking status (X1) is the underlying cause of both

2. Conditioning on a common effect

No bias



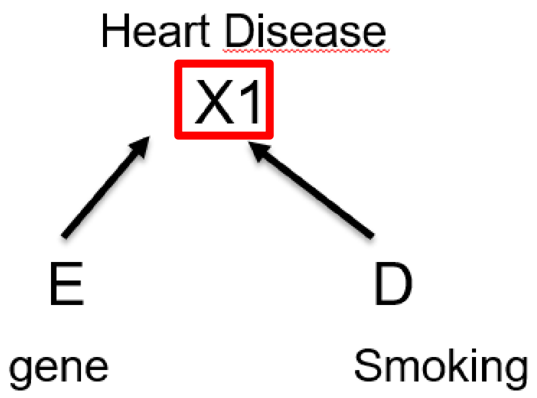
Information on E does not give information on D.

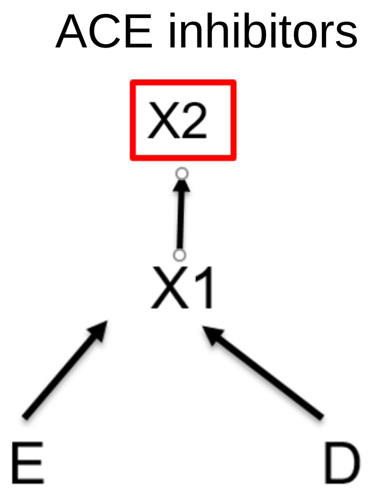
Example: A gene (E) and smoking status (D) are independent causes of hearth disease (X1)

2. Conditioning on a common effect

- Selection bias
- Collider-stratification bias
- Bias due to conditioning on a collider

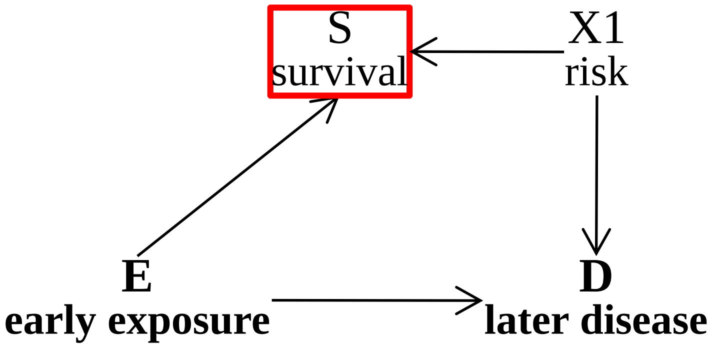
Bias!





Conditioning on X1=1 or X1=0: Information on E does give information on D.
- If E and D are the only two causes of X1, conditioning on X1=1 means that if and E=0, then D has to be=1

Survival bias (a type of selection bias)



Paths:

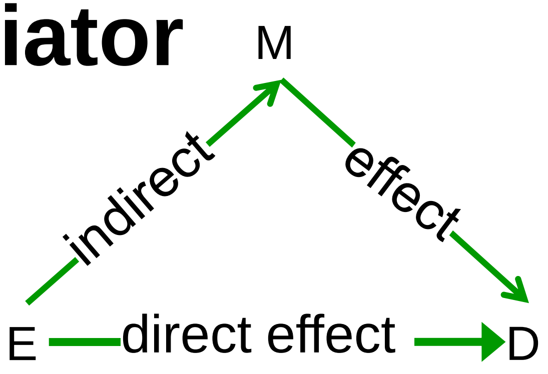
E→D	Causal	Open
E→[S]←X1→D	Non-causal	Open

Conclusion:

Have survival bias

Must adjust for X1 to remove the bias

3. Conditioning on a mediator



- Have found a cause (E)
- How does it work?
  - Mediator (M)
  - Paths

$$Total\ effect = indirect + direct$$

$$Mediated\ proportion = \frac{indirect}{total}$$

Use ordinary regression methods (with and without M) if:
no E-M interaction and collapsible measures*.
Otherwise, need new methods

Strong conditions of non-confounding

*Risk difference and risk ratio are collapsible
Odds ratio and rate ratio are collapsible if rare disease

(VanderWeele 2016)

How to identify and block backdoor (biasing) paths

- Pay attention to the direction of the arrow (causal path, don’t want to condition on this)

- 4 possible permutations of arrows around a variable (node)

$\longrightarrow X1 \longrightarrow \qquad \longleftarrow X1 \longleftarrow \qquad \longleftarrow X1 \longrightarrow \qquad \longrightarrow X1 \longleftarrow$

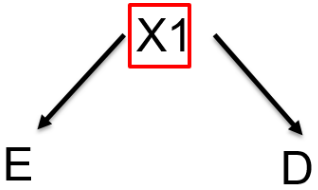
- Are there any existing natural blocks?

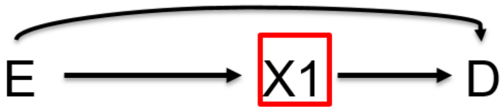
$\perp$ = independent

D- separation

- Conditional indepencence (epidemiology)
- Simply means that one variable (E) is independent of another variable (D), given another variable (X) or a set of other variables(X1-X_t).

$$E \perp D | X1$$





- The opposite is «d-connected»

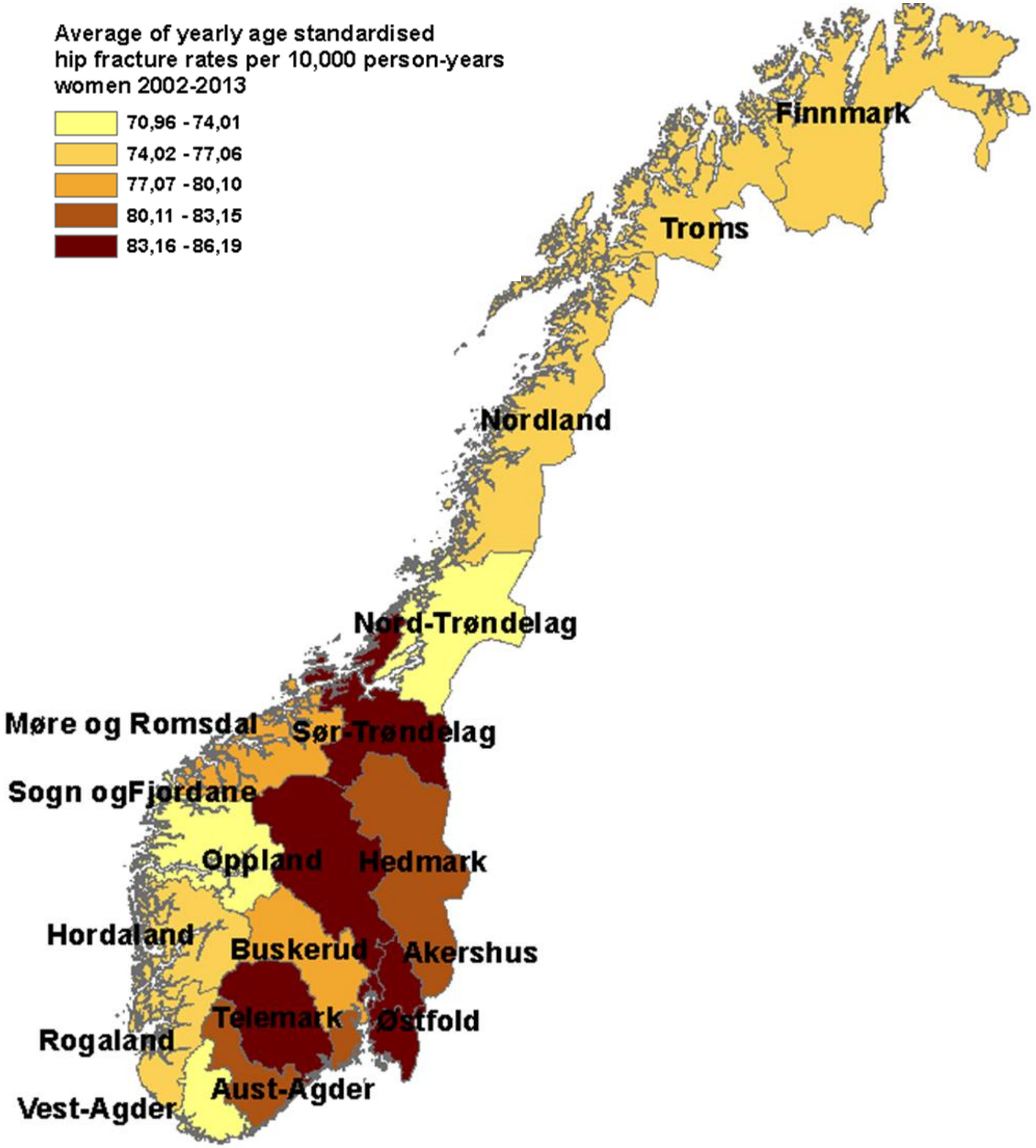
Nodes

Open (E and D are «d-connected»)	Blocked (E and D are d-separated)		Action (total effect)
$E \longrightarrow X1 \longrightarrow D$	$E \longrightarrow \boxed{X1} \longrightarrow D$	Causal	Do not adjust for X1
$E \longleftarrow X1 \longrightarrow D$	$E \longleftarrow \boxed{X1} \longrightarrow D$	Non-causal	Adjust for X1
$E \longleftarrow X1 \longleftarrow D$	$E \longleftarrow \boxed{X1} \longleftarrow D$	Causal (reverse)	Switch E and D
$E \longrightarrow \boxed{X1} \longleftarrow D$	$E \longrightarrow X1 \longleftarrow D$	Non-causal	Do not adjust for X1

Goal:
Want to keep all causal pathways open
Close or keep closed the non-causal pathways (backdoor paths)

Example: Outside temperature and the risk of bone fracture in older adults

- Bone fracture incidence has been found to vary by season in Norway and in other countries, higher incidence in wintertime vs. summertime
- There is also a variation by latitude, in general a higher incidence at higher latitudes (in Sweden and in the rest of Europe), but no clear variation by latitude in Norway
- In Norway we see a higher incidence inland compared to the coast
- Can outside temperature be the «causal factor» for this pattern?



Forsén et al (unpublished)

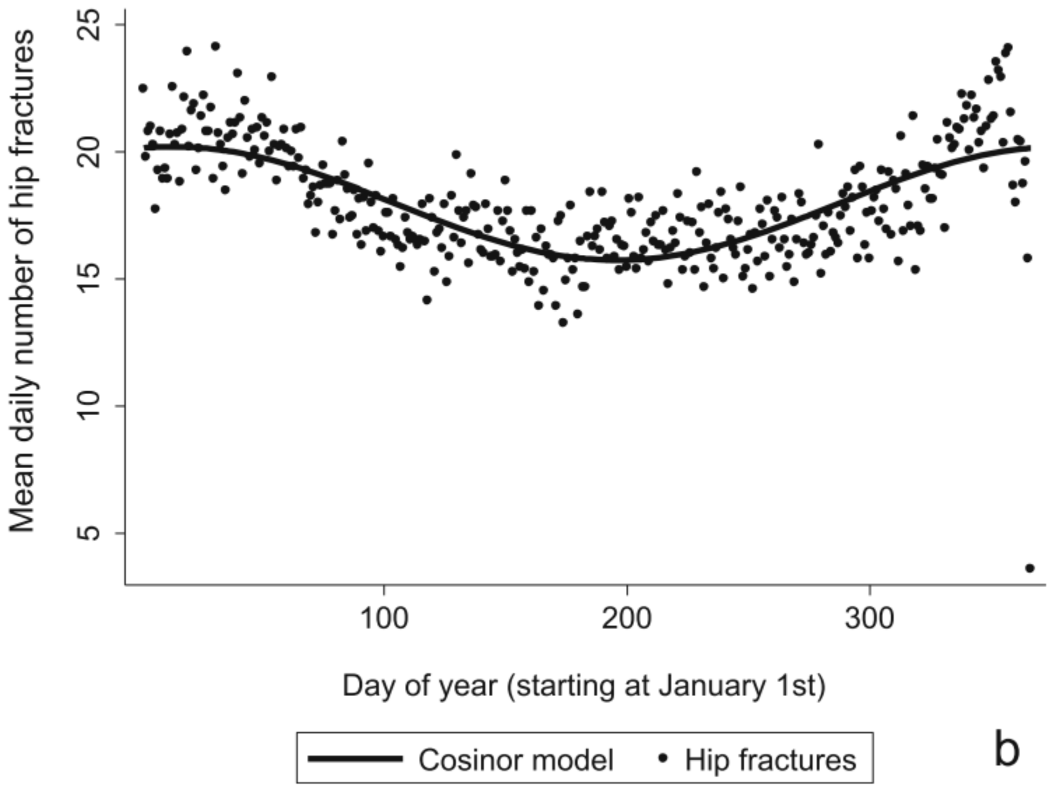
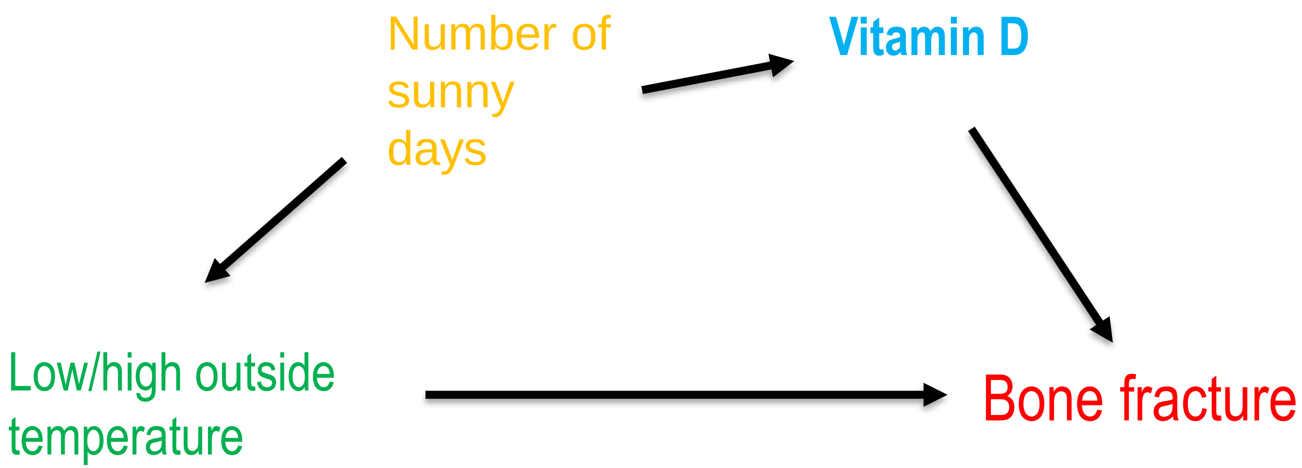


Fig. 2 Mean daily number of hip fractures during 1994–2008 (Norwegian patients aged 50–103 years) in **a** men and **b** women. A time series model (Cosinor model) is fitted. Norwegian Epidemiologic Osteoporosis Studies (NOREPOS)

(Solbakken et al. 2014)

Example: Outside temperature and the risk of bone fracture in older adults



Urbanization degree

Air pollution

Number of sunny days

Season

Latitude

Distance to coast

Relocation

Vitamin D

Low/high outside temperature

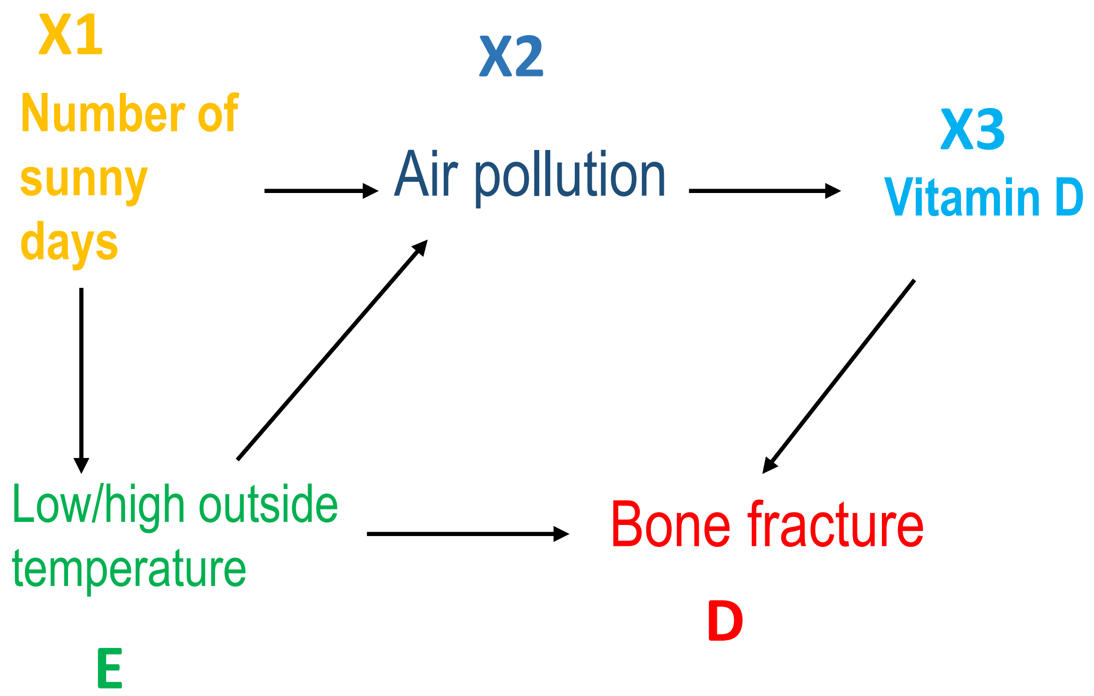
Bone mineral density (BMD)

Body Mass Index

Ethnicity

Socioeconomic status (SES)

Exercise 1a: Temperature and the risk of bone fracture in older adults



Low temperature= Mean yearly temp<10°C
High temperature=Mean yearly temp >10°C

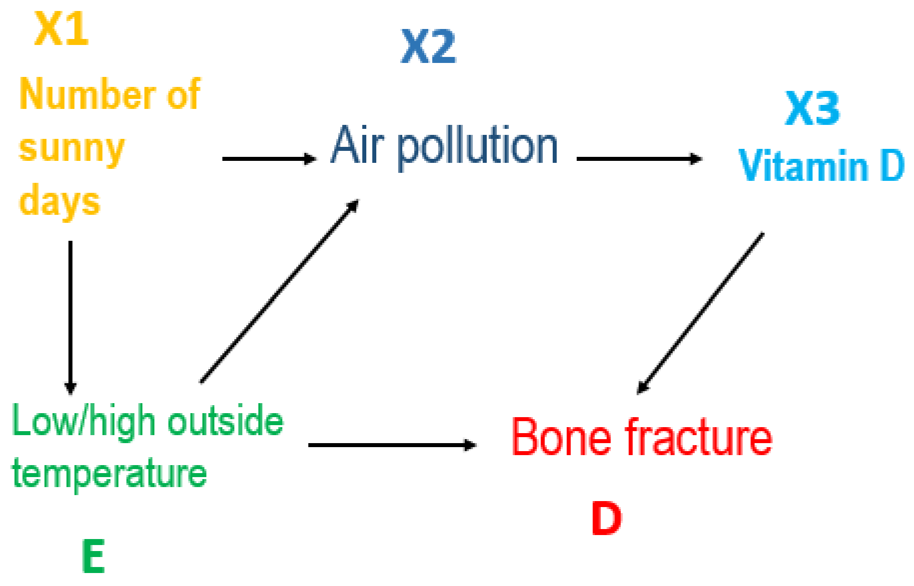
1. Write down the paths
2. Are they causal/non-causal, open, closed?

3. How would you get the
   a) total effect
   b) direct effect

10 minutes

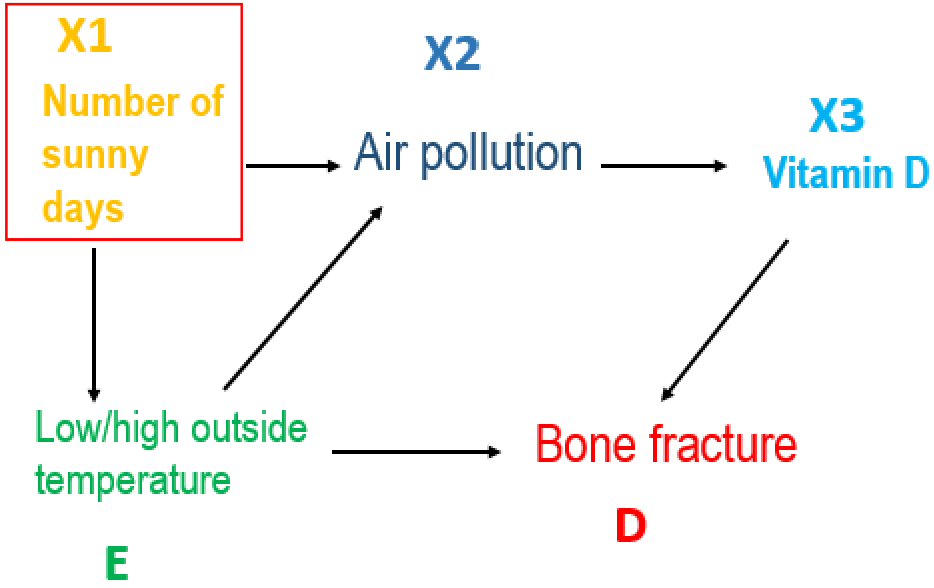
Solution 1a: Temperature and the risk of bone fracture in older adults

3.



Path	Causal/non-causal	Open/closed
E $\longrightarrow$ D	Causal	Open
E $\longrightarrow$ X2 $\longrightarrow$ X3 $\longrightarrow$ D	Causal (indirect)	Open
E $\longleftarrow$ X1 $\longrightarrow$ X2 $\longrightarrow$ X3 $\longrightarrow$ D	Non-causal	Open

Solution 1a: Temperature and the risk of bone fracture in older adults



3.

a) Total effect: Adjust for X1

Path	Causal/non-causal	Open/closed
E ⟶ D	Causal	Open
E ⟶ X2 ⟶ X3 ⟶ D	Causal (indirect)	Open
E ⟵ [X1] ⟶ X2 ⟶ X3 ⟶ D	Non-causal	Closed

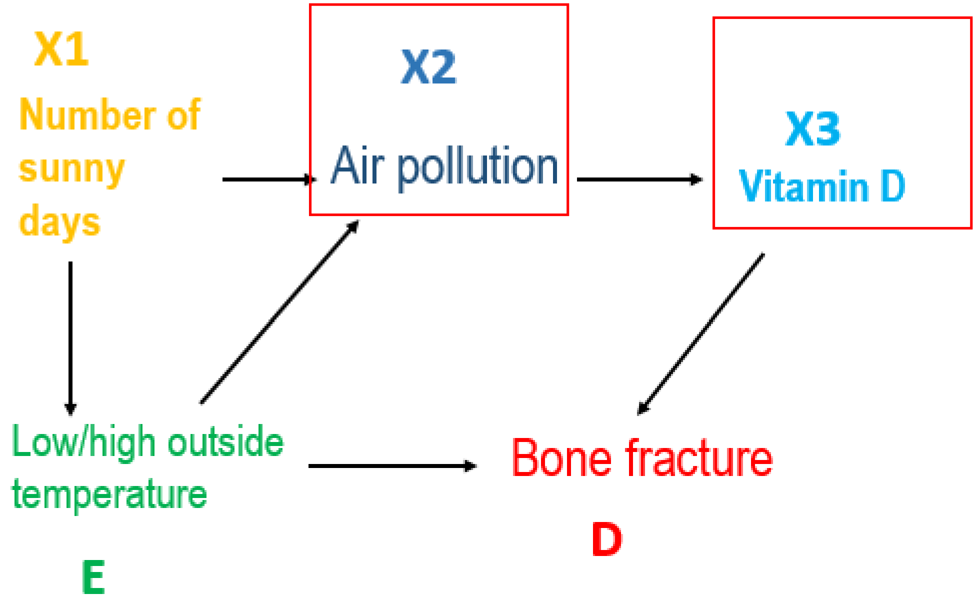
Solution 1a: Temperature and the risk of bone fracture in older adults

3.

a) Total effect: Adjust for X1

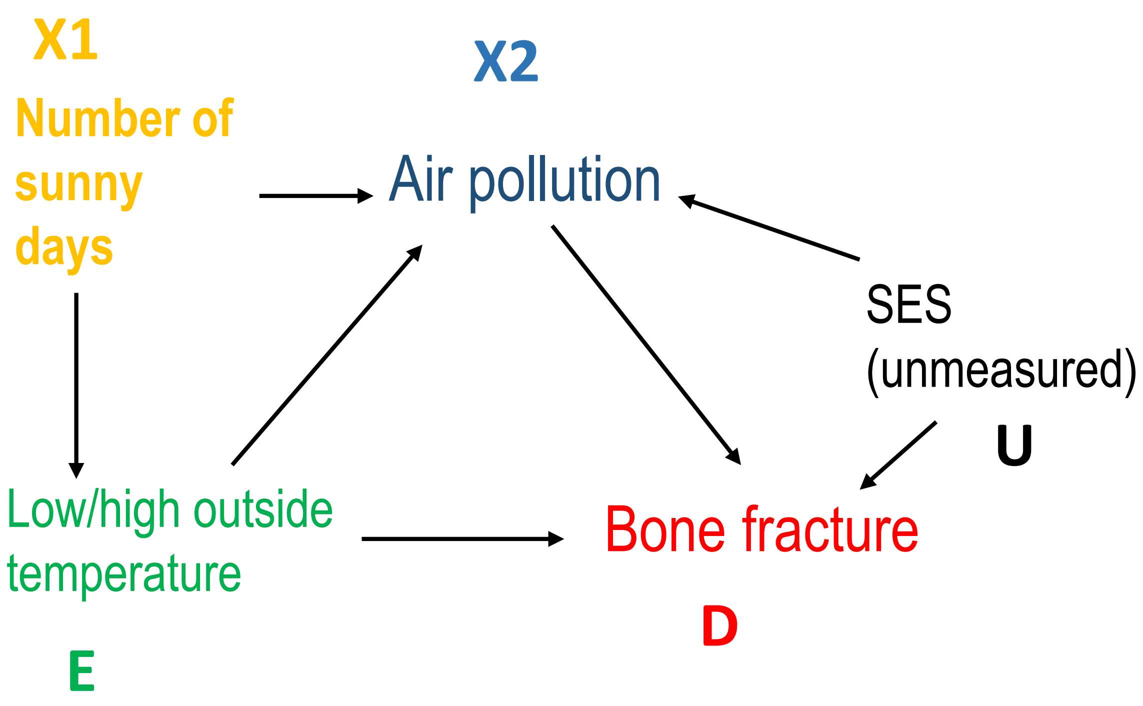
b) Direct effect: adjust for **X2** or **X3**

Also X1??



Path	Causal/non-causal	Open/closed
E ⟶ **D**	Causal	Open
E ⟶[X2]⟶ **X3** ⟶ **D**	Causal (indirect)	Closed
E ⟵**X1**⟶[X2]⟶ **X3** ⟶ **D**	Non-causal	Closed

Exercise 1b: Temperature and the risk of bone fracture in older adults



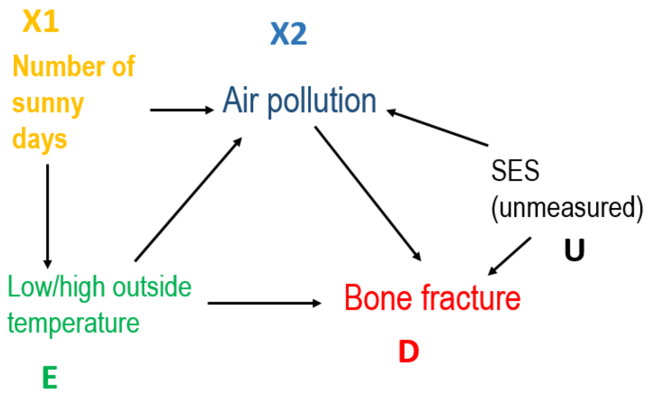
Low temperature= Mean yearly temp<10°C
High temperature=Mean yearly temp >10°C

1. Write down the paths
2. Are they causal/non-causal, open, closed?

3. How would you get the
    1. total effect
    2. direct effect

10 minutes

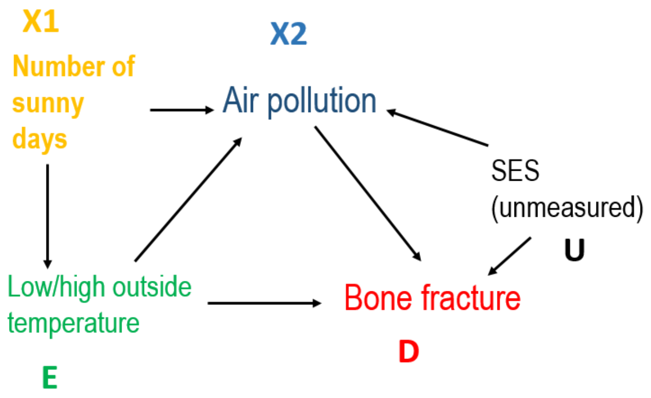
Solution 1b: Temperature and the risk of bone fracture in older adults



3.
a)
b)

Path	Causal/non-causal	Open/closed
E ⟶ D	Causal	Open
E ⟶ X2 ⟶ D	Causal (indirect)	Open
E ⟶ X2 ⟵ U ⟶ D	Non-causal	Closed (Collider)
E ⟵ X1 ⟶ X2 ⟶ D	Non-causal	Open
E ⟵ X1 ⟶ X2 ⟵ U ⟶ D	Non-causal	Closed (Collider)

Solution 1b: Temperature and the risk of bone fracture in older adults



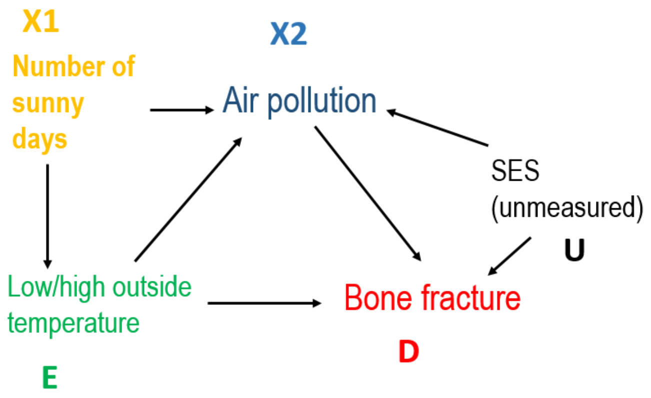
3.

a) Total effect: Adjust for X1

b) Direct effect:

Path	Causal/non-causal	Open/closed
E ⟶ D	Causal	Open
E ⟶ X2 ⟶ D	Causal (indirect)	Open
E ⟶ X2 ⟵ U ⟶ D	Non-causal	Closed
E ⟵ [X1] ⟶ X2 ⟶ D	Non-causal	Closed
E ⟵ [X1] ⟶ X2 ⟵ U ⟶ D	Non-causal	Closed

Solution 1b: Temperature and the risk of bone fracture in older adults



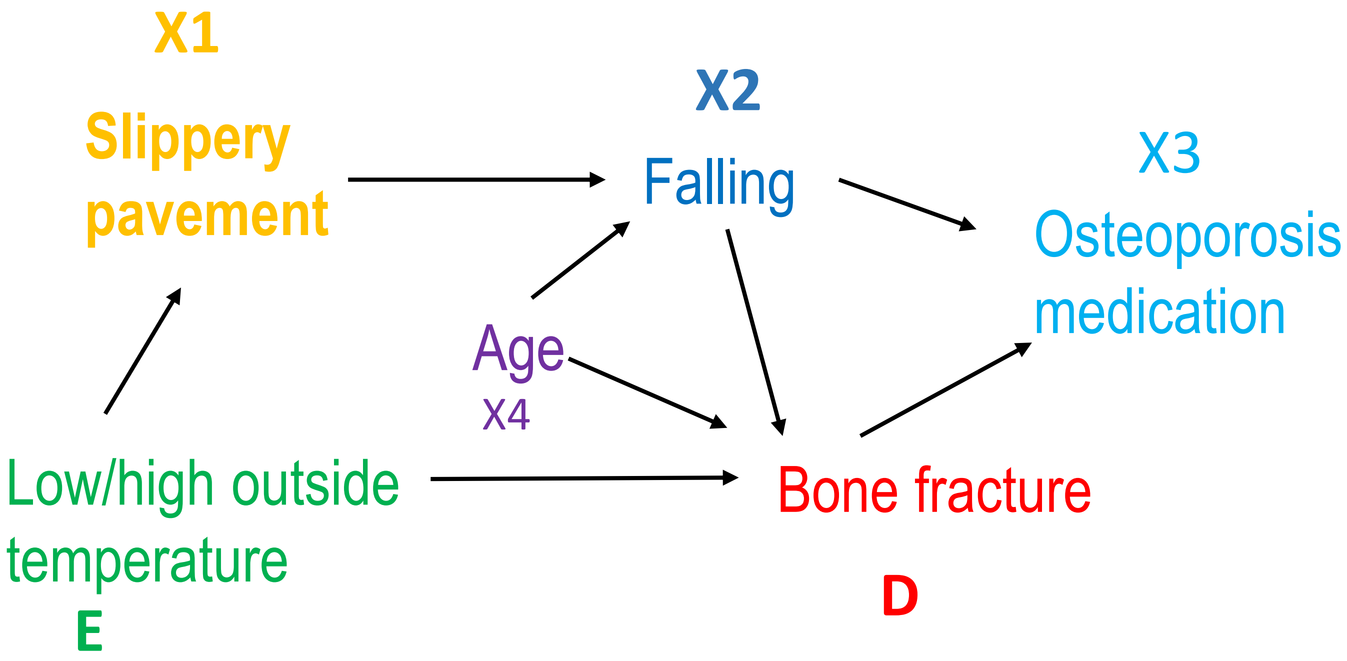
3.

a)Total effect: Adjust for X1

b) Direct effect: not possible

Path	Causal/non-causal	Open/closed
E ⟶ D	Causal	Open
E ⟶ [X2] ⟶ D	Causal (indirect)	Closed
E ⟶ [X2] ⟵ U ⟶ D	Non-causal	**Open** BIAS!
E ⟵ [X1] ⟶ [X2] ⟶ D	Non-causal	Closed
E ⟵ [X1] ⟶ [X2] ⟵ U ⟶ D	Non-causal	Closed (confounder adjustment)

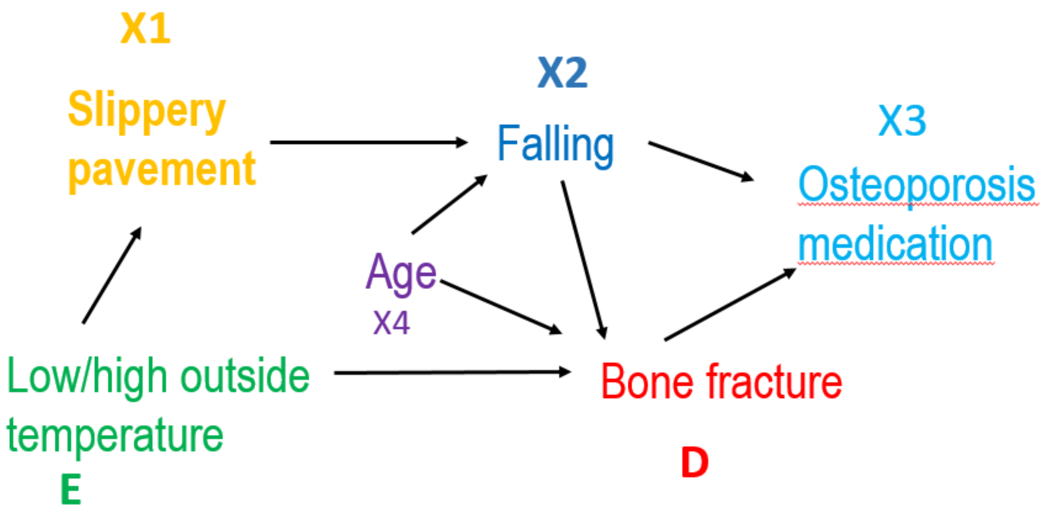
Exercise 2. Temperature and the risk of bone fracture in older adults



1. Write down all the paths
2. Are they open or closed, causal or non-causal?
3. How would you get the total effect of Outside temperature on Bone fracture?
4. Optional: How would you get the direct effect ?

5 minutes

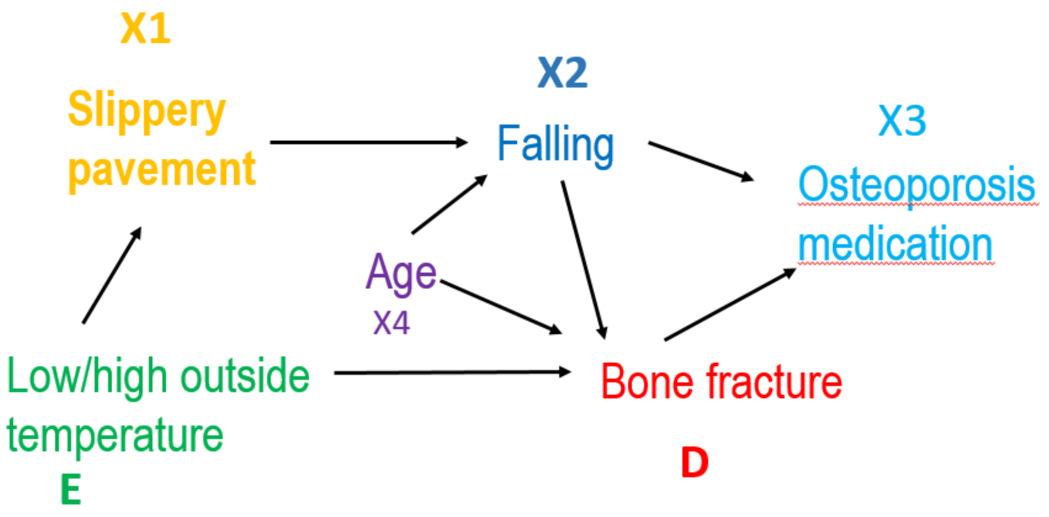
Solution 2. Temperature and the risk of bone fracture in older adults



3. Total effect: No adjustment necessary

Path	Causal/non-causal	Open/closed
E ⟶ D	Causal	Open
E ⟶ X1 ⟶ X2 ⟶ D	Causal	Open
E ⟶ X1 ⟶ X2 ⟶ X3 ⟵ D	Non-causal	Closed
E ⟶ X1 ⟶ X2 ⟵ X4 ⟶ D	Non-causal	Closed

Solution 2. Temperature and the risk of bone fracture in older adults

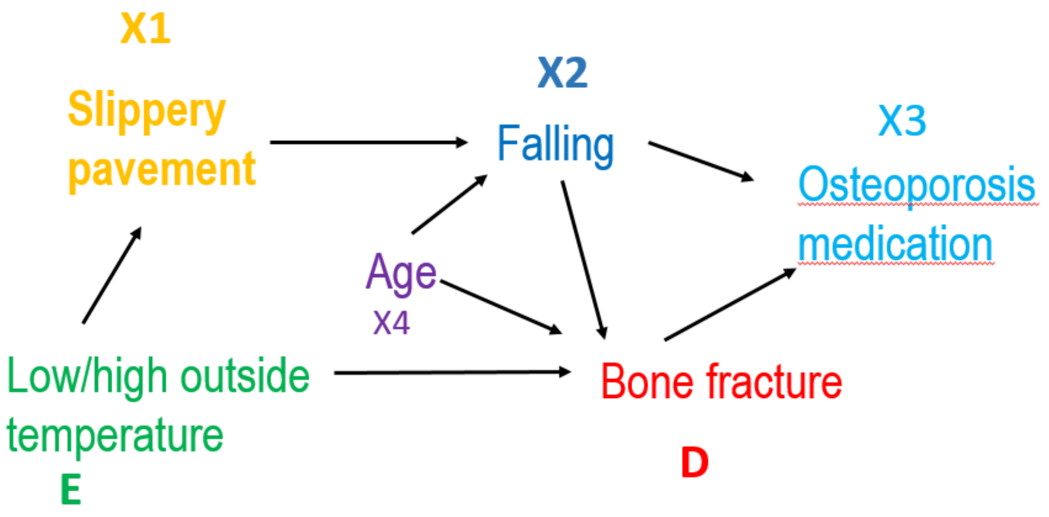


3. Total effect: No adjustment necessary

4. Direct effect (optional): Adjust for X1 **or** for X2 and X4

Path	Causal/non-causal	Open/closed
E → D	Causal	Open
E →[X1] →X2→D	Causal	Closed
E →[X1] →X2 →X3 ←D	Non-causal	Closed
E →[X1]→ X2 ←X4→D	Non-causal	Closed

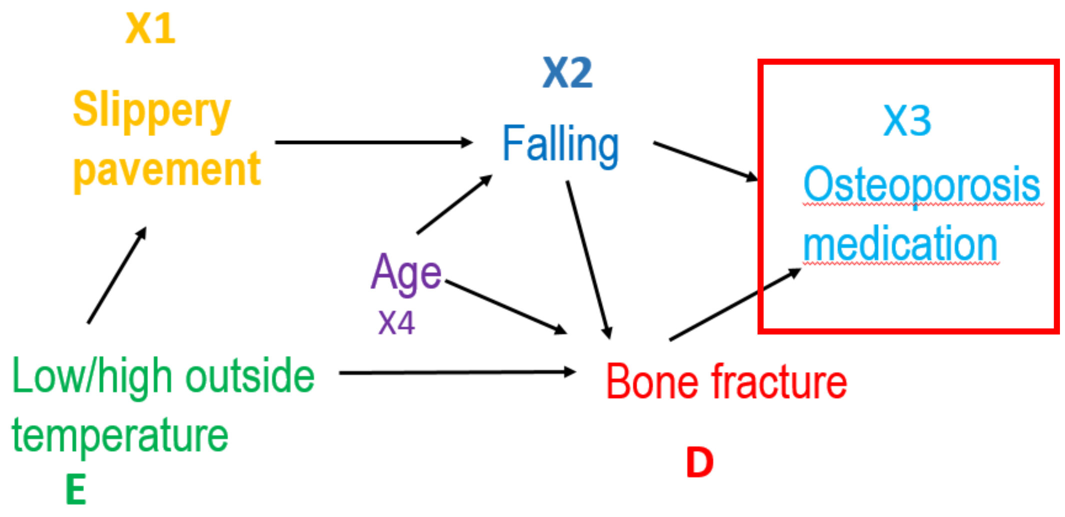
Solution 2. Temperature and the risk of bone fracture in older adults



3. Total effect: No adjustment necessary
4. Direct effect (optional):
Adjust for X1 **or** for X2 and X4

Path	Causal/non-causal	Open/closed
E ⟶ D	Causal	Open
E ⟶ X1 ⟶ [X2] ⟶ D	Causal	Closed
E ⟶ X1 ⟶ [X2] ⟶ X3 ⟵ D	Non-causal	Closed
E ⟶ X1 ⟶ [X2] ⟵ [X4] ⟶ D	Non-causal	Closed

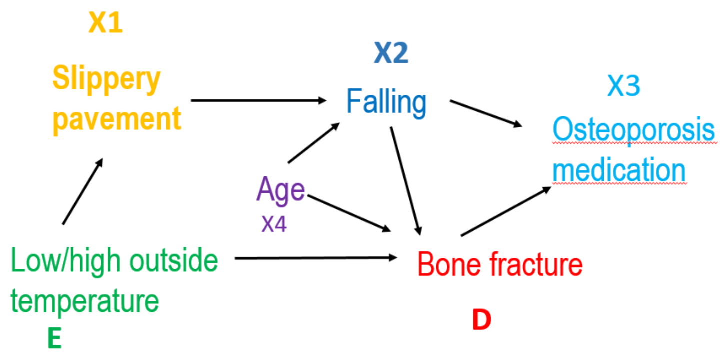
Solution 2. Temperature and the risk of bone fracture in older adults



3. Total effect: No adjustment necessary
4. Direct effect (optional):
Adjust for X1 **or** for X2 and X4

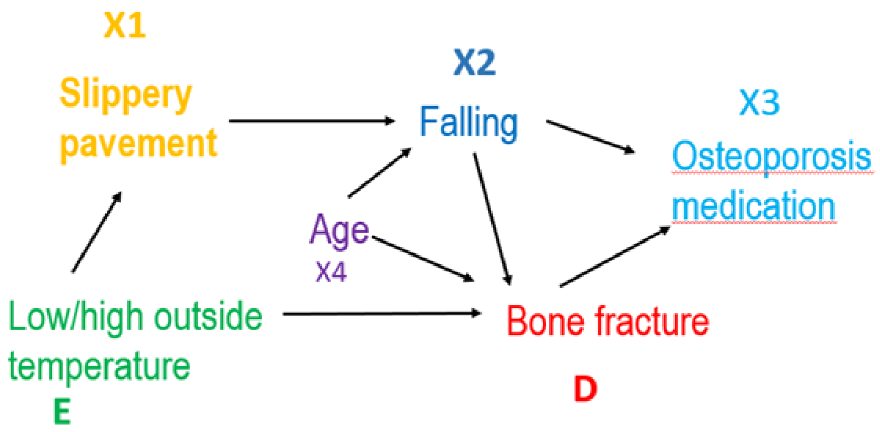
Path	Causal/non-causal	Open/closed
E → D	Causal	Open
E → X1 → X2 → D	Causal	Open
E → X1 → X2 → [X3] ← D	Non-causal	**Open**
E → X1 → X2 ← X4 → D	Non-causal	Closed

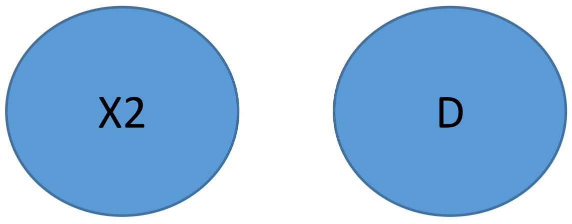
Hypothetical analysis



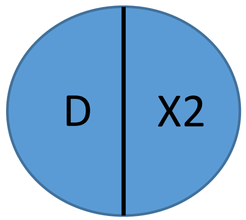
	Bone fracture				
	Yes	No	Total personyears	Rate	RD
Low outside temperature	84	9,916	10,000	0.0084	0.0
High outside temperature	84	9,916	10,000	0.0084	

Sufficient causes for osteoporosis medication

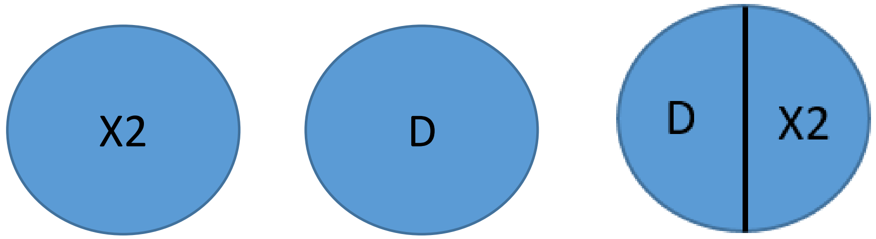




Fall (as a result of low temperature) or bone fracture



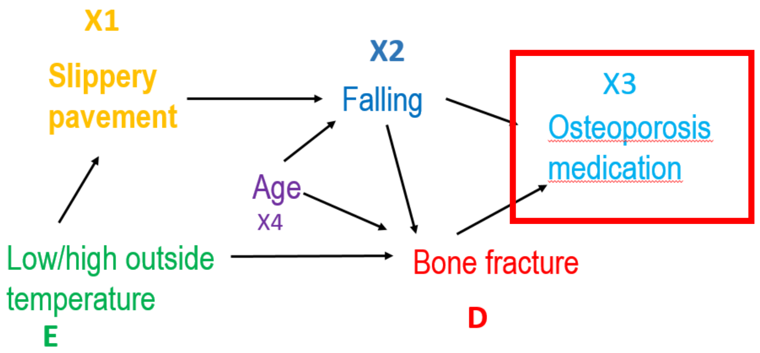
Fall and bone fracture



Fall or bone fracture and fall or bone fracture

(Assuming that the effect of age on falling is small, and that there are no other causes of bone fractur)

Hypothetical analysis -restricting on X3



Bone fracture					
	Yes	No	Total personyears	Rate	RD
Low outside temperature	84	9,916	10,000	0.0084	0.0
High outside temperature	84	9,916	10,000	0.0084	
Condition on X3 (medication use can only be due to falling as a result of slippery pavement from low outside temperature, or having experienced a bone fracture)					
Medication=yes					
Low outside temperature	84	9,916	10,000	0.0084	-0.99
High outside temperature	84	**0**	84	1.0	

How to select an adjustment set?

- Adjustment set: *minimum* set of variables to include in analysis in addition to our exposure and outcome

- By hand: Write down all paths between exposure and outcome, and between covariates, close the open non-causal paths
  - Can be difficult!

- Rather: Use a program!
  - The program will give a suggestion on the variables to include to obtain an unbiased association (if possible)

http://dagitty.net/

UiO : Universitetet i Oslo

http://dagitty.net/

Welcome to DAGitty!

Launch	Download	Learn	Code
Launch DAGitty online in your browser	Download DAGitty's source for offline use	Learn more about DAGs and DAGitty	The R package "dagitty" is available on CRAN or github

What is this?

DAGitty is a browser-based environment for creating, editing, and analyzing causal mode
or causal Bayesian networks). The focus is on the use of causal diagrams for minimizing
and other disciplines. For background information, see the "learn" page.

DAGitty is developed and maintained by Johannes Textor (Tumor Immmunology Lab anc
Sciences, Radboud University Nijmegen). The algorithms implemented in DAGitty were c
Maciej Liśkiewicz and Benito van der Zander, University of Lübeck, Germany (see literatu

DAGitty development happens on GitHub. You can download all source code from there

How can I get help?

If you encounter any problems using DAGitty, or would like to have a certain feature impl
textor {at} gmx {dot} de". Your feedback and bug reports are very welcome and contribute
for everyone. Past contributors are acknowledged in the manual.

Is it free?

Because the main purposoe of DAGitty is facilitating the use of causal models in empirica
software (both "free as in beer" and "free as in speech"). You can copy, redistribute, and
general public license. Enjoy!

DAGitty development has been sponsored by the Leeds Institute for Data Analytics and b
(DFG), grant 273587939.

Deutsche

Versions

The following versions of DAGitty are available:

- Development version
  Recent development snapshot. May contain new features, but could also contain new bugs.
- Experimental version
  Most recent development snapshot. May not even work.
- 2.3: Released 2015-08-19
- 2.2: Released 2014-10-30
- 2.1: Released 2014-02-06
- 2.0: Released 2013-02-12
- 1.1: Released 2011-11-29
- 1.0: Released 2011-03-24
- 0.9b: Released 2010-11-24
- 0.9a: Released 2010-09-01

News on Twitter

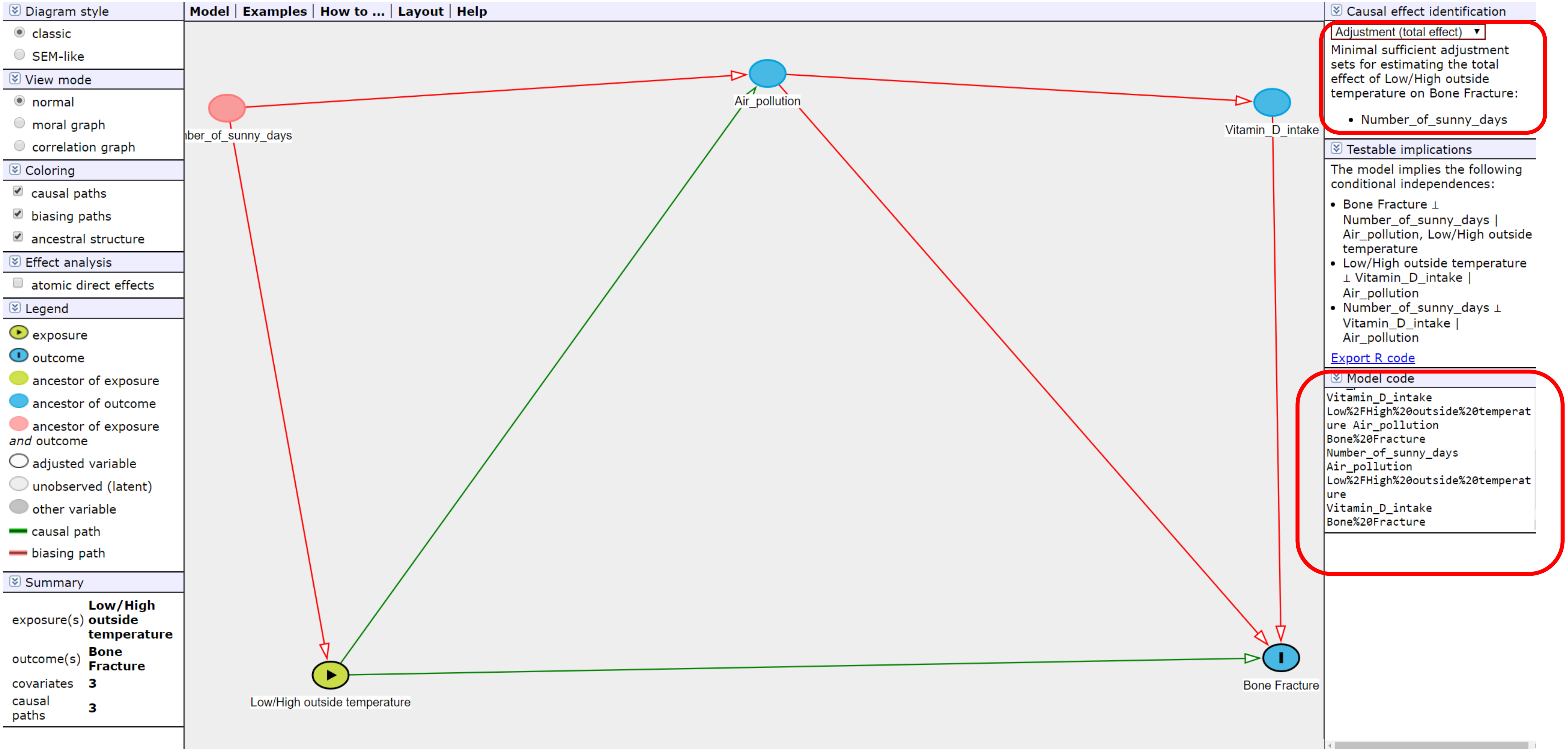
#dagitty

Wolfhart Feldmeier Retweeted

Malcolm Barrett
@malco_barrett

🎉 Now on CRAN: ggdag 0.1.0! Create and analyze tidy causal DAGs in #rstats using ggplot2. Powered by @JohannesTextor's amazing dagitty pkg with graph 💪 from ggraph by @thomasp85. #causalinference #dagitty malco.io/2018/03/28/ggd...

Draw, Analyze, Test



(Textor, Hardt et al. 2011)

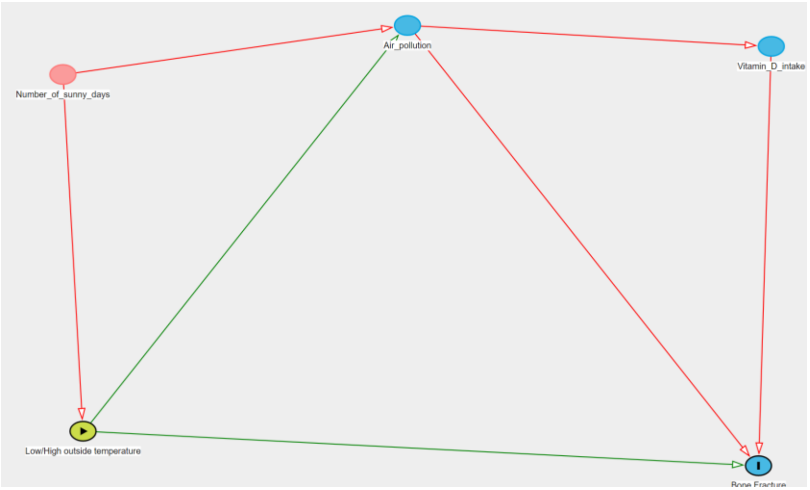
Draw model

- Draw new model
  - Model>New model, Exposure, Outcome
- New variables, connect
  - nnew variable (or double click)
  - c connect (hit c over V1 and over V2 to connect)
  - r rename
  - ddelete
- Status (toggle on/off)
  - uunobserved
  - aadjusted

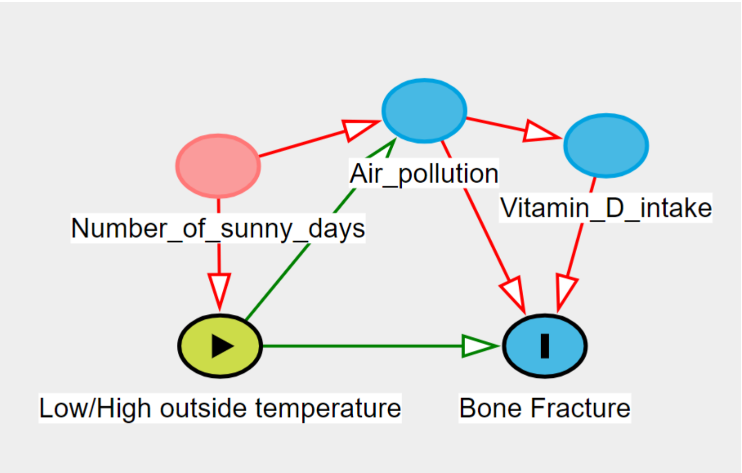
Export DAG

- Export to Word or PowerPoint
  - "Zoom" the DAGitty drawing first (*Ctrl-roll*)
  - Use "Snipping tool" or
  - use Model>Export as PDF

Without zooming



With zooming



Daggity: Draw all causal relationships

- Draw all variables/ factors that may influence your *outcome*
- Consider whether these factors also affect your *exposure*
- Are there any arrows between *cofactors*?
- Also put in common *ancestors of any two variables included,* also variables that are *unobserved*

- Remember*: Absence* of arrow is a strong statement
  - Omitting an arrow will explicitly state that there is no association between variables in *any* of your participants

- Now: Use Daggity to draw and analyse DAGs in your own research question.
- Next: Discussion of student examples
- Short summary

Summing up

- Data driven analyses do not work. Need (causal) information from outside the data.
- DAGs are intuitive and accurate tools to display that information.
- Paths show the flow of causality and of bias and guide the analysis.
- DAGs clarify concepts like confounding and selection bias, and show that we can adjust for both.

Better discussion based on DAGs

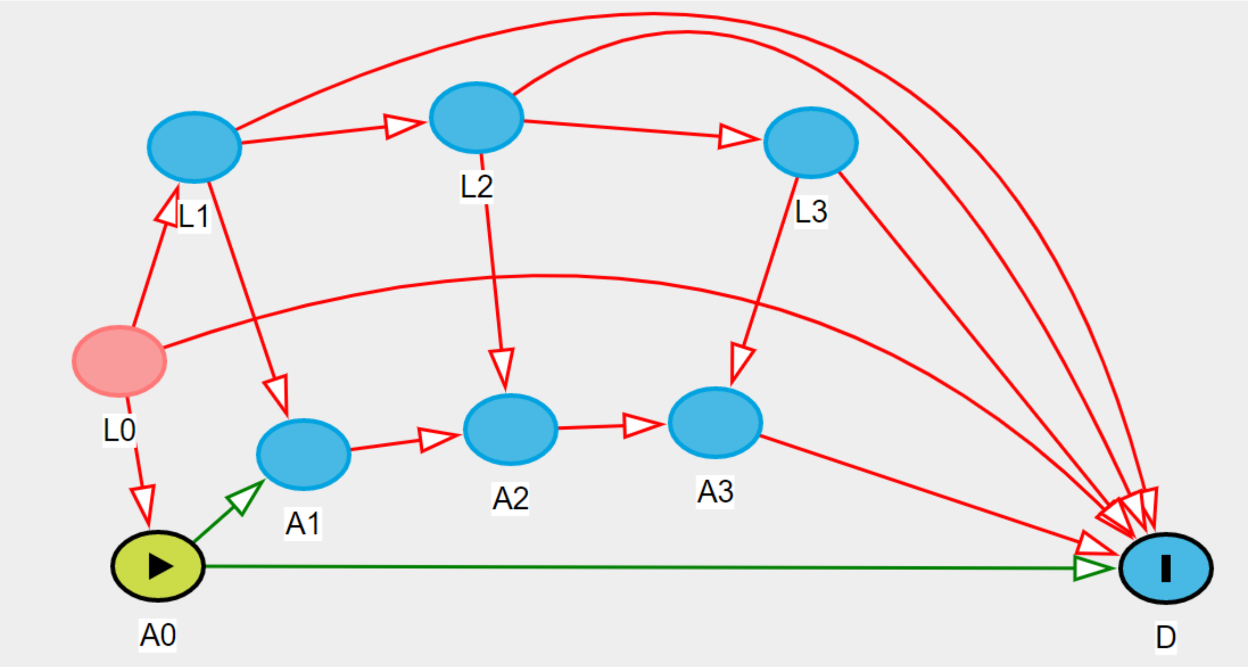
Draw your assumptions before your conclusions

Recommended DAG reading

- Books
  - Hernan, M. A. and J. Robins. *Causal Inference*. Web:https://www.hsph.harvard.edu/miguel-hernan/causal-inference-book/
  - Rothman, K. J., S. Greenland, and T. L. Lash. *Modern Epidemiology,* 2008.
  - Morgan and Winship, *Counterfactuals and Causal Inference*, 2009
  - Pearl J, *Causality – Models, Reasoning and Inference*, 2009
  - Veierød, M.B., Lydersen, S. Laake,P. Medical Statistics. 2012
- Papers
  - Greenland, S., J. Pearl, and J. M. Robins. *Causal diagrams for epidemiologic research,* Epidemiology 1999
  - VanderWeele TJ. 2016. Mediation analysis: A practitioner's guide. Annual Review of Public Health, Vol 37 37:17-32.
  - Hernandez-Diaz, S., E. F. Schisterman, and M. A. Hernan. *The birth weight "paradox" uncovered?* Am J Epidemiol 2006
  - Hernan, M. A., S. Hernandez-Diaz, and J. M. Robins. *A structural approach to selection bias,* Epidemiology 2004
  - Berk, R.A. *An introduction to selection bias in sociological data*, Am Soc R 1983
  - Greenland, S. and B. Brumback. *An overview of relations among causal modeling methods,* Int J Epidemiol 2002
  - Weinberg, C. R. *Can DAGs clarify effect modification?* Epidemiology 2007

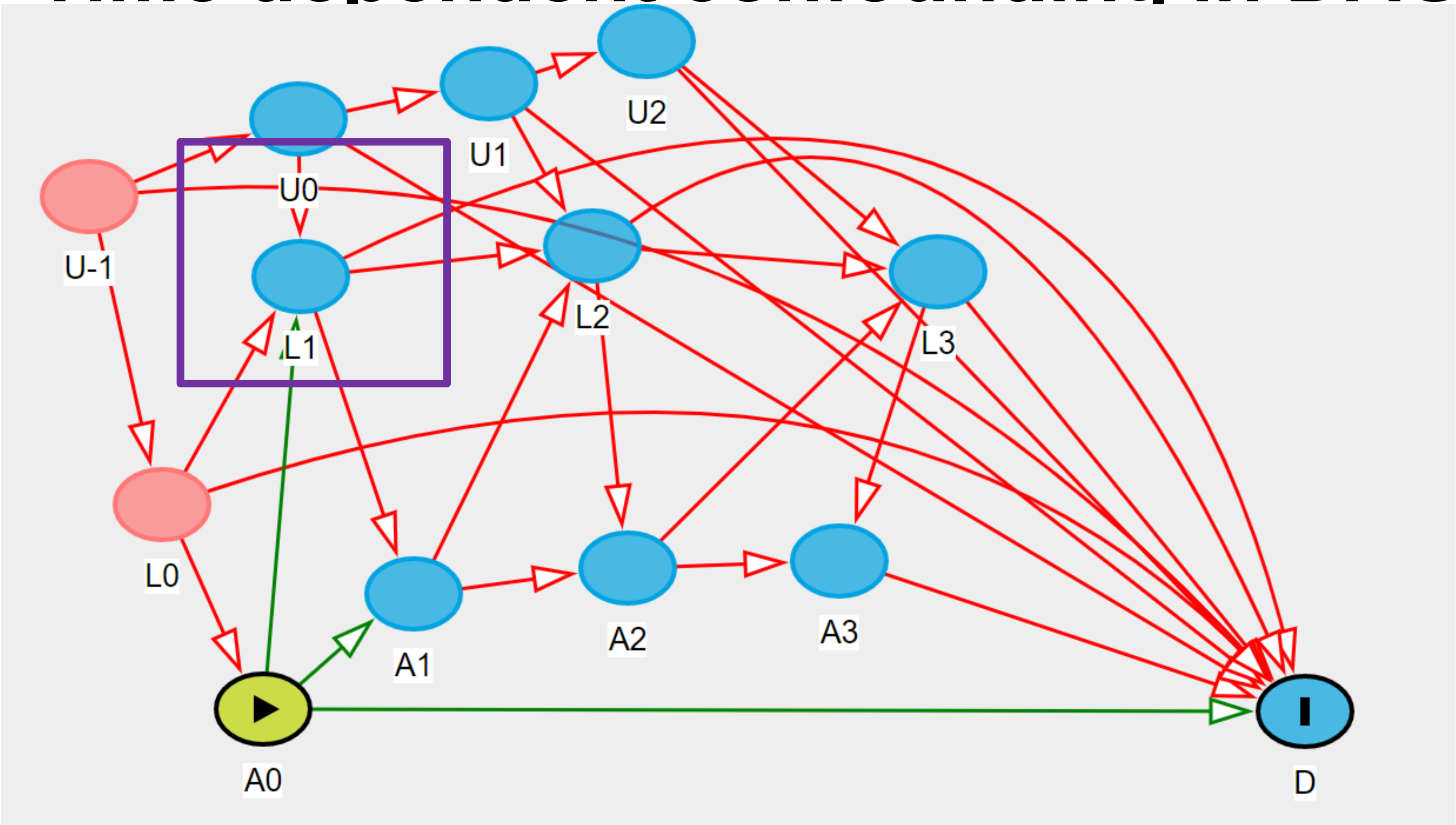
EXTRA

Time dependent confounding in DAGs



Generally okay when the treatment (A) does not have an effect on the diagnosis criteria (L) that determines the next treatment

Time dependent confounding in DAGs



More difficult when the treatment (A) has an effect on the diagnosis criteria (L) that determines the next treatment (A), because then you open collider paths and need to adjust for U (which is unknown). Can use IPW, g-estimation.