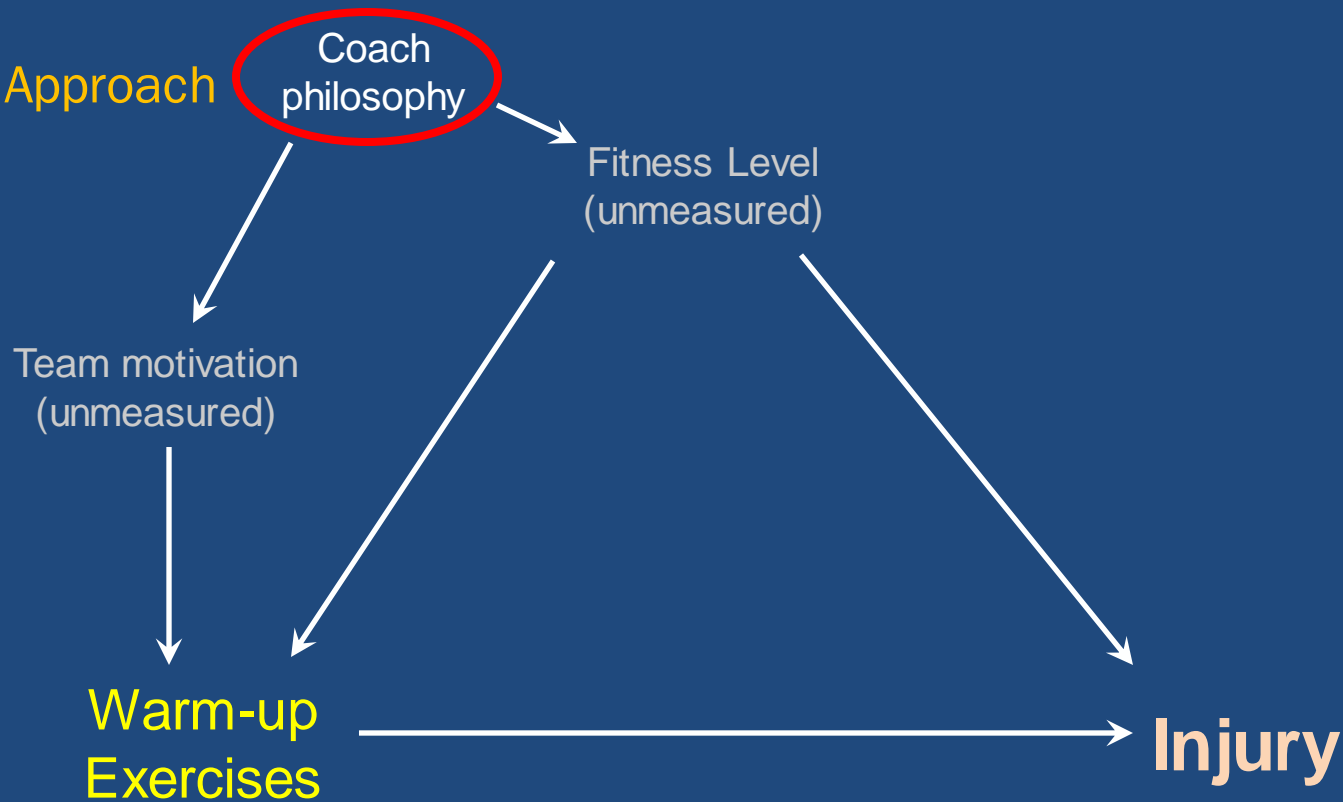


Causal Diagrams: A tool to better understand results from epidemiological and machine learning analyses

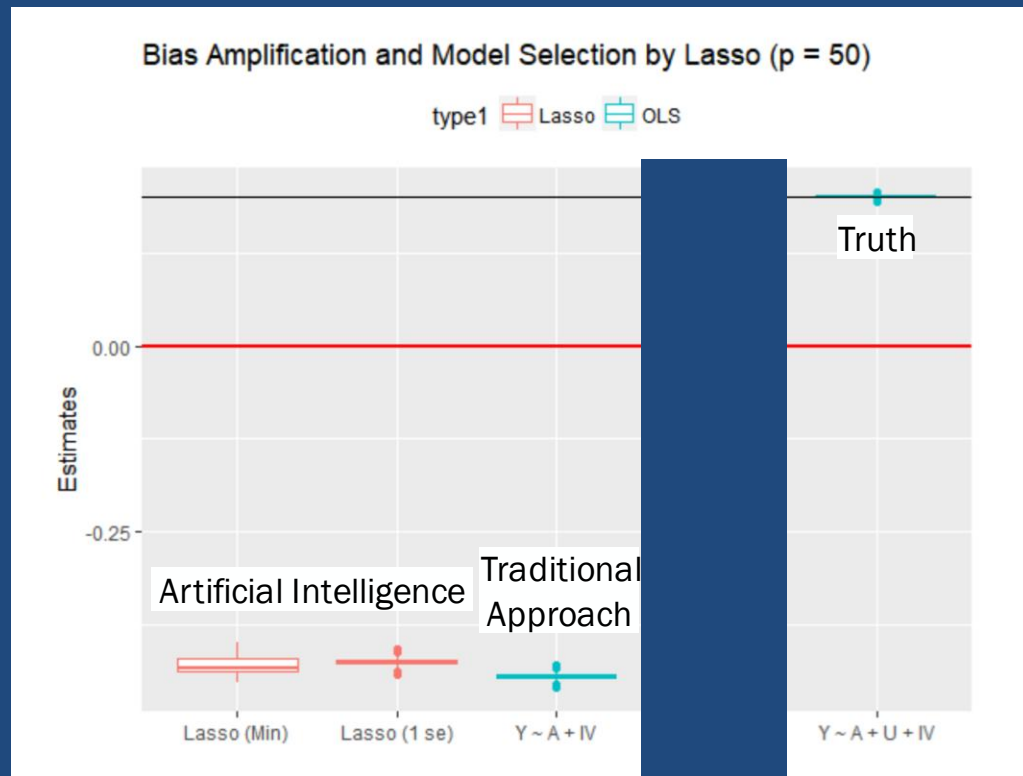
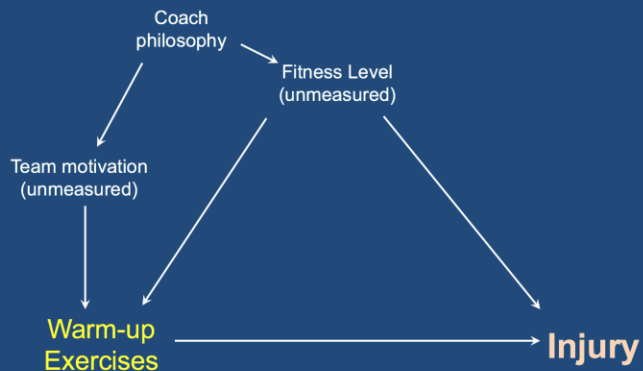
Ian Shrier MD, PhD, Dip Sport Med (FACSM)
Centre for Clinical Epidemiology,
Lady Davis Institute, Jewish General Hospital, McGill University

AI versus Causal Inference

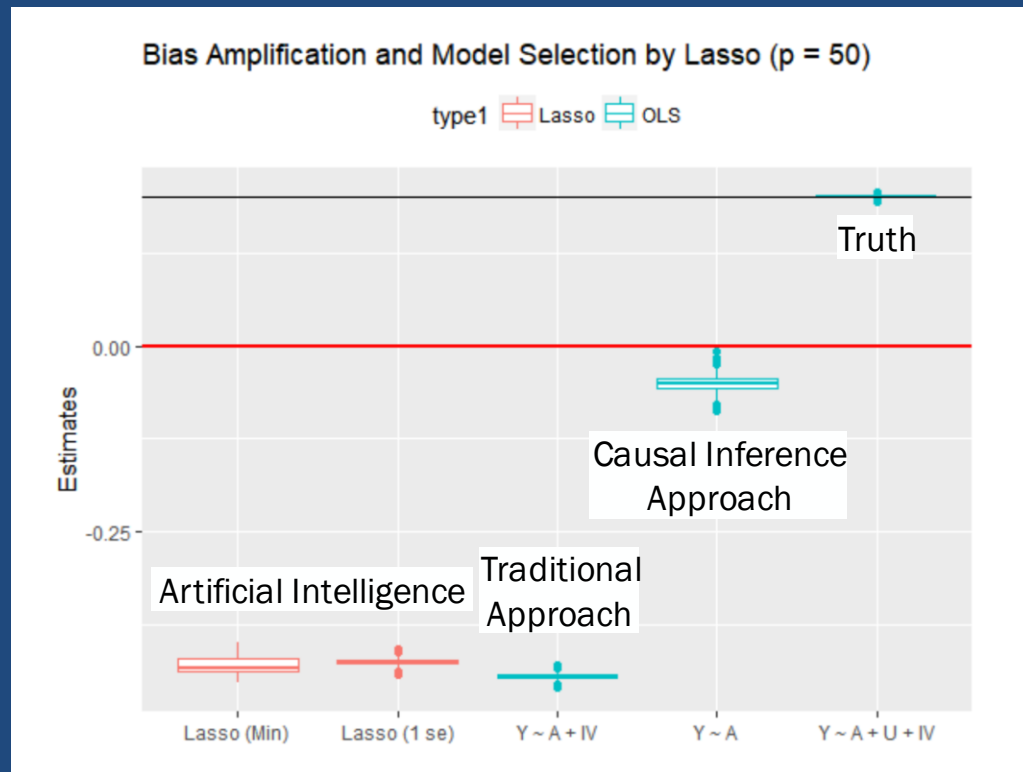
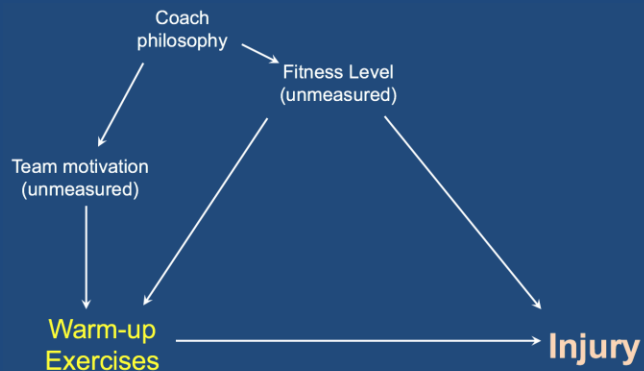
Both Traditional & AI Approach



AI versus Causal Inference

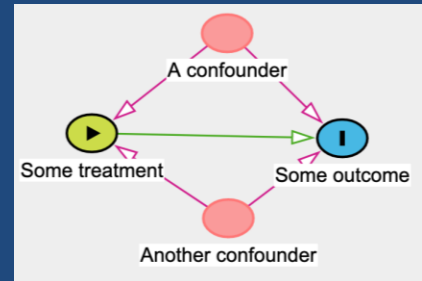
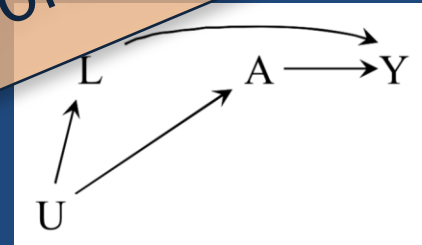


AI versus Causal Inference

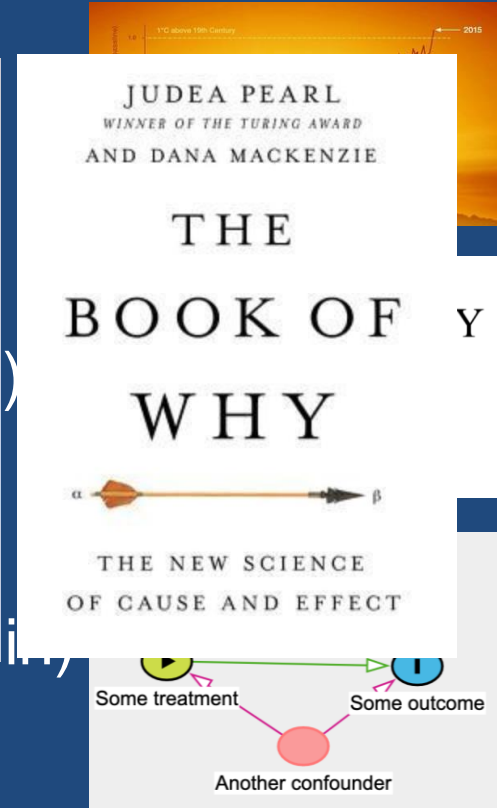
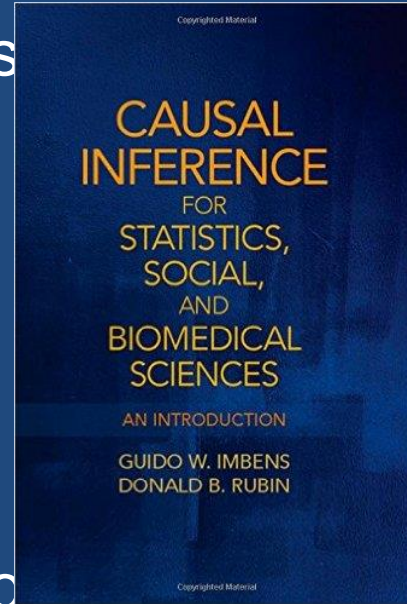
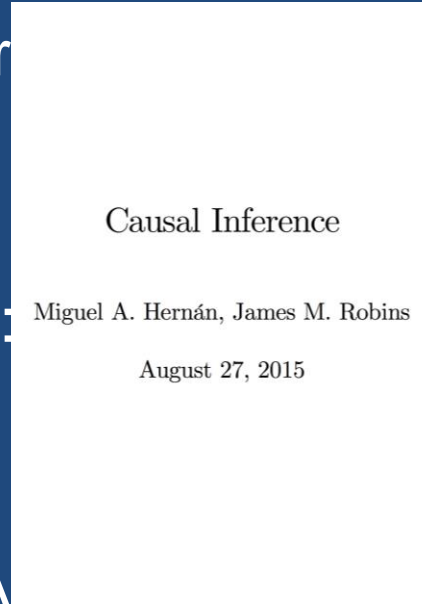
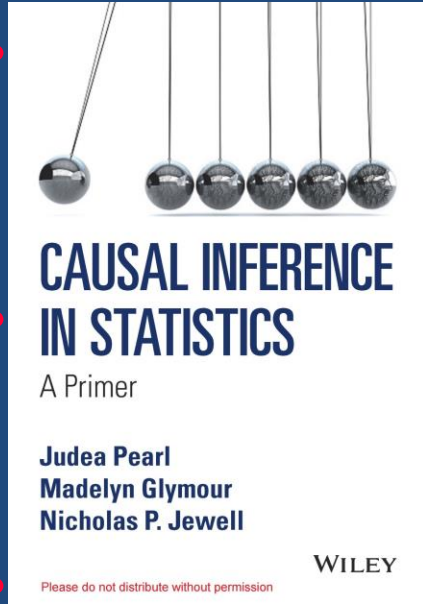


Overview

- First Period: Prediction vs causation (15 min)
If you don't have time to do it right,
When will you have time to do it over?
(John Wooden)
- Second Period: Introduction to DAGs (20 min)
- Third Period: An introduction to Dagitty (10 min)



Overview



Prediction – $\text{VO}_{2\text{max}}$ (regression)

The Bruce Protocol Treadmill Test for Athletes

A Fitness Evaluation Used to Measure VO_2 Max

The Bruce Protocol Formula for Estimating VO_2 Max

These are the formulas used:

- For men VO_2 max = $14.8 - (1.379 \times T) + (0.451 \times T^2) - (0.012 \times T^3)$
- For women VO_2 max = $4.38 \times T - 3.9$
- T = Total time on the treadmill measured as a fraction of a minute (a test time of 9 minutes 30 seconds would be written as T=9.5).

Prediction – Image Recognition

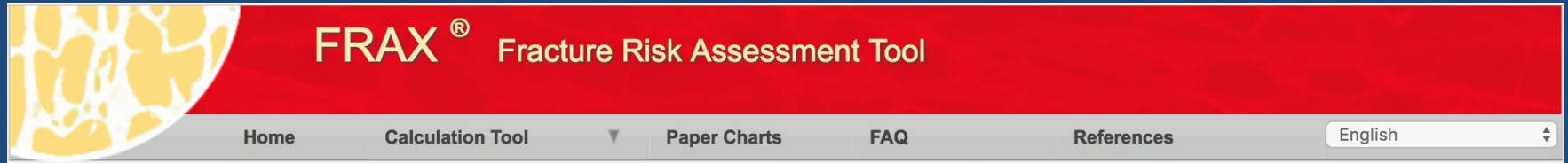


"A refrigerator filled with lots of food and drinks"



"A yellow school bus parked in a parking lot"

Prediction – Fracture Risk (regression)



Country: **Canada** Name/ID: david [About the risk factors](#)

Questionnaire:

1. Age (between 40 and 90 years) or Date of Birth
Age: Date of Birth: Y: M: D:

2. Sex Male Female

3. Weight (kg)

4. Height (cm)

5. Previous Fracture No Yes

6. Parent Fractured Hip No Yes

7. Current Smoking No Yes

8. Glucocorticoids No Yes

9. Rheumatoid arthritis No Yes

10. Secondary osteoporosis No Yes

11. Alcohol 3 or more units/day No Yes

12. Femoral neck BMD (g/cm²)
Select BMD

BMI: 26.6
The ten year probability of fracture (%)

Major osteoporotic	4.4
Hip Fracture	0.4

Country: **Canada** Name/ID: david [About the risk factors](#)

Questionnaire:

1. Age (between 40 and 90 years) or Date of Birth
Age: Date of Birth: Y: M: D:

2. Sex Male Female

3. Weight (kg)

4. Height (cm)

5. Previous Fracture No Yes

6. Parent Fractured Hip No Yes

7. Current Smoking No Yes

8. Glucocorticoids No Yes

9. Rheumatoid arthritis No Yes

10. Secondary osteoporosis No Yes

11. Alcohol 3 or more units/day No Yes

12. Femoral neck BMD (g/cm²)
Select BMD

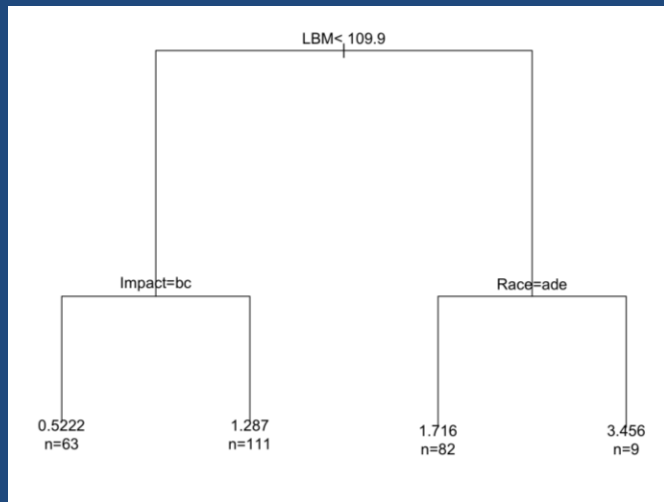
BMI: 26.6
The ten year probability of fracture (%)

Major osteoporotic	3.7
Hip Fracture	0.3

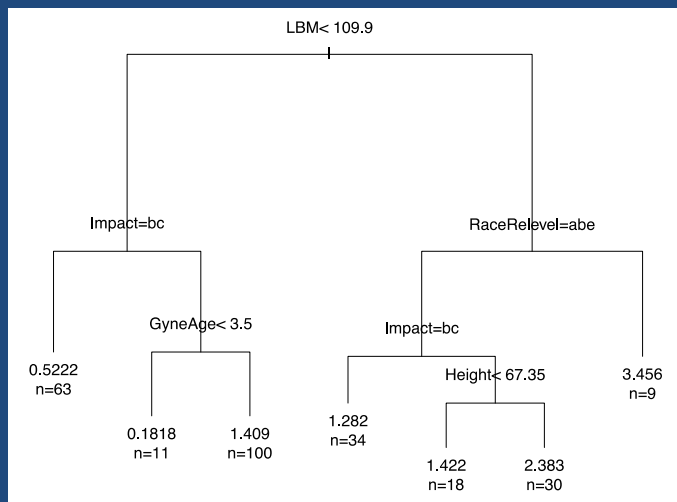
Prediction – Total Body BMD (CART)

Frame First menses Gyne Age Infrequent menses BCP	Fam Hx Stress # Fam Hx Osteopenia Prev Stress # Impact Activity Race	Weight Fat Mass Lean Body Mass AG Ratio BMI
---	--	---

Cut = 0.5



Cut = 0.25



Generalizability from a study requires Causal approach



Prediction – Fracture Risk (regression)

FRAX[®] Fracture Risk Assessment Tool

Home Calculation Tool Paper Charts FAQ

Country: Canada Name/ID: david About the risk factors

Questionnaire:

- Age (between 40 and 90 years) or Date of Birth
Age: 58 Y: 1960 M: D:
- Sex Male Female
- Weight (kg) 76.8
- Height (cm) 170
- Previous Fracture No Yes
- Parent Fractured Hip No Yes
- Current Smoking No Yes
- Glucocorticoids No Yes
- Rheumatoid arthritis No Yes
- Secondary osteoporosis No Yes
- Alcohol 3 or more units/day No Yes
- Femoral neck BMD (g/cm²)
Select BMD

BMI: 26.6
The ten year probability of fracture (%)

without BMD	
Major osteoporotic	3.7
Hip Fracture	0.3

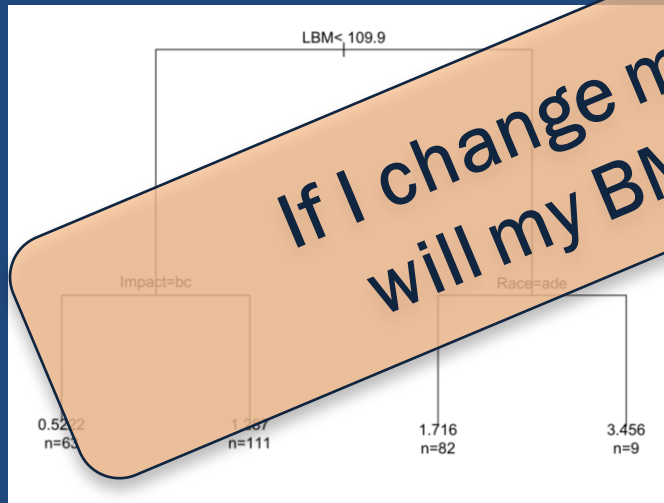
If I lower my alcohol consumption, will my risk of fracture decrease?

<https://www.sheffield.ac.uk/FRAX/tool.asp?country=19>

Prediction – Total Body BMD (CART)

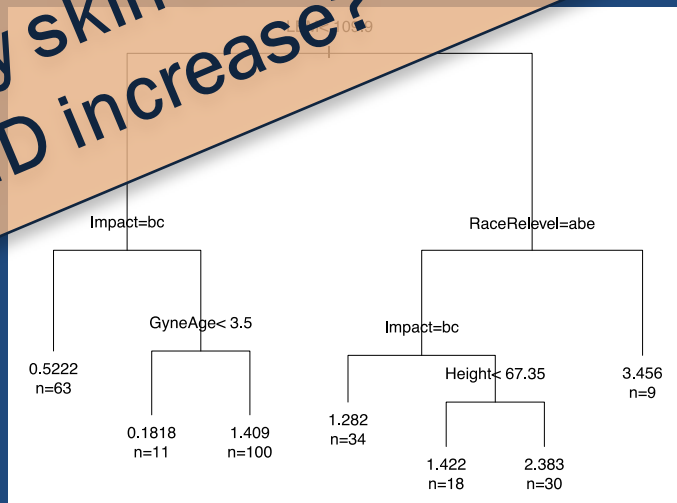
Frame	Fam Hx Stress #	Weight
First menses	Fam Hx Osteopenia	Fat Mass
Gyne Age	Prev Stress #	Lean Body Mass
Infrequent menses	Impact Activity	AG Ratio
BCP	Race	BMI

Cut = 0.5

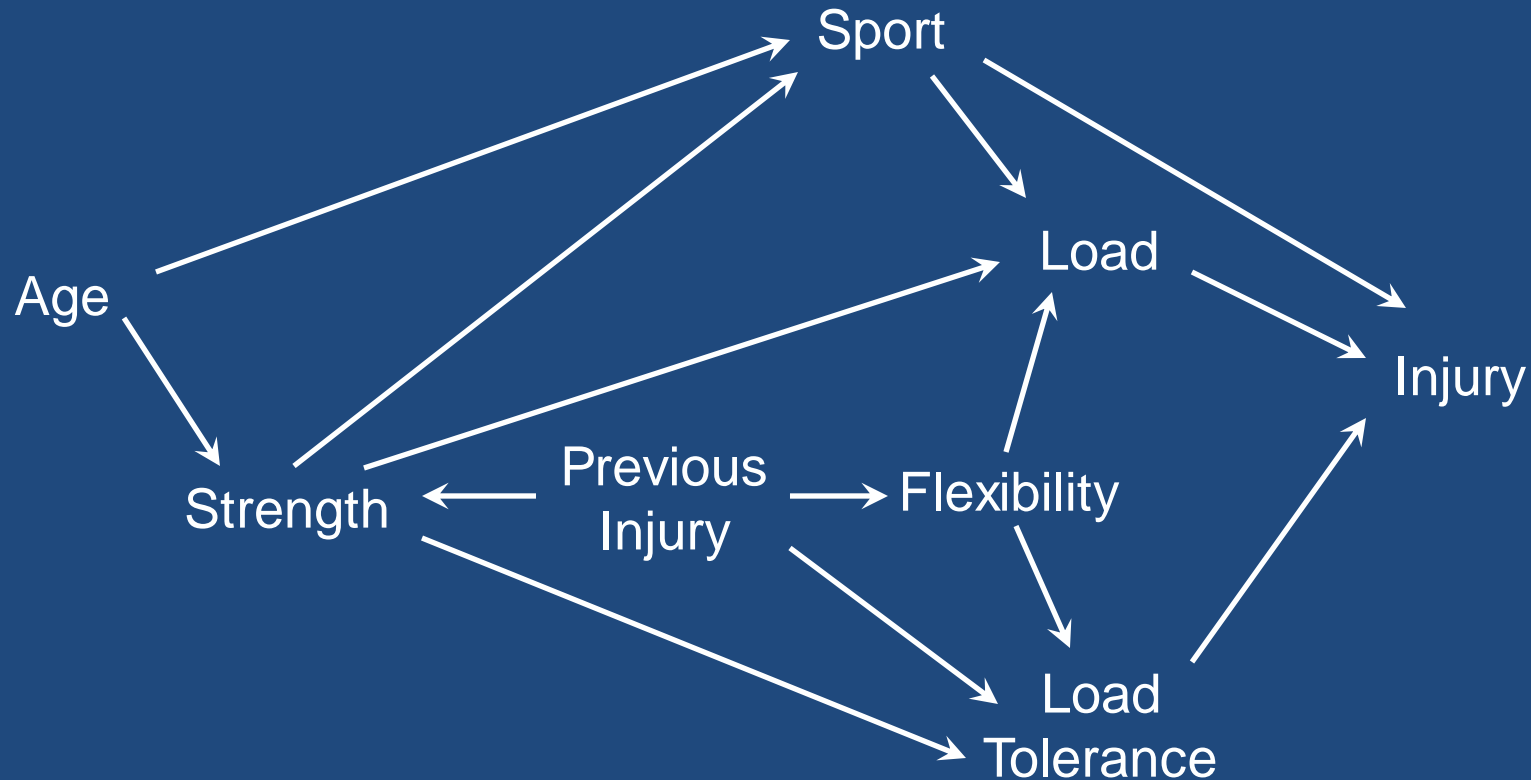


**If I change my skin colour,
will my BMD increase?**

Cut = 0.25



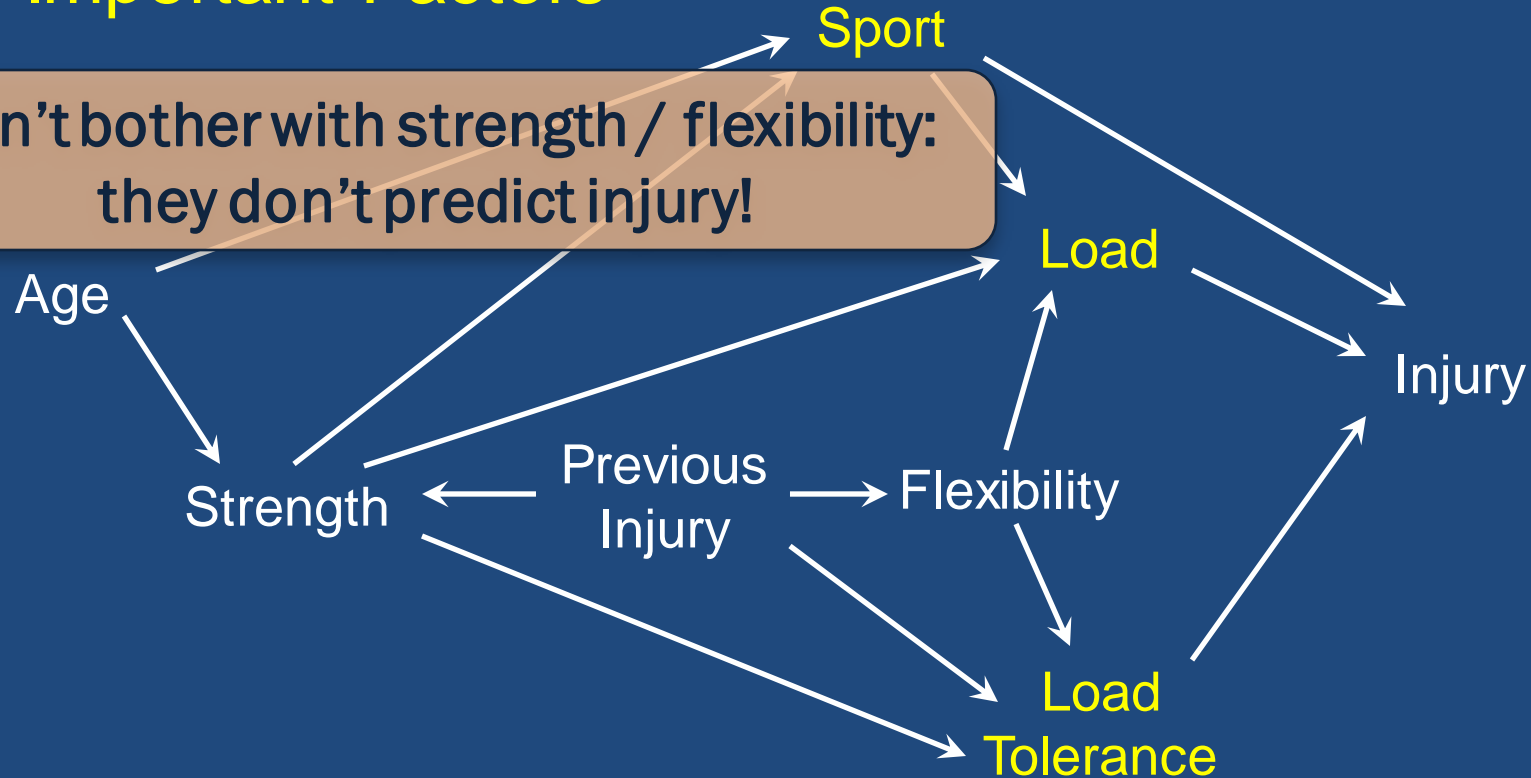
Prediction – Risk Factors for Injury



Prediction – Risk Factors for Injury

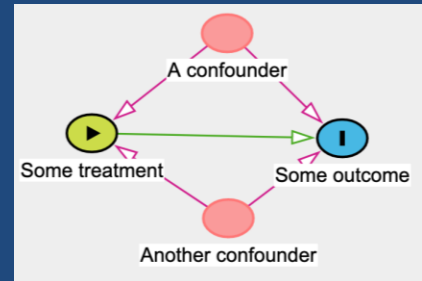
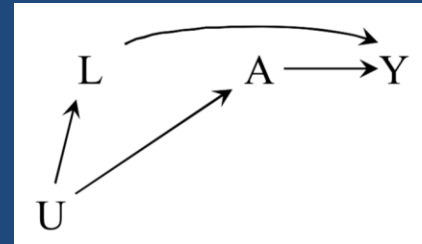
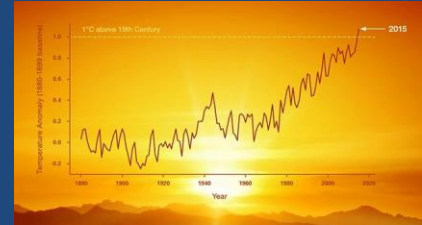
AI: Important Factors

Don't bother with strength / flexibility:
they don't predict injury!



Overview

- First Period: Prediction vs causation (15 min)
- Second Period: Introduction to DAGs (20 min)
- Third Period: An introduction to Dagitty (10 min)



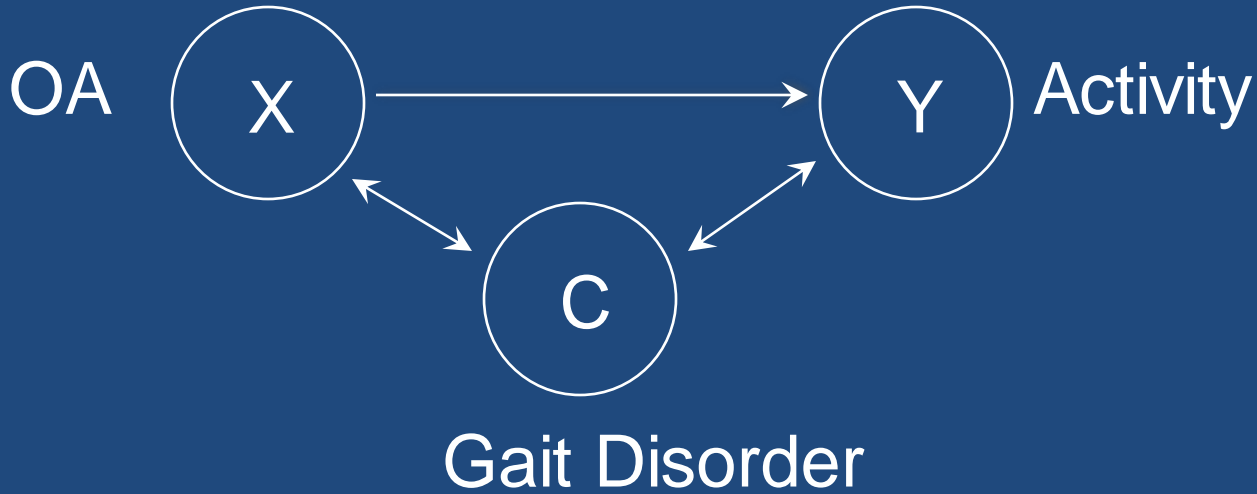
Identifiability vs Estimation

- Causal DAGs are part of a method to determine if a causal effect is “identifiable”
- A causal effect is identifiable if there is one and only one possible answer from the observed data to predict the outcome
- The \sqrt{X} is not identifiable:
 - $\Rightarrow \sqrt{4} = (2)^2$
 - $\Rightarrow \sqrt{4} = (-2)^2$
- Estimation may or may not be challenging

“Standard” Confounder

- A variable may (i.e. potential confounder) affect the magnitude or direction of the estimated effect if it:
 - ⇒ Is associated with the exposure
 - ⇒ Is associated with the outcome independent of exposure
 - ⇒ Is not affected by exposure: in particular, it does not lie on the hypothesized causal path

Confounder?



“Standard” Confounder

- A variable may affect the magnitude or direction of the estimated effect if it:
 - ⇒ Is associated with the exposure
 - ⇒ Is associated with the outcome independent of exposure
 - ⇒ Is not affected by exposure: in particular, it does not lie on the hypothesized causal path
- Must also
 - ⇒ Not be a marker for a variable on the causal path
 - ⇒ Not be caused by the outcome
 - ⇒ Cause disease in the “unexposed” group
 - ⇒ Yield true risk of disease in exposed and unexposed when included in the model

“Standard” Confounder



⇒ Must cause the exposure, or be a marker for a cause of the exposure



- ⇒ Not be caused by the outcome
- ⇒ Not be affected by the exposure
- ⇒ Not be a marker for a variable affected by exposure

“Standard” Confounder



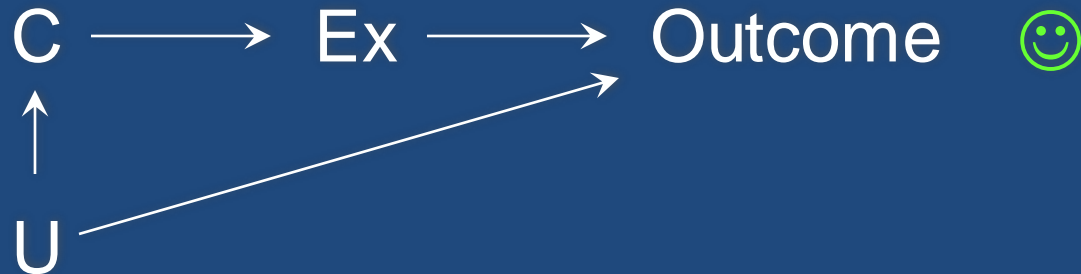
Directed Acyclic Graph (causal diagram)

Potential Confounder?



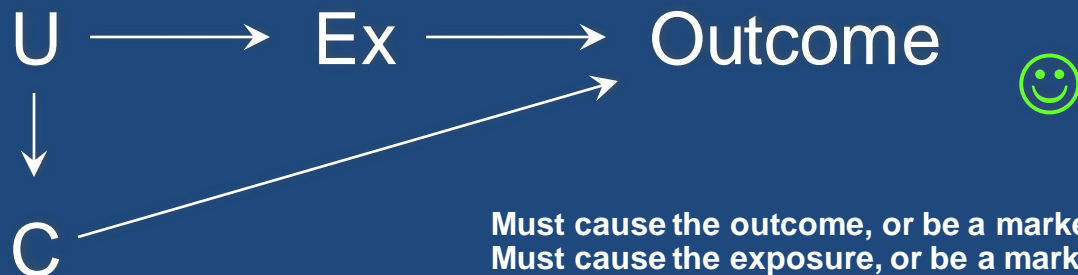
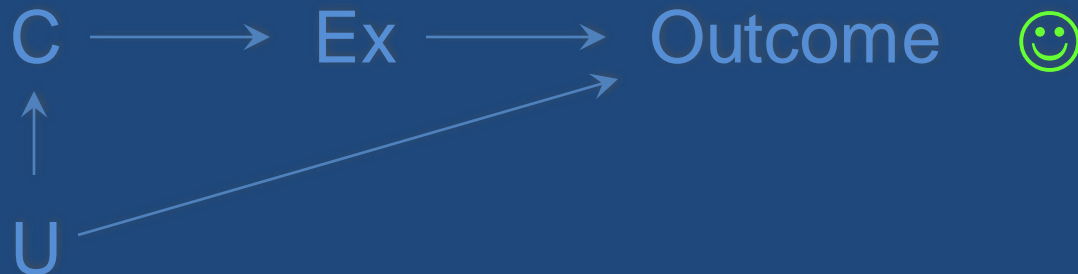
Must cause the outcome, or be a marker for a cause of the outcome
Must cause the exposure, or be a marker for a cause of the exposure

Potential Confounder?



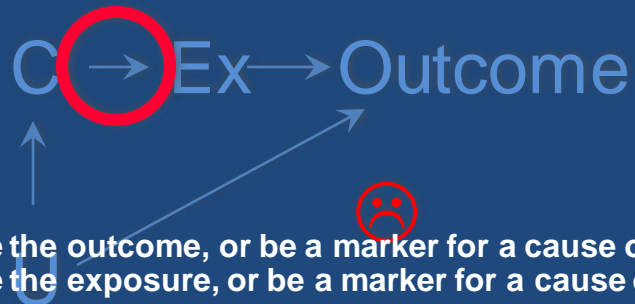
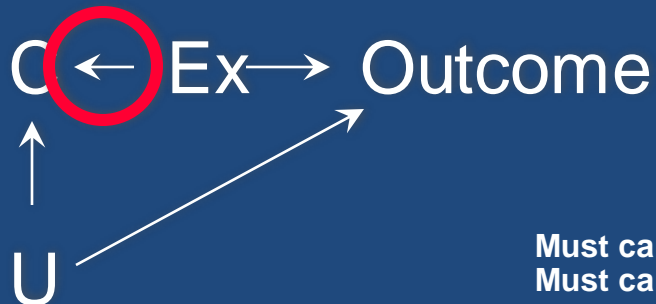
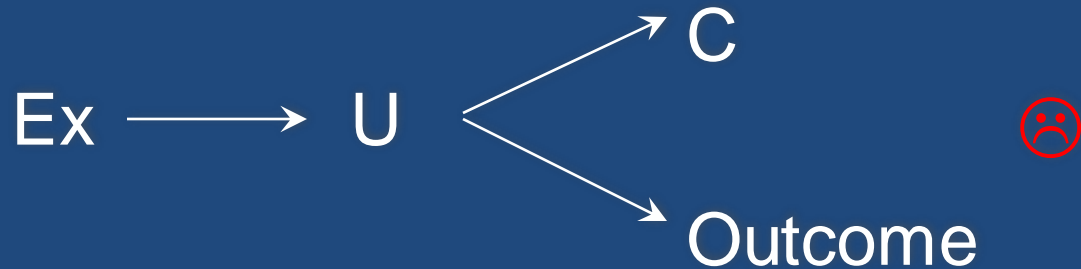
Must cause the outcome, or be a marker for a cause of the outcome
Must cause the exposure, or be a marker for a cause of the exposure

Potential Confounder?



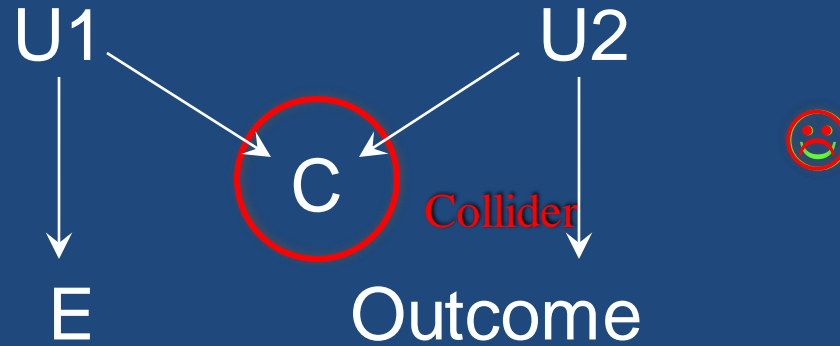
Must cause the outcome, or be a marker for a cause of the outcome
Must cause the exposure, or be a marker for a cause of the exposure

Potential Confounder?



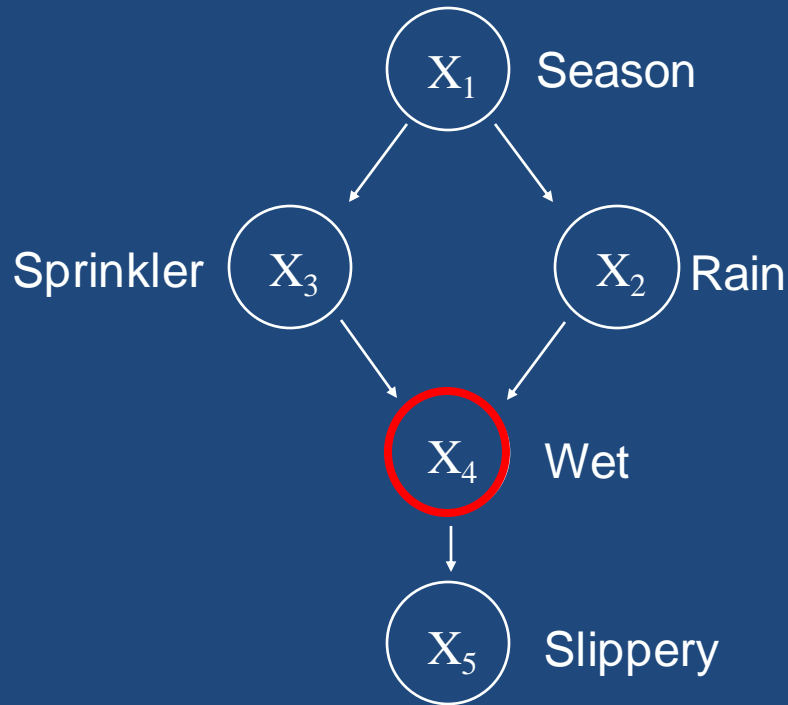
Must cause the outcome, or be a marker for a cause of the outcome
Must cause the exposure, or be a marker for a cause of the exposure

Potential Confounder?



Must cause the outcome, or be a marker for a cause of the outcome
Must cause the exposure, or be a marker for a cause of the exposure

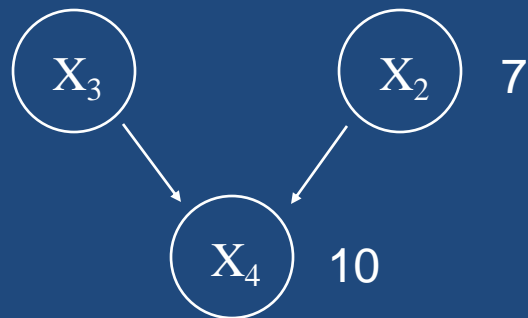
Pearl's Rules - Explanation



If one knows the value of the “collider”, the parents are associated.

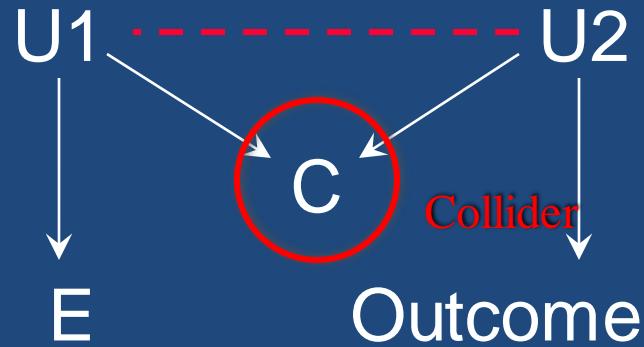
If wet: the sprinkler is more likely to be on if there was no rain.

Pearl's Rules - Explanation



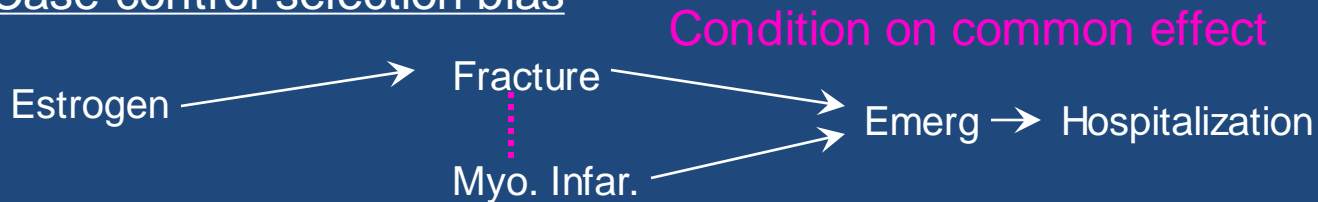
$$X_4 = X_2 + X_3$$

Potential Confounder vs. Collider?

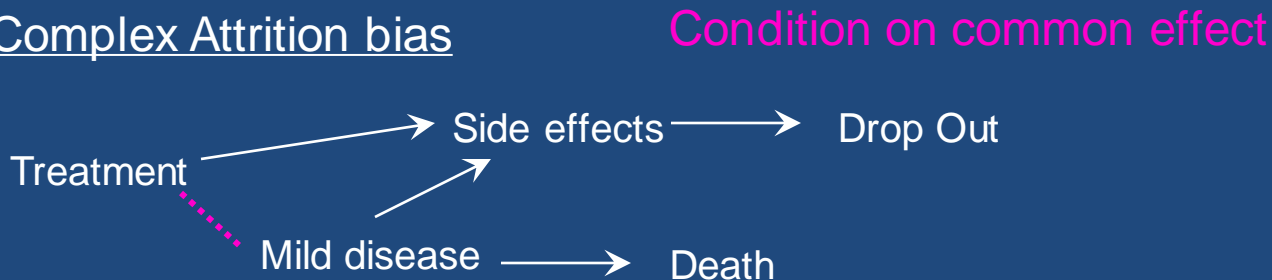


Collider-Stratification Bias

Case-control selection bias



Complex Attrition bias



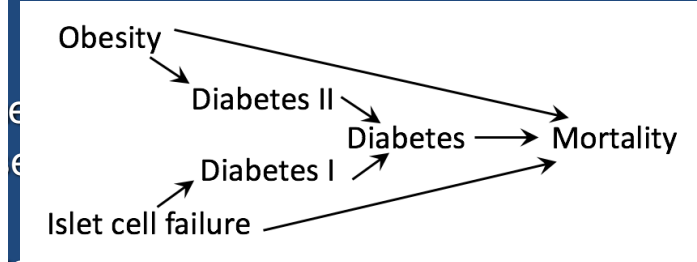
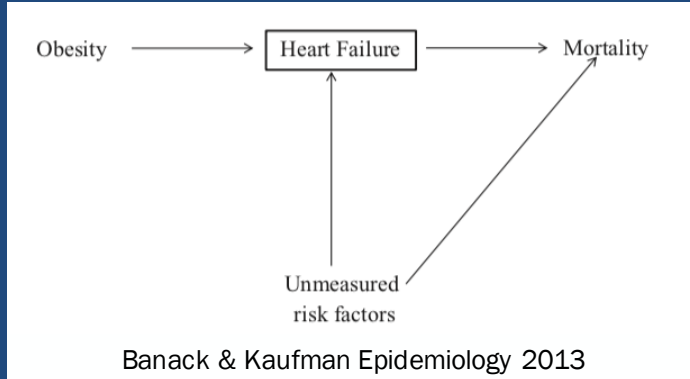
Obesity Paradox

1. Obesity causes heart failure
2. Heart failure causes death
3. Obesity causes death independent of heart failure

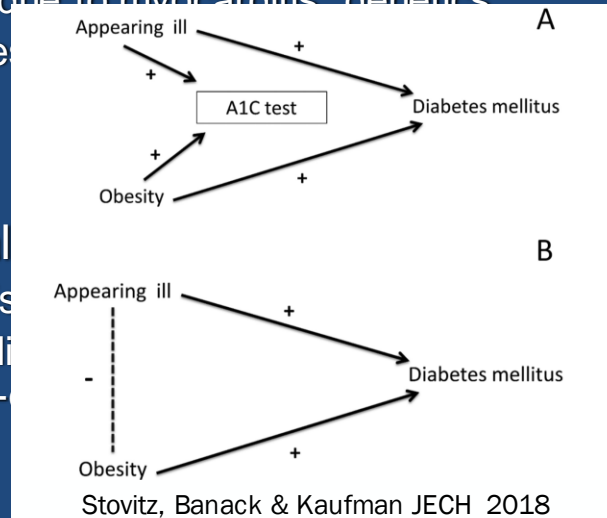
The risk of death in obese patients with heart failure is less than the risk of death in non-obese patients with heart failure.

Draw a causal diagram

Obesity Paradox



⇒ Obesity have milder form of HF compared to HF due to myocarditis, genetics



- Diabetes Anal

⇒ Obesity is

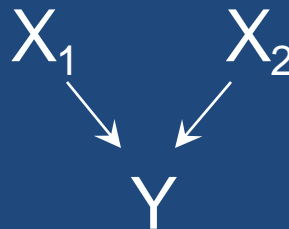
⇒ Among di
than non-

better

(non-insulin) do better

What does \rightarrow on a DAG mean?

- Graph Theory (for directed acyclic graph)
 - \Rightarrow Vertices (nodes, variables) joined by directed edges (arrows)
 - \Rightarrow Parents cause descendants (children, grandchildren)
 - \Rightarrow Children are caused by ancestors
 - \Rightarrow Variables are defined in space-time: Blood pressure measured at two different times is two different variables



What does \rightarrow on a DAG mean?

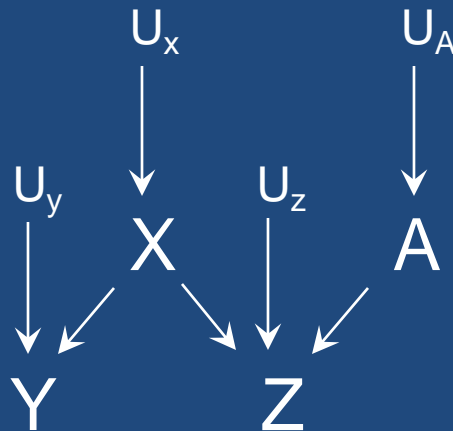
- Structural Causal Models

\Rightarrow Child is *some* function of parents

$\Rightarrow f_x: X = u_x$

$\Rightarrow f_y: Y = f(u_y, X)$

$\Rightarrow f_z: Z = f(u_z, X, A)$



What does \rightarrow on a DAG mean?

- Structural Causal Models

$\Rightarrow f_x$

$\Rightarrow X = u_x$ (or maybe $3 \cdot u_x + 10$)

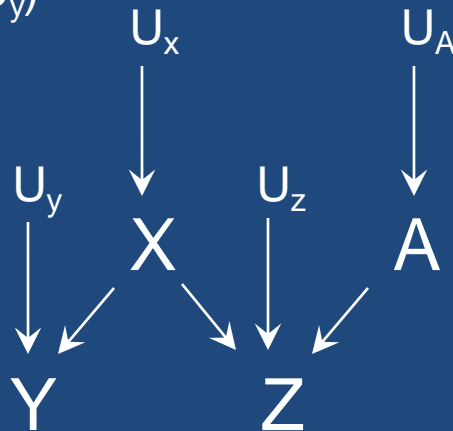
$\Rightarrow f_y:$

$\Rightarrow Y = X \cdot U_y$ (or maybe $X^2 \cdot U_y$)

$\Rightarrow f_z$

\Rightarrow If $X=0$, $Z=U_z \cdot 2$

\Rightarrow If $X=1$, $Z=A \cdot X$



What does \rightarrow on a DAG mean?

- Conditional Independence

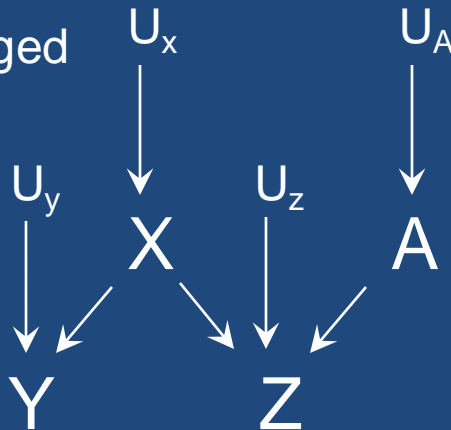
\Rightarrow Absence of an arrow means variable is NOT a function of the “parent” – independent

\Rightarrow A is independent of X

\Rightarrow Z is independent of Y given X (fix X)

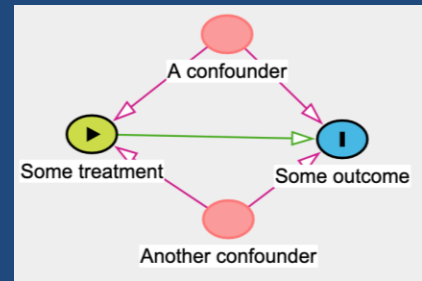
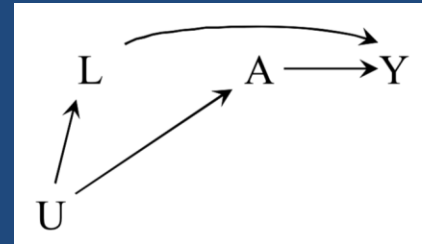
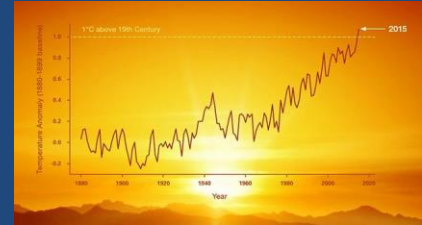
\Rightarrow Z does not change if Y is changed

\Rightarrow Z is not caused by Y

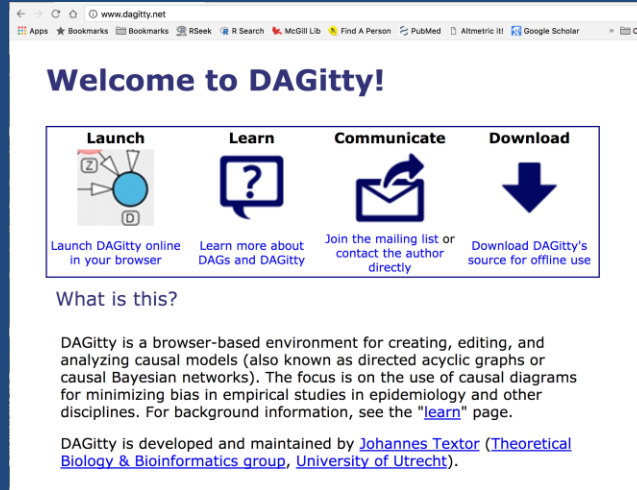


Overview

- First Period: Prediction vs causation (15 min)
- Second Period: Introduction to DAGs (20 min)
- Third Period: An introduction to Dagitty (10 min)



Dagitty



The screenshot shows the homepage of the Dagitty website. At the top, there is a navigation bar with various search engines and tools like RSeek, R Search, McGill Lib, Find A Person, PubMed, Altmetric It, and Google Scholar. Below the navigation bar, the main heading reads "Welcome to DAGitty!". Underneath, there are four columns of action items: "Launch" with a blue circular node icon and the text "Launch DAGitty online in your browser"; "Learn" with a question mark icon and the text "Learn more about DAGs and DAGitty"; "Communicate" with an envelope icon and the text "Join the mailing list or contact the author directly"; and "Download" with a downward arrow icon and the text "Download DAGitty's source for offline use". Below these columns, the text "What is this?" is followed by a paragraph explaining that Dagitty is a browser-based environment for creating, editing, and analyzing causal models (directed acyclic graphs or causal Bayesian networks). It mentions the focus on minimizing bias in empirical studies in epidemiology and other disciplines, and refers to a "learn" page for background information. At the bottom, it states that Dagitty is developed and maintained by Johannes Textor (Theoretical Biology & Bioinformatics group, University of Utrecht).


- Coloring
- Adjustment Sets
- Testable Implications
- Model Code

Humans: What would happen if...

Pearl's 3rd Rung: Imagining / counterfactuals

**Now At Number #2 We
have Great Middle
School Football Play
"This Kid Was
Untouched"**


Sufficient Causal Sets

 American Journal of Epidemiology
© The Author 2007. Published by the Johns Hopkins Bloomberg School of Public Health.
All rights reserved. For permissions, please e-mail: journals.permissions@oxfordjournals.org.

Vol. 166, No. 9
DOI: 10.1093/aje/kwm179
Advance Access publication August 16, 2007

Practice of Epidemiology

Directed Acyclic Graphs, Sufficient Causes, and the Properties of Conditioning on a Common Effect



Tyler J. VanderWeele¹ and James M. Robins²

Eur J Epidemiol (2011) 26:347–357
DOI 10.1007/s10654-011-9568-3

METHODS

Identification of operating mediation and mechanism in the sufficient-component cause framework

Etsuji Suzuki · Eiji Yamamoto · Toshihide Tsuda

 American Journal of Epidemiology
© The Author 2011. 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.

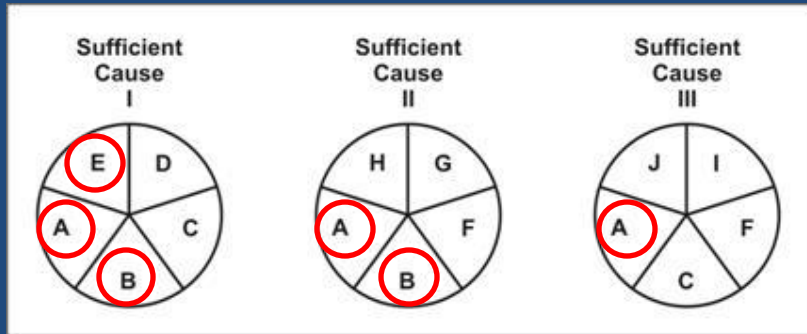
Vol. 185, No. 11
DOI: 10.1093/aje/kwr083
Advance Access publication: May 23, 2011

Invited Commentary

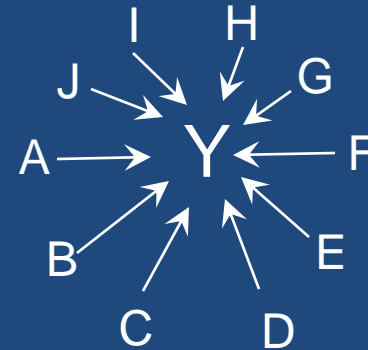
Invited Commentary: The Continuing Need for the Sufficient Cause Model Today

Tyler J. VanderWeele*

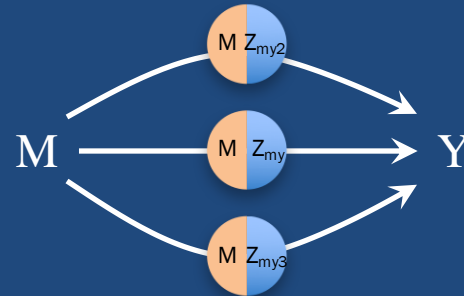
(Rothman AJE 1976)



Understanding Causes



Suff. Causes Sets and Graphical Diagrams



Suff. Causes Sets and Graphical Diagrams

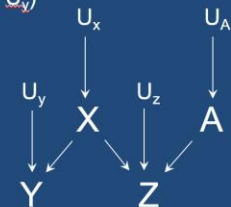


- What is the causal effect of M on Y?

Unknown from the diagram alone

- Structural Causal Models

- ⇒ f_x
⇒ $X = u_x$ (or maybe $3 \cdot u_x + 10$)
- ⇒ f_y
⇒ $Y = X \cdot u_y$ (or maybe $X^2 \cdot u_y$)
- ⇒ f_z
⇒ If $X=0$, $Z=U_z \cdot 2$
⇒ If $X=1$, $Z=A \cdot X$



Suff. Causes Sets and Graphical Diagrams

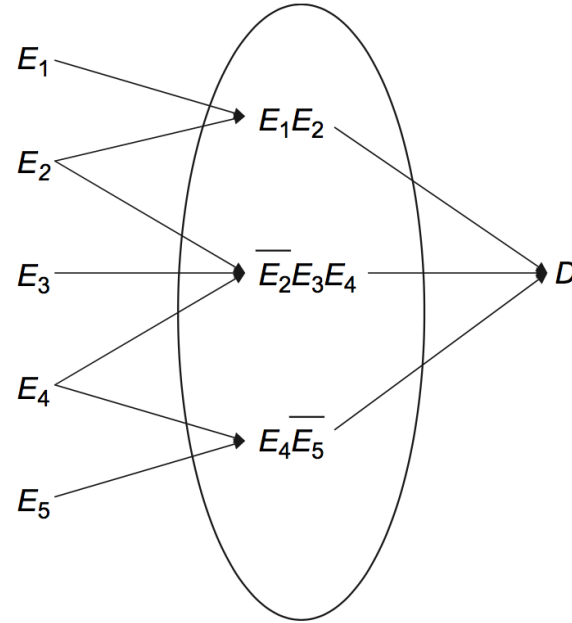
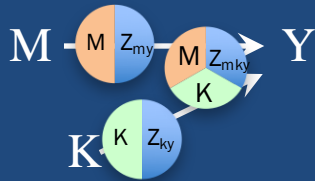


FIGURE 3. A causal directed acyclic graph with a sufficient causation structure.

Summary

- Dagitty makes life much easier
- Collider-stratification bias may be more common than believed (Obesity paradox, Hamstring paradox)
- Causal arrows examined through the lens of:
 - ⇒ Conditional independence
 - ⇒ Sufficient causal sets