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 DAGs?
- Short- Need to know
- Analyzing a DAG (with examples)
- Daggity- the «easy way» to draw and analyze a DAG

- Practice drawing your own DAGs, discussion

WHY DO WE NEED 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 (in the unexposed)

3. Not on the causal pathway between exposure and outcome

Criteria for a variable to be a con[illegible]der:

1. Associated with the exposure in the so[illegible]ation

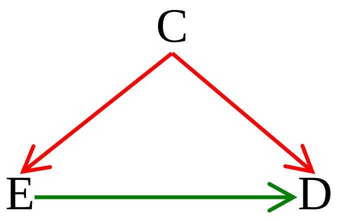
2. Associated with the outcome

3. Not on the causal path[illegible]een exposure and outcome

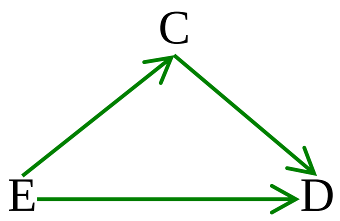
Not a sufficient definition

Adjust for C: cause versus association

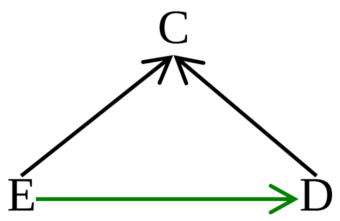
Cause:



Confounder

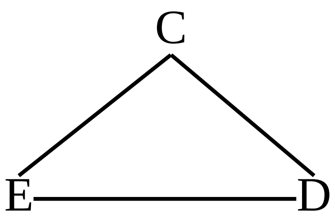


Mediator



Collider

Association:



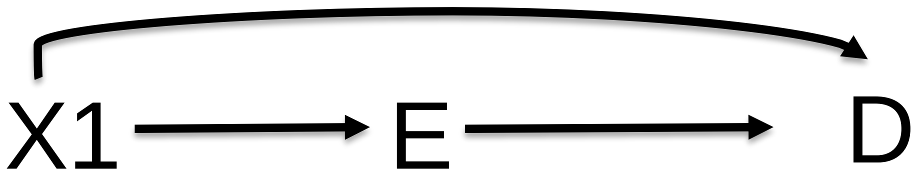
Statistical criteria:
likelihood ratio, AIC, 10% change in estimate
cannot differentiate between
Confounder, Mediator or Collider

Need causal model to do a proper analysis

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 Y

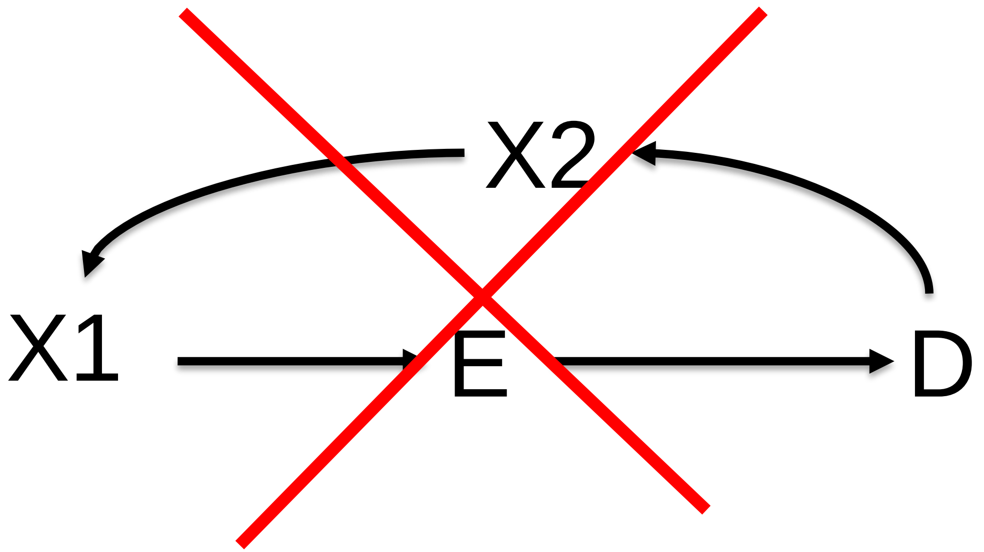


Parent → Child

Ancestor → Descendant

Acyclic

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



Components

- Variables represented as «nodes» or «vertices»

- Arrows (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. Association paths

E and D are associated through X1

X1 is a cause of E and D

Complete 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 if they are part of the causal structure

  – Did we include all common ancestors of two variables?

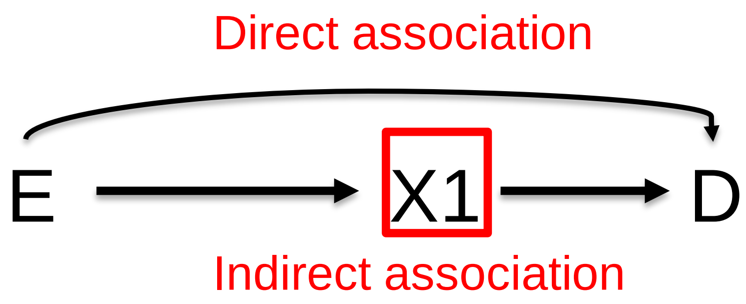
  – Did we include all the arrows?
    - Absense of an arrow is a strong statement

ANALYZING A DAG

3 reasons why two variables may be associated

- One causes the other
- They share common causes
- They share a common effect and the analysis is restricted to certain level of that common effect (or of its descendants)

Conditioning on a mediator



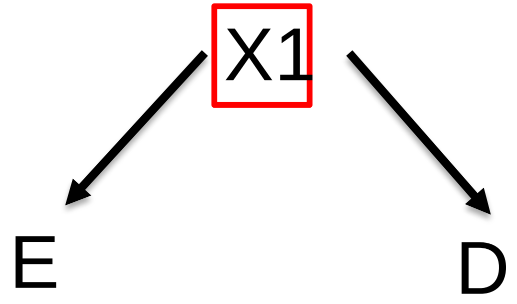
Conditioning blocks the indirect path- only direct association left.

May require special «mediation analysis»

Example: Aspirin (E) affects risk of heart disease (D) through reducing platelet aggregation (X1)

Conditioning on a common cause

- Confounding



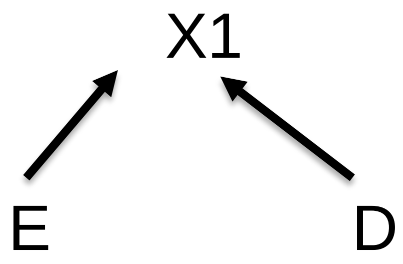
Unconditioned-
Information on E could give information on E through X1

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

Conditioning on a common effect

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

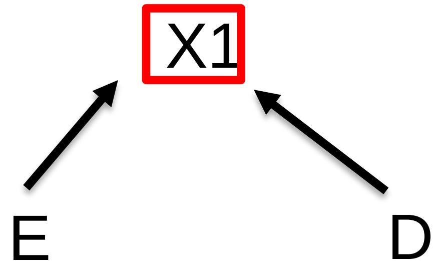


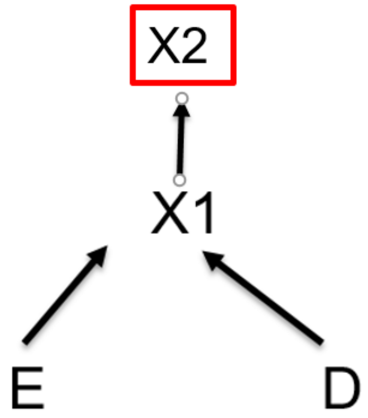
Information on E does not give information on D.

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

Conditioning on a common effect

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





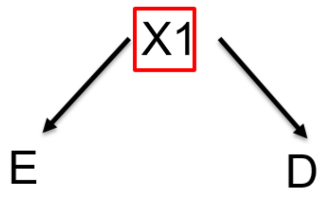
Conditioning on X1=1 or X1=0:
Information on E does give information on D.
- If E and D are the only two causes of X1, conditioning on X1=1 means that if and E=0, then D has to be=1

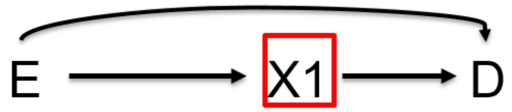
$\perp$ = independent

D- separation

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

$E \perp D \mid X1$





- The opposite is «d-connected»

How to identify and block backdoor 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

$\longrightarrow X1 \longrightarrow$ $\quad$ $\longleftarrow X1 \longleftarrow$ $\quad$ $\longleftarrow X1 \longrightarrow$ $\quad$ $\longrightarrow X1 \longleftarrow$

- Are there any existing natural blocks?

Paths

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

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

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

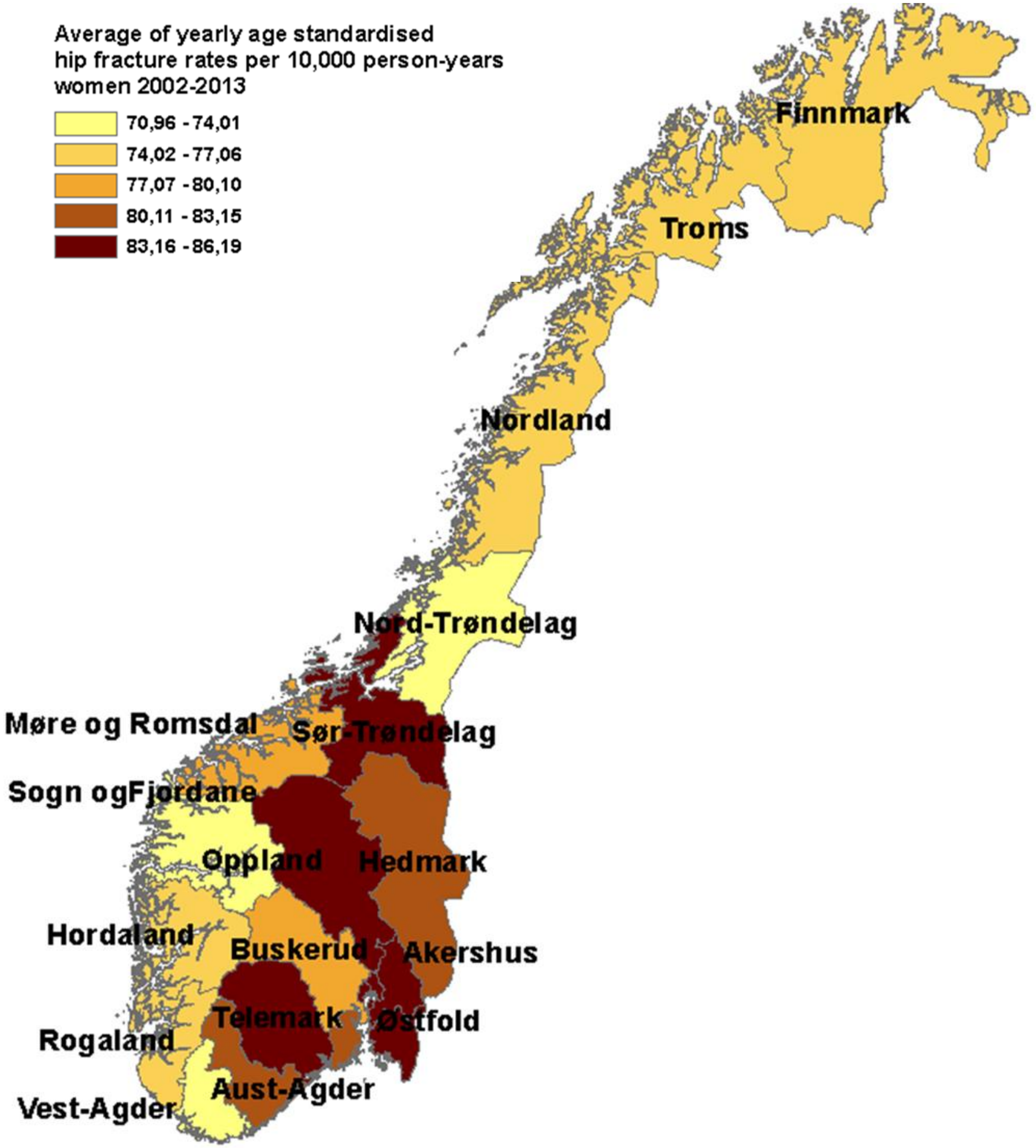
- Bone fracture incidence has been found to vary by season in Norway and in other countries, higher incidence in wintertime vs. summertime

- There is also a variation by latitude, in general a lower 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 a «causal factor» in this pattern?

UiO : Universitetet i Oslo



Forsén et al (unpublished)

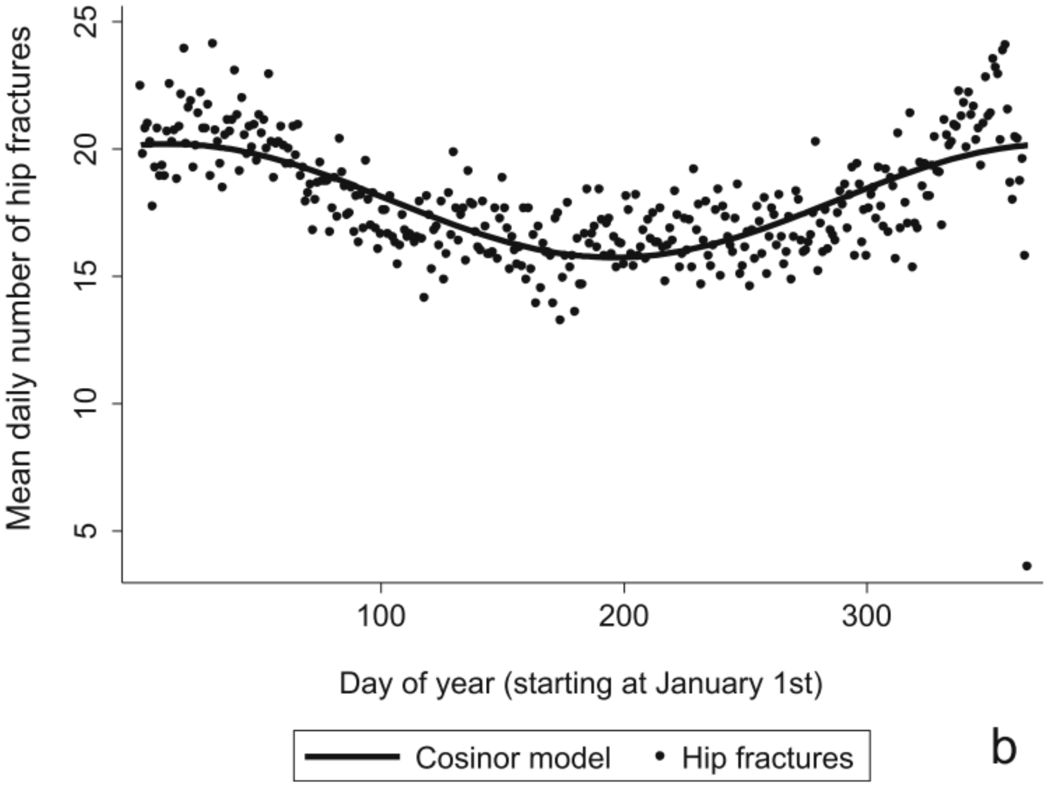
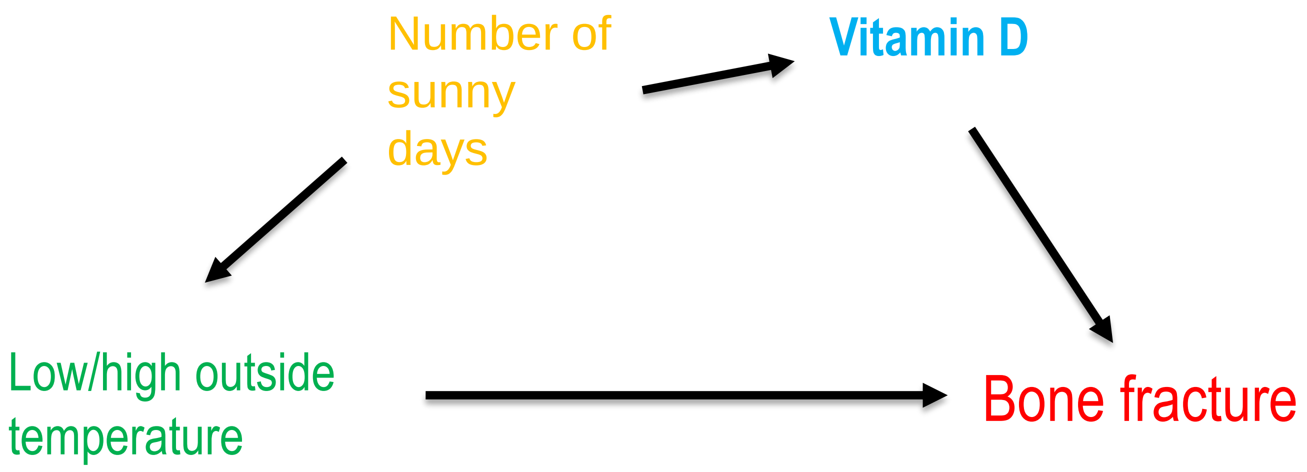


Fig. 2 Mean daily number of hip fractures during 1994–2008 (Norwegian patients aged 50–103 years) in **a** men and **b** women. A time series model (Cosinor model) is fitted. Norwegian Epidemiologic Osteoporosis Studies (NOREPOS)

(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

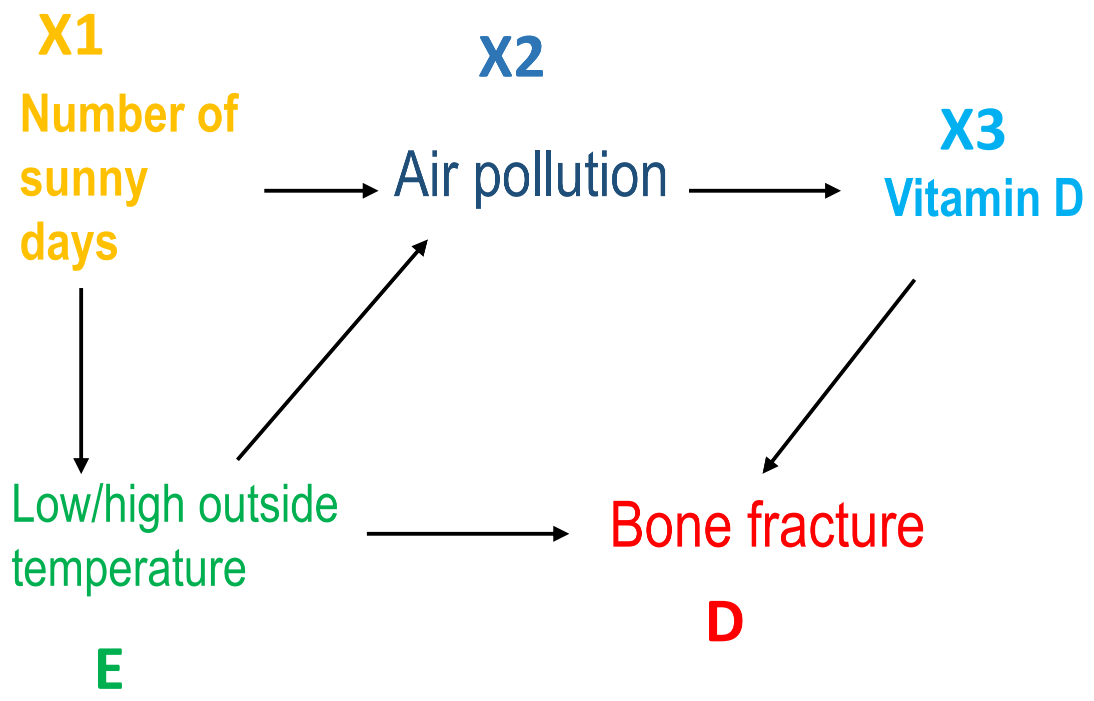
Bone mineral density (BMD)

Body Mass Index

Ethnicity

Socioeconomic status (SES)

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

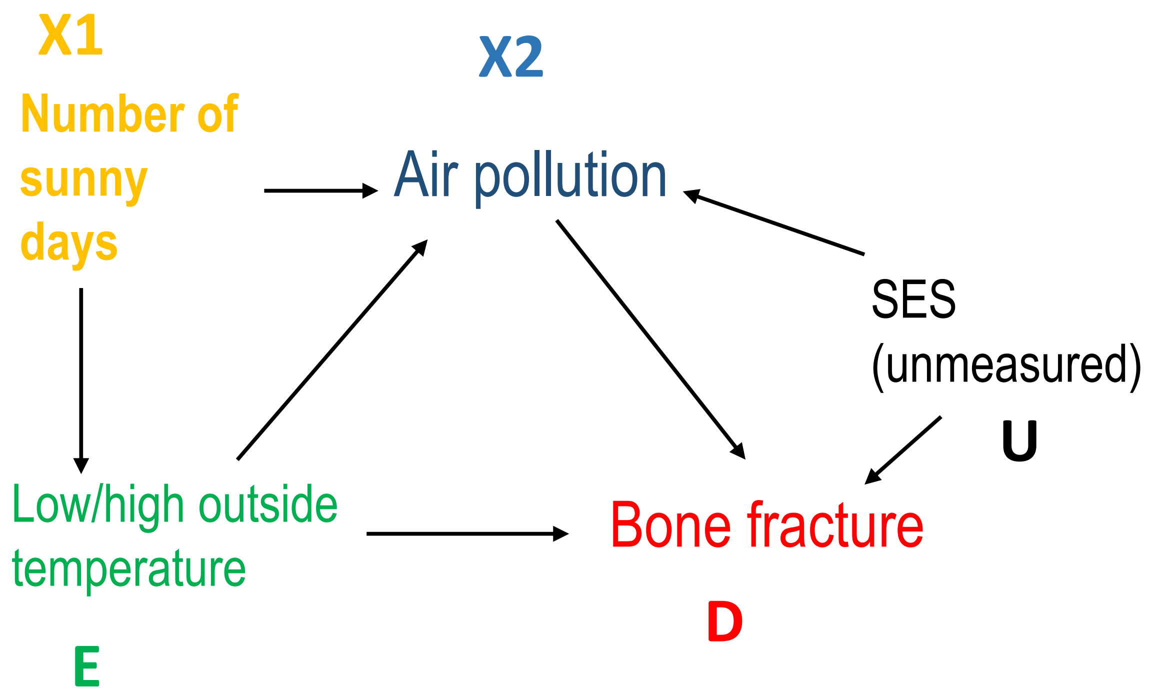


Low temperature= Mean yearly temp<10°C
High temperature=Mean yearly temp >10°C

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

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

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

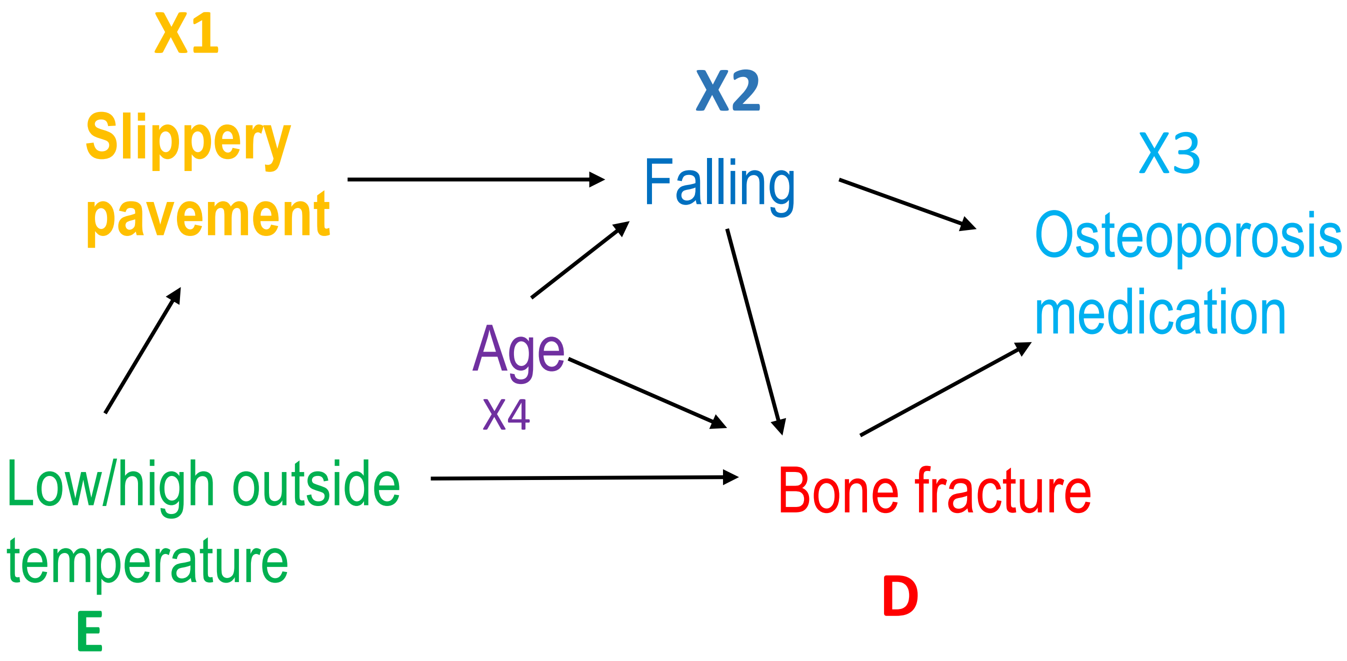


Low temperature= Mean yearly temp<10ºC
High temperature=Mean yearly temp >10ºC

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

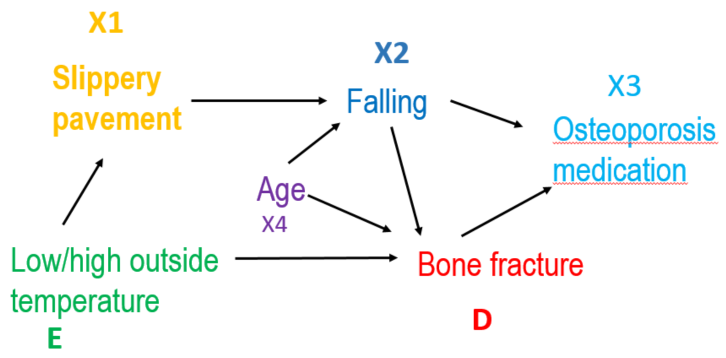
3. How would you get the
   1. total effect
   2. direct effect

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 ?

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	

What happens if you restrict on osteoporosis medication? (Will be covered in class)

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 and 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.

DFG 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

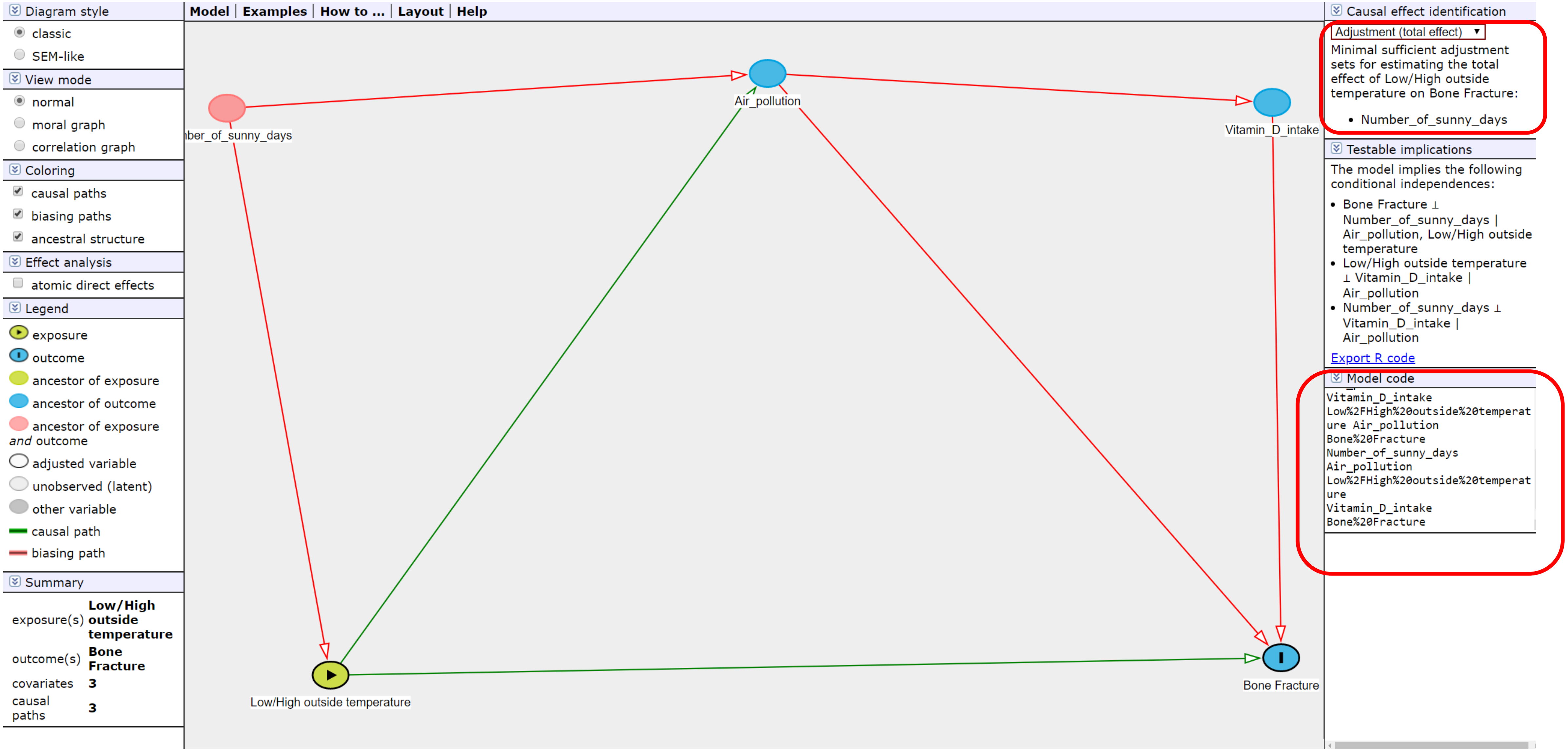
#dagitty

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)

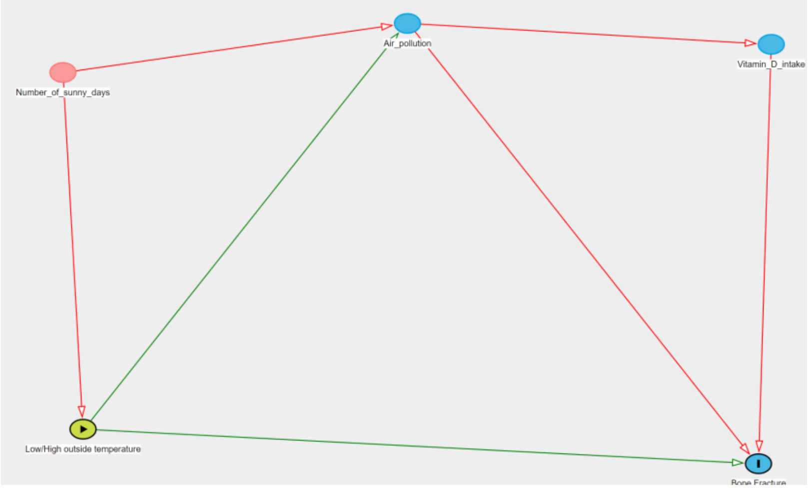
Draw model

- 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

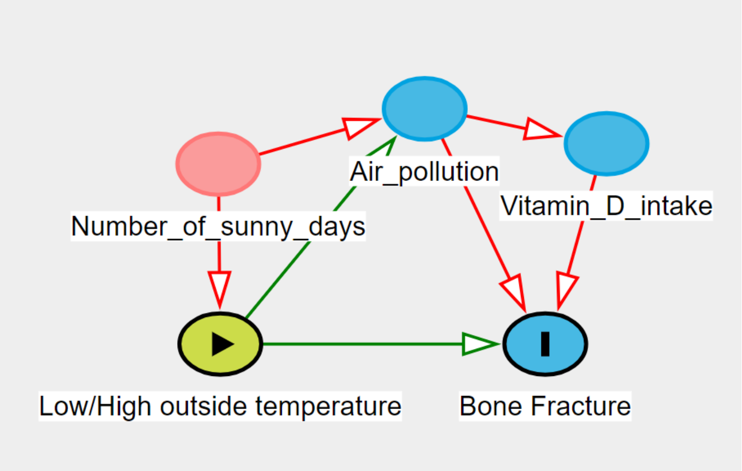
Export DAG

- 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