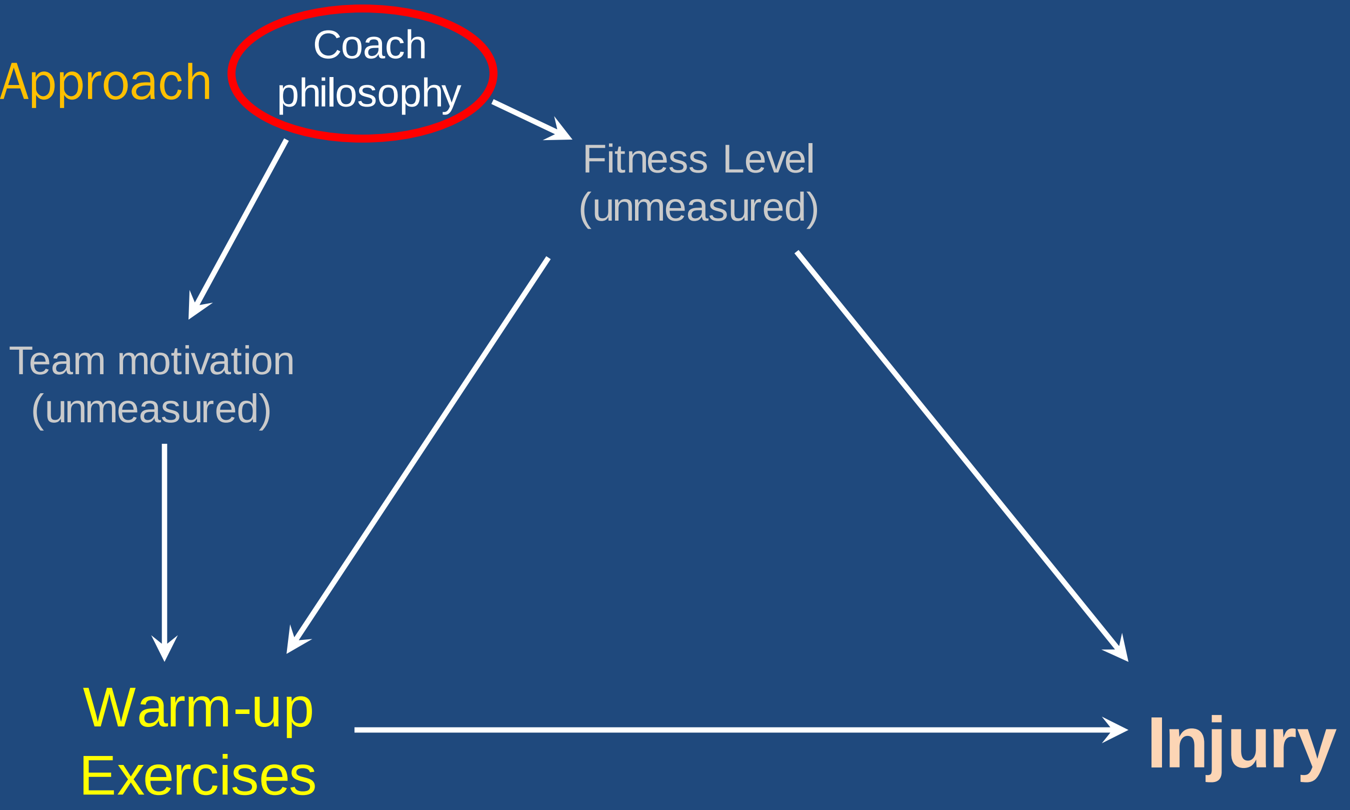
Causal Diagrams:
A tool to better understand results from epidemiological and machine learning analyses

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

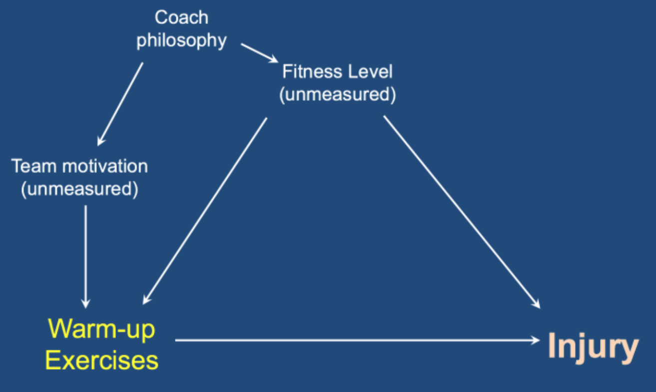
AI versus Causal Inference

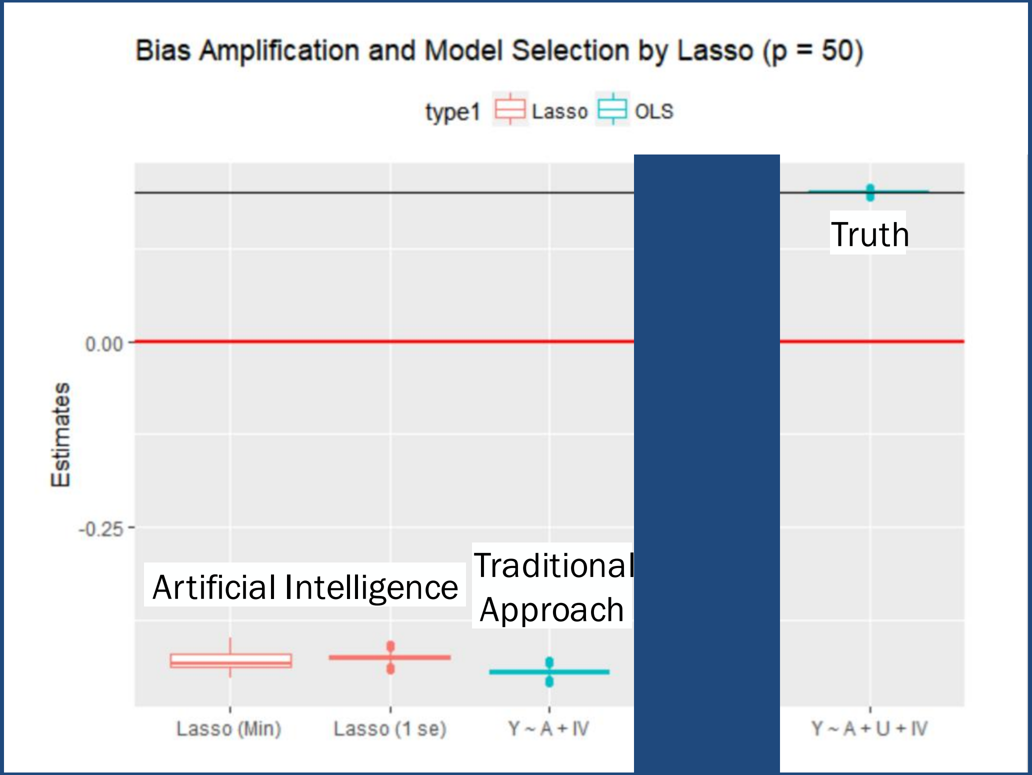
Both Traditional & AI Approach



(Stokes, Stat Methods Med Res 2021)

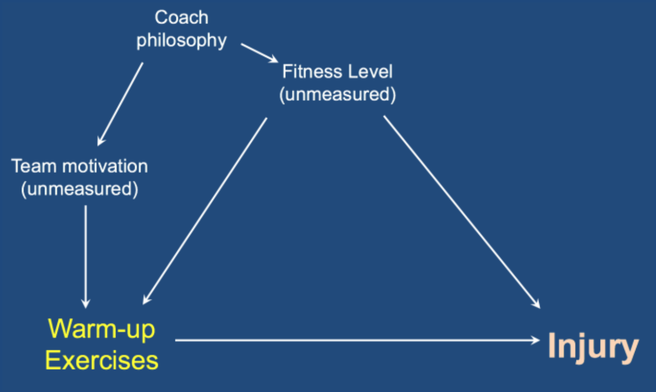
AI versus Causal Inference

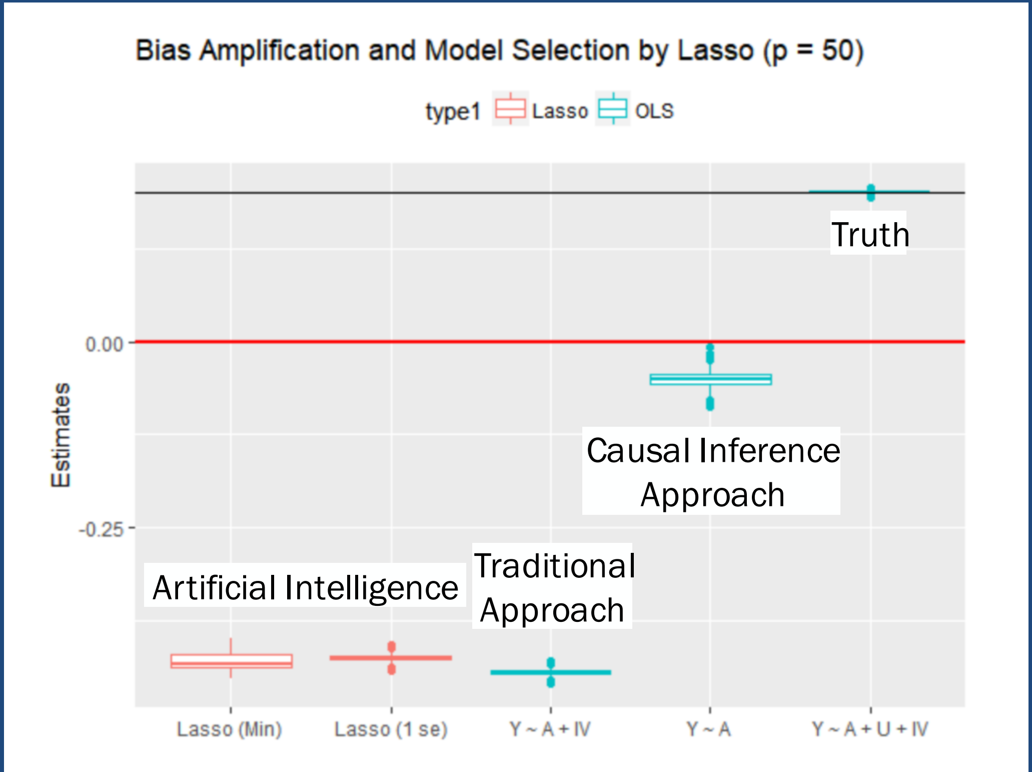




(Stokes, Stat Methods Med Res 2021)

AI versus Causal Inference



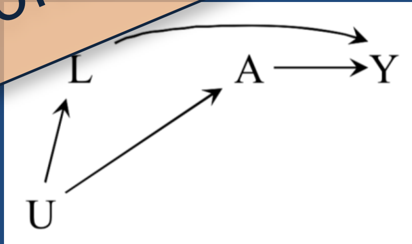


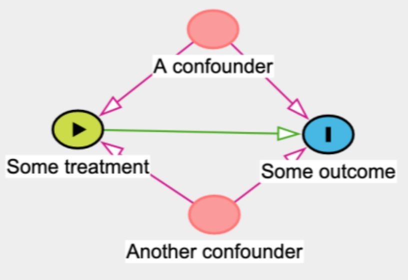
(Stokes, Stat Methods Med Res 2021)

Overview

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

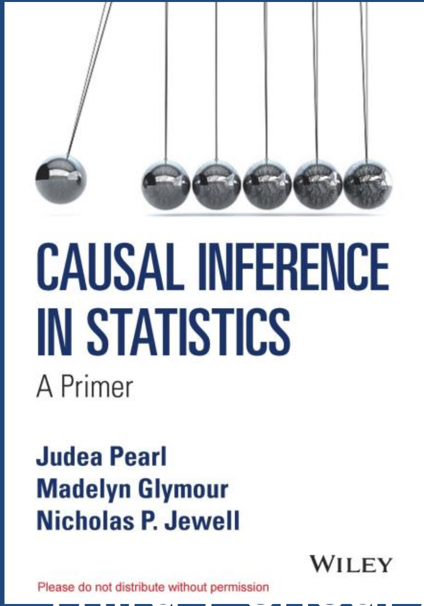
If you don’t have time to do it right,
When will you have time to do it over?
(John Wooden)

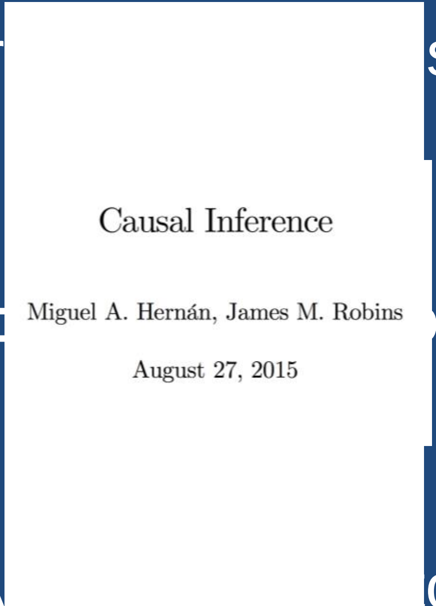


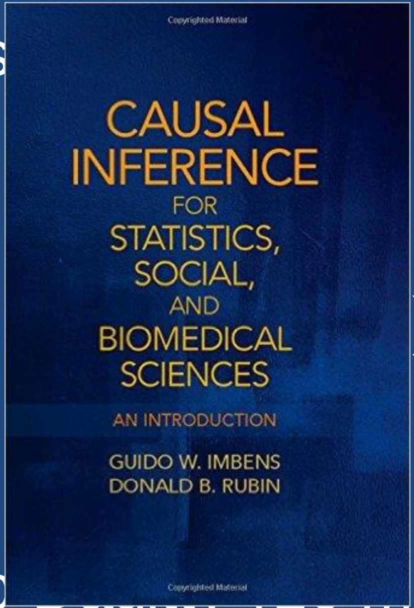


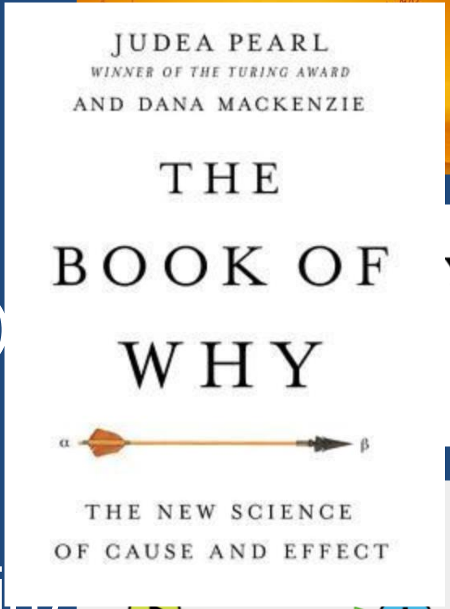
Overview

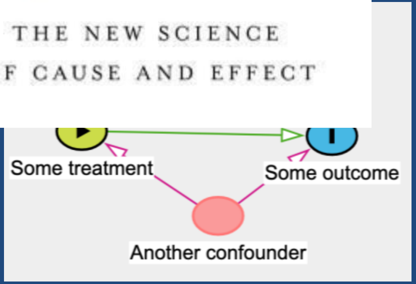
- Pr s
- d:)
- A o)











Prediction – VO_{2max} (regression)

The Bruce Protocol Treadmill Test for Athletes

A Fitness Evaluation Used to Measure VO2 Max

The Bruce Protocol Formula for Estimating VO2 Max

These are the formulas used:

- For men VO2 max = 14.8 – (1.379 x T) + (0.451 x T^2) – (0.012 x T^3)
- For women VO2 max = 4.38 x 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).

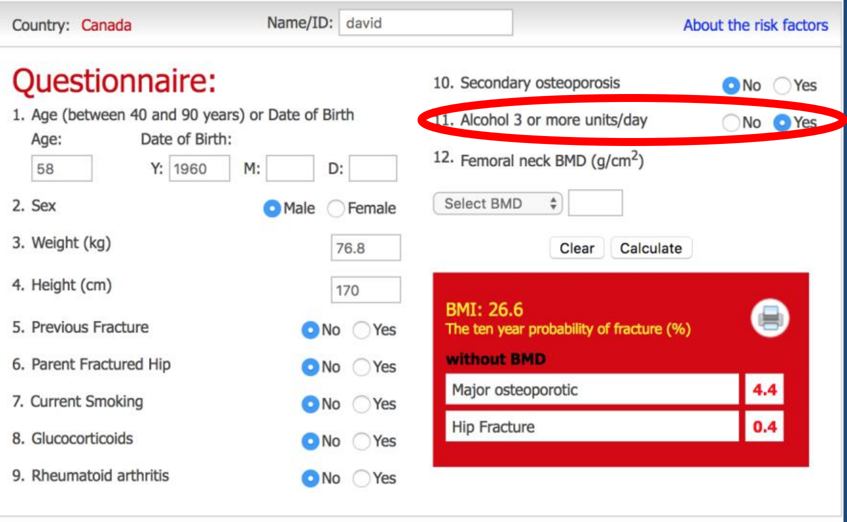
(https://www.verywellfit.com/the-bruce-treadmill-test-protocol-3120269)

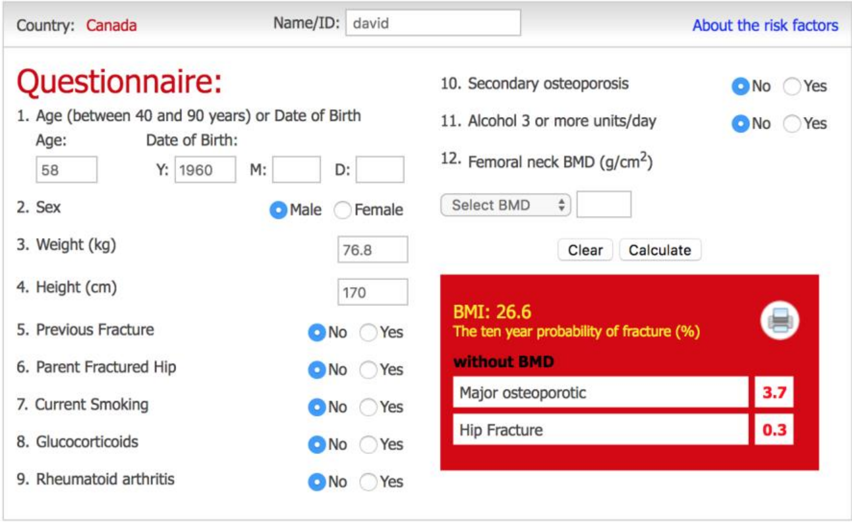
Prediction – Image Recognition

"A refrigerator filled with lots of food and drinks"

"A yellow school bus parked in a parking lot"

Prediction – Fracture Risk (regression)



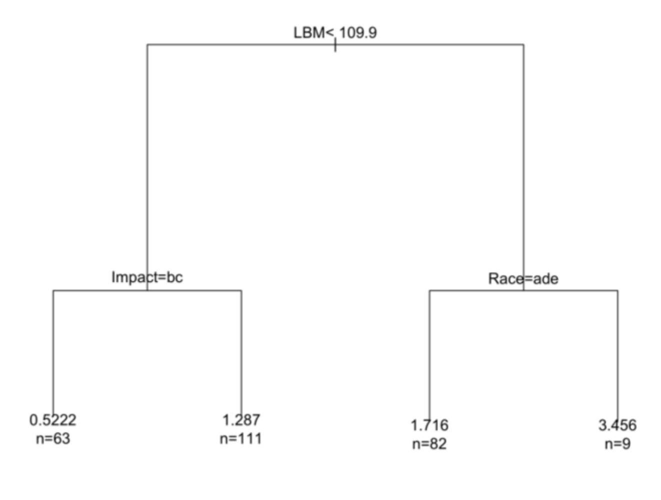


(https://www.sheffield.ac.uk/FRAX/tool.aspx?country=19)

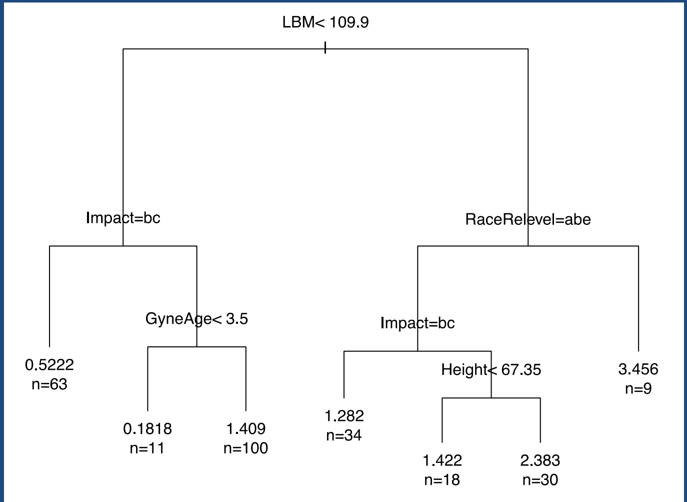
Prediction – Total Body BMD (CART)

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



Cut = 0.25

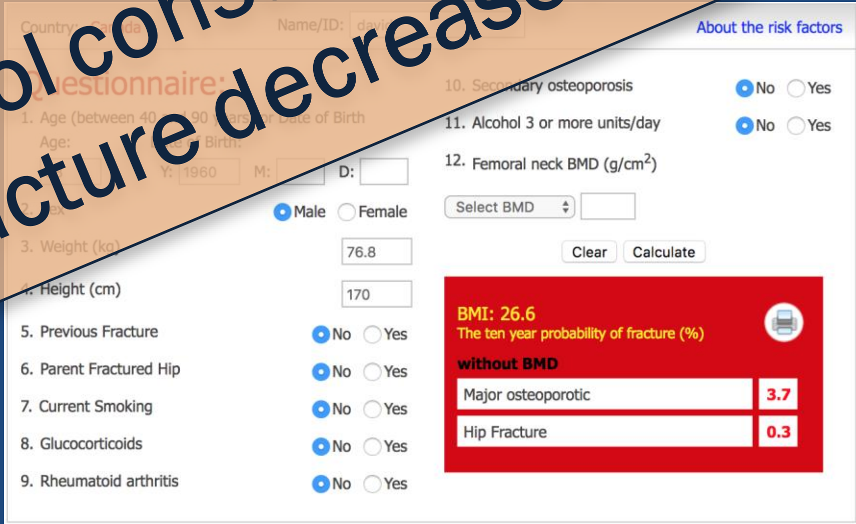


(Stanford University data, 2011)

Generalizability from a study requires Causal approach

(Bareinboim and Pearl)

Prediction – Fracture Risk (regression)



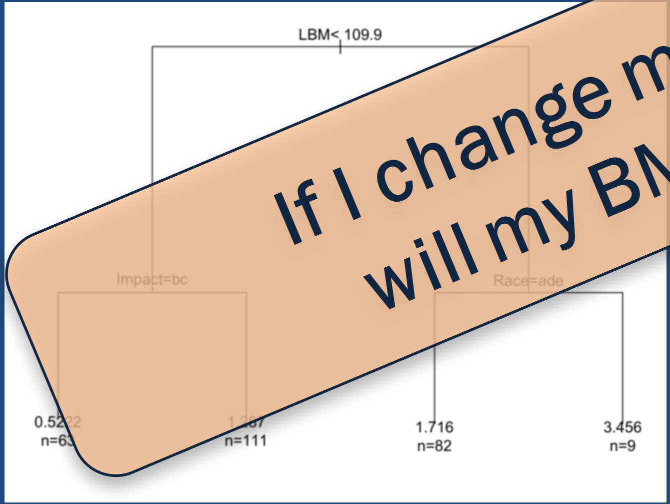
If I lower my alcohol consumption, will my risk of fracture decrease?

(https://www.sheffield.ac.uk/FRAX/tool.aspx?country=19)

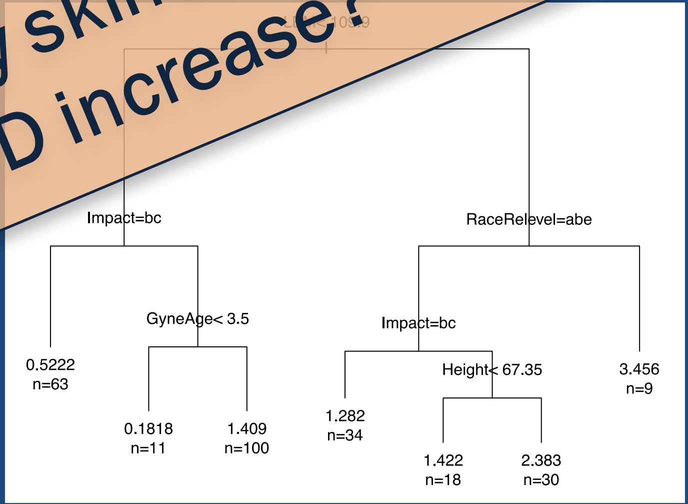
Prediction – Total Body BMD (CART)

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



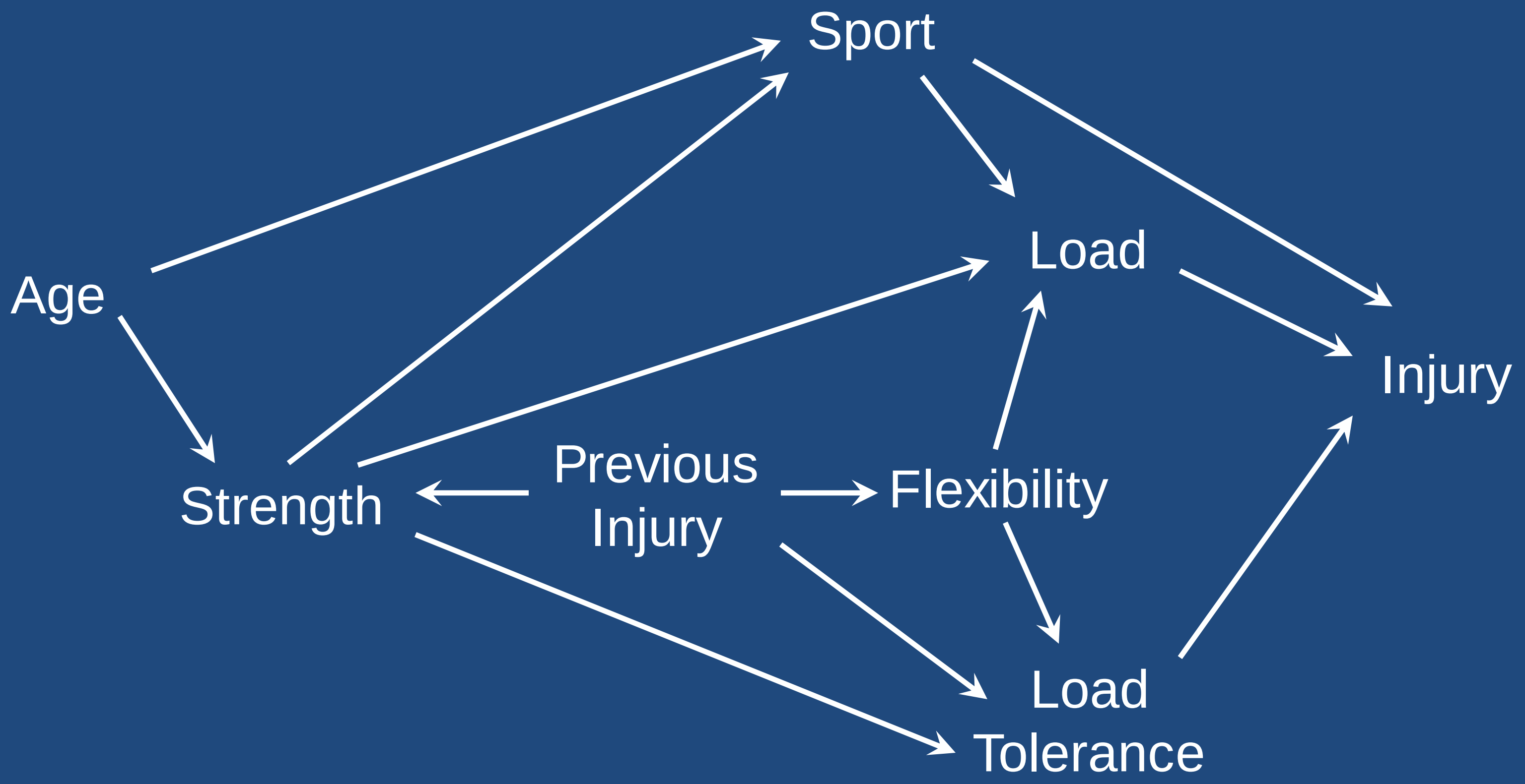
Cut = 0.25



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

(Stanford University data)

Prediction – Risk Factors for Injury

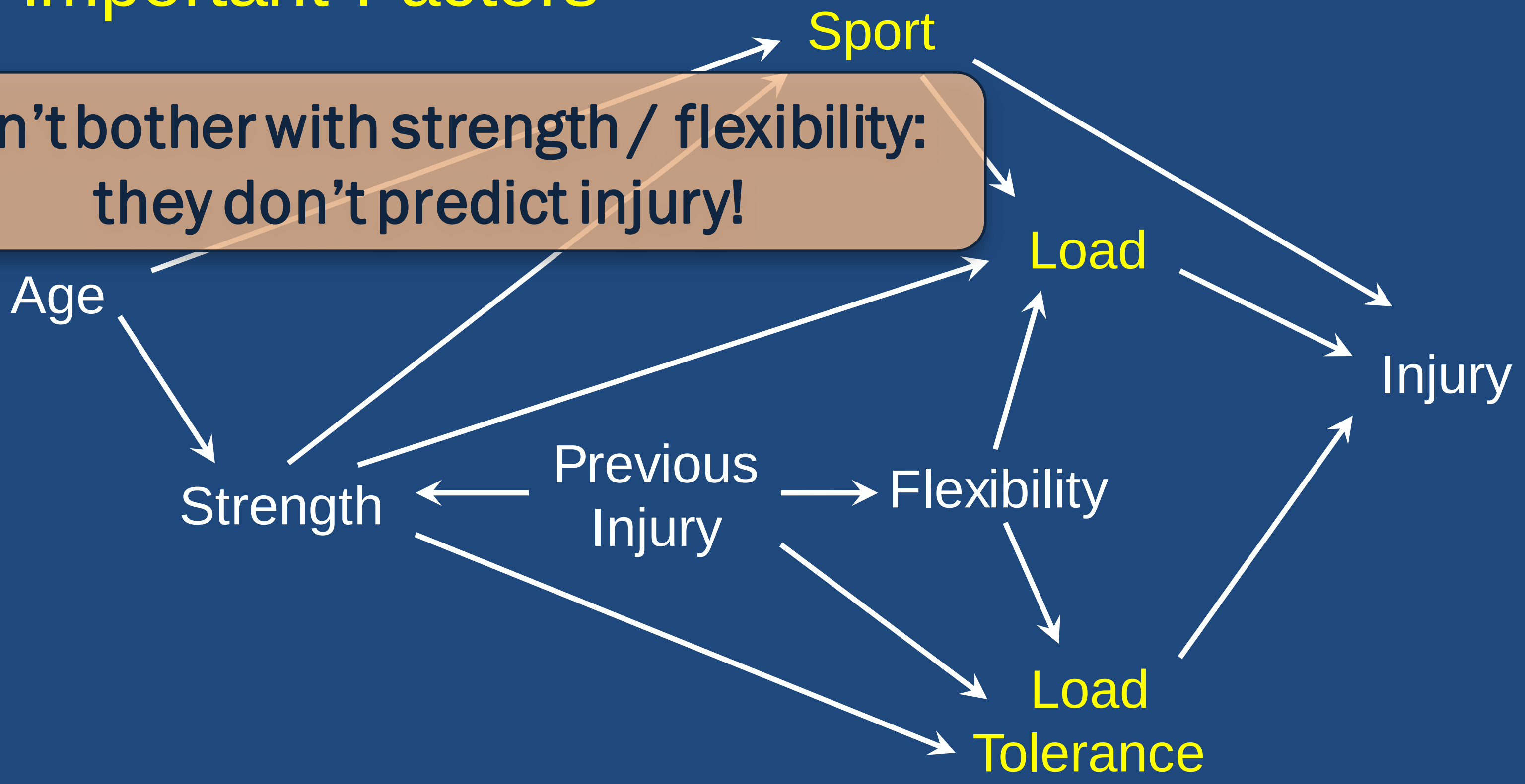


(Dutch Study, in preparation)

Prediction – Risk Factors for Injury

AI: Important Factors

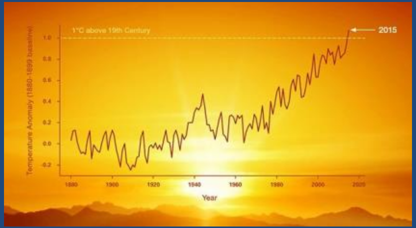
Don't bother with strength / flexibility: they don't predict injury!



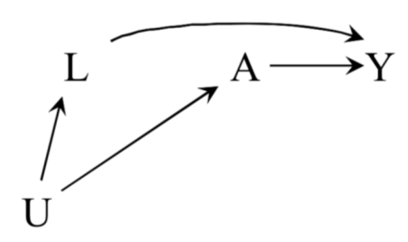
(Dutch Study, in preparation)

Overview

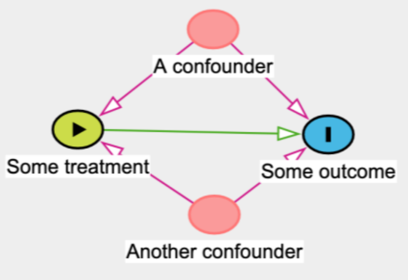
- First Period: Prediction vs causation (15 min)



- 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:

  $\Rightarrow \sqrt{4} = (2)^2$

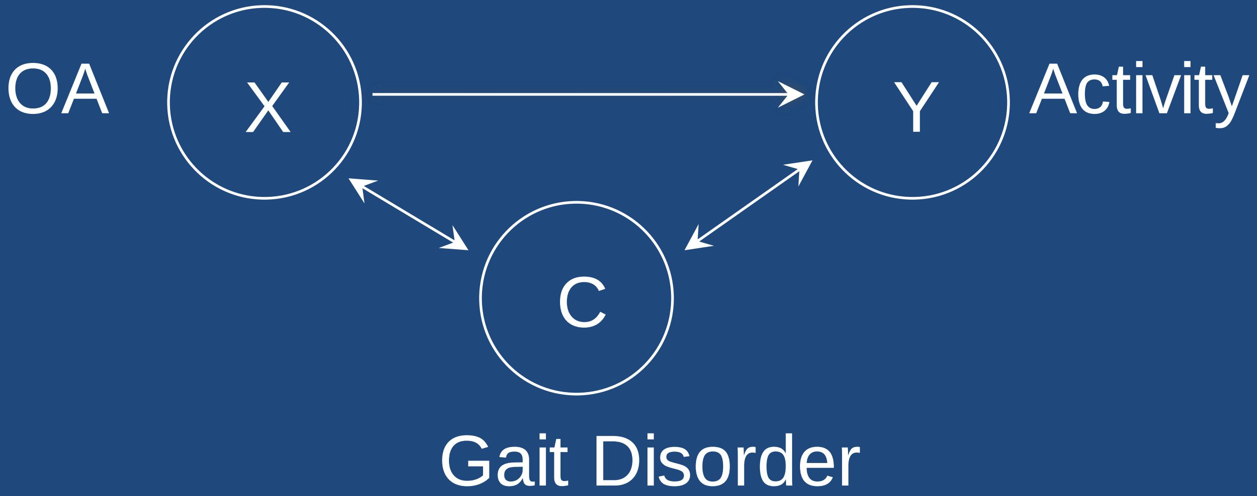
  $\Rightarrow \sqrt{4} = (-2)^2$

- 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

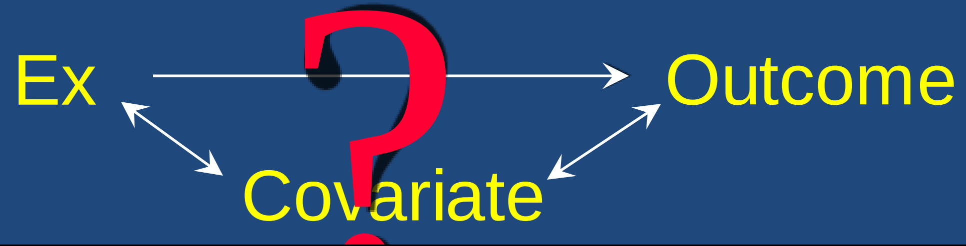
Confounder?



"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)

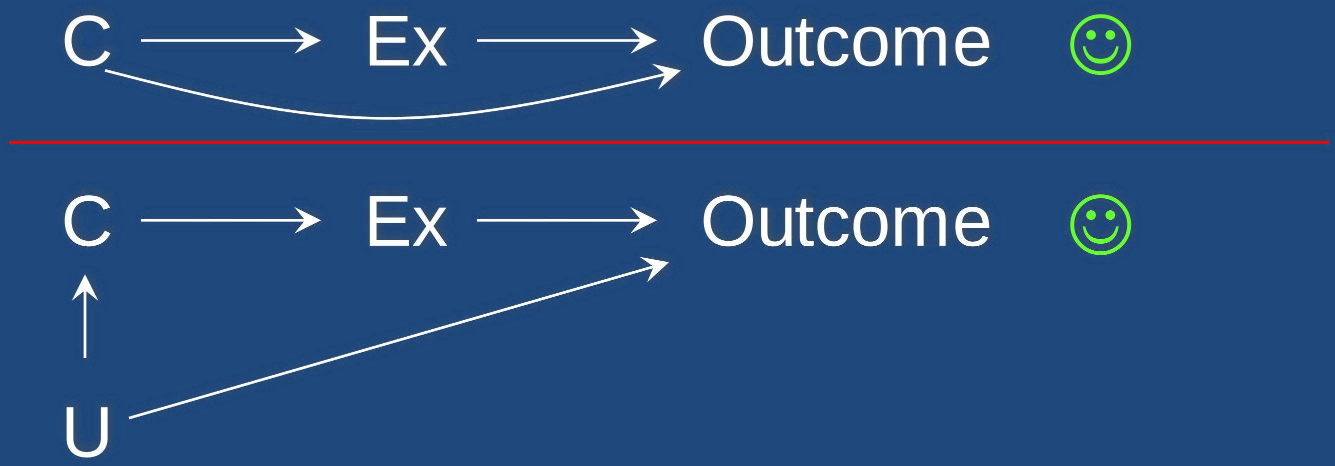
Potential Confounder?

$$\begin{tikzcd} C \arrow[r] \arrow[rr, bend right] & Ex \arrow[r] & Outcome \end{tikzcd}$$ ☺

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

(Hernán Am J Epid 2002)

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

(Hernán Am J Epid 2002)

Potential Confounder?

$$\begin{tikzcd} C \arrow[r] \arrow[rr, bend right] & Ex \arrow[r] & \text{Outcome} \end{tikzcd}$$ ☺

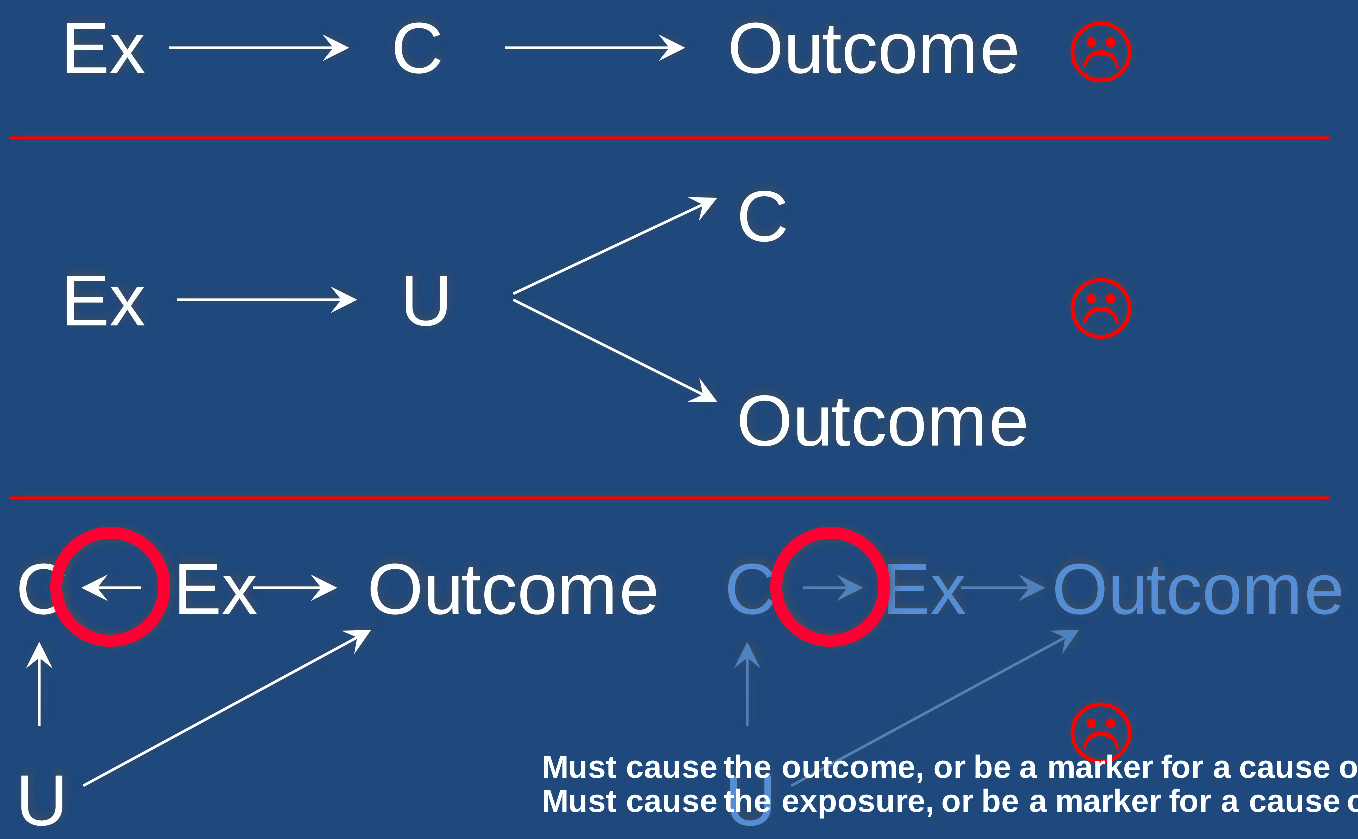
$$\begin{tikzcd} C \arrow[r] & Ex \arrow[r] & \text{Outcome} \\ U \arrow[u] \arrow[urr] & & \end{tikzcd}$$ ☺

$$\begin{tikzcd} U \arrow[r] \arrow[d] & Ex \arrow[r] & \text{Outcome} \\ C \arrow[urr] & & \end{tikzcd}$$ ☺

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

(Hernán Am J Epid 2002)

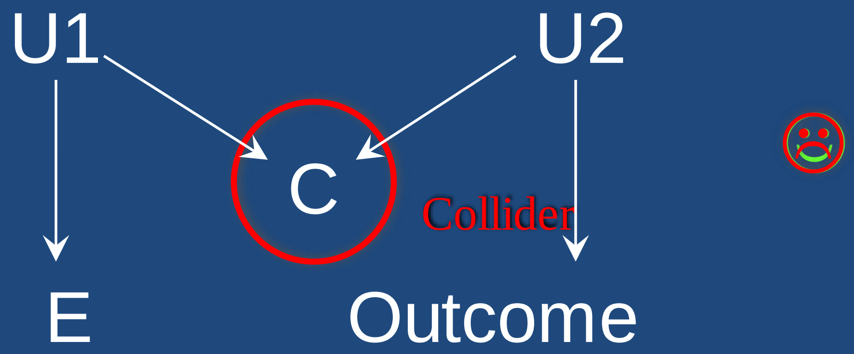
Potential Confounder?



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

(Hernán Am J Epid 2002)

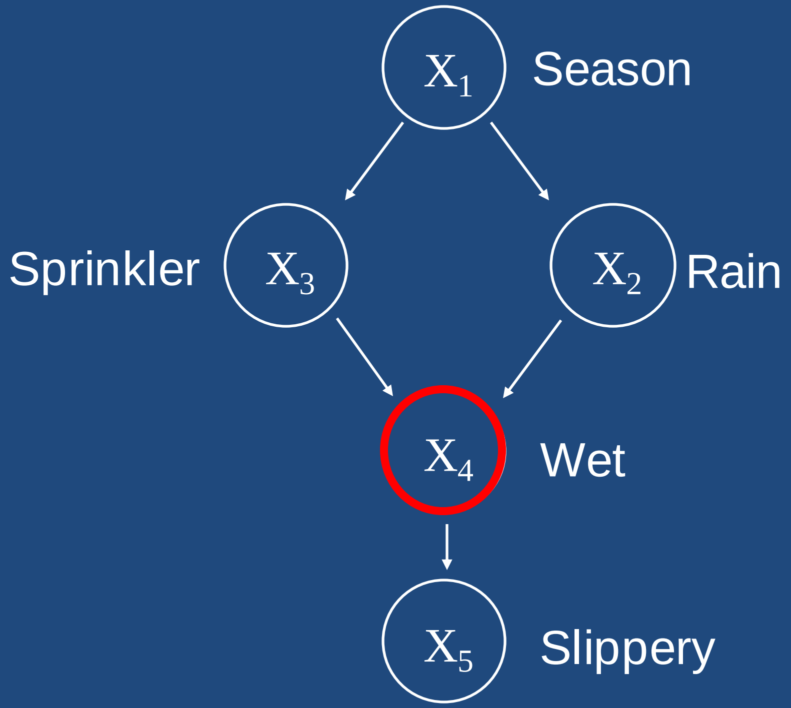
Potential Confounder?



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

(Hernán Am J Epid 2002)

Pearl's Rules - Explanation

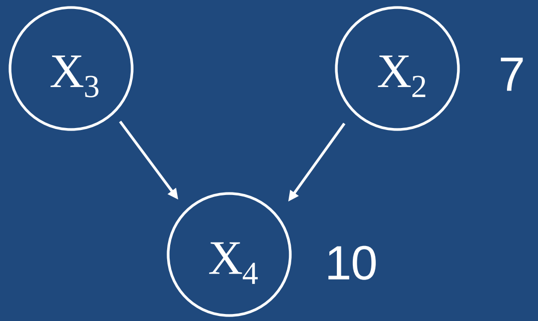


If one knows the value of the "collider", the parents are associated.

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

(Shrier & Platt, 2008)

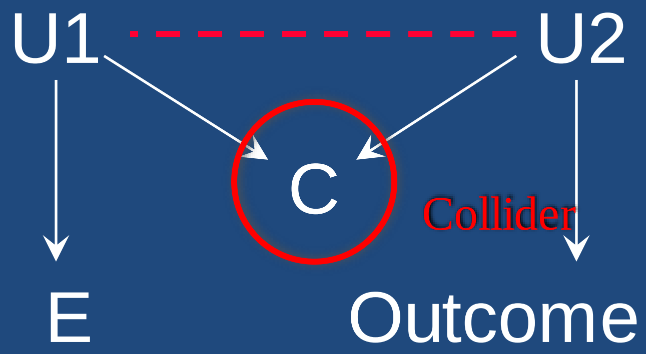
Pearl’s Rules - Explanation



$$X_4 = X_2 + X_3$$

(Pearl, Glymour and Jewell. Causal Inference in Statistics. A primer 2016)

Potential Confounder vs. Collider?



(Cole & Hernán Int J Epid 2002)

Collider-Stratification Bias

Case-control selection bias

Condition on common effect

Estrogen → Fracture → Emerg → Hospitalization

Myo. Infar. → Emerg

Complex Attrition bias

Condition on common effect

Treatment → Side effects → Drop Out

Mild disease → Side effects

Mild disease → Death

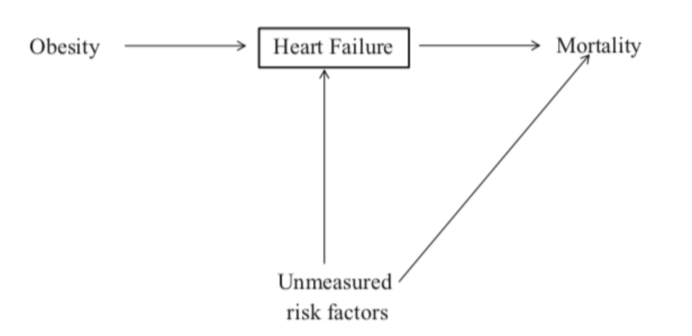
Obesity Paradox

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

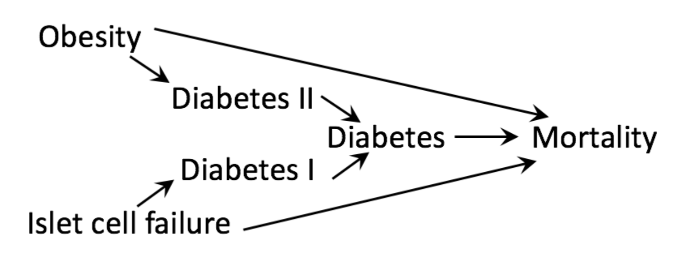
The risk of death in obese patients with heart failure is less than the risk of death in non-obese patients with heart failure.

Draw a causal diagram

Obesity Paradox



Banack & Kaufman Epidemiology 2013

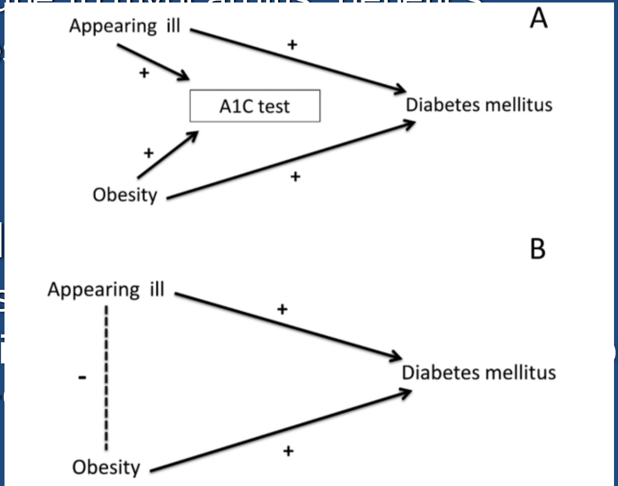


e

e

ave milder form of HF compared to HF due to myocarditis, genetics

⇒ Obe ... better



Stovitz, Banack & Kaufman JECH 2018

- Diabetes Anal
  - ⇒ Obesity is
  - ⇒ Among di ... n-insulin) do better than non-

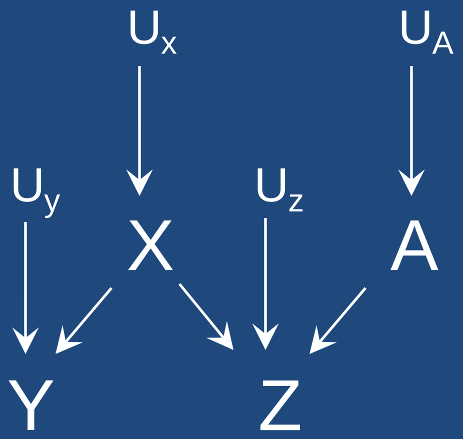
What does → on a DAG mean?

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

$$\begin{tikzcd}
X_1 \arrow[dr] & & X_2 \arrow[dl] \\
& Y &
\end{tikzcd}$$

What does → on a DAG mean?

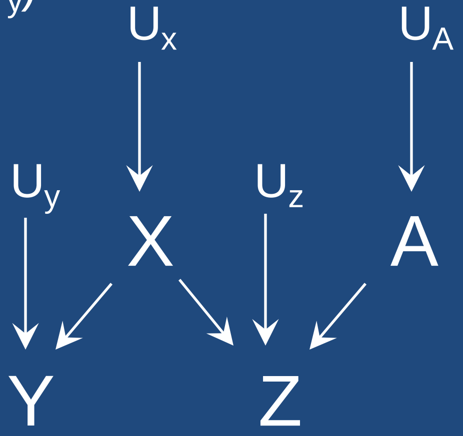
- Structural Causal Models
  - ⇒ Child is *some* function of parents
  - ⇒ f_x: $X = u_x$
  - ⇒ f_y: $Y = f(u_y, X)$
  - ⇒ f_z: $Z = f(u_z, X, A)$



(Pearl, Glymour and Jewell. Causal Inference in Statistics. A primer 2016)

What does ➛ on a DAG mean?

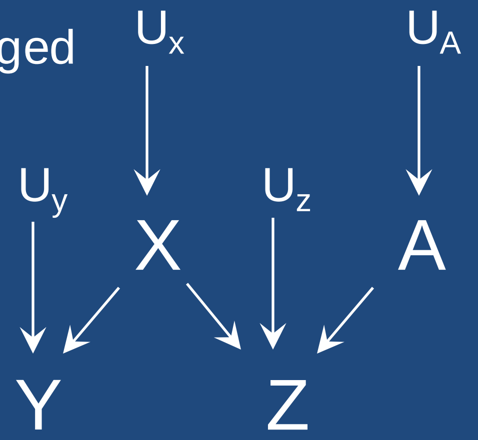
- Structural Causal Models
  - ⇒ **f_x**
    - ⇒ $X = u_x$ (or maybe $3*u_x + 10$)
  - ⇒ f_y:
    - ⇒ $Y = X*U_y$ (or maybe X^2*U_y)
  - ⇒ f_z
    - ⇒ If $X=0$, $Z=U_z*2$
    - ⇒ If $X=1$, $Z=A*X$



(Pearl, Glymour and Jewell. Causal Inference in Statistics. A primer 2016)

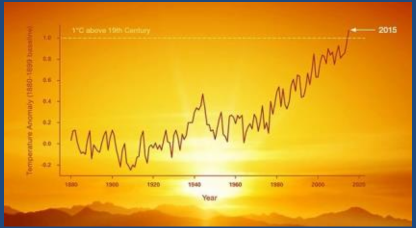
What does ➛ on a DAG mean?

- Conditional Independence
  - ⇒ Absence of an arrow means variable is NOT a function of the “parent” – independent
  - ⇒ A is independent of X
  - ⇒ Z is independent of Y given X (fix X)
  - ⇒ Z does not change if Y is changed
  - ⇒ Z is not caused by Y

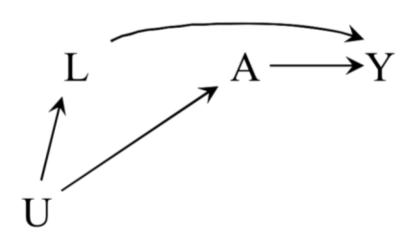


Overview

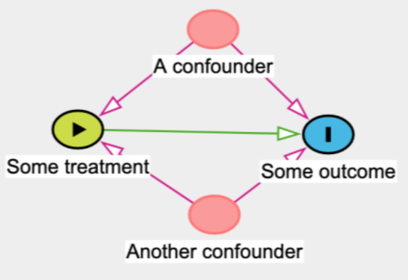
- First Period: Prediction vs causation (15 min)



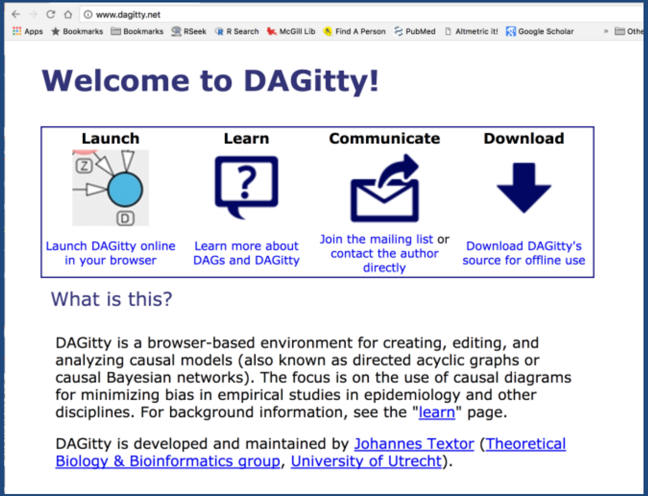
- Second Period: Introduction to DAGs (20 min)



- Third Period: An introduction to Dagitty (10 min)



Dagitty



- Coloring
- Adjustment Sets
- Testable Implications
- Model Code

Humans: What would happen if…

Pearl's 3rd Rung: Imagining / counterfactuals

Sufficient Causal Sets

American Journal of Epidemiology
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. VanderWeele[1] and James M. Robins[2]

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

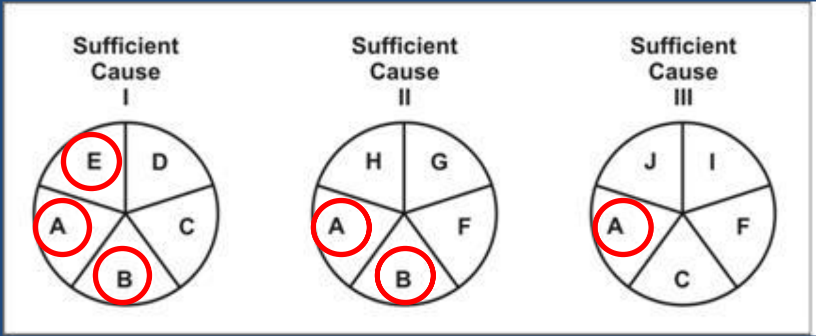
American Journal of Epidemiology
Vol. 185, No. 11
DOI: 10.1093/aje/kwx083
Advance Access publication: May 23, 2017

Invited Commentary

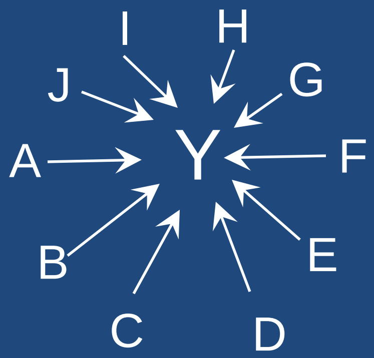
Invited Commentary: The Continuing Need for the Sufficient Cause Model Today

Tyler J. VanderWeele*

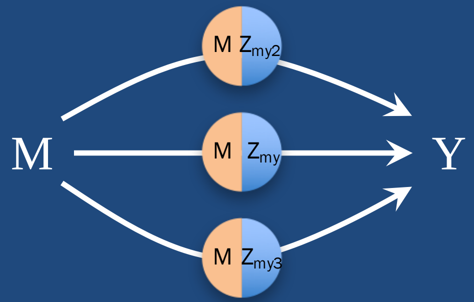
(Rothman AJE 1976)



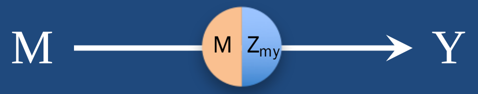
Understanding Causes



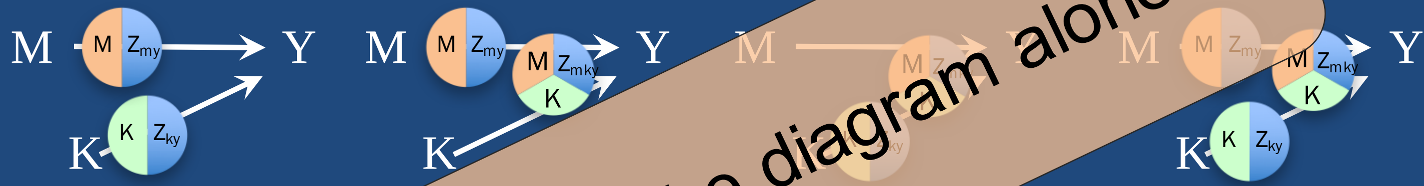
Suff. Causes Sets and Graphical Diagrams



Suff. Causes Sets and Graphical Diagrams



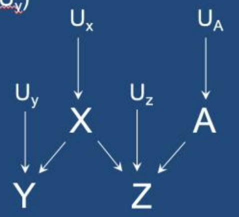
(monotonicity)



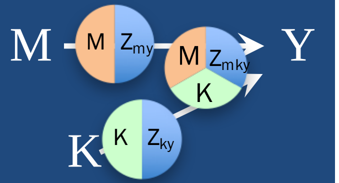
Unknown from the diagram alone

- What is the causal effect of M on Y?

- Structural Causal Models
  - ⇒ f_x
    - ⇒ $X = u_x$ (or maybe $3*u_x + 10$)
  - ⇒ f_y:
    - ⇒ $Y = X*U_y$ (or maybe X^2*U_y)
  - ⇒ f_z
    - ⇒ If $X=0$, $Z=U_z*2$
    - ⇒ If $X=1$, $Z=A*X$



Suff. Causes Sets and Graphical Diagrams



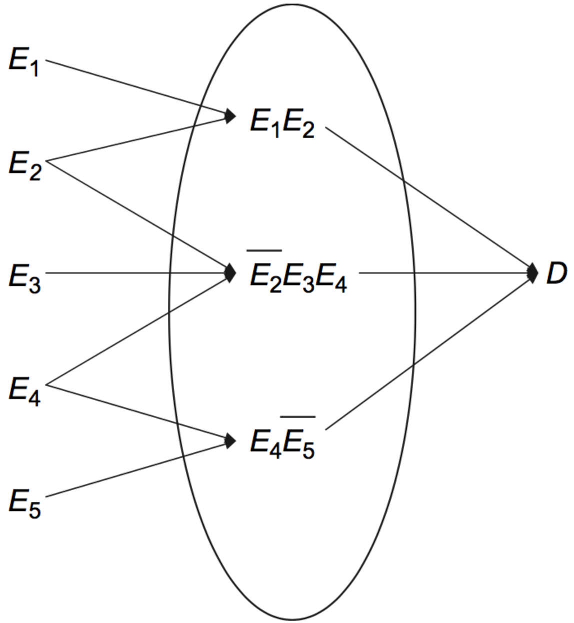


FIGURE 3. A causal directed acyclic graph with a sufficient causation structure.

(VanderWeele and Robins. Am J Epidemiol 2007)

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