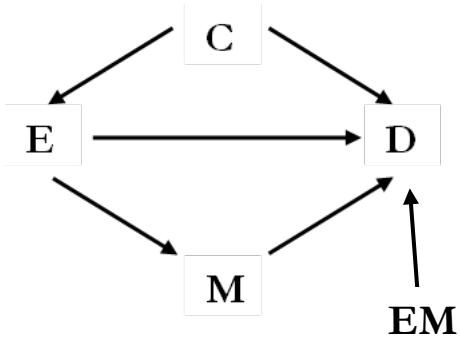
Quick Guide on Directed Acyclic Graph (DAG)

By: Andrew Jergel



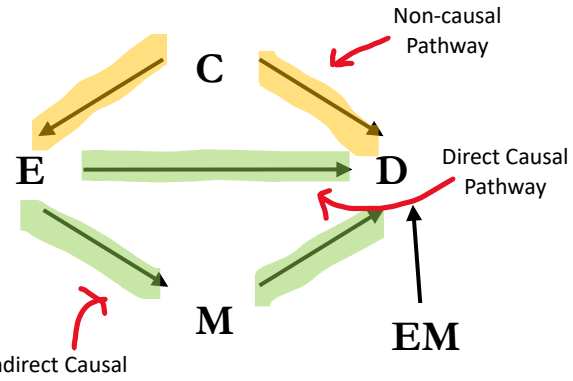
When to use DAG?

1. If you are a visual learner and or have trouble explaining the relationship between variables
2. Junior PIs who are not yet familiar with variable selection or picking the most appropriate variables
3. It is difficult to narrow down your variables

Why use DAG?

1. Express your ideas to others – especially if they are complex and or niche and the listener is not experienced in the topic (i.e., biostatistician, grant reviewer, etc.)
2. Focus research questions, aims, and hypothesis.
3. Visualize the relationship similarly to a mind map or other flow charts.
4. Reduces confusion around ambiguous language or terms
5. Universal language – all anyone on the project would need to know is the variable names themselves and the workings of a DAG.

The Fundamentals of DAG

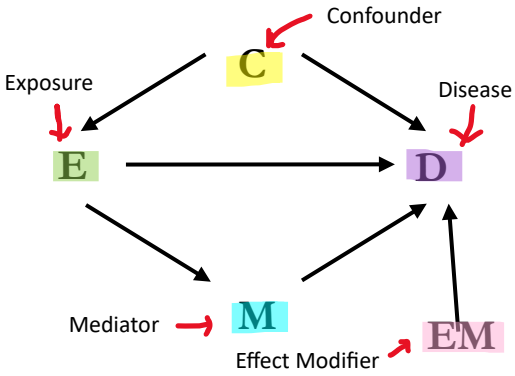


Directed**AG** – each arrow or edge is directed in ONE way

- E↔D ☒
- E→D ☑

DAcyclic**G** – without circles – you cannot go from one variable to the next to the next and return to your starting variable.

- E→D ☑
- E→D→E ☒



Causal Pathway (The relationship of interest from exposure to outcome.)
All arrows point in same direction
E → D
E → M → D

Non-Causal Pathway (A pathway that connects from exposure to outcome but is affected or obstructed/blocked by some variable)
NOT all arrows point in the same direction
E ← C → D

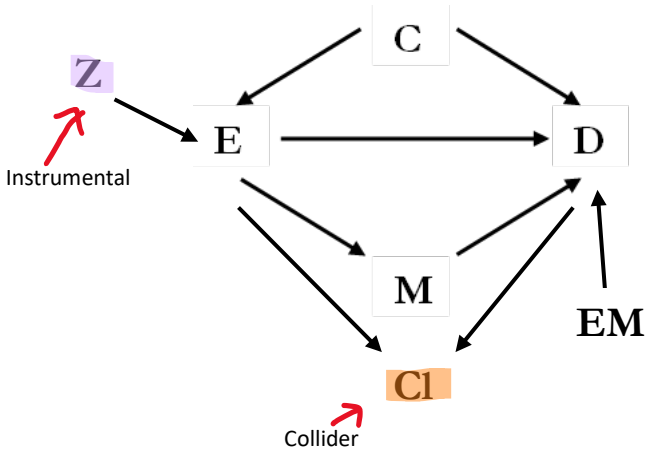
Additional Terms for DAG

Z = Instrumental Variable

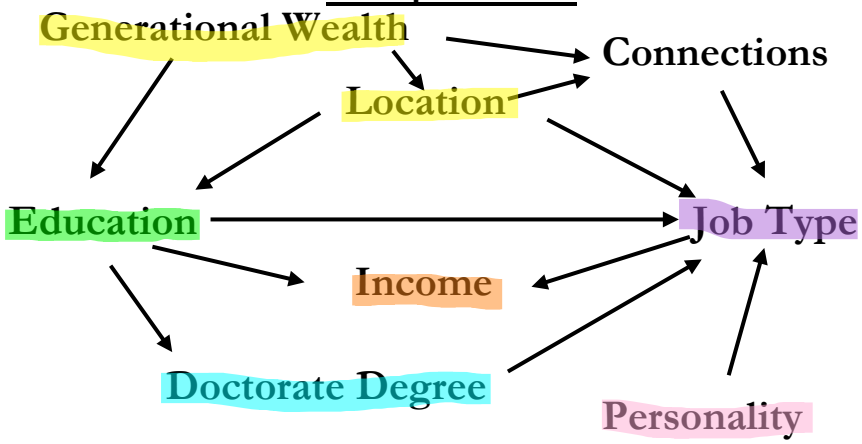
- A variable that influences the exposure but not the disease. An example of this would be randomization of treatment.

Cl = Collider

- A variable that is caused by both the exposure and disease. An example of this would be "depression" in the relationship between income and negative health outcomes.
  - This is a non-causal pathway and is already "blocked"
  - It is important to recognize these as they should not be controlled for. Doing so may bias your results.



Hypothetical Complex Example of DAG



Causal Pathways:
Education → Job Type (direct exposure to outcome)
Education → Doctorate Degree → Job Type (indirect – mediated)

Non-Causal Pathways:
Education ← Loc. → Job Type
Education → Income ← Job Type
Education ← Loc. → Conn. → Job Type
Education ← Gen. Wealth → Loc. → Job Type
Education ← Gen. Wealth → Conn. → Job Type
Education ← Gen. Wealth → Loc. → Conn. → Job Type

We can now look at this and determine how we want to proceed with the analysis (without doing model selection techniques)
Controlling for Generational wealth and location would clear up any confounding. Note Income is a collider and therefore this pathway is automatically blocked. Controlling for it would induce bias.
We could investigate the mediated effect of a doctorate degree, and we could also take into account the interaction effect of personality