Introduction to Causal Directed Acyclic Graphs

Amber W. Trickey, PhD, MS, CPH
Senior Biostatistician

S-SPIRE Works in Progress
January 28, 2019

S-SPIRE Stanford–Surgery Policy Improvement Research & Education

@StanfordSPIRE

Overview

- What are DAGs & why do we need them?
- DAG rules & conventions
- How to construct a DAG
  - Which variables should be included?
  - How to determine covariates for adjustment?
- Examples: manual + DAG online tool
- Build your own DAG

Observational Health Services Research

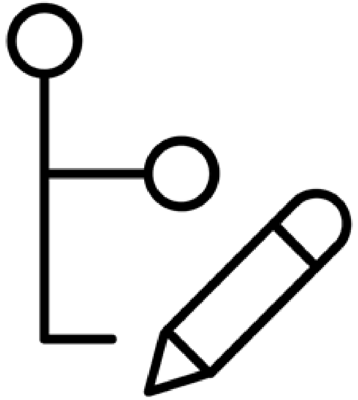
- Big HSR datasets are observational
  - Medicare
  - HCUP: NIS, NEDS, NRD, SIDs
  - Truven, Optum
  - EMR: STARR
  - Clinical Registries: NSQIP, VQI
- Observational comparative effectiveness [1]
  - Treatments not assigned, determined by mechanisms of routine practice
  - Actual mechanisms are often unknown
  - However researchers can (and should) speculate on the treatment assignment process or mechanism
- Problem: correlation ≠ causation

[1] *2013 AHRQ Developing a protocol for observational comparative effectiveness research: a user's guide*

S-SPIRE

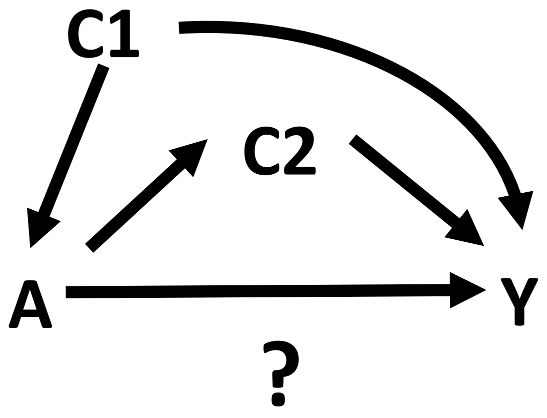
Causal Graphs: Helpful Tools

1. Illustrate sources of bias
2. Determine whether the effect of interest can be identified from available data
3. Causal graphs are based on assumptions (but so are analytic models)



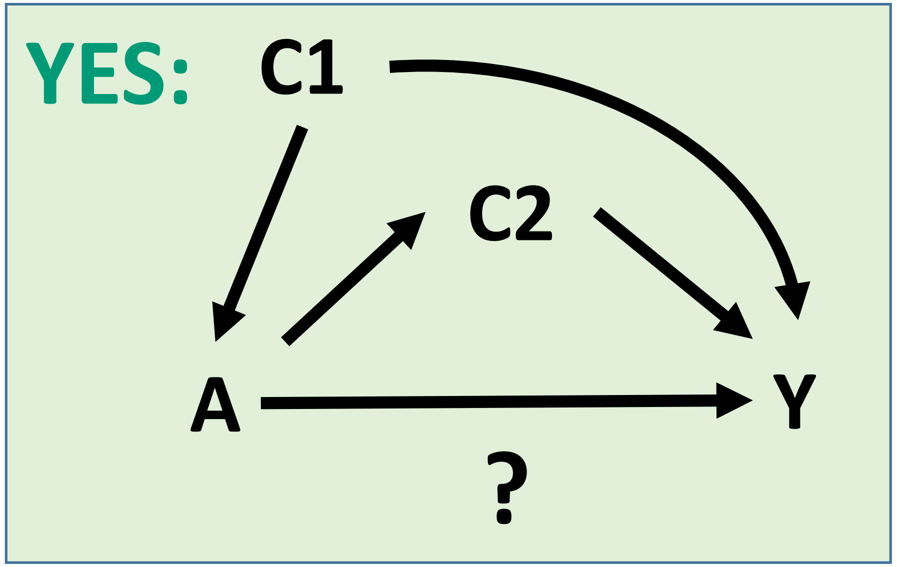
What are Directed Acyclic Graphs?

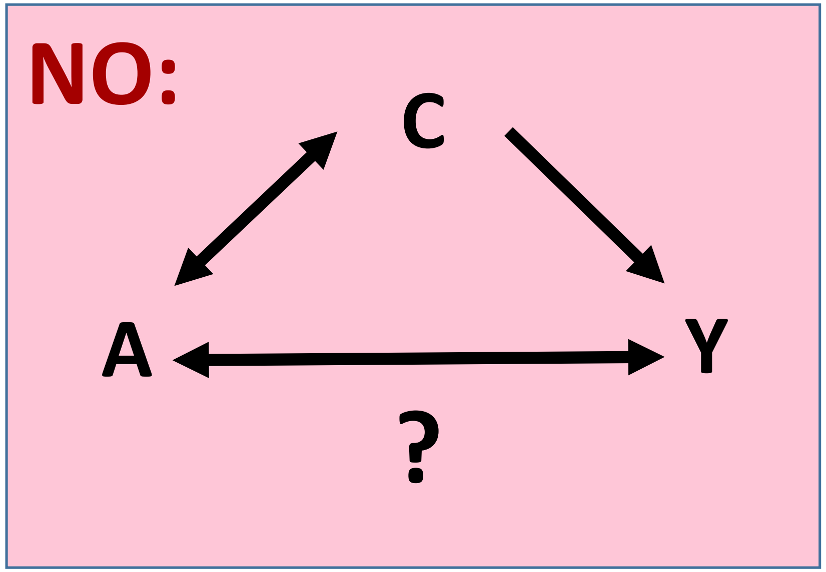
- Computer science: data structure
- Markov models: visualization
- Epidemiology: Causal DAGs are systematic representation of causal relationships
- Useful tools to represent assumptions & known relationships
  - plan analytic approach
  - reduce bias



Directed & Acyclic

- Directed: point from cause to effect
  - Causal effects cannot be bidirectional
- Acyclic: no directed path can form a closed loop





Why do we need DAGs?

- Clarify study question & relevant concepts
- Explicitly identify assumptions
- Reduce bias
    - Separate individual effects
    - Ascertain appropriate covariates for statistical analysis
- Estimate required analysis time

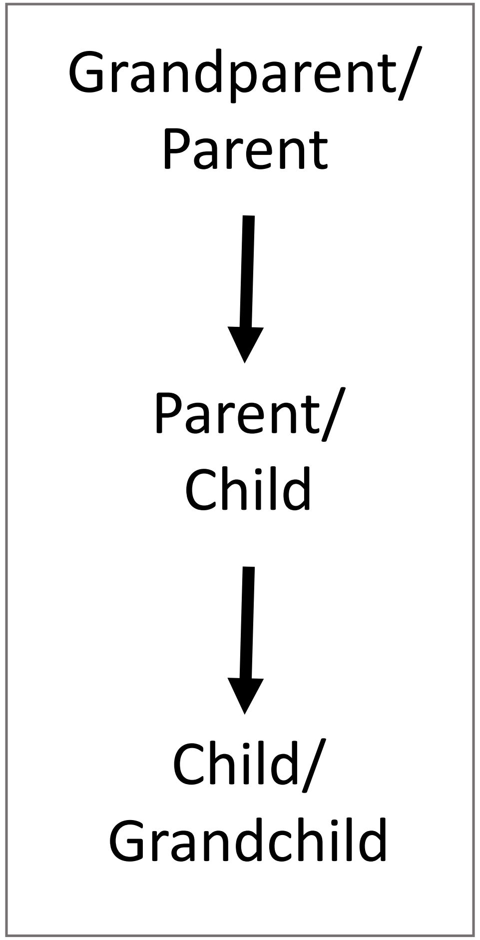
We can assist with DAG creation and covariate assessment

DAG Rules

- All common causes are represented
- No arrow = no causal effect
- Time flows left to right
    - A (or E) = Exposure / Treatment / Intervention / Primary IV
    - Y (or D) = Outcome
    - C = Covariates / Confounders
    - U = Unmeasured relevant variables
- Confounders can be grouped for notation

Glossary – Genealogy

- **Parent**: a direct cause of a particular variable
- **Ancestor**: a direct cause (i.e. parent) or indirect cause (e.g. grandparent) of a particular variable
- **Child**: the direct effect of a particular variable, i.e. the child is a direct effect of the parent
- **Descendant**: a direct effect (i.e. child) or indirect effect (e.g. grandchild) of a particular variable
- **Common Cause**: a covariate that is an ancestor of two other covariates



S-SPIRE

How to Construct a DAG: Variables to Include

How to construct a DAG

Step 1: Articulate the research question

- Start the DAG with your:
  - A / treatment / exposure / primary IV (cause)
  - Y / dependent variable / endpoint / outcome (effect)
- Indicate the research question with a “?”

Exposure / Treatment
“Cause”
A

$\xrightarrow{\quad ? \quad}$

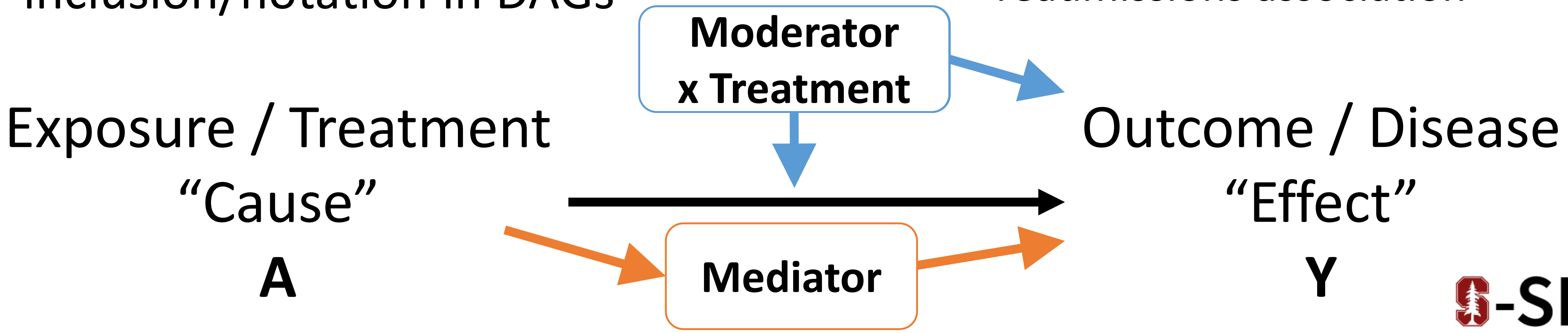
Outcome / Disease
“Effect”
Y

How to construct a DAG

Step 2: Consider important variables embedded in the question

- **Moderator**: affects the direction and/or strength of the relation between A & Y (AKA effect modifier, statistical interaction)
  - e.g. **gender** differences in surgical history-opioid relationship
- Some disagreement on inclusion/notation in DAGs [2]

- **Mediator**: a variable within the causal pathway between the treatment and outcome. Treatment (A) influences the mediator, which in turn influences the outcome.
  - e.g. **complications** in frailty-readmissions association

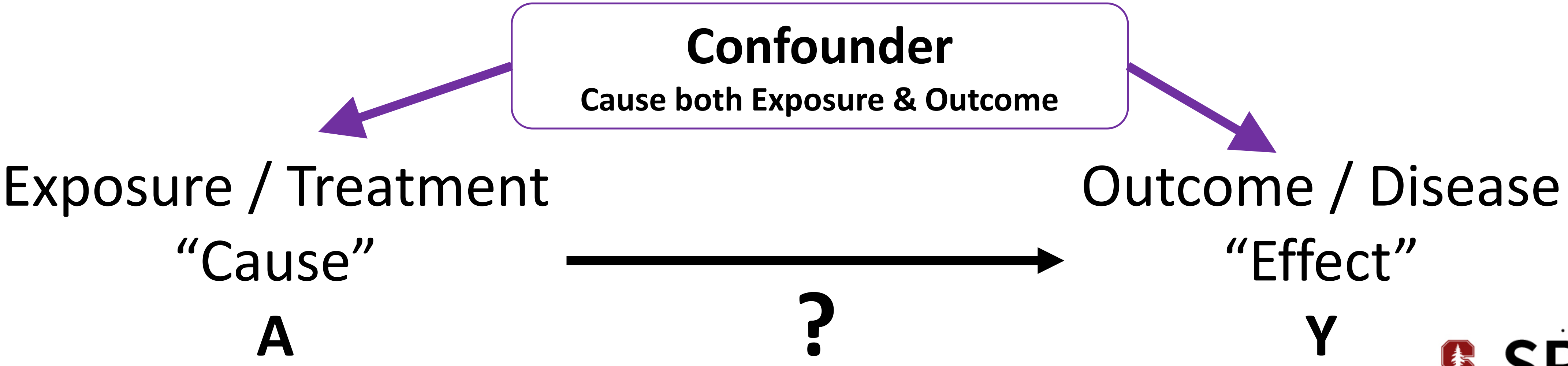


S-SPIRE

How to construct a DAG

Step 3: Consider confounding variables

- Variables that **confound** the relationship you are evaluating
- Confounders are causes of both the treatment (A) & the outcome
  - e.g. **age, gender, race, insurance**
- Add confounders to DAG, considering causal mechanism



How to construct a DAG

Step 4: Consider other relevant variables

Which Variables Should be Included?

- All common causes of any 2 variables in the DAG
- Unmeasured (and unmeasurable) common causes
- Selection variables, i.e. inclusion criteria

Not Required in Causal DAGs:

- Variables that cause Y but not A
  - May be included if desired, e.g. for comparison to other studies which adjusted for the variable

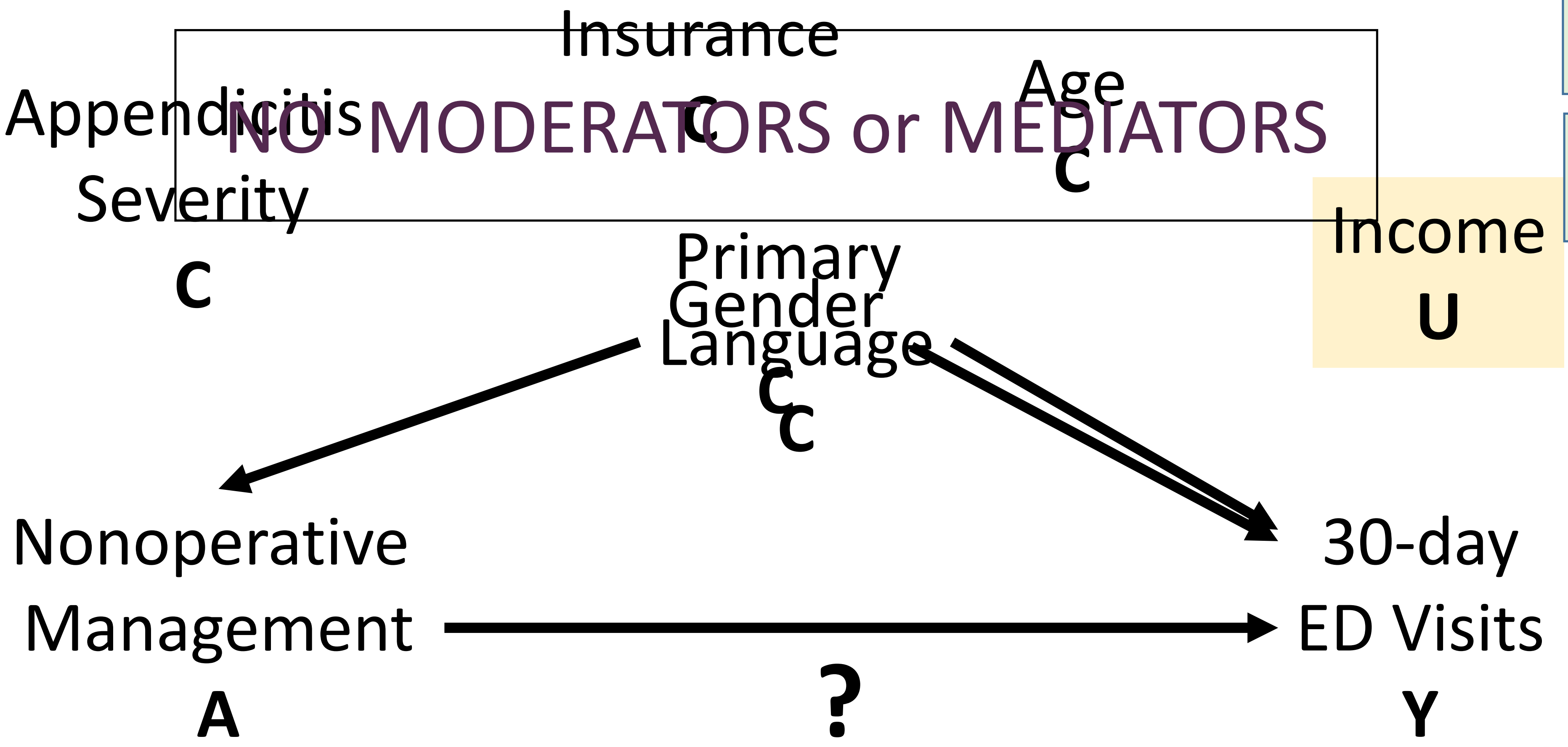
Example: Uncomplicated Appendicitis[3]

Step 1: Research Q

Step 2: Med/Mods

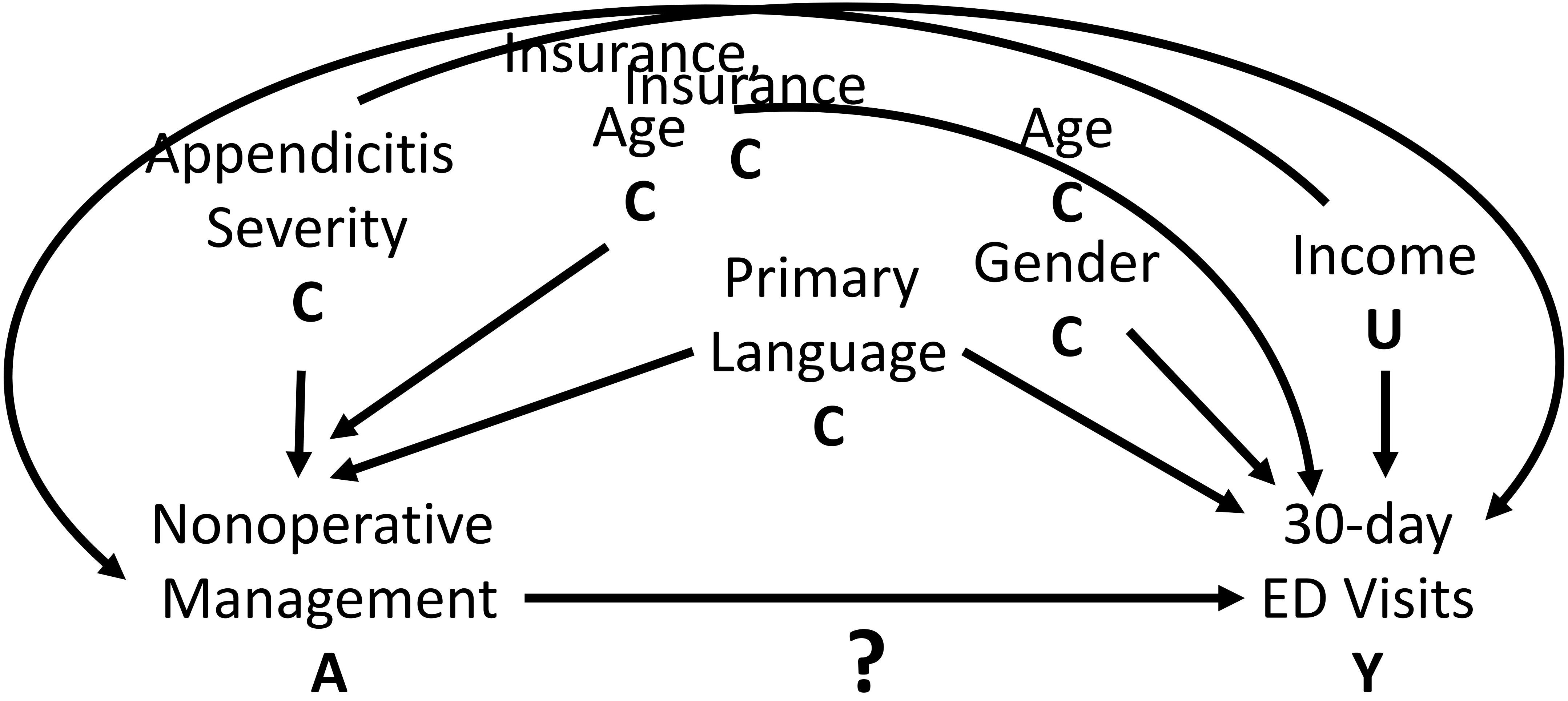
Step 3: Confounds

Step 4: Others



S-SPIRE

Example: Uncomplicated Appendicitis[3]



S-SPIRE

Keep in mind

- Assumptions must be made
- Every analysis has built-in assumptions
- DAGs make them explicit, represent *your* model of relationships between variables
- Often more than 1 appropriate DAG
- Alternate DAGs can make excellent sensitivity analyses

S-SPIRE

How to Construct a DAG: Determine Covariates for Adjustment

Glossary – Causes, Effects, Associations

- **Back Door Path**: a connection between A & Y that does not follow the path of the arrows.

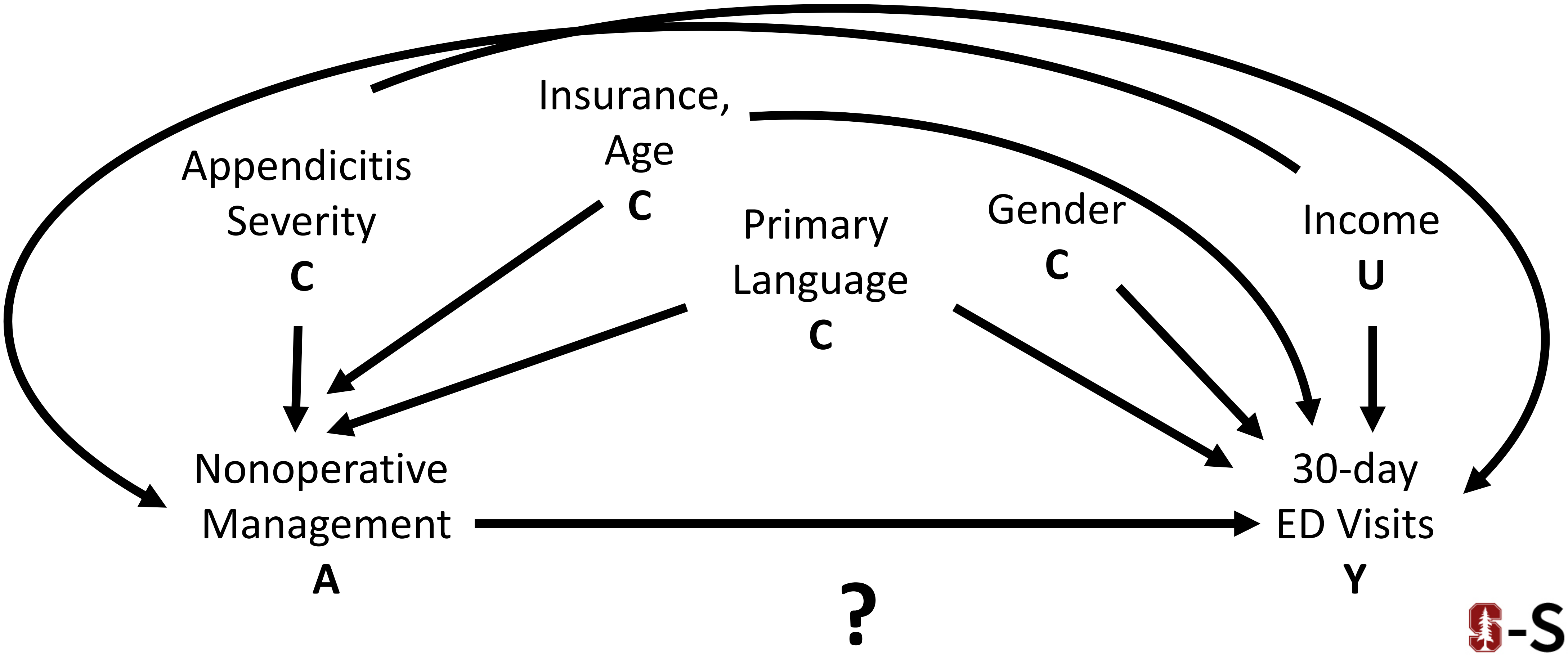
- **Common Effect** (also known as **Collider**): a covariate that is a descendant of two other covariates. The term collider is used because the two arrows from the parents "collide" at the descendant node.

- **Conditioning**: Conditioning on a variable means using either sample restriction, stratification, adjustment or matching to examine the association of A & Y within levels of the conditioned variable.
  - Other terms "adjusting" or "controlling" suggest a misleading interpretation of the statistical model. [4]

S-SPIRE

Back Door Path

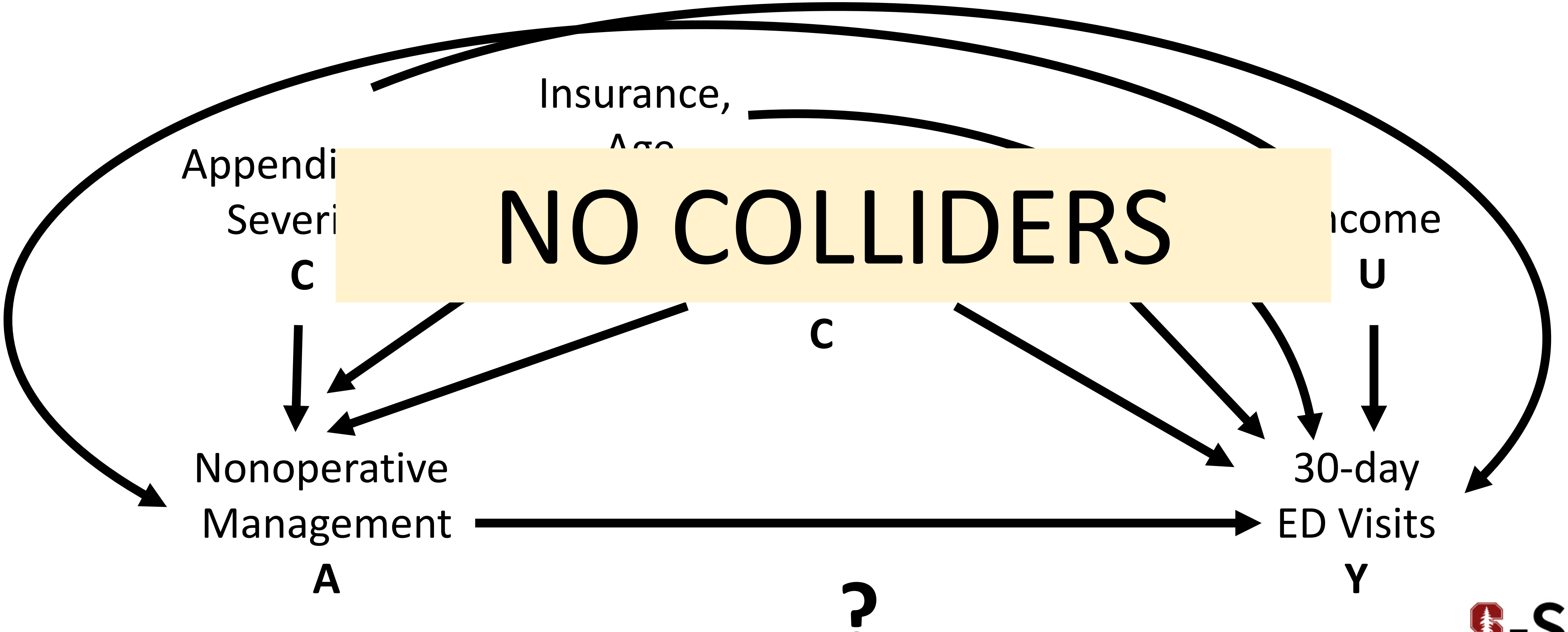
- A connection between A & Y that does not follow the path of the arrows.
- Open back door path → confounding bias, i.e. association b/t A & Y ≠ causal effect



S-SPIRE

Common Effect or Collider

- A covariate descendant of two other covariates.
- Two arrows from the parents "collide" at the descendant node.



NO COLLIDERS

Conditioning

Conditioning can mean:

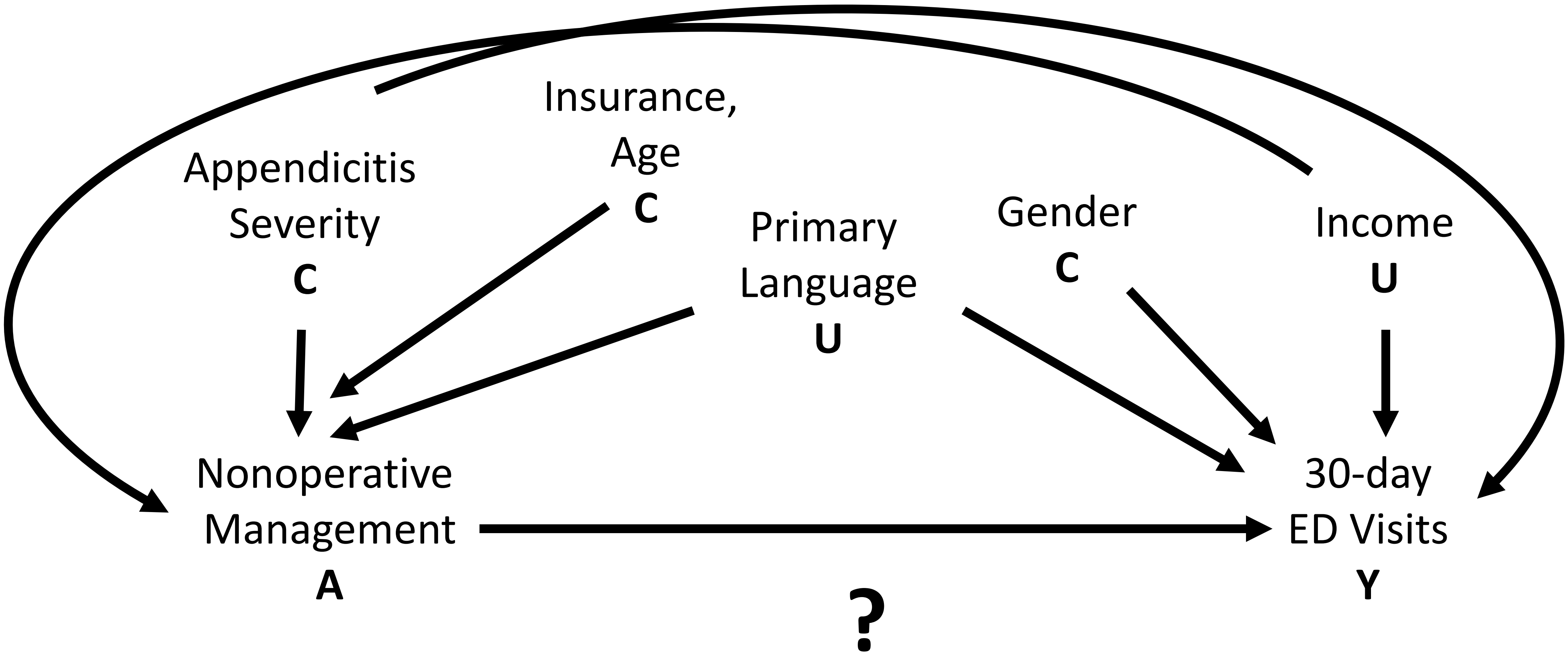
- adjusting, restricting, stratifying, matching

Draw a $\boxed{\text{box}}$ around the conditioned variables

1. Conditioning on a variable in an open backdoor path removes the non-causal association (i.e. controls for confounding)
2. Conditioning on a collider opens the path that the collider was blocking
3. Conditioning on a variable in the causal pathway (mediator) removes part of the causal effect

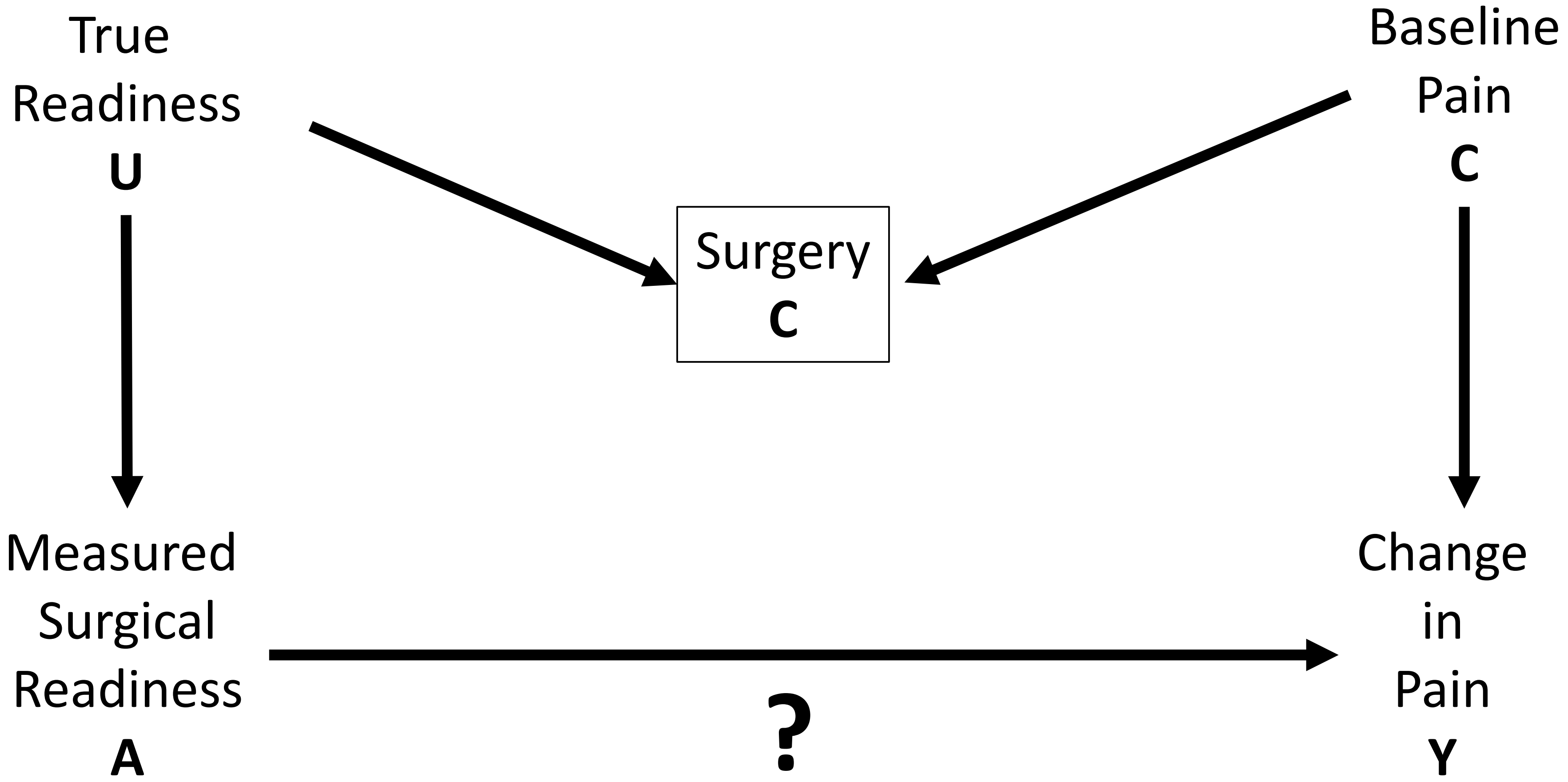
Conditioning

Remaining potential confounding due to unmeasured language & income



S-SPIRE

Collider Bias Example: Surgery for Low Back Pain



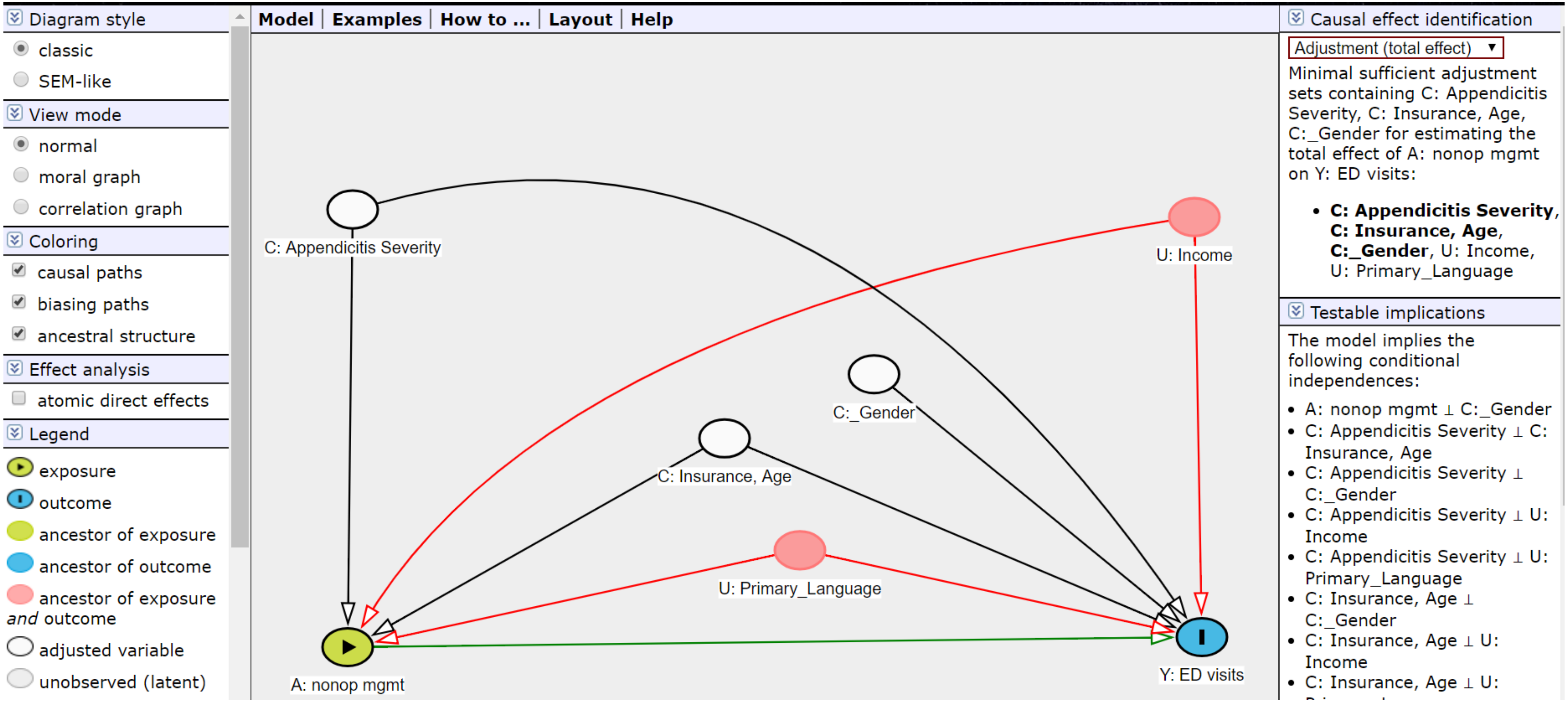
Glossary – Structural Approach to Bias

- **Confounding bias**: common cause of A & Y that is not "blocked“ by conditioning on other specific covariates
- **Collider bias:** general phenomenon involving conditioning on common effects.
- **Selection bias**: a particular type of collider bias in which the common effect is selection into the study; occurs when a common effect is conditioned such that there is now a conditional association between A & Y (e.g. Berkson's Bias, loss to f/u, missing data, healthy worker bias)
- **Berkson’s bias:** a particular type of selection bias in which selection of cases into the study depends on hospitalization, and the treatment is another disease, or a cause of another disease, which also results in hospitalization [5]

S-SPIRE

Online Tool [6] www.dagitty.net

Save your code!
(.docx or .txt)



S-SPIRE

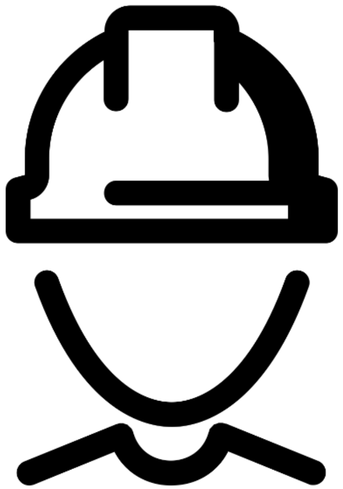
Summary: AHRQ CER User's Guide [7]

Guidance	Key Considerations
Develop a simplified DAG to illustrate concerns about bias.	• Use a DAG to illustrate and communicate known sources of bias, such as important well known confounders and causes of selection bias.
Develop complete DAG(s) to identify a minimal set of covariates.	• Construction of DAGs should not be limited to measured variables from available data; they must be constructed independent of available data. • The most important aspect of constructing a causal DAG is to include on the DAG any common cause of any other 2 variables on the DAG. • Variables that only causally influence 1 other variable (exogenous variables) may be included or omitted from the DAG, but common causes must be included for the DAG to be considered causal. • Identify a minimal set of covariates that blocks all backdoor paths and does not inadvertently open closed pathways by conditioning on colliders or descendants.

*Sauer, VanderWeele. "Supplement 2: Use of directed acyclic graphs." (2013).

S-SPIRE

Questions?



Build your own DAG!

References

1. Velentgas P, Dreyer NA, Nourjah P, Smith SR, Torchia MM, editors. Developing a protocol for observational comparative effectiveness research: a user's guide. Government Printing Office; 2013 Feb 21.
2. Weinberg CR. Can DAGs clarify effect modification?. Epidemiology 2007 Sep;18(5):569.
3. Sceats LA, Trickey AW, Morris AM, Kin C, Staudenmayer KL. Nonoperative Management of Appendicitis: A Retrospective Cohort Analysis of Privately Insured Patients. JAMA Surgery, in press.
4. Shrier I, Platt RW. Reducing bias through directed acyclic graphs. BMC medical research methodology. 2008 Dec;8(1):70.
5. Pearce N, Richiardi L. Commentary: three worlds collide: Berkson's bias, selection bias and collider bias. International journal of epidemiology. 2014 Feb 28;43(2):521-4.
6. Textor J, van der Zander B, Gilthorpe MS, Liśkiewicz M, Ellison GT. Robust causal inference using directed acyclic graphs: the R package 'dagitty'. International journal of epidemiology. 2016 Dec 1;45(6):1887-94.
7. Sauer B, VanderWeele TJ. "Use of directed acyclic graphs." In: Velentgas P, Dreyer NA, Nourjah P, Smith SR, Torchia MM, editors. Developing a protocol for observational comparative effectiveness research: a user's guide. Government Printing Office; 2013 Feb 21.
8. Greenland S, Pearl J, Robins JM. Causal diagrams for epidemiologic research. Epidemiology. 1999 Jan 1:37-48.
9. LaCalle E, Rabin E. Frequent users of emergency departments: the myths, the data, and the policy implications. Annals of emergency medicine. 2010 Jul 1;56(1):42-8.

S-SPIRE

Additional Resources

1. Howards PP, Schisterman EF, Poole C, Kaufman JS, Weinberg CR. “Toward a clearer definition of confounding” revisited with directed acyclic graphs. American journal of epidemiology. 2012 Aug 17;176(6):506-11.
2. Williams TC, Bach CC, Matthiesen NB, Henriksen TB, Gagliardi L. Directed acyclic graphs: a tool for causal studies in paediatrics. Pediatric research. 2018 Jun 4.
3. Suttorp MM, Siegerink B, Jager KJ, Zoccali C, Dekker FW. Graphical presentation of confounding in directed acyclic graphs. Nephrology Dialysis Transplantation. 2014 Oct 16;30(9):1418-23.
4. Baron RM, Kenny DA. The moderator–mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. Journal of personality and social psychology. 1986 Dec;51(6):1173.

S-SPIRE