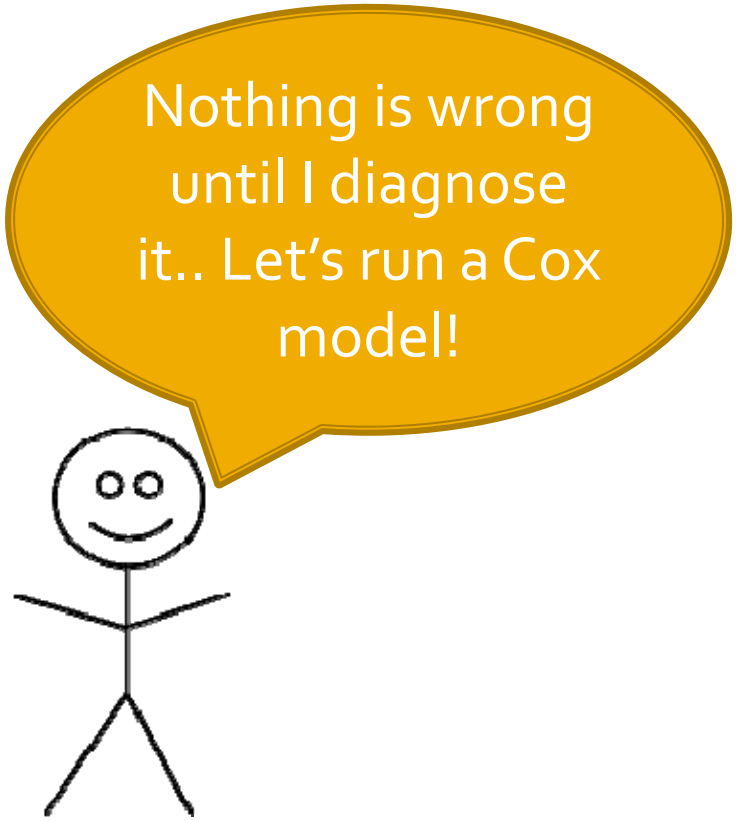
Organization of the workshop

1. Talks on methods
2. Talks on data: HAAS, WHICAP, and REGARDS
3. Meet with a work group to develop a precise research question and propose an analytic plan
4. Work with the data enough to realize why your plan won't work
5. Revise….
6. Find the people who can solve your problem and sit next to them at dinner
7. By Thursday: present the proposed research, preliminary results, next steps
8. By Friday morning: nail down logistics with your group

What's different this year?

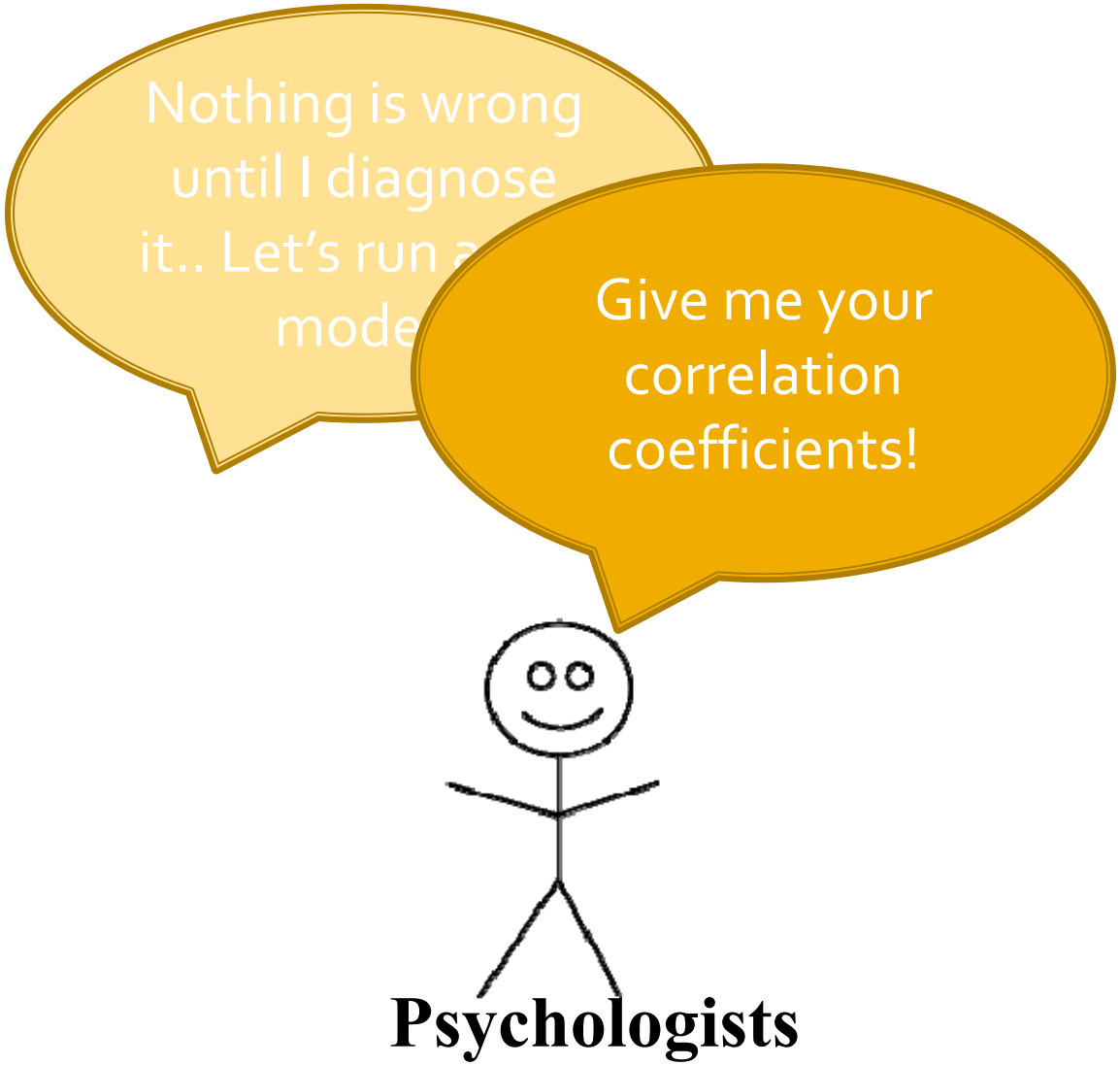
- This workshop was conceptualized and launched to solve measurement problems
- (My impression): many of the people who attend have causal problems, and measurement problems exacerbate their causal inference problems
- Dementia researchers draw from neuropsychology, neurology, neuropathology, epidemiology, biostatistics, computer science, and…
- Common language is a barrier

My wildly simplified cartoon of research on cognitive outcomes in old age

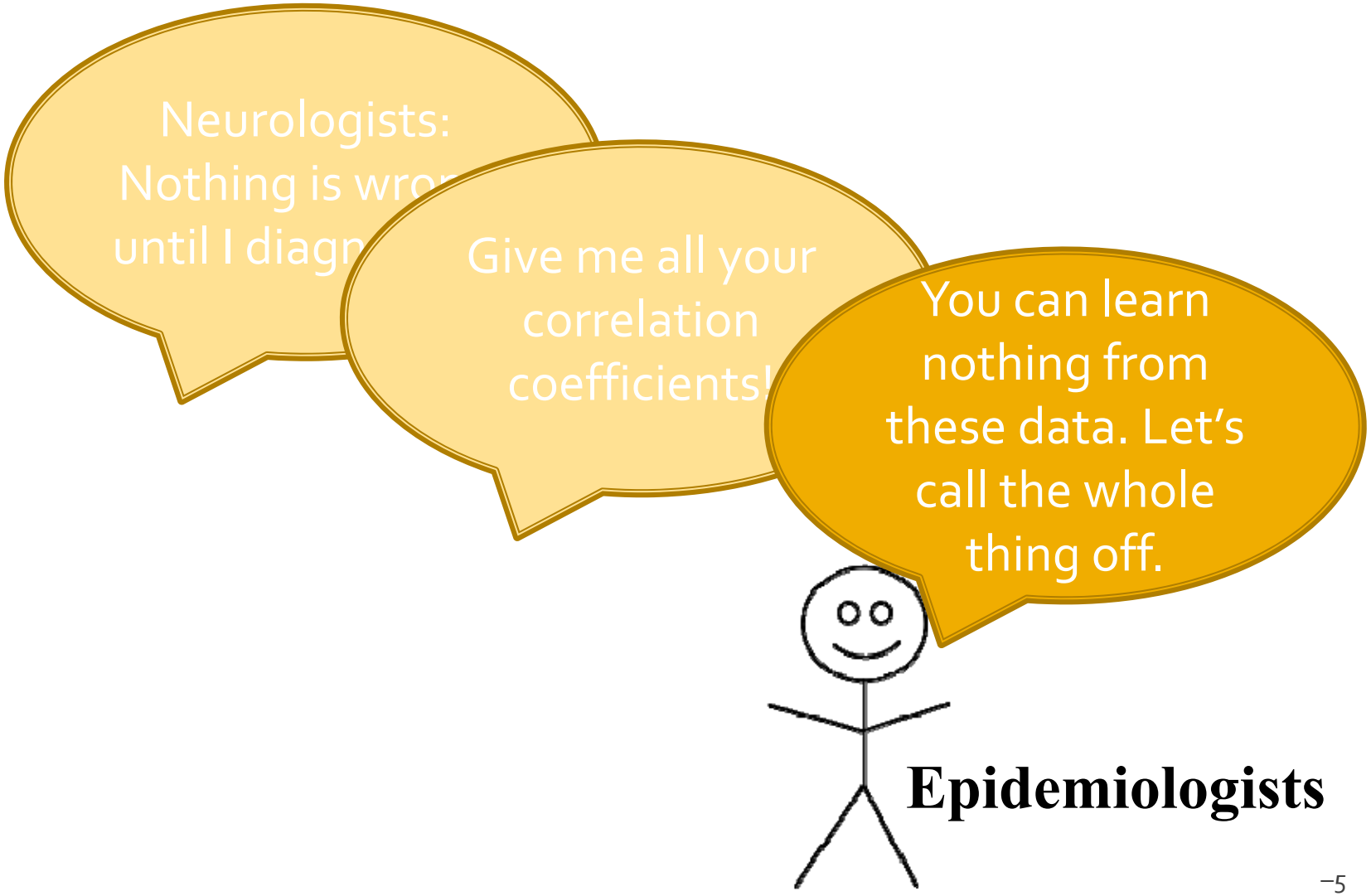


Neurologists

My wildly simplified cartoon of on cognitive outcomes in old age



My wildly simplified cartoon of on cognitive outcomes in old age



We have crummy data and a lot of glorified correlation coefficients. What *can* we learn?

- Since the 1980s (?) there has been a sea change in how we thought and talked about causal inference
- Once: we can learn nothing about causation without a randomized experiment. It represented sloppy thinking to refer to causation except to acknowledge it as an impossible dream.
- Now: we can (and do) learn lots about causation by measuring associations (correlations) in data
- The challenge is to be explicit about the assumptions you must invoke to make the leap from your data to your inferences

We have crummy data and a lot of glorified correlation coefficients. What *can* we learn?

- Now: we can (and do) learn lots about causation by measuring associations (correlations) in data
- The challenge is to be explicit about the assumptions you must invoke to make the leap from your data to your inferences

We have crummy data and a lot of glorified correlation coefficients. What *can* we learn?

- Most of the methods discussion is to:
  - Articulate the causal parameter you want (eg, what trial would you run if you could run a perfect trial?)
  - Articulate the assumptions you need to estimate that parameter from your data
  - Find an analysis that would deliver that parameter from your data under the most plausible set of assumptions

Introduction to Causal Inference Methods: Using Counterfactuals and Directed Acyclic Graphs

Maria Glymour
Department of Epidemiology & Biostatistics
University of California, San Francisco

Organization

1. Counterfactual account of causation
2. Drawing and using DAGs
3. Contrasting IV-based methods and covariate adjustment/propensity scores

Organization

- Counterfactual account of causation
  - What's a counterfactual?
  - The fundamental problem of causal inference
  - Causation vs association: when is prediction good enough?
  - Confounding per counterfactuals
  - Objections

Causal Inference

Very commonly, we wish to know about causal relations…

If we increased X, would Y also increase?

But we observe only statistical associations…

People with high values of X also have high values of Y.

Counterfactuals or Potential Outcomes

- Everyone has a well-defined outcome value (Y), under all possible values of the exposure (X), but we only get to observe one of the possible outcomes for each individual.
- Specifically, we observe each individual's outcome Y under the value of exposure X that the individual *actually* received.
- For individuals who actually received X=1, then we observe the value Y would take under exposure X=1, but not the value Y would take for any other values of X
- For individuals who actually received X=0, then we observe the value Y would take under exposure X=0, but not the value Y would take for any other values of X
- X is a cause of Y if the value of Y *would have been* different if X had been different.

Counterfactuals or Potential Outcomes

- X is a cause of Y if the value of Y *would have been* different if X had been different.
- Sometimes we invoke the idea of "do X" or "set X"
- Notation: $E(Y_{X=1})$ is the expected value of the counterfactual value of Y setting X=1. Sometimes written: $E(Y^{X=1})$
- Distinguish from E(Y| X=1)

Counterfactuals or Potential Outcomes

- Not deterministic:
  - X is a cause of Y if X is a cause of Y *for some people in the population* or
  - X is a cause of Y if Y has different probability distributions under different values of X
- X can be a cause of Y even if it is neither necessary nor sufficient to produce Y for some individuals

Counterfactuals or Potential Outcomes

What is the *effect* of living in poverty while aged 23-30 (X) on risk of developing dementia before age 80 (Y)?
Suppose that if you were omniscient, you would know that:

- If Earnest lives in poverty, he will develop dementia by age 80.
  - $Y_{X=1}=1$
- If he doesn't live in poverty, he will *not* get dementia by age 80.
  - $Y_{X=0}=0$

Earnest actually does live in poverty, and he actually does develop dementia. Non-omniscient researchers never get to see what would have happened to him if he'd taken that Wall Street job right out of college and avoided poverty. This is the *fundamental problem of causal inference.*

Estimating counterfactual values from observed values

Instead we observe the dementia status of Francis, who's a lot like Earnest but avoided poverty. Francis did not develop dementia. We assume that Francis and Earnest are "exchangeable", and conclude that Earnest developed diabetes because of his poverty.

On a population level, we want to know:

$$E(Y_{X=1}-Y_{X=0})$$

But we observe only:

$$E(Y|X=1)-E(Y|X=0)$$

Estimating counterfactual values from observed values

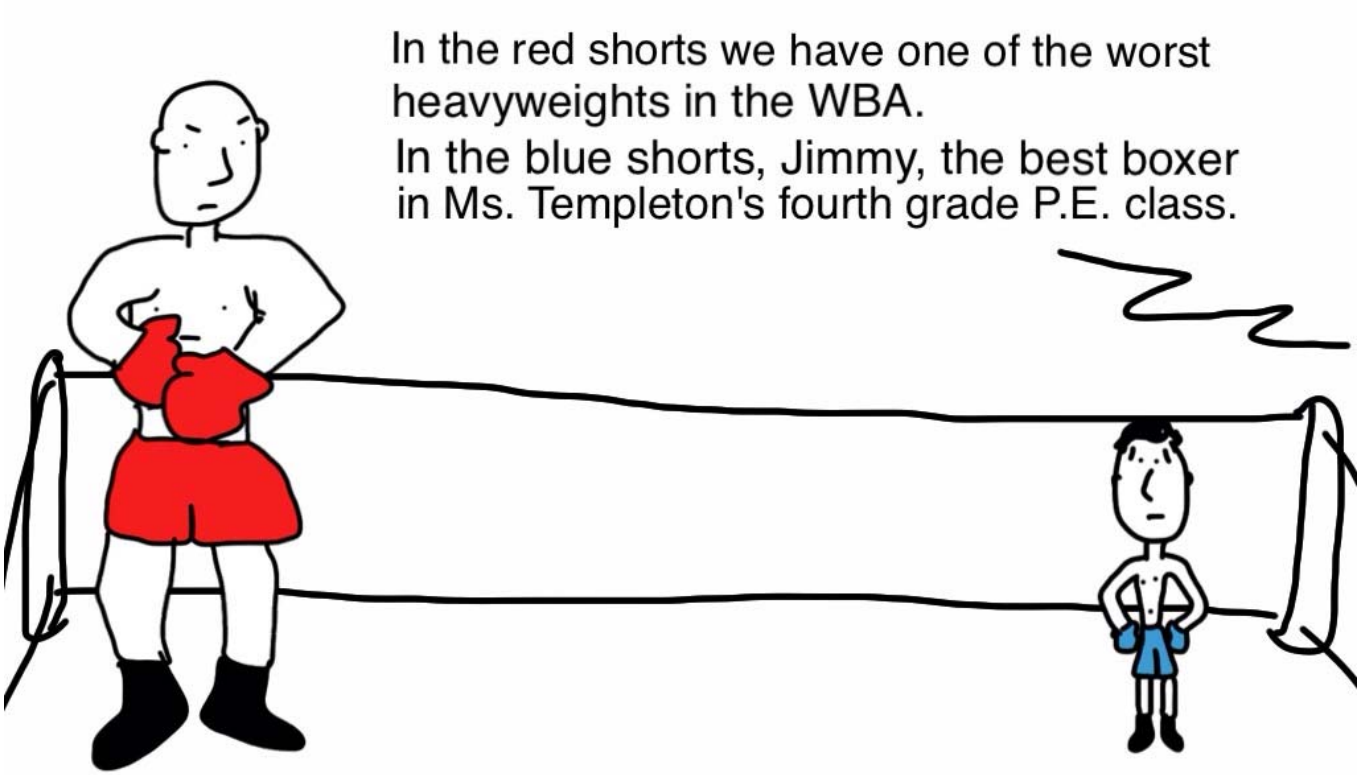
We observe the statistical association between poverty and dementia, and hope that the dementia outcomes of people who were not impoverished represent the outcomes people who were impoverished **would have had** if they hadn't been poor.

Estimating counterfactual values from observed values

"Confounding is present if our substitute imperfectly represents what our target would have been like under the counterfactual condition."

- Maldonado and Greenland (2002)

Choosing an appropriate substitute to estimate unobserved counterfactuals



freshspectrum.com

Hint: not good substitutes

Inferring Causation from Association

"Confounding is present if our substitute imperfectly represents what our target would have been like under the counterfactual condition."

- Maldonado and Greenland (2002)

- For any analysis, ask: What is the identifying contrast? Which groups of people am I using to represent one another's counterfactual outcomes?

- What causal structures could have generated the statistical associations we observe? We are interested in knowing whether X caused Y – what are the range of causal structure consistent with the observed statistical associations?

Statistical versus Causal Language

Statistical concepts:

- Correlation
- Regression
- Conditional independence
- Association
- Likelihood
- Risk ratio
- Odds ratio

Causal concepts:

- Influences
- Effect
- Confounding
- Exogeneity

Statistical versus Causal Language

Statistical claims:

- X and Y are correlated
- X predicts Y
- X predicts Y conditional on (adjusting for or stratifying on) Z
- The prevalence of Y among those with X is twice as high as the prevalence of Y among those without X.

Causal claims:

- X causes Y
- X affects Y
- X increases (or decreases) Y
- X induces M, which induces Y

Statistical versus Causal Language

"The distinction between causal and statistical parameters is crisp and fundamental. Causal parameters can be discerned from joint distributions only when special assumptions are made, and such assumptions *must* have causal concepts to them."

-Pearl *Causality* 1st edition pg 39

Why try to draw causal inferences?

"One commonly heard argument is that epidemiologic studies are about association, not causation. According to this proposition, epidemiologists should not worry too much about fishy causal concepts but rather focus their efforts on estimating correct associations. This is certainly a safer strategy but also a dangerous one because it can make much of epidemiology close to irrelevant for both scientists and policy makers."

-Hernán (2005)

Inferring Causation From Association

Statistical *association* between two variables X and Y may be due to:

1. Random fluctuation
2. X caused Y
3. Y caused X
4. X and Y share a common cause
5. The statistical association was induced by conditioning on a common effect of X and Y (as in selection bias).

How we can use this

- Eliminating four of these explanations is usually the goal of an analysis.
- Knowing these five sources of statistical association helps identify the (set of) causal structure(s) that could have generated the observed statistical associations.

Organization

- Drawing and using Causal DAGs
  - Basic rules
  - The d-separation criterion
  - A tangent on collider bias
  - Criteria for testing the null and identifying effects

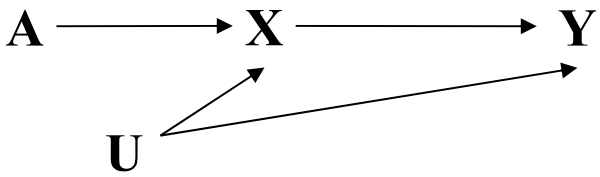
Causal Directed Acyclic Graphs

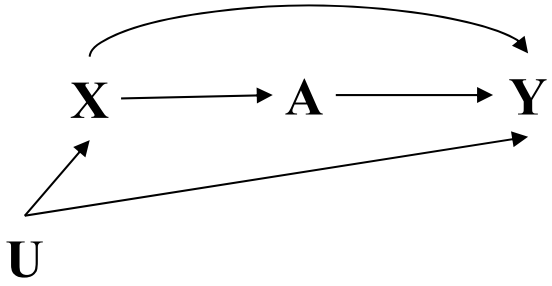
Show your assumptions about the causal relationships among X, Y, and possible covariates in a causal diagram:

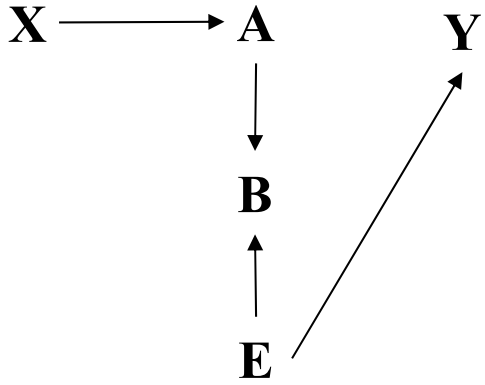
•If two variables shown in the graph have a common cause, you must show the cause in the graph.

•Do not allow causal "loops".

•If you use SEM: draw an SEM, take out the shapes around your variables, and replace correlations with "some association" on the paths





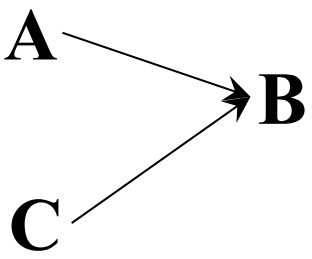


Terminology

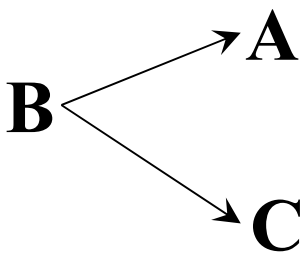
- Descendants
  - The direct or indirect effects of a variable
- Paths
  - A sequence of lines (edges) between two variables, regardless of direction of arrows
  - Not retracing any line segments
- Colliders
  - Common effect of two variables in a path: where the arrows 'collide'.
  - The two causes must both be "on the path".
  - Any variable on a path that is not a collider is a "non-collider".
- Conditioning
  - Examining the distribution of one variable within levels of another
  - Regression adjustment, stratification, restriction

Colliders vs Non-Colliders

Colliders: common *effects*



Non-Colliders:
common *causes* (=confounders)



Or mediators

D-separation

- The assumptions shown in a causal diagram imply that a variable X will be independent of a variable Y, after conditioning on a set of variables {Z} if every path between X and Y is blocked by {Z}.

- The set of covariates {Z} blocks a path if and only if either:
  1. The path contains a non-collider that is in {Z} , or
  2. The path contains a collider which is not in {Z} , and no descendant of the collider is in {Z} .

- If there is an unblocked path linking X and Y, then X and Y will typically be statistically dependent (unless there is a perfectly offsetting balance between two paths).

D-separation: intuition

- There may be many reasons that two variables are associated (some confounding, some mediated causation etc).
- Adjusting for a confounder of the two variables blocks that source of association between two variables
- Adjusting for a mediator between the two variables blocks that source of association between two variables
- Adjusting for a common effect of the two variables *creates* an association between the two variables

A collider anecdote

Some tall people are fast, and some are slow.

Some short people are fast, and some are slow.

Knowing that somebody in the general population is short does not give you information about whether they are fast or slow.

NBA ball players must be either very tall, or very fast.

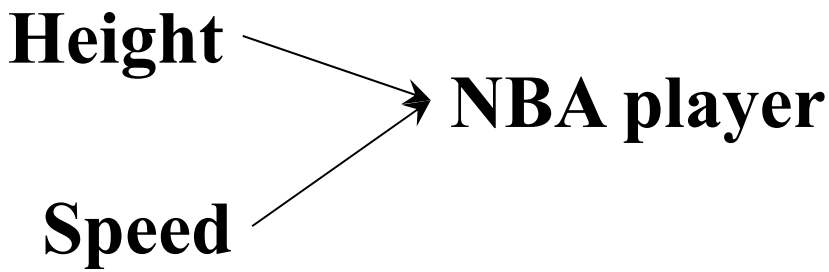
If you know an NBA ball player is short… what do you know about his speed?

A collider anecdote

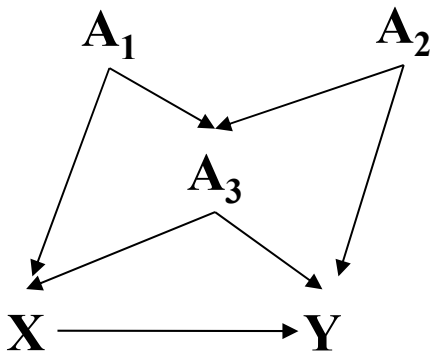
Some tall people are fast, and some are slow.
Some short people are fast, and some are slow.
Knowing that somebody in the general population is short does not give you information about whether they are fast or slow.

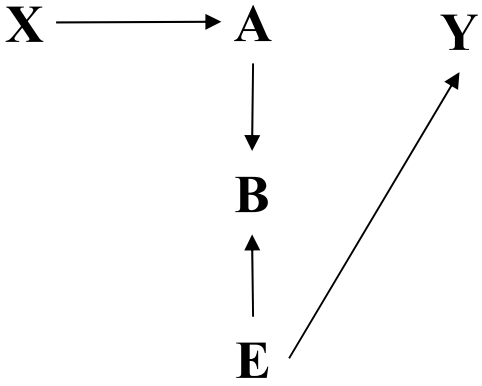
NBA ball players must be either very tall, or very fast.

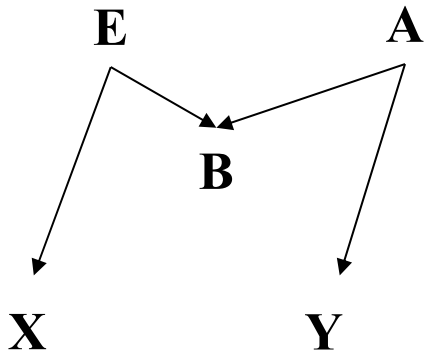
If you know an NBA ball player is short… what do you know about his speed?

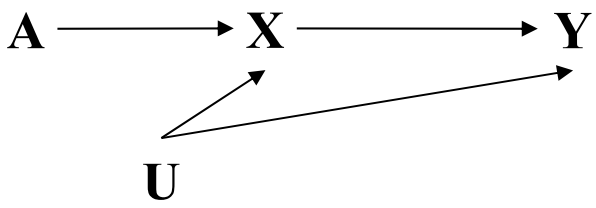


Example Causal Diagrams









DAG example

- Choose a partner and draw a DAG for the exercise…