

Directed acyclic graphs in epidemiology

JA Labrecque

Erasmus Medical Center, Rotterdam, The Netherlands

23 mei, 2018

Directed acyclic graphs (DAGs)

- ▶ Draw your assumptions before you draw your conclusions

Directed acyclic graphs (DAGs)

- ▶ Draw your assumptions before you draw your conclusions
- ▶ “Directed” because the edges have a direction

Directed acyclic graphs (DAGs)

- ▶ Draw your assumptions before you draw your conclusions
- ▶ “Directed” because the edges have a direction
- ▶ “Acyclic” because there can be no cycles, i.e. a variable cannot cause itself

Directed acyclic graphs (DAGs)

- ▶ Draw your assumptions before you draw your conclusions
- ▶ “Directed” because the edges have a direction
- ▶ “Acyclic” because there can be no cycles, i.e. a variable cannot cause itself
- ▶ A causal DAG requires that:

Directed acyclic graphs (DAGs)

- ▶ Draw your assumptions before you draw your conclusions
- ▶ “Directed” because the edges have a direction
- ▶ “Acyclic” because there can be no cycles, i.e. a variable cannot cause itself
- ▶ A causal DAG requires that:
 - ▶ a lack of an arrow between two nodes implies the absence of a causal effect

Directed acyclic graphs (DAGs)

- ▶ Draw your assumptions before you draw your conclusions
- ▶ “Directed” because the edges have a direction
- ▶ “Acyclic” because there can be no cycles, i.e. a variable cannot cause itself
- ▶ A causal DAG requires that:
 - ▶ a lack of an arrow between two nodes implies the absence of a causal effect
 - ▶ all common causes of variables on the graph are themselves on the graph

Directed acyclic graphs (DAGs)

- ▶ Draw your assumptions before you draw your conclusions
- ▶ “Directed” because the edges have a direction
- ▶ “Acyclic” because there can be no cycles, i.e. a variable cannot cause itself
- ▶ A causal DAG requires that:
 - ▶ a lack of an arrow between two nodes implies the absence of a causal effect
 - ▶ all common causes of variables on the graph are themselves on the graph
 - ▶ any variable is a cause of its descendents

Directed acyclic graphs (DAGs)

- ▶ Draw your assumptions before you draw your conclusions
- ▶ “Directed” because the edges have a direction
- ▶ “Acyclic” because there can be no cycles, i.e. a variable cannot cause itself
- ▶ A causal DAG requires that:
 - ▶ a lack of an arrow between two nodes implies the absence of a causal effect
 - ▶ all common causes of variables on the graph are themselves on the graph
 - ▶ any variable is a cause of its descendents
- ▶ To map a causal DAG to observed data we require the causal Markov assumption that a node is independent of all nodes which it does not cause conditional on its parents

Directed acyclic graphs (DAGs)

- ▶ Draw your assumptions before you draw your conclusions
- ▶ “Directed” because the edges have a direction
- ▶ “Acyclic” because there can be no cycles, i.e. a variable cannot cause itself
- ▶ A causal DAG requires that:
 - ▶ a lack of an arrow between two nodes implies the absence of a causal effect
 - ▶ all common causes of variables on the graph are themselves on the graph
 - ▶ any variable is a cause of its descendents
- ▶ To map a causal DAG to observed data we require the causal Markov assumption that a node is independent of all nodes which it does not cause conditional on its parents
- ▶ We can then use DAGs to explore the causal relationship between variables using our observational data

Directed acyclic graphs (DAGs)



george davey smith

@mendel_random

Following



Epidemiologists rejoice! "the task of selecting an appropriate set of covariates to control for confounding has been reduced to a simple "roadblocks" puzzle manageable by a simple algorithm" arxiv.org/abs/1801.04016 and



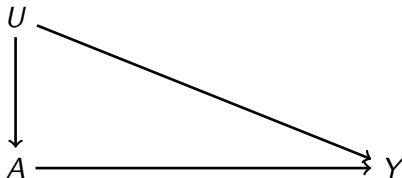
Comments on: The tale wagged by the DAG | International ...

I am grateful to the editors for the opportunity to comment on Nancy Krieger and George Davey Smith's article, 'The tale wagged by the DAG', which appeared in academic.oup.com

This is true! But the DAG has to be correctly specified which is the hard (impossible) part.

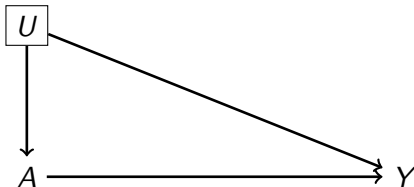
Rules of DAGs

- Show conditioning on a node with a box



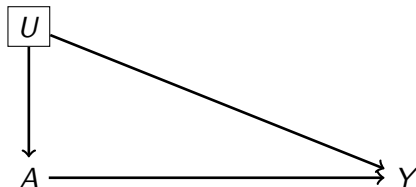
Rules of DAGs

- Show conditioning on a node with a box



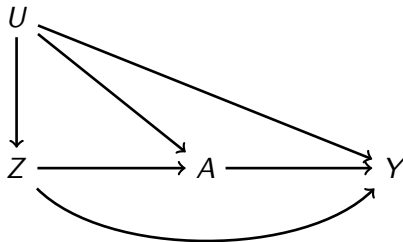
Rules of DAGs

- ▶ Show conditioning on a node with a box
- ▶ The effect of A on Y is confounded because $E[Y|A=1] - E[Y|A=0] \neq E[Y^{a=1}] - E[Y^{a=0}]$



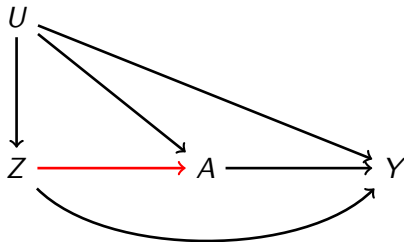
Showing the IV assumptions in a DAG

- This DAG makes the least amount of assumptions about the observed data. Every node is connected to every other



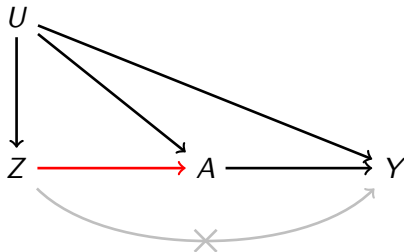
Showing the IV assumptions in a DAG

- The relevance assumption is represented by the red arrow, showing that there is a relationship between Z and A



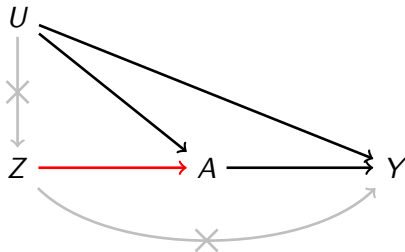
Showing the IV assumptions in a DAG

- ▶ The exclusion restriction (ER) assumption assumes that the direct edge between Z and A is not present



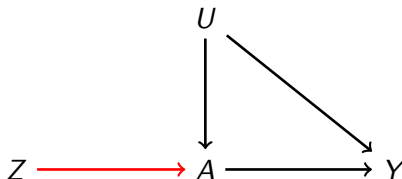
Showing the IV assumptions in a DAG

- The exchangeability assumption assumes that Z is not confounded with Y



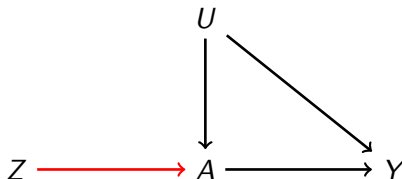
Showing the IV assumptions in a DAG

- ▶ With our three assumptions, we end up with our classic IV DAG



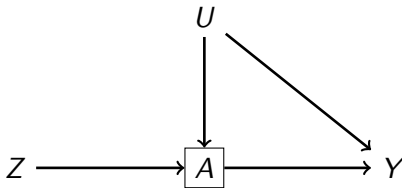
Showing the IV assumptions in a DAG

- ▶ With our three assumptions, we end up with our classic IV DAG
- ▶ Note: monotonicity and homogeneity cannot be drawn on a DAG because they are parametric concepts



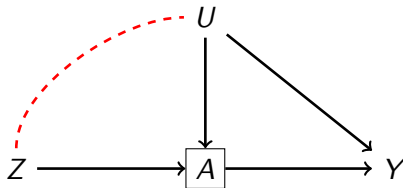
Conditioning on exposure in IV analyses

- Sometimes people like to stratify or restrict on exposure



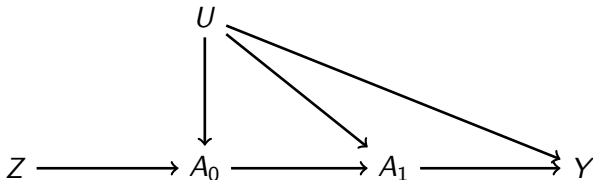
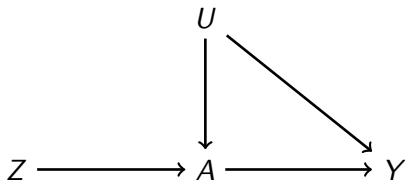
Conditioning on exposure in IV analyses

- ▶ But we create a new association between Z and U therefore violating the exchangeability assumption



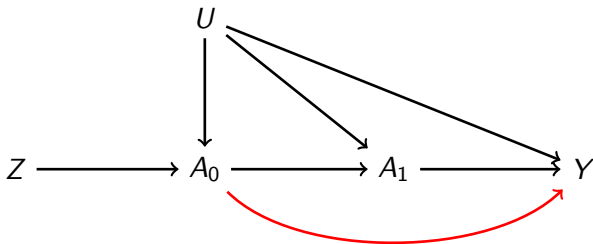
Using DAGs to think about IV longitudinally

- We can depict a time-varying exposure by splitting the A node into separate nodes for each time point



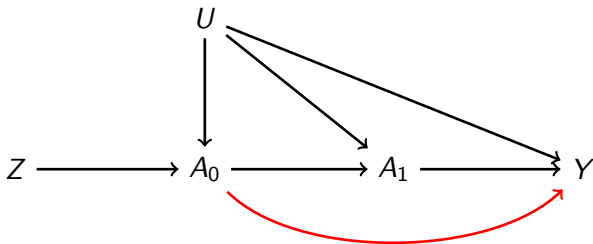
Using DAGs to think about IV longitudinally

- We have assumed that A_0 does not have a direct effect on Y



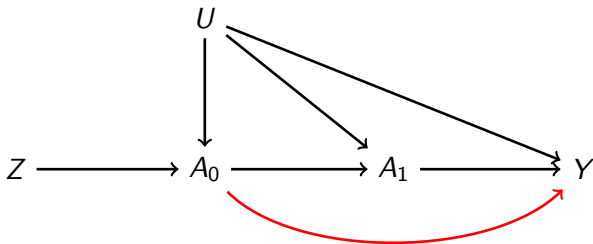
Using DAGs to think about IV longitudinally

- ▶ We have assumed that A_0 does not have a direct effect on Y
- ▶ If A_0 does affect Y then we can no longer estimate the effect of A_1 on Y



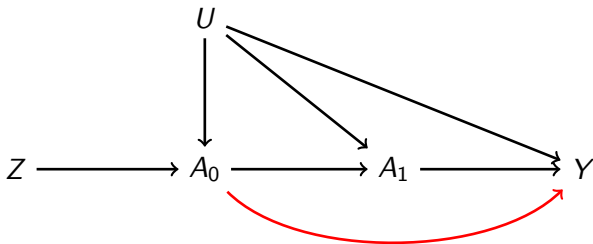
Using DAGs to think about IV longitudinally

- ▶ We have assumed that A_0 does not have a direct effect on Y
- ▶ If A_0 does affect Y then we can no longer estimate the effect of A_1 on Y
- ▶ This is because, although the ER holds for A as a whole, it does not hold for A_1



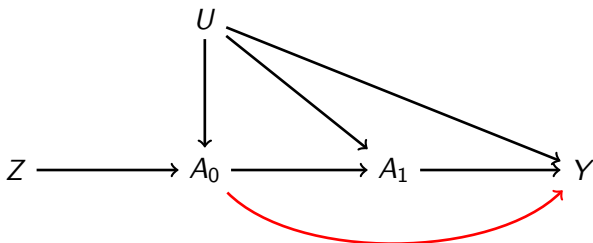
Using DAGs to think about IV longitudinally

- ▶ We have assumed that A_0 does not have a direct effect on Y
- ▶ If A_0 does affect Y then we can no longer estimate the effect of A_1 on Y
- ▶ This is because, although the ER holds for A as a whole, it does not hold for A_1
- ▶ So if we measure A_0 we're alright?



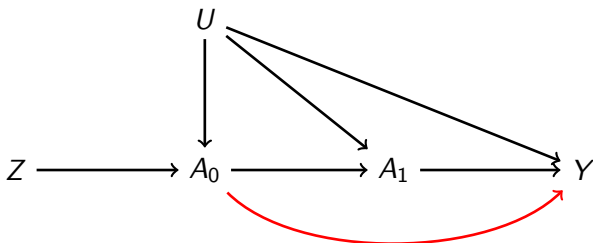
Using DAGs to think about IV longitudinally

- We have also assumed that Z does not have a direct effect on A_1



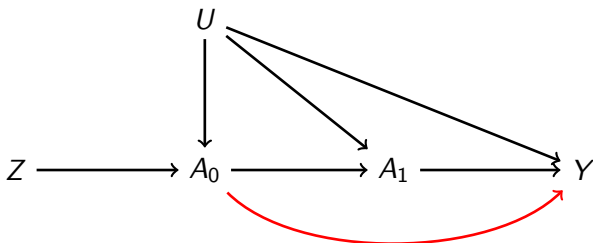
Using DAGs to think about IV longitudinally

- ▶ We have also assumed that Z does not have a direct effect on A_1
- ▶ If Z does affect A_1 then we can no longer estimate the effect of A_0 on Y



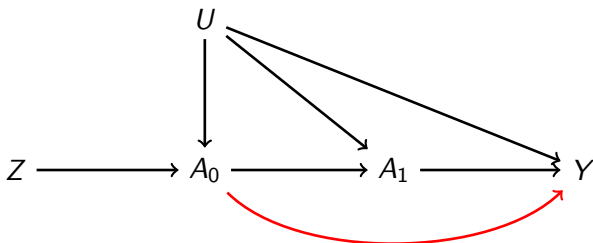
Using DAGs to think about IV longitudinally

- ▶ We have also assumed that Z does not have a direct effect on A_1
- ▶ If Z does affect A_1 then we can no longer estimate the effect of A_0 on Y
- ▶ This is because, although the ER still holds for A as a whole, it does not hold for A_0



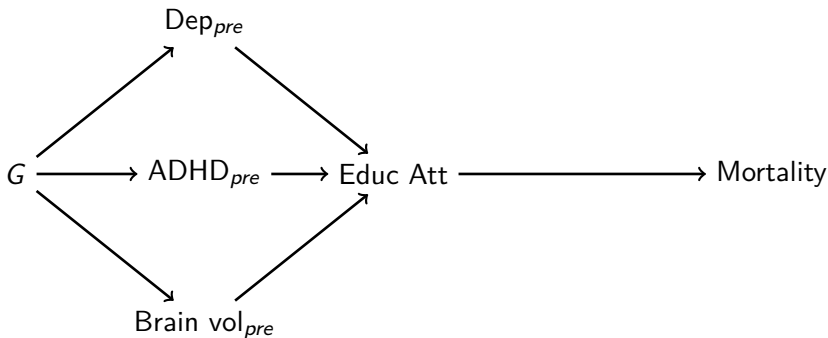
Using DAGs to think about IV longitudinally

- ▶ We have also assumed that Z does not have a direct effect on A_1
- ▶ If Z does affect A_1 then we can no longer estimate the effect of A_0 on Y
- ▶ This is because, although the ER still holds for A as a whole, it does not hold for A_0
- ▶ All this to illustrate that, in contexts such as Mendelian randomization, important thought must be given to how exposures change longitudinally



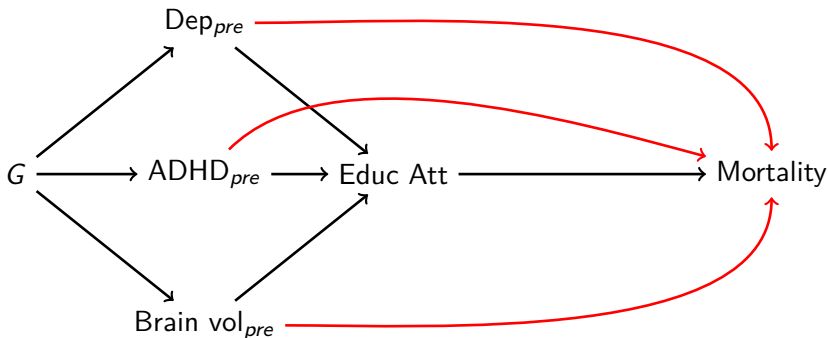
Using DAGs to think about Mendelian randomization

- DAG for a MR study of the effect of educational attainment on mortality



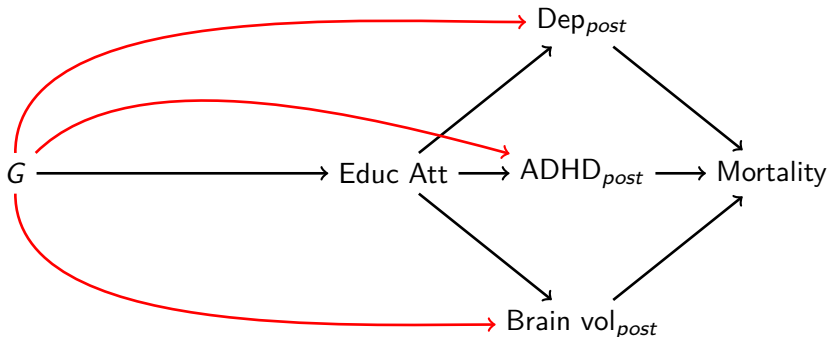
Using DAGs to think about Mendelian randomization

- ▶ But what if the intermediates between G and education attainment have a direct effect on mortality?



Using DAGs to think about Mendelian randomization

- Or what if G has a direct affect on the intermediates after educational attainment



Using DAGs to think about Mendelian randomization

- ▶ We have also assumed that Z does not have a direct effect on A_1

