Causal Models

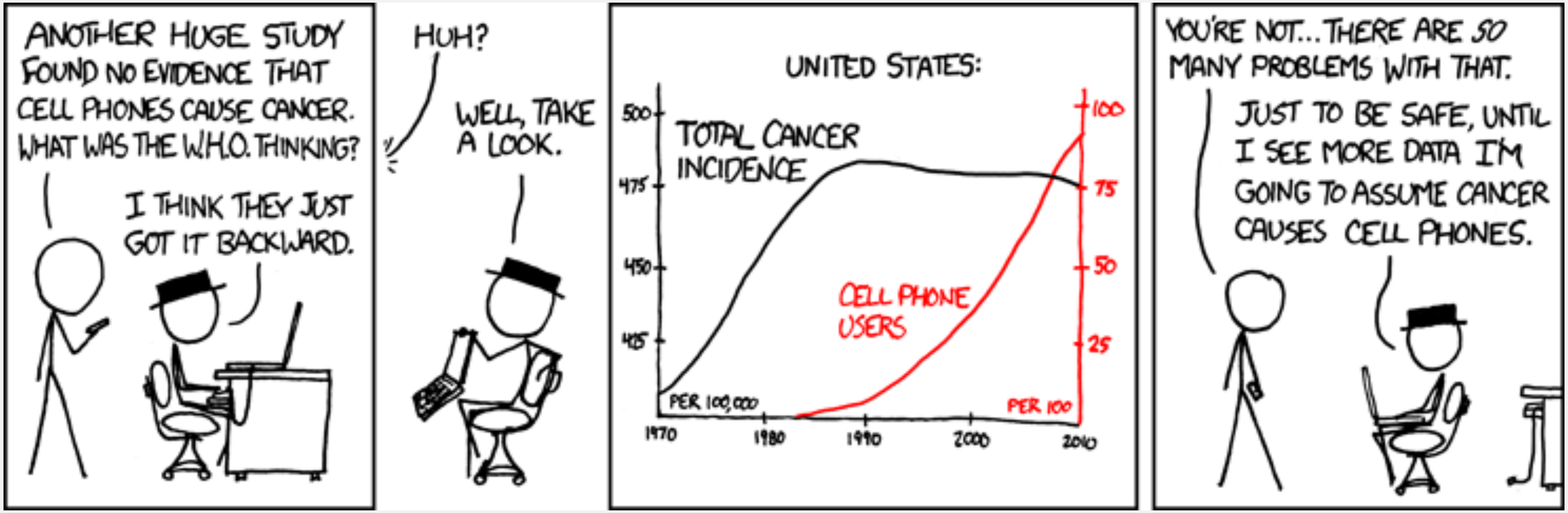
Vasant Honavar
Artificial Intelligence Research Laboratory
Informatics Graduate Program
Computer Science and Engineering Graduate Program
Bioinformatics and Genomics Graduate Program
Neuroscience Graduate Program

Center for Big Data Analytics and Discovery Informatics
Huck Institutes of the Life Sciences
Institute for Cyberscience
Clinical and Translational Sciences Institute
Northeast Big Data Hub
Pennsylvania State University

vhonavar@ist.psu.edu
http://faculty.ist.psu.edu/vhonavar
http://ailab.ist.psu.edu

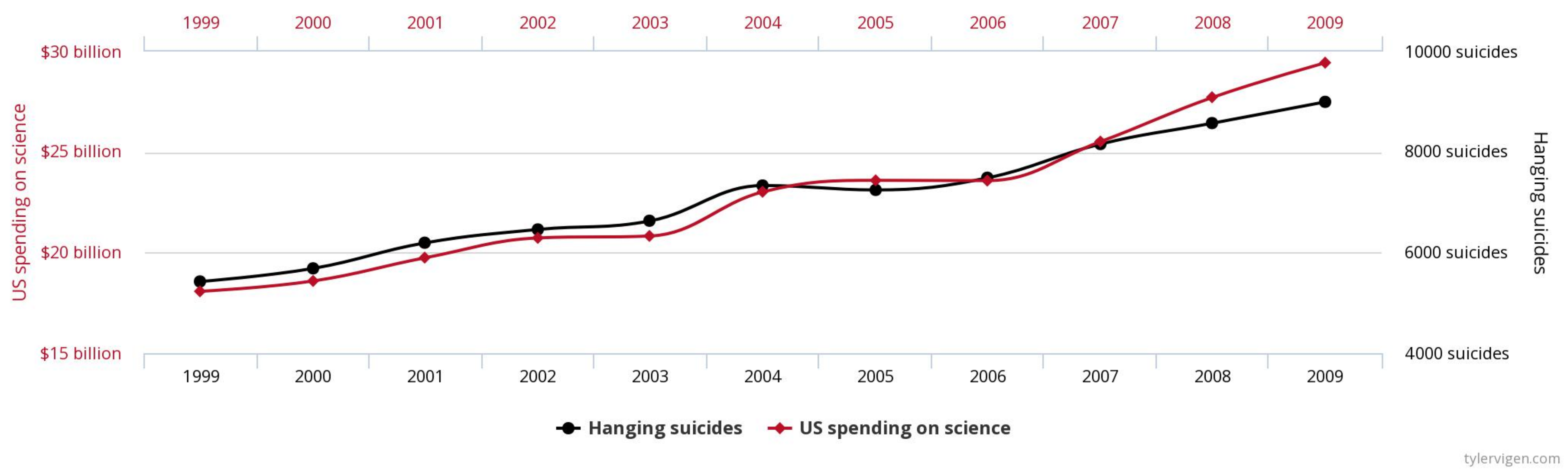
Perils of fishing for wisdom in oceans of data

Does cancer cause cell phone use?



Data Science and informatics have to offer more than a facility in finding correlations!

Perils of fishing for wisdom in oceans of data



Correlation between science funding and hanging suicides is over 0.99!

Eliminate science funding to save American lives!

Big data Fallacy

CHRIS ANDERSON, The end of theory: The data deluge makes the scientific method obsolete, *Wired Magazine* 16.07 (June 23, 2008). http://www.wired.com/science/discoveries/magazine/16-07/pb_theory

- "Petabytes allow us to say: "Correlation is enough." We can stop looking for models. We can analyze the data without hypotheses about what it might show. We can throw the numbers into the biggest computing clusters the world has ever seen and let statistical algorithms find patterns where science cannot."
- "Correlation supersedes causation, and science can advance even without coherent models, unified theories, or really any mechanistic explanation at all."
- We will argue that causal models are extremely important for making sense of big data!

Causality in Biomedical and Health Sciences

The central concern of all sciences, biomedical and health sciences included, has to do with the discovery of causal relationships

- Understanding mechanisms
- Predicting the results of interventions
- Controlling events

Causality

- Is there a causal calculus?
  - What language can we use to represent and reason about causal relationships?
- What tools do we have for
  - Answering causal questions from causal models?
  - Learning causal models from observational and experimental data?
  - Generalizing a causal account beyond the setting in which it was obtained?

Statistics 1834–2018

- "The object of statistical methods is the reduction of data" (Fisher 1922)
  - Statistical concepts are those expressible in terms of joint distribution of observed variables
  - All others are:
    - "substantive matter," "domain dependent," "metaphysical," "ad hockery,"
    - i.e., outside the province of statistics
    - thus ruling out many interesting questions
- Traditional statistics education promotes a fear of venturing to answer causal questions (Pearl, 2000)

Things are changing

- “More has been learned about causal inference in the last few decades than the sum total of everything that had been learned about it in all prior recorded history.” (Gary King, Harvard, 2014)
- From liability to respectability
  - Many workshops, including one at NAS
  - An NIH BD2K Center for Causal Modeling and Discovery
  - Papers in AAAI, NIPS, ICML, UAI, PNAS, JSSM…
- Causal revolution makes it fun to solve important problems that Pearson, Fisher, Neyman, and most of of their successors. . . were not able to articulate, let alone solve!

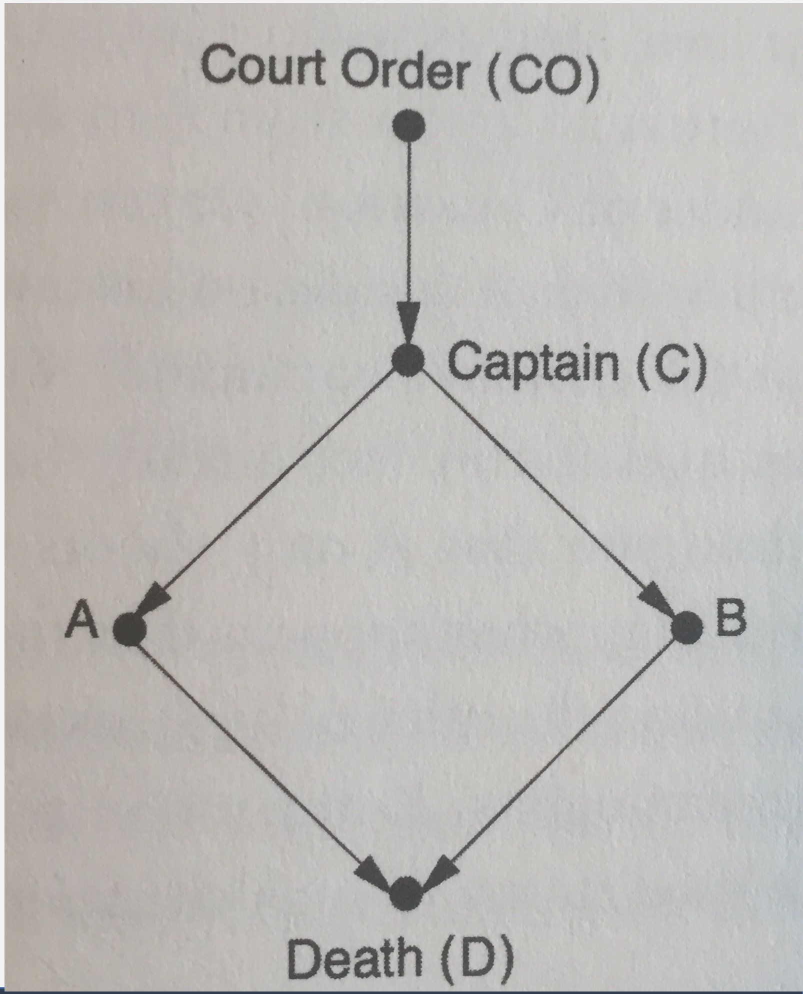
What can a causal reasoner do?

- Association
  - Activity: Seeing (Observation)
  - Question: How would seeing X change my belief about Y?
  - Methods: Statistics, Traditional machine learning
  - Powerful methods for summarizing data!
- Intervention
  - Activity: Doing (Intervention)
  - Question: What would Y be if I do X?
  - Statistics and traditional machine learning don’t offer the means to even pose the question, let alone answer it!
- Counterfactuals
  - Activity: Imagining (Retrospection)
  - Question: What would Y be if I had not done X?
  - Statistics and traditional machine learning don’t offer the means to even pose the question, let alone answer it

What does a causal reasoner need?

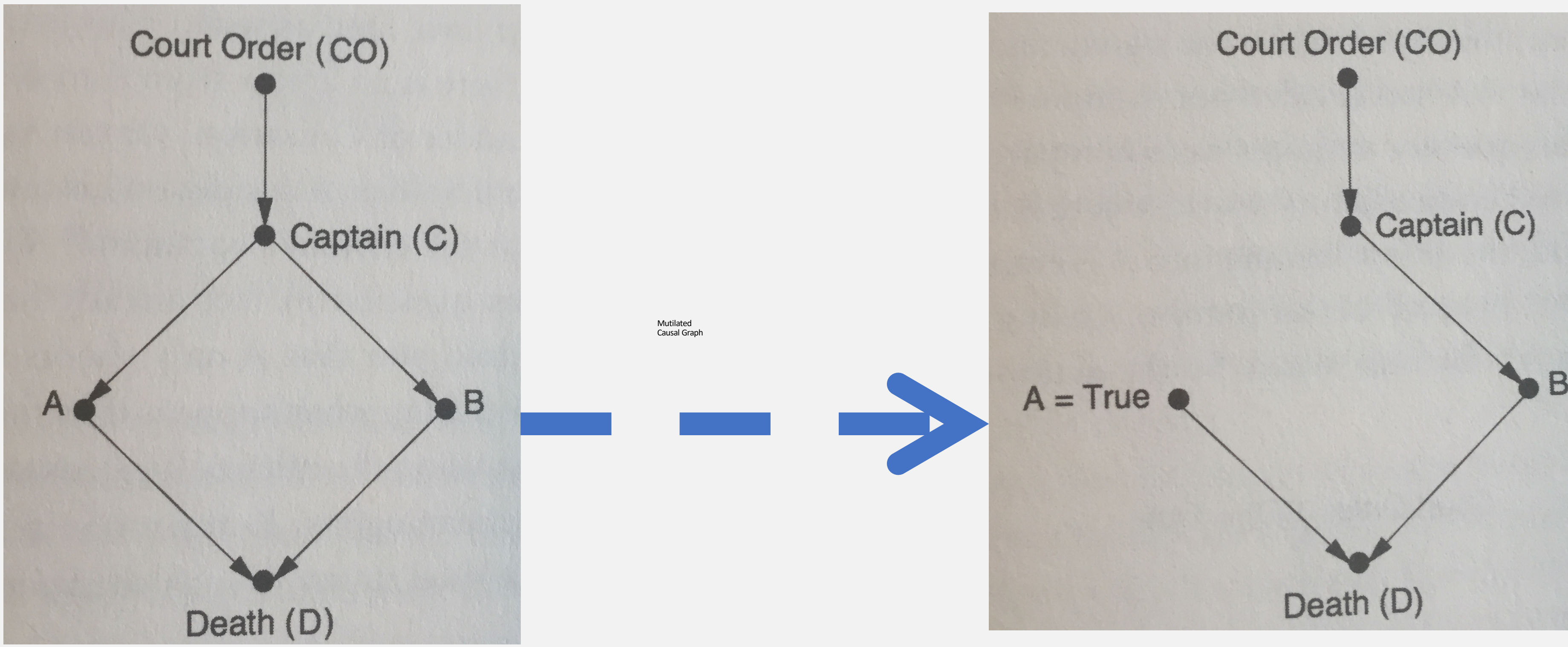
AI Mantra: Representation before anything else

- How can we represent causal knowledge?
- Causal diagrams
  - Nodes denote variables
  - Links denote direct causes
- Suppose all variables are Boolean
- If a court order is given captain orders soldiers A and B to fire. If at least one fires, prisoner dies.
- Association: Prisoner is found dead. Was court order given?



What does a causal reasoner need?

- Intervention: If soldier A is compromised and is forced to shoot (without caption's order, would the prisoner die?



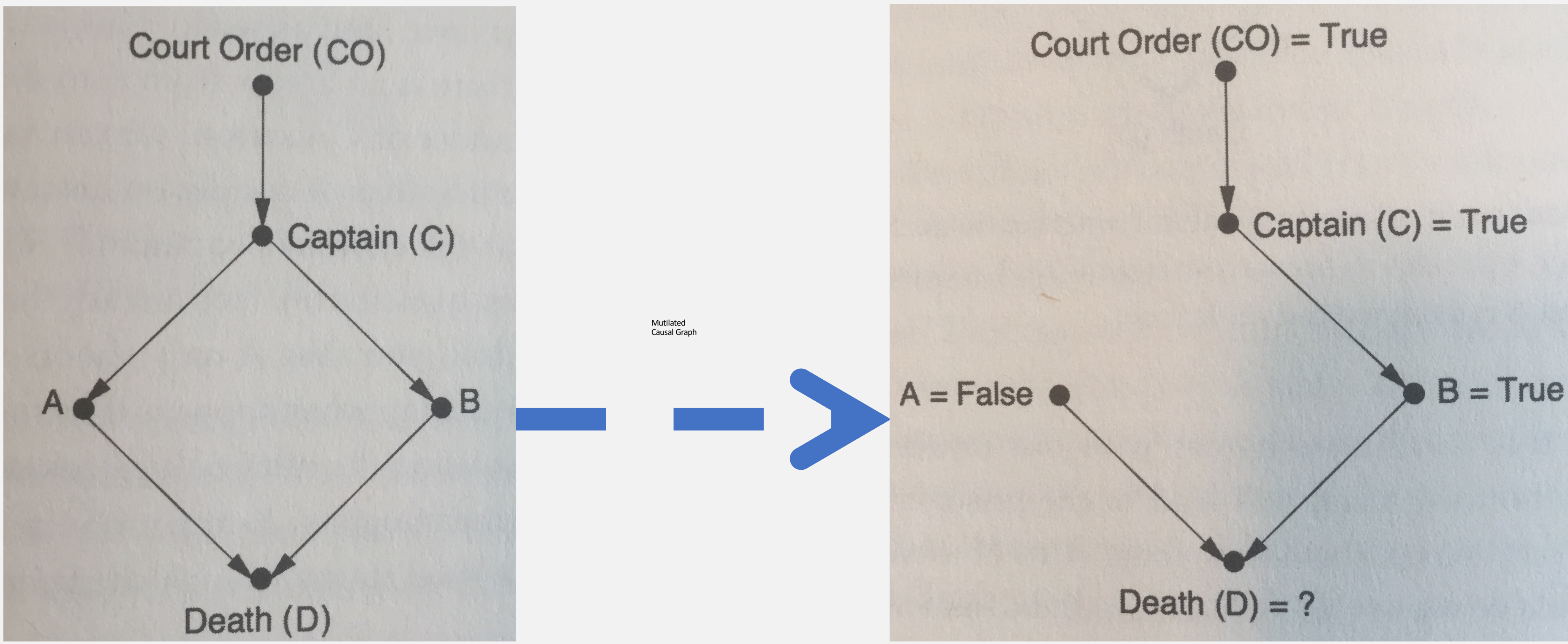
Seeing ≠ Doing!

Seeing: If A shoots, we can conclude that B shoots too

Doing: If A is forced to shoot, we can't say what B does

What does a causal reasoner need?

- Counterfactual: Suppose the prisoner is found dead. Would he have died had A's gun failed to shoot?



Seeing ≠ Imagining!

Seeing: If D is dead, A and B must have shot (captain's order, court order)

Imagining: If A failed to shoot, and D is dead, B must have shot….

Causal reasoning under uncertainty

Data:

- We learn that 99 kids who were vaccinated die of a reaction to vaccine
- And 40 who were not vaccinated die of smallpox
- More children die from vaccine than those that die from smallpox

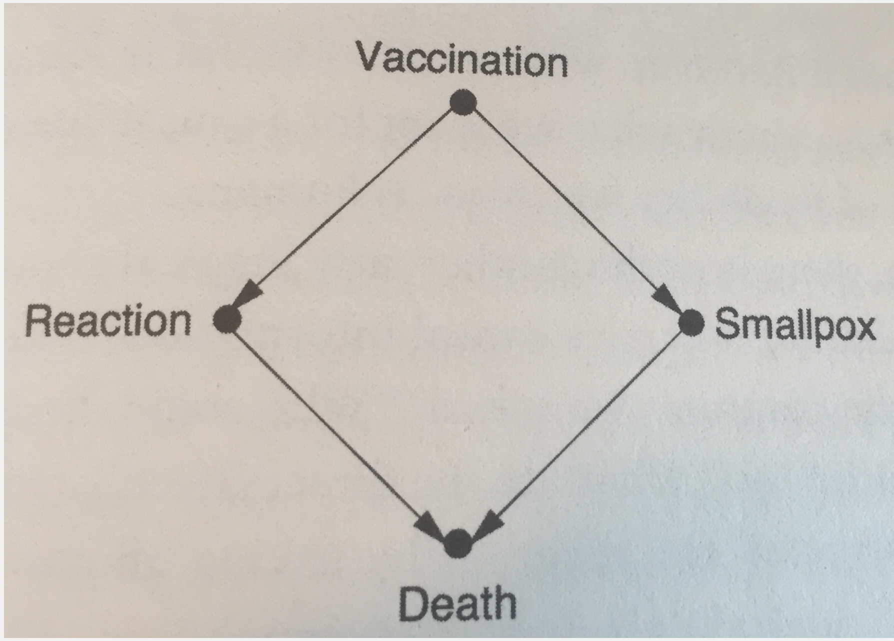
Question:

- Should we ban vaccination?
  - Big data answer?
  - CNN headlines?
  - Your answer?

Causal reasoning under uncertainty

Suppose we know: 99% of children are vaccinated, 1% are not

- A vaccinated child has a 1 in 100 chance of a reaction; and a reaction has a 1 in 100 chance of being fatal
- A child who is not vaccinated has 0 chance of reaction, but 1 in 50 chance of smallpox which is fatal in 1 in 5 cases



Data:

- Out of 1 million kids, 990,000 are vaccinated; 9900 have reaction, 99 of whom die
- 10,000 are not vaccinated, 200 get smallpox, 40 die of smallpox

Fact:

- More children die from vaccine than those that die from smallpox

Question:

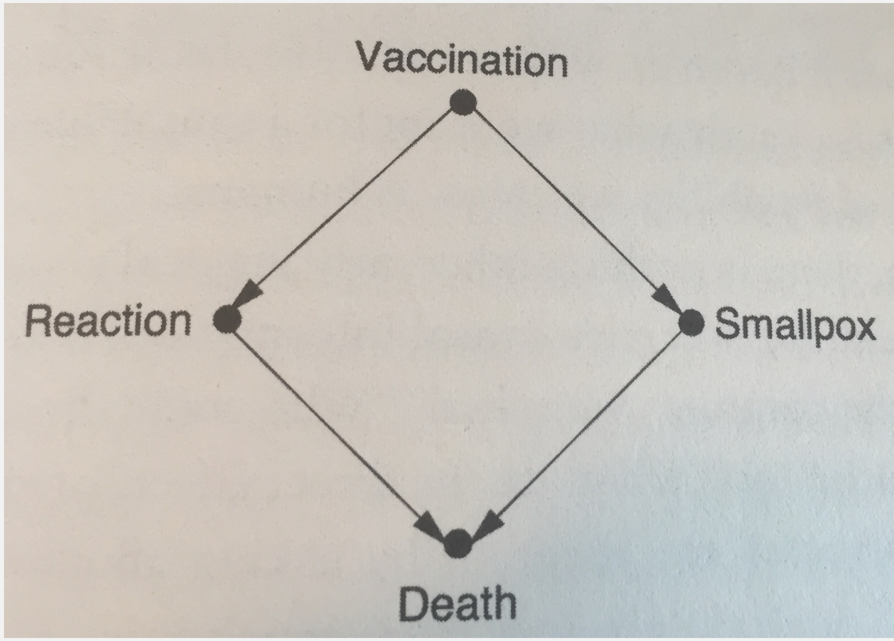
- Should we ban vaccination?

Pearl, 2018

Causal reasoning under uncertainty

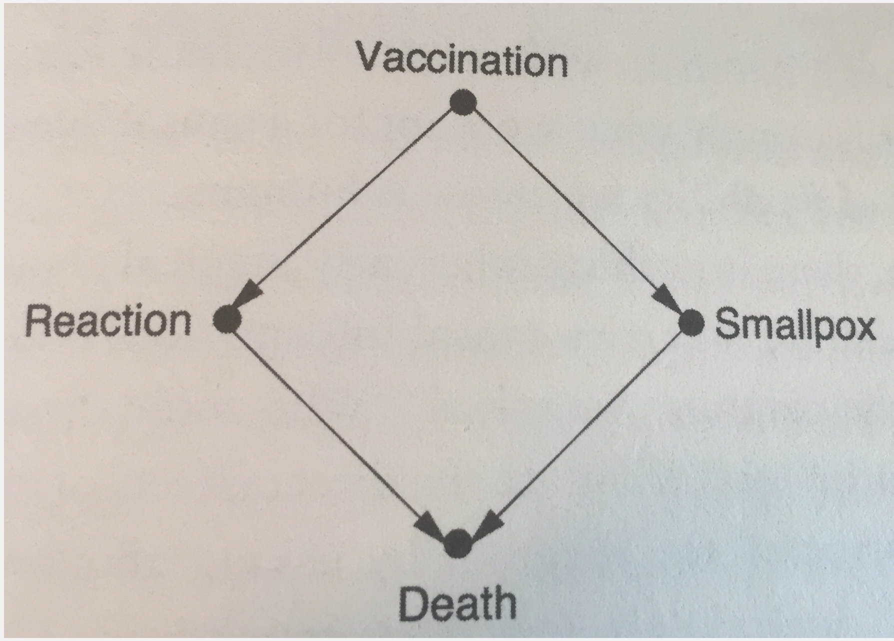
Should we ban vaccination?

- Depends.
- On what?
- On the answer to the counterfactual question
  - How many children out of 1 million would have died if none had been vaccinated?
- If no child is vaccinated,
  - No child would have a reaction, and there would be no reaction related fatalities
  - We expect 1 in 50, or 20,000 out of 1 million small pox cases, of which 1 in 5 or 4000 would result in death



Causal reasoning under uncertainty

Should we ban vaccination?



Data informed by causal model:

- Out of 1 million children
  - If none were vaccinated, 4000 would have died
  - If 99% were vaccinated, 99+40 = 139 would have died

Fact:

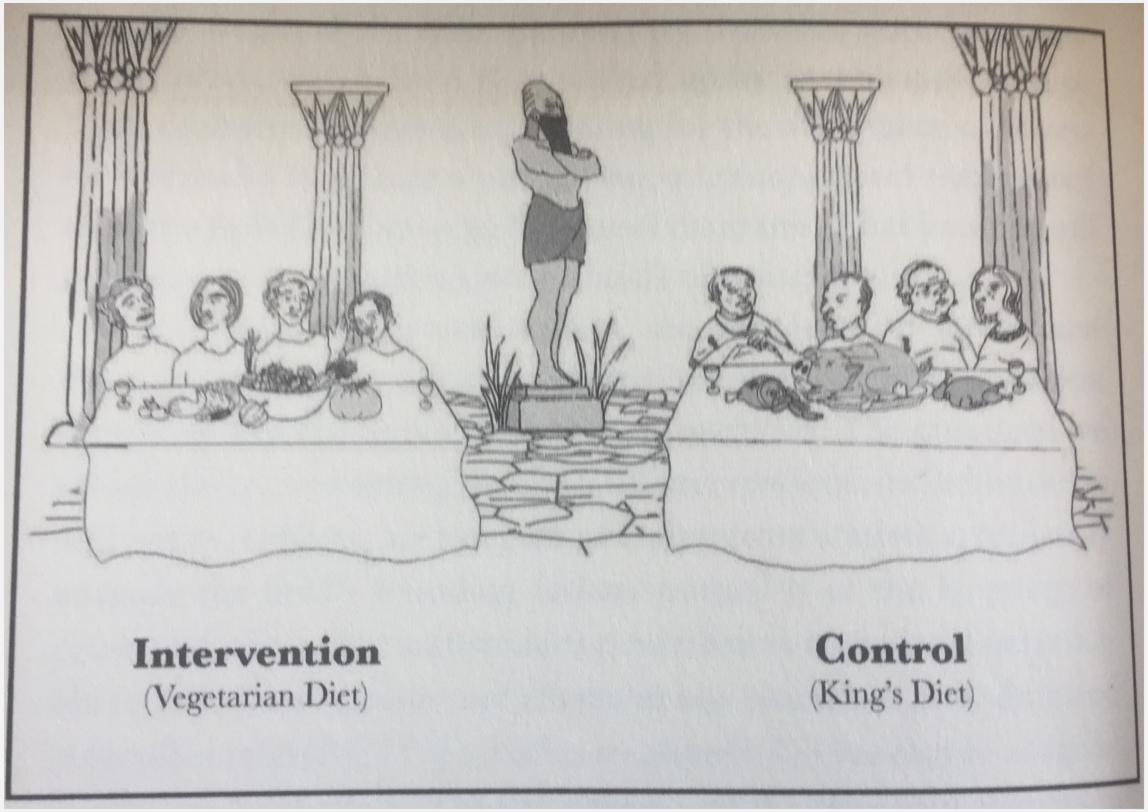
- Fewer children (139) die if the vaccination policy is in place than if vaccination were banned (4000)

Question:

- Should we ban vaccination?
- Obviously not!

Note: We could not have answered this question from observational data alone in the absence of a causal model

Controlled experiments



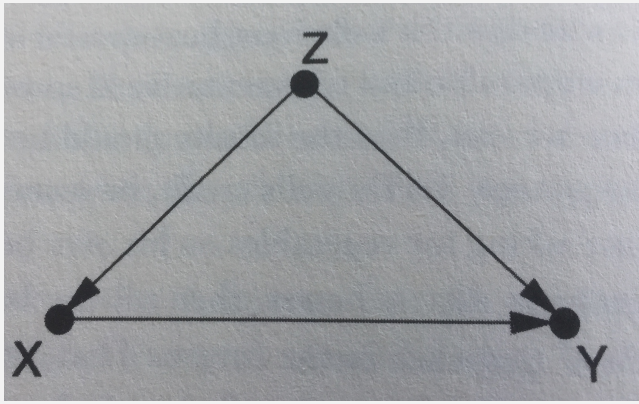
- Prospectively choose two groups of individuals
  - One is treated (veg diet)
  - The other is not
- Compare the two groups
- Do you see a problem?

Observations versus Controlled Experiments

- Suppose you want to see whether exercise helps reduce heart disease
- Prospectively choose two groups of individuals
  - One group exercised
  - The other did not
- Compare the two groups on incidence of heart disease
- Do you see a problem with this setup?

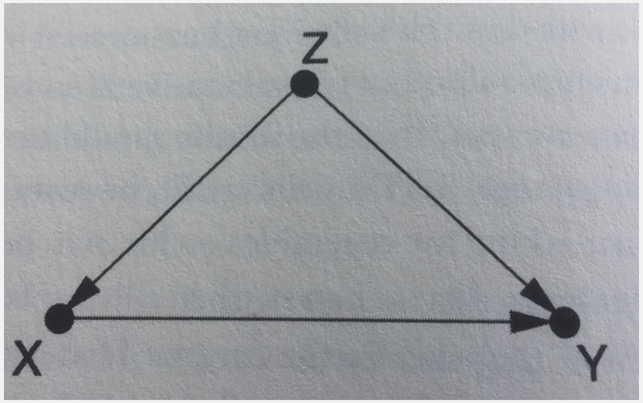
- Solution – Randomized control trial
  - But RCT is not always feasible

Confounding bias



- Suppose the treated group is healthier than the control group to start with
- Confounding bias arises whenever a variable influences both who is selected for treatment and the outcome of the experiment
  - Sometimes the confounders are known
  - Sometimes the confounders are suspected
- The most basic version of confounding
  - The true causal effect X → Y is mixed with the spurious correlation induced by the fork X ←Z→ Y
  - Example:
    - We are testing a drug but give it to patients who are younger, but not to those who are older
    - age becomes a confounder

Confounding bias



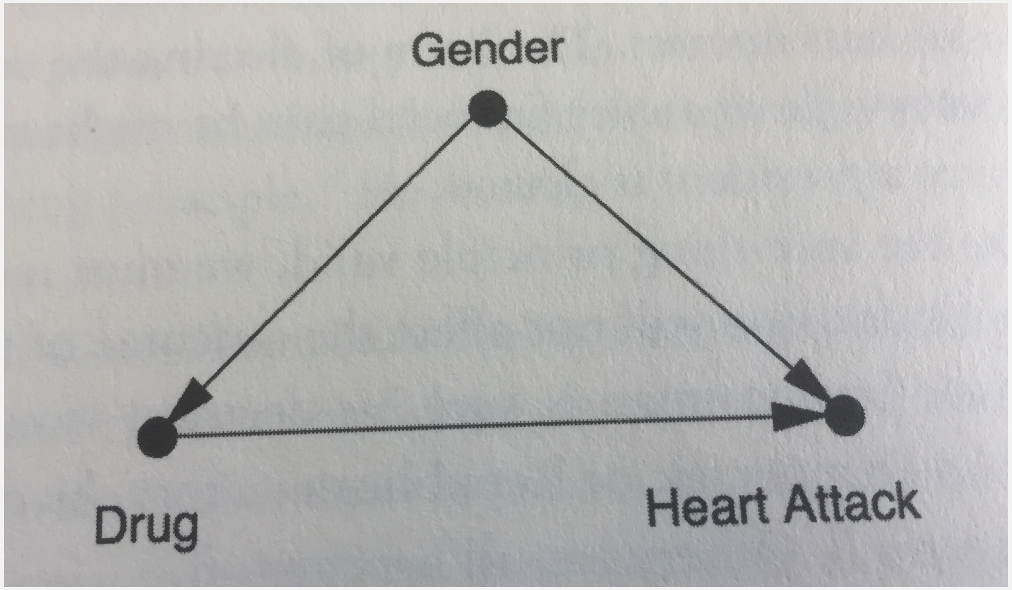
- The most basic version of confounding
  - The true causal effect X → Y is mixed with the spurious correlation induced by the fork X ←Z→ Y
  - Example: We are testing a drug but give it to patients who are younger, but not to those who are older – age becomes a confounder
- If we have measurements on the confounder, it is very easy to de-confound the true and spurious causal effects – `adjusting for Z'
  - Simply compare treatment and control groups for each value of Z
  - Take a weighted average where the weights correspond to the fraction of the population represented by each value of Z

The puzzle of a drug that is bad for men, bad for women, but good for people

	Control Group (No Drug)		Treatment Group (Took Drug)	
	Heart attack	*No heart attack*	*Heart attack*	*No heart attack*
Female	1	19	3	37
Male	12	28	8	12
Total	13	47	11	49

- The data are from an observational study
- The data present instance of Simpson's paradox which has puzzled statisticians since 1956
- There dozens of papers and PhD theses have been written to "explain" the Simpson's paradox, the most recent one in 2017

The puzzle of a drug that is bad for men, bad for women, but good for people



- Suppose gender is unaffected by the drug
- Suppose gender affects both heart attack risk and whether the patient chooses to take the drug
- Gender is a confounder that needs to be controlled for in assessing the effect of the drug on heart attack

The puzzle of a drug that is bad for men, bad for women, but good for people

	Control Group (No Drug)		Treatment Group (Took Drug)	
	Heart attack	*No heart attack*	*Heart attack*	*No heart attack*
Female	1	19	3	37
Male	12	28	8	12
Total	13	47	11	49

- For women, the rate of heart attack was 1 in 20 (5%) without the drug and 3 in 40 (7.5%) with the drug: **The drug is bad for women**
- For men, the rate of heart attack was 12 in 40 (30%) without the drug and 8 in 20 (40%) with the drug: **The drug is bad for men**
- Adjusting for the confounder, given that the proportion of men and women is the same, we simply average the gender-specific heart attack rates to get population rates
- Among the population at large, the rate of heart attacks is 17.5% without the drug and 23.75% with the drug: The drug is bad for people
- Paradox resolved!

Eliciting causal effects: The story so far

- We can predict the effect of an intervention by adjusting for the confounders if we have data on a sufficient set of variables (de-confounders) to block all backdoor paths between the intervention and the outcome
- What if the confounders are not observable?
  - All previous methods fail, and we need Randomized Control Trials
  - What if RCT are not feasible? due to practical or ethical reasons?

What is wrong with adjusting for confounders?

- Adjusting for confounders is both:
  - Overrated by standard statistical practice because of controlling for many more variables than needed, including variables that should not be controlled for
    - Statistical methodology provides little guidance for what variables to control for
    - You can end up controlling for the very thing you are trying to measure
  - Underrated by standard statistical practice because even when the confounders have been properly identified and controlled for, conclusions stop short of making causal claims even if they are justified

Confounding redefined

- Many definitions, almost all flawed (Pearl, 2018)
- Correct definition using the language of causal calculus
  - Confounder is any factor that leads makes

$$P(Y \mid X) \neq P(Y \mid do(X))$$

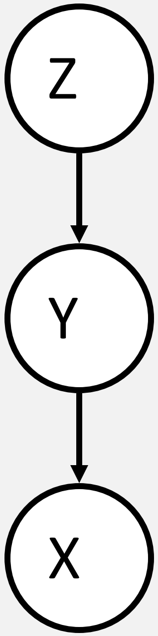
Confounding in the language of causal calculus

- Confounder is any factor that leads makes $P(Y|X) \neq P(Y|do(X))$
- In causal graphs, information flows in causal direction (along the arrows) as well as non-causal direction
- Non-causal paths are the source of confounding
- How can we stop the flow of information along non-causal paths?
  - In a chain A→B→ C, controlling for B makes A and C independent
  - In a fork A ←B → C, controlling for B makes A and C independent
  - In a collider A → B ← C, A and C are independent to start with. But if you control for B or one of its descendants, you make A and C dependent (so B becomes a confounder although it was not in the absence of adjustment)!
    - The usual statistical practice of controlling for every variable you can think of is not just wasteful, but fatally flawed!

Reasoning about Independence in Causal Graphs

What Independences does a causal graph model?

Example:



Given Y, does learning the value of Z tell us nothing new about X?

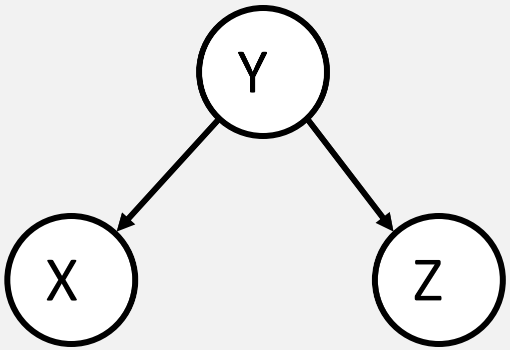
i.e., is $P(X|Y, Z)$ equal to $P(X \mid Y)$?

Yes. Since we know the value of all of X's parents (namely, Y), and Z is not a descendant of X, X is conditionally independent of Z.

Also, since independence is symmetric, $P(Z|Y, X) = P(Z|Y)$.

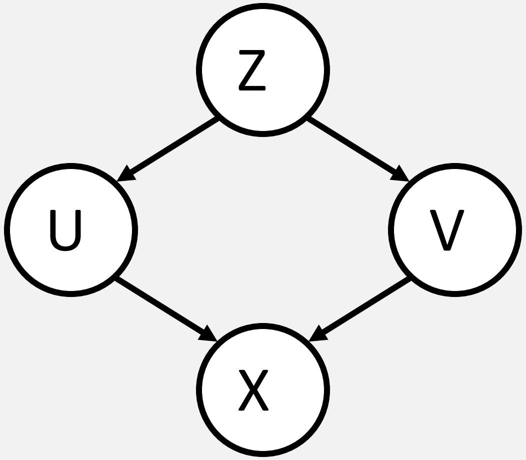
What Independences does a causal network model?

- Let $I(X,Y,Z)$ represent *X* and *Z* being conditionally independent given *Y*.



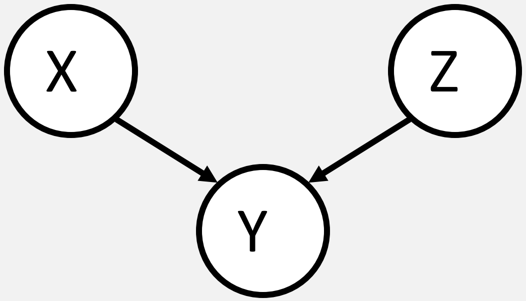
- $I(X,Y,Z)$? Yes, just as in previous example: All X' s parents given, and Z is not a descendant.

What Independences does a Bayes Network model?



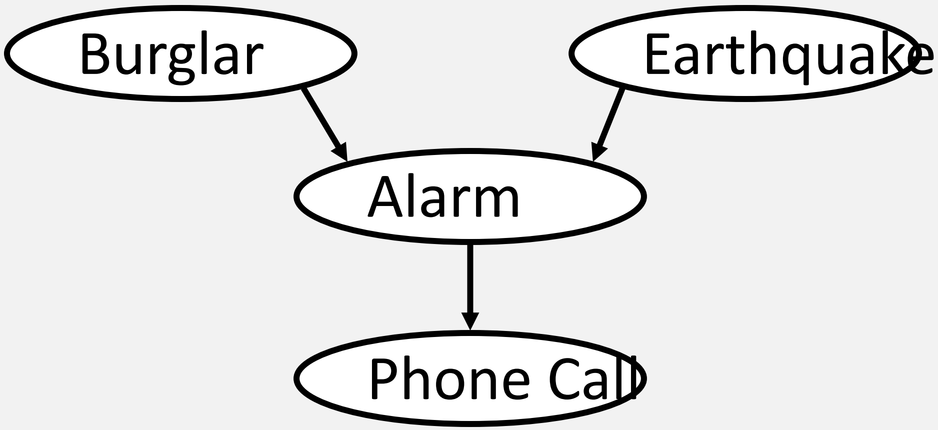
- $I(X,\{U\},Z)$? No.
- $I(X,\{U,V\},Z)$? Yes.

Dependency induced by V-structures

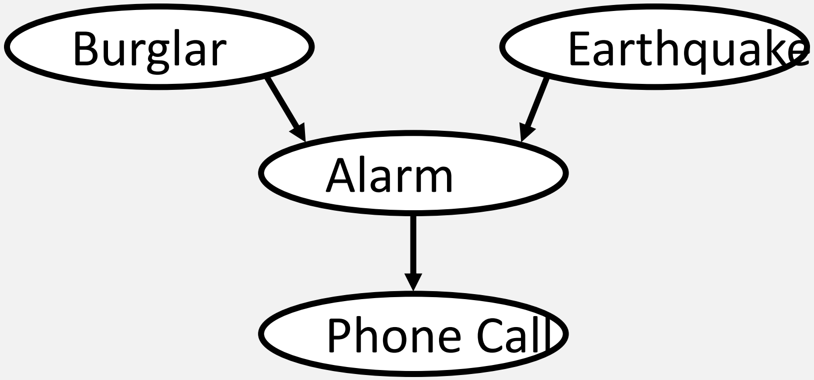


- X has no parents, so we know all its parents' values trivially
- Z is not a descendant of X
- So, $I(X,\{\},Z)$, even though there is a undirected path from X to Z through an unknown variable Y.
- What if we do know the value of Y? Or one of its descendants?

The Burglar Alarm example



- Your house has a twitchy burglar alarm that is also sometimes triggered by earthquakes.
- Earth arguably doesn’t care whether your house is currently being burgled
- While you are on vacation, one of your neighbors calls and tells you your home’s burglar alarm is ringing.



- But now suppose you learn that there was a medium-sized earthquake in your neighborhood. …Probably not a burglar after all.
- Earthquake "explains away" the hypothetical burglar.
- But then it must NOT be the case that

  $I(\text{Burglar}, \{\text{Phone Call}\}, \text{Earthquake})$,

  even though $I(\text{Burglar}, \{\}, \text{Earthquake})$!

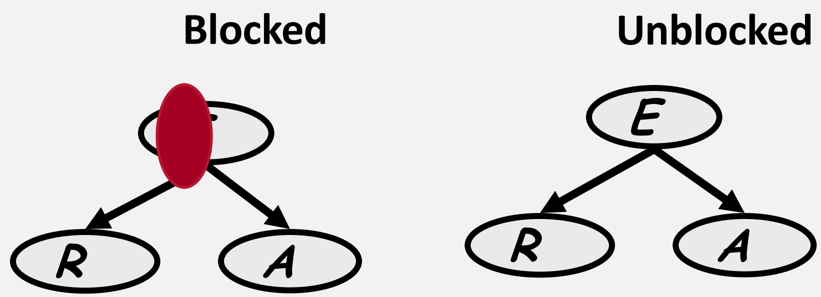
d-separation

- Fortunately, there is a relatively simple algorithm for determining whether two variables in a Bayesian network are conditionally independent given some other variables:
  - *d-separation*.
- Two variables are independent if all paths between them are blocked by evidence
- Three cases:
  - Common cause
  - Intermediate cause
  - Common Effect

d-separation

Evidence may be transmitted through a diverging connection unless it is instantiated.

- Two variables are independent if all paths between them are blocked by evidence
- Three cases:
  - Common cause
  - Intermediate cause
  - Common Effect

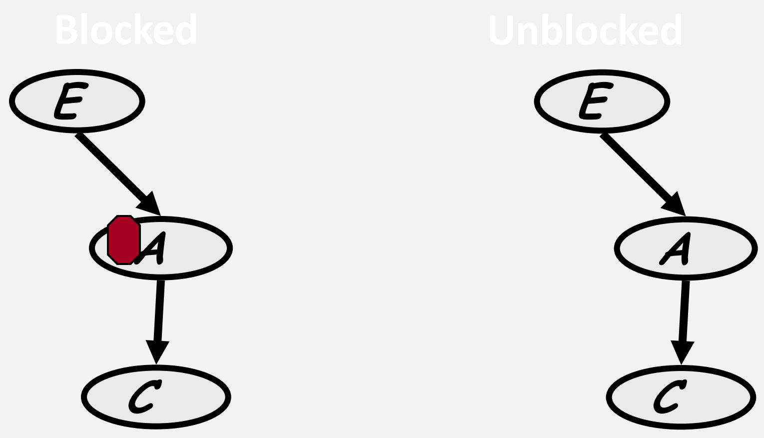


- If we do not know whether an earthquake occurred, then radio announcement can influence our belief about the alarm having gone off.
- If we know that earthquake occurred, then radio announcement gives no information about the alarm

d-separation

Common cause
Intermediate cause
Common Effect

Blocked Unblocked



Evidence may be transmitted through a serial connection unless it is blocked

d-separation

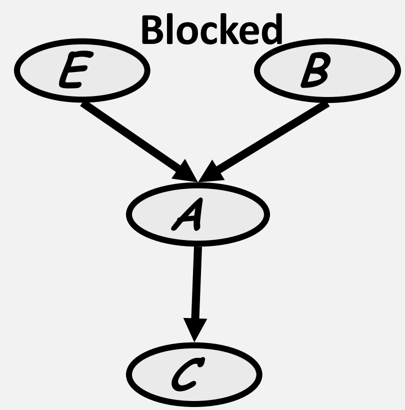
Blocked

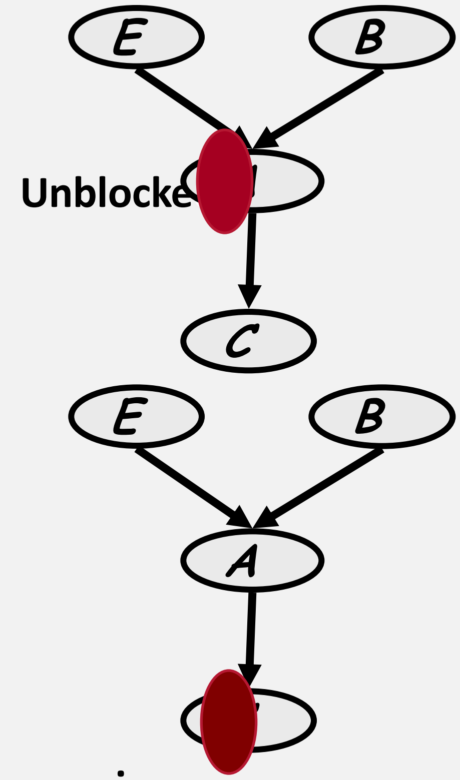
Unblocked

Common cause

Intermediate cause

Common Effect



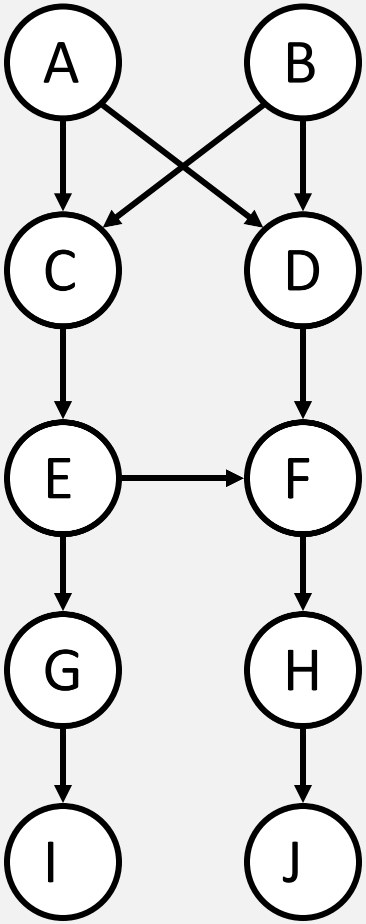


Evidence may be transmitted through a converging connection only if either the variable or one of its descendants has received evidence

d-separation

- Definition: *X* and *Z* are *d-separated* by a set of evidence variables *E* iff every undirected path from *X* to *Z* is "blocked" by evidence *E*

d-separation

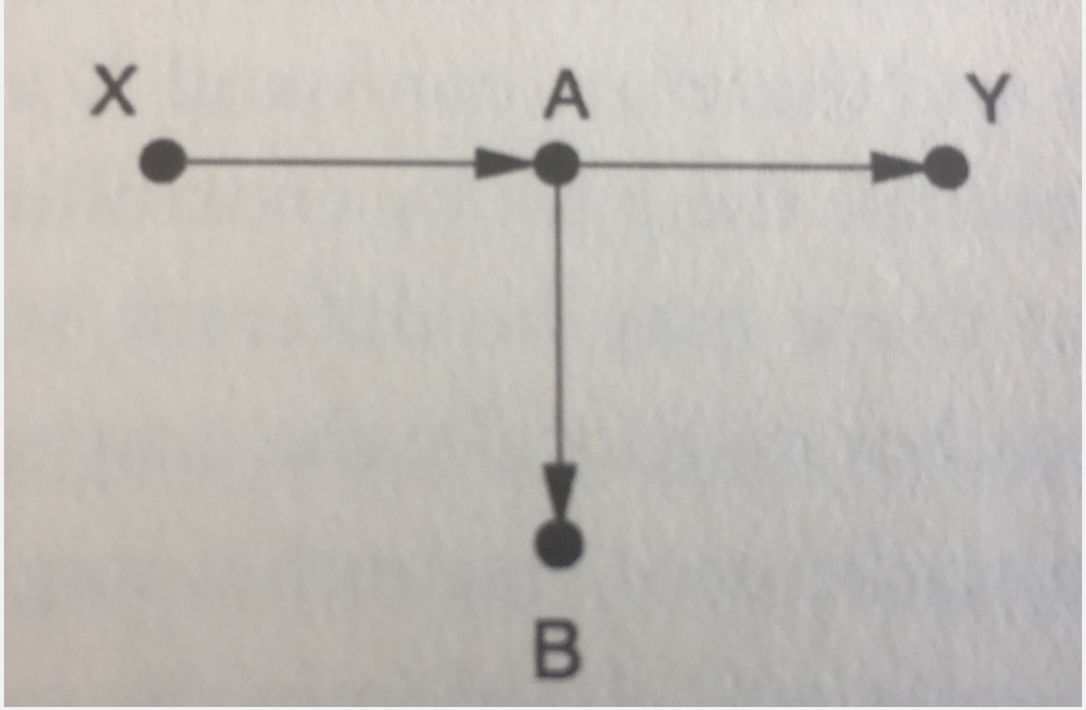


I(C, {}, D)?
I(C, {A}, D)?
I(C, {A, B}, D)?
I(C, {A, B, J}, D)?

Confounding through the lens of causal calculus

- Confounder is any factor that leads makes $P(Y|X) \neq P(Y|do(X))$
- To de-confound two variables X and Y
  - We need to block all non-causal paths between X and Y without perturbing any causal paths
    - A backdoor path is any path from X to Y that starts with an arrow pointing into X
    - X and Y will be de-confounded if we block every such backdoor path
    - If we do this by controlling for some variables Z, we need to make sure that no member of Z is a descendent of X on a causal path
  - That is all there is to it!

Confounding through the lens of causal calculus



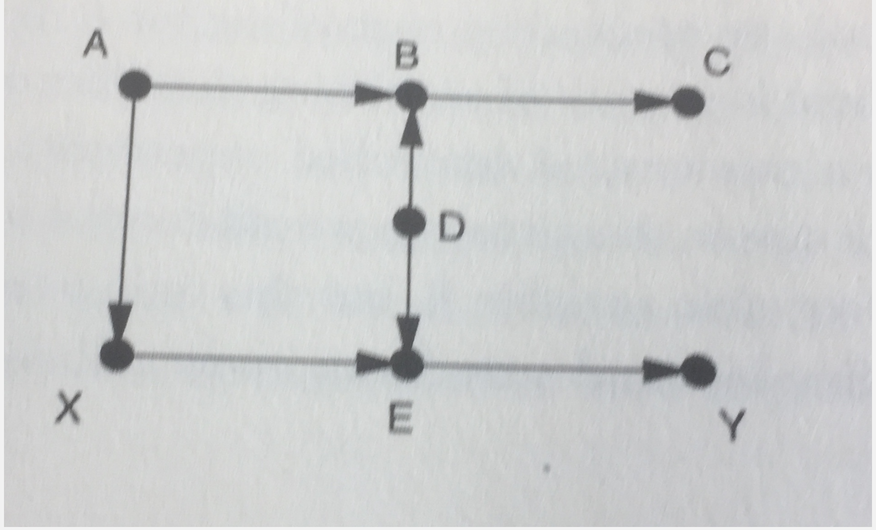
What do we need to control for in order to de-confound X and Y?

- B passes a classical epidemiological definition of confounding
- But if we control for B, we introduce confounding rather than eliminating it!

Confounding through the lens of causal calculus

- If we can identify and measure the confounders, we can control for them
- But as Pearl's work has shown, standard criteria for identifying confounders are flawed
- Both false positive and false negative confounders can yield misleading conclusions
- Causal calculus and tools based on graph theoretic criteria like d-separation provide effective methods for identifying the confounders (and only the confounders)

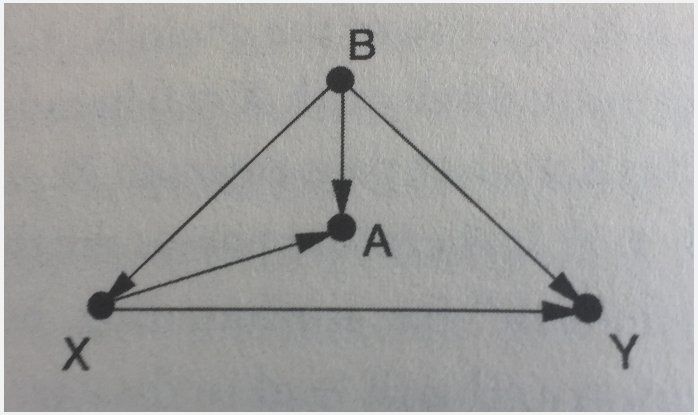
Confounding through the lens of causal calculus



What do we need to control for in order to de-confound X and Y?

- A, B, C, D are pre-treatment variables, X is the treatment
- The only backdoor path X ←A → B ←D→ E→Y is already blocked by the collider B, so no need to control for anything!
- Standard statistical practice would be to control for B and C
  - "To avoid conditioning on some observed covariates … is scientific ad hockery"
  - Controlling for B and C introduces confounding (unless we control for A or D as well)

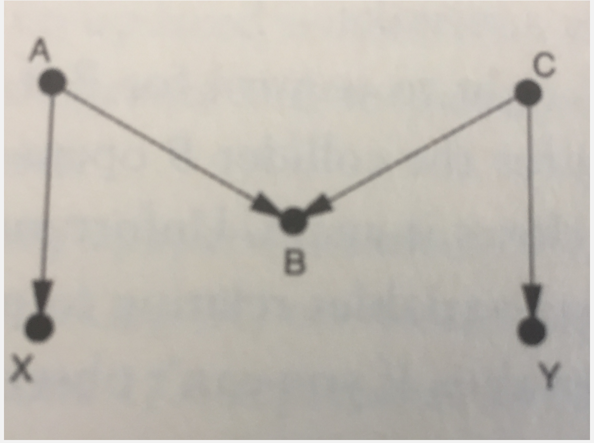
Confounding through the lens of causal calculus



What do we need to control for in order to de-confound X and Y?

- There is a backdoor path X ←B →Y
- We can block it only by blocking B
- If B is observable, we are all set
- If B is unobservable
  - We cannot control for it, so there is no way we can de-confound X and Y, so there is no way to estimate the causal effect of X on Y without running a RCT
  - Current statistical practice would advocate controlling for A, a proxy of B – but this only partially eliminates the confounding bias and introduces a collider biasl

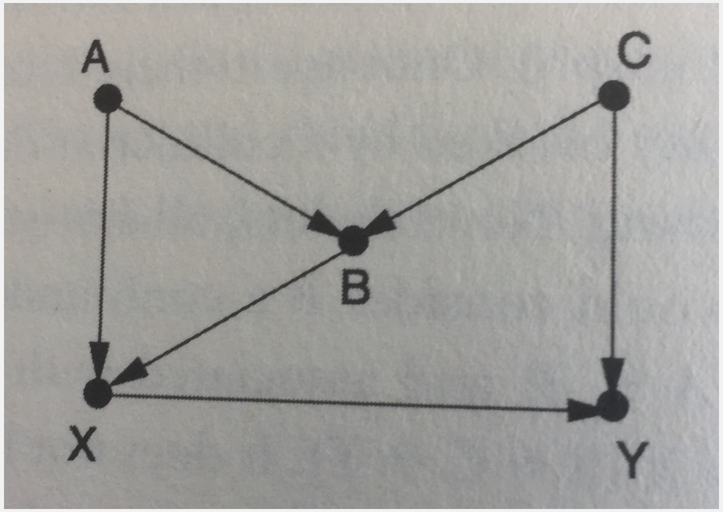
Confounding through the lens of causal calculus



What do we need to control for in order to de-confound X and Y?

- There is a backdoor path X ←A →B ←C → Y which is already blocked by B
- Some of the correlation based statistical definitions of confounding would identify B as a confounder!
- B becomes a confounder when we control for it!
- Example
  - B – Seatbelt use, X – Smoking, A – Attitude towards societal norms, C – Attitude towards safety and health related measures, Y – lung cancer
  - A 2006 study found B to be correlated with both X and Y

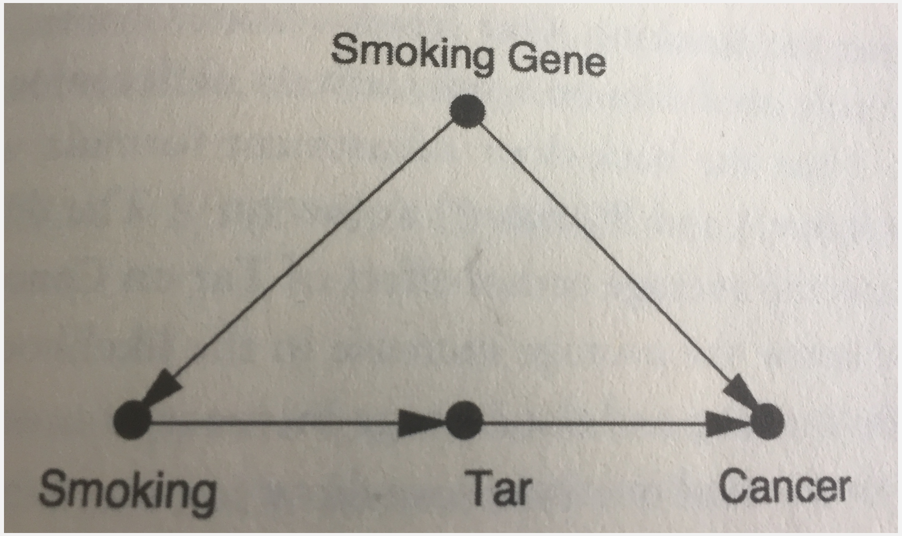
Confounding through the lens of causal calculus



What do we need to control for in order to de-confound X and Y?

- There is a backdoor path X ←A →B ←C → Y which is already blocked by B
- There is a second backdoor path X ←B ←C →Y
  - If we control B to block this path, we need to block A and C as well to ensure that the first backdoor path does not get unblocked
  - But blocking C alone suffices to block both the paths

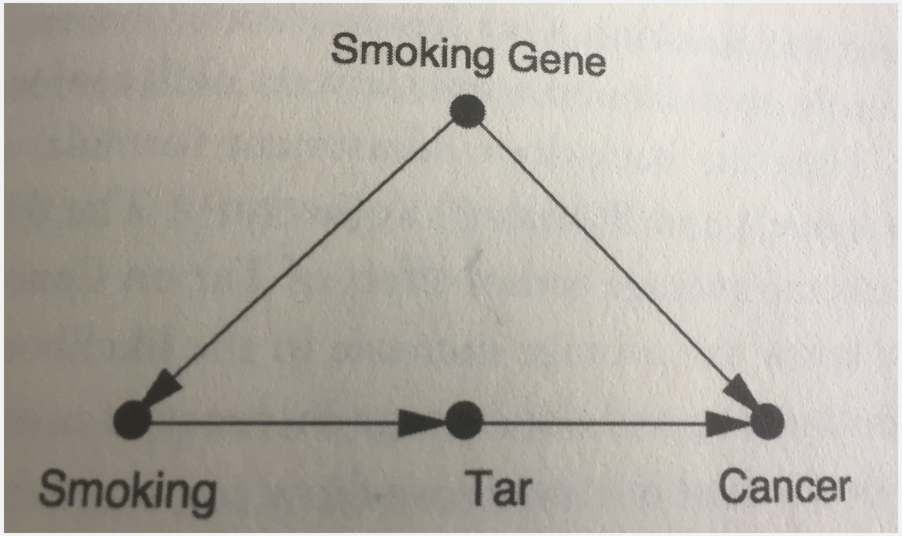
Knocking on the front door



- Observational study collects data about Smoking, Lung Tar, and Cancer
- Suppose we want to rule out the possibility that a "smoking gene" is responsible for both the tendency to smoke and susceptibility to cancer

- Lacking data on Smoking Gene, we cannot block the backdoor path Smoking ←Smoking Gene → Cancer
- We try the front door Smoking → Tar → Cancer
- Since we have data on all three variables on the front door path, suppose we estimate the effect of Smoking on Tar and Tar on Cancer and from these, obtain the effect of Smoking on Cancer
- The Smoking ←Smoking Gene →Cancer ←Tar path is blocked by the collider Cancer – there is no need for backdoor adjustment!
- We can simply observe $P\left(Tar\,|\,Smoking\right)$and $\mathrm{P}\left(Cancer\,|\,Tar\right)$ and take their difference to compute the causal effect of Smoking on Tar

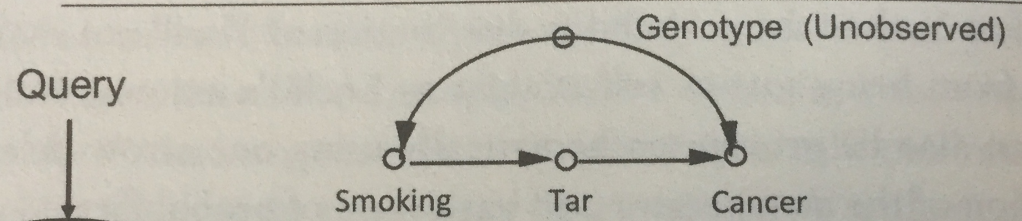
Knocking on the front door



- Observational study collects data about Smoking, Lung Tar, and Cancer
- Suppose we want to rule out the possibility that a "smoking gene" is responsible for both the tendency to smoke and susceptibility to cancer

- Lacking data on Smoking Gene, we cannot block the backdoor path Smoking ←Smoking Gene → Cancer
- We try the front door Smoking → Tar → Cancer
- Since we have data on all three variables on the front door path, suppose we estimate the effect of Smoking on Tar and Tar on Cancer and from these, obtain the effect of Smoking on Cancer
- The Smoking ←Smoking Gene →Cancer ←Tar path is blocked by the collider Cancer – there is no need for backdoor adjustment!
- We can simply observe $P(Tar \mid Smoking)$ and $\mathrm{P}(Cancer \mid Tar)$ and take their difference to compute the causal effect of Smoking on Tar

DO-CALCULUS AT WORK

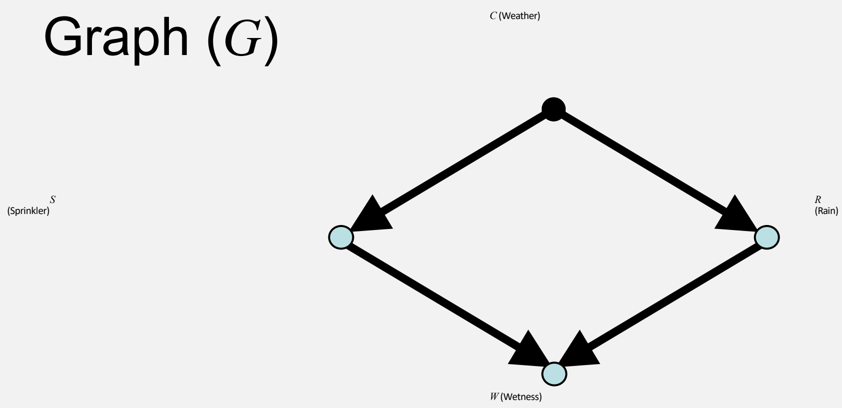


$P(c \mid do(s)) = \sum_t P(c \mid do(s), t) P(t \mid do(s))$	Probability Axioms
$= \sum_t P(c \mid do(s), do(t)) P(t \mid do(s))$	Rule 2
$= \sum_t P(c \mid do(s), do(t)) P(t \mid s)$	Rule 2
$= \sum_t P(c \mid do(t) P(t \mid s)$	Rule 3
$= \sum_{s'} \sum_t P(c \mid do(t), s') P(s' \mid do(t)) P(t \mid s)$	Probability Axioms
$= \sum_{s'} \sum_t P(c \mid t, s') P(s' \mid do(t)) P(t \mid s)$	Rule 2
$= \sum_{s'} \sum_t P(c \mid t, s') P(s') P(t \mid s)$	Rule 3

Estimand

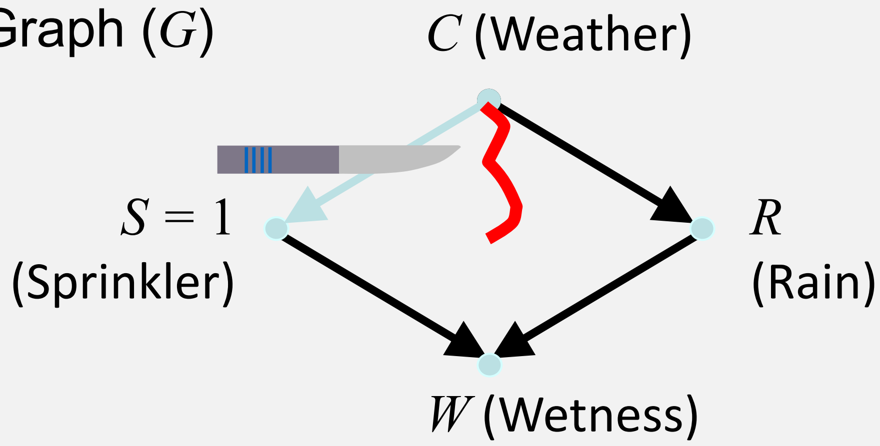
Deriving counterfactuals from a causal model

Graph (G)



Deriving counterfactuals from a causal model

Graph (G)



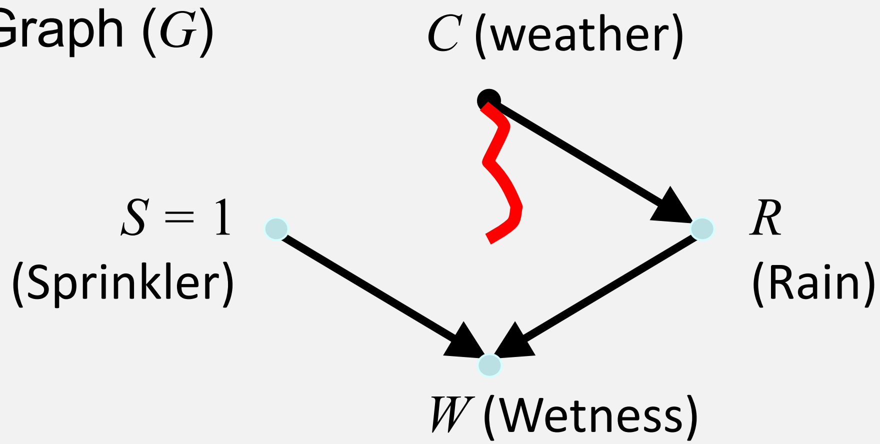
Mutilated Model ($M_{S=1}$)

What is the probability that we find the pavement is wet if we turn the sprinkler ON?

Find if $P(W_{S=1} = 1) = P(W = 1 \mid do(S = 1))$

Deriving counterfactuals from a causal model

Graph (G)

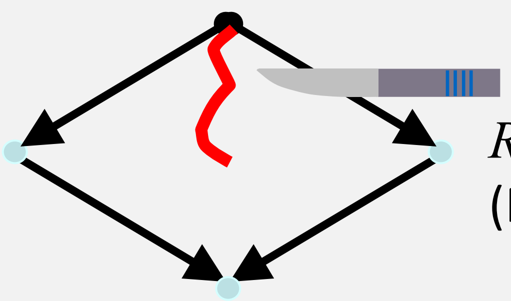


Mutilated Model ($M_{S=1}$)

Would it rain if we turn the sprinkler ON?
Not necessarily, because $R_{S=1} = R$

Deriving counterfactuals from a causal model

Graph (G)



Mutilated Model ($M_{R=1}$)

Would the pavement be wet had it rained?
Find if $W = 1$ in $M_{R=1}$

Every counterfactual question has an answer in M

Counterfactuals and algorithmic Fairness

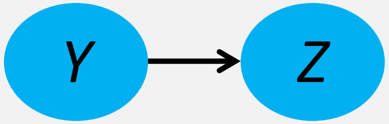
- Is there gender-based discrimination in hiring?
  - We know that Jane was not hired
  - Was she discriminated against because of gender?
- Would the hiring decision be different had the candidate been female (or male) as opposed to male (female)?
- In other word

$$P(hiring|do(gender = female)) \neq P(hiring|do(gender = male))?$$

[Kusner et al., 2017, Khademi et al., 2018]

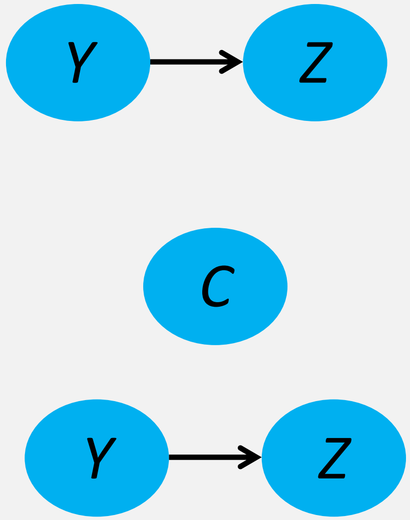
Traditional method for learning causal relationships from observational data

- Question: Does *Y* cause *Z*?



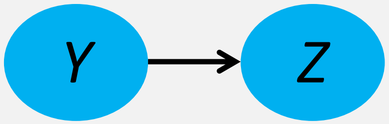
Traditional Method for Learning Causal Relationships from Observational Data

- Question: Does *Y* cause *Z*?
- Standard recipe
  - Identify all confounders of *Y* and *Z*
  - Measure them

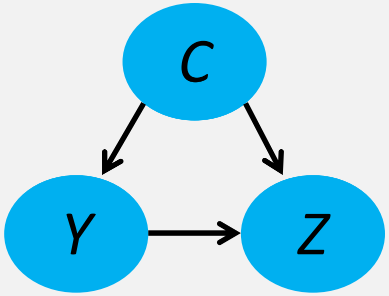


Traditional Method for Learning Causal Relationships from Observational Data

- Question: Does *Y* cause *Z*?

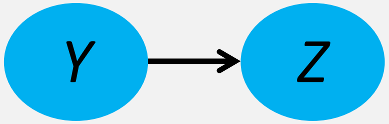


- Identify all confounders of *Y* and *Z*
- Measure them
- Condition on them

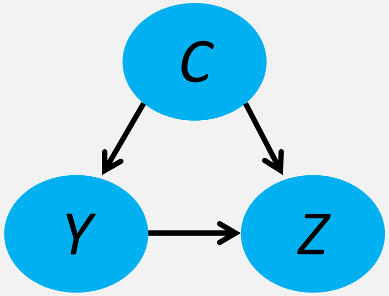


Traditional Method for Learning Causal Relationships from Observational Data

- Question: Does *Y* cause *Z*?

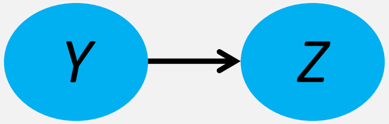


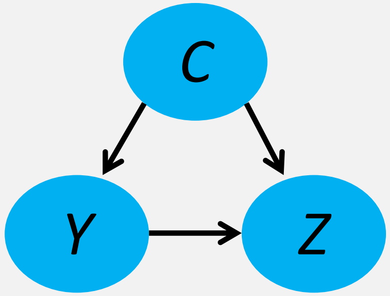
- Identify all confounders of *Y* and *Z*
- Measure them
- Condition on them

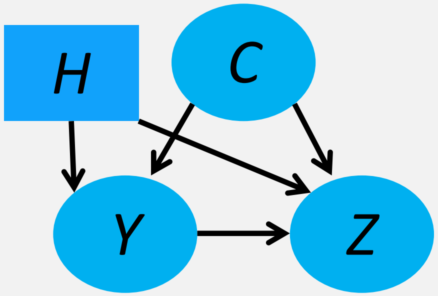


Traditional Method for Learning Causal Relationships from Observational Data

- Question: Does *Y* cause *Z*?
  - Identify all confounders of *Y* and *Z*
  - Measure them
  - Condition on them
- The recipe fails when
  - The traditional methods of identifying confounders fail to identify the confounders
  - Not all confounders are measureable







Confounding redefined using causal calculus

- Many definitions, almost all flawed
- Except the potential outcomes framework of Greenland and Robins (1986) which introduces exchangeability as a requirement for absence of confounding
  - Treatment group would behave identical to the control group if it were not treated
- Correct definition using the language of causal calculus
  - Confounder is any factor that leads makes $P(Y|X) \neq P(Y|do(X))$

Causal Network Methods for Learning Causal Relationships from Observational Data

- Constraint-based
- Bayesian
- Other

Causal Network Methods for Learning Causal Relationships from Observational Data

- Constraint-based
- Bayesian
- Other

The Constraint-Based Learning of Causal Models

1. Determine constraints that hold among the nodes (e.g., independence conditions based on statistical tests)
2. Use the patterns of constraints to narrow the causal possibilities

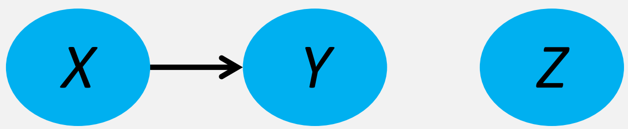
Constraint-Based Search for a Causal Model: Example

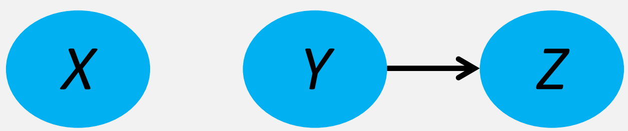
- Three binary variables *X*, *Y*, *Z*
- Suppose time ordering is known (we can relax this condition):

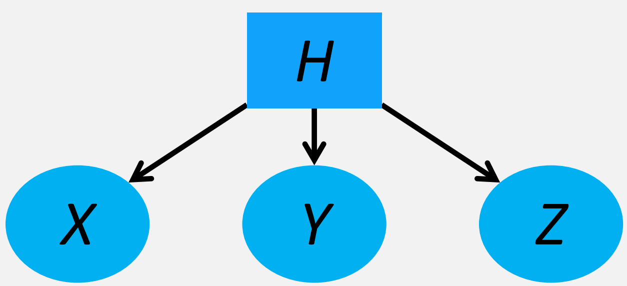
  X occurs before *Y* which occurs before *Z*
- For instance
  - *X*: economic circumstances
  - *Y*: environmental risk
  - *Z*: disease
- Question: Does *Y* cause *Z*?

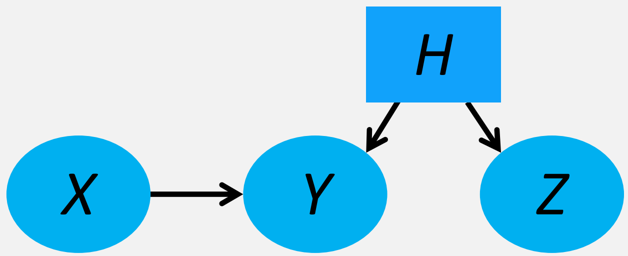
Constraint-Based Search for a Causal Model: Example

- Suppose statistical testing yields the following constraints
  dep(X, Y), dep(Y, Z), dep(X, Z), ind(X, Z | Y)
- Consider the consistency of these constraints with respect to the following causal models:

X


X


X




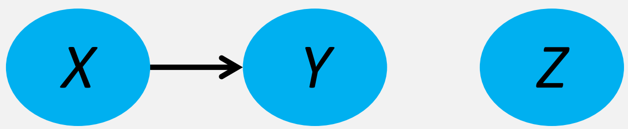
Constraint-Based Search for a Causal Model: Example

- Suppose statistical testing yields the following constraints

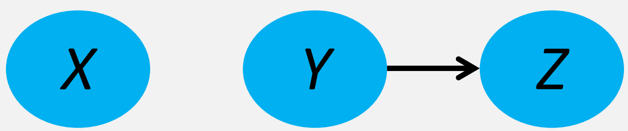
  dep(*X*, *Y*), dep(*Y*, *Z*), dep(*X*, *Z*), ind(*X*, *Z* | *Y*)

- Consider the consistency of these constraints with respect to the following causal models:

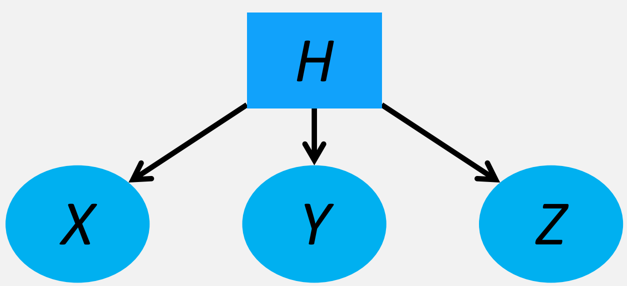
X



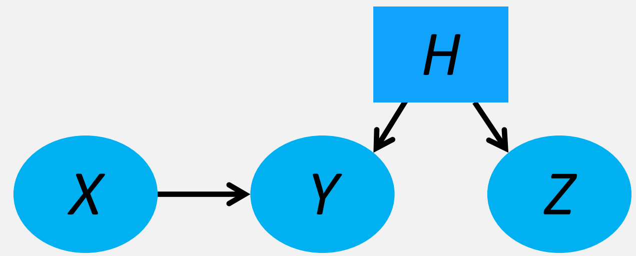
X

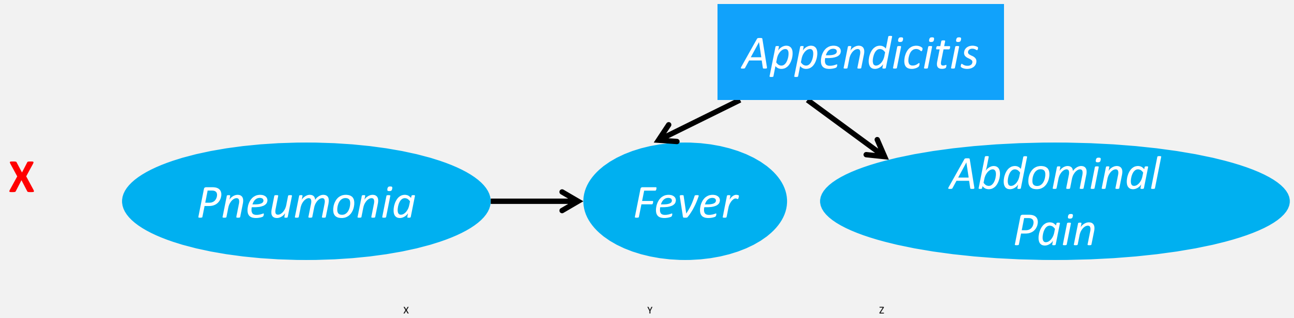


X



X





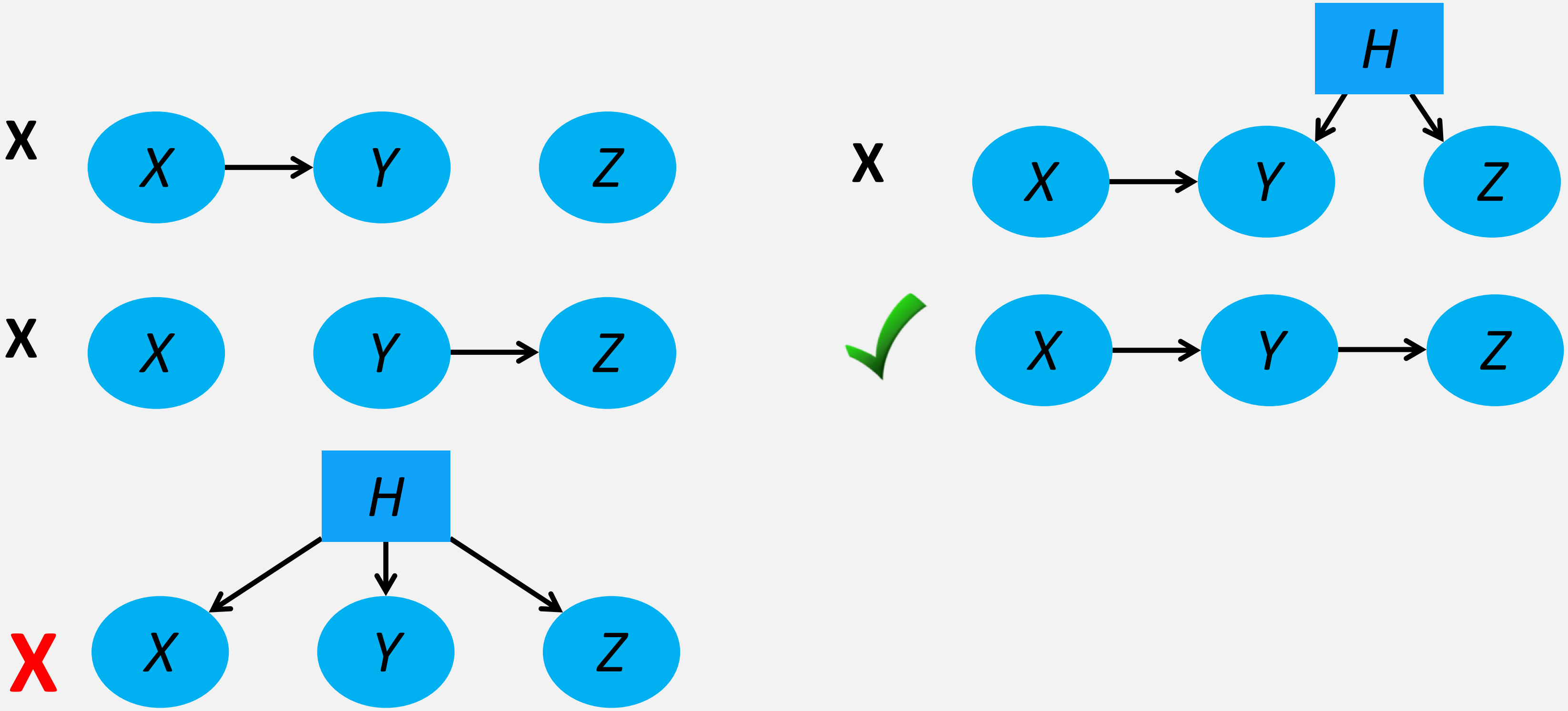
- Given Fever = present, if Pneumonia = present then Appendicitis Is unlikely and therefore Abdominal Pain is unlikely.
- Given Fever = present, if Pnemonia = absent then Appendicitis is likely and therefore Abdominal Pain is likely.
- Thus, when fever is present, Pneumonia and Abdominal Pain have an inverse statistical relationship
- This causal model is not consistent with the known constraint ind(X, Z | Y).
- The pneumonia story is more complicated because pneumonia does lead to toxemia which leads to abdominal pain

Constraint-Based Search for a Causal Model: Example

- Suppose statistical testing yields the following constraints

  dep(*X*, *Y*), dep(*Y*, *Z*), dep(*X*, *Z*), ind(*X*, *Z* | *Y*)

- Consider the consistency of these constraints with respect to the following causal models:

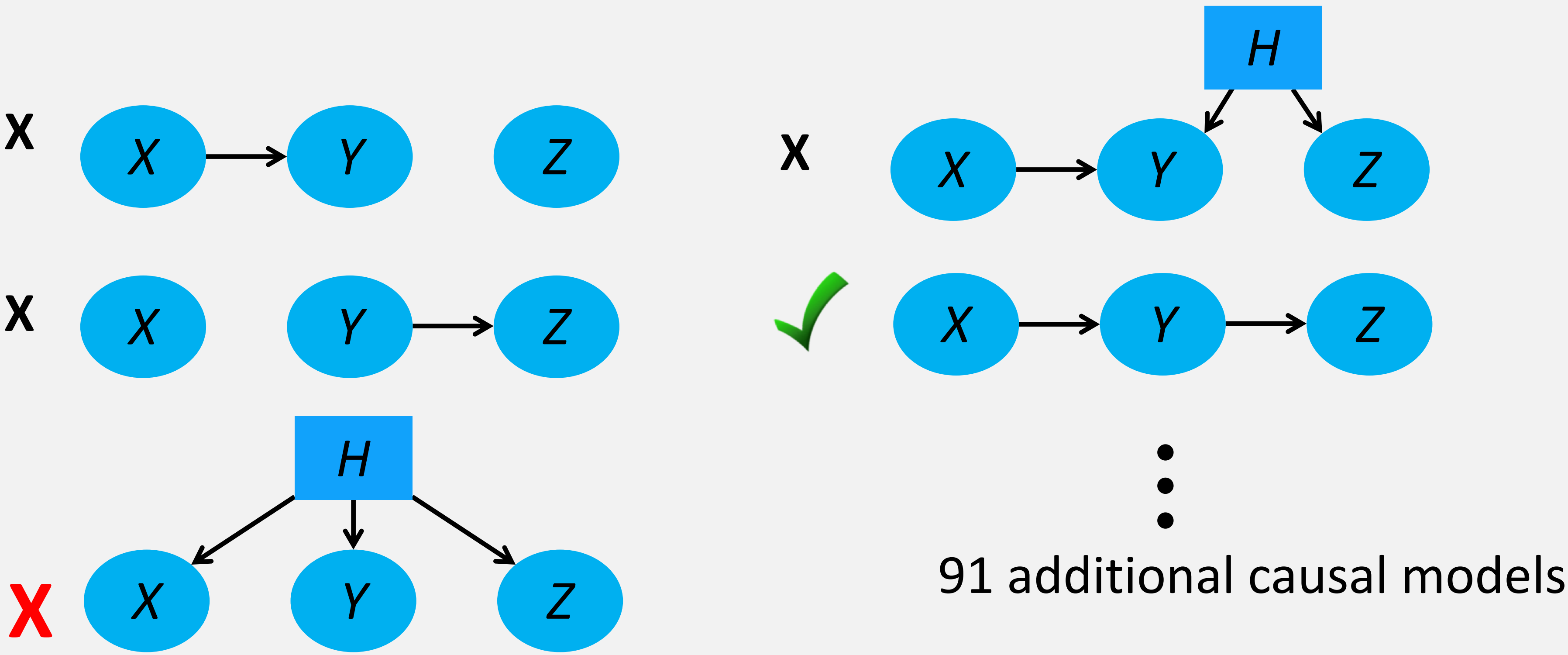


Constraint-Based Search for a Causal Model: Example

- Suppose statistical testing yields the following constraints

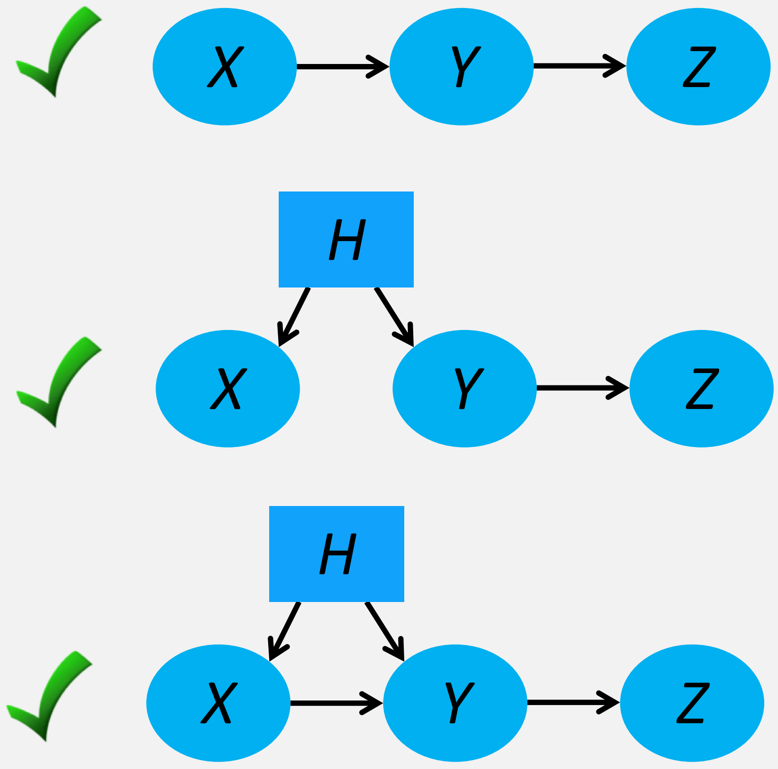
  dep(*X*, *Y*), dep(*Y*, *Z*), dep(*X*, *Z*), ind(*X*, *Z* | *Y*)

- Consider the consistency of these constraints with respect to the following causal models:



Constraint-Based Search for a Causal Model: Example

- The following models are the only ones consistent with the constraints:



- In all of these models, *Y* causes *Z* and there is no confounding of *Y* and *Z*.

Constraint-Based Search for a Causal Model: Example

- Reduce the large number of causal model possibilities to just those models consistent with the constraints obtained from the data
- Look for causal relationships that are invariant across those models (e.g., Y → Z).

Constraint-Based Causal Discovery Algorithms

- They find general patterns of statistical dependency among the measured variables that are consistent with the causal models that they output
- They make the following assumptions:
  - Causal Markov Condition: Causality is local.

Constraint-Based Causal Discovery Algorithms

- They find general patterns of statistical dependency among the measured variables that are consistent with the causal networks that they output
- They make the following assumptions:
  - Causal Markov Condition: A node is independent of its non-effects given it direct causes. A → B → C
  - Causal Faithfulness Condition: The only independence among nodes is due to the Causal Markov Condition.
  - Test accuracy: The tests of statistical independence are correct.

What else can we do with causal models?

- Causal transportability[1]
  - Suppose we have run a study in Chicago and learned a causal relationship, say between poverty and obesity
  - Suppose we want to see if the relationship is true in some form in Los Angeles
    - Los Angeles is different from Chicago in some respects, e.g., demographics
- We now have tools to answer if the causal relationship which we learned from a study in Chicago can be tweaked in some way so that it applies to Los Angeles

[1] Bareinboim and Pearl, 2012; Lee and Honavar, 2013, Bareinboim, Lee, Honavar, and Pearl, 2013

Beyond simple cases

- Learning Causal Models from Relational Data[1]
  - The methods considered so far assume that the observations are independent and identically distributed
  - Not true in the case when individuals are connected, e.g., through family relationship
- We now have tools to learn causal models from relational data
- We will soon have tools to learn causal models from temporal relational data
- We will soon have tools for counterfactual inference from observational data

[1]Lee and Honavar, 2016, Lee and Honavar, 2017a; Lee and Honavar, 2017b

Key lessons in causality

1. Correlation ≠ causation
2. Causality is more - not less - important in the era of big data
3. Every causal inference task must rely on assumptions beyond the information supplied by data
4. We have ways of encoding those assumptions mathematically and testing their implications
5. We have a mathematical machinery to combine the assumptions with data to answer to questions of interest
6. We now have a way of doing (2) and (3) in a language that permits us to judge the scientific plausibility of our assumptions and to derive their ramifications
7. Items (2)-(4) make causal inference manageable and useful

Summary

- Causal models play a pivotal role in science
- We now have tools to learn and reason with causal models from observations and experiments
- We ought to use such tools to advance discovery
  - By answering associational, interventional, and counterfactual questions
  - By integrating experimental and observational data
  - By optimizing experiments…
- Across a variety of domains including
  - Biomedical Sciences
  - Social sciences
  - Environmental sciences ..