Pearls of Wisdom in Causal Analysis: The Directed Acyclic Graph

Roderick A. Rose, PhD
November 24, 2014
Tate Lecture Series

The Challenge of Causal Inference from Statistics

- Statistics: strength and sign of relationships.
- Not whether a relationship is causal, or direction of causality.
- Need: assumptions, with information that supports (does not falsify) these assumptions.
- Usually use randomization; based on assumptions.
- What if we cannot randomize?

Causal Frameworks

- Campbell (Validity)
- Rubin (also Neyman; Potential outcomes model)
- Pearl (Directed acyclic graphs)
- Other frameworks include the econometric tradition and the structural causal model developed by Pearl.

The Directed Acyclic Graph (DAG)

- A graphical technique for representing our assumptions about the data.
- Use rules specified by Pearl.
- Test features of the DAGs to falsify our assumptions.
- What follows are the basic rules.

Chains and Forks

- A chain is given by

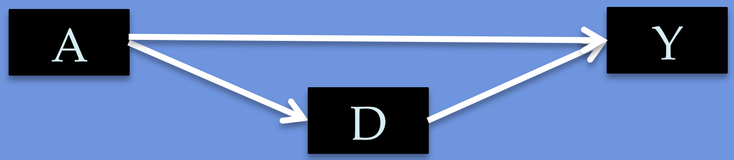
- This is represented in text by A → D → Y.
- D mediates the effect of A on Y.

- A fork, alternatively, is given by

- Which is represented in text by A ← D → Y.
- In this case, D is a common cause of both A and Y.

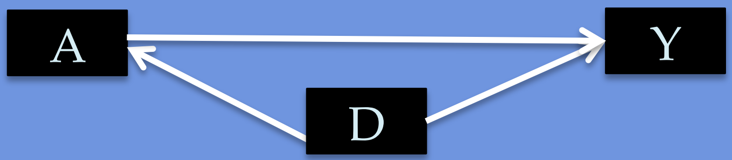
Combining Chains and Forks: 1

- Let's say we want to estimate the effect of A (course attendance) on Y (grades). We also measure D (doing homework).
- The DAG below has a familiar form. What does it look like?
- Note that there is both a chain and a fork.
- Can we estimate a causal effect of A on Y?



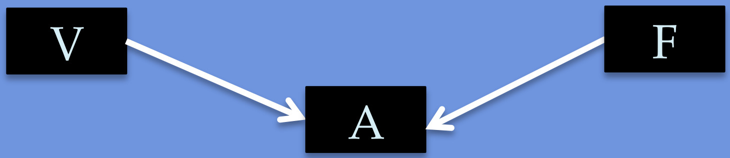
Combining Chains and Forks: 2

- The DAG below connotes the presence of confounding.
- Note that there is both a chain and a fork.
- Can a causal effect of A on Y be estimated here?



Inverted Fork or Collider

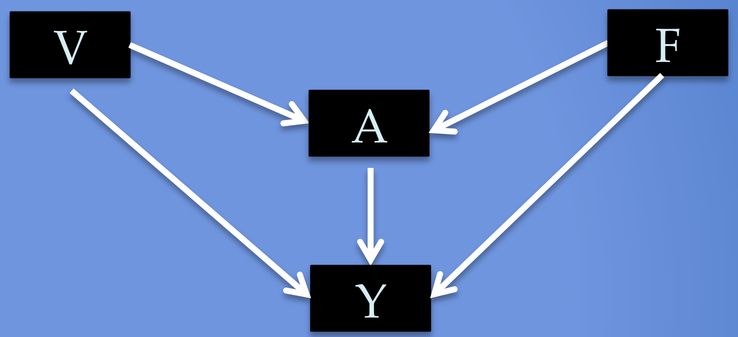
- In an inverted fork V → A ← F, two variables V and F are independent if A is <u>not conditioned on</u>; excluding A blocks the path between V and F.



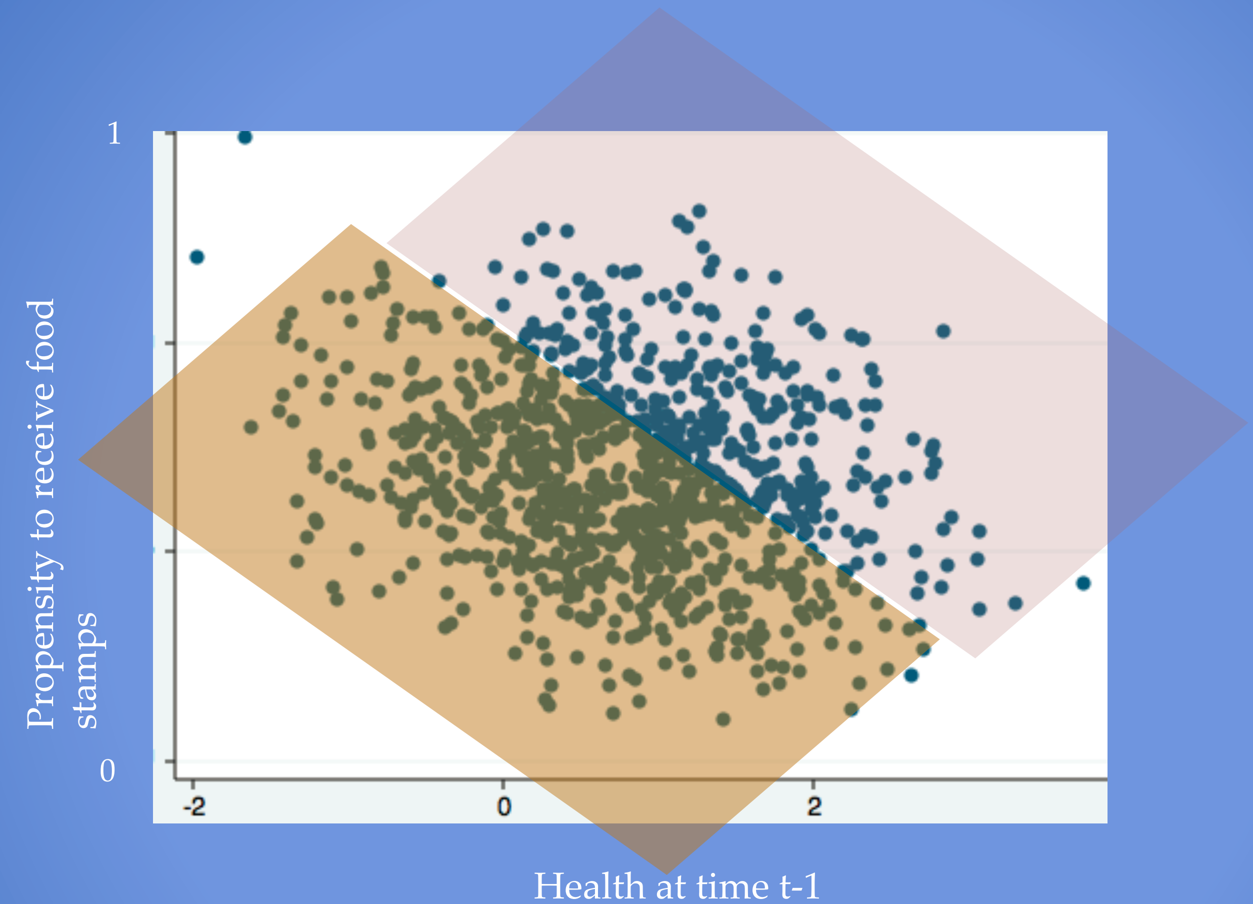
- Colliders present a challenge because inadvertently conditioning on a collider (in an attempt to remove confounding) actually <u>introduces confounding</u>.

The Collider Problem in Practice

- We want to estimate the causal effect of V on Y.
- If we leave out A, we would actually succeed, based on what we know about mediation.
- If we include A, to estimate the indirect causal effect, we render V → A → Y biased by V → A ← F → Y.



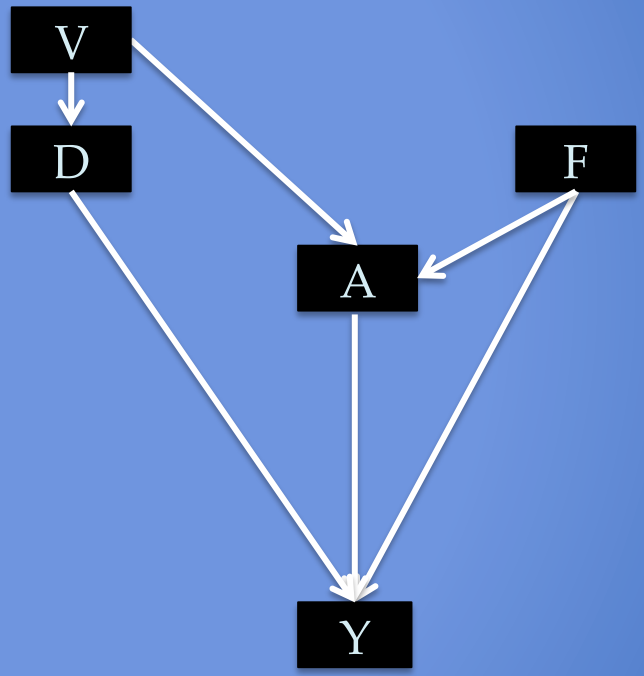
Collider Intuition



Orange: does not have public health insurance
Pink: has public health insurance.

Example 1

- Which variables in the DAG at right should be modeled in order to estimate the causal effect of D on Y?



Examples

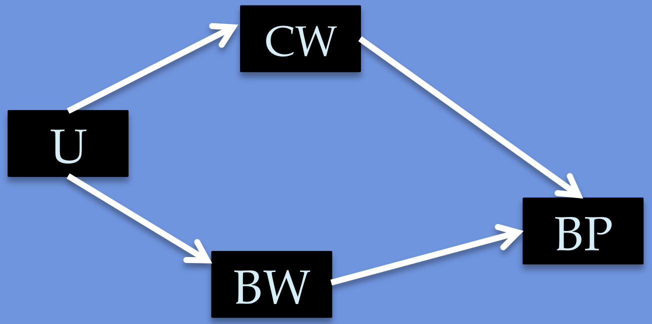
• • •

See how well you understand DAGs.

Most of these examples are from

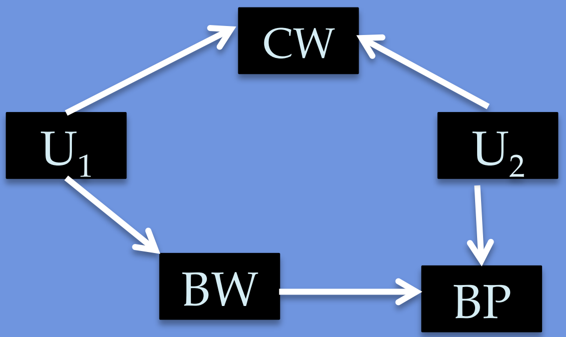
Arah, O. A. (2008). The role of causal reasoning in understanding Simpson's paradox, Lord's paradox, and the suppression effect: covariate selection in the analysis of observational data. *Emerging Themes in Epidemiology 5*(5).1-5.

Arah Figure 3



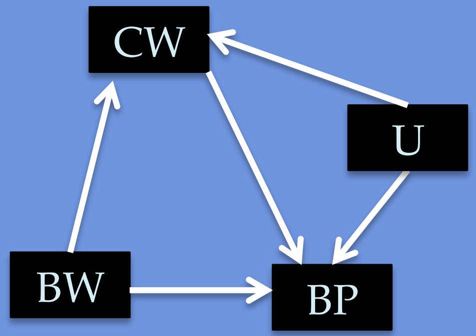
U = confounder (unmeasured)
BW = birth weight
CW = current weight
BP = blood pressure

Arah Figure 4



U = confounder (unmeasured)
BW = birth weight
CW = current weight
BP = blood pressure

Arah Figure 7



U = confounder (unmeasured)
BW = birth weight
CW = current weight
BP = blood pressure

DAGS are:

- Didactic tools
- Frameworks for understanding what data to collect and model
- Frameworks for testing relationships and falsifying claims of causality
- A positivist approach