DAGs and Professional Development Q&A

Questions from Last Class?

- Syllabus updated!

Problem Set 1

- Problem Set 1 has been posted on GitHub
- Includes two required readings, also in that folder
- Due 2/22, as we do not have class next week
- Feel free to email me if you have questions
- Reminder: Solo or group work is acceptable for problem sets

DAGs

- Directed Acyclic Graphs
- Useful as a tool for thinking about causal relationships, dealing with confounding and variable selection, etc.
- Are networks (which we'll cover later) but most people don't really deal with them as if they are
- Variables are nodes while relationships are edges

Directed

- Relationships go in one direction
- X -> Y
- Not X <-> Y
- This is where we depart from a lot of less rules-based conceptual diagrams

Acyclic

- No feedback loops – there's no cycles within the graph
- No X -> Y -> X
- You can break this out like:
  - X_1 -> Y -> X_2
  - This is often a representation of time, and outcomes reinforcing behavior
  - Examples?

Graph

- The mathematics term for network

Notes

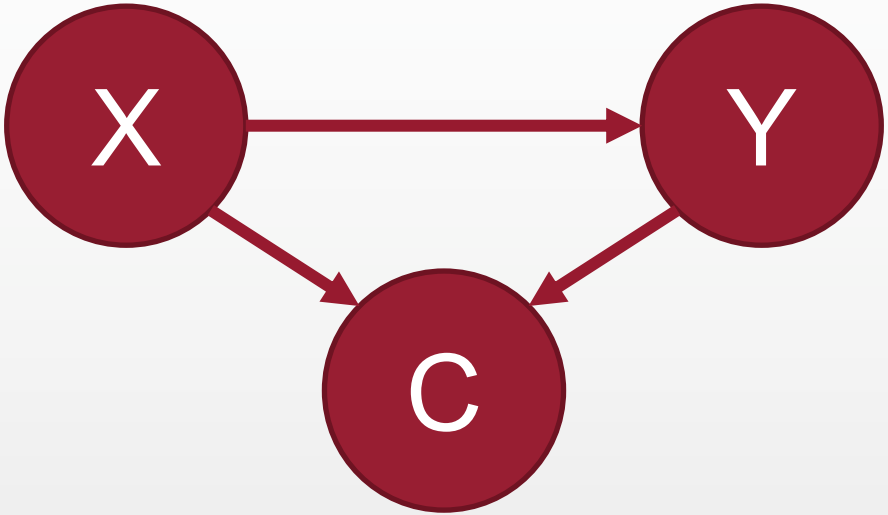
- DAGs are non-parametric – they do not specify the distribution or magnitude of an effect (save that it is there)
- The absence of an arrow is a strong assertion that there is no relationship *for anyone*
  - This is just as strong a causal assertion as asserting there is

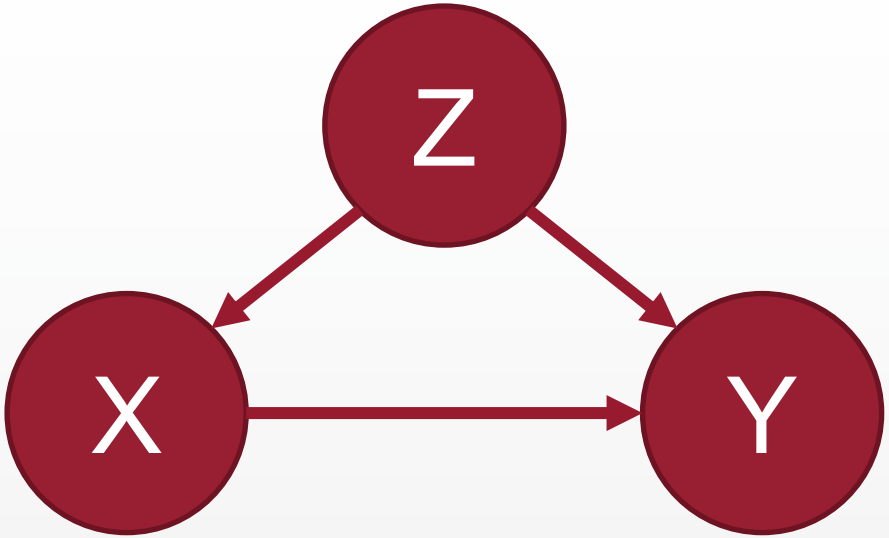
Other Helpful Terminology

- “Backdoor Path”
  - “Non-causal path”
  - Any path where at least one arrow points against the flow of time

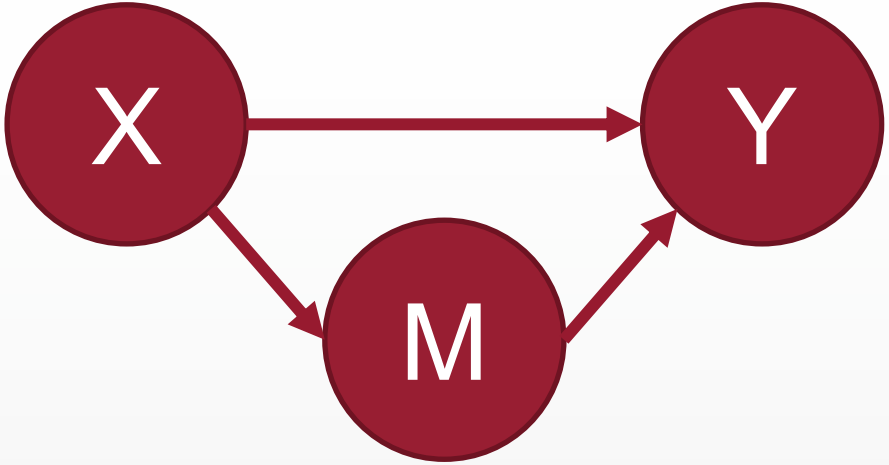
A Simple DAG

Colliders





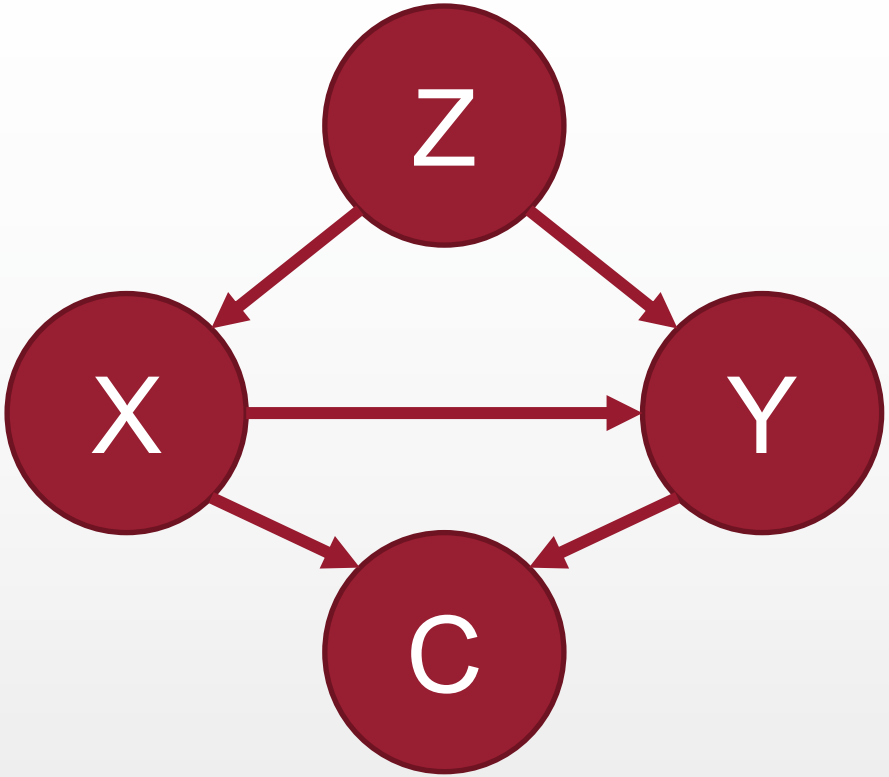
- Z is a common cause of X and Y
- We call these “confounders”, and they can introduce spurious associations between X and Y, or distort a true association’s magnitude

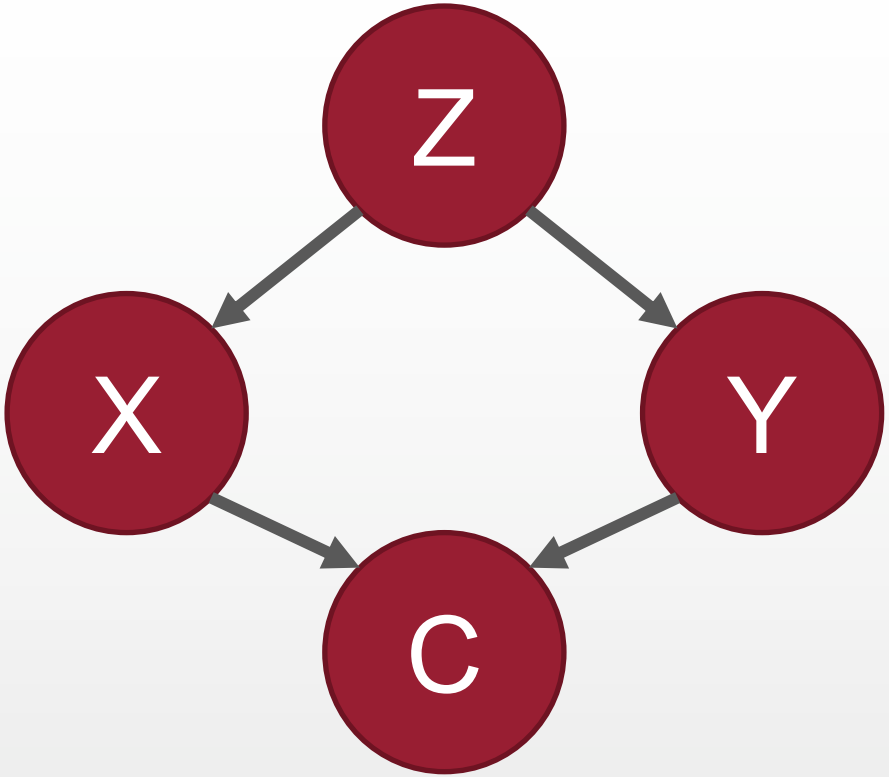


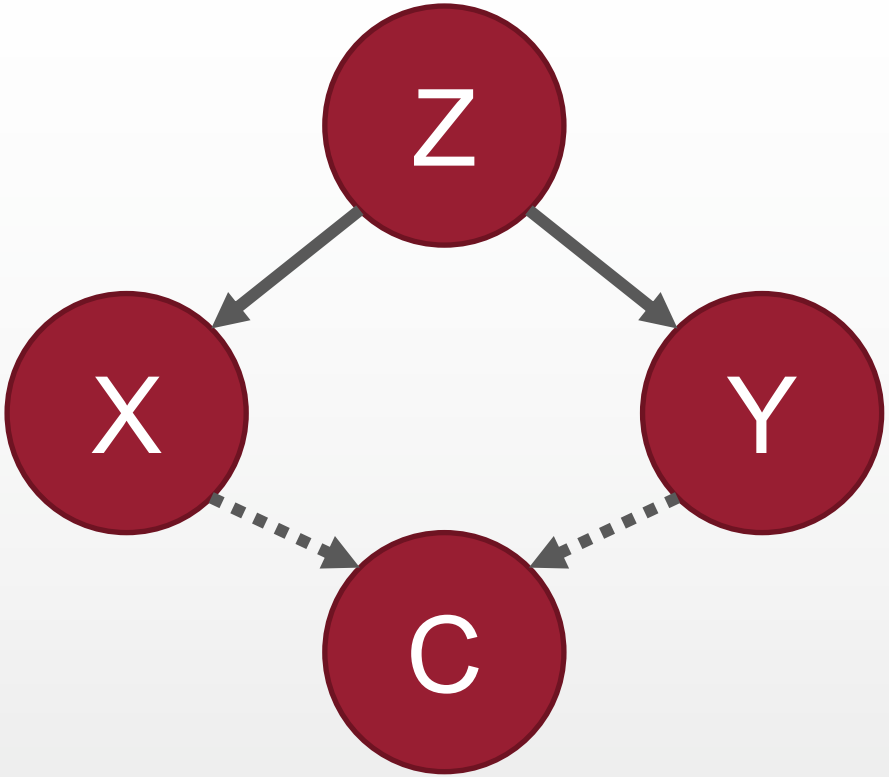
- M is a *mediator* of X and Y
- Some (or all) of the effect of X on Y goes through M
- The direct effect: X -> Y
- The indirect effect: X -> M – Y
- Total effect: X -> Y + X -> M -> Y

Basic DAG Analysis

- Remove arrows from Exposure to Outcome
- Identify all backdoor paths
- Identify whether each path is “Open” or “Closed”
  - Closed paths have a collider on them
- All remaining “Open” paths are confounded, and the variables on them are confounders

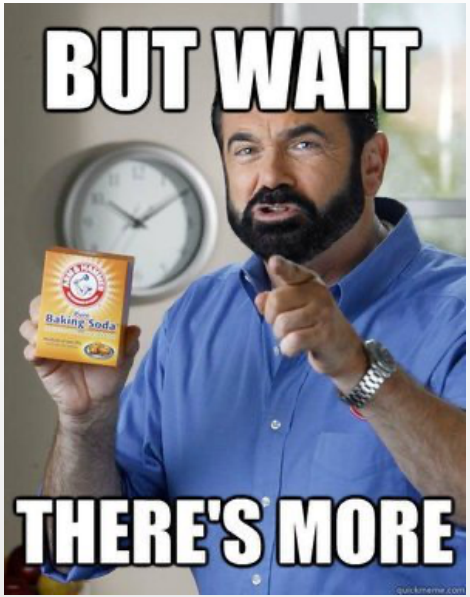

Z
X
Y
C


Z
X
Y
C


Z
X
Y
C

We Did It!

- We all understand DAGs!



What Do We Do About Confounders?

- We need to condition on confounders in order to remove the association between them and the exposure/outcome
  - Think of this as erasing one of the arrows
- Lots of ways to do this, which we'll discuss in the regression-related classes

Do I Need to Get All of Them?

- Nope!
- Controlling just the confounders needed to close backdoors paths is enough
- Controlling for more is statistically inefficient
  - “Minimum cut set”

What Do We Do About Colliders?

- We don't do anything
- Colliders are already blocking their path – leave them to it
- Conditioning on a collider can *induce* a relationship that isn't there
- This is why "just adjust for everything" is so dangerous statistically

Conditioning on Colliders

- Example by Sackett:
  - 257 hospitalized patients
  - Association between locomotor disease and respiratory disease (OR ~ 4.0)
  - In the general population, this association disappears
  - Locomotor disease and Respiratory disease are *independent causes* of hospitalization
  - By only looking in one strata of “Hospitalized”, Sackett conditioned on a collider

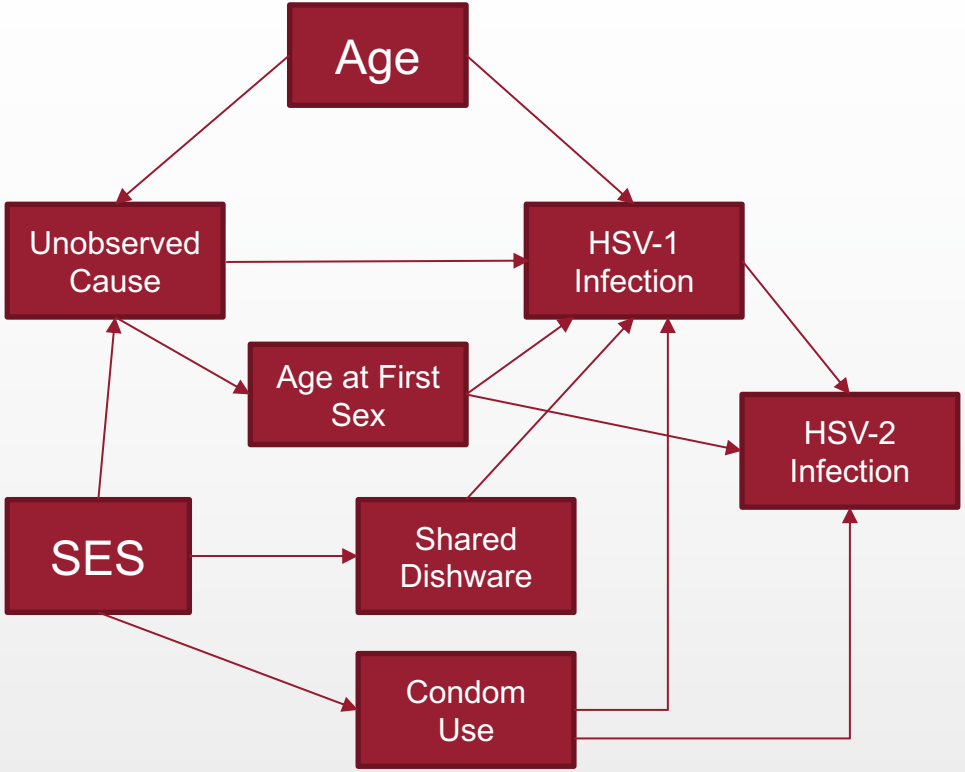
Conditioning on Colliders = Selection Bias

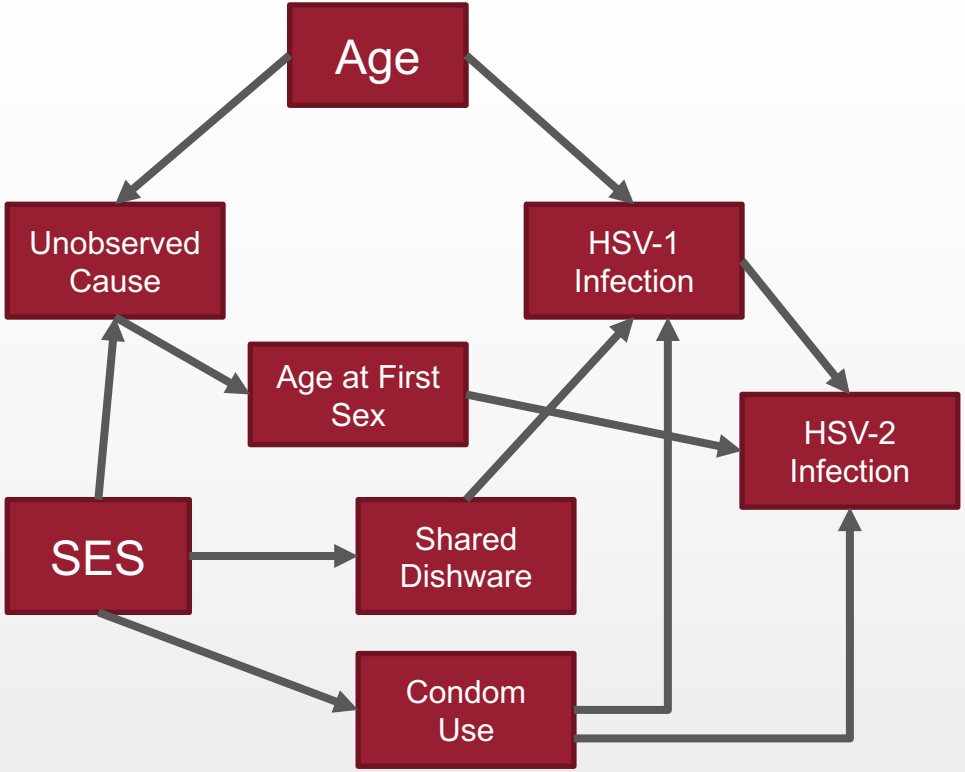
What About Mediators?

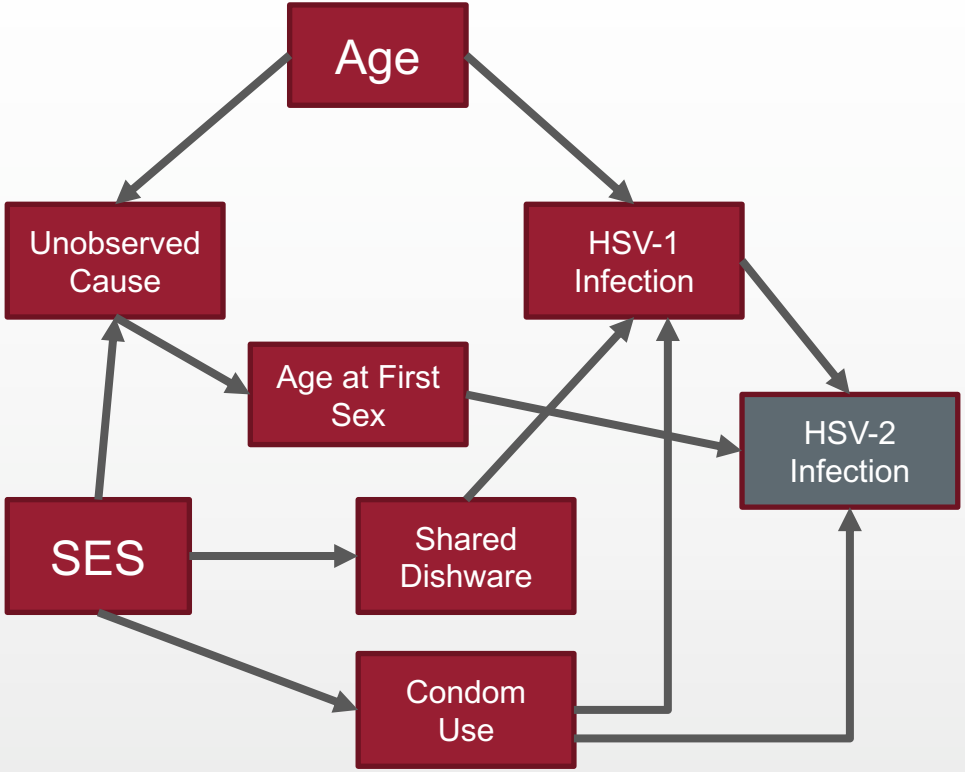
- The analysis of causal mediators is a vibrant and active field
- There is a temptation to want to control for mediators to estimate “the direct effect” without all that other junk in the way
- But how much are you losing?
- What policy tools are you skipping?
- Generally, err on the side of estimating total effects

A More Complicated Example

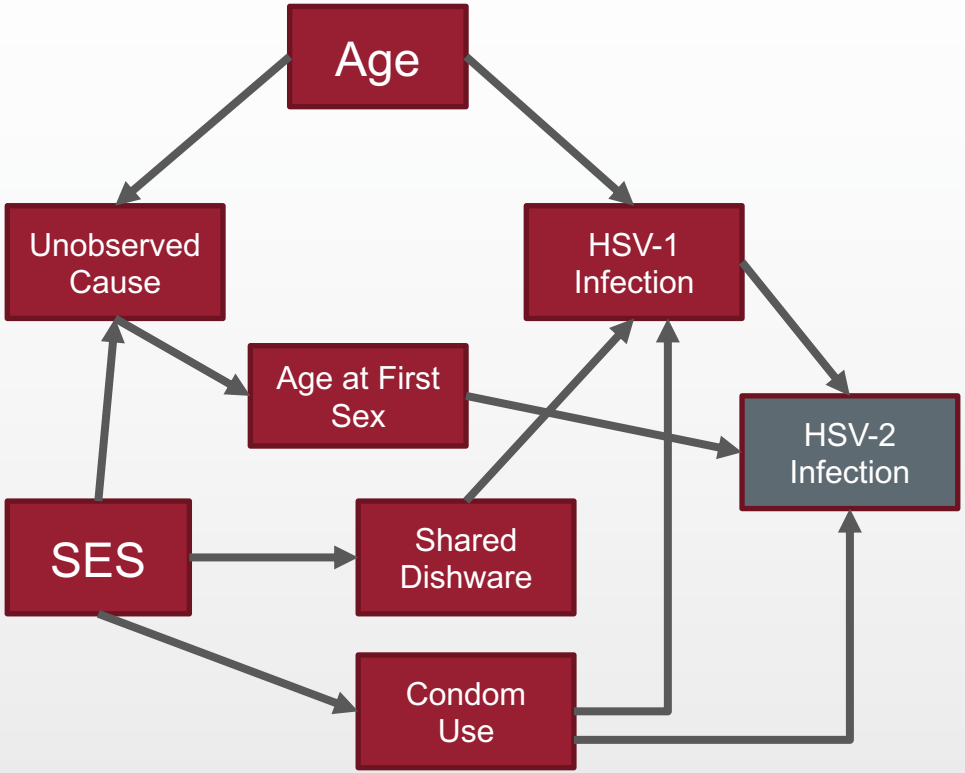
- A study observed an elevated incidence of HSV-1 seroprevalence in one study site as compared to another in Thailand
- This was an ancillary finding in an HPV cohort, so most of the variables that were collected with STI-related


Age
Unobserved Cause
HSV-1 Infection
Age at First Sex
HSV-2 Infection
SES
Shared Dishware
Condom Use


Age
Unobserved Cause
HSV-1 Infection
Age at First Sex
HSV-2 Infection
SES
Shared Dishware
Condom Use


Age
Unobserved Cause
HSV-1 Infection
Age at First Sex
HSV-2 Infection
SES
Shared Dishware
Condom Use

What Do We Adjust For?



Where Do We Get These Variables?

- Subject-matter expertise
- Previous studies
- DAGs are admittedly of limited utility in very emerging fields
- Should we use our data?
  - Arguably no!
  - DAGs are causal arguments – if you ask if you should control for the X -> Z relationship, you now have to make a new DAG for X & Z, etc.

What If I Can't Decide?

- Disagreements as to which way an arrow goes could make a big difference – is this a confounder or a collider?
- Which cut-set to use – is one variable measured better, is someone anticipating a reviewer will dig their heels in?
- Two members of my committee are fighting!

WHY NOT BOTH?

Multiple DAGs

- The use of two or more DAGs can help you estimate an effect in the presence of structural/causal uncertainty
- This then leads to a causal “Choose Your Own Adventure”
- Is an effect still strong despite several approaches?
- Is it dependent on only one possible causal setup being true?
- This is, admittedly, harder to publish.

Software!

- Daggity
  - R and Standalone
- ggDAG
  - Just in R
- Good for both plotting and some tools to help you deal with complex DAGs

Are DAGS Useful?

- This seems like a lot of work
  - It is!
- But DAGs are powerful tools for both study design and analysis – though there are constraints as to how well they can work
- On occasion, the complexity can seem overwhelming
  - This is a hint that you are working in a complex space
- DAGs are *very* powerful as explanatory tools to explore studies
  - Why might a weird finding have arose?
  - "Yes Reviewer 2, we thought of that, but it's a collider, so we're not controlling for it."

Questions?

Break Time

Questions?

Professional Development Q&A

- Normally, there's a casual professional development aspect to Vet Clin 570, the Infectious Diseases Journal Club
- That's been canceled this semester, both because of time constraints and because the journal club format doesn't lend itself well to Zoom
- We have some time for questions this class

Recommended Resources

- #epitwitter
  - Admittedly somewhat less useful during the pandemic
- Acadames (Podcast)
  - Expressly targeted toward women in academia
  - I think this is a podcast that's worth listening to regardless of who you are
- Shiny Epi People (Podcast)
  - Because epidemiologists are human beings