Causal inference with Directed Acyclical Graphs

CARMA Webcast presentation
January 26, 2024

Heiko Breitsohl

What are Directed Acyclical Graphs (DAGs)?

UNIVERSITÄT KLAGENFURT

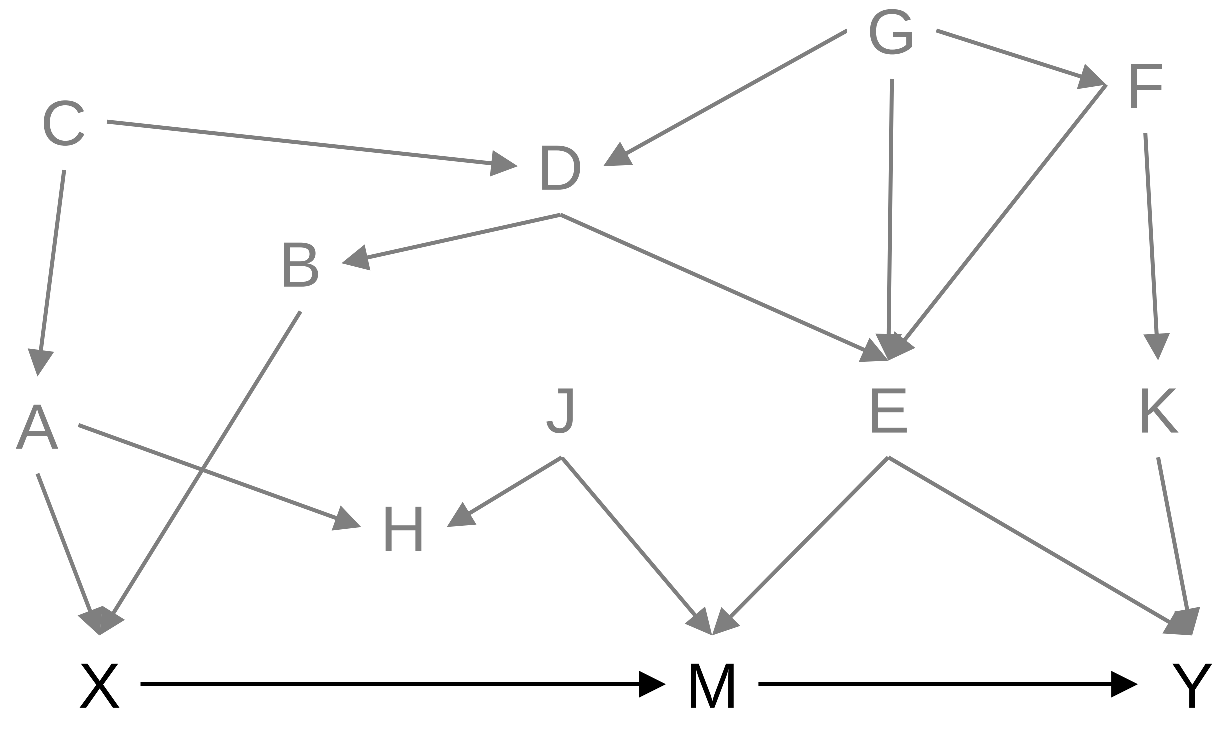
- Visual representations of causal assumptions about some phenomenon

- Purpose:
  To identify requirements for valid empirical causal inference of an effect of interest
  1. Build a graph of the causal network surrounding the effect of interest
  2. Identify potential sources of bias (and therefore false conclusions)
  3. Identify options for removing bias (if any)
  4. Consider implications for study design

Basic elements of DAGs: Variables and arrows

UNIVERSITÄT KLAGENFURT

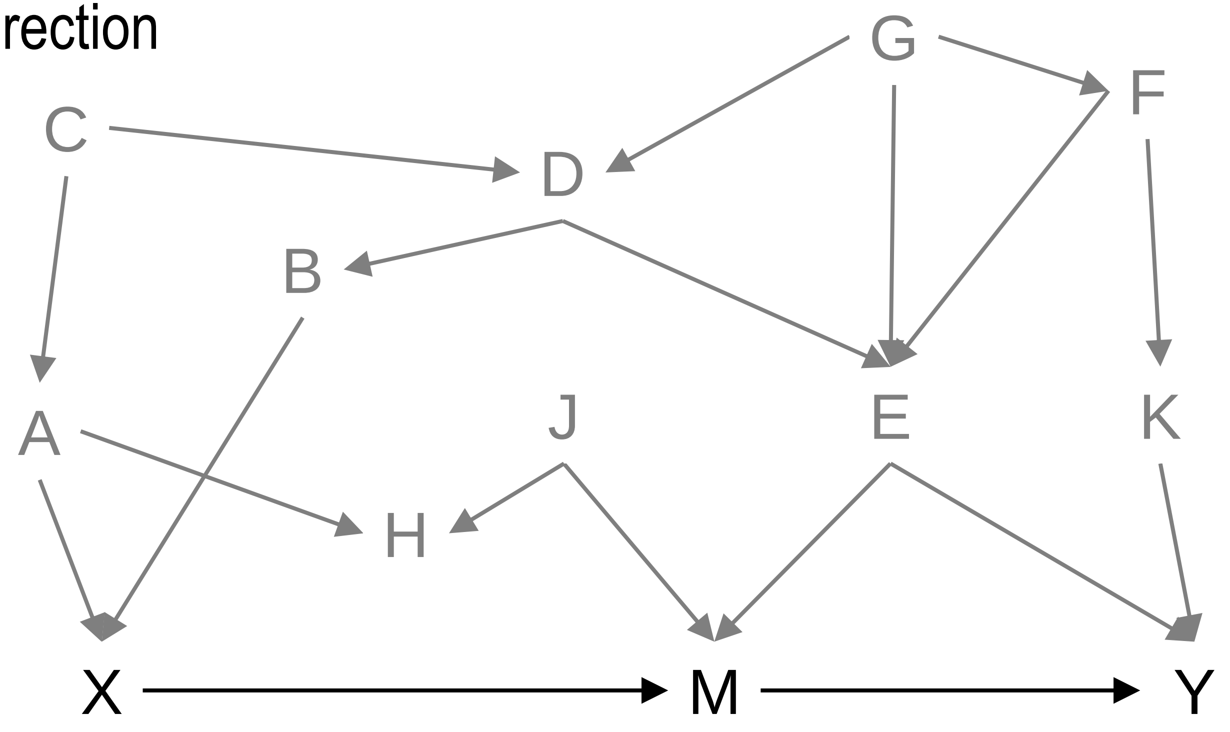
- Variables (nodes, vertices) relevant to the causal network
- Arrows (edges) representing causal effects
- Strength of assumptions
  - Presence of variable/arrow:
    Weak assumption of *possible* relevance
  - Absence of variable/arrow:
    Strong assumption of *certain* irrelevance



Basic elements of DAGs: Paths

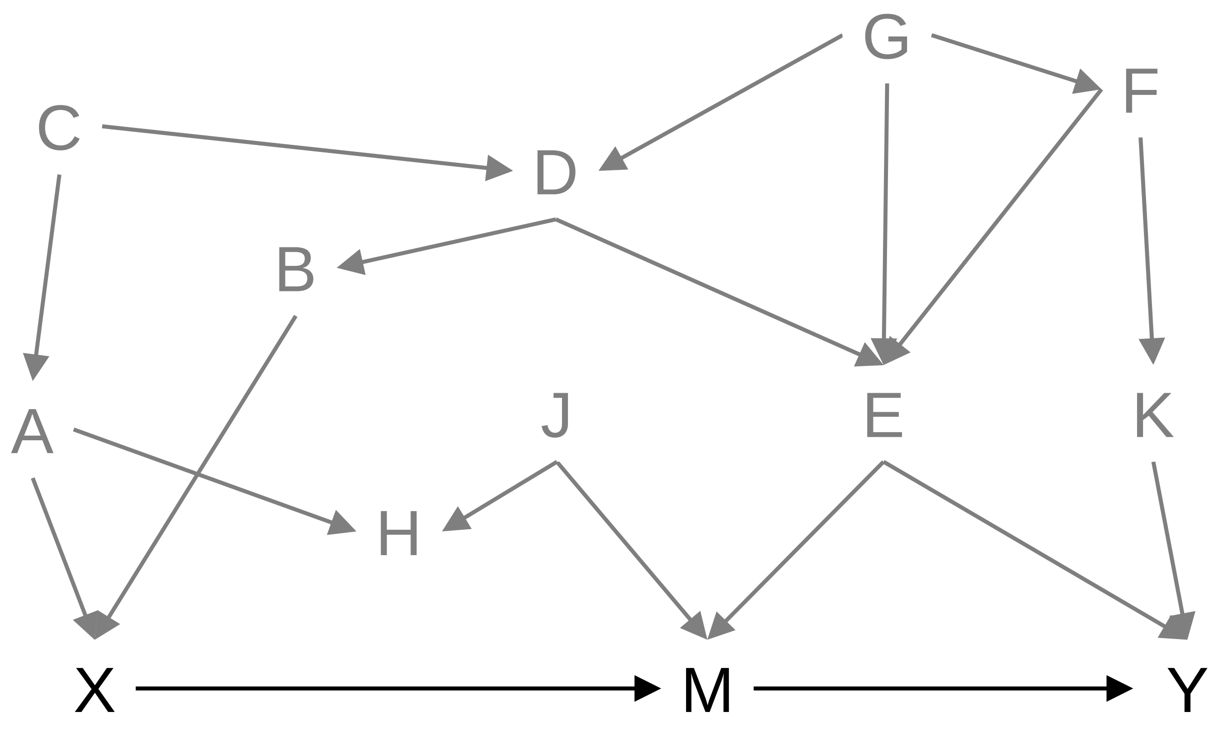
- Path: Sequence of variables connected by arrows (of *any* direction)

- *Causal* path: Arrows all point in the same direction
- *Non-causal* path: At least two arrows point in different directions
  - <u>Open</u> (unblocked, active, d-connected): Generates a *spurious* association between its endpoints
  - <u>Closed</u> (blocked, inactive, d-separated): Does *not* generate a spurious association



Basic elements of DAGs: “Roles” of variables

- Specific “roles” played by variables in a specific DAG:
  - *Mediator*: A variable in the middle of a “chain”, e.g.: X -> **M** -> Y
  - *Confounder*: A variable in the middle of a “fork”, e.g., A <- **C** -> D
  - *Collider*: A variable in the middle of an “ inverted fork”, e.g., A -> **H** <- J



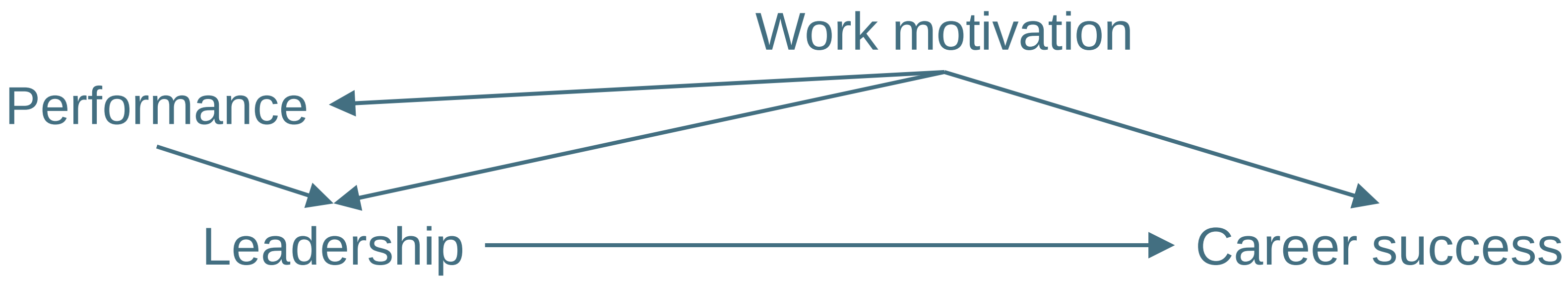
- Core principles:
  - *Directed*: No double-headed arrows
  - *Acyclical*: No variable may cause itself
  - Non-parametric

Simple, fictional, example

- What is the causal effect of leadership behavior on follower career success?

- What (unstated) assumptions are we making here?

- What paths (from L to C) are in this DAG?
  - L -> C
  - L <- W -> C
  - L <- P <- W -> C



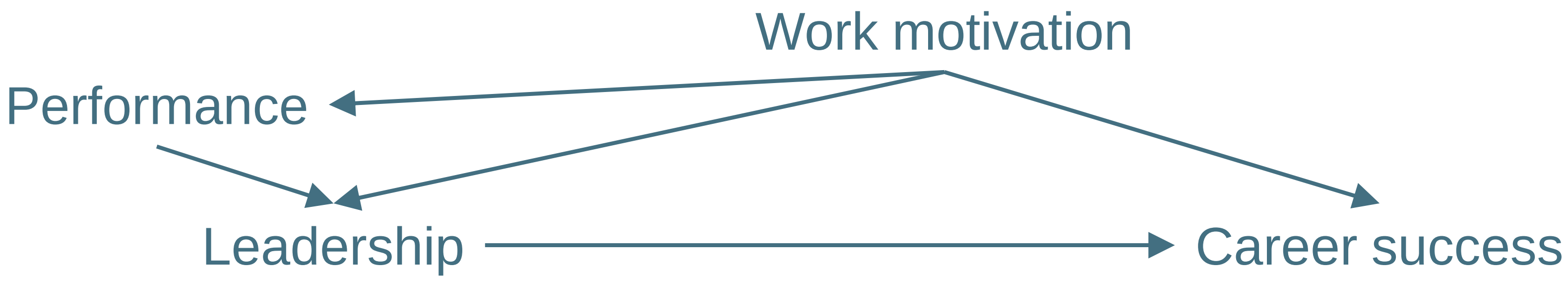
Confounding as a key issue for causal inference

- Confounding
  - Induces a spurious (biased) correlation between cause and outcome
  - Can be identified based on ***backdoor paths***:
    - Paths starting with an arrow into the cause and ending with an arrow into the outcome
  - Backdoor paths can (hopefully) be *blocked* to remove bias
    - Design (e.g., randomly manipulating the cause of interest)
    - Modeling (e.g., covariate adjustment, stratification, …)
- Key requirements:
  - All relevant variables are included in the DAG
  - All relevant arrows are included (in the correct direction)

Simple, fictional, example

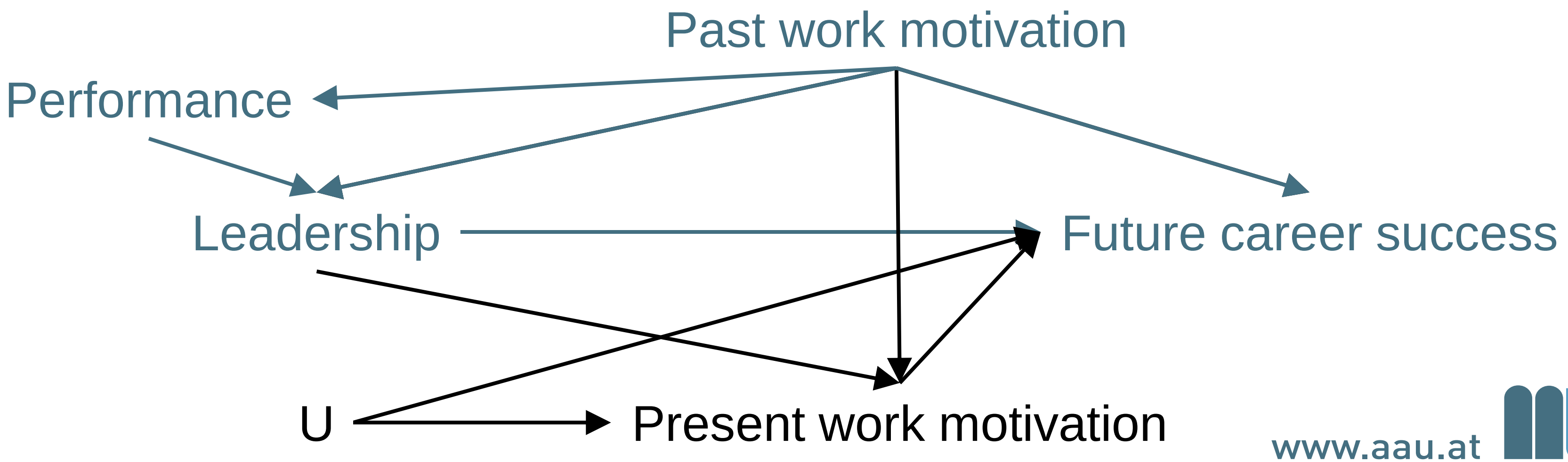
- What is the causal effect of leadership behavior on follower career success?

- Backdoor paths in this DAG
  - L <- W -> C
  - L <- P <- W -> C
- W is a ***confounder***
- How might we block the backdoor paths?
  - Manipulate L *or*
  - Adjust for W



Not so simple, but still fictional, example

- Controlling for past work motivation closes backdoor paths.
- Should we also control for *present* work motivation?
  - PresWM is a ***mediator***: Adjustment *blocks* a causal path L -> PresWM -> FCS
  - PresWM is a ***collider***: Adjustment *opens* a non-causal path L <- U -> FCS



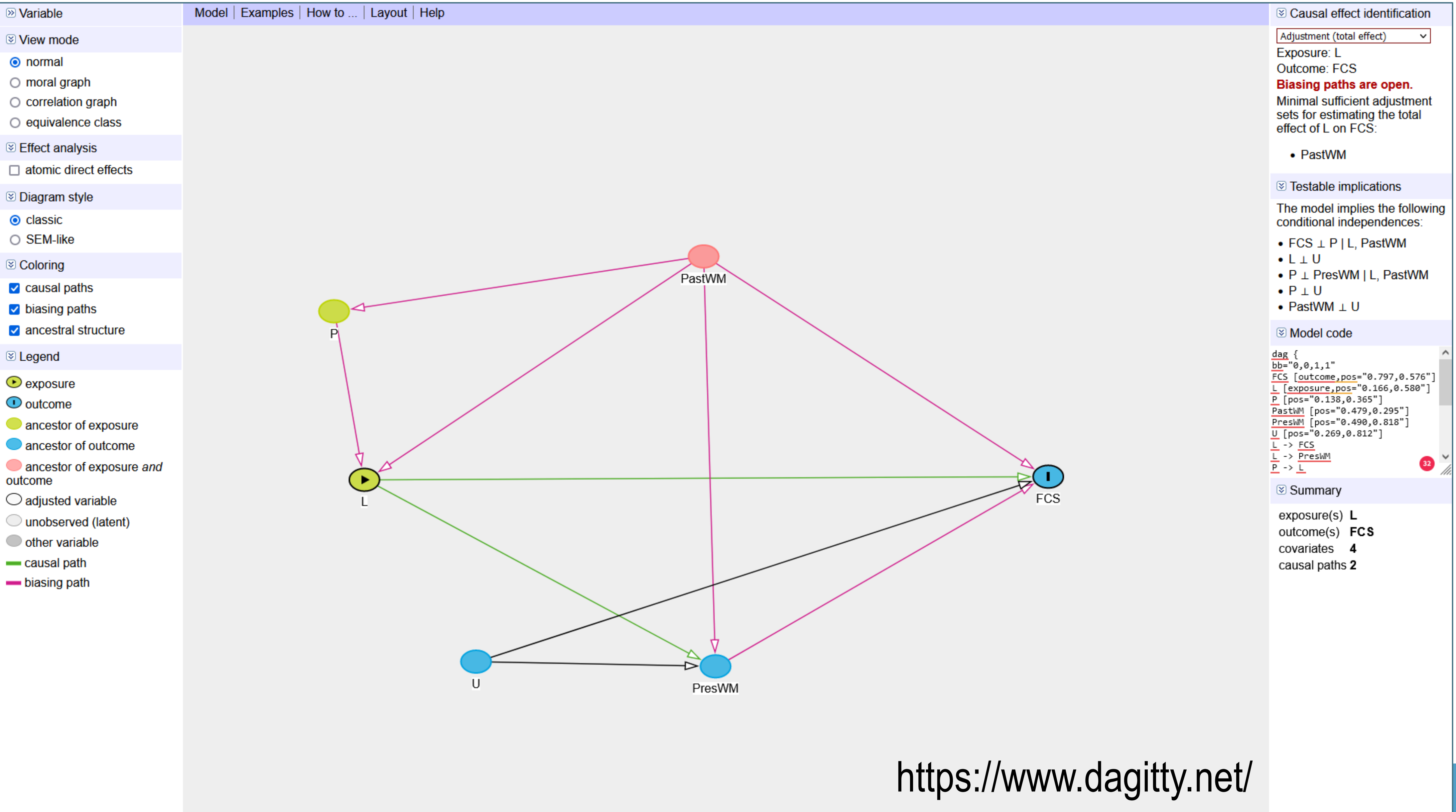
Lessons learned so far

- Directed, ***but*** "bi-directed" arrows = unmeasured confounders: PresWM <- U -> FCS
- Acyclical, ***but*** mutual causation *over time*: PastWM -> L -> PresWM

- Spurious effects (biased estimates) can result from
  - Leaving backdoor (non-causal) paths open: *Confounding*
  - Blocking causal paths containing mediators: *Overcontrol bias*
  - Blocking paths containing colliders: *Collider bias* (e.g., selection bias)

- DAGs for real research questions…
  - Tend to be much more complex
  - May not have a (single) solution!

Software: DAGitty



https://www.dagitty.net/

Take-aways

UNIVERSITÄT KLAGENFURT

- Strengths: DAGs are helpful in…
  - … identifying potential sources of bias
  - … anticipating unwanted side effects of study design and statistical modeling
  - … preventing false conclusions
- “Limitation”: One must assume that the DAG is correct

- When in doubt, draw a DAG!
  - To better understand a study you read
  - To summarize a field of research
  - To prepare your own empirical work

Further reading

- Rohrer, J. M. (2018). Thinking clearly about correlations and causation: Graphical causal models for observational data. *Advances in Methods and Practices in Psychological Science, 1*(1), 27-42.
- Elwert F. (2013). Graphical causal models. In Morgan S. L. (Ed.), *Handbook of causal analysis for social research* (pp. 245–273). Dordrecht, The Netherlands: Springer.
- Pearl J., Glymour M., & Jewell N. P. (2016). *Causal inference in statistics: A primer.* Chichester, England: John Wiley & Sons.
- Pearl, J. (2009). *Causality: Models, reasoning and inference.* Cambridge University Press.
- Morgan, S. L. & Winship, C. (2015). *Counterfactuals and causal inference.* Cambridge University Press.

- Pearl, J. & Mackenzie, D. (2018). *The book of why: The new science of cause and effect.* New York: Basic Books.

Thank you for your attention!

heiko.breitsohl@aau.at