# THE ROLE OF CAUSAL INFERENCE IN MACHINE LEARNING

Ricardo Silva
**ricardo@stats.ucl.ac.uk**

Department of Statistical Science
Centre for Computational Statistics and Machine Learning

University College London

# Causality

- Knowing cause-effect relationships is useful. Machine learning should have something (or lots of) to say about it.

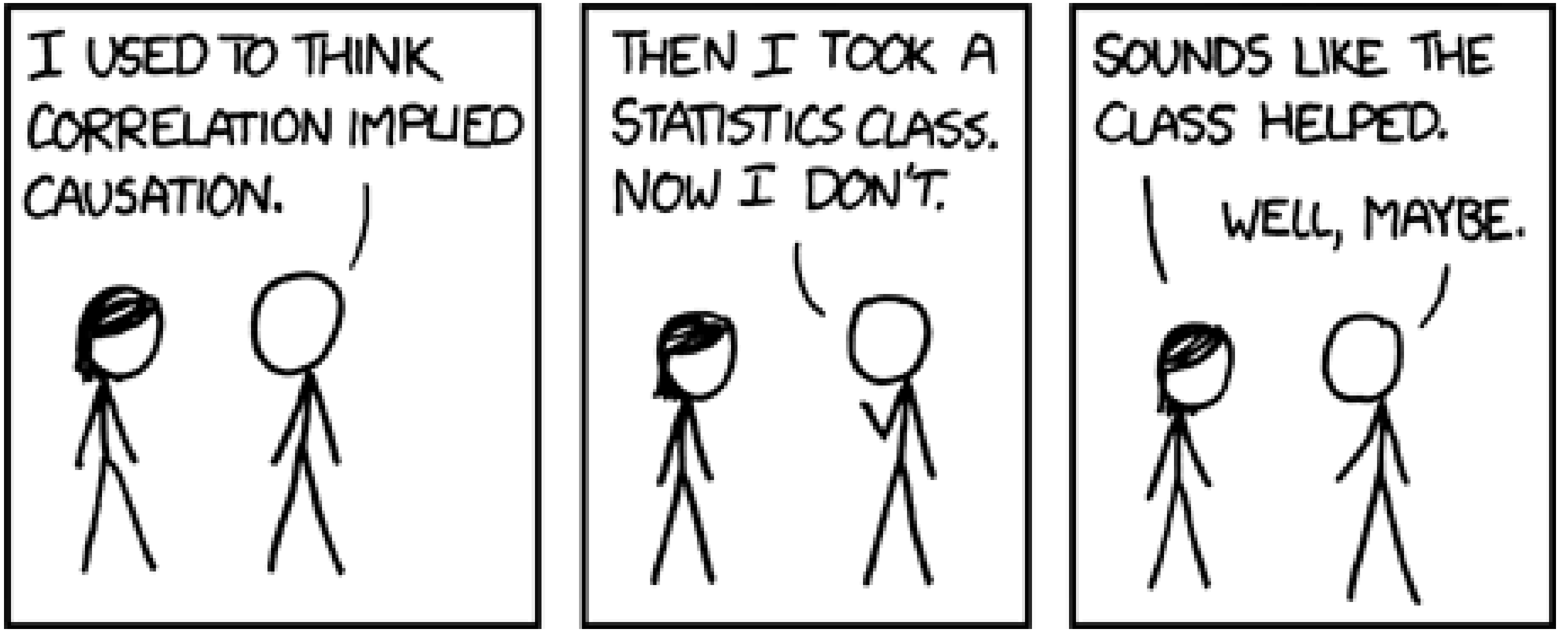
- What is the “horse” factor? See also, **external validity**.

- I will discuss background on causal inference, some machine learning aspects of it, some validity aspects. Depending on time, maybe even details of a particular algorithm (don’t hold your breath).

# On Causation and Prediction

- There are tasks of **prediction**, and tasks of **control**.
- Prediction is bog-standard in machine learning, statistics, predictive analytics etc.
- Control is about **taking actions** to achieve a particular outcome.

# Introducing: Observational Studies


I USED TO THINK CORRELATION IMPLIED CAUSATION.
THEN I TOOK A STATISTICS CLASS. NOW I DON'T.
SOUNDS LIKE THE CLASS HELPED.
WELL, MAYBE.

Compulsory XKCD strip

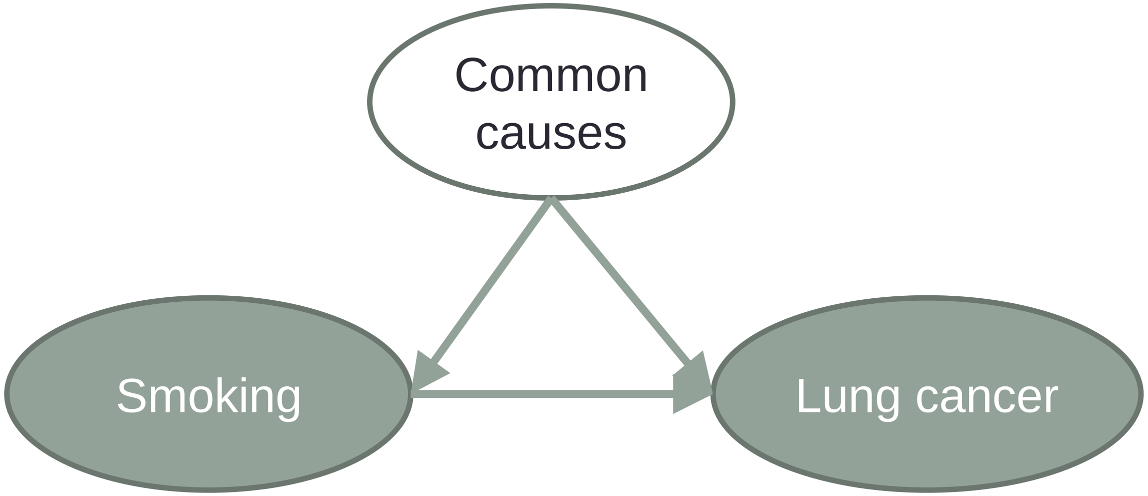
# Out of Control

- In an observational study, the quantity we deem as the “treatment” is not under any designer’s control.

- Case in point, **smoking** as treatment, **lung cancer** as outcome.

- How would one apply the framework of experimental design to the smoking and lung cancer problem?

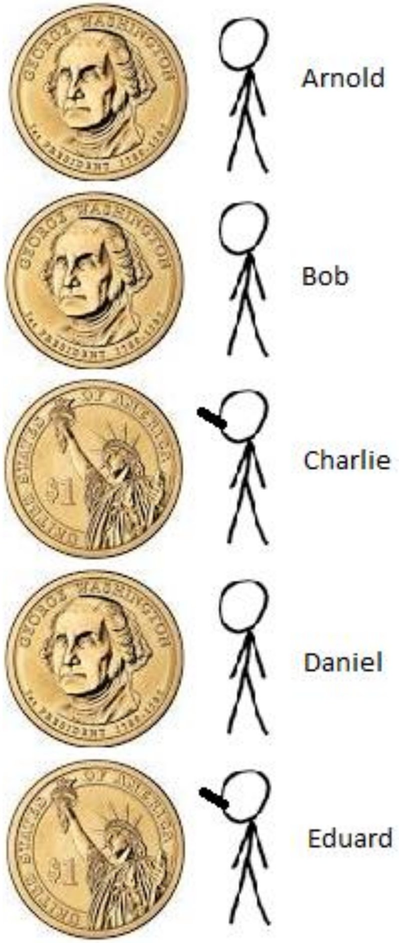
# Where Do Treatments Come From?

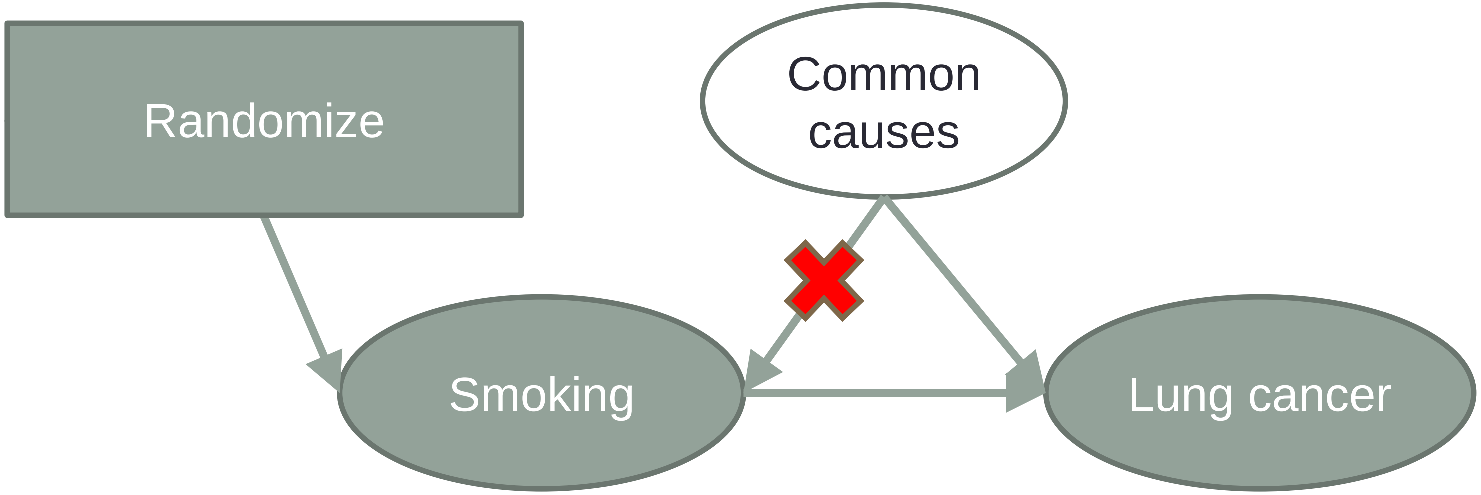

Common causes

Smoking

Lung cancer

# Running a Controlled Trial





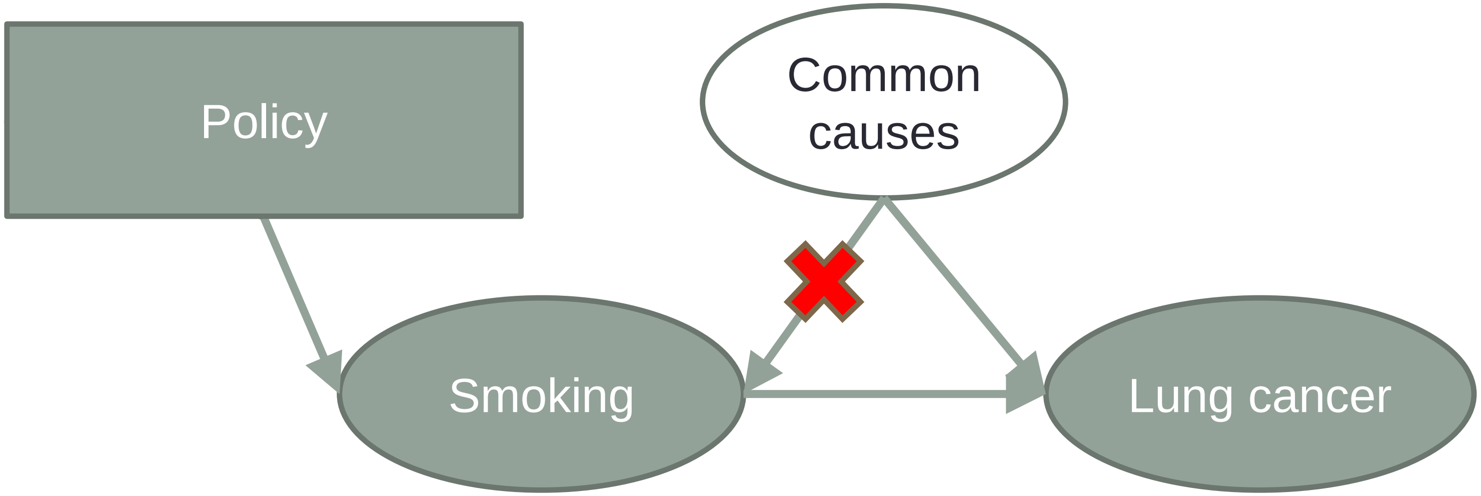
Randomize

Common causes

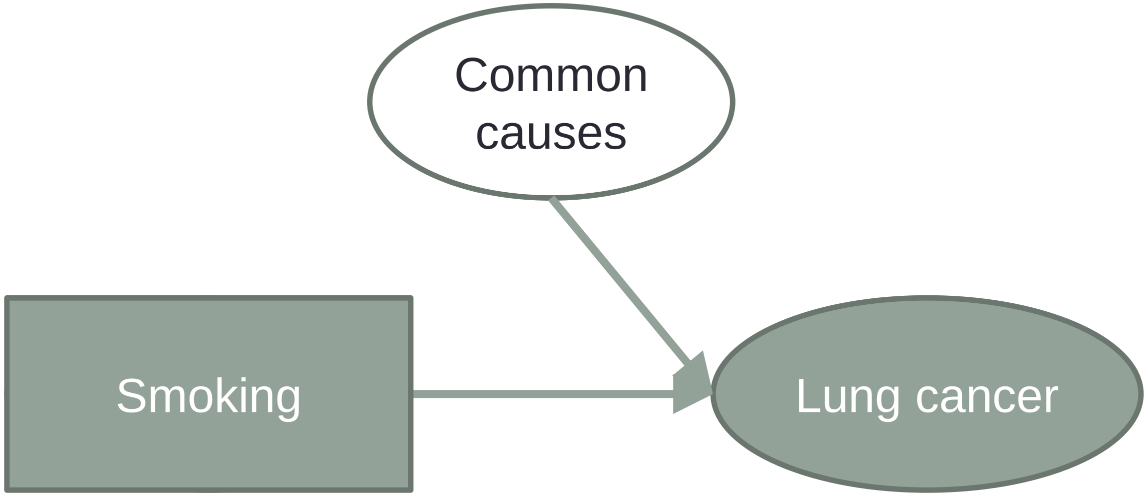
Smoking

Lung cancer

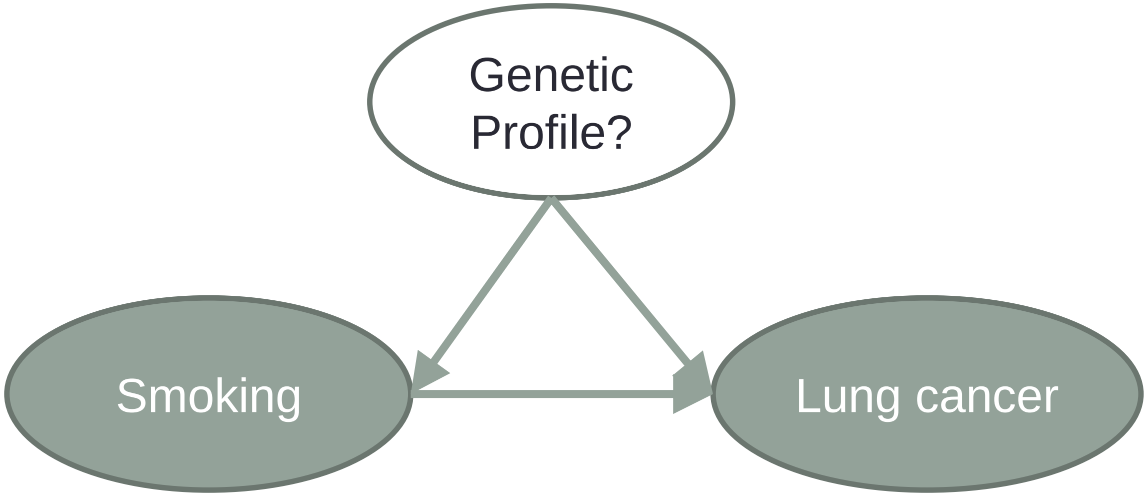
# Exploiting the Knowledge Learned from a Controlled Trial



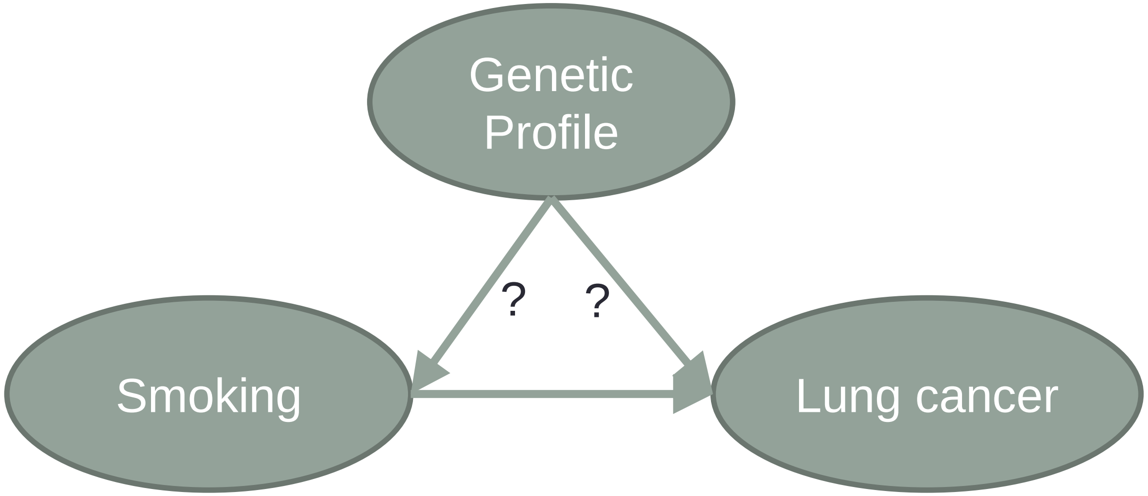
# Exploiting the Knowledge Learned from a Controlled Trial



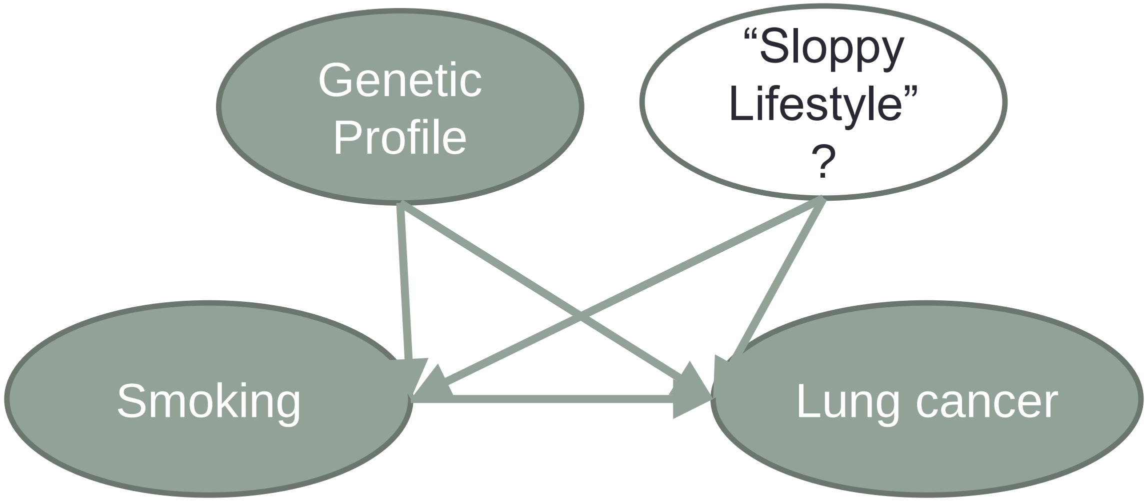
# But… We Can’t Randomize



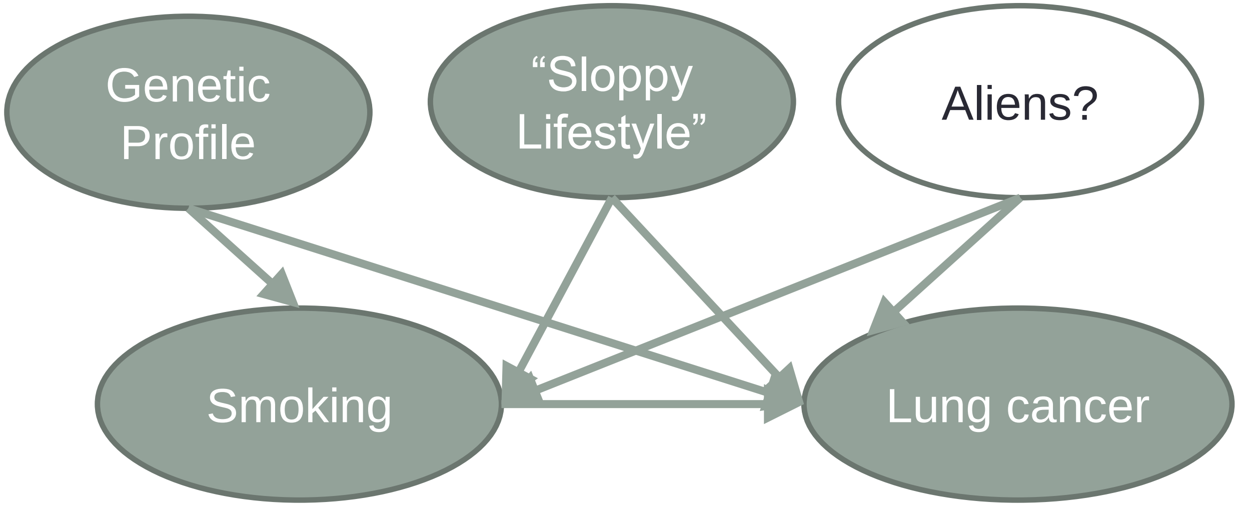
# “Adjust”



# But… What If?...



# And So On


Genetic Profile
"Sloppy Lifestyle"
Aliens?
Smoking
Lung cancer

# Observational Studies

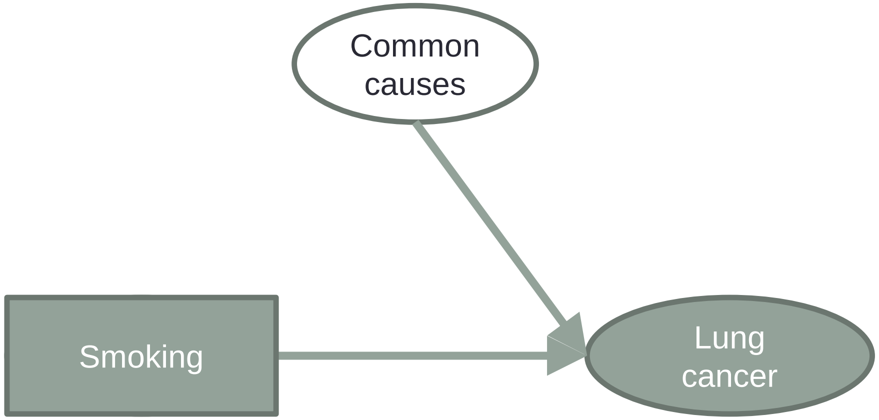
- The task of learning causal effects when we do not control the treatment, which instead comes in a “**natural regime**”, or “**observational regime**”.

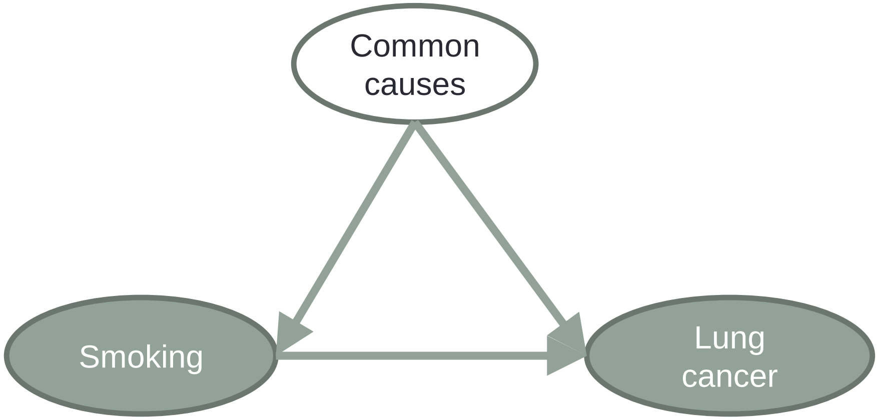
- The aim is to relate use the data in the observational regime to infer effects in the **interventional regime**.

# That Is

We would like to infer P(Outcome | Treatment) in a “world” (regime) like this

All we have is (lousy?) data for P(Outcome | Treatment) in a “world” (regime) like this instead





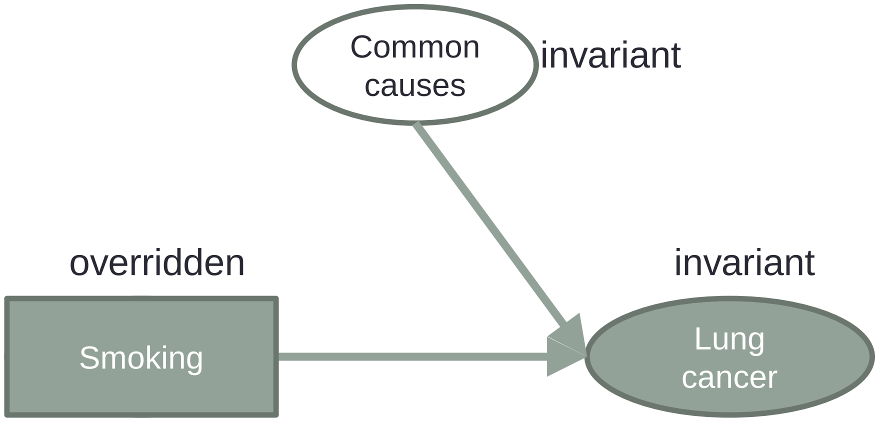
# What Now?

- To do “smoothing” across regimes, we will rely on some **modularity assumptions** about the underlying causal processes.

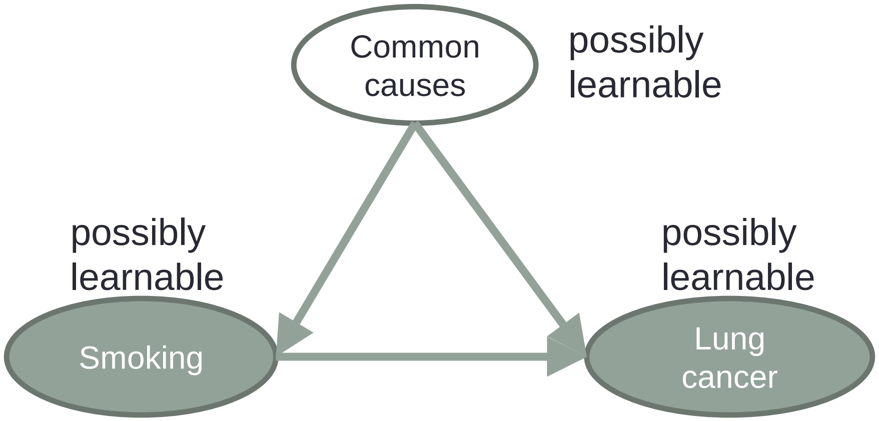
- We just have the perfect tool for the job: graphical models.

# What Now?

- The jump to causal conclusions from observational data requires some “smoothing” assumptions **linking different regimes**.



**Interventional Regime:**
**P(Outcome | do(Treatment))**



**Observational Regime:**
**P(Outcome | Treatment)**

# Task

- Say you have some **treatment X** and some **outcome Y**.

- Say you have some **background variables Z** you do observe in your data, and which may (or may not) block all paths along common causes of X and Y.

- **Find me a measure of how Y changes when I intervene on X at different levels.**

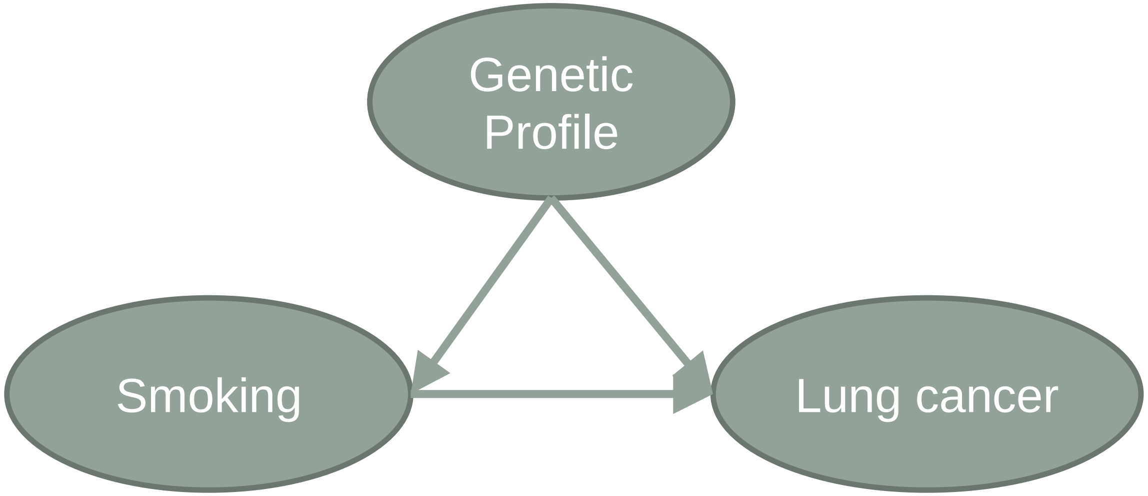
# What is a Perfect Intervention?

- A perfect intervention on some X is an independent **cause** of X that sets it to a particular value, **all other things remain equal**.
- …

# What is a Perfect Intervention?

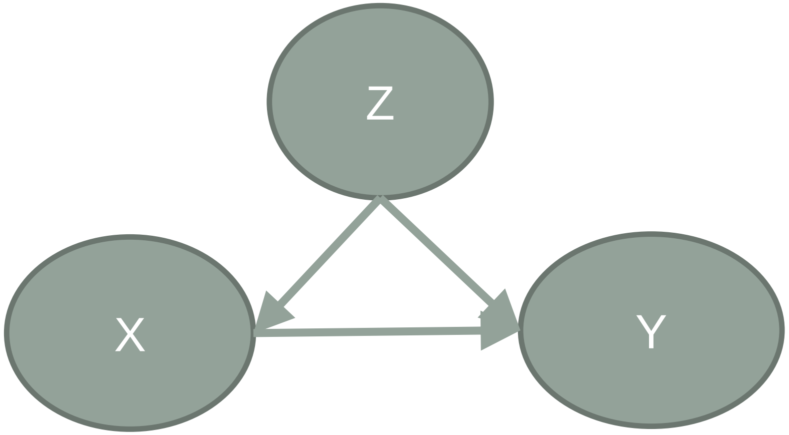
- We won’t define it. **We will take it as a primitive**.
- “I know it when I see it.”
- Operationally, this just wipes out all edges into X and make it a constant, **all other things remain equal**.
- How is it related to randomization?

# Trick 1: “Adjust” (a.ka., “The Backdoor Adjustment”)



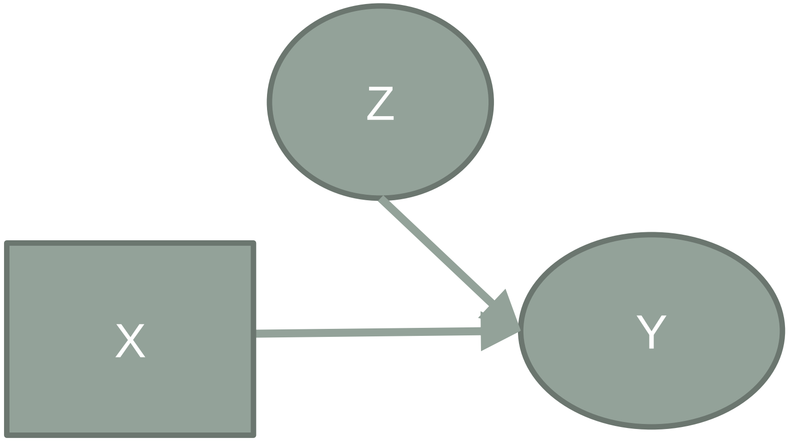
# Why It Works

- Estimand: P(Y | **do(X = x)**), not P(Y | X = x)
- Model:



- Relation to estimand:
  - $P(Y \mid do(x)) = \sum_{z} P(Y \mid do(x), Z = z)\, P(Z = z \mid do(x))$

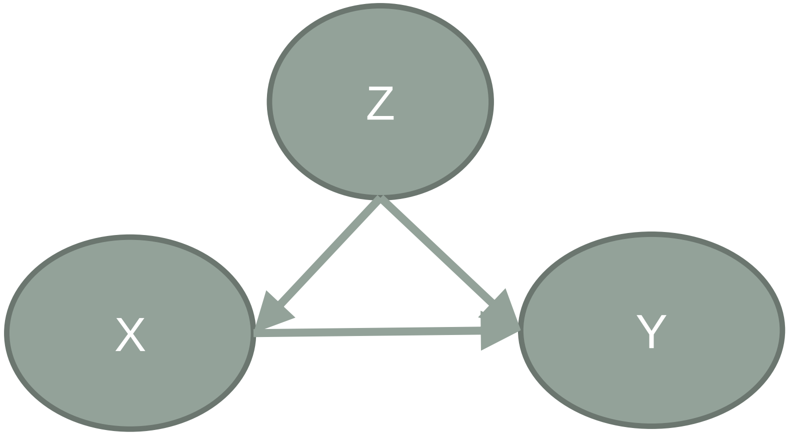
# Why It Works



$$P(Y \mid do(x)) = \sum_{z} P(Y \mid \underset{\text{invariance}}{\cancel{do(x)}}, Z = z)\; P(Z = z \mid \underset{\text{invariance}}{\cancel{do(x)}})$$

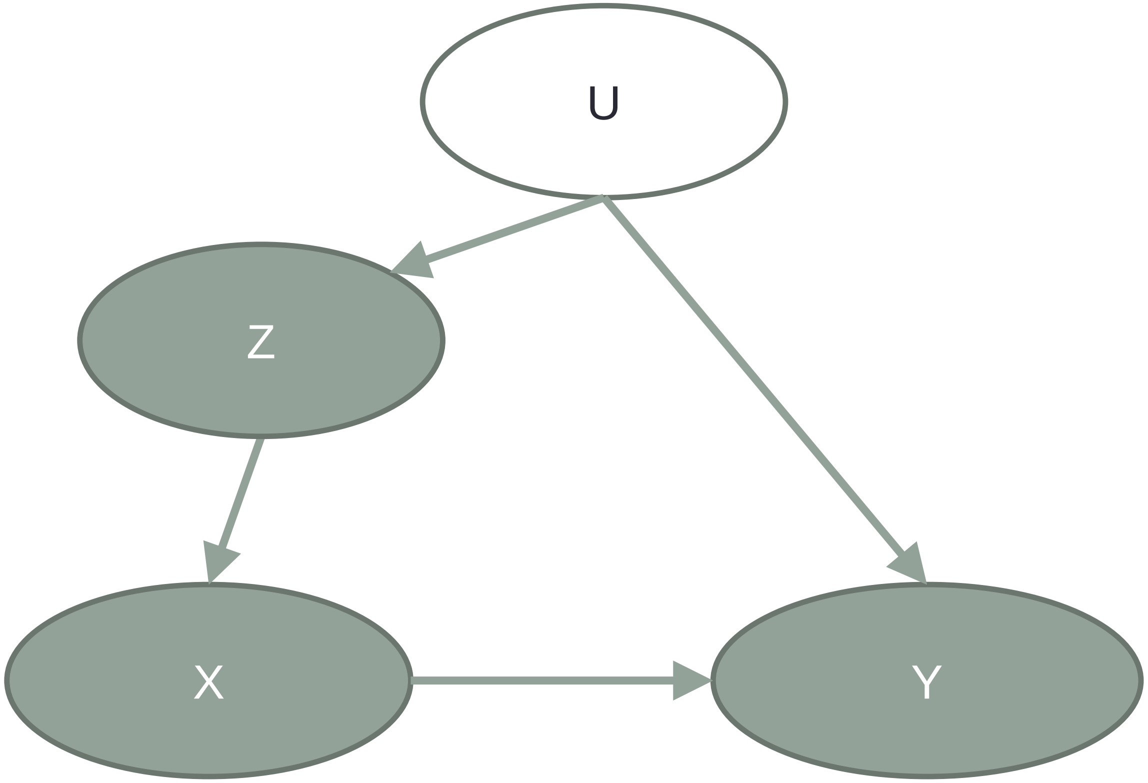
$$= \sum_{z} P(Y \mid X = x, Z = z)\; P(Z = z)$$

# Contrast!



$$P(Y \mid X = x) = \sum_{z} P(Y \mid X = x, Z = z)\, P(Z = z \mid X = x)$$

# Note: We don’t really need “all” hidden common causes

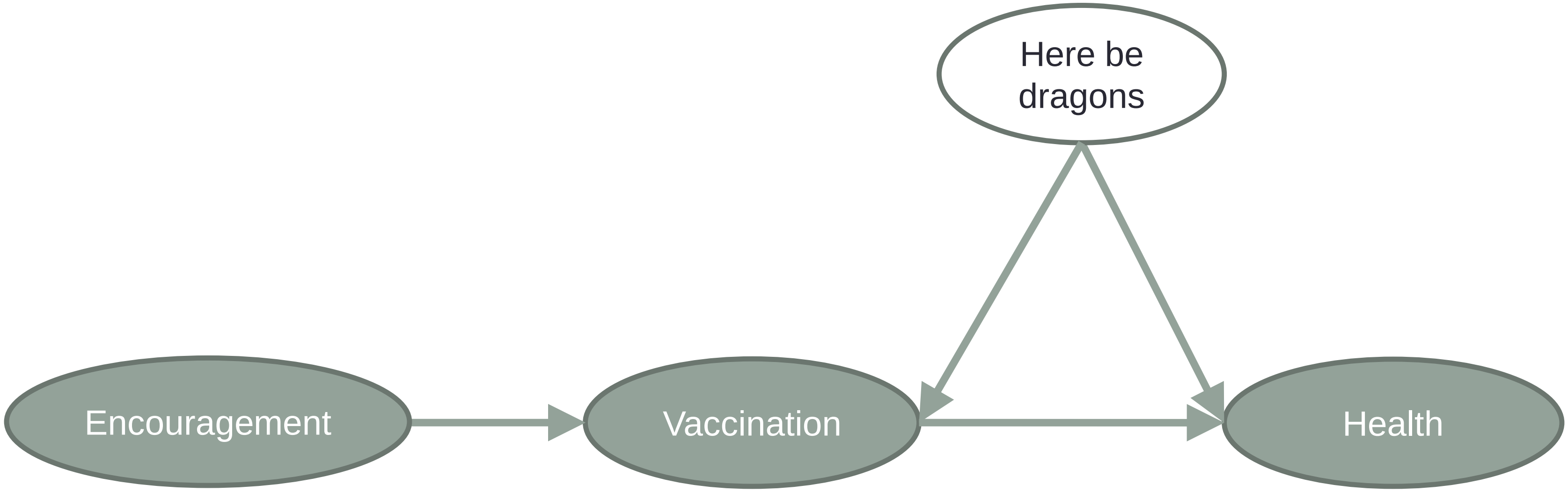


# Criticisms

- What if I don’t buy the assumption we were able to block all hidden common causes? Is there any hope?

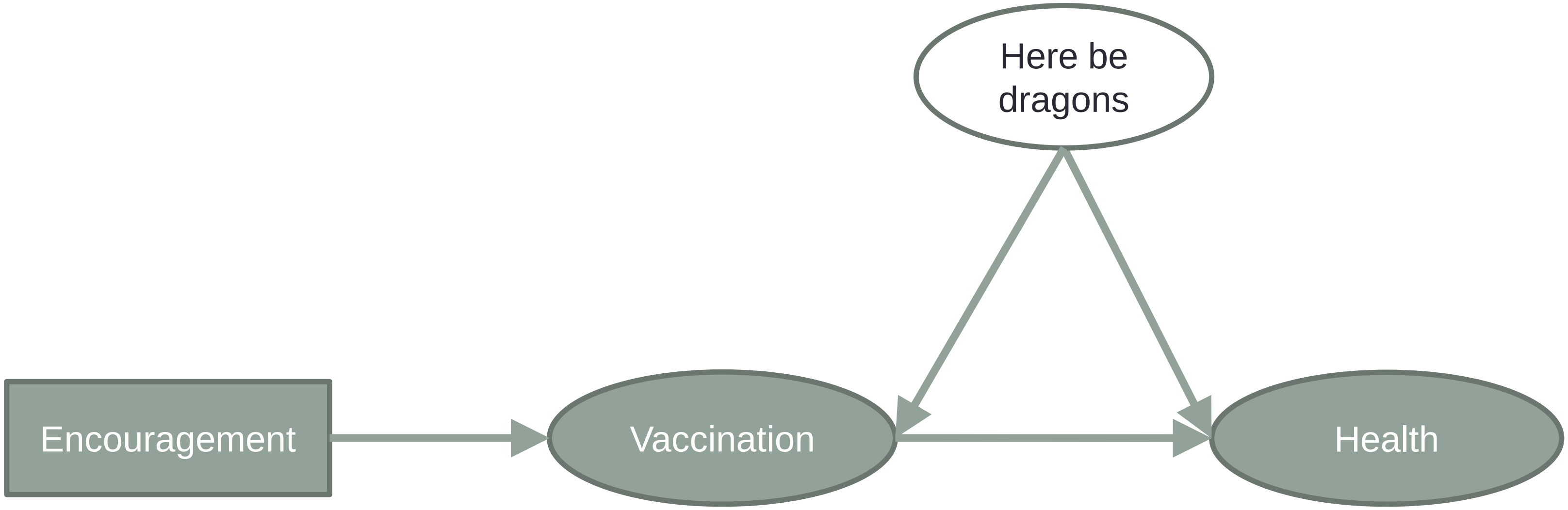
# Trick 2: Instrumental Variables

- Variables that can act as “surrogate” experiments.
- Sometimes they *are* surrogate experiments.
- Let’s look at some vaccination data.



# Instrumental Variables

- Variables that can act as “surrogate” experiments.
- Sometimes they *are* surrogate experiments.
- Let’s look at some vaccination data.



# Why Do We Care?

- Instrumental variables **constraint** the distribution of the **hidden common causes**
- It can be used to infer **bounds** on causal effects or, **under further assumptions, the causal effects** even if hidden common causes are out there.

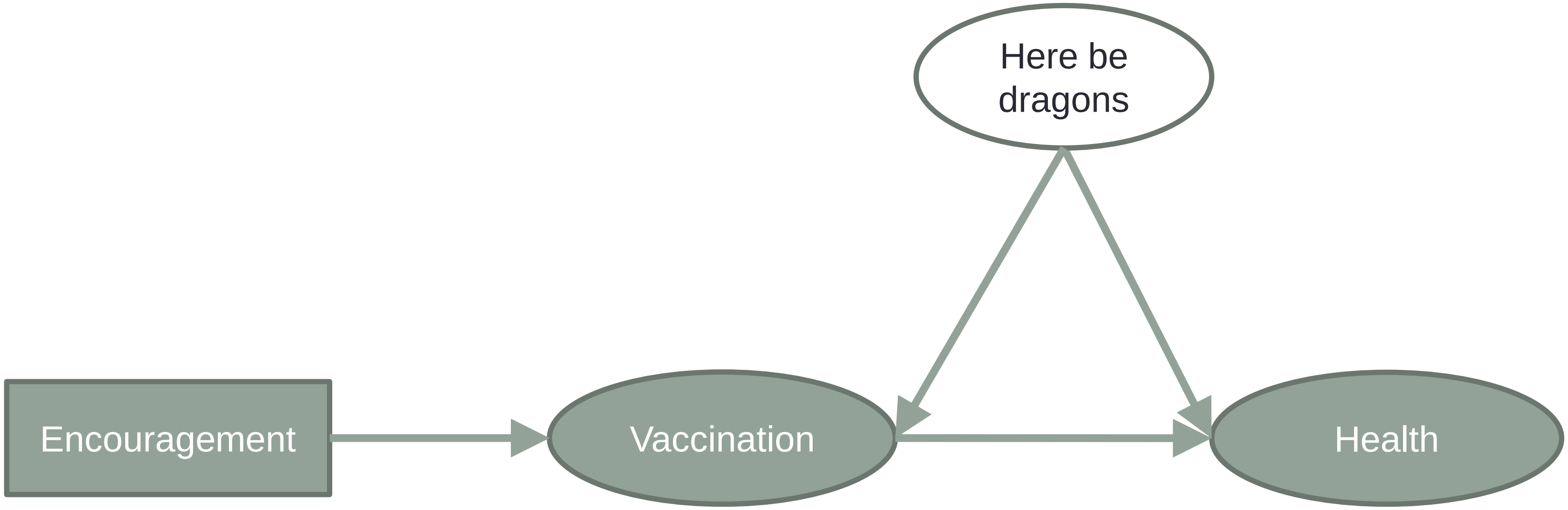
# In the Real World

“It turns out the department of correction's software was improperly giving some inmates credit for good behavior.”

http://www.npr.org/2016/01/01/461700642/computer-glitch-leads-to-mistaken-early-release-of-prisoners-in-washington?utm_campaign=storyshare&utm_source=facebook.com&utm_medium=social

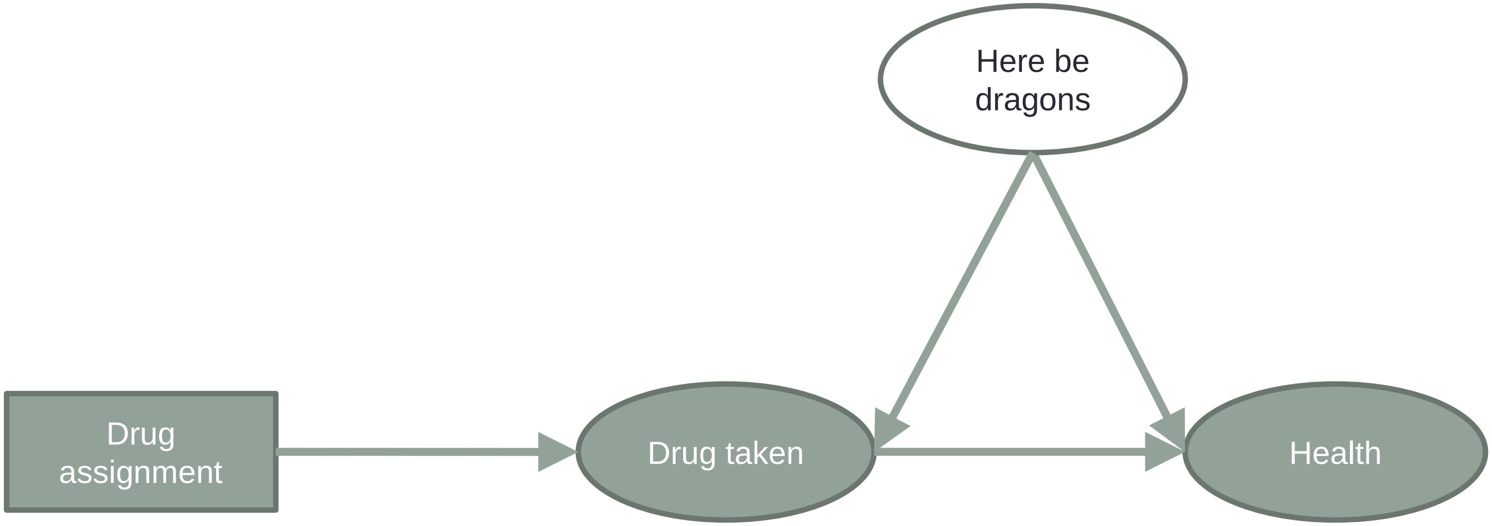
# Horses Appearing?

- But it looks like we can get the effect of encouragement on health. Isn’t this enough?
  - Also known in the literature as **intention to treat** effect.



# Instrumental Variables and "Broken Experiments"

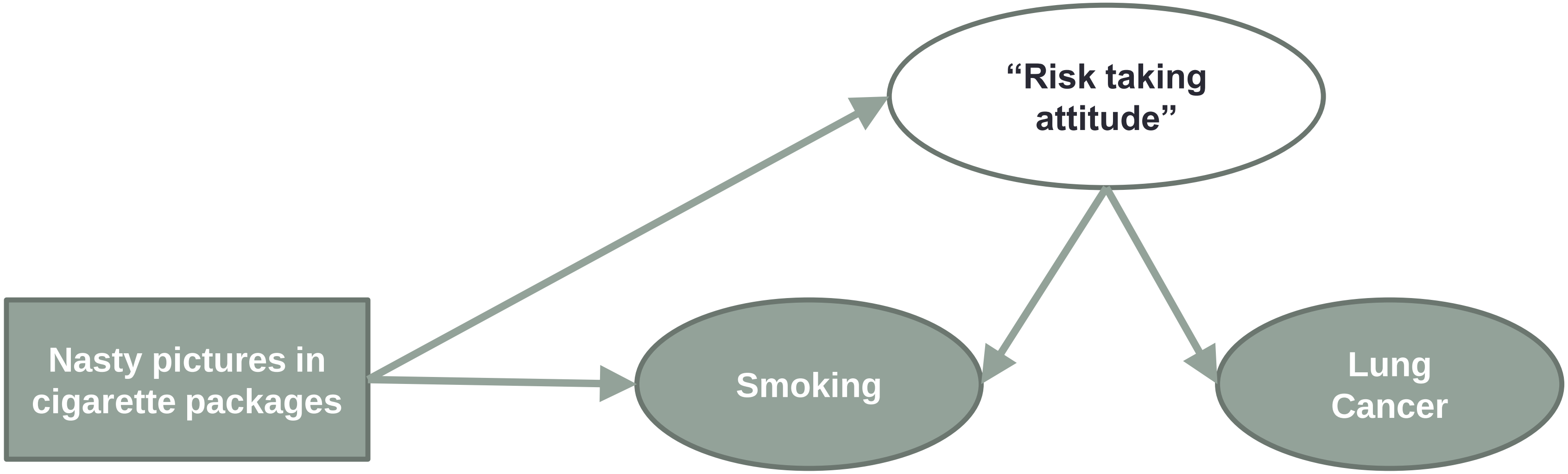
- Even randomized controlled trials might not be enough.
- Another reason why the machinery of observational studies can be so important.
- Consider the **non-compliance problem** more generally.



# Intention-to-Treat, Policy Making, and Horses

- From the RCT, we can indeed get the intention-to-treat effect.

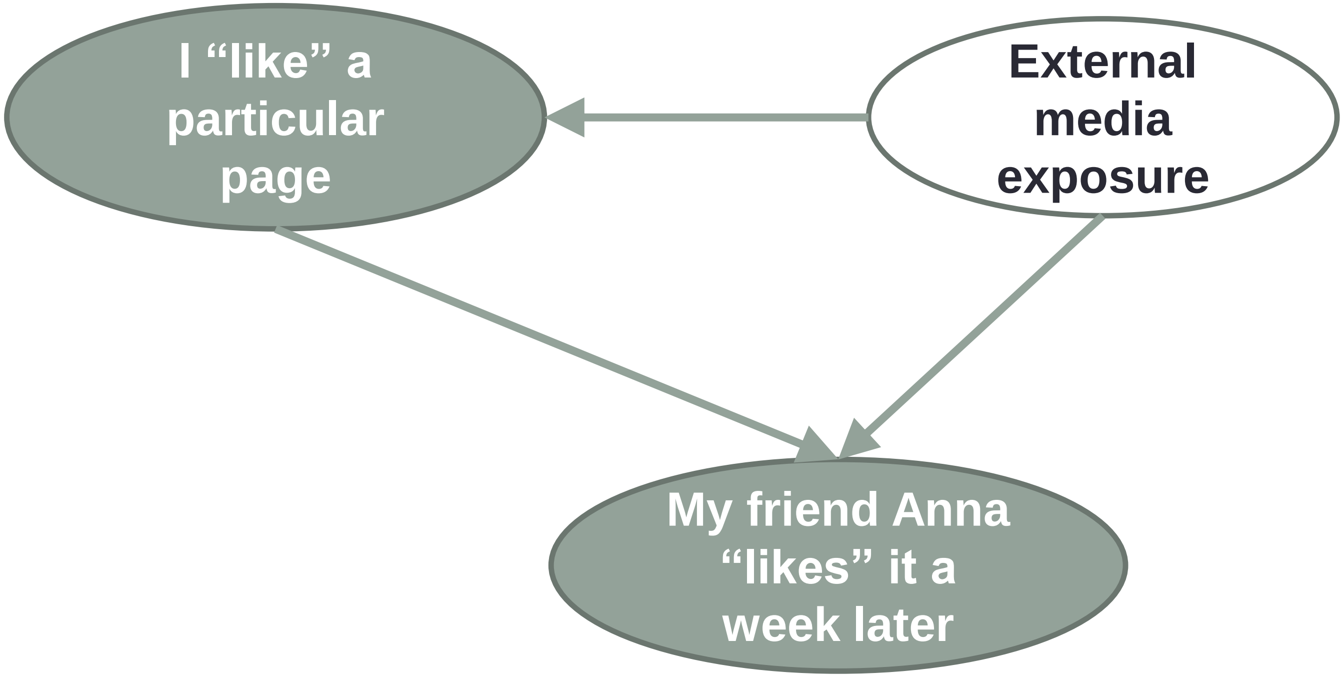
- From the point of view of policy making, would that be enough?



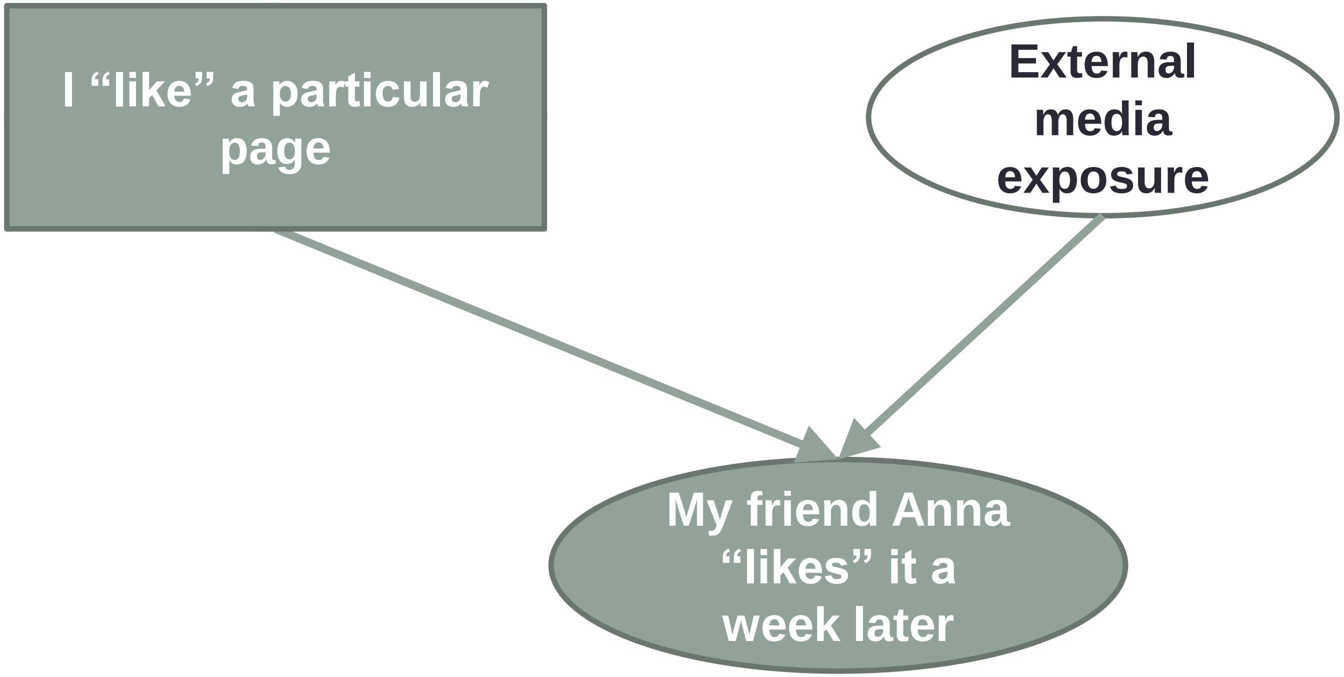
# A Modern Example

- What is the social influence of an individual or organization?
- It is pointless to define it without causal modelling.
  - Orwellian frame: “If we control the source, we control the followers.”
- Much social influence analysis out there is not necessarily wrong, but it may certainly be naïve.
- Time ordering is very far from enough.
  - Time of measurement is not the same as time of occurrence!
  - What are the common causes?

# Broken Experiments of Social Influence


I "like" a particular page
External media exposure
My friend Anna "likes" it a week later

# What Facebook-like Companies Would Love to Do

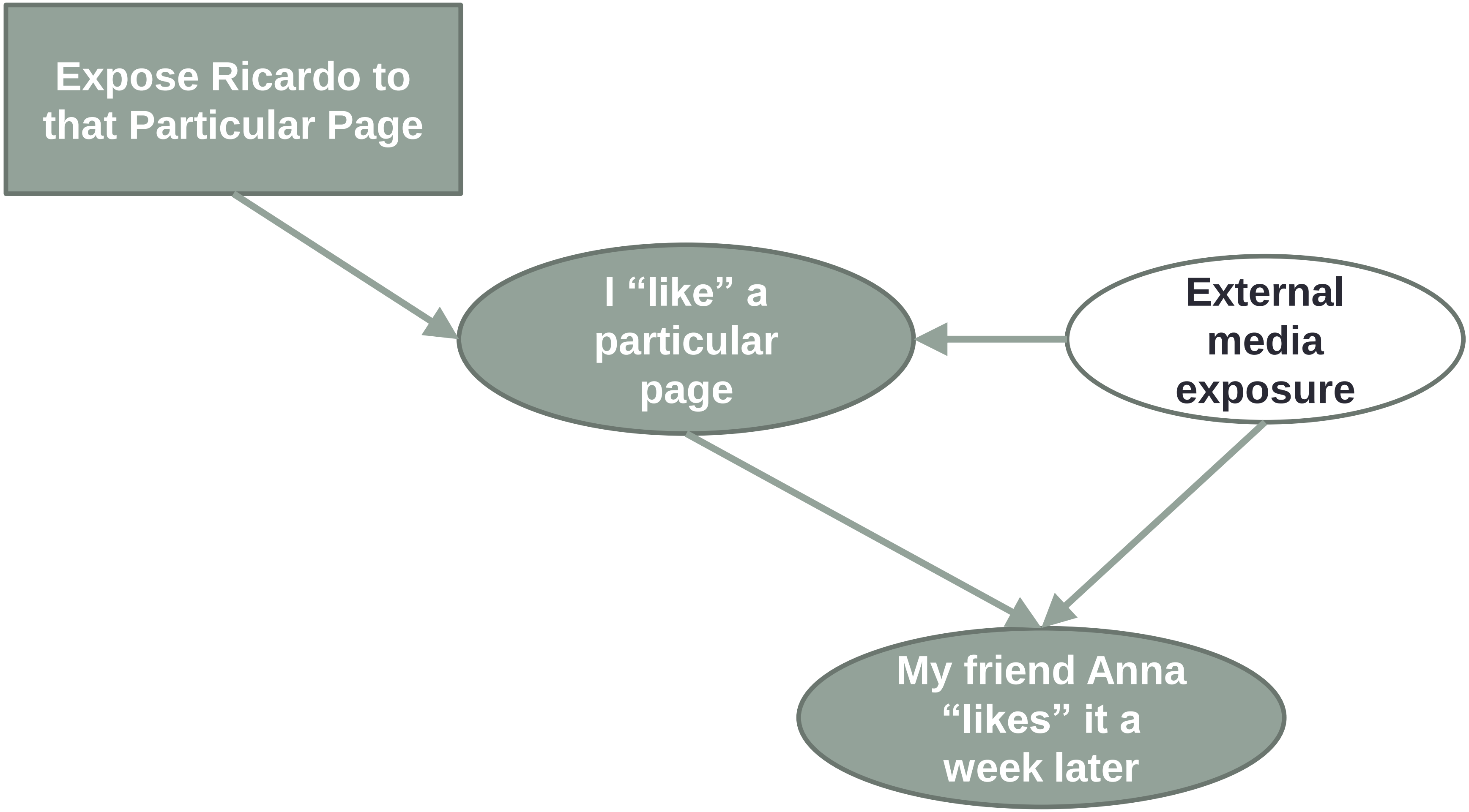


I “like” a particular page

External media exposure

My friend Anna “likes” it a week later

# What They Can Actually Do



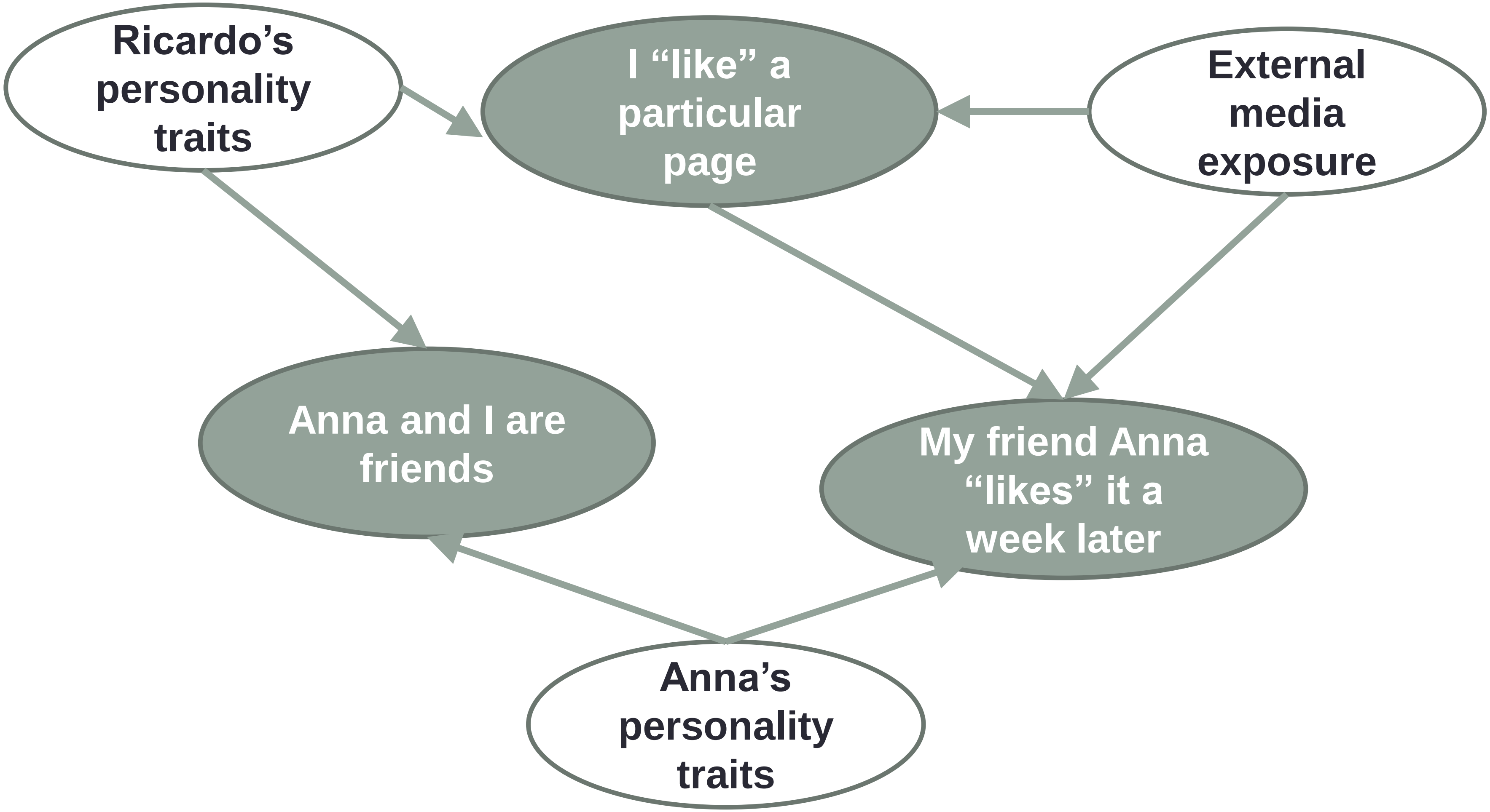
Expose Ricardo to that Particular Page

I “like” a particular page

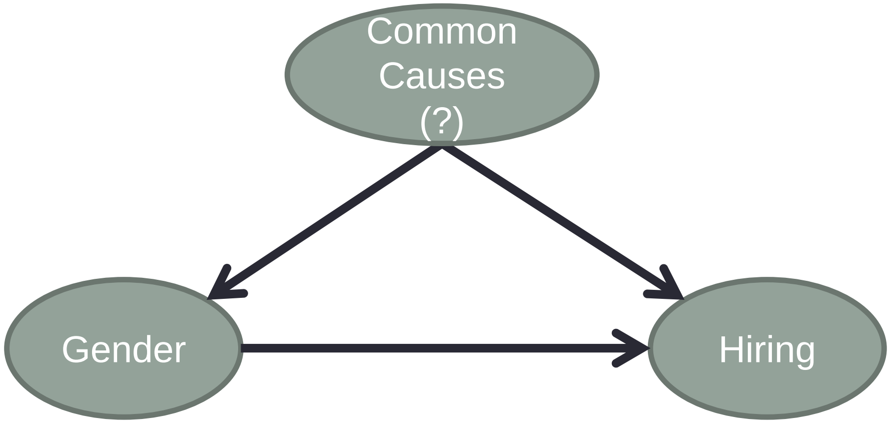
External media exposure

My friend Anna “likes” it a week later

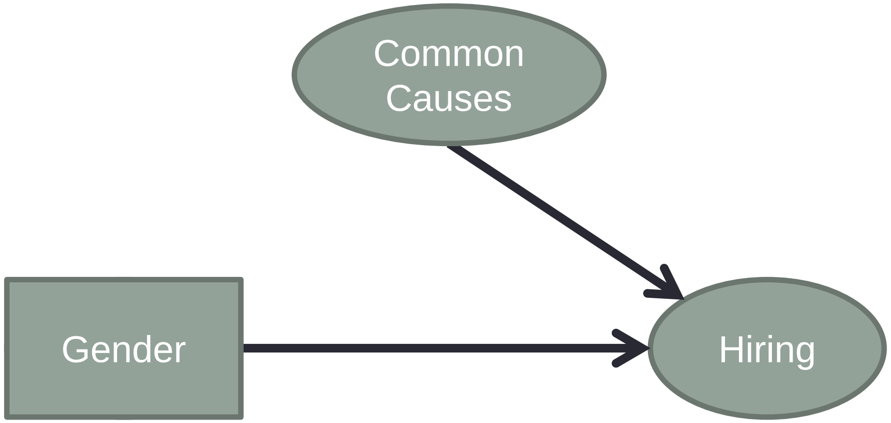
# Wait, It Gets Worse


Ricardo's personality traits
I "like" a particular page
External media exposure
Anna and I are friends
My friend Anna "likes" it a week later
Anna's personality traits

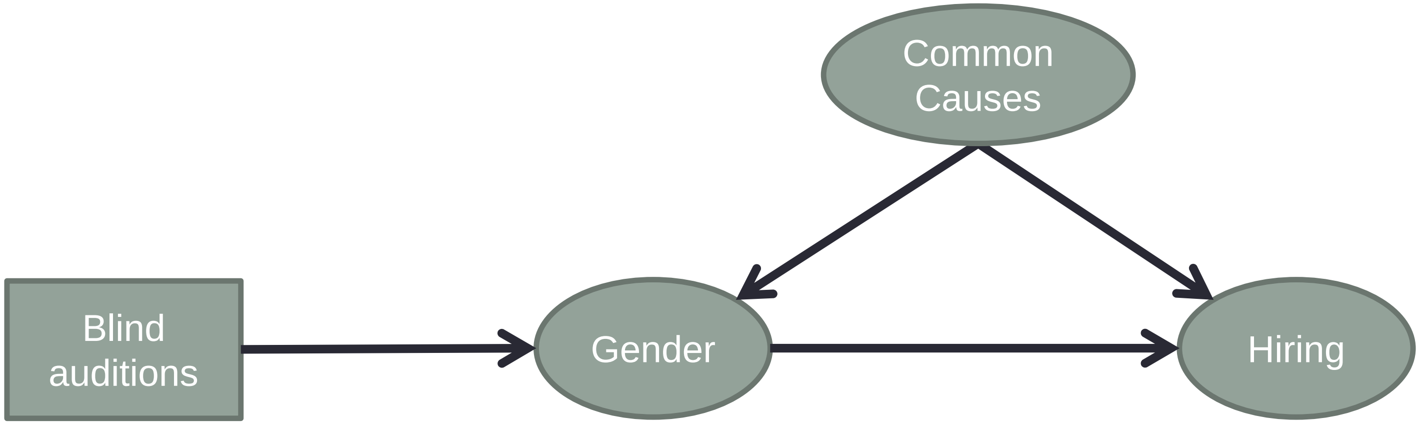
# It Gets “Worser”: Don’t Take Your Measurements and Interventions for Granted



# What Does That Mean?

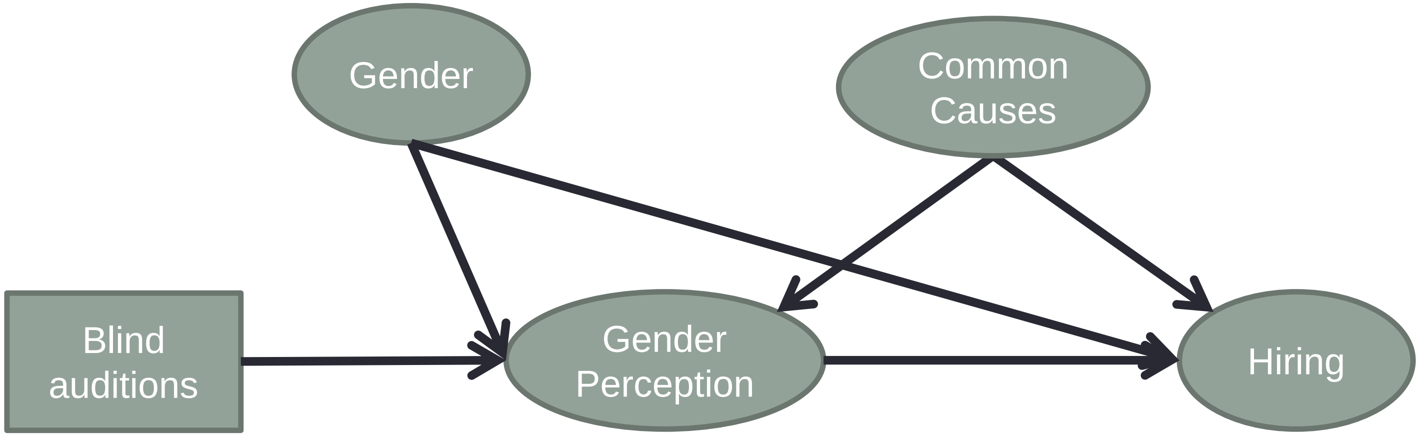


# What About This?



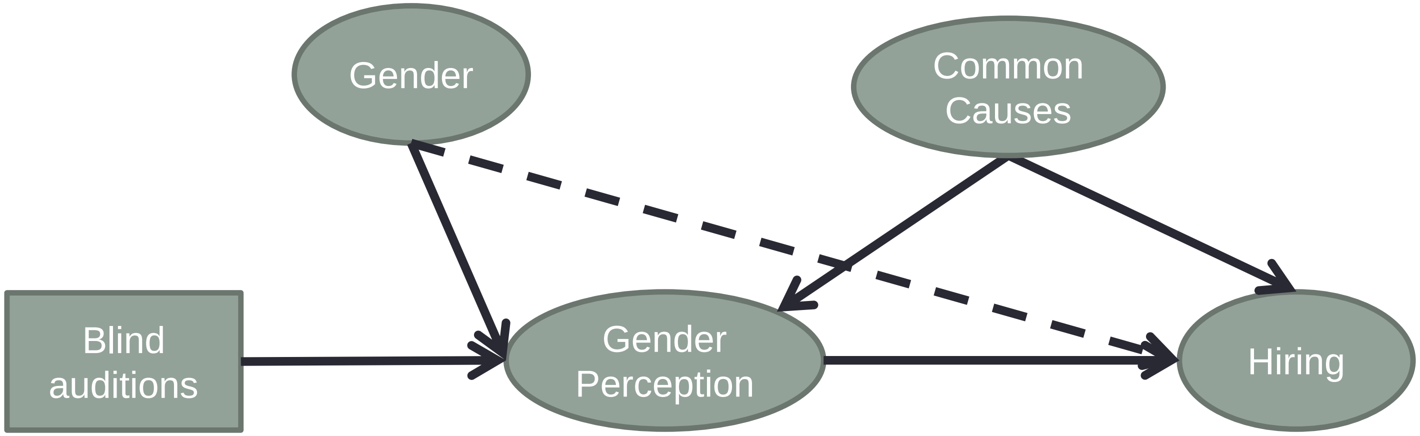
http://www.theguardian.com/women-in-leadership/2013/oct/14/blind-auditions-orchestras-gender-bias

# What About This?

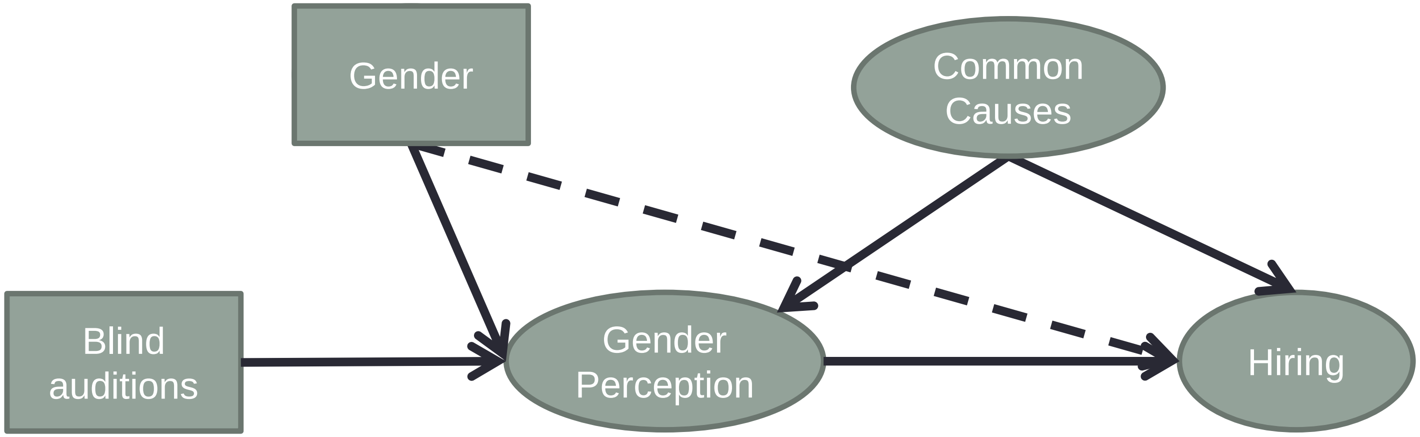


# What About (Lack of) “Direct Effects”?

- I’d appeal to further assumptions and see how Gender and Hiring can be made independent by Gender Perception and other covariates.



# But What Does That Mean???



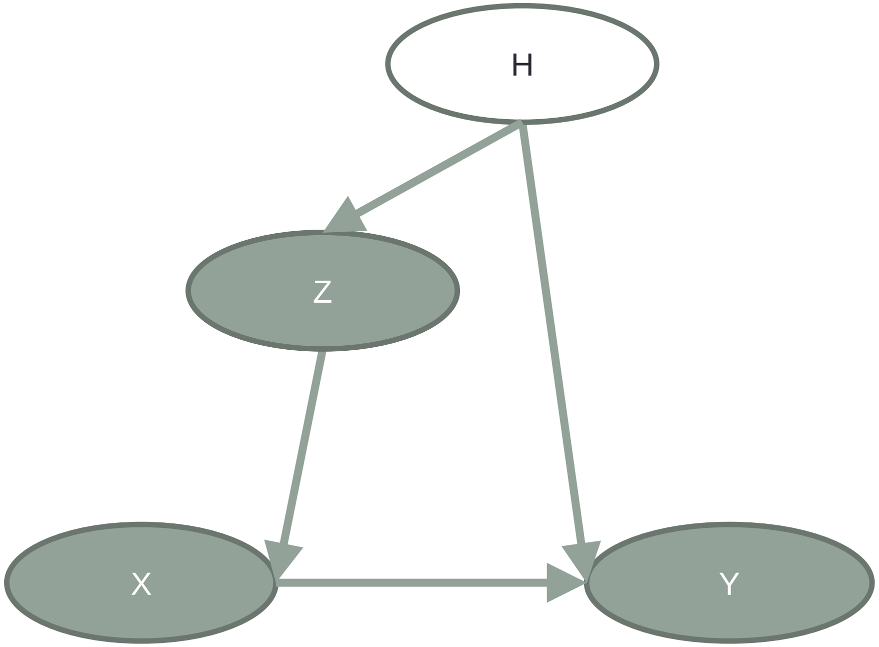
# FROM DATA TO GRAPHS

An Algorithm for Bounding Causal Effects

Joint work with Robin Evans (Oxford, Statistics)

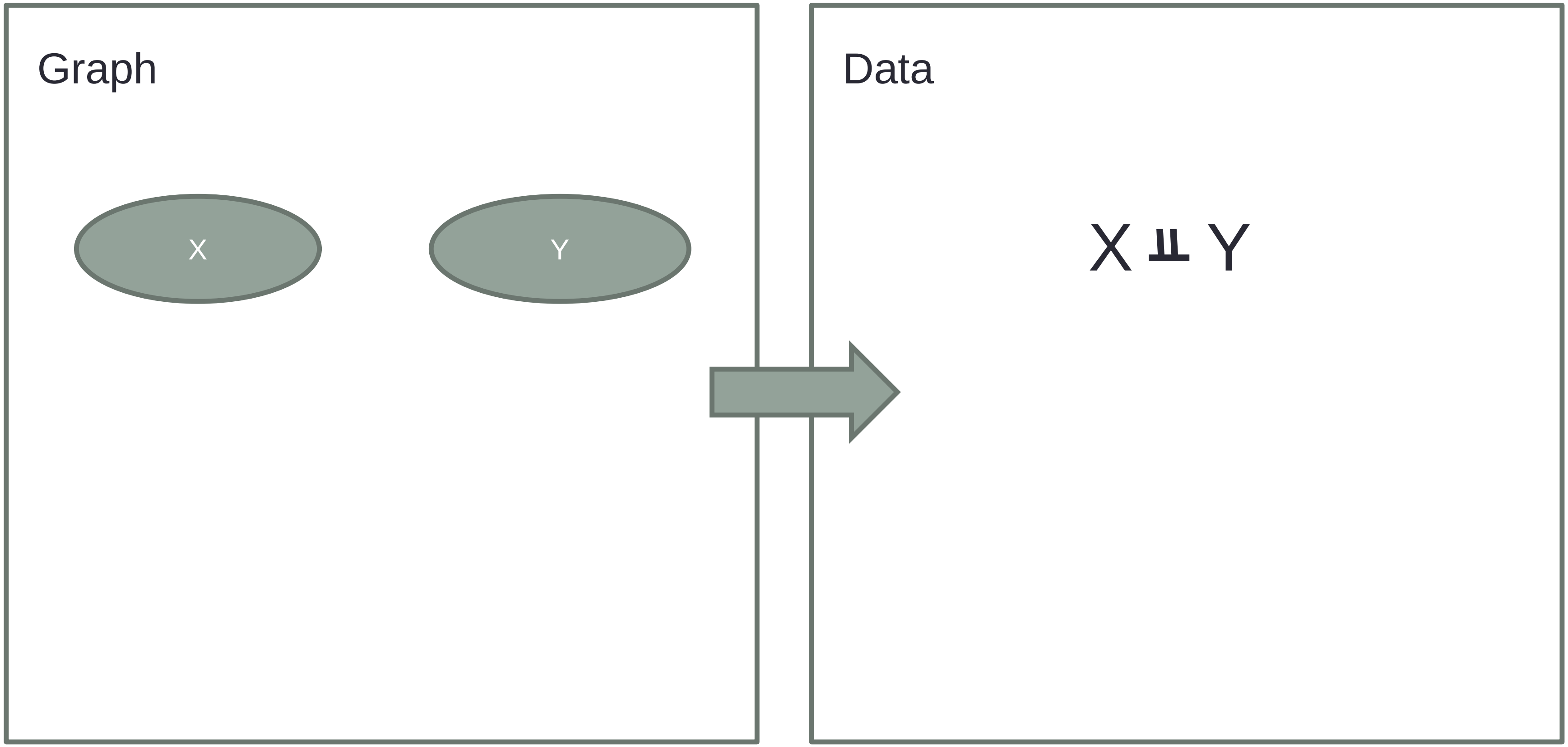
# Those Back-door Adjustments

- Can we get some **proof** or **certificate** we are doing the right thing using data, not only background knowledge?



# Structure Learning

- Inferring graphs from testable observations



Graph

X

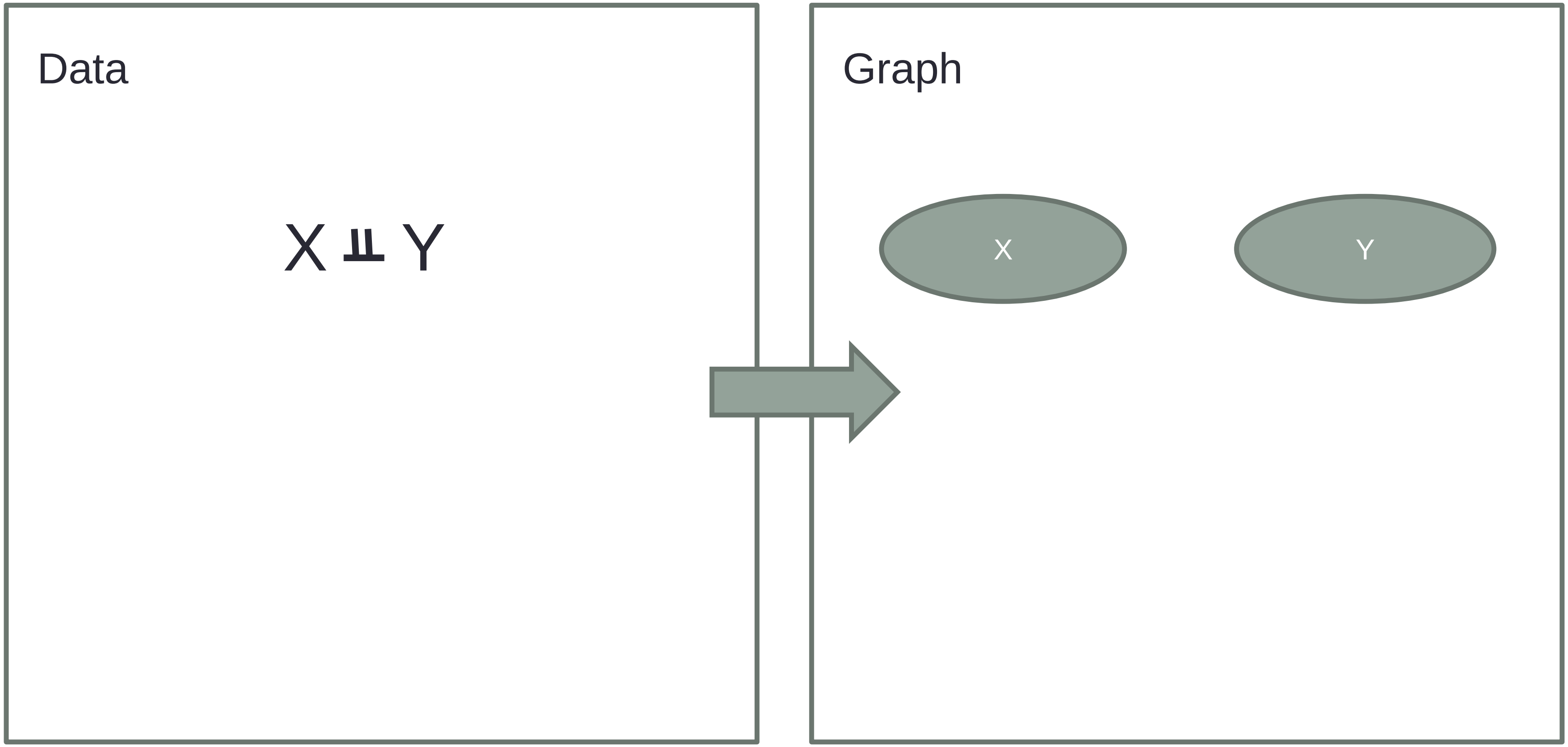
Y

Data

$X \perp\!\!\!\perp Y$

# Structure Learning

- Inferring graphs from testable observations



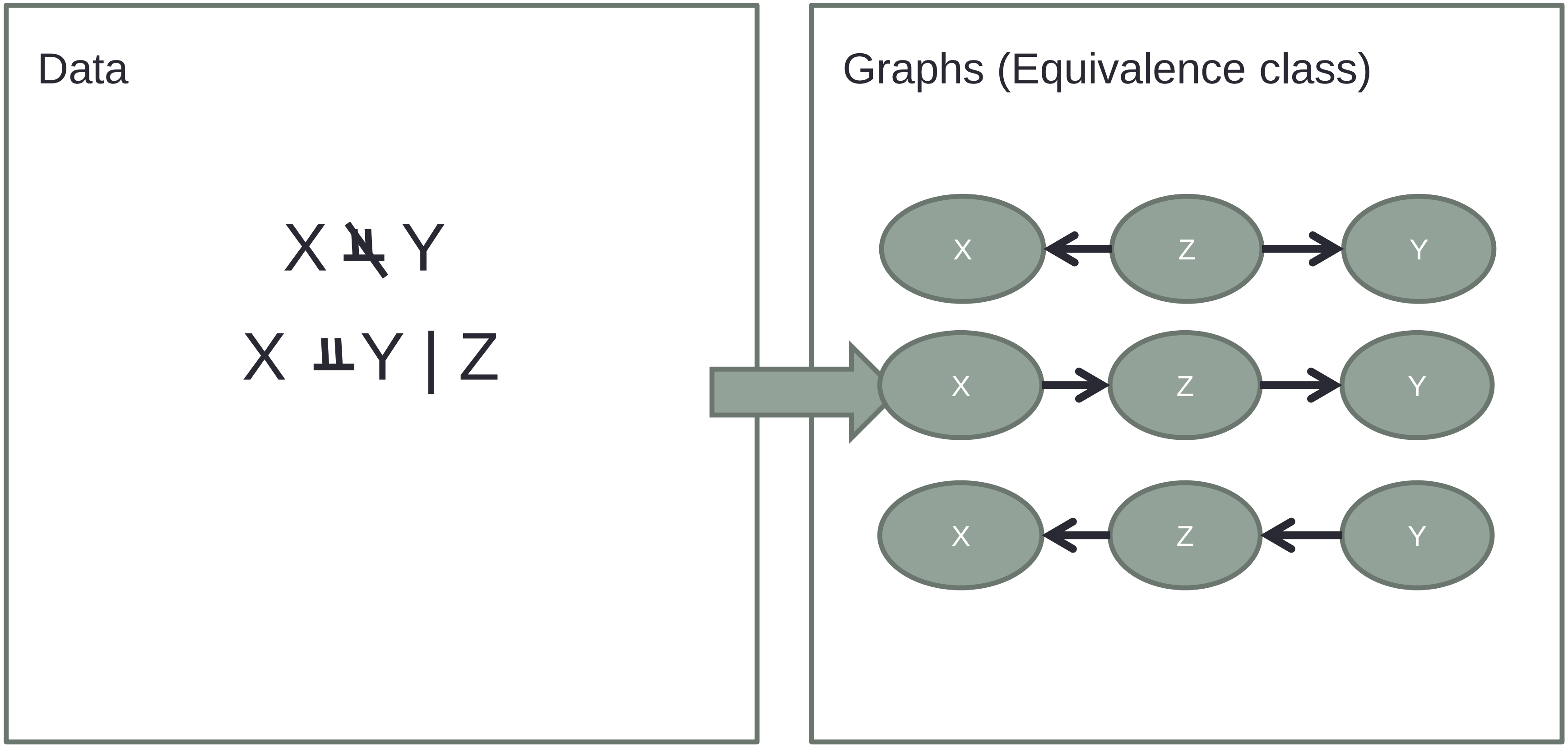
Data

$$X \perp\!\!\!\perp Y$$

Graph

# Structure Learning

- Inferring graphs from testable observations



Data

$X \not\perp\!\!\!\perp Y$

$X \perp\!\!\!\perp Y \mid Z$

Graphs (Equivalence class)

# Equivalence Class?

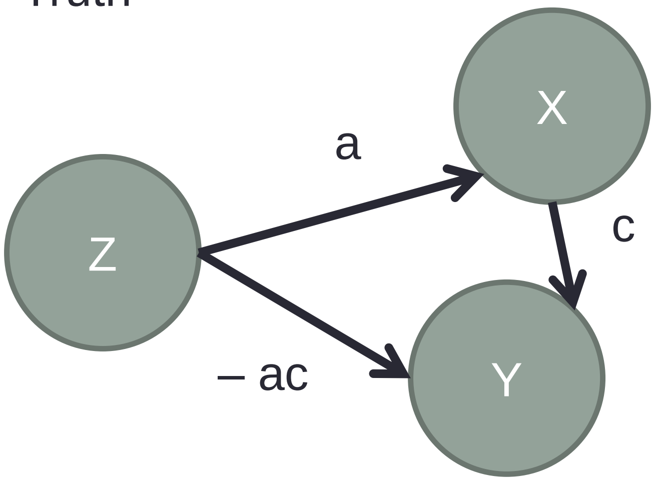
- Just life effect identification, graph identification might not be possible. It will depend on which assumptions we are willing to make.

- For instance,
  - Partial ordering
  - Parametric relationships, like linear effects

# Main Assumption: **Faithfulness**

- "Non-structural independencies do not happen."

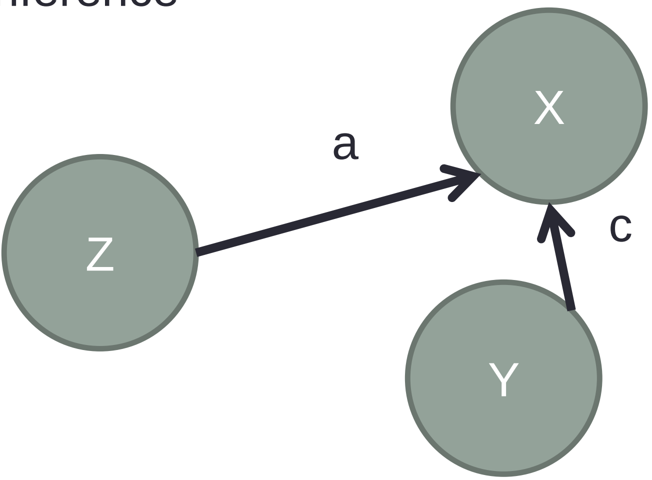
Truth



$Y \perp\!\!\!\perp Z$

$Y \not\!\perp\!\!\!\perp Z \mid X$

Inference

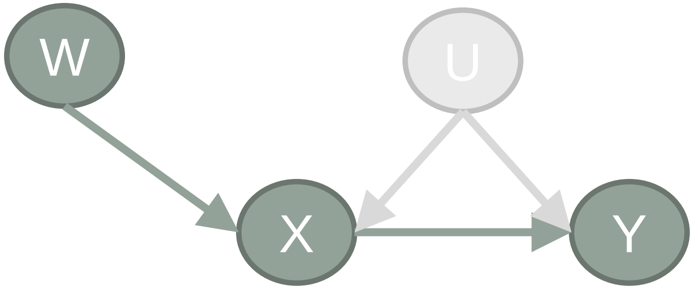


$Y \perp\!\!\!\perp Z$

$Y \not\!\perp\!\!\!\perp Z \mid X$

# Example

- W not caused by Y nor Y, assume ordering X → Y
- $W \not\perp\!\!\!\perp X$, $W \perp\!\!\!\perp Y \mid X$ + Faithfulness. Conclusion?



No unmeasured confounding

- Naïve estimation works:
  Causal effect = $P(Y = 1 \mid X = 1) – P(Y = 1 \mid X = 0)$

- This super-simple nugget of causal information has found some practical uses on large-scale problems.

# Application

- Consider “the genotype at a fixed locus *L* is a random variable, whose random outcome occurs before and independently from the subsequently measured expression values”

- Find genes $T_i$, $T_j$ such that $L \rightarrow T_i \rightarrow T_j$

Chen, Emmert-Streib and Storey (2007)
Genome Biology, 8:R219

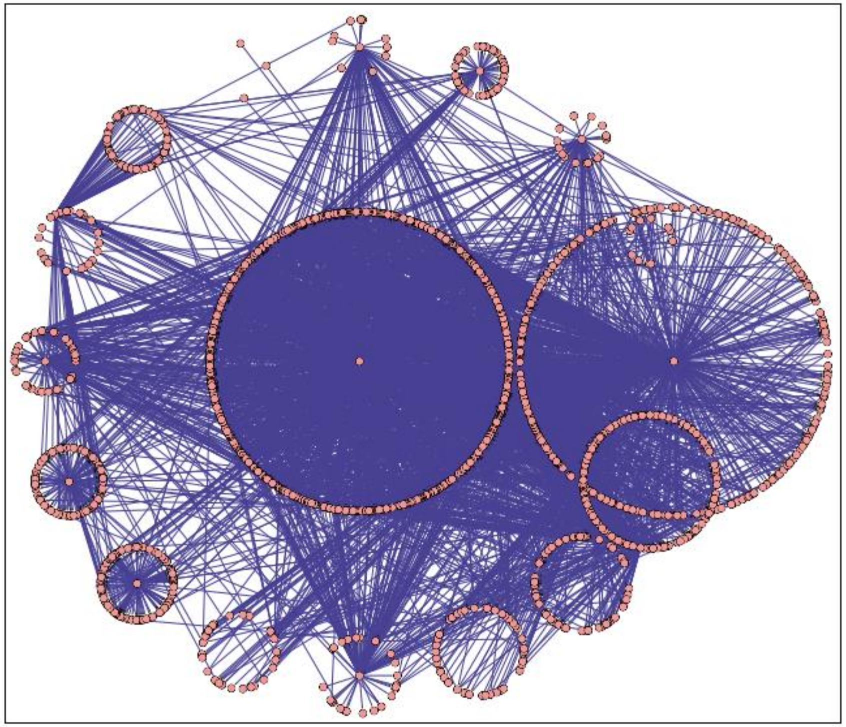
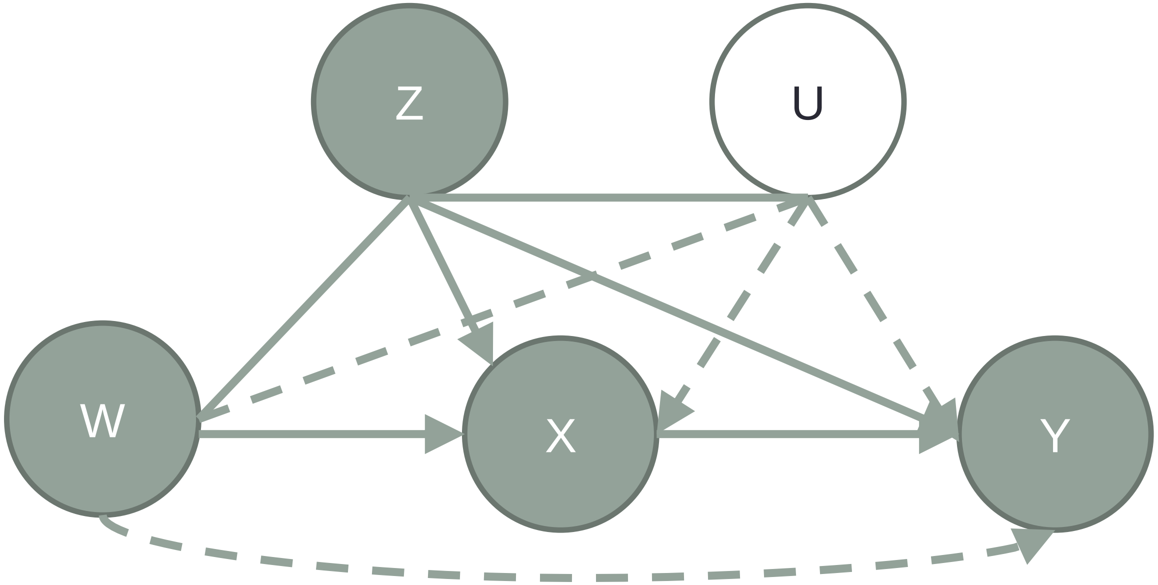


**Figure 2**
A transcriptional regulatory network drawn from a Trigger probability threshold of 90%. The network consists of 4,394 genes, 2,145 causal relationships, and 127 causal genes. Genes are represented by orange circles and causal relationships are represented by directed edges with black arrows.

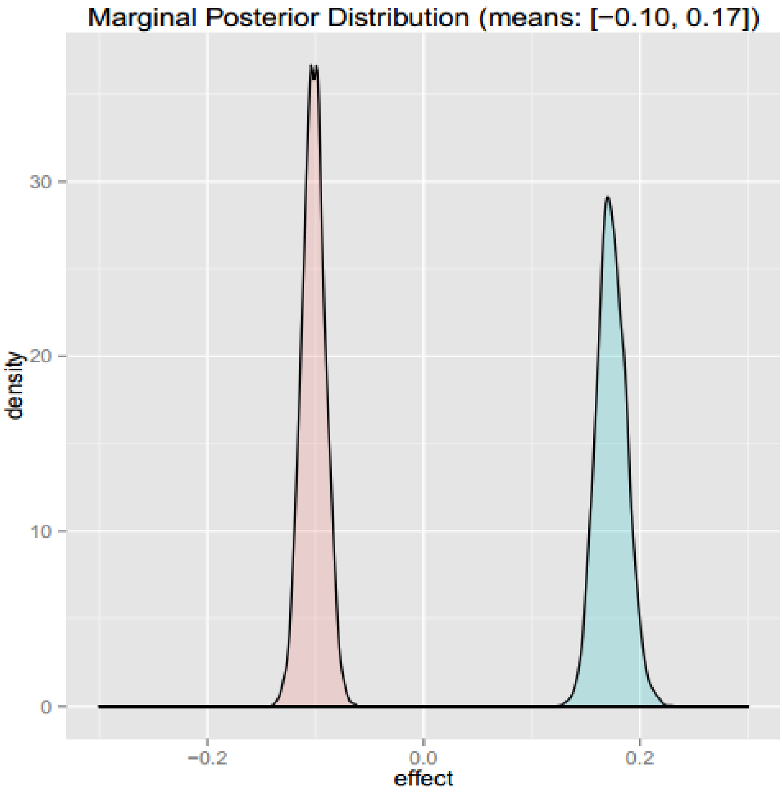
# A More General Method (Silva and Robins, JMLR2016)

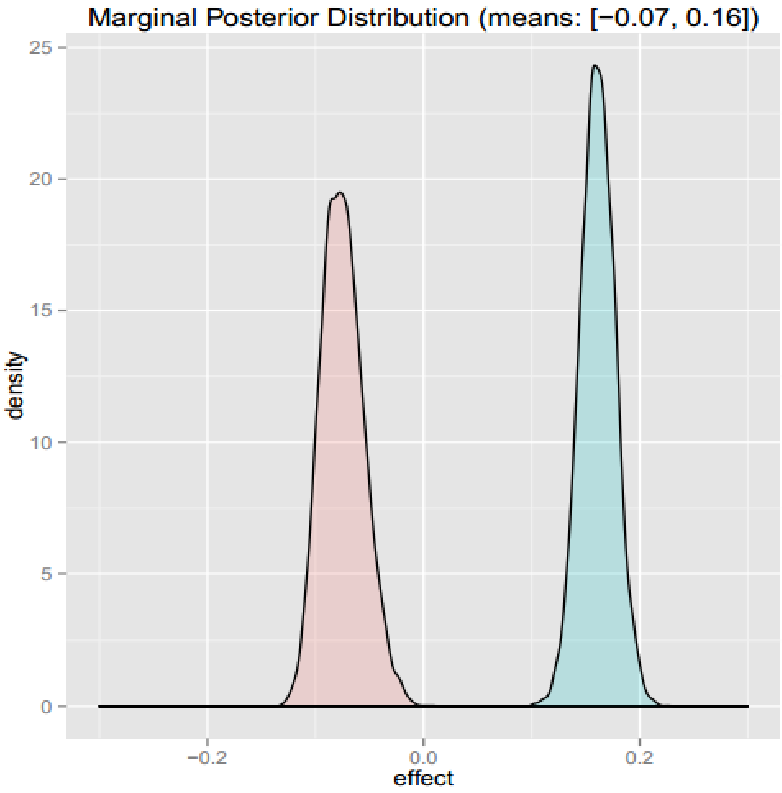
- We look at independence constraints that suggest “almost instruments” for the effect of X on Y, which allows for (weak) violations of faithfulness.
- We use it to learn average **bounds** on causal effects in discrete data.

$$ACE = P(Y = 1 \mid do(X = 1)) - P(Y = 1 \mid do(X = 0))$$



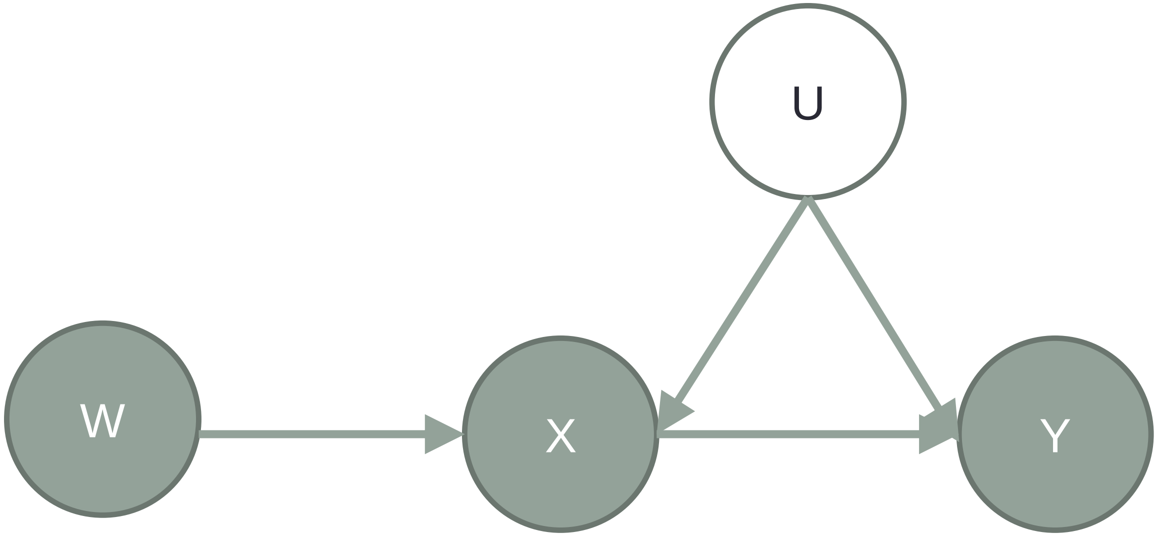
# Influenza Data: Example of Output


Marginal Posterior Distribution (means: [−0.10, 0.17])


Marginal Posterior Distribution (means: [−0.07, 0.16])

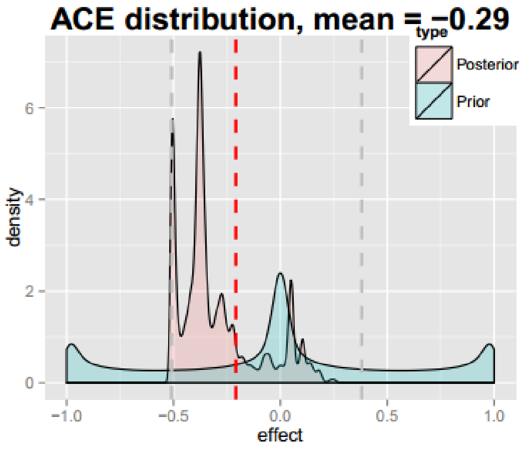
# Final Comment: A Self-Inflicted Horse

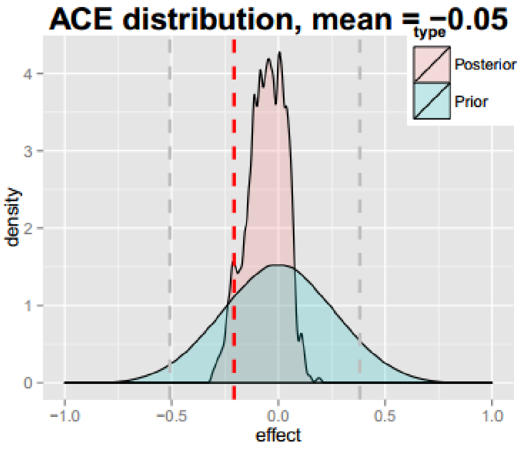
- Why don’t I put priors on this latent variable model and turn the crank of Bayesian inference?

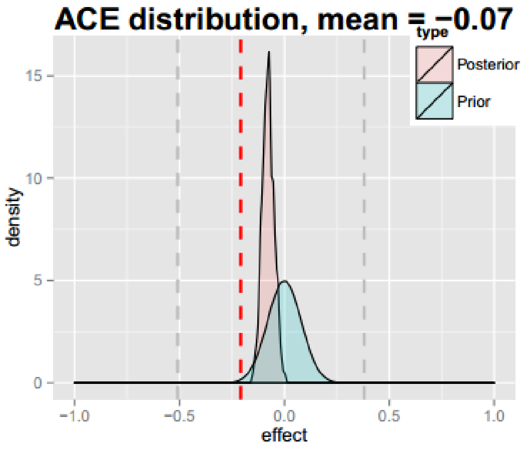


# Final Comment: A Self-Inflicted Horse

- However, model is unidentifiable  == results extremely sensitive to priors







# CONCLUSIONS

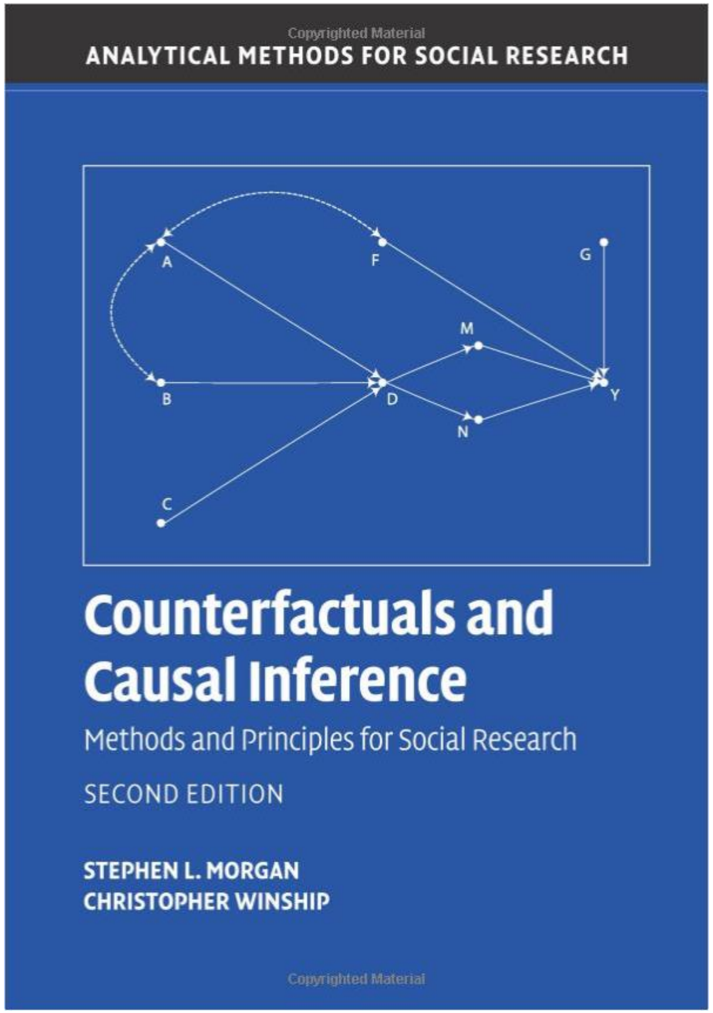
# Yes, It is Hard, But:

- Pretending the problems don’t exist won’t make them go away.
- There is a world out there to better explored by combining experimental and observational data.
- In particular, how to “design experimental design”.
- The upside of many causal inference problems is that **getting lower bounds and relative effects instead of absolute effects might be good enough**.

# Main Advice

**Don’t rely on a single tool.** If you can derive similar causal effects from different sets of assumptions, great. If they contradict each other, this is useful to know too. Make use of your background knowledge to disentangle the mess.

# Textbooks


Copyrighted Material

ANALYTICAL METHODS FOR SOCIAL RESEARCH

Counterfactuals and Causal Inference

Methods and Principles for Social Research

SECOND EDITION

STEPHEN L. MORGAN
CHRISTOPHER WINSHIP

Copyrighted Material

Excellent, but be warned: verbose

In press (soonish):

Hernán MA, Robins JM (2016). **Causal Inference**. Boca Raton: Chapman & Hall/CRC, forthcoming.

http://www.hsph.harvard.edu/miguel-hernan/causal-inference-book/

Shalizi, C. (2015?). **Advanced Data Analysis from an Elementary Point of View**. Cambridge University Press.

http://www.stat.cmu.edu/~cshalizi/ADAfaEPoV/

# Thank You, and Shameless Ad

## What If? Inference and Learning of Hypothetical and Counterfactual Interventions in Complex Systems

**A NIPS 2016 Workshop**
**Centre Convencions Internacional Barcelona, Barcelona, Spain**
**December 10th 2016**

**Deadline: October 31st**

**https://sites.google.com/site/whatif2016nips/call-for-papers**