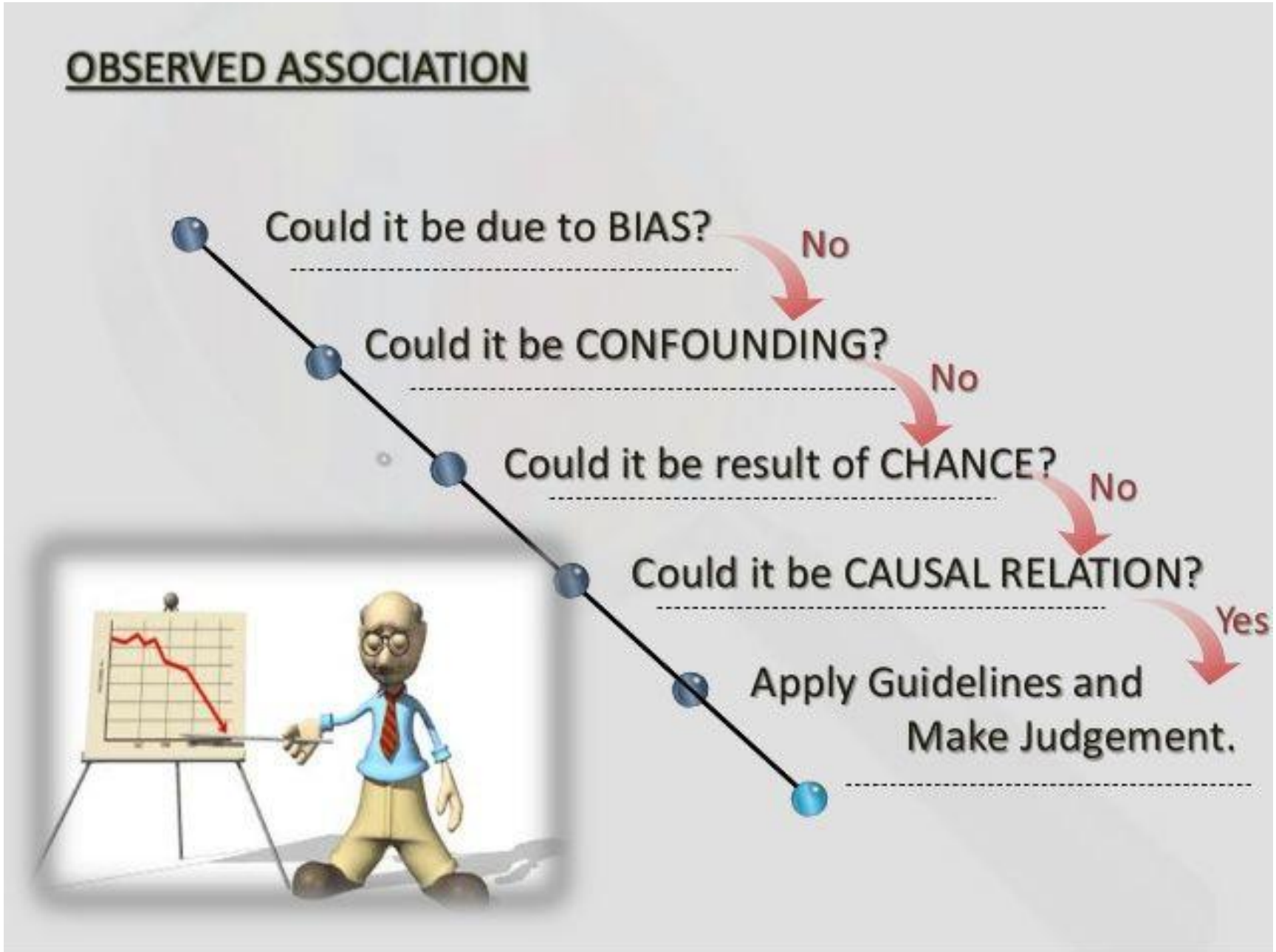
Causal Inference and Confounding

Elena Losina, PhD
Robert W. Lovett Professor of Orthopedic Surgery
Director, Orthopedic and Arthritis Center for Outcomes Research (OrACORe)

VERITY Course, 4/2/2024

Objectives of epidemiologic studies

- Describe health phenomenon
- Predict outcomes
- Identify causes



https://www.slideshare.net/VishnuYenganti/association-causation

Types of studies in epidemiology

- Observational
  - Traditional statistical methods: establish association
  - Causal inference: establish causal effects
- Interventions
  - ‘Perfect’ (sterile) condition: causality could be determined using traditional statistical methods
  - ‘Imperfect’ (differential attrition, imperfect adherence): traditional statistical methods ‘fail’ to provide evidence of causality

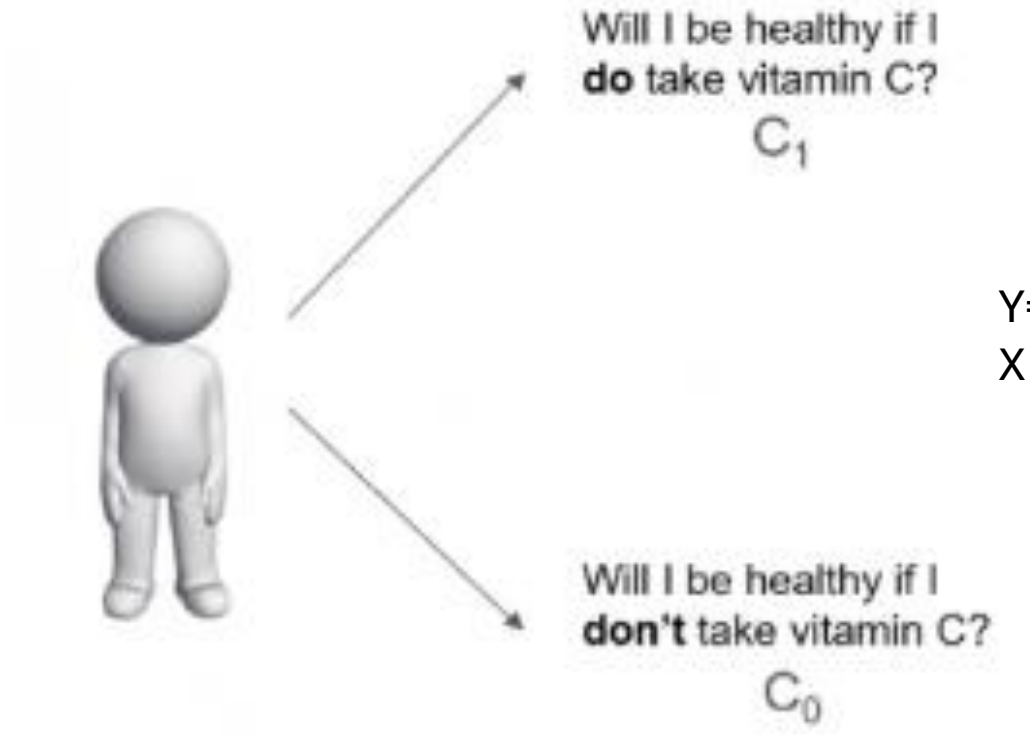
Interventions are not always ‘possible’

- Intervention is harmful – randomizing to it would be unethical
  - We would like to estimate the magnitude of the harm
- Disease is rare (outcome is rare): conducting RCT is not feasible
- Observational data may be the only available data source
  - Treatment (intervention) is assigned by ‘choice’
  - If difference in outcomes in two treatment groups is observed: how confident are we that such difference is due to treatment vs.
    - Pretreatment difference in pretreatment predisposition (risk factors distribution) to outcomes across groups
    - Confounding is a well known ‘threat’ to validity of causal inferences in observational studies

Causality

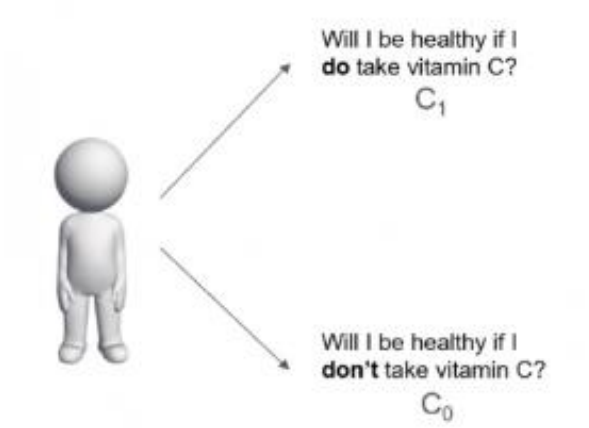
- Allows us reason the world
- Plays key role in decision making
- Essential in medicine
  - Will medicine M impact outcome D
- What type of evidence we need to establish causality?
- When do we ‘feel’ (know?) there is enough evidence to establish causality

Association and Causality



Y=outcome (1=healthy, 0=not healthy)
X =treatment (1=taking vitamin C, 0 not taking vitamin C)

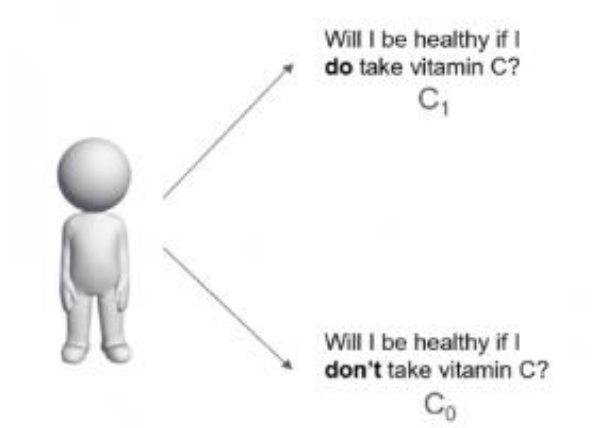
Association and Causality



Association: E(Y|X=1)-E(Y|X=0)
Observed difference between outcomes with and without vitamin C

Y=outcome (1=healthy, 0=not healthy)
X =treatment (1=taking vitamin C, 0 not taking vitamin C)

Association and Causality



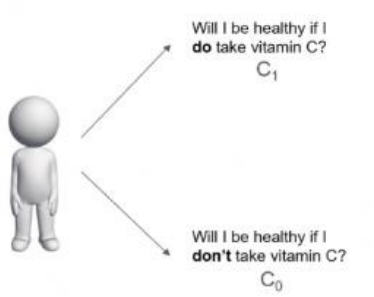
Association: E(Y|X=1)-E(Y|X=0)
Observed difference between outcomes with and without vitamin C

Causal effect: E(C1)-E(C0)
Difference between expected outcome of taking and not taking vitamin C (C1, C0)

Y=outcome (1=healthy, 0=not healthy)
X =treatment (1=taking vitamin C, 0 not taking vitamin C)

Association and Causality

Individual	X (taking vitamin C)	Y (being healthy)	C1 (**potential** outcome with taking C)	C0(**potential** outcome without taking vitamin C)
1	0	0	0	0
2	0	0	0	0
3	0	0	0	0
4	1	1	1	1
5	0	0	0	0
6	1	1	1	1
7	1	1	1	1
8	1	1	1	1



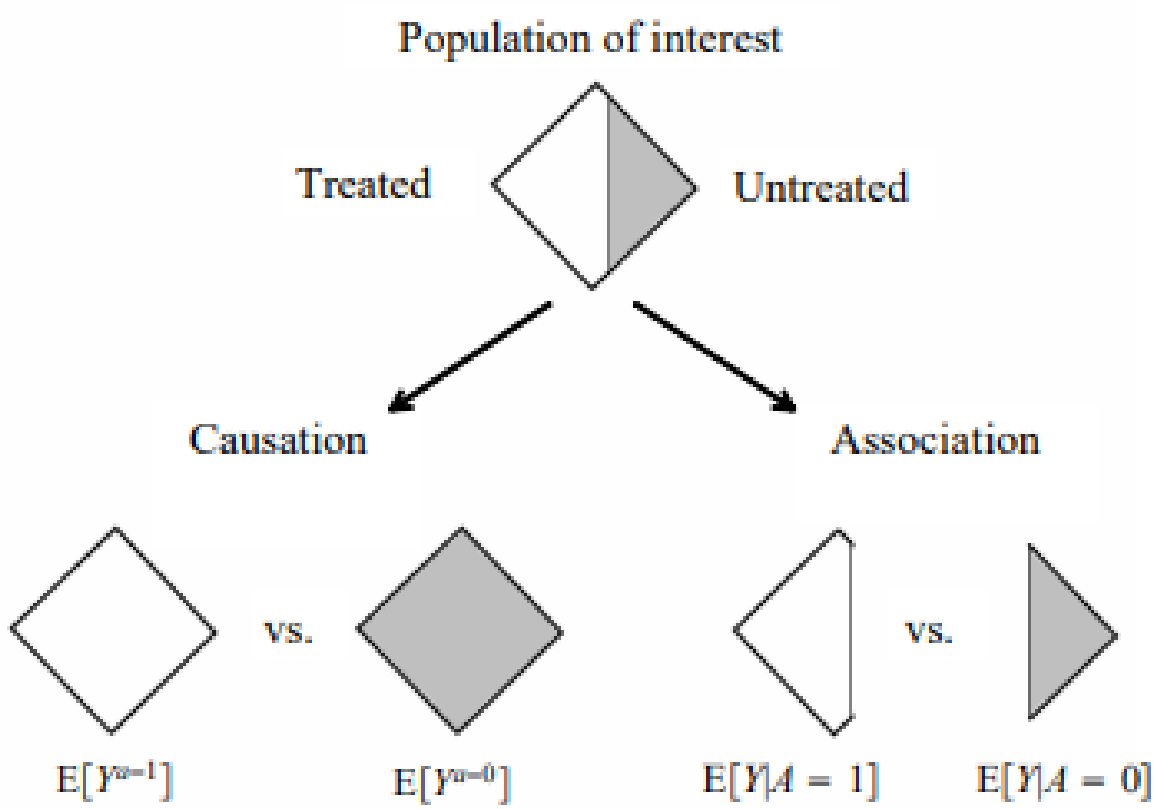
Causal effect: E(C1)-E(C0)
Difference between expected outcome of taking and not taking vitamin C

E(C1)=4/8=0.5
E(C0)=4/8=0.5
Causal Effect: E(C1)-E(C0)=0

Association: E(Y|X=1)-E(Y|X=0)
Observed difference between outcomes with and without vitamin C

E(Y|X=1)=(1+1+1+1)/4=1
E(Y|X=0)=(0+0+0+0)/4=0
Association: E(Y|X=1)-E(Y|X=0)=1

Association vs. Causality



Reproduced from *Causal inference What if,* M. Hernan and J. Robbins, Fig 1.1

Definition of Causal Effect

- The causal effect of receiving treatment for a person i is a comparison of potential outcomes
- Definition of causal effects does not depend on the actual treatment taken

Assumptions of Causal Inference

- Ignorability
  - All variables affecting exposure (treatment) and outcome are observed and can be controlled for
- Positivity
  - The data should contain treated and untreated subjects
    - ~~Always treated or never treat with trt X~~ vs sometimes treat with X
- Consistency
  - Intervention is well defined
    - Does poverty cause severe RA vs. Does household income below poverty level causes RAPID3 global >6
  - Not an attribute of an individual
  - Individual's outcomes are not affected by treatment status of another person
- Exchangeability (conditional)
  - You can exchange the treated and untreated for the same result

Causal Effects: Example

- A person i with RA could receive bDMARD or not
- Outcome: Yi is defined as achieving RA remission at 3 months (binary)
- Causal effect of receiving bDMARD is defined as the comparison of potential outcome for the person with bDMARD and without bDMARD
  - Potential outcomes do not depend on what treatment the person gets

Key Concept of Causal Inference: Counterfactual

- ‘What if’: what would happen under condition other than observed

Umbrella prevented me from getting wet

Counterfactual (deterministic): If I would not bring umbrella, I would get wet

Counterfactual (probabilistic): if I would not bring umbrella, I would most likely get wet

Counterfactuals

- Let’s define observed outcome of a person i as Yi
- Let’s define potential outcome of a person i with treatment as Y1i and without treatment as Y0i
- We can define observed outcome Yi in terms of the potential outcome as
    - $Y_i = Y_{0i} + (Y_{1i} - Y_{0i})D_i$
- If Di = 1 then the observed outcome is Yi = Y1i .
- If Di = 0 then the observed outcome is Yi = Y0i
- If a person receives the treatment the observed outcome is Y1i and counterfactual outcome is Y0i
- For treated group, Y0i is a counterfactual while Y1i is observed.
    - For treated: Y1i = Yi
- For the control group,Y1i is a counterfactual, Y0i is observed
    - For controls: Y0i=Yi
- **A person either receives the treatment or not, never both**

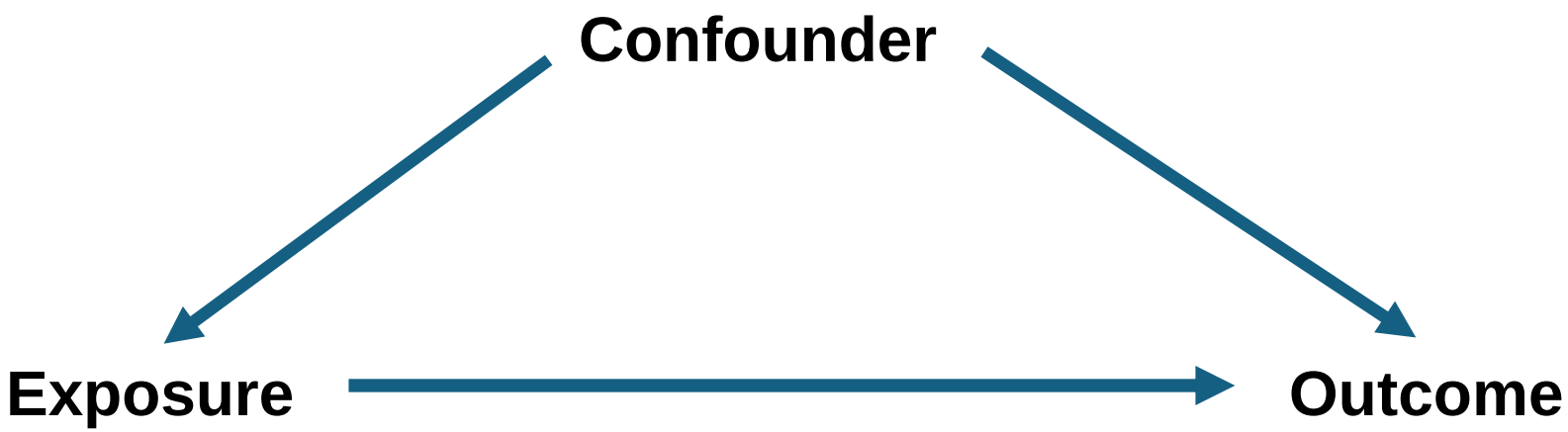
Problem

- Can not observe both conditions at the same time
- Solution: observe a ‘substitute’ population whose experience represents that of ‘exposed (treated)’ without exposure (treatment)
- Challenges: how to identify such population?
  - Ignorability
  - Positivity
  - Exchangeability

Directed Acyclic Graphs (DAGs)

- Used to visualize prior knowledge about variables related to specific causal question
- Help to identify variables that need (or NOT need!) to be controlled in design or analysis
  - To minimize distortions due to confounding
  - To prevent controlling for mediators or colliders
- Point to sources of bias

Confounders

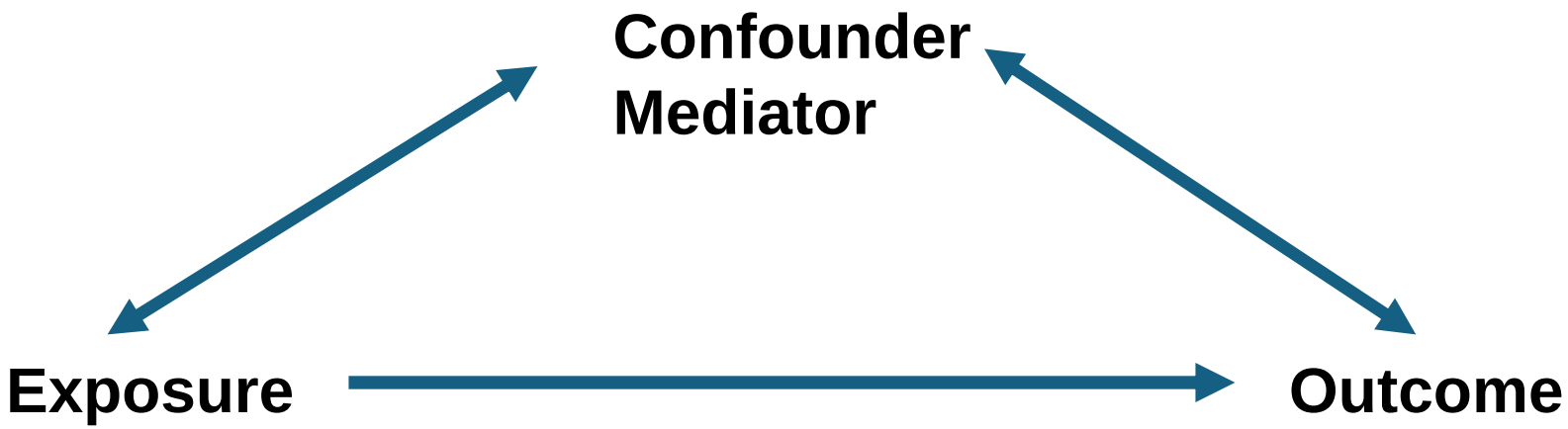


- A confounder is a variable which causes both the exposure and the outcome
- It is necessary to adjust for all confounders – which blocks these **back-door paths** to obtain an unbiased causal estimate of the exposure–outcome association

Mediators

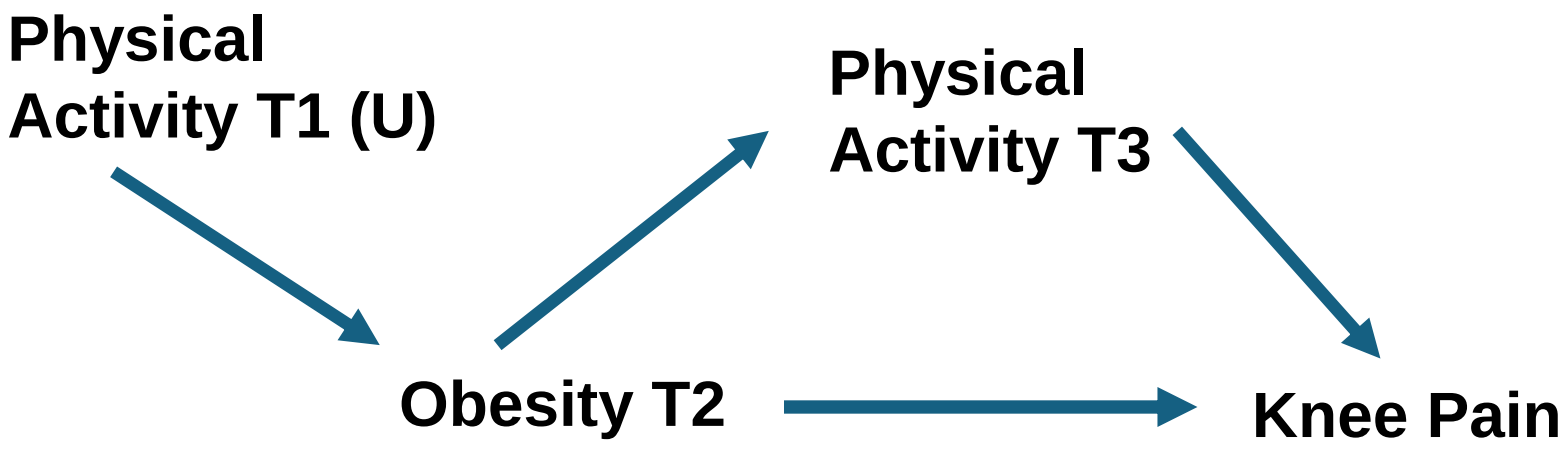
- A mediator is a variable which is caused by the exposure, which in turn causes the outcome
- Mediators are part of the pathway by which the exposure causes the outcome
- Adjusting for a mediator will result in a biased estimate of the exposure–outcome association

Confounders and Mediators

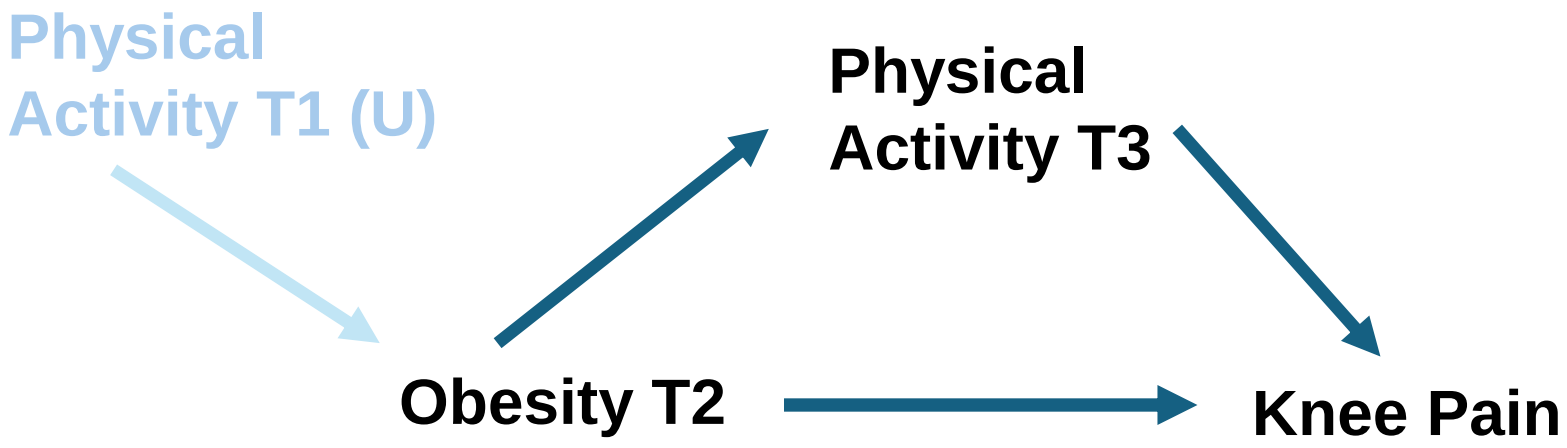


- Some covariates may be both confounders and mediators
- Implications: in the absence of longitudinal, repeated, data – calculating causal estimates may be even more difficult as both adjusting and not adjusting for the covariate will result in bias.

Illustration of bi-directionality

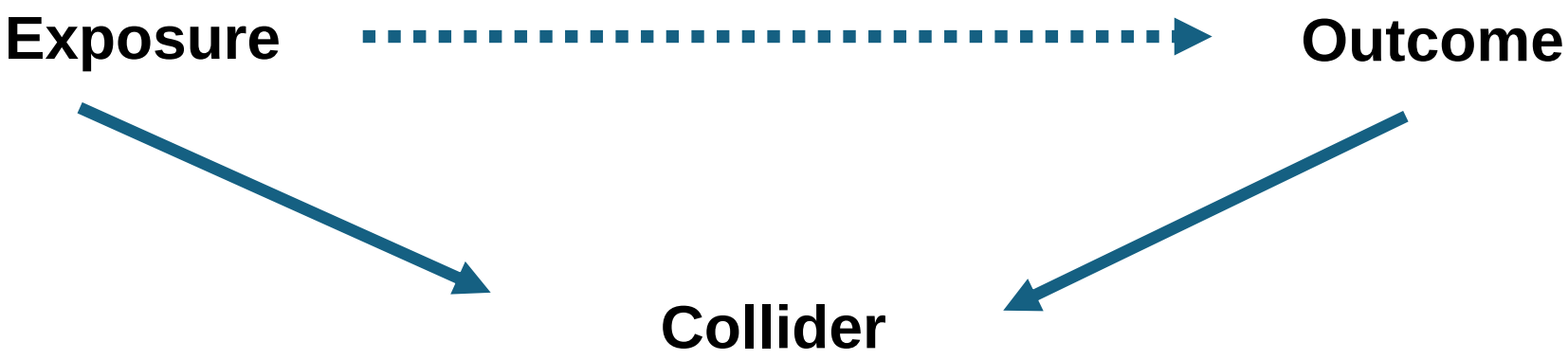


Challenges of model misspecification



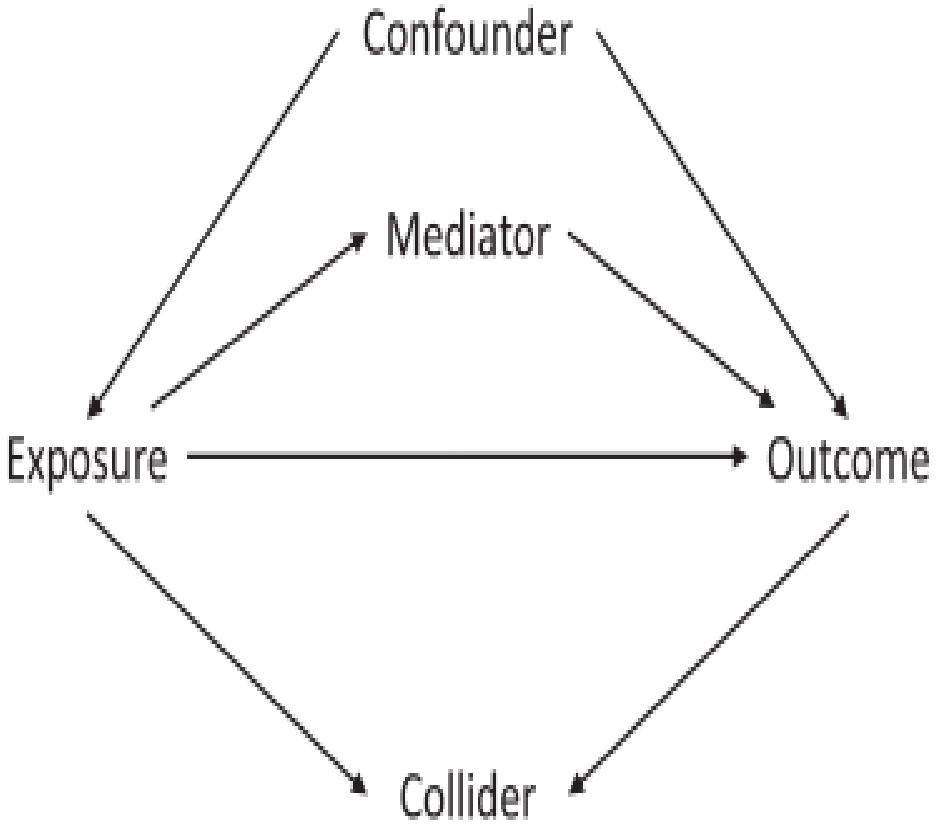
- PA at T3 is a mediator – do not adjust
- PA at T1 is a confounder – adjust!
- If PA at T1 is unmeasured – we can't estimate unbiased causal relationship between obesity and knee pain using standard statistical methods
- Solution: a set of sensitivity analyses (over and underadjustment)

Collider



- A collider is a variable which is caused by both the exposure and the outcome
- Colliders ‘block’ the flow of information between other variables
- Adjusting for a collider, however, opens these pathways, potentially resulting in biased associations
  - Introducing association between E and O where they are none

Putting Puzzle Together: Confounders, Mediators, Colliders



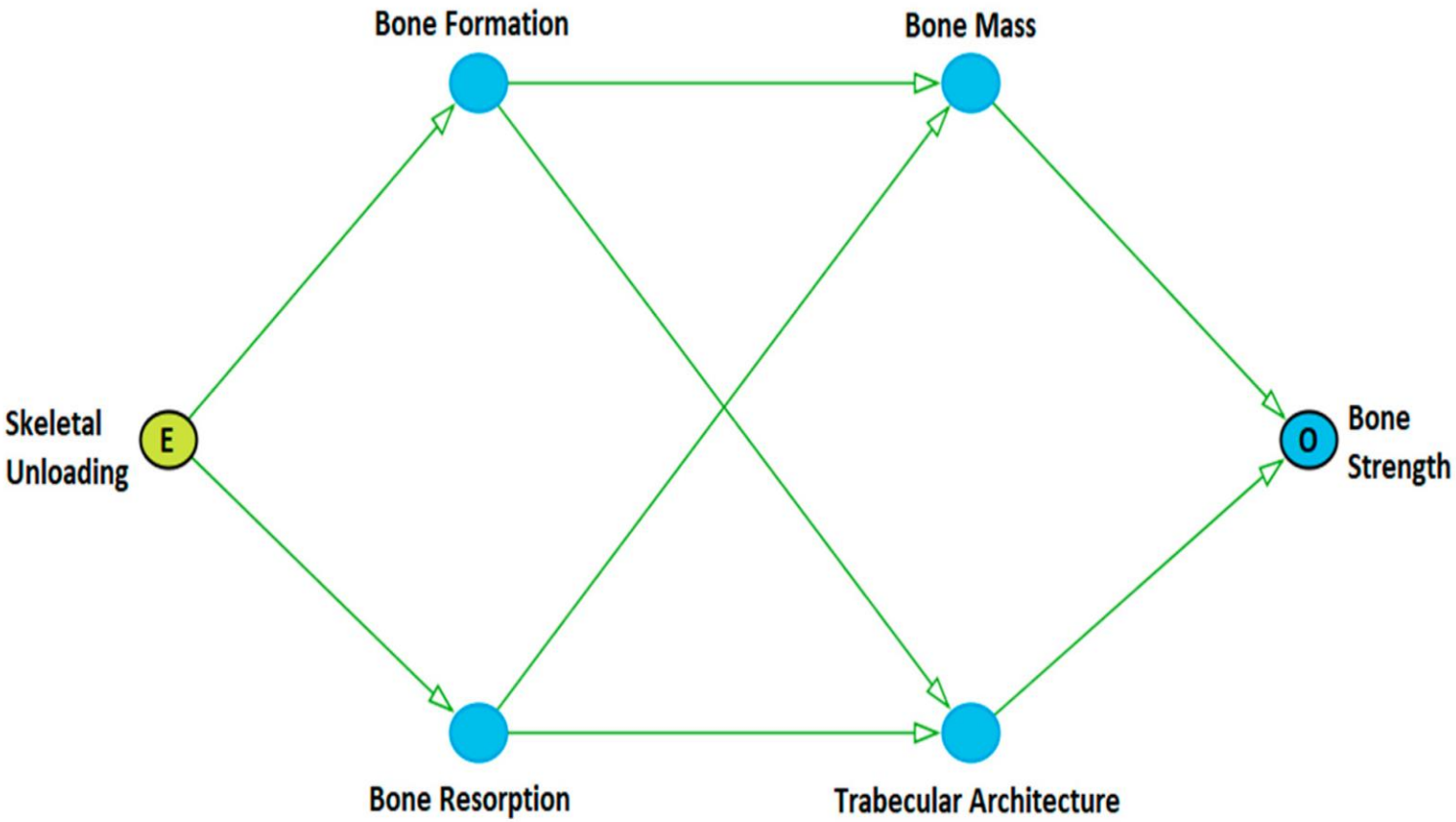
<u>Causal paths</u>:
Exposure→Outcome
Exposure→Mediator→ Outcome

<u>Non-causal paths linking E and O and how to block them</u>
Exposure←**C**onfounder→Outcome (control for Confounder)
Exposure→Collider←Outcome (don’t control for Collider)

Assumptions to inform appropriate statistical analysis to derive unbiased causal estimates

- Selecting variables for adjustment: Identifying confounders, mediators and collider variables
- Residual confounding (RC)
  - Unmeasured confounders
  - Confounders measured with errors
  - Hard to model
  - Solution: sensitivity analysis
    - Estimating the magnitude of residual confounding to alter conclusions of the analysis
      - If conclusions are robust unless the magnitude of RC large (larger than what would consider to be reasonable) – results are unlikely to be biased by RC
      - If small levels of RC lead to drastic changes in conclusions – more likely that results are subject to RC
      - These sensitivity analysis help to assess the direction is plausible, even in presence of RC
- Selection bias: establishing causal inference relies of absence of selection bias
  - Selection bias occurs when selection into the sample is not random with respect to target population

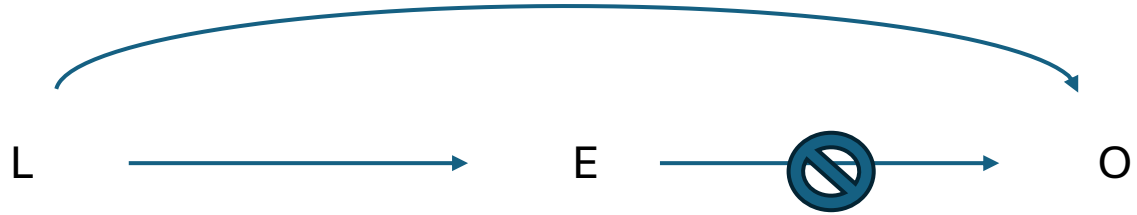
DAGs terminology: exercise



Ascendants: parents, grandparents
Descendants: children, grandchildren

No cyclicity

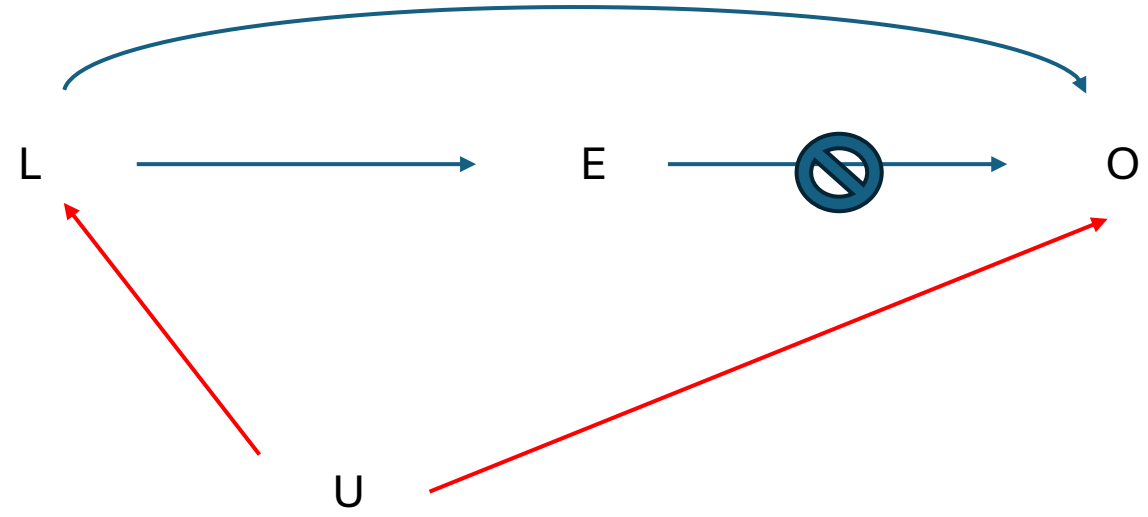
Back-Door Criterion



A back-door path is non-causal path between exposure (E) and outcome (O) that remains even if all descendants of E are removed

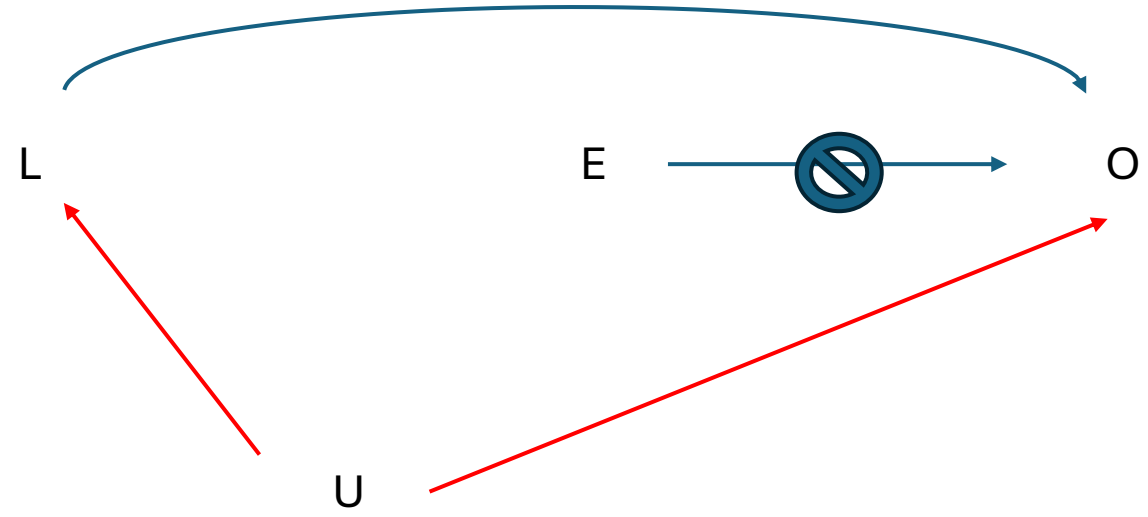
O ←L→E

Back-Door Criterion



Unmeasured variable U jeopardizes deriving the causal inference regarding E and O, but conditioning on L will block all back door pathways between E and O

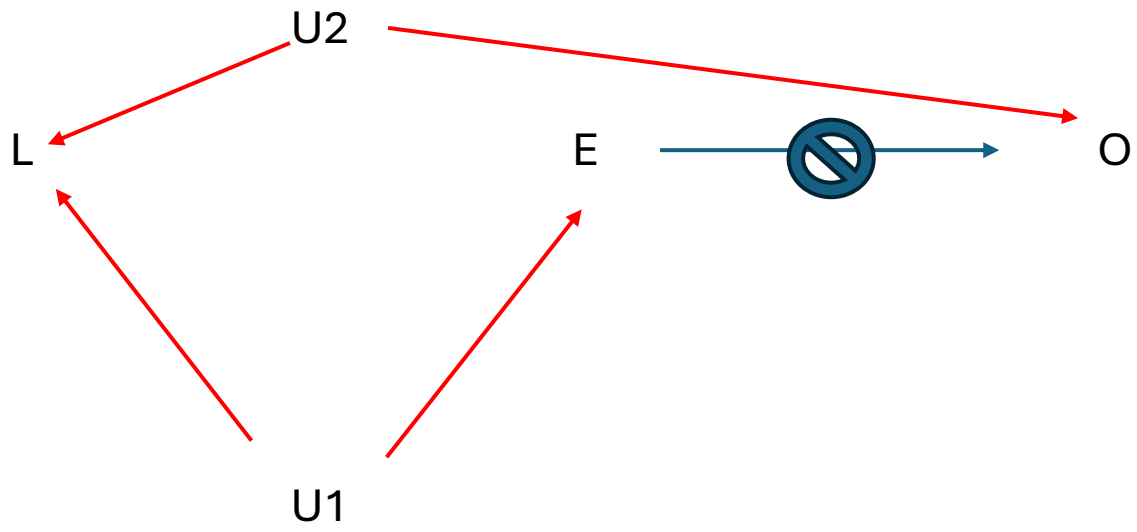
Back-Door Criterion



- Unmeasured variable U jeopardizes deriving the causal inference regarding E and O, but conditioning on L will block all back door pathways between E and O

- All back door pathways could be blocked by controlling for measured L, we can estimate causal relationship between E and O

Back-Door Criterion



There is no back door criteria that need to be blocked, no confounding. Relationship between E and O are causal

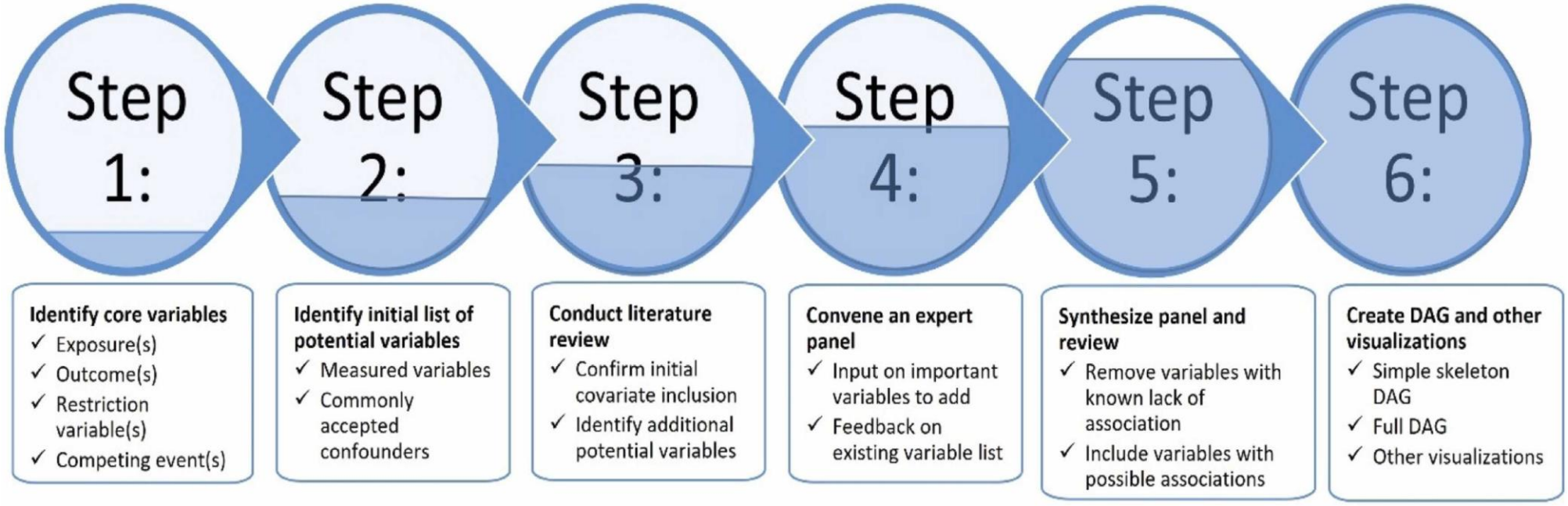
Controlling for confounders

- Solving the fundamental problem of causal inference:
  - Randomization
    - Challenges are due to post-randomization drop outs
  - Causal Inference methods based on observational studies rely on uncheckable assumption:
    - **all potential confounders are measured**
    - **All back door paths could be blocked by controlling for measured confounders**
    - **Conditional exchangeability**

Controlling for confounding: statistical methods

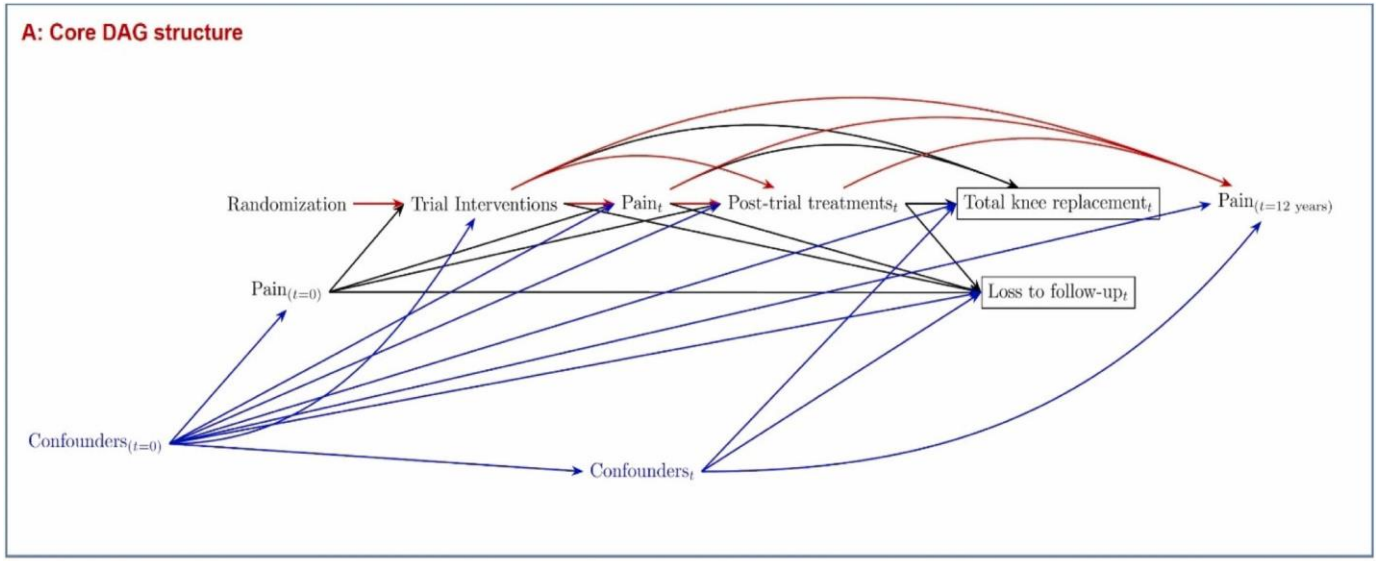
- Methods requiring conditional exchangeability
  - G-methods:
    - Standardization
    - inverse probability weighting
    - g-estimation
      - Most recommended with time-varying confounders
  - Stratification-based methods
    - Stratification (restriction)
    - Matching
- Conditional exchangeability is hard to achieve
  - Expert knowledge is helpful
- Other methods (rely on unverifiable assumptions, can't be used for time varying confounders)
  - Difference-in-difference
  - Instrumental variable estimation

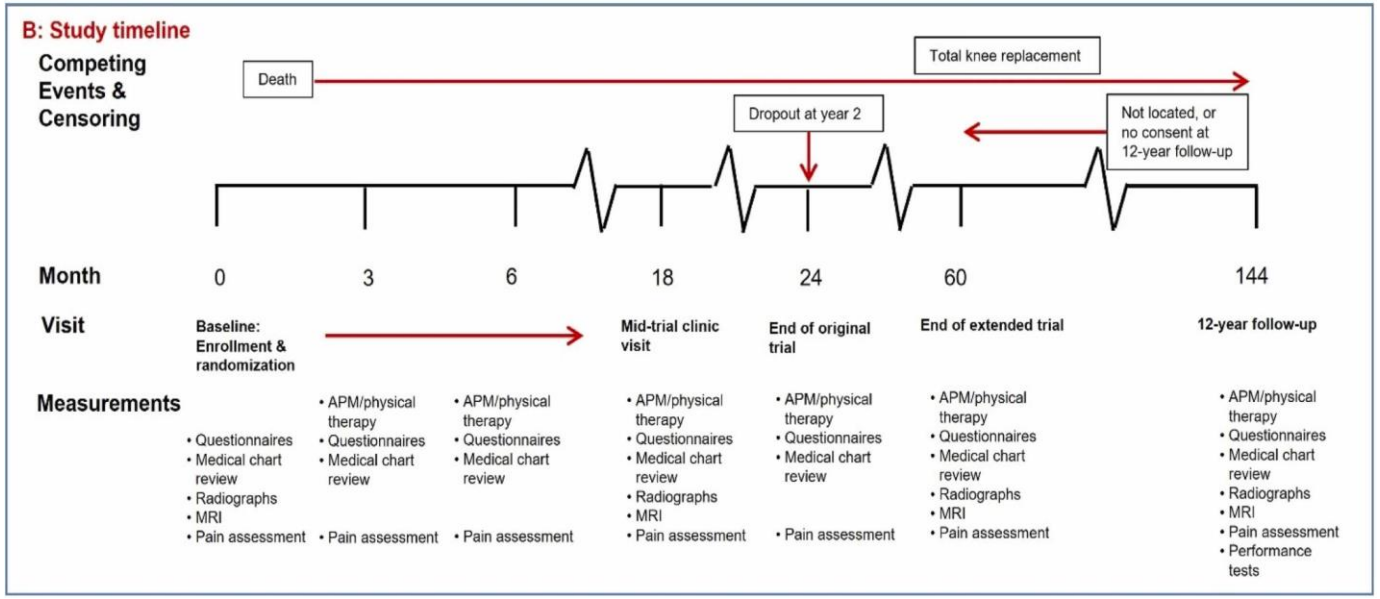
Process of building DAGs



Osteoarthritis and Cartilage 2024 32319-328DOI: (10.1016/j.joca.2023.10.007)

DAG structure and study timeline





Osteoarthritis and Cartilage 2024 32319-328DOI: (10.1016/j.joca.2023.10.007)

Take home message(s)

- Association ≠ Causality
- Randomization is not always possible
- Causal inference methods help to define ‘pathway to causality’ using observational data
- Constructing DAGs could help to define confounders, colliders and mediators
- Controlling for confounders is important, controlling for colliders and mediators may introduce bias
- Building DAGs based on expert knowledge helps to establish all causal and non-causal pathways between treatment (exposure) and outcome
- Statistical methodologies exist to address causality in observational studies

Parting thoughts

- All models are wrong, some are useful

(George Box)

- If you torture data long enough, they will confess to anything (Ronald Coase)

A few useful references

- Causal Inference: What If (M. Hernan, J. Robins) (https://www.hsph.harvard.edu/miguel-hernan/causal-inference-book/)
- Causal Inference (https://blog.ml.cmu.edu/2020/08/31/7-causality/)
- An introduction to g methods (doi: 10.1093/ije/dyw323)
- Directed acyclic graphs for clinical research: a tutorial

(https://doi.org/10.7602/jmis.2023.26.3.97)

- Association or causation? How do we ever know?

(https://catalogofbias.org/2019/03/05/association-or-causation-how-do-we-ever-know/)

- Tutorial on Directed Acyclic Graphs (doi:10.1016/j.jclinepi.2021.08.001)
- Causal relationships between pain, medical treatments, and knee osteoarthritis: A graphical causal model to guide analyses (doi: 10.1016/j.joca.2023.10.007.)
- Validating Causal Diagrams of Human Health Risks for Spaceflight: An Example Using Bone Data from Rodents https://doi.org/10.3390/biomedicines10092187