Practical Causal Inference

AI4ER MRes Workshop

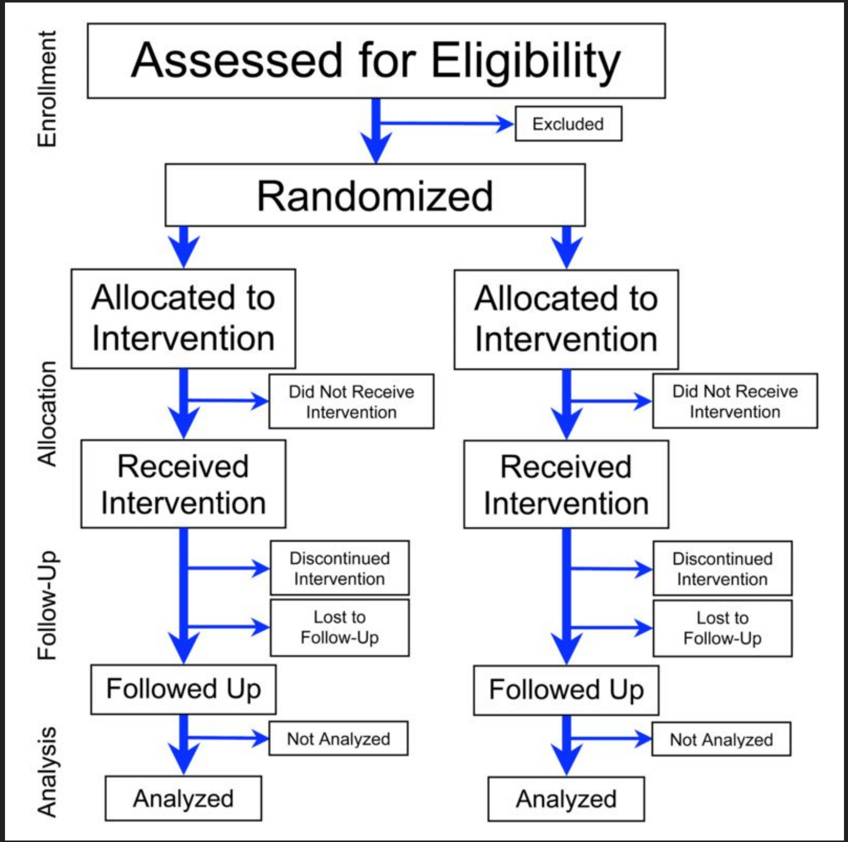
So…how do we actually determine causal effects?

- In a perfect world, how do you think we should try to determine a causal effect?
- Ideas?

The Gold Standard

Randomised Controlled Trials

Blind, randomised, large numbers etc.



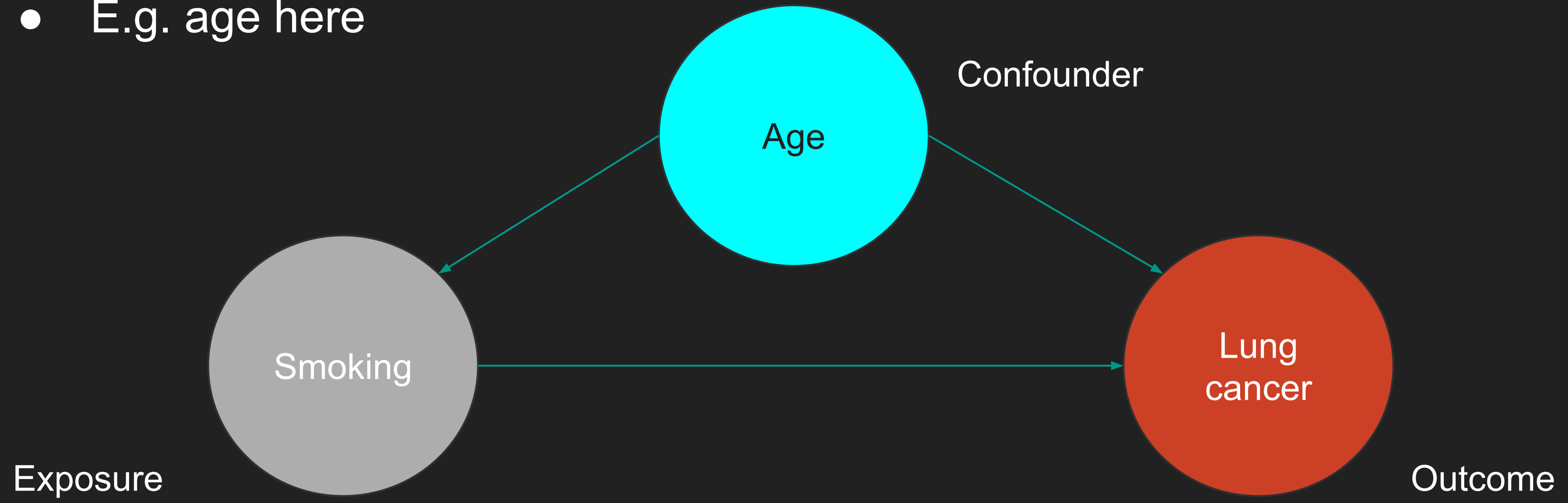
Why are RCTs the gold standard?

This is a key question!

What is about RCTs that make us sure we are isolating a causal effect?

In order to identify a causal effect…

- We need to remove or minimise the effect of **confounders**
- **What is a confounder?**
- A confounder is something that may affect both our **exposure** and **outcome**
- E.g. age here



But we can’t always do RCTs…

- When might we be **unable** to do a randomised controlled trial?

Potential reasons not to do an RCT?

Unethical

Impossible

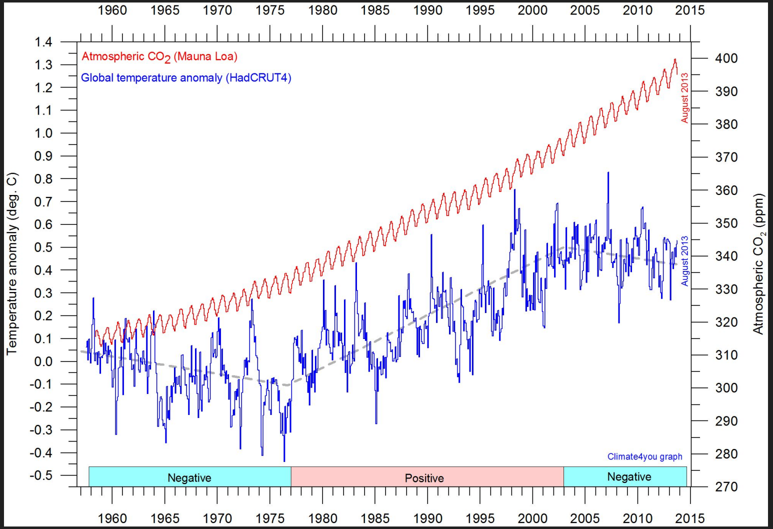
Expensive

All of the above are usually true for the climate!

So what are we left with?

We are left with purely observational data.

This is all we have in many examples, particularly in those relating to climate - and so the question is, can we use observational data to draw causal conclusions?!



What do we need from our observational data?

Essentially, in causal inference methods, we are trying to **simulate** an RCT as best we can, using our observational data.

Can we think of strategies that might help us to do this?

- ?
- ?
- ?

We will return to this later!

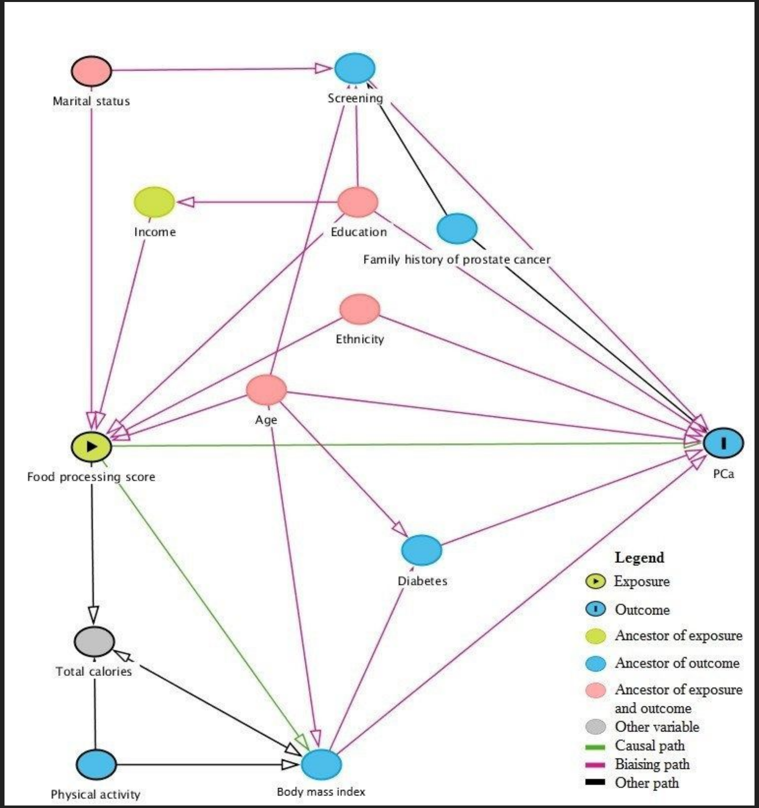
Pause…

Introducing Directed Acyclic Graphs

DAGs are a key tool for helping us to do causal inference. They help us to visualise causal relationships.

Firstly, some graph theory basics!

A DAG helps us to visualise causal relationships



A DAG helps us to work out what we need to control!

There is a mathematical framework for working out what to control for once we have a DAG. This framework was clearly described by Judea Pearl in the 1990s, and it is one his seminal contributions.

But the question is how do we come up with this DAG in the first place?

- Expert knowledge
- Causal discovery - discovering the DAG from data!

We will return to causal discovery later.

Expert knowledge to build a DAG

Expert knowledge allows us to build our DAG.

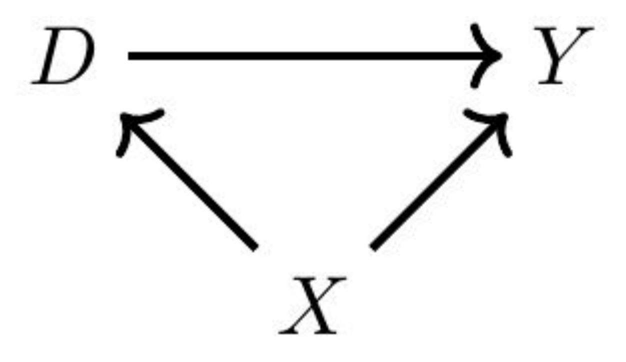
However it requires a number of assumptions. First, we have to assume that our causal structure is correct! Secondly, we have to assume that we are not missing any (latent, unobserved) variables!

These are **strong assumptions.** Furthermore we **cannot** test if these assumptions are valid.

They can be relaxed in certain circumstances.

Two key concepts: **backdoor paths** and *collider bias*

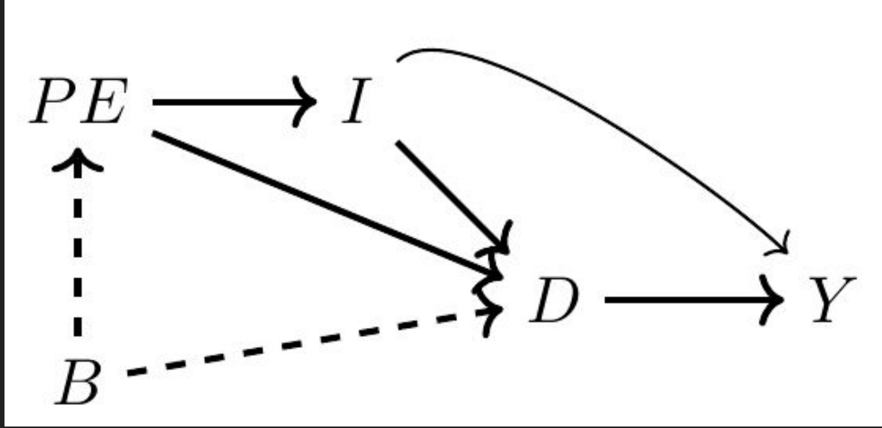
Firstly, backdoor paths.



In this DAG, we have three random variables: X, D, and Y. There is a direct path from D to Y, which represents a causal effect. That path is represented by D→Y. But there is also a second path from D to Y called **the backdoor path**. **The backdoor path** is D←X→Y. While the direct path is a causal effect, **the backdoor path** is not causal. Rather, it is a process that creates spurious correlations between D and Y that are driven solely by fluctuations in the X random variable. X is a confounder!

Let’s look at the causal effect of college education on future earnings…

DAGs show us up!

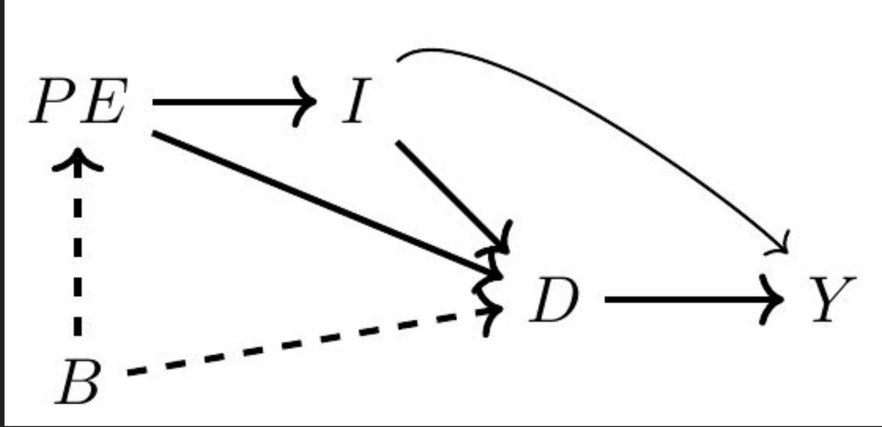


- D: college education
- Y: earnings
- PE: parental education
- I: family income
- B: unobserved factors, such as genetics, family environment, desire to earn…

This DAG makes very clear our assumptions.

What might possibly be wrong in this DAG? What assumptions do we make?

Backdoor paths



D→Y (the causal effect of education on earnings)

D←I→Y (backdoor path 1)

D←PE→I→Y (backdoor path 2)

D←B→PE→I→Y (backdoor path 3)

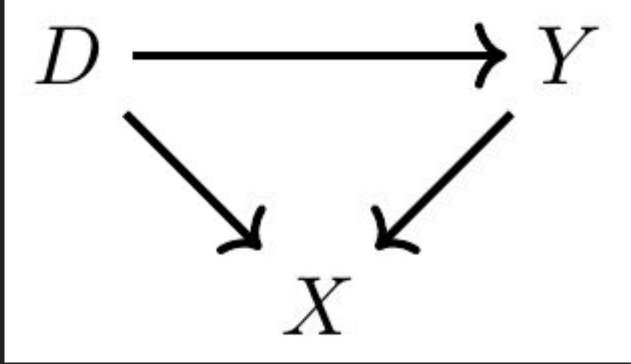
If we **close** these paths, then we can isolate the direct causal effect. We can close the path by **conditioning** on variables along the backdoor paths

Assuming our graph is a valid representation of the world…

Is it always this ‘simple’?

- Sadly not…due to our second key concept, ***colliders***

What is a collider?



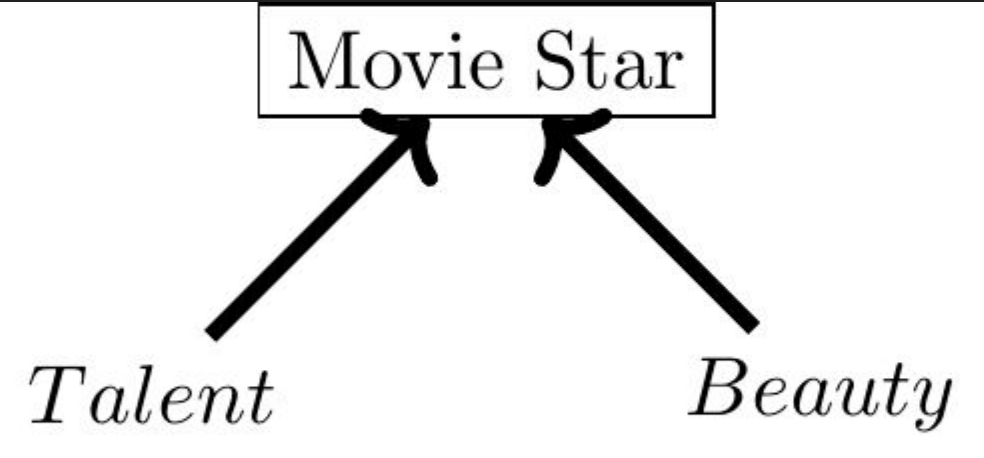
Arrows to/from X are going the other way!

Colliders are special in part because when they appear along a backdoor path, that backdoor path is closed simply because of their presence. Colliders, when they are left alone, always close a specific backdoor path, i.e. we don't need to control for them to isolate the causal effect of D on Y.

Before we look at an example case…

- Let’s look at some of the errors that can arise from colliders.

Sample selection and collider bias



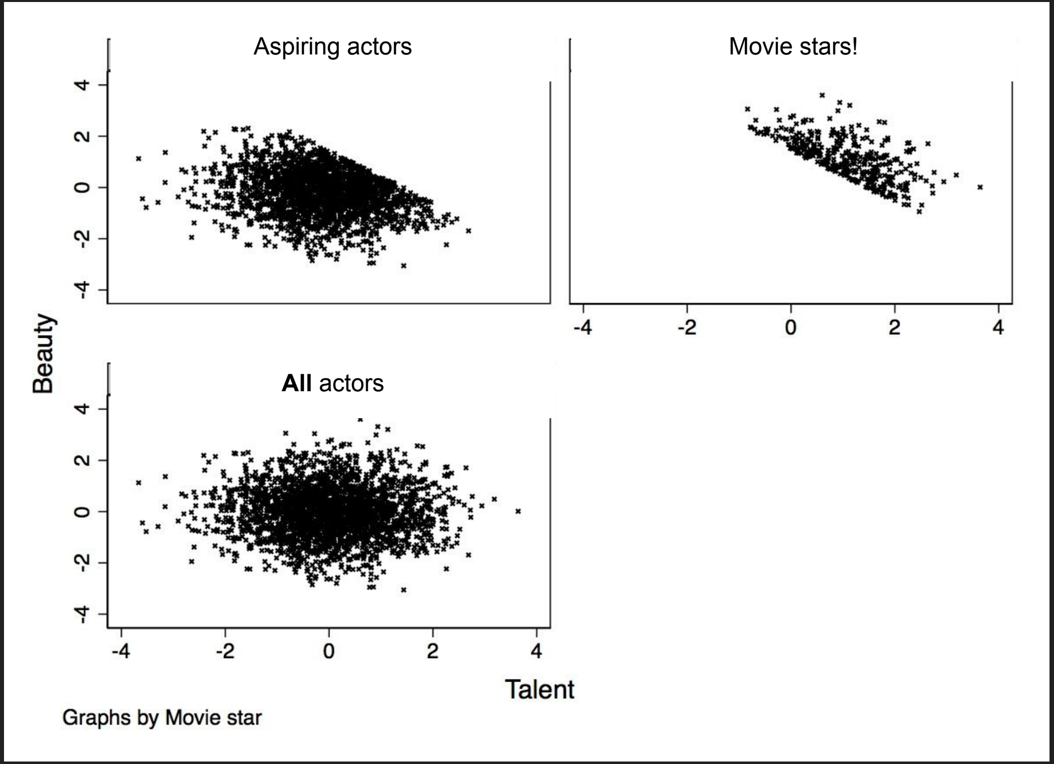
What will happen when we **condition** on ‘movie star’?

- Condition just means **isolate** the subset of people who are big movie stars

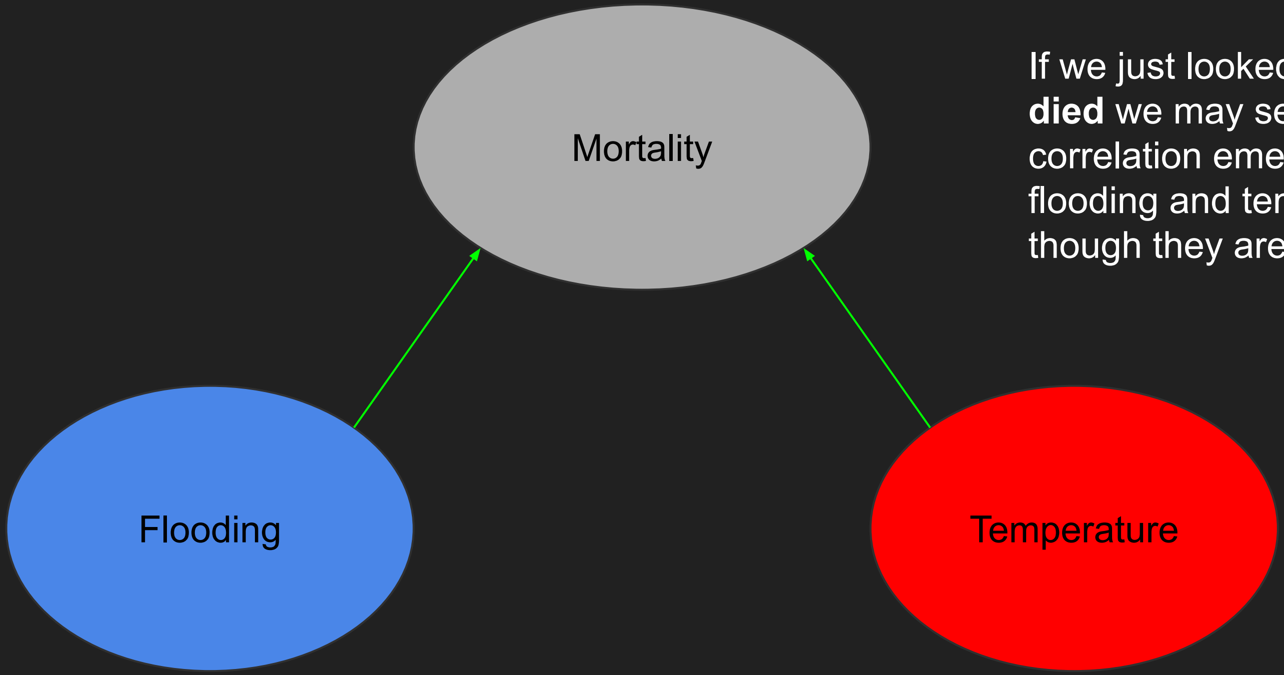
What will happen?

- We might **induce a correlation** between beauty and talent that doesn’t really exist in the wider population
- This is a difficult leap - let’s illustrate it.

We see a negative correlation between beauty and talent amongst the most beautiful and talented (movie stars)!



We would see a similar thing in this environmental case…



If we just looked **at those who died** we may see a negative correlation emerge between flooding and temperature, even though they are unrelated!

Closing paths?

To identify an individual causal effect from our graph, we need to **close** the paths.

How do we actually do this?

1. If we have an open backdoor path, we can close it by **conditioning** on that variable, or **controlling** for that variable. This means essentially holding that variable fixed by methods including **subclassification, matching, or regression.**
2. If we have a collider in our backdoor path, that path is already closed! Hooray!

Returning to our previous example, how do we isolate the direct causal effect?

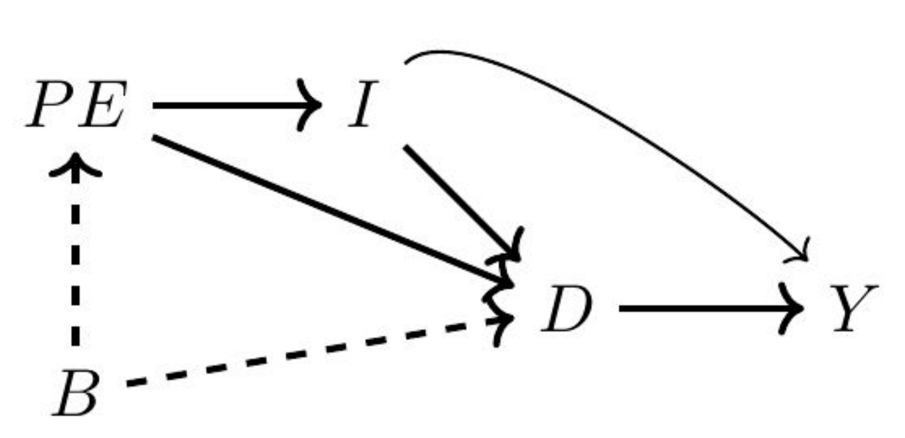
Backdoor paths:

D→Y (the causal effect of education on earnings)

D←I→Y (backdoor path 1)

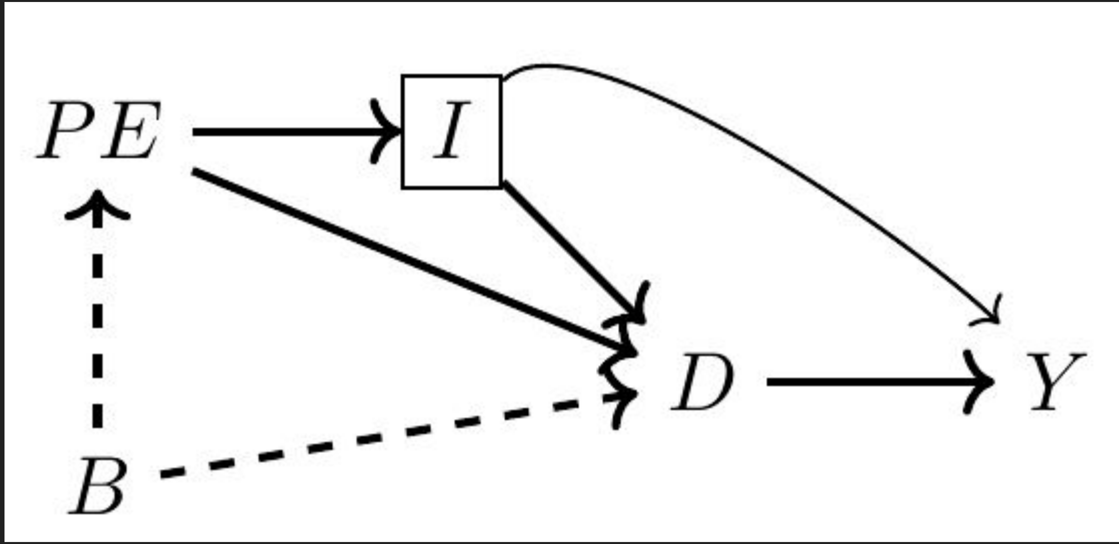
D←PE→I→Y (backdoor path 2)

D←B→PE→I→Y (backdoor path 3)



Condition on I!

- This allows us to isolate our causal effect of D (education) on Y (earnings!).



So how do we actually do conditioning?

1. **Subclassification:** this is splitting our dataset into strata in order to remove the effect of e.g. age
2. **Matching:** finding an identical example which did receive the intervention in order to determine the causal effect
3. **Approximate matching:** finding a very similar example to determine the causal effect
4. **Difference-in-differences:** a ‘natural experiment’ where we find two comparable groups, one exposed to the exposure and one not, to determine the causal effect

We will explore these in code examples!

Questions?