Psych 253
Advanced Statistical Modeling

Graphical causal models

Daniel Yamins
Wu Tsai Neurosciences Institute
Departments of Psychology and Computer Science
Stanford University

Russ Poldrack
Department of Psychology
Stanford University

Everyone knows that we can’t infer causality from correlation, right?

Causality and correlation: An example

Does excessive alcohol consumption cause heart disease?

Alcohol and heart disease

The effects of alcohol on the heart, looking at the risks and also the potential benefits claimed by some researchers.

- What is coronary heart disease?
- The effects of alcohol
- The effects of binge drinking
- Alcohol's benefits?
- The facts

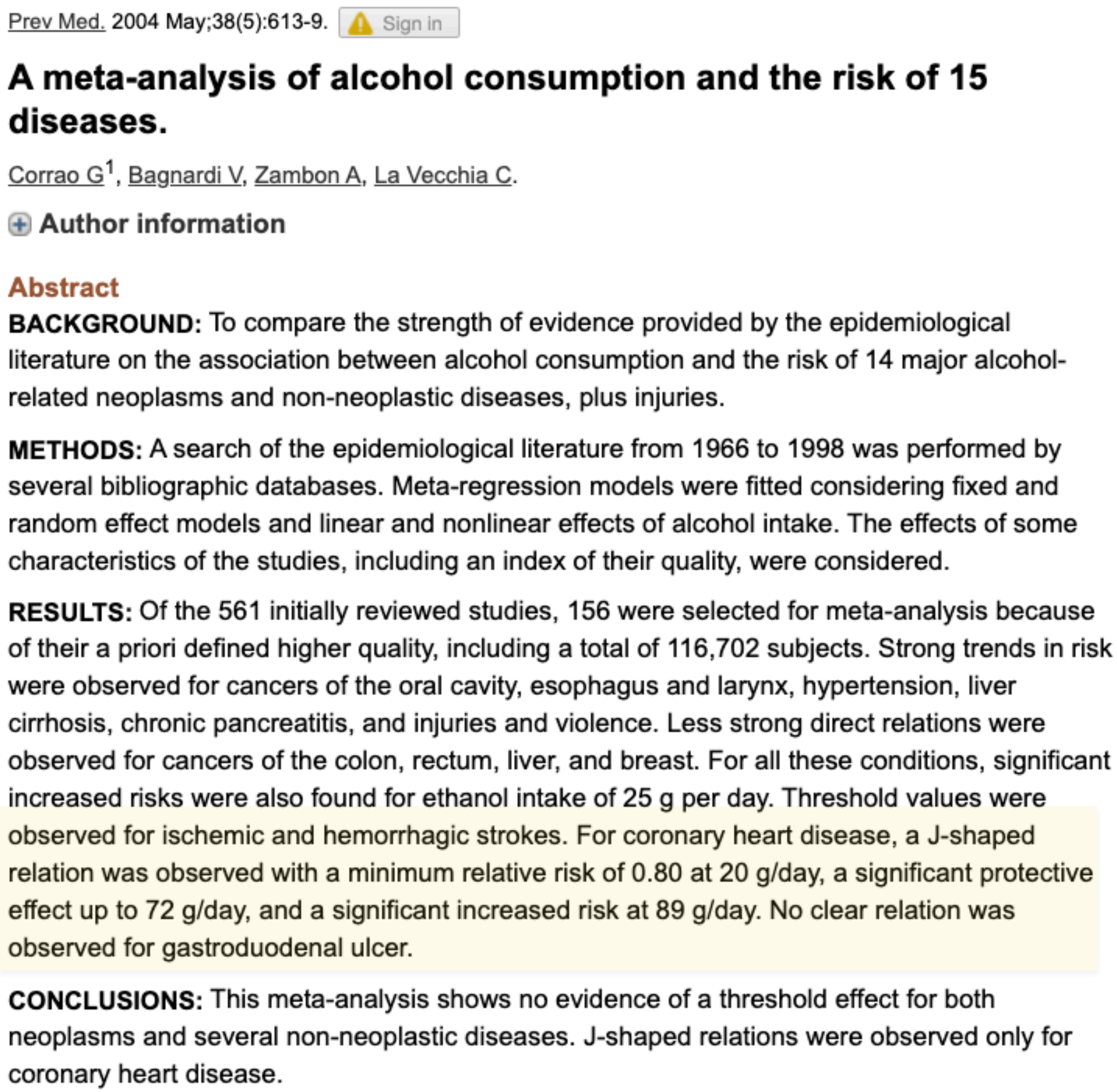
Alcohol and the Heart

increasing your blood pressure and risks to your heart...

https://www.drinkaware.co.uk/alcohol-facts/health-effects-of-alcohol/diseases/alcohol-and-heart-disease/

Causality and correlation: An example

Does excessive alcohol consumption cause heart disease?



Prev Med. 2004 May;38(5):613-9.

A meta-analysis of alcohol consumption and the risk of 15 diseases.

Corrao G[1], Bagnardi V, Zambon A, La Vecchia C.

⊕ **Author information**

Abstract

BACKGROUND: To compare the strength of evidence provided by the epidemiological literature on the association between alcohol consumption and the risk of 14 major alcohol-related neoplasms and non-neoplastic diseases, plus injuries.

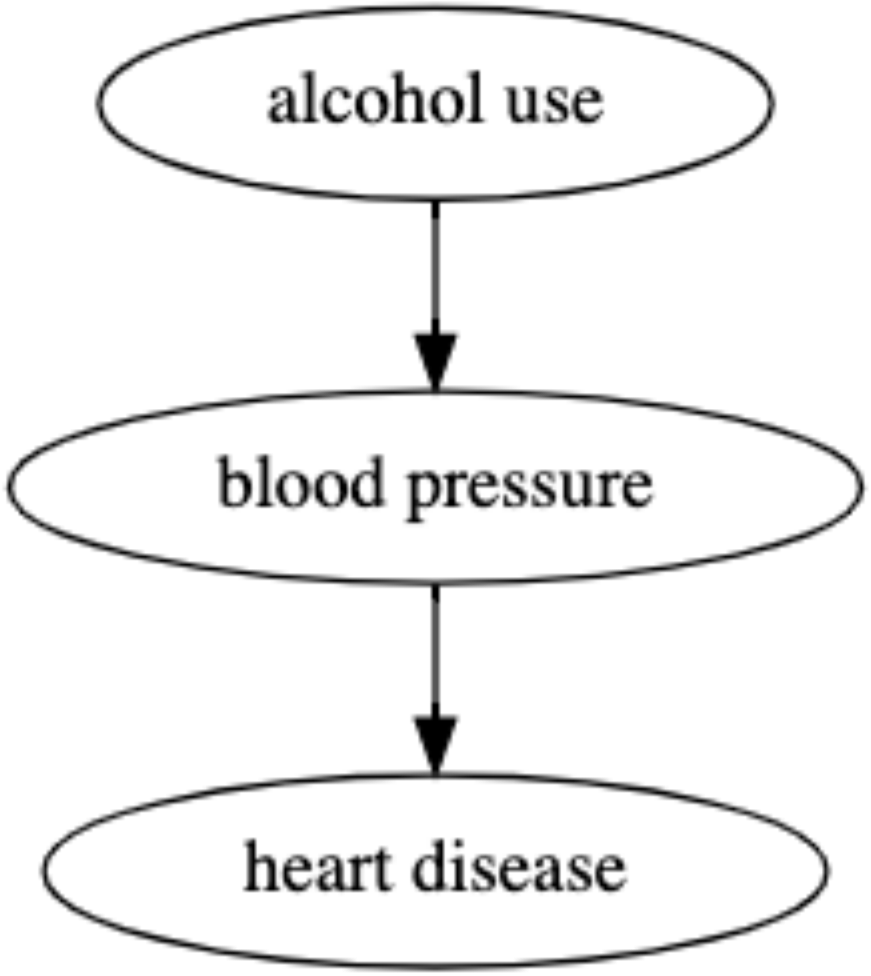
METHODS: A search of the epidemiological literature from 1966 to 1998 was performed by several bibliographic databases. Meta-regression models were fitted considering fixed and random effect models and linear and nonlinear effects of alcohol intake. The effects of some characteristics of the studies, including an index of their quality, were considered.

RESULTS: Of the 561 initially reviewed studies, 156 were selected for meta-analysis because of their a priori defined higher quality, including a total of 116,702 subjects. Strong trends in risk were observed for cancers of the oral cavity, esophagus and larynx, hypertension, liver cirrhosis, chronic pancreatitis, and injuries and violence. Less strong direct relations were observed for cancers of the colon, rectum, liver, and breast. For all these conditions, significant increased risks were also found for ethanol intake of 25 g per day. Threshold values were observed for ischemic and hemorrhagic strokes. For coronary heart disease, a J-shaped relation was observed with a minimum relative risk of 0.80 at 20 g/day, a significant protective effect up to 72 g/day, and a significant increased risk at 89 g/day. No clear relation was observed for gastroduodenal ulcer.

CONCLUSIONS: This meta-analysis shows no evidence of a threshold effect for both neoplasms and several non-neoplastic diseases. J-shaped relations were observed only for coronary heart disease.

Graphical causal models

A way to visualize our assumptions about the causal structure that relates a set of variables



- Nodes (ellipses) refer to variables
- Arrows refer to causal influences relating the variables

This is a directed acyclic graph (DAG)

- *directed*: all edges are directed
- *acyclic*: a directed path never leads back to the same variable where it started

We will focus on DAGs in this lecture

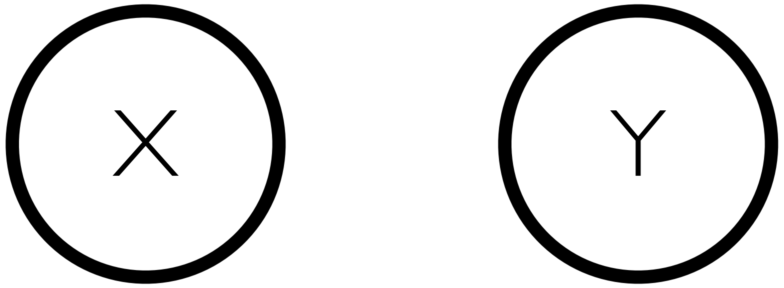
Causal vs. statistical language

"Correlation is not causation, but it sure is a hint."
— Edward Tufte

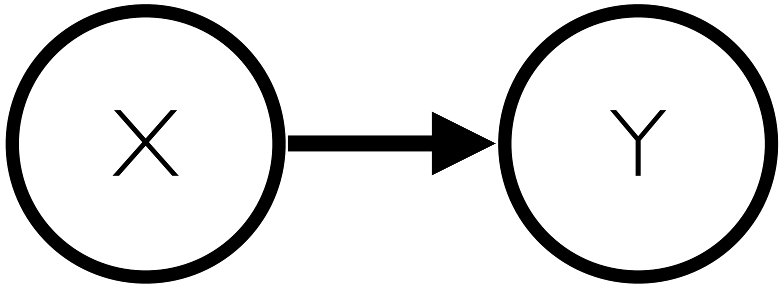
If we were to observe a correlation between two variables x and y, what does this imply regarding causation?

Causal implication of statistical association: 5 possibilities

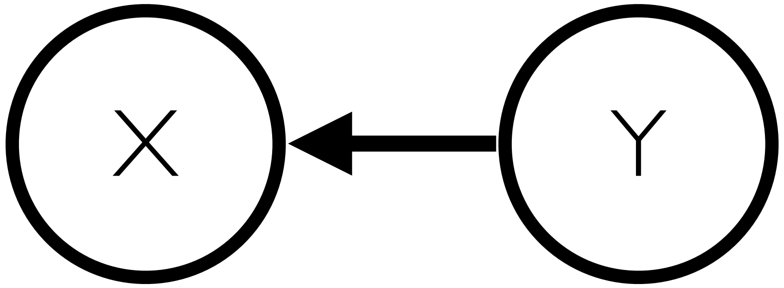
There are no causal relations between X and Y; the association arose by random fluctuation



X causes Y

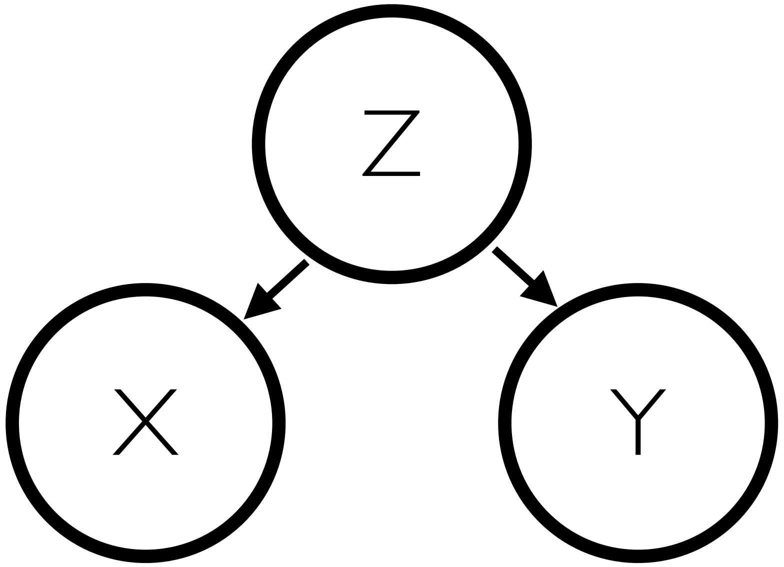


Y causes X

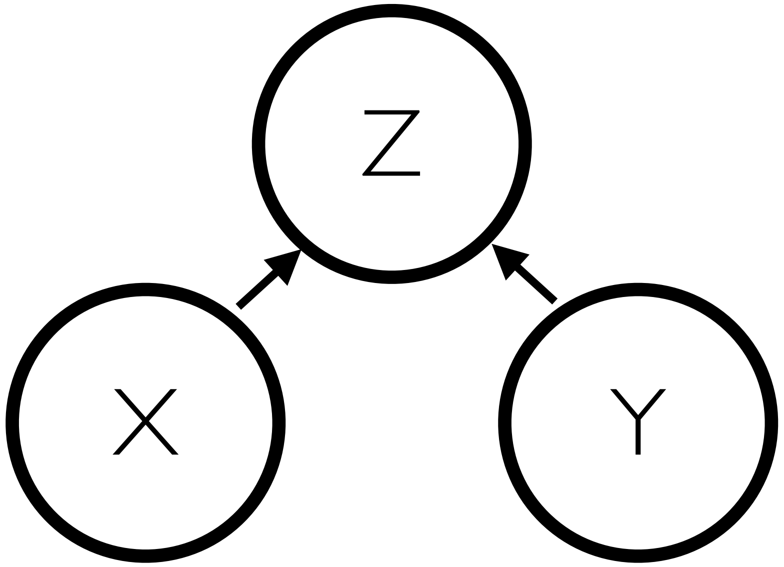


(maybe via an indirect path)

The association is confounded - that is, a third factor Z causes both X and Y



The association is spurious, due to conditioning on a common effect of X and Y



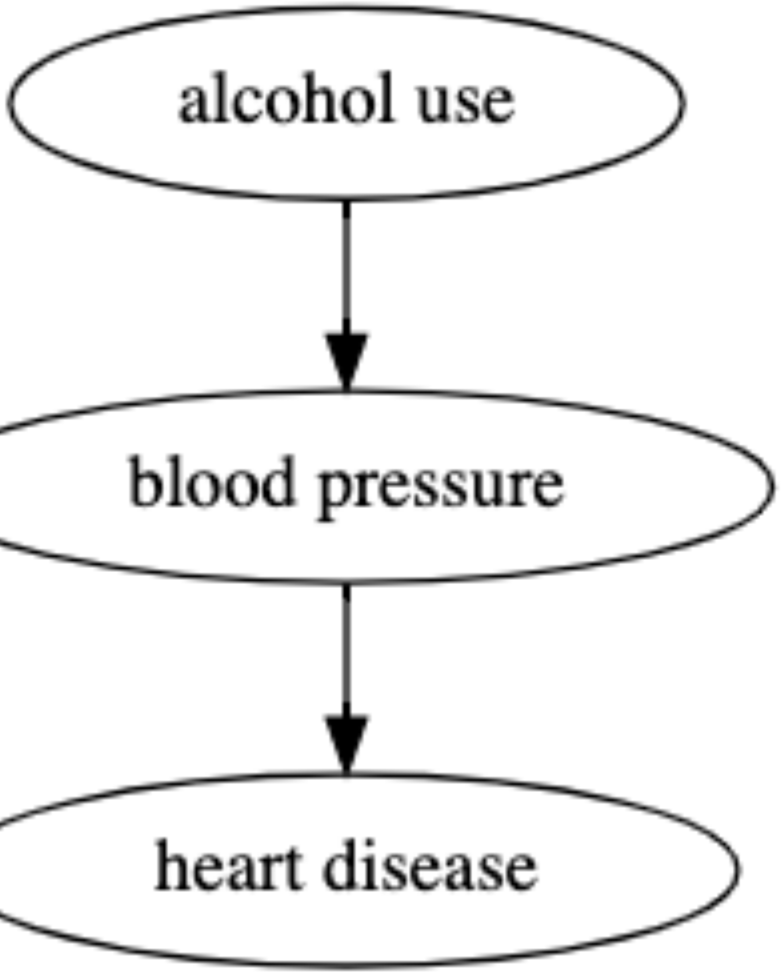
From statistical to causal language

Fundamental insight: The structure of a causal graph has implications for the statistical relations between the variables

Pearl's "do-calculus":

Provides the ability to go from statements about interventions ("do(X =x)") to statements about the probability distributions of observed data (with no intervention)

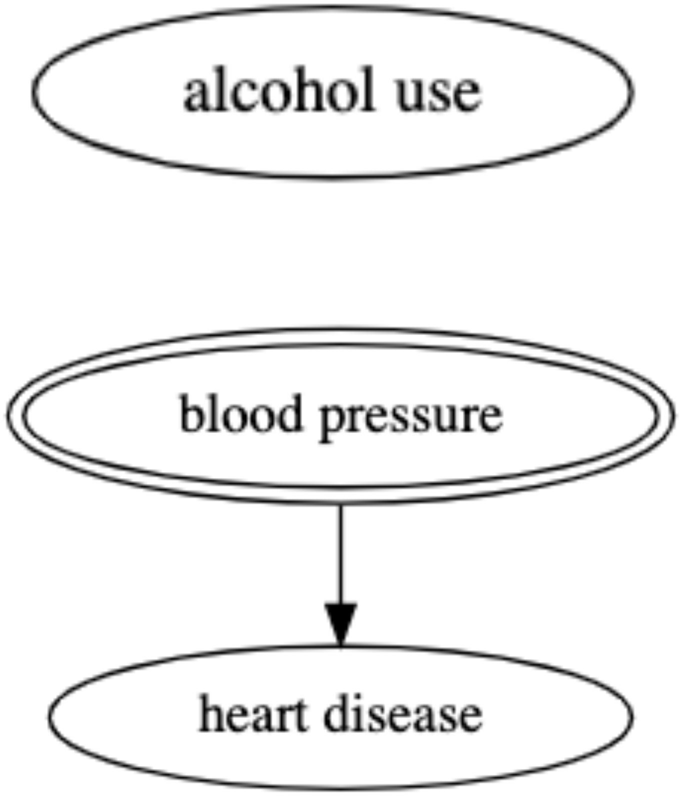
From statistical to causal language



Changing alcohol use has an effect on heart disease, via Y

r(alc, BP), r(BP, HD), and r(alc, HD) all positive

From statistical to causal language

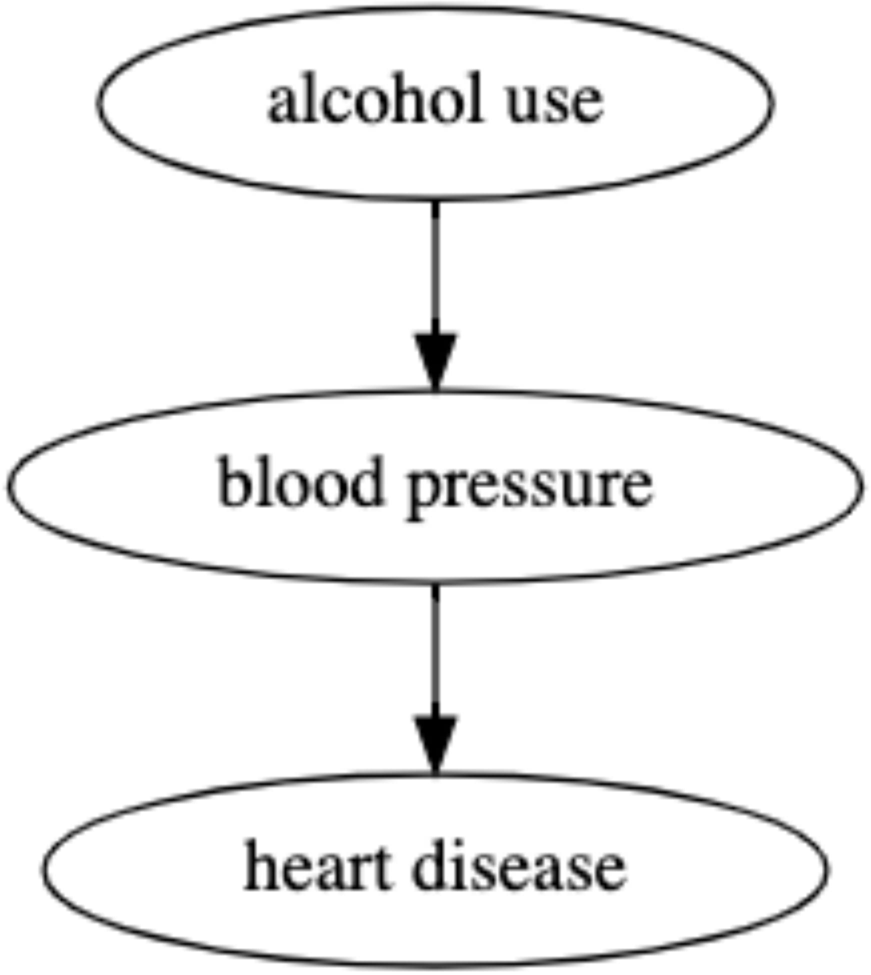


do(BP=120)

Holding BP constant, alcohol use no longer has an effect on heart disease

- i.e. we can change alcohol use but it will not change heart disease, once BP is fixed

From statistical to causal language



lm(HD ~ alc + BP)

After conditioning on BP, there is no statistical relation between alcohol use and heart disease

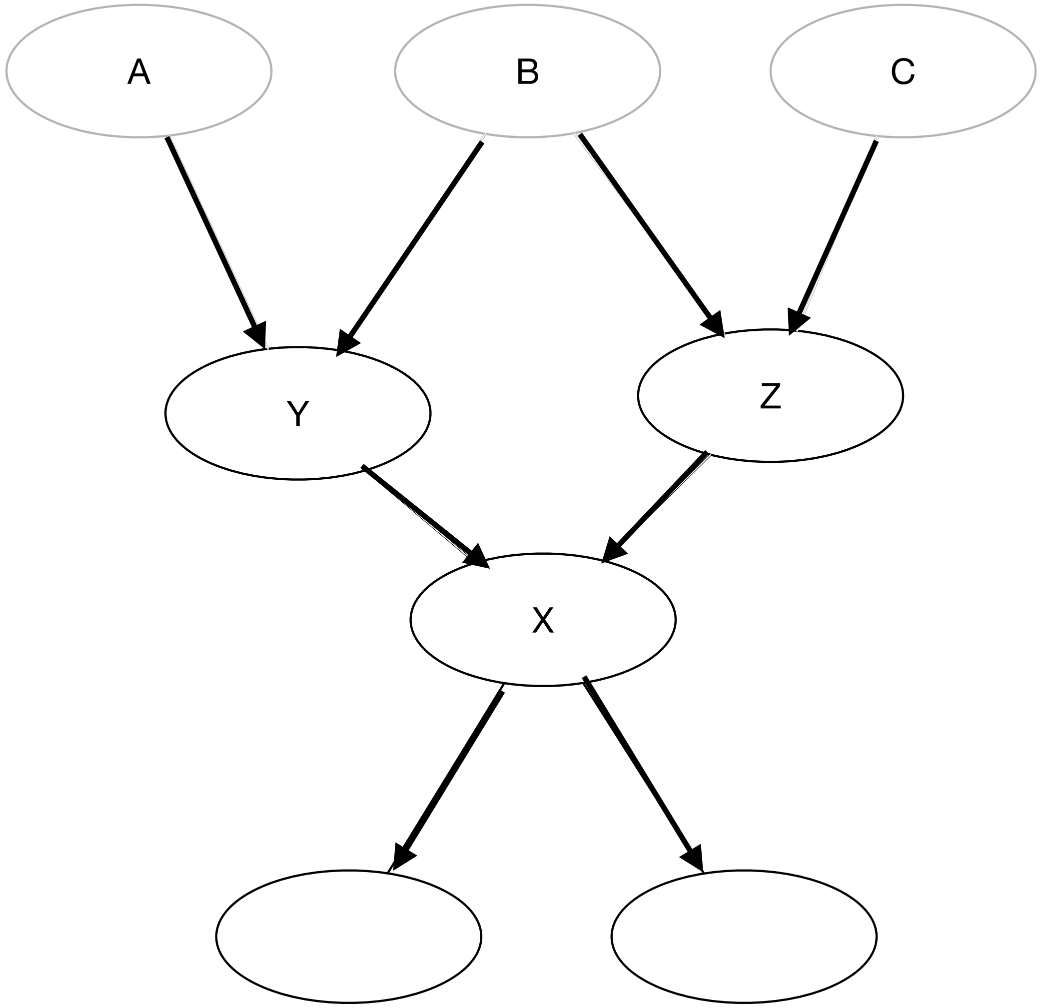
they are "conditional independent" given BP

$$\rho(alc, HD|BP) = 0$$

(ρ: "partial correlation")

Assumptions of graphical causal modeling approach

- Causal Markov condition
  - Every variable in a DAG is independent of all of its non-descendents, conditional on its parents.
- This implies that there are no other variables outside of the graph!



$X \perp \{A, B, C\} \mid \{Y, Z\}$

Glymour, 2006

Assumptions of graphical causal modeling approach

- Faithfulness assumption:
  - The statistical relationships in the data faithfully reflect the causal structure. In particular, positive and negative causal effects do not perfectly cancel each other out.
- Negligible randomness:
  - We assume that any observed statistical relationships did not arise due to chance.

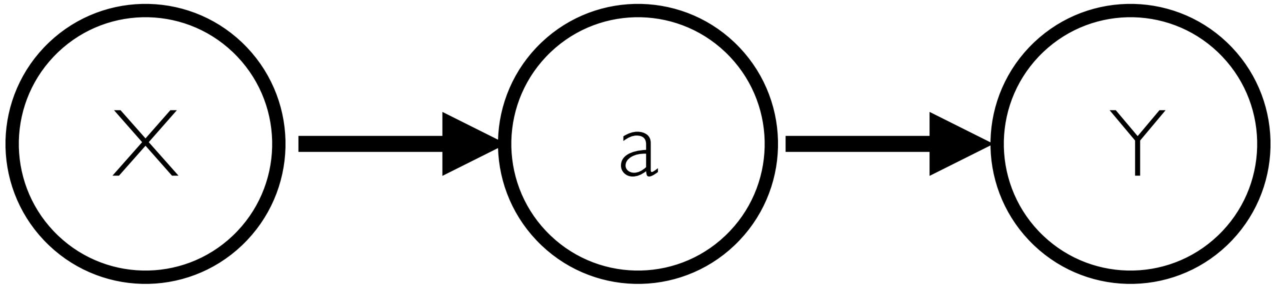
Glymour, 2006

Inferring conditional independencies using d-separation

Two variables X and Y are d-separated by another set of variables Z if and only if Z blocks every path from X to Y

This can happen two ways:

1. Every path from X to Y includes a non-collider that is in Z



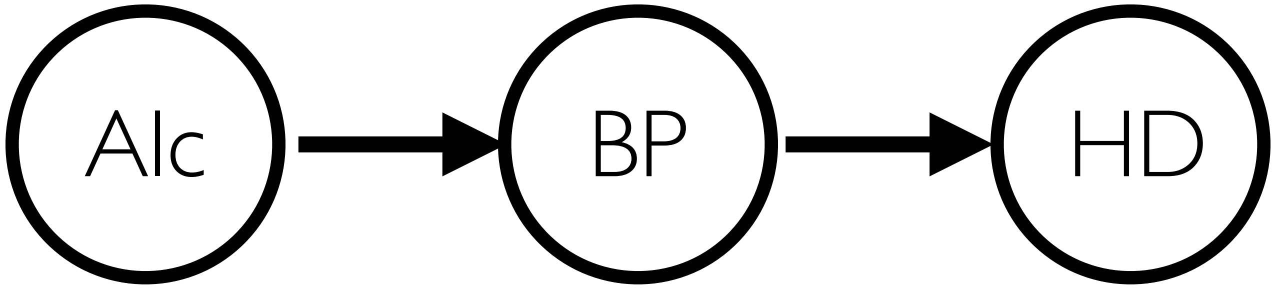
Z: {a}

Inferring conditional independencies using d-separation

Two variables X and Y are d-separated by another set of variables Z if and only if Z blocks every path from X to Y

This can happen two ways:

1. Every path from X to Y includes a non-collider that is in Z



Z: {BP}

Alcohol use and heart disease are d-separated by blood pressure

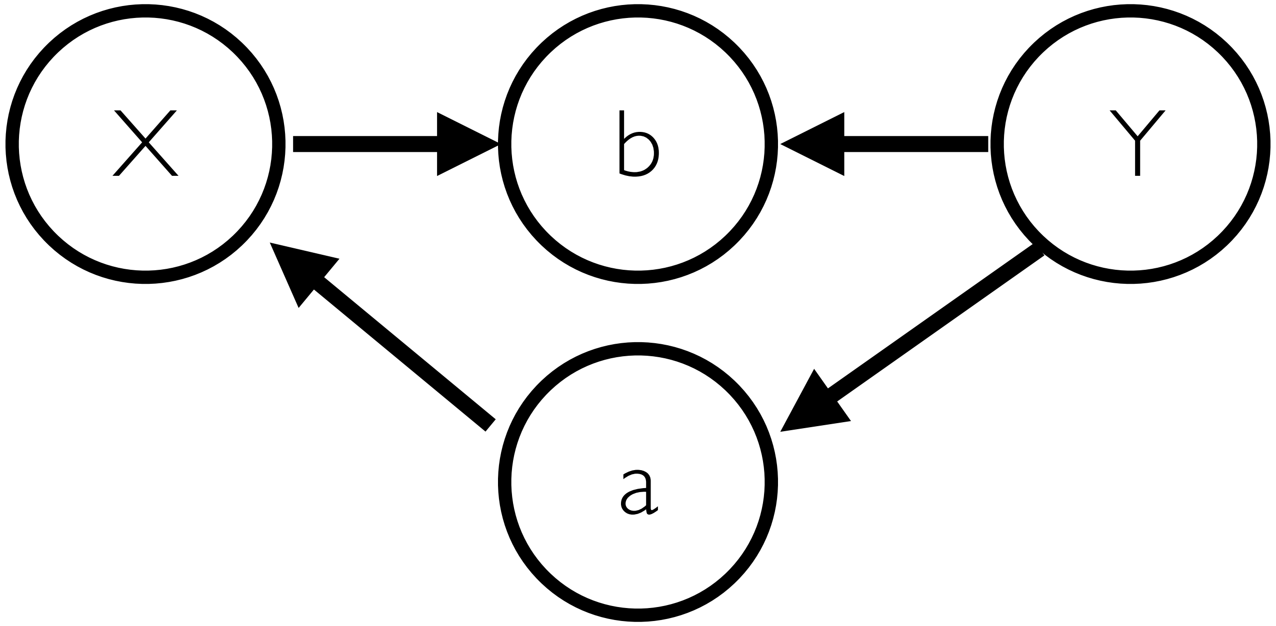
Inferring conditional independencies using d-separation

Two variables X and Y are d-separated by another set of variables Z if and only if Z blocks every path from X to Y

This can happen two ways:

2. There is a collider on the path, but neither the collider nor any of its descendants is in Z

b is a "collider"



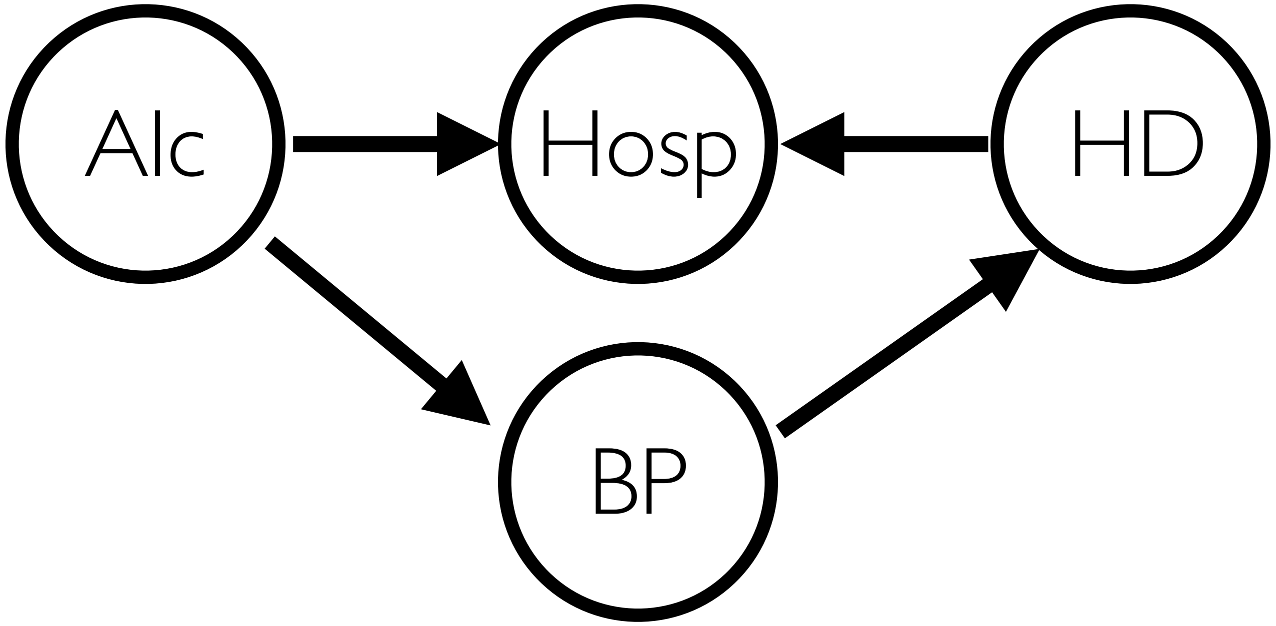
Z: {a}

Inferring conditional independencies using d-separation

Two variables X and Y are d-separated by another set of variables Z if and only if Z blocks every path from X to Y

This can happen two ways:

2. There is a collider on the path, but neither the collider nor any of its descendants is in Z

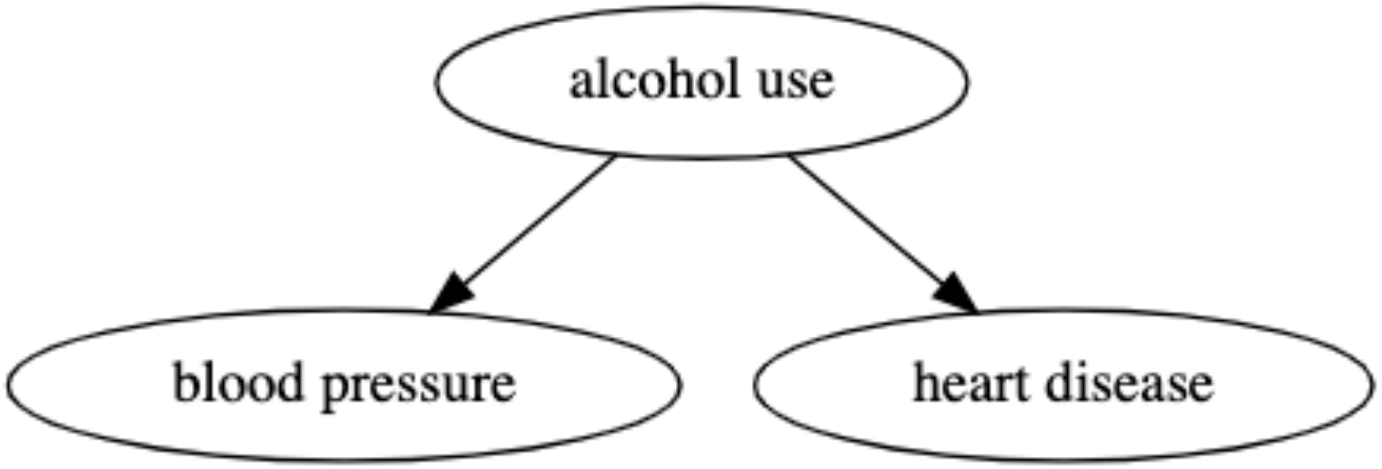


Z: {BP}

Inferring conditional independencies using d-separation

The d-separation algorithm allows us to determine the set of conditional independencies implied by any graph

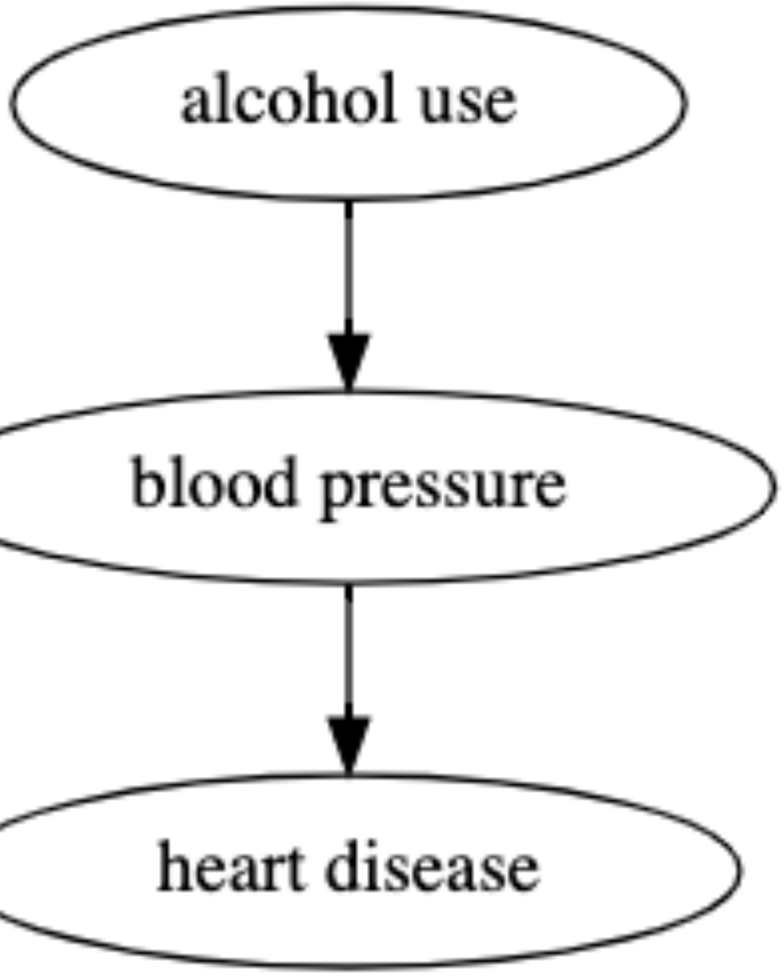
“fork”



```
Implied conditional Independence Relationship:
[('blood pressure', 'heart disease', {‘alcohol use'})]
```

Inferring conditional independencies using d-separation

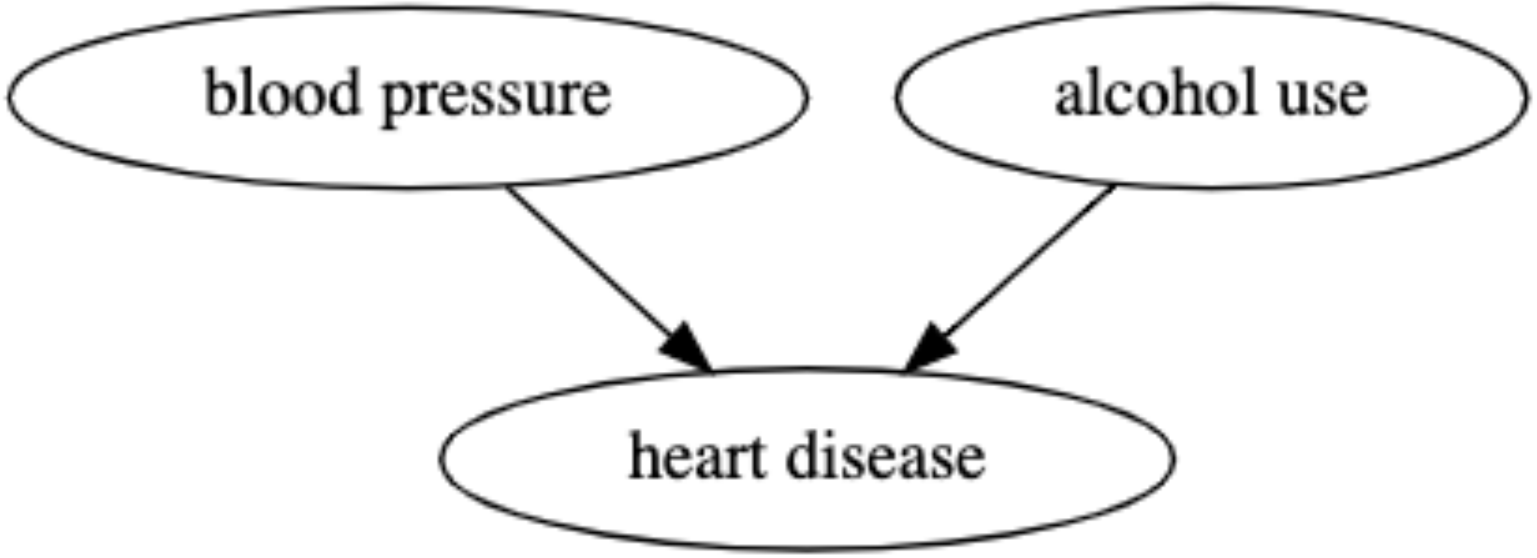
“chain”



```
Implied conditional Independence Relationship:
[('heart disease', 'alcohol use', {'blood pressure'})]
```

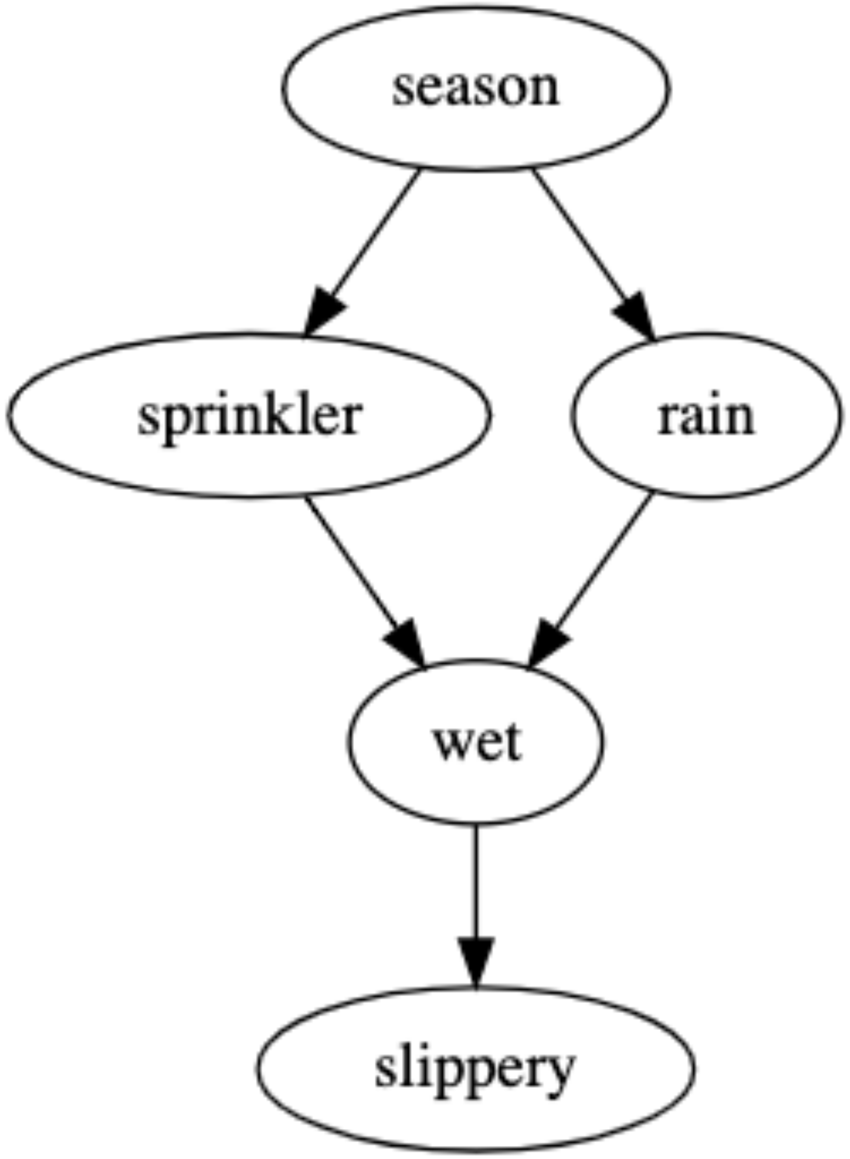
Inferring conditional independencies using d-separation

“collider”



```
Implied conditional Independence Relationship:
[('blood pressure', 'alcohol use', set())]
```

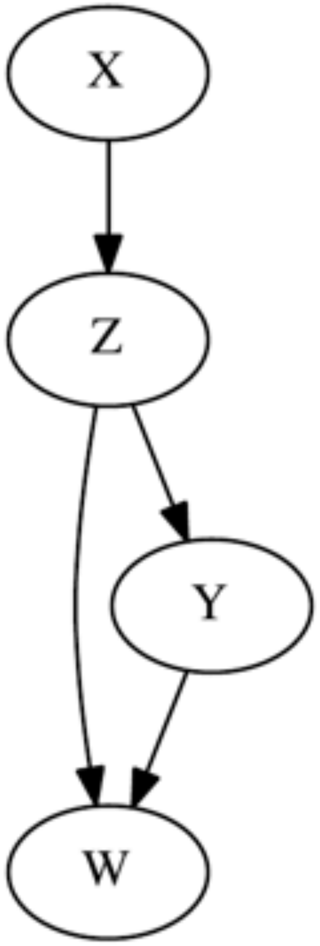
A more complex example

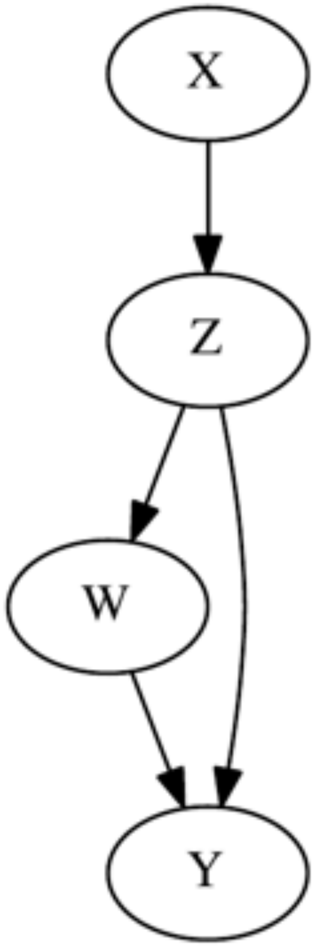


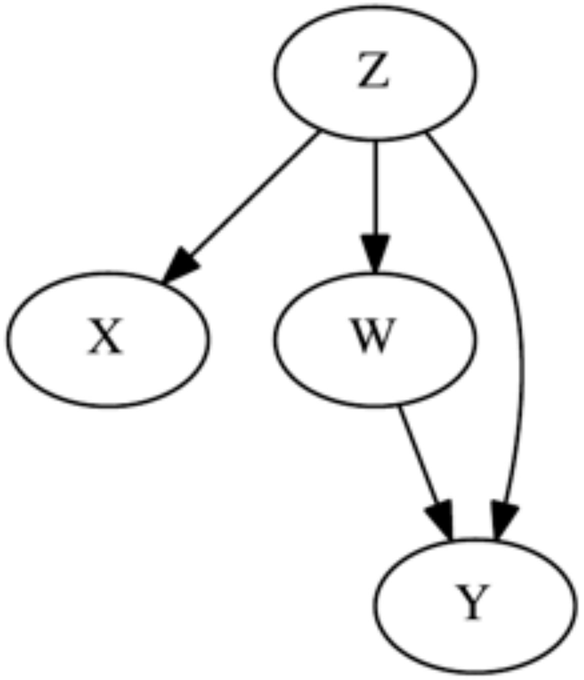
```
sprinkler.get_all_independence_relationships()

[(‘wet', 'season', {'rain', 'sprinkler'}),
 ('wet', 'season', {'rain', 'slippery', 'sprinkler'}),
 ('sprinkler', 'slippery', {'wet'}),
 ('sprinkler', 'slippery', {'rain', 'wet'}),
 ('sprinkler', 'slippery', {'season', 'wet'}),
 ('sprinkler', 'slippery', {'rain', 'season', 'wet'}),
 ('sprinkler', 'rain', {'season'}),
 ('slippery', 'season', {'wet'}),
 ('slippery', 'season', {'rain', 'sprinkler'}),
 ('slippery', 'season', {'rain', 'wet'}),
 ('slippery', 'season', {'sprinkler', 'wet'}),
 ('slippery', 'season', {'rain', 'sprinkler', 'wet'}),
 ('slippery', 'rain', {'wet'}),
 ('slippery', 'rain', {'sprinkler', 'wet'}),
 ('slippery', 'rain', {'season', 'wet'}),
 ('slippery', 'rain', {'season', 'sprinkler', 'wet'})]
```

Markov equivalence

Different graphs (particularly, differently oriented edges) can have equivalent implied conditional independences



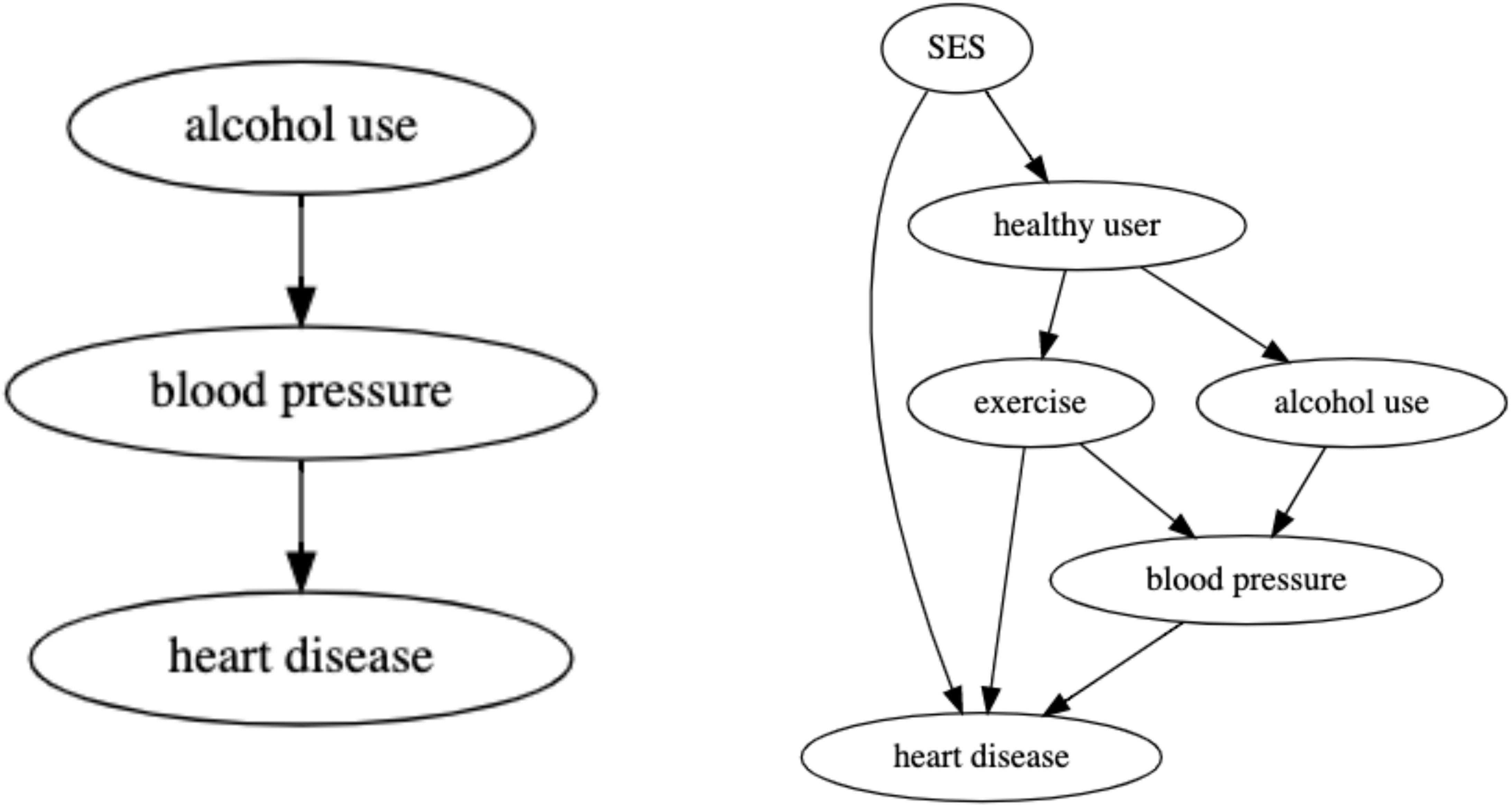




```
('X', 'W', {'Z'}),
('X', 'W', {'Z', 'Y'}),
('X', 'Y', {'Z'}),
('X', 'Y', {'Z', 'W'})
```

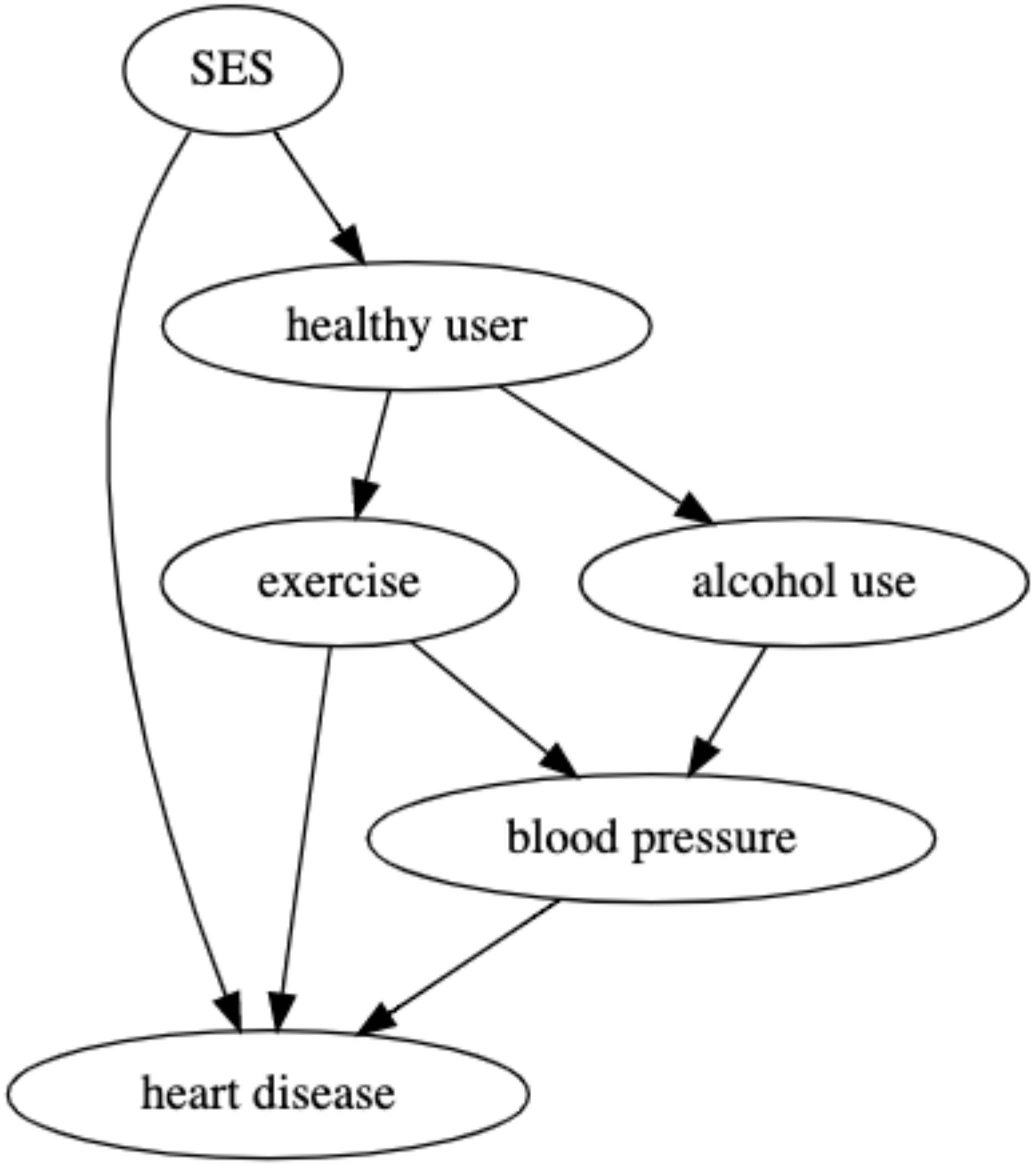
Unobserved causes and confounding

A correct graphical model must include both observed and unobserved causes that are known to exist - otherwise the assumptions fail due to possible confounding



Back-door paths

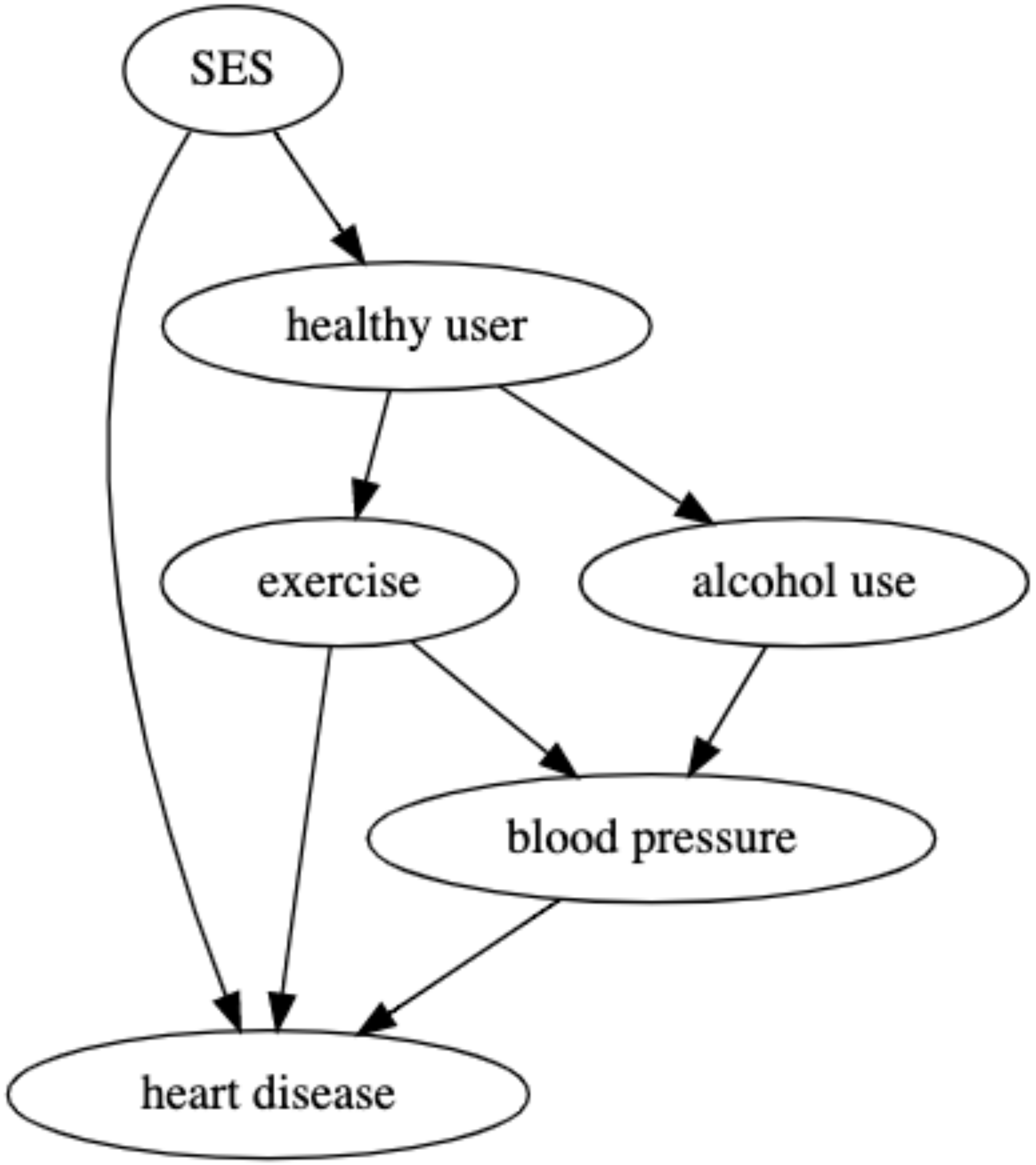
Graphical modeling tools can help identify potential pathways for confounding (known as *back-door paths*)



```
hd2.get_all_backdoor_paths(
'alcohol use', 'heart disease')

 [['alcohol use',
   'healthy user',
   'exercise',
   'blood pressure',
   'heart disease'],
  ['alcohol use', 'healthy user',
    'exercise', 'heart disease'],
  ['alcohol use', 'healthy user',
    ‘SES', 'heart disease']]
```

Back-door paths

They can also tell us what we would need to control for in order to eliminate potential confounds



```
hd2.get_all_backdoor_adjustment_sets(
    ‘alcohol use’, 'heart disease')
```

```
({set({'exercise', 'healthy user'}),
        set({'SES', 'healthy user'}),
        set({'SES', 'exercise'}),
        set({'healthy user'}),
        set({'SES', 'exercise',
             'healthy user'})})
```

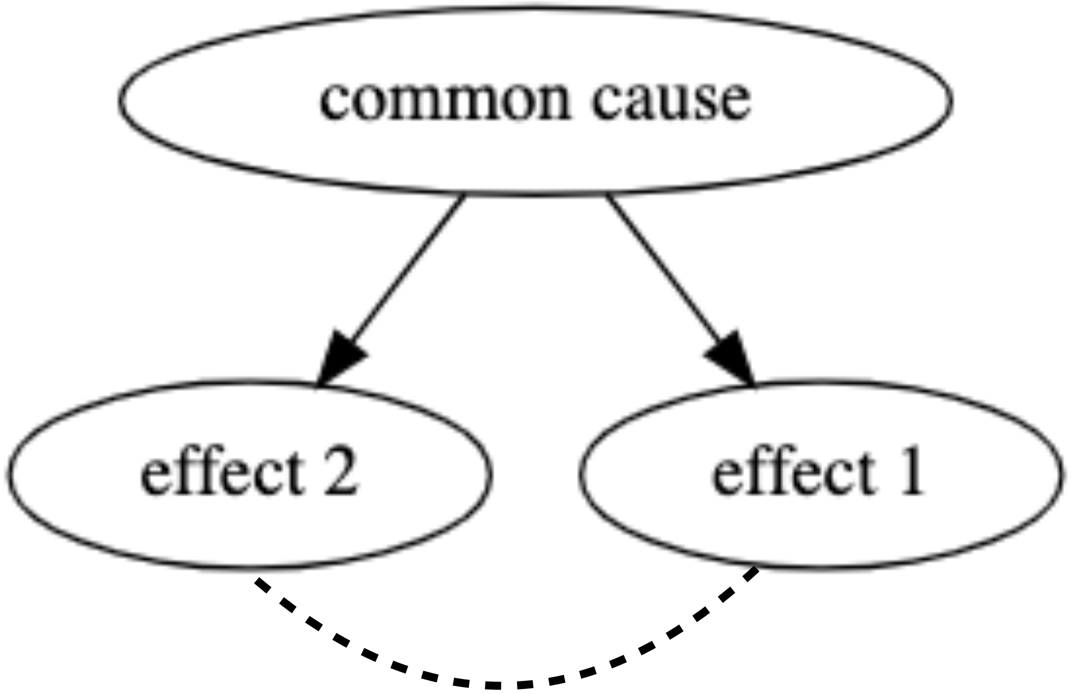
Psych 253
Advanced Statistical Modeling

Berkson's paradox and causal models

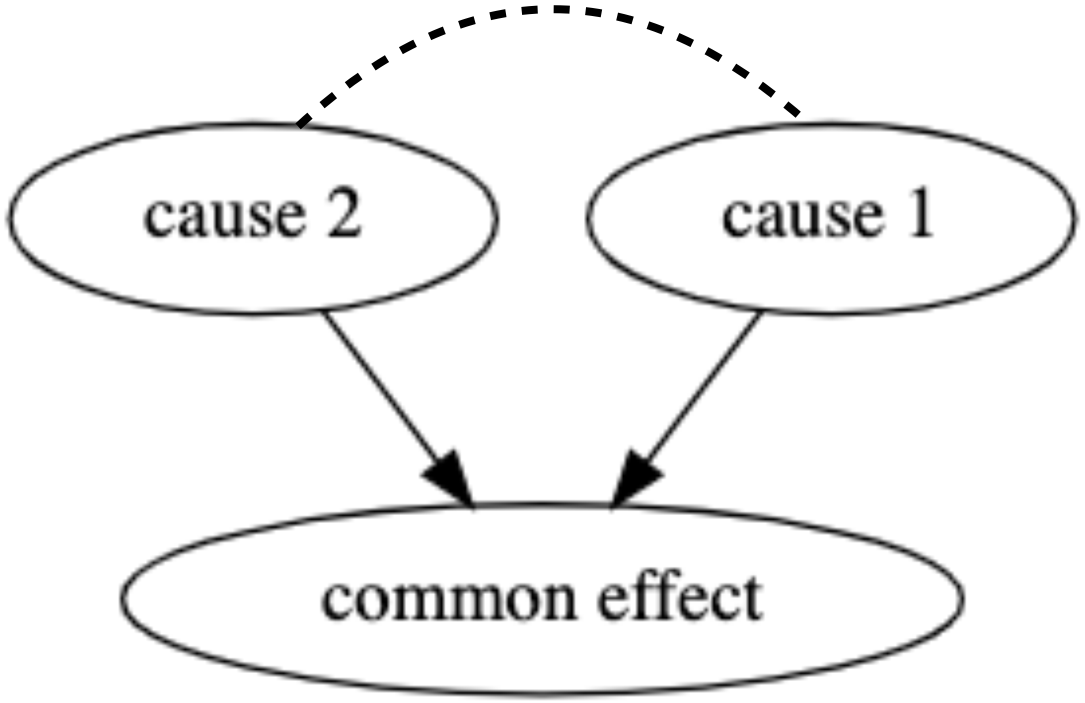
Daniel Yamins
Wu Tsai Neurosciences Institute
Departments of Psychology and Computer Science
Stanford University

Russ Poldrack
Department of Psychology
Stanford University

Conditioning on common causes vs. common effects



Failing to condition on
a common cause results
in a spurious association
between effects



Conditioning on
a common effect results
in a spurious association
between causes

https://catalogofbias.org/biases/collider-bias/

Conditioning on a collider: An example

- 100 people attend a meeting
  - 10 were infected with flu (but pre-symptomatic)
  - Lunch is randomly assigned
    - 50 eat chicken
    - 50 eat egg salad
- What is the risk of flu?

	flu	no flu	risk
chicken	5	45	0.1
egg salad	5	45	0.1

Cole et al., 2010

Conditioning on a collider

- It turns out that the egg salad was tainted, and everyone who ate it gets a fever
  - Assume that flu and food poisoning are the only way get a fever
- Let's look at the flu risk depending on whether or not one has a fever

Fever

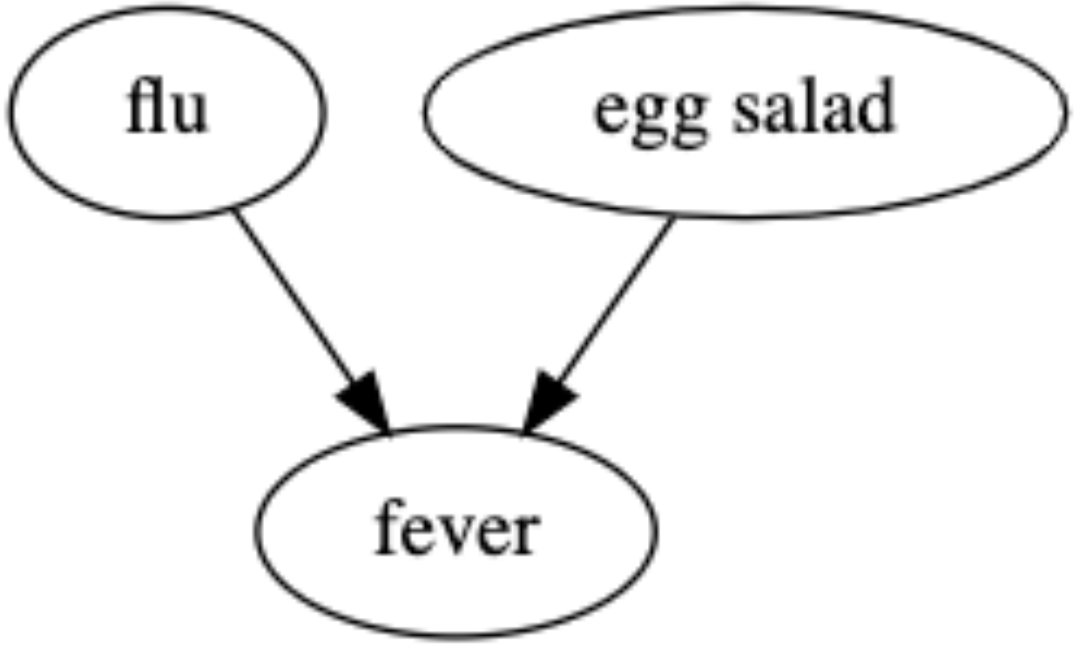
	flu	no flu	risk
chicken	5	0	1.0
egg salad	5	45	0.1

No fever

	flu	no flu	risk
chicken	0	45	0
egg salad	0	0	NA

If we only look at people who have a fever, it would appear that eating egg salad was protective from the flu!

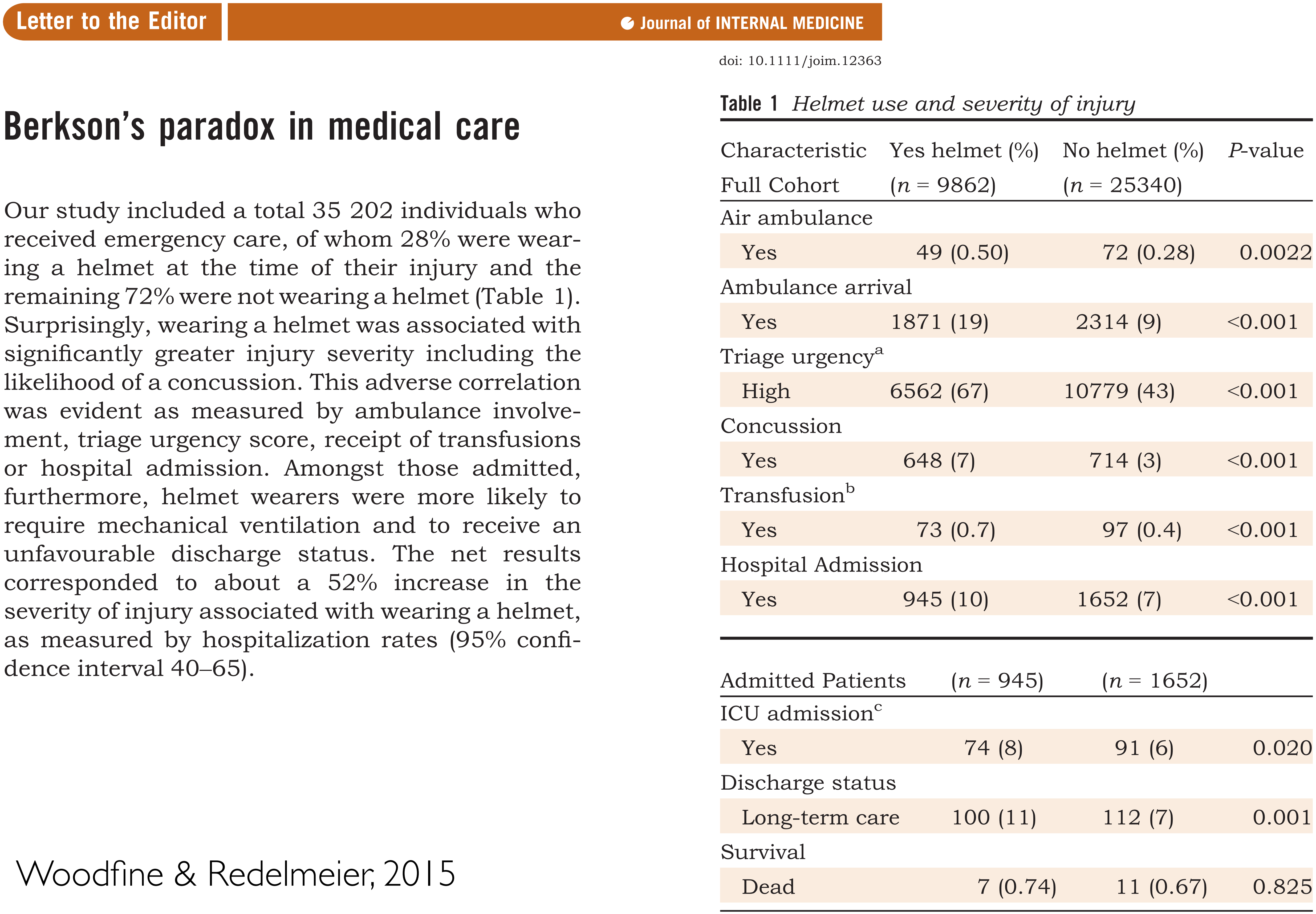
Cole et al., 2010



Conditioning on a collider (selecting based on fever in this case) induces a spurious association between originally independent variables

```
Implied conditional Independence Relationship:
[(‘flu', 'egg salad', set())]
```

Berkson’s paradox



Letter to the Editor

Journal of INTERNAL MEDICINE

doi: 10.1111/joim.12363

Berkson’s paradox in medical care

Our study included a total 35 202 individuals who received emergency care, of whom 28% were wearing a helmet at the time of their injury and the remaining 72% were not wearing a helmet (Table 1). Surprisingly, wearing a helmet was associated with significantly greater injury severity including the likelihood of a concussion. This adverse correlation was evident as measured by ambulance involvement, triage urgency score, receipt of transfusions or hospital admission. Amongst those admitted, furthermore, helmet wearers were more likely to require mechanical ventilation and to receive an unfavourable discharge status. The net results corresponded to about a 52% increase in the severity of injury associated with wearing a helmet, as measured by hospitalization rates (95% confidence interval 40–65).

Table 1 *Helmet use and severity of injury*

Characteristic	Yes helmet (%)	No helmet (%)	*P*-value
Full Cohort	(*n* = 9862)	(*n* = 25340)	
Air ambulance			
Yes	49 (0.50)	72 (0.28)	0.0022
Ambulance arrival			
Yes	1871 (19)	2314 (9)	<0.001
Triage urgency[a]			
High	6562 (67)	10779 (43)	<0.001
Concussion			
Yes	648 (7)	714 (3)	<0.001
Transfusion[b]			
Yes	73 (0.7)	97 (0.4)	<0.001
Hospital Admission			
Yes	945 (10)	1652 (7)	<0.001
Admitted Patients	(*n* = 945)	(*n* = 1652)	
ICU admission[c]			
Yes	74 (8)	91 (6)	0.020
Discharge status			
Long-term care	100 (11)	112 (7)	0.001
Survival			
Dead	7 (0.74)	11 (0.67)	0.825

Woodfine & Redelmeier, 2015

Berkson’s paradox

OBSERVED CRASHES

		Helmet	
		Yes	No
Hospitalized	Yes	945	1652
	No	8917	23 688

Relative risk = 1.52
Confidence interval: 1.40 to 1.65

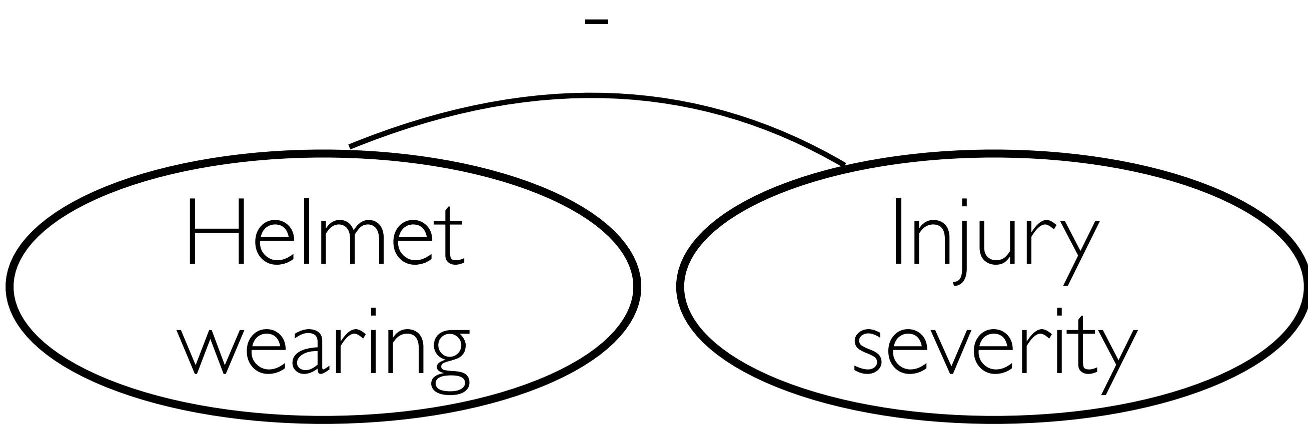
UNOBSERVED CRASHES

		Helmet	
		Yes	No
Hospitalized	Yes	0	0
	No	8917	0

TOTAL CRASHES

		Helmet	
		Yes	No
Hospitalized	Yes	945	1652
	No	17 834	23 688

Relative risk = 0.76
Confidence interval: 0.70 to 0.83


-

Helmet wearing

Injury severity

Berkson’s paradox

OBSERVED CRASHES

		Helmet	
		Yes	No
Hospitalized	Yes	945	1652
	No	8917	23 688

Relative risk = 1.52
Confidence interval: 1.40 to 1.65

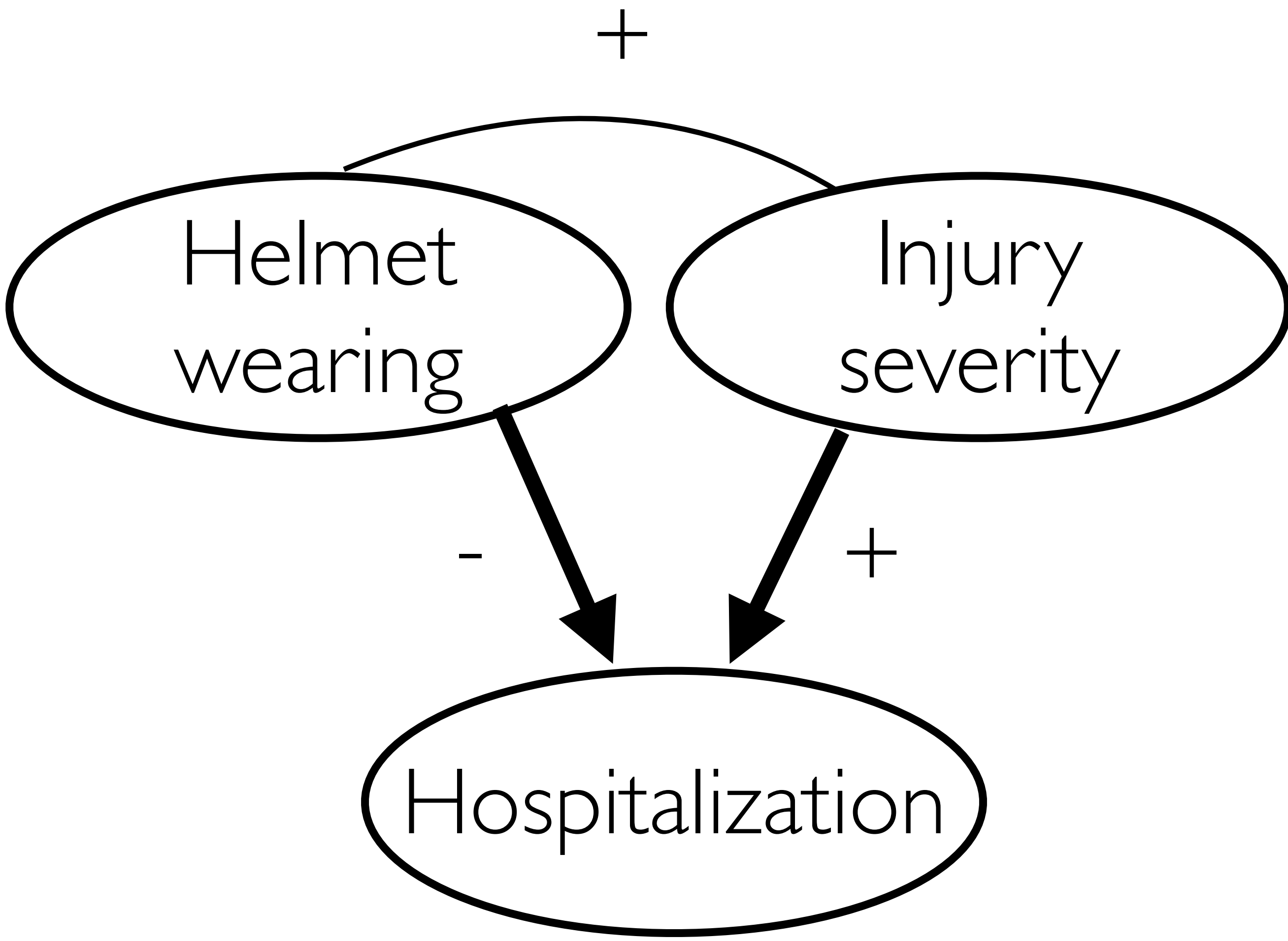
UNOBSERVED CRASHES

		Helmet	
		Yes	No
Hospitalized	Yes	0	0
	No	8917	0

TOTAL CRASHES

		Helmet	
		Yes	No
Hospitalized	Yes	945	1652
	No	17 834	23 688

Relative risk = 0.76
Confidence interval: 0.70 to 0.83



Implications of Berkson's paradox

One must be very careful about including covariates in a regression analysis

- Including a common effect of X and Y as a covariate can induce a spurious association between X and Y

Selection bias can induce spurious associations

- if the selection variable is a common effect of other variables

Regression versus causal modeling

There is no mathematical reason that one can't simply flip the x and y variables in regression

$$y_i = \alpha_y + \beta_{yx} * x_i + \epsilon_{yi}$$

Can be rewritten as:

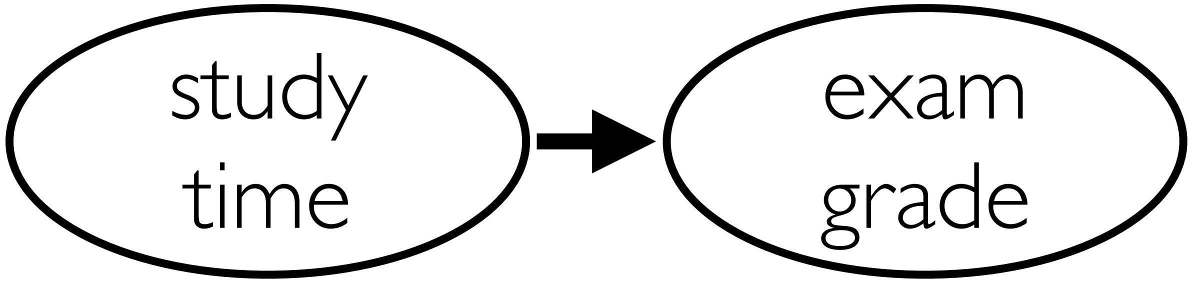
$$x_i = \alpha_x + \beta_{xy} * y_i + \epsilon_{xi}$$

$$\alpha_x = \frac{-\beta_{yx}}{\alpha_y}, \ \beta_{xy} = \frac{1}{\beta_{yx}}, \quad \epsilon_{xi} = \frac{\epsilon_{yi}}{-\beta_{yx}}$$

Regression versus causal modeling

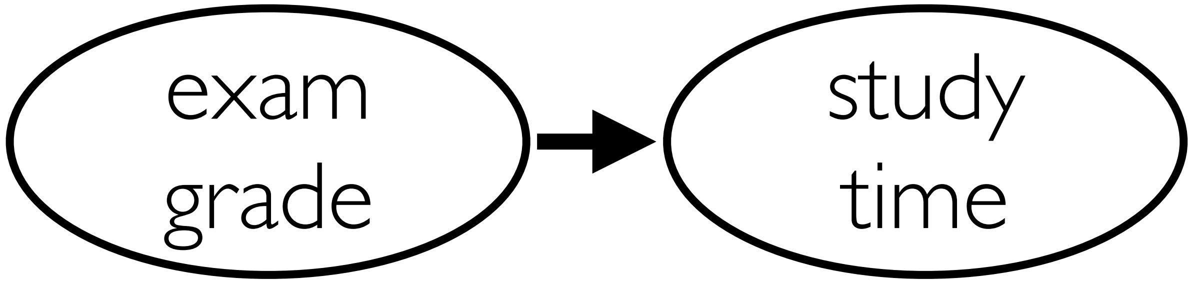
But this often doesn't make sense with regard to causality. For example, this regression makes sense to us as a causal graph:

$$\text{exam grade} = \beta_0 + \text{study time} * \beta_1$$



Whereas this one, while mathematically fine, doesn't make sense as a causal graph:

$$\text{study time} = \beta_0 + \text{exam grade} * \beta_1$$



Thus, a causal model involves a set of equations along with a set of causal assumptions

SEM as a causal model

Whereas graphical models generally do not make any particular claims about the mathematical structure of the model, SEM is framed in terms of a set of linear models.

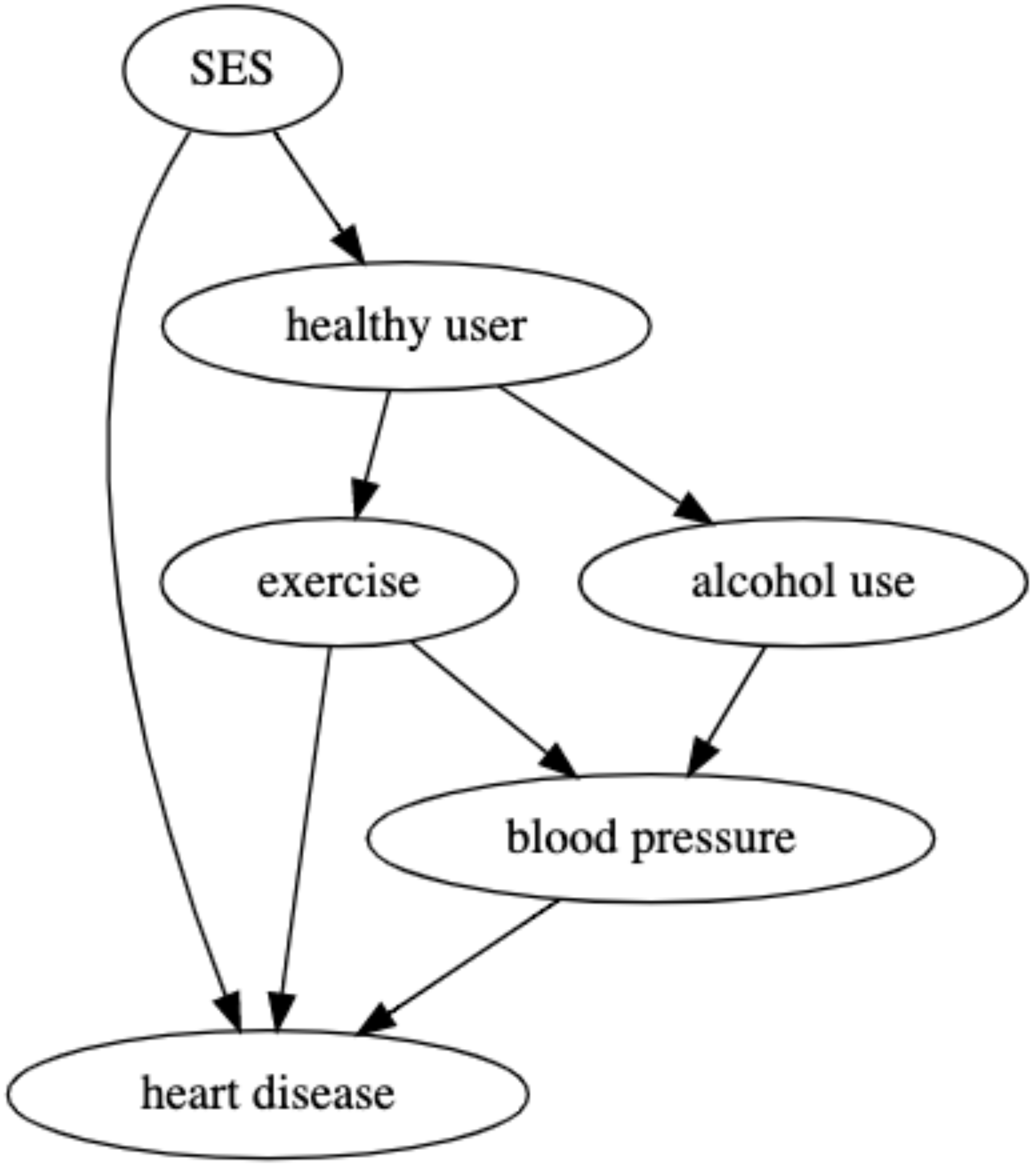
That is, graphical models are *nonparametric* whereas SEMs are a particular *parametric* instantiation of those models.

Pearl (2012) lays out the logic of structural equation modeling (SEM) in the context of graphical causal models.

SEM as a causal model

The inputs to the model are:

1. A set $\boldsymbol{A}$ of causal assumptions (justified on scientific grounds) that are encoded in a model $\boldsymbol{M}$ expressed as a directed graph.

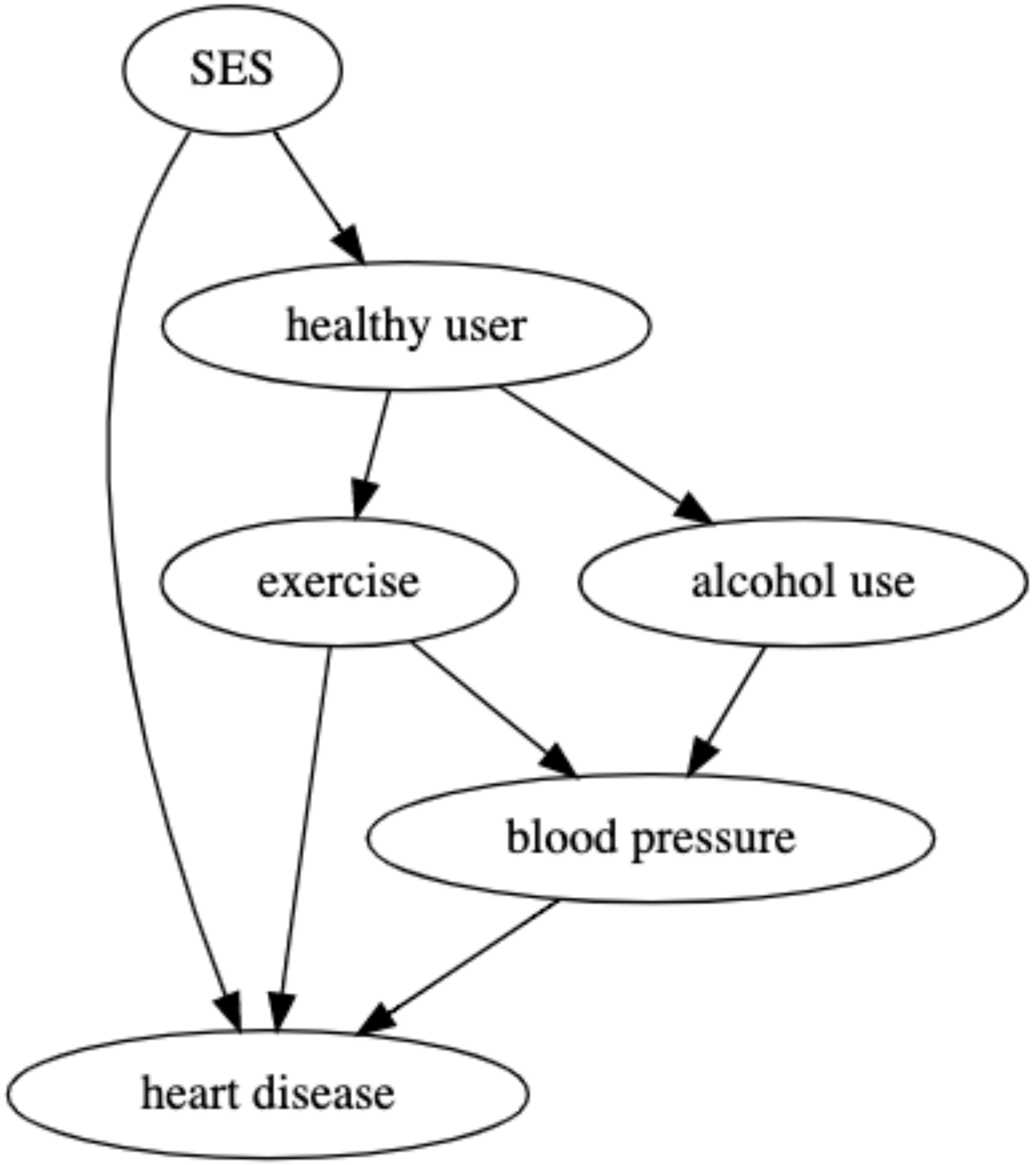


SEM as a causal model

The inputs to the model are:

2. A set $\boldsymbol{Q}$ of questions concerning causal relations between the variables in the model ---

   e.g. "What is the effect of reducing alcohol use on heart disease?’’

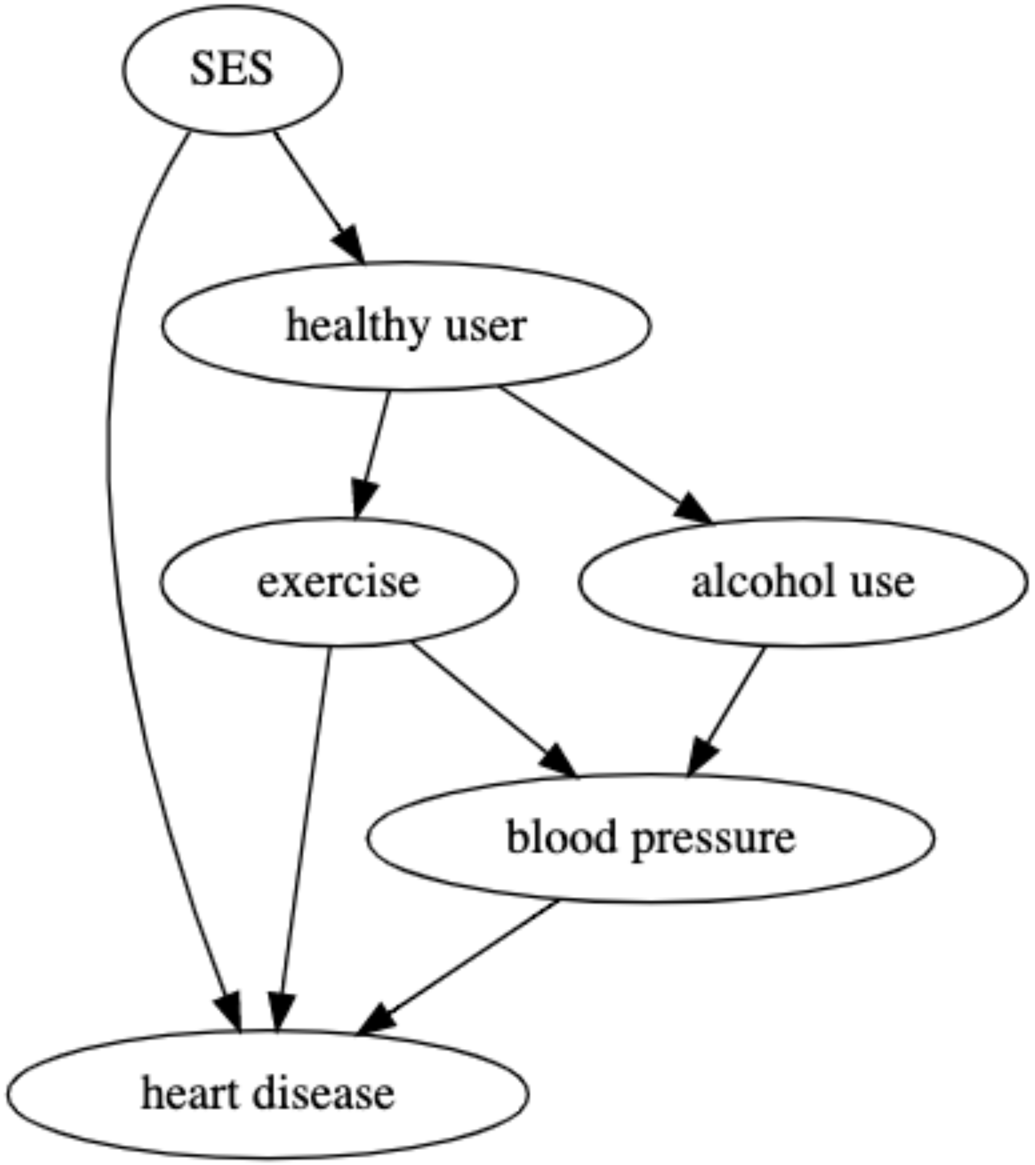


SEM as a causal model

The inputs to the model are:

3. A set ***D*** of data (experimental or observational) presumably generated by a process consistent with assumptions ***A***

- e.g. results from a randomized controlled trial to reduce alcohol consumption

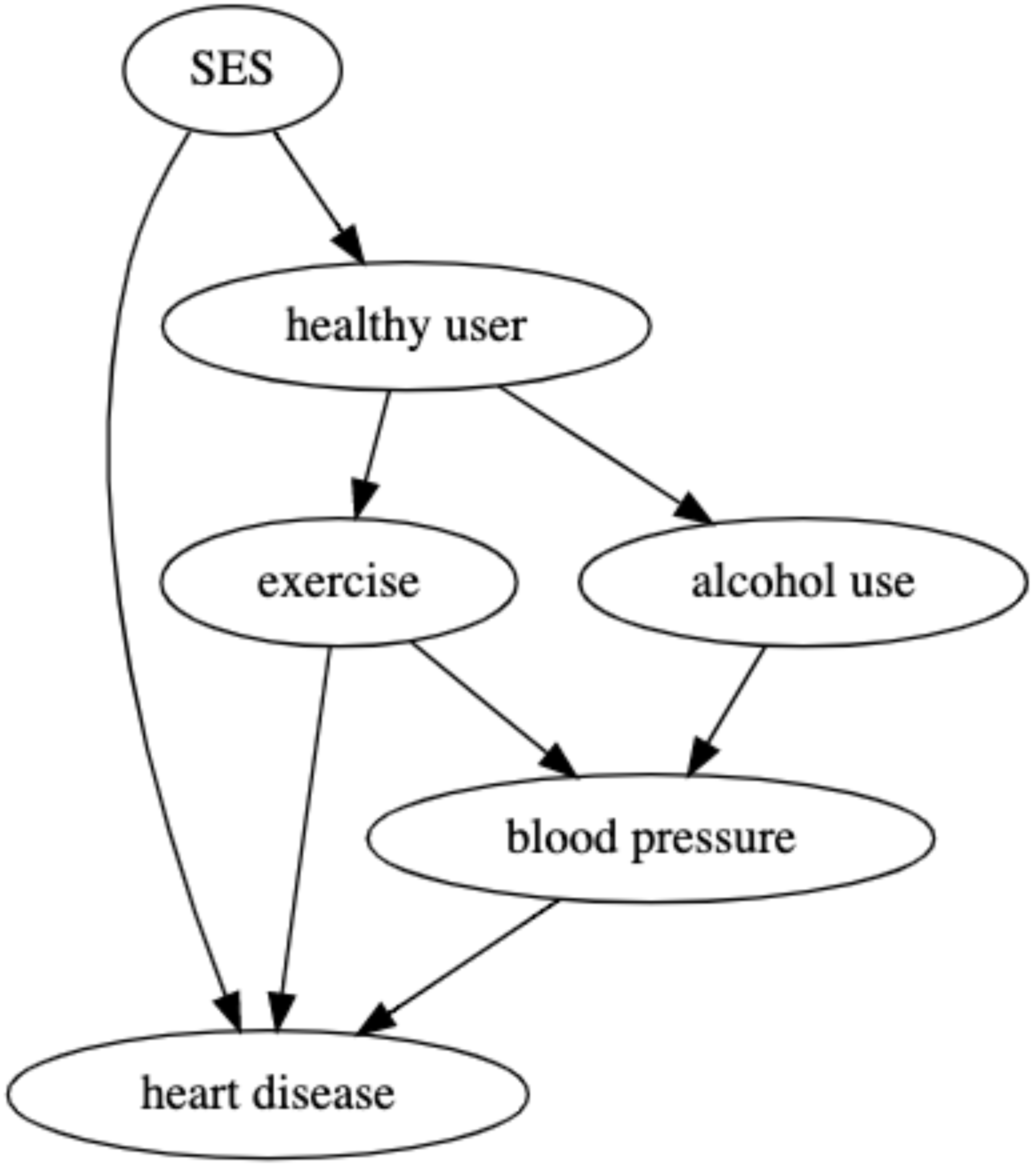


SEM as a causal model

The outputs from the process of modeling are:

1. A set $\boldsymbol{A*}$ of statements regarding the logical implications of $\boldsymbol{A}$, separate from the data, such as the conditional independences implicated by the graph.

```
('heart disease',
  'alcohol use',
  {'blood pressure', 'exercise', 'healthy user'}),
 ('heart disease', 'alcohol use', {'SES', 'blood pressure', 'exercise'}),
 ('heart disease',
  'alcohol use',
  {'SES', 'blood pressure', 'exercise', 'healthy user'}),
 ('heart disease', 'healthy user', {'SES', 'blood pressure', 'exercise'}),
 ('heart disease', 'healthy user', {'SES', 'alcohol use', 'exercise'}),
 ('heart disease',
  'healthy user',
  {'SES', 'alcohol use', 'blood pressure', 'exercise'}),
 ('exercise', 'alcohol use', {'healthy user'}),
 ('exercise', 'alcohol use', {'SES', 'healthy user'}),
 ('exercise', 'SES', {'healthy user’}),
…
```

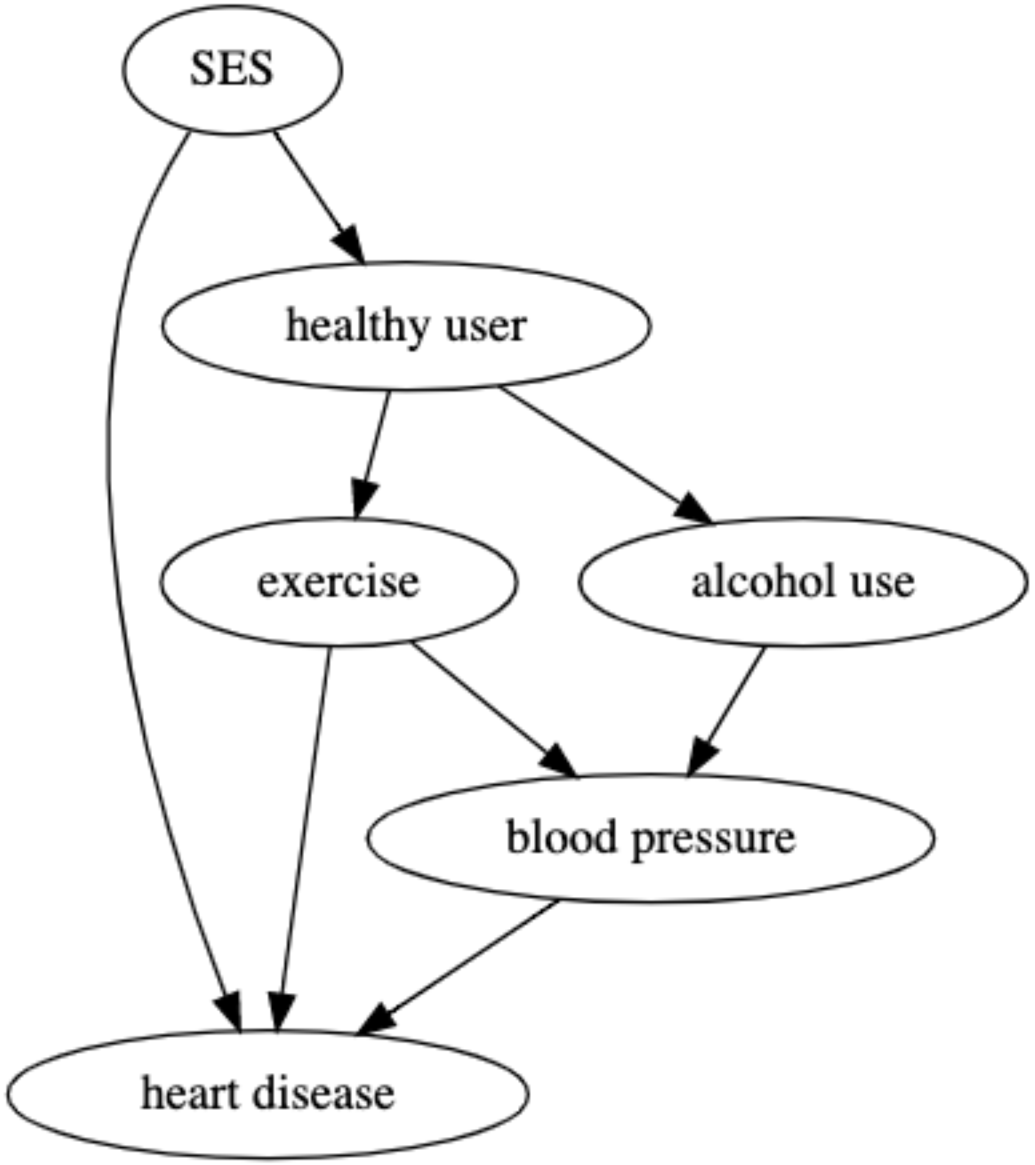


SEM as a causal model

The outputs from the process of modeling are:

2. A set $\boldsymbol{C}$ of data-based claims regarding the magnitudes or likelihoods of the questions in $\boldsymbol{Q}$ given the data $\boldsymbol{D}$ and assumptions $\boldsymbol{A}$

e.g. observed partial correlation b/w alcohol use and heart disease, conditioning on particular covariates to eliminate back-door paths

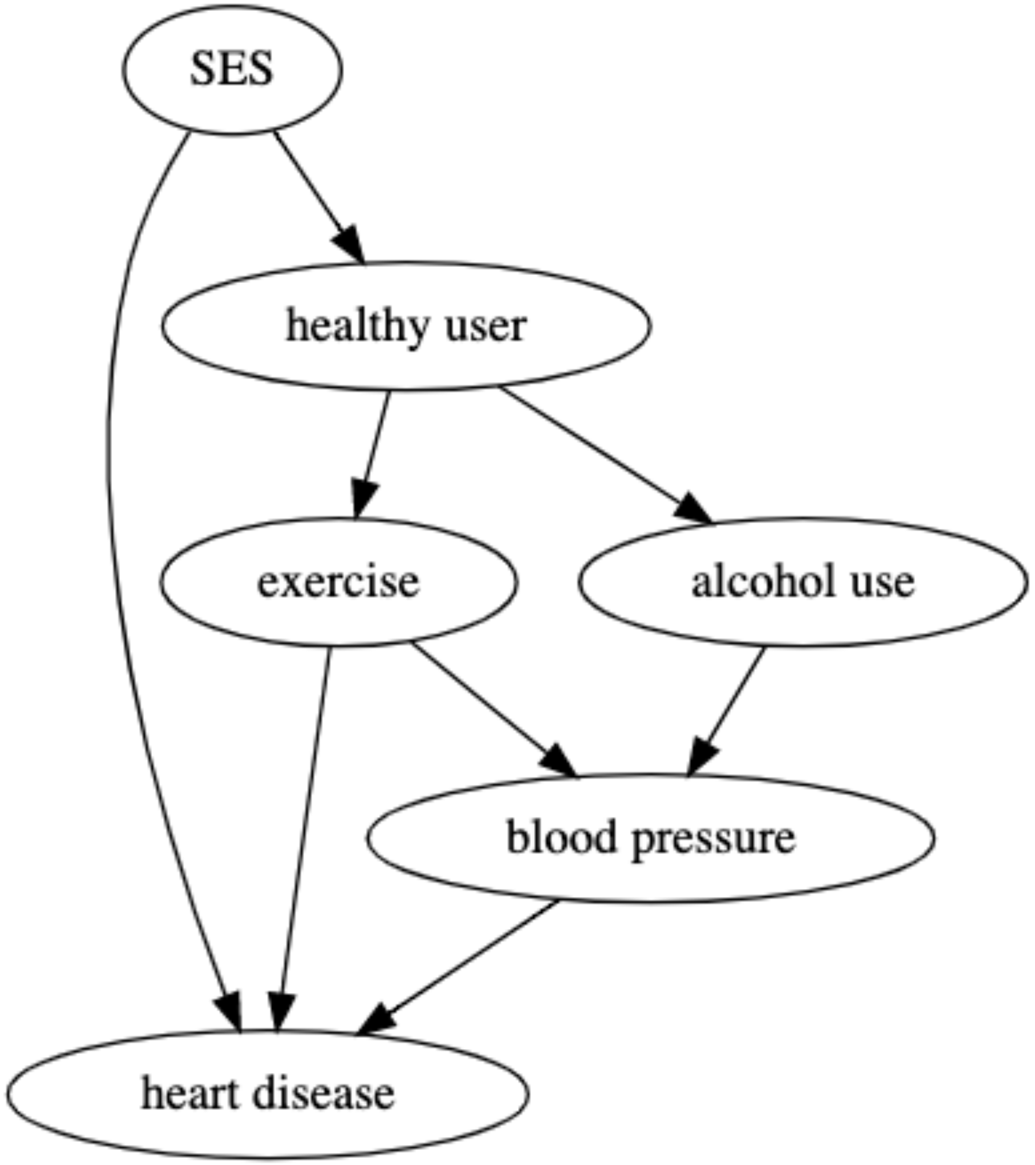


SEM as a causal model

The outputs from the process of modeling are:

3. A set $\boldsymbol{T}$ of testable statistical implications of $\boldsymbol{A}$, such as the predicted vanishing of particular partial correlations that is implied by conditional independence relationships.

e.g. test for whether r(alcohol use, heart disease | blood pressure) is zero



SEM as a causal model

SEM can ***never*** provide evidence in favor of the causal assumptions

"Lest there be any doubt: SEM does not aim to establish causal relations from associations alone. Perhaps the best way to make this point clear is to state formally and unambiguously what SEM does aim to establish. SEM is an inference engine that takes in two inputs, qualitative causal assumptions and empirical data, and derives two logical consequences of these inputs: quantitative causal conclusions and statistical measures of fit for the testable implications of the assumptions. Failure to fit the data casts doubt on the strong causal assumptions of zero coefficients or zero covariances and guides the researcher to diagnose, or repair the structural misspecifications. Fitting the data does not "prove" the causal assumptions, but it makes them tentatively more plausible. Any such positive results need to be replicated and to withstand the criticisms of researchers who suggest other models for the same data"— Bollen and Pearl (2012)

Instrumental variables

Does alcohol use cause heart disease?

Why can't we simply rely on observational data?

$$\text{heart disease} = \text{alcohol use} * \beta_1 + \epsilon$$

Least squares requires assumption that errors are uncorrelated with the covariates (including unobserved covariates!)

But unobserved causes will be correlated with error

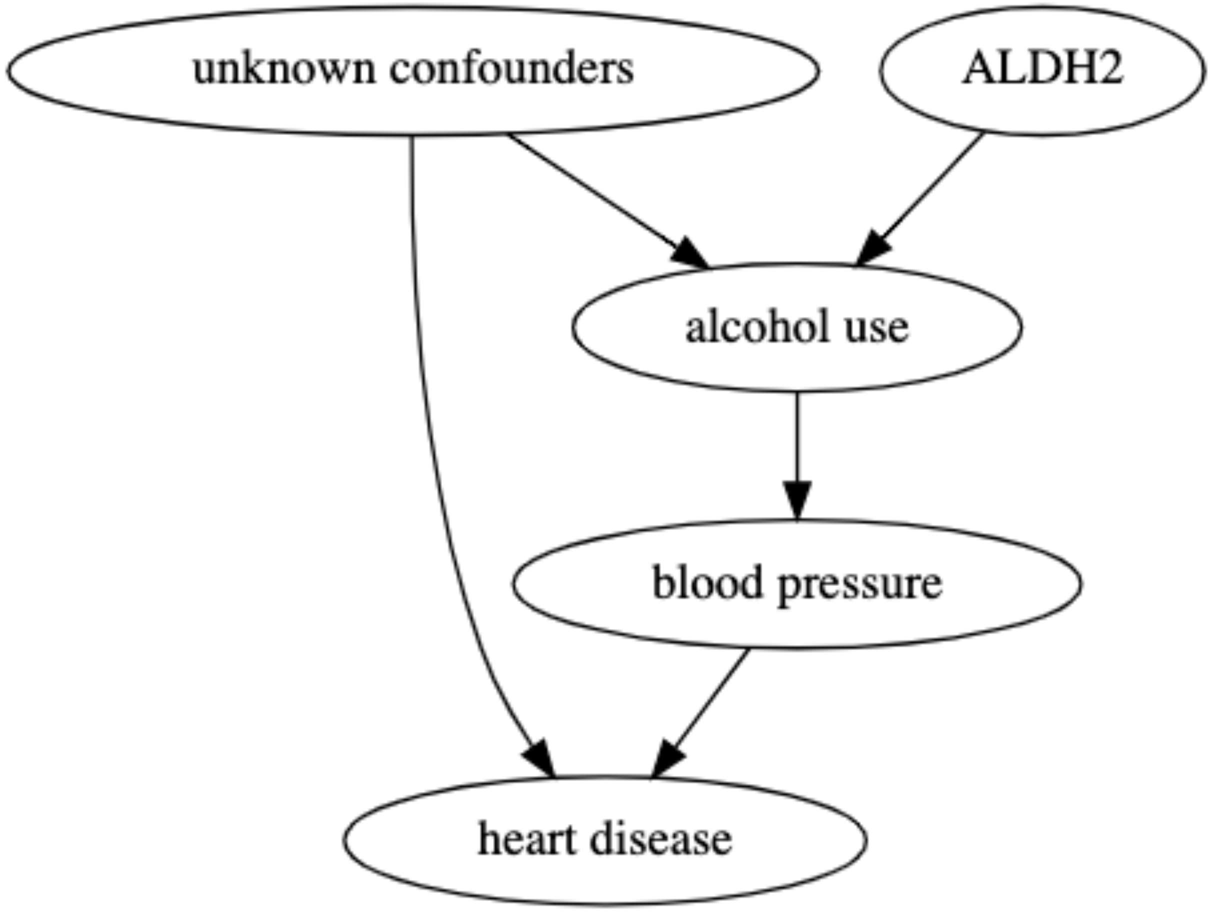
Instrumental variables as "natural experiments"

An instrumental variable is a a variable that is related to the putative cause, but independent of all other error terms (holding the putative cause constant)

Example: "Mendelian randomization"

- variant in ALDH2 gene leads to "alcohol flush" reaction
- Associated with substantially lower alcohol use

Causal effect of alcohol use on heart disease (assuming linearity):



$$\text{causal effect} = \frac{r(\text{ALDH2}, \text{heart disease})}{r(\text{ALDH2}, \text{alcohol use})}$$

Psych 253
Advanced Statistical Modeling

Gaussian graphical models

Daniel Yamins
Wu Tsai Neurosciences Institute
Departments of Psychology and Computer Science
Stanford University

Russ Poldrack
Department of Psychology
Stanford University

Gaussian graphical models

If our data are multivariate Gaussian, then we can use a known fact to move from the data to the underlying (undirected) graph:

If the inverse covariance of two variables X_i and X_j is zero, then X_i and X_j are independent given all other variables

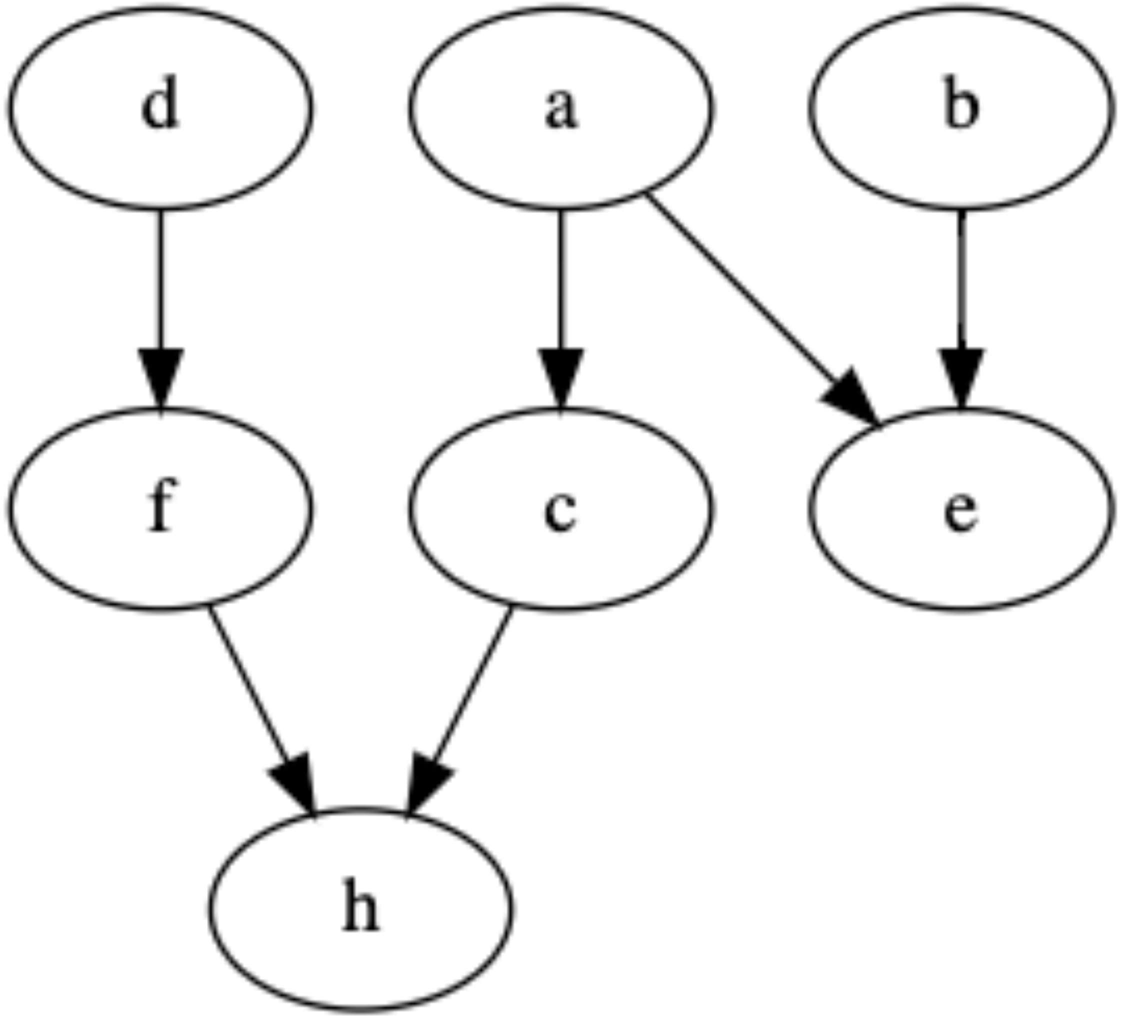
$$X = N(\mu, \Sigma)$$

$$\Theta = \Sigma^{-1}$$

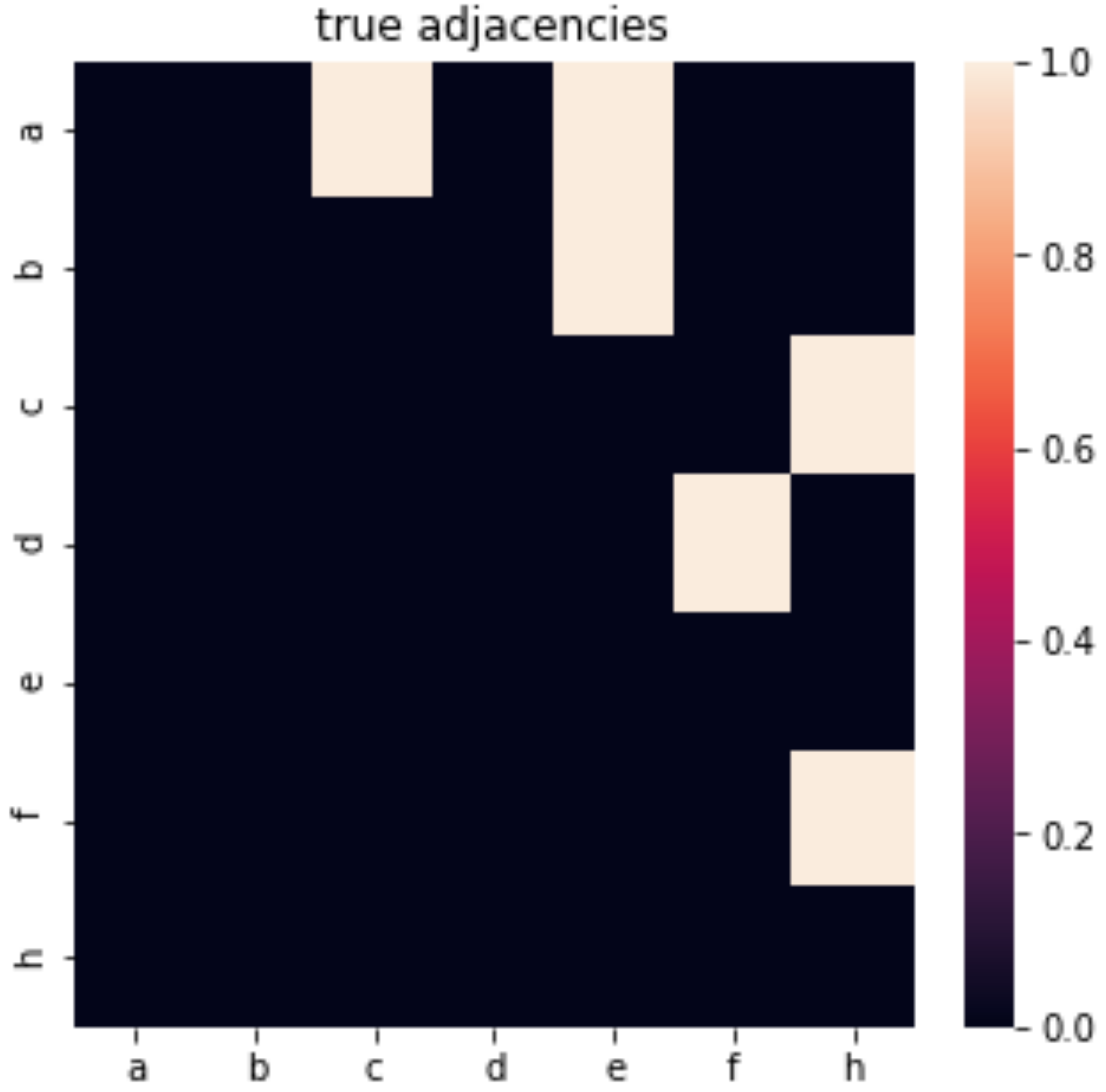
- inverse covariance (aka "precision")

$$\Theta_{i,j} = 0 \implies X_i \perp X_j | X_{\setminus\{i,j\}}$$

Gaussian graphical models: An example



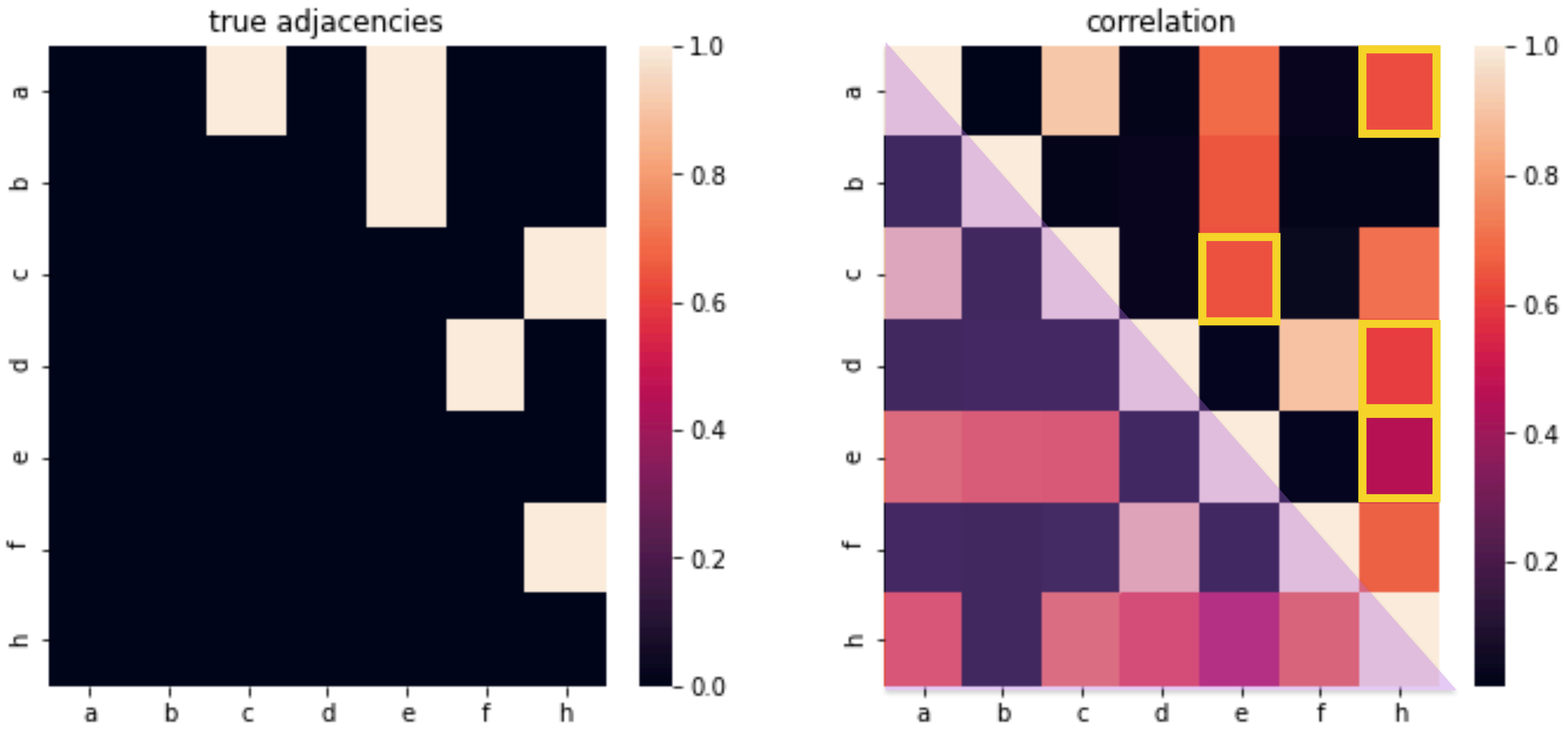
Adjacency matrix



Source node

Target node

Gaussian graphical models: An example

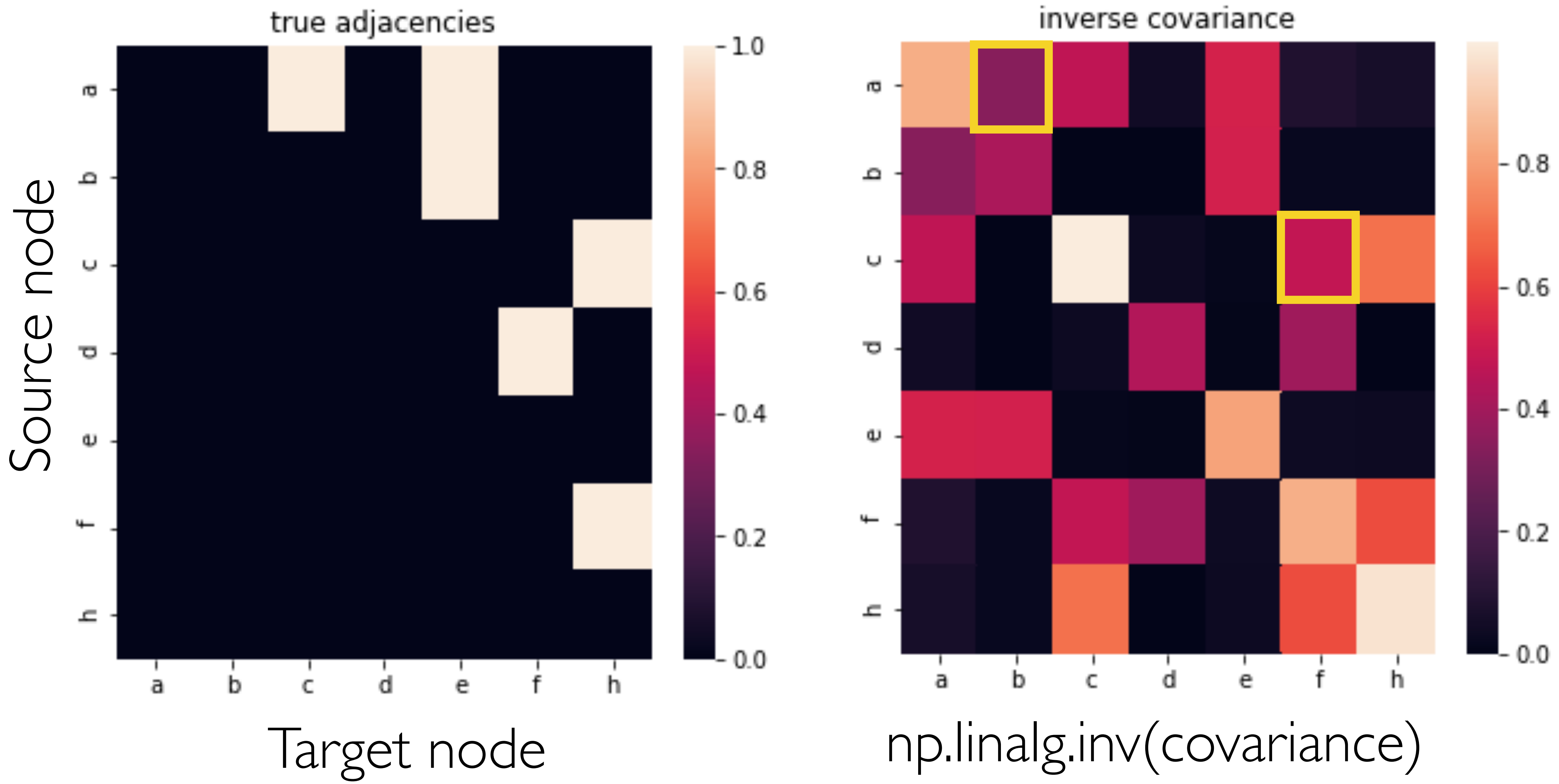


□ - correlations due to indirect connections

🟪 - correlations due to backward connections

Gaussian graphical models: An example

Inverse covariance identifies an *undirected* graph



☐ - spurious direct connections

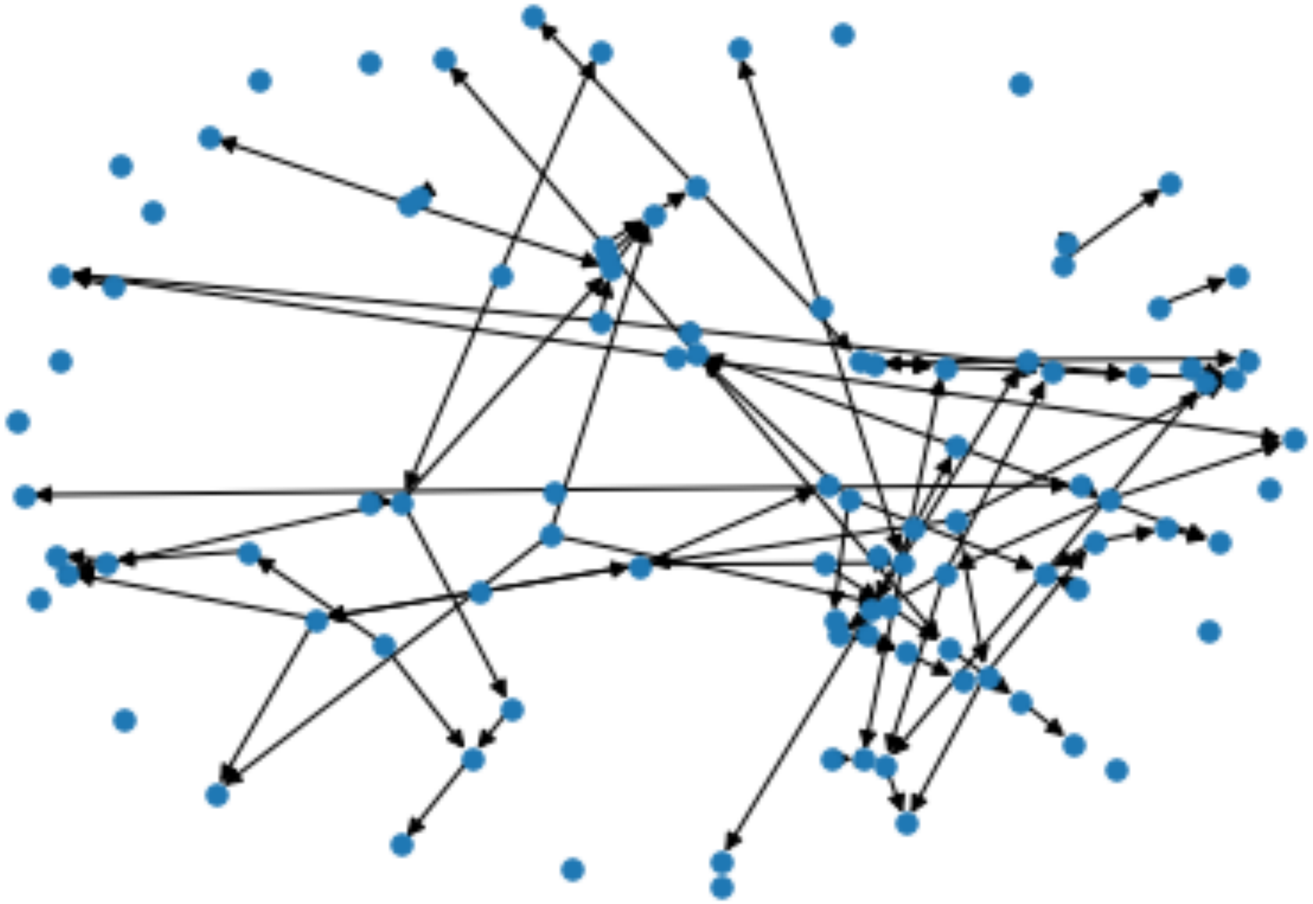
Challenges for inverse covariance estimation

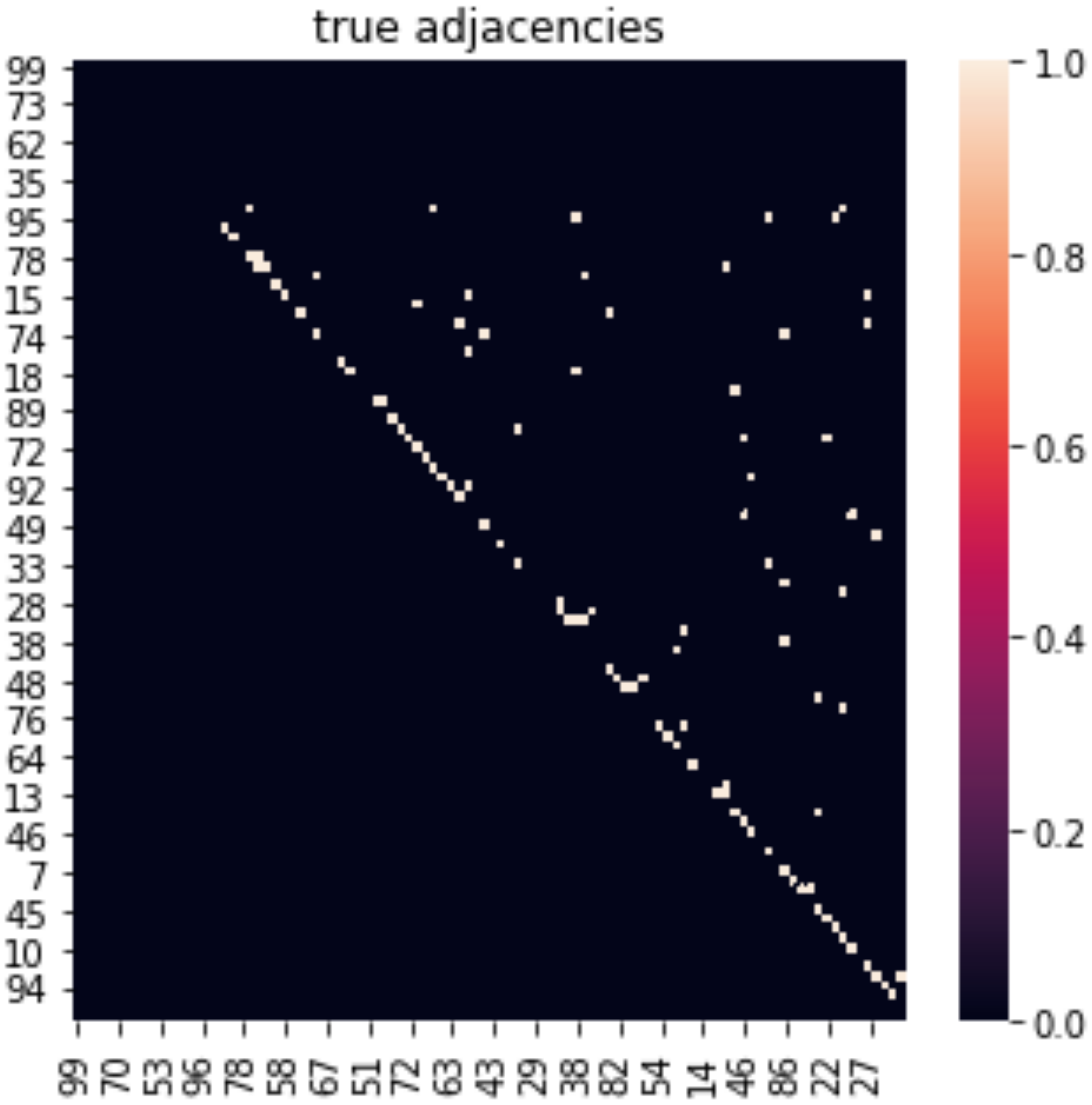
When number of nodes gets large, inversion of the covariance matrix can become ill-conditioned (especially if the number of observations is not large)

This can be addressed by using regularized inverse covariance estimation

Inverse covariance estimation: Example

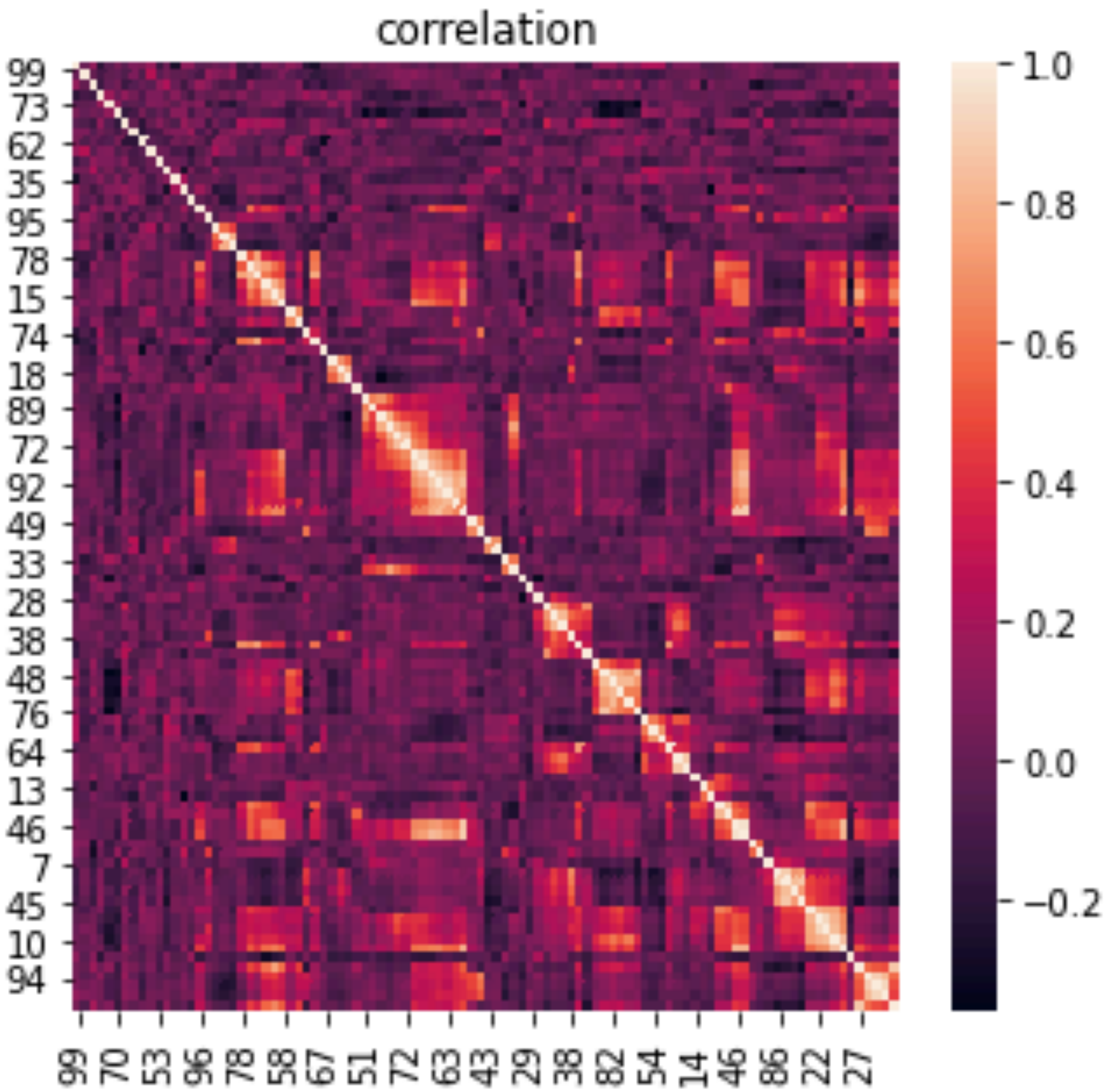
random DAG - 100 nodes, 100 edges



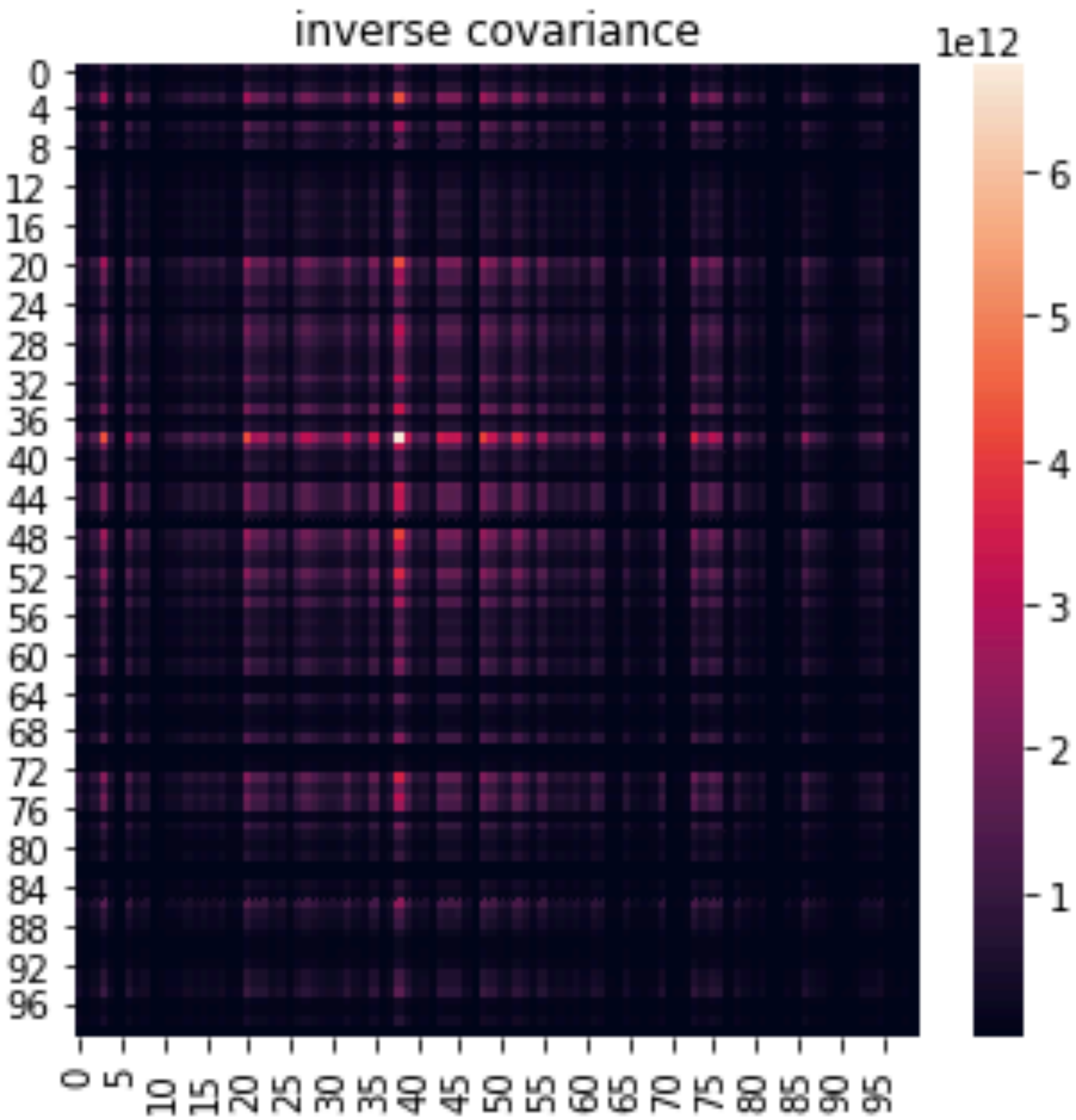


Inverse covariance estimation: Example

Using random data generated from this graph:



np.linalg.inv(covariance)



Note scale of values in the inverse covariance estimate…

Inverse covariance estimation: Example

How can we compare the observed inverse covariance to the true adjacencies?

It is continuous, so we need to threshold in order to make it binary

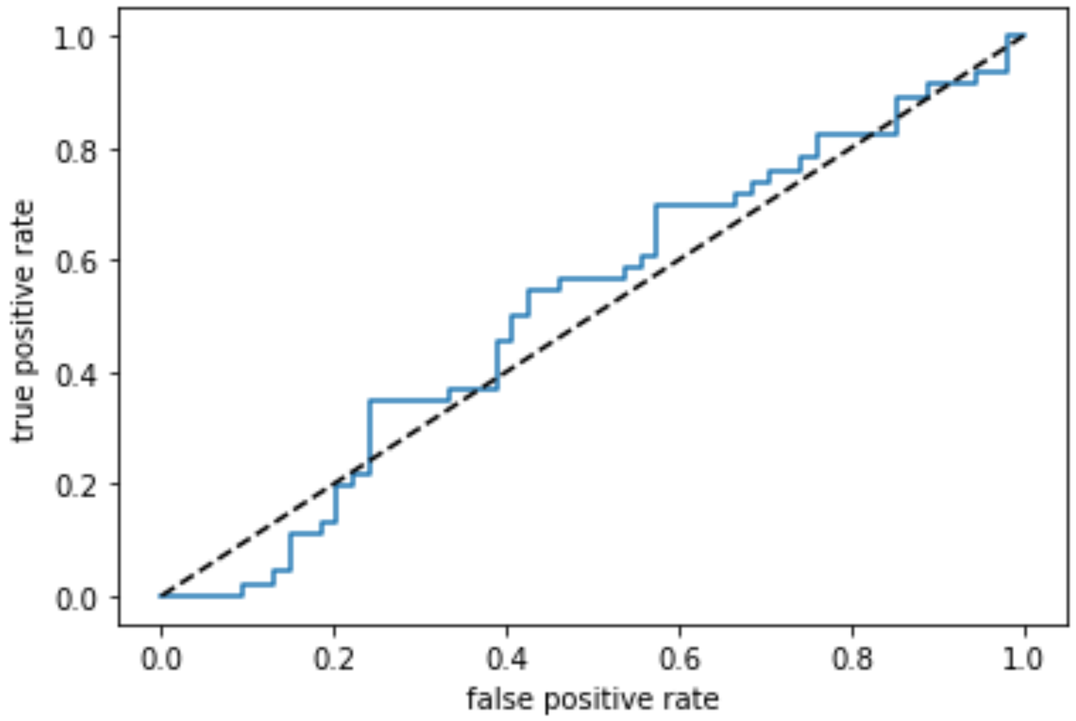
- but what threshold do we use?

- The receiver operating characteristic (ROC) curve allows us to look at performance across a range of thresholds

ROC curves

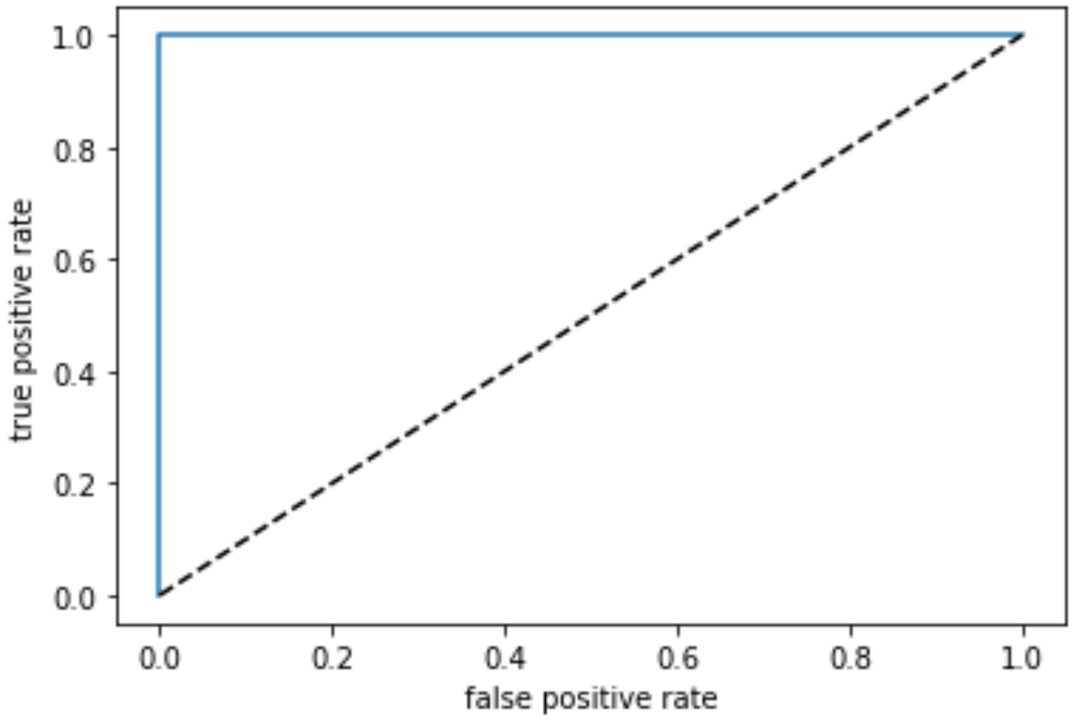
Random relationship

```
AUC: 0.5177133655394525
```

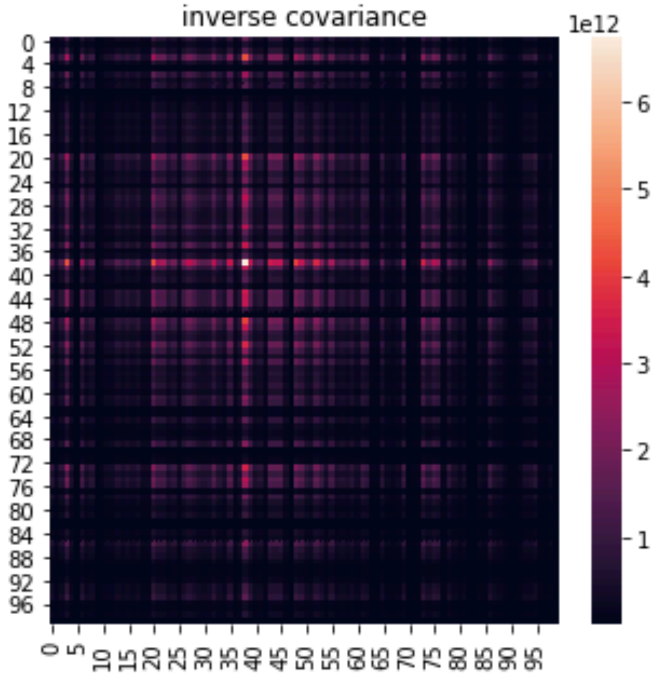


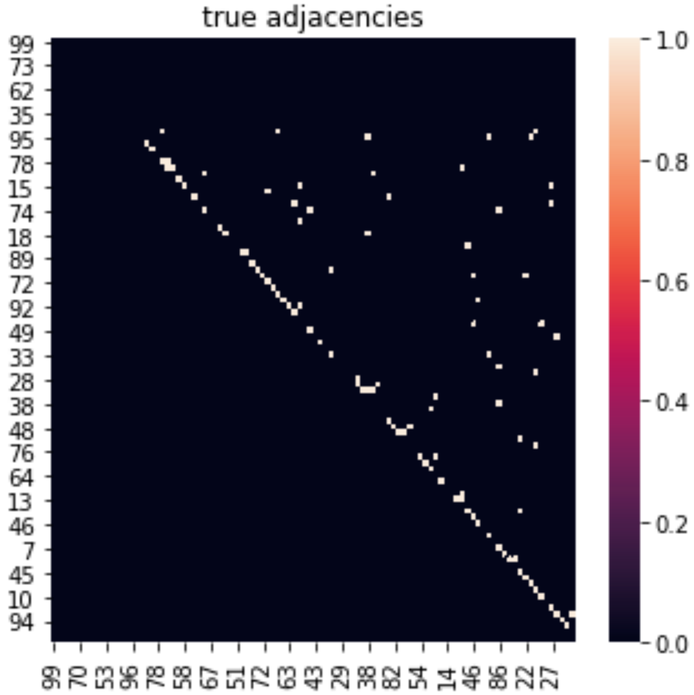
Perfect relationship

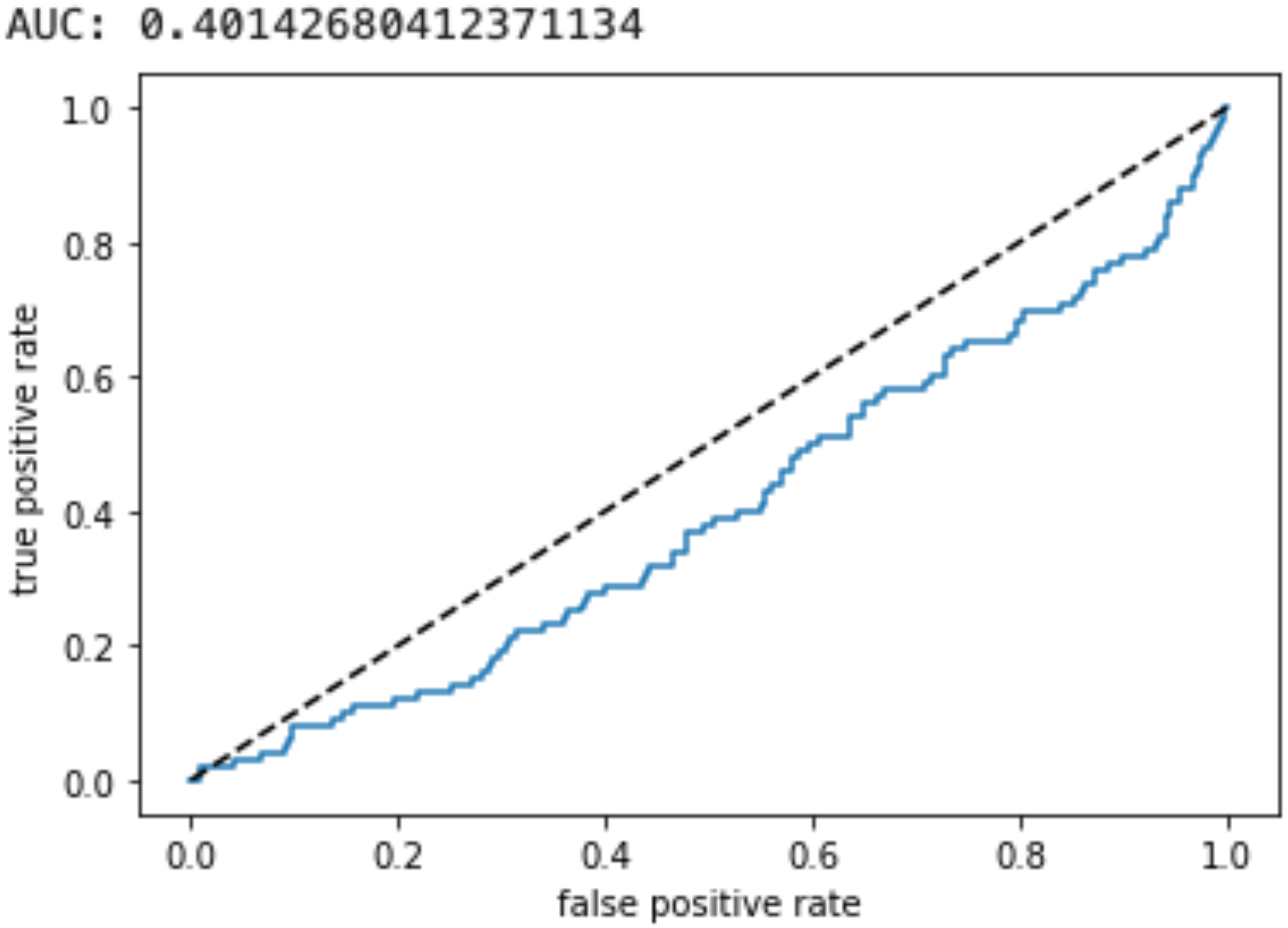
```
AUC: 1.0
```



ROC curve for estimated inverse covariance vs true graph







AUC: 0.40142680412371134

The graphical lasso (Friedman et al., 2007)

assume that $\boldsymbol{\Theta}$ is sparse (i.e. most entries are zero)

$$loss = -log|\Theta| + tr(\mathbf{S}\Theta) + \lambda\|\Theta\|_1$$

negative log determinant of estimated inverse covariance (larger when inverse covariances are large)

divergence of predicted and actual covariances

L1 penalty on inverse covariance parameters (larger when there are more nonzero parameters)

log-likelihood of multivariate Gaussian covariance

S: observed covariance

λ: penalty for L1 norm (i.e. sum of absolute values of $\boldsymbol{\Theta}$)

The graphical lasso (Friedman et al., 2007)

Graphical lasso accurately recovers the (undirected) graph structure

```
model = GraphicalLassoCV(alphas=10,
        n_refinements=4, max_iter=1000)
model.fit(simulated_data)
```