Resources mainly from Brady Neal's "Introduction to Causal Inference".
Additional reference from: Marcelo Coca Perraillon's "Week 2: Causal Inference" and David Rawlinson's "An introduction to Causal Inference with Python"

Towards the Frond-door Adjustment in Causal Learning

Wed. reading group

Presenter: Yong-Min Shin
jordan7186.github.io

10th Jan. 2024

수학계산학부(계산과학공학)
School of Mathematics and Computing
(Computational Science and Engineering)
MIDaS LAB
Machine Intelligence & Data Science

Concepts & Important Topics

1. *Recap: Why does causal inference matter?*
    a. *What is causal inference?*
    b. *Simpson's paradox*
    c. *Confounders*

Recap

---

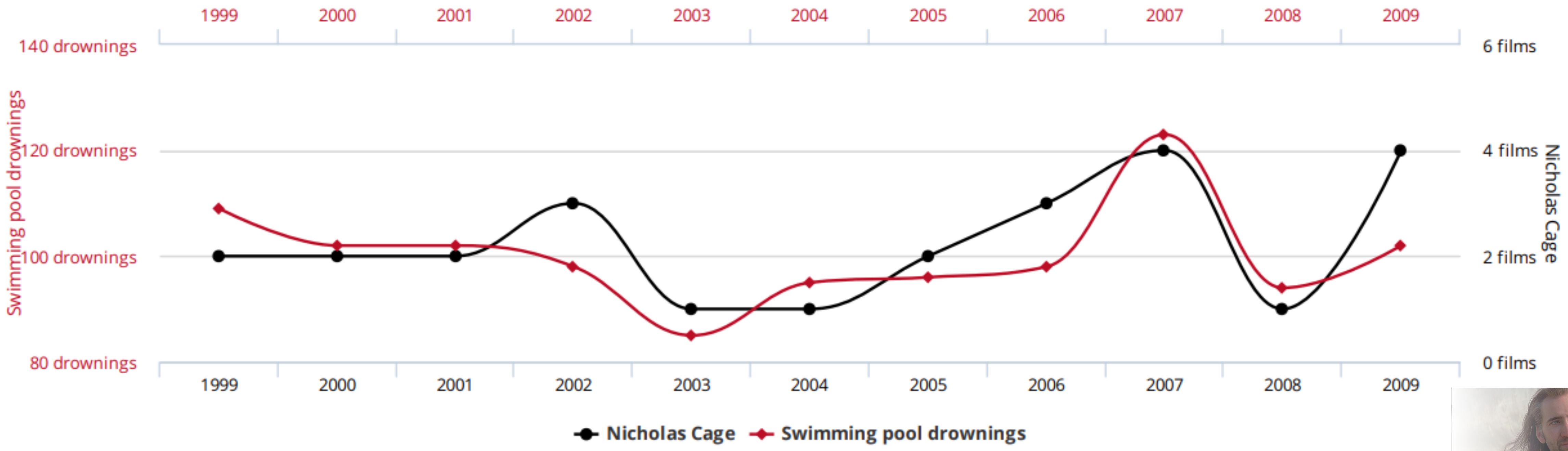
2. *Controlling counfounding factors with causal graphs*
    a. *Basics of causal graphs*
    b. *d-separation imples association is causation*
    c. *The do-operator*
    d. *The front-door adjustment* ← Eventual goal

Why does causal inference matter?

What is causal inference?

- Causation: Refers to the relationship between **cause** and **effect**.
  - "A happened because of B"
  - "B happened, therefore A has happened as a result"
  - Heavily used in economics, medical research, and recently, **machine learning**.

- "Correlation does not imply Causation"
  - Critical difference between statistic association and causal association.
  - Example data: Nicholas Cage vs. Swimming pool drownings
  - *Did Nicholas Cage **cause** the national swimming pool drowning pandemic?*



Why does causal inference matter?

Simpson's paradox

- Knowing the causal structure of the data provides a deep understanding of the problem.
- Example dataset: Administering a drug to cure a patient

Table: Rate of death in patients after drug administeration.

Treatment =
What type of drug? (A or B)

Treatment	Condition: Mild	Severe	Total
A	15% (210/1400)	30% (30/100)	**16%** (240/1500)
B	**10%** (5/50)	**20%** (100/500)	19% (105/550)

Looking at each conditions:
We should use drug B.

Looking at the 'total' data:
We should use drug A.

Which drug is more effective in reducing mortality rate?

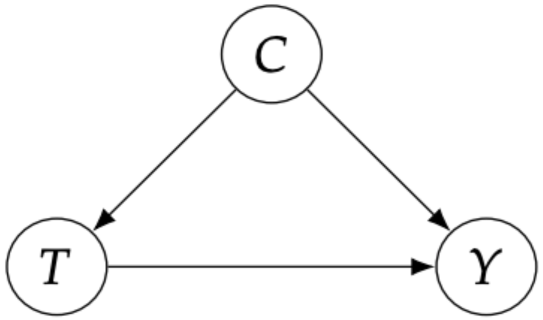
Why does causal inference matter?

Simpson's paradox

- Answer can be either A or B, **depending on the causal structure** of the problem.

C = Some cause | T = Treatment: 1 (A) or 0 (B) | Y = Mortality: 1 (live) or 0 (die)

Scenario 1

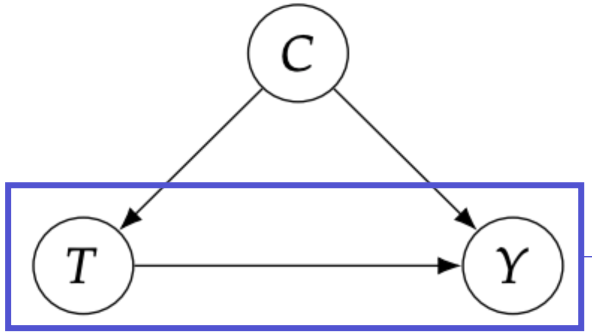


C = Condition of the patient

- Example: Doctor prescribes drug based on patient's condition.
- (C = Mild patient condition)
  - Doctor prescribes drug A ($C \rightarrow T$)
  - Mild patients usually live ($C \rightarrow Y$)
  - vice versa
- Therefore, B is more effective in treating since the patients taking A probably has a mild condition in the first place.

Why does causal inference matter?

Confounders



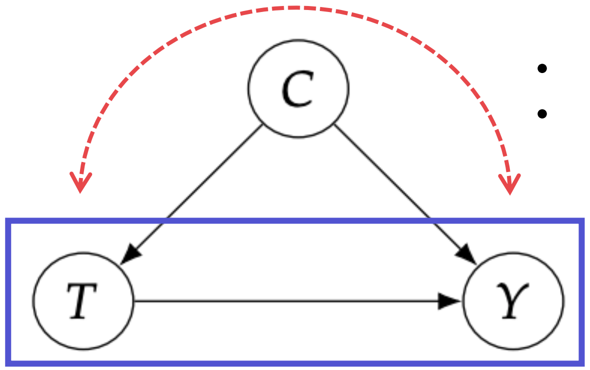
C = Condition of the patient

- Example: Doctor prescribes drug based on patient's condition.
- (C = Mild)
  - Doctor prescribes drug A ($C \rightarrow T$)
  - Mild patients usually live ($C \rightarrow Y$)
  - vice versa
- Therefore, B is more effective in treating since the patients taking A probably has a mild condition in the first place.

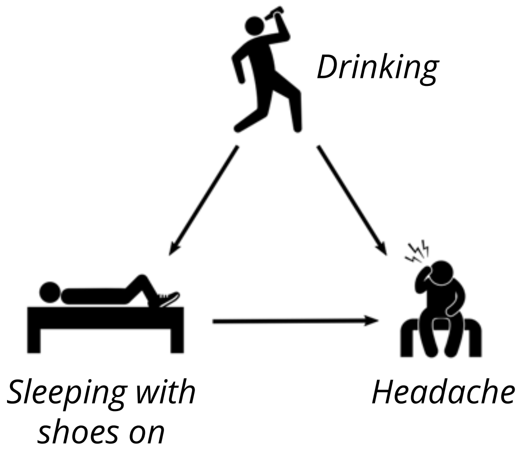
- This implies: When we only have data for T and Y, we may conclude a **causal relation** of $T \rightarrow Y$.
- However, there is **no guarantee** that is the case.
- There are definitely **statistical associations**.

Why does causal inference matter?

Confounders



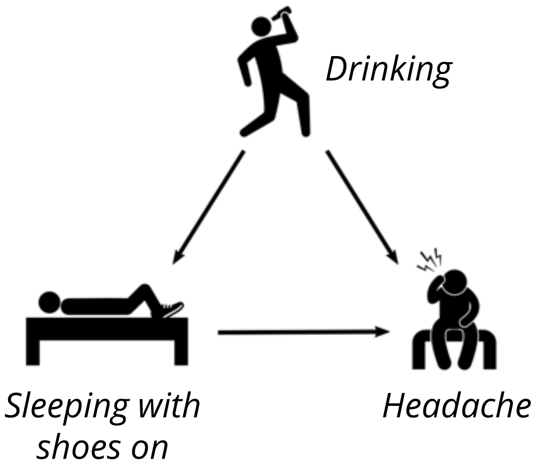
- *C* is a **confounding variable**, causing a confounding association.
- Counfounding path: $T \leftarrow C \rightarrow Y$



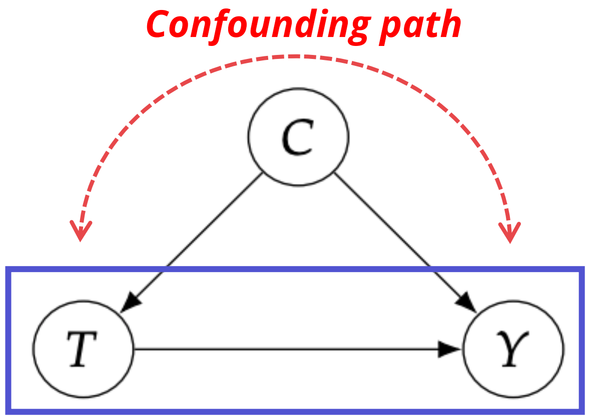
- This implies: When we only have data for *T* and *Y*, we may conclude a **causal relation** of $T \rightarrow Y$.
- However, there is **no guarantee** that is the case.
- There are definitely **statistical associations**.
- **Main problem: How can we turn causal associations into statistical associations?**

Why does causal inference matter?

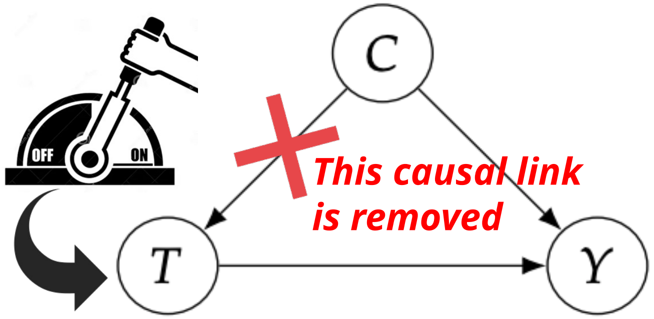
Confounders



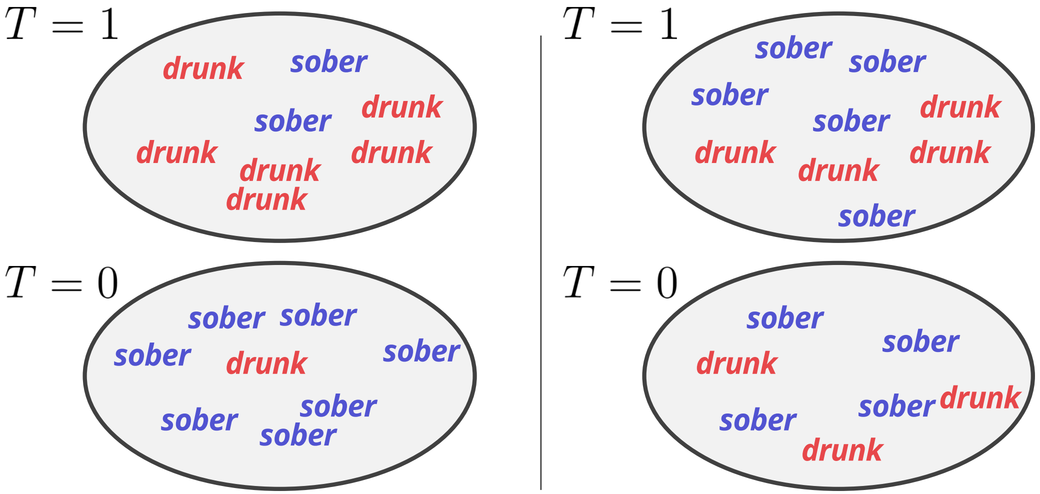
T=1: Sleep with shoes on / *T=0*: Sleep without shoes



(True) Causal association



Intervention by "forcing" subjects how to sleep



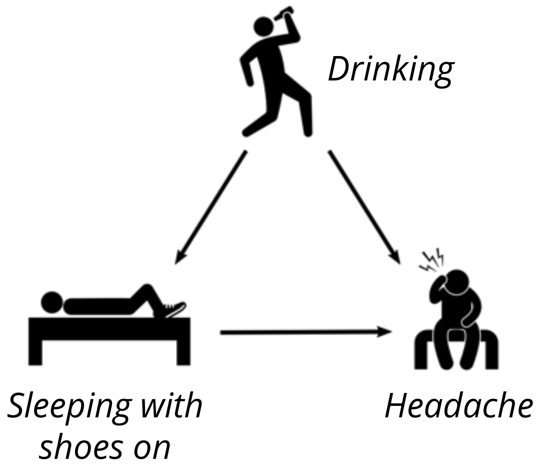
High selection bias (not a fair comparison)

No selection bias (fair comparison)

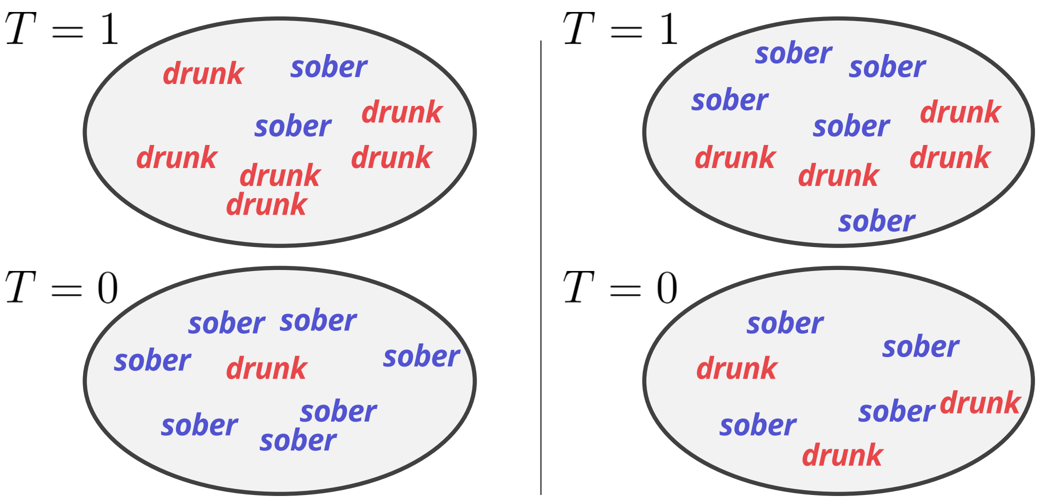
- Previous talk: Randomized Control Trials (RCTs) can eliminate the counfounding factor.
- Basic idea: If we can assume the counfounding factor has affected the two groups (T=0 & T=1) equally via randomization, we can safely compare the two groups without counfoundinig.
- Conceptually, this is the same as intervening subjects by randomly forcing them to either sleep with/without their shoes on.

Why does causal inference matter?

Confounders

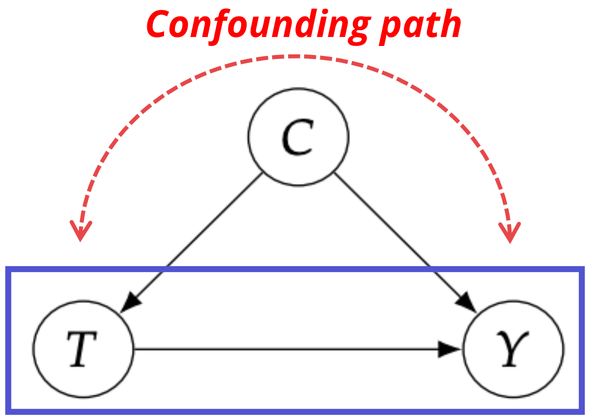


T=1: Sleep with shoes on / T=0: Sleep without shoes

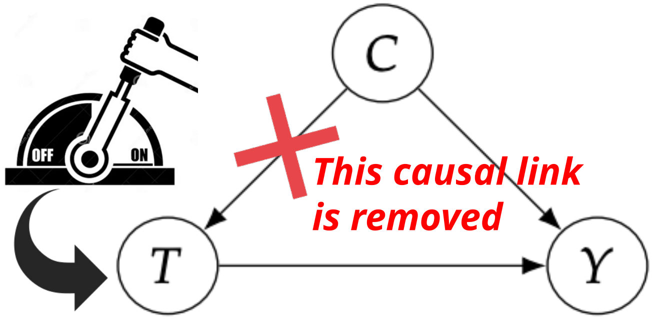


High selection bias (not a fair comparison)

No selection bias (fair comparison)



(True) Causal association



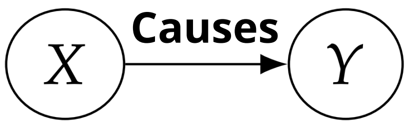
Intervention by "forcing" subjects how to sleep

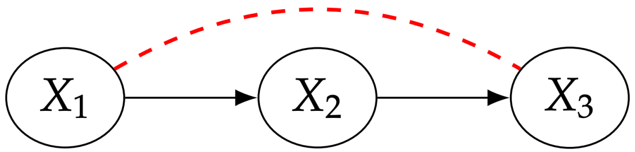
- **What if we <u>can't</u> perform RCTs?**
  - Moral issues: You can't deny a person their medications just because you want to perform RCTs
  - Highly impractical: You can't perform multiple RCTs to test your nation-level policies
- **Is there a better way to get rid of the counfounding factor?**

Controlling confounding factors with causal graphs

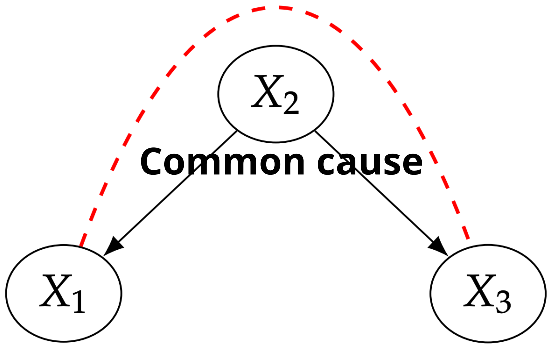
Basics of causal graphs: Chains & Forks

Flow of association
(*Statistical dependence*)





Chain



Fork

- (Causality in graphs) A variable ***X*** is said to be a **cause** of a variable ***Y*** if *Y* can change in response to changes in *X* (i.e., *Y* 'listens' to *X*)

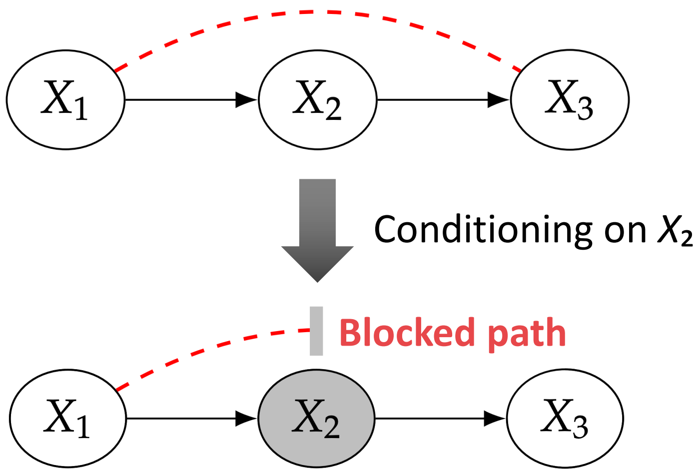
- X_1 and X_3 are usually statistically dependent.
- X_1 changes X_2, which changes X_3.

- <u>X_1 and X_3 are usually statistically dependent</u>.
- The same X_2 determines both X_1 and X_3 (common cause).
- We have already seen this as the confounder.

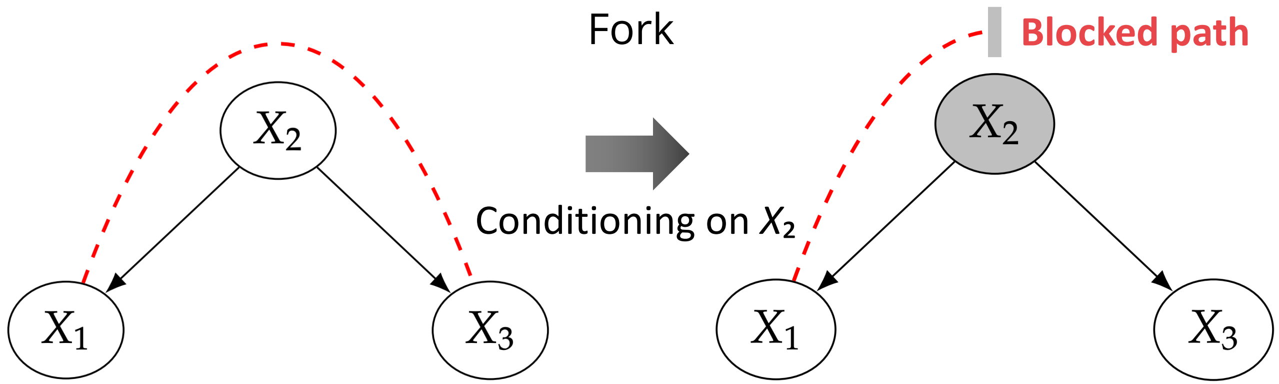
Controlling confounding factors with causal graphs

Basics of causal graphs: Conditioning

Chain



X_1 and X_3 are conditionally independent on X_2.



$$
\begin{aligned}
P(x_1, x_3 \mid x_2) &= \frac{P(x_1)\,P(x_2|x_1)\,P(x_3|x_2)}{P(x_2)} && \text{(Markov assumption \& Bayes rule)} \\
&= \frac{P(x_1, x_2)}{P(x_2)}\,P(x_3|x_2) && \text{(Bayes rule)} \\
&= P(x_1|x_2)\,P(x_3|x_2)
\end{aligned}
$$

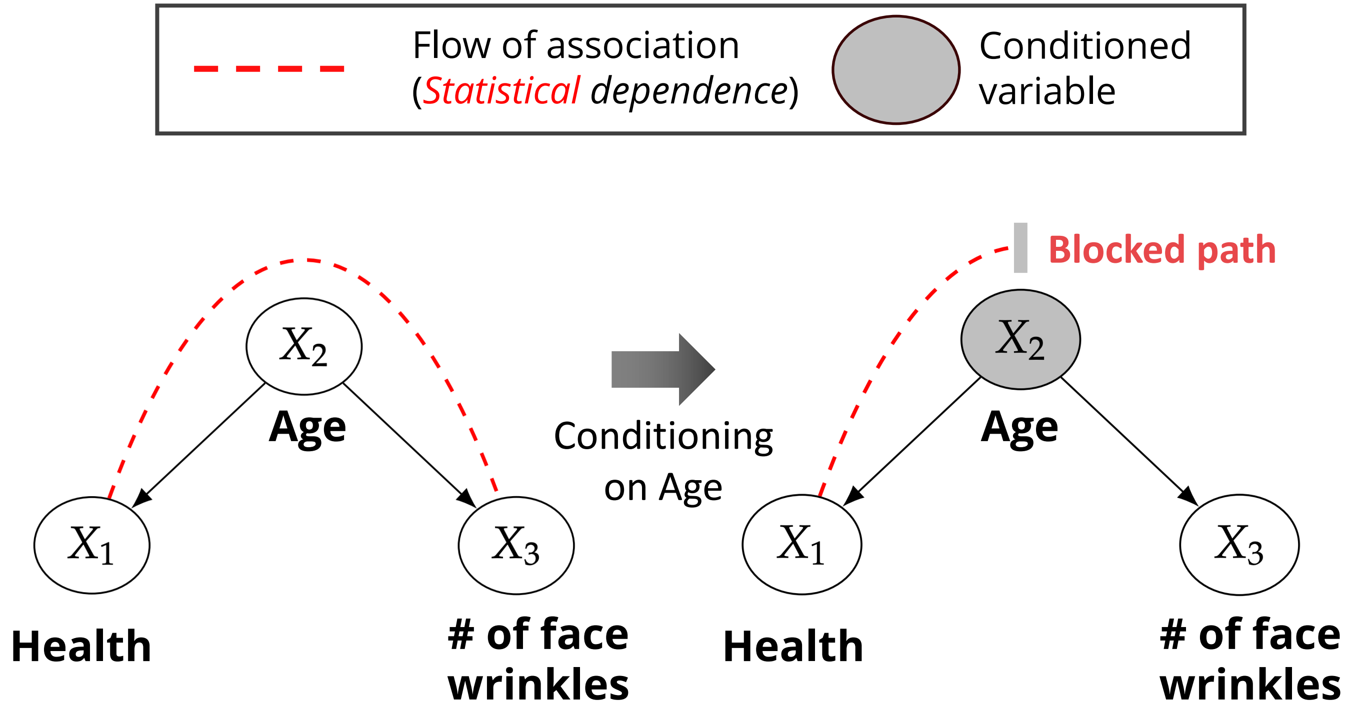
$$
\begin{aligned}
P(x_1, x_3|x_2) &= \frac{P(x_1, x_2, x_3)}{P(x_2)} \\
&= P(x_1|x_2)P(x_3|x_2)
\end{aligned}
$$

(Markov assumption & Bayes rule)

Conditioning blocks the flow of association for chains & forks.

Controlling confounding factors with causal graphs

Basics of causal graphs: Conditioning (Example)



Unconditioned

- Healthier people tend to have less wrinkles
- Non-healthy people tend to have more wrinkles
- **"Seems like"** health and wrinkes are associated!

Conditioned to elderly people

- Just looking at the elderly population, health and wrinkles have no association with each other

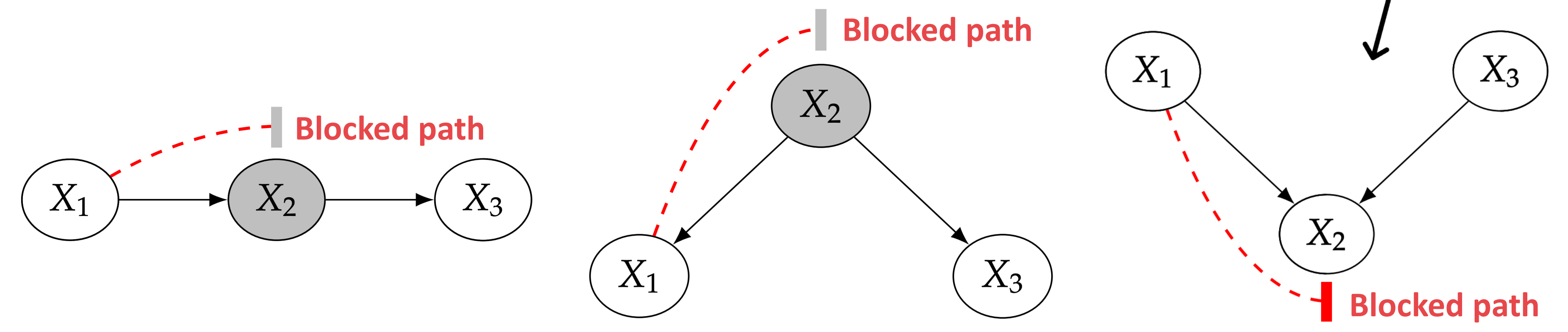
Conditioning blocks the flow of association for chains & forks.

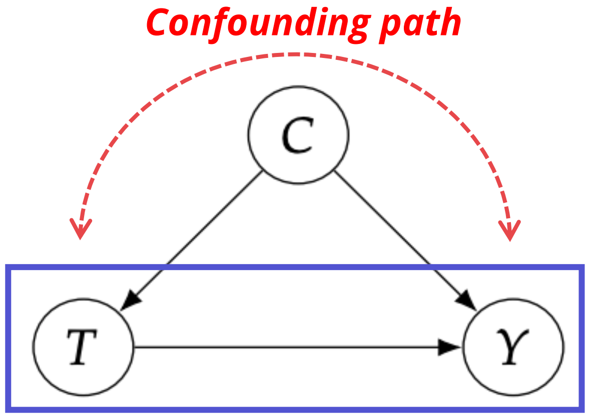
Controlling confounding factors with causal graphs

d-separation implies association is causation

X_1 and X_3 are ***d-separated**

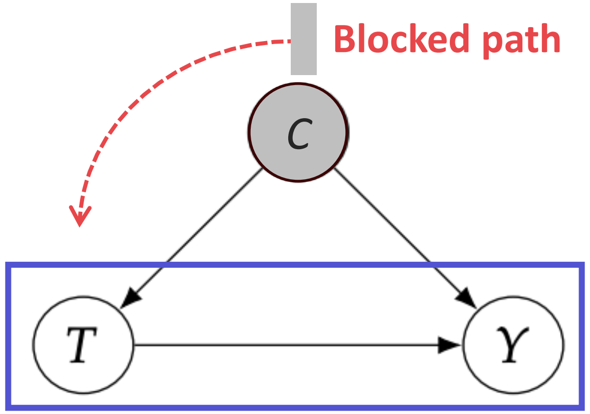
This type is called <u>immorality</u>, excluded for the sake of time





(True) Causal association

Q. How can we isolate causal association?
A. Only leave the causal association and <u>d-separate all non-causal associations</u>.

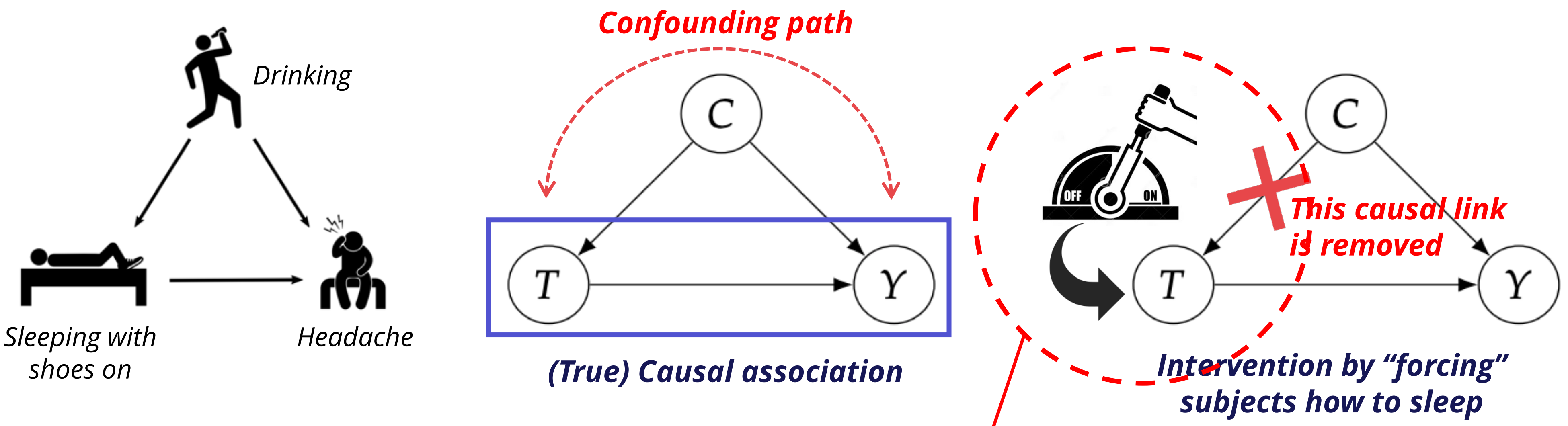


Now, association implies causation.

**d stands for dependency*

Controlling confounding factors with causal graphs

The do-operator



We have a mathematical operator for these interventions: ***do-operator***

do-operator

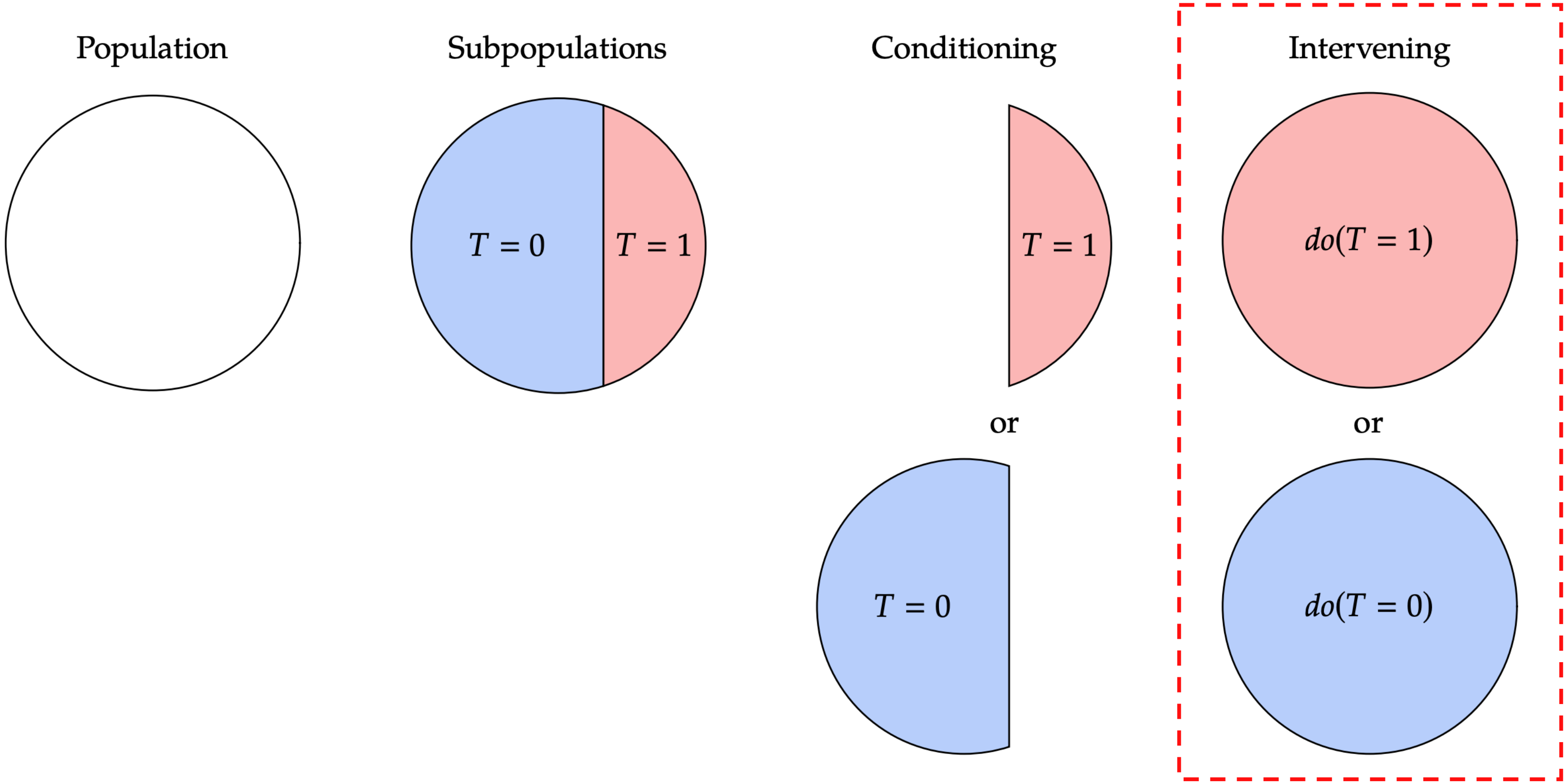
$$P(Y(t) = y) \triangleq P(Y = y \mid do(T = t)) \triangleq P(y \mid do(t))$$

Controlling confounding factors with causal graphs

The do-operator: Difference from conditioning

do-operator

$$P(Y(t) = y) \triangleq P(Y = y \mid do(T = t)) \triangleq P(y \mid do(t))$$

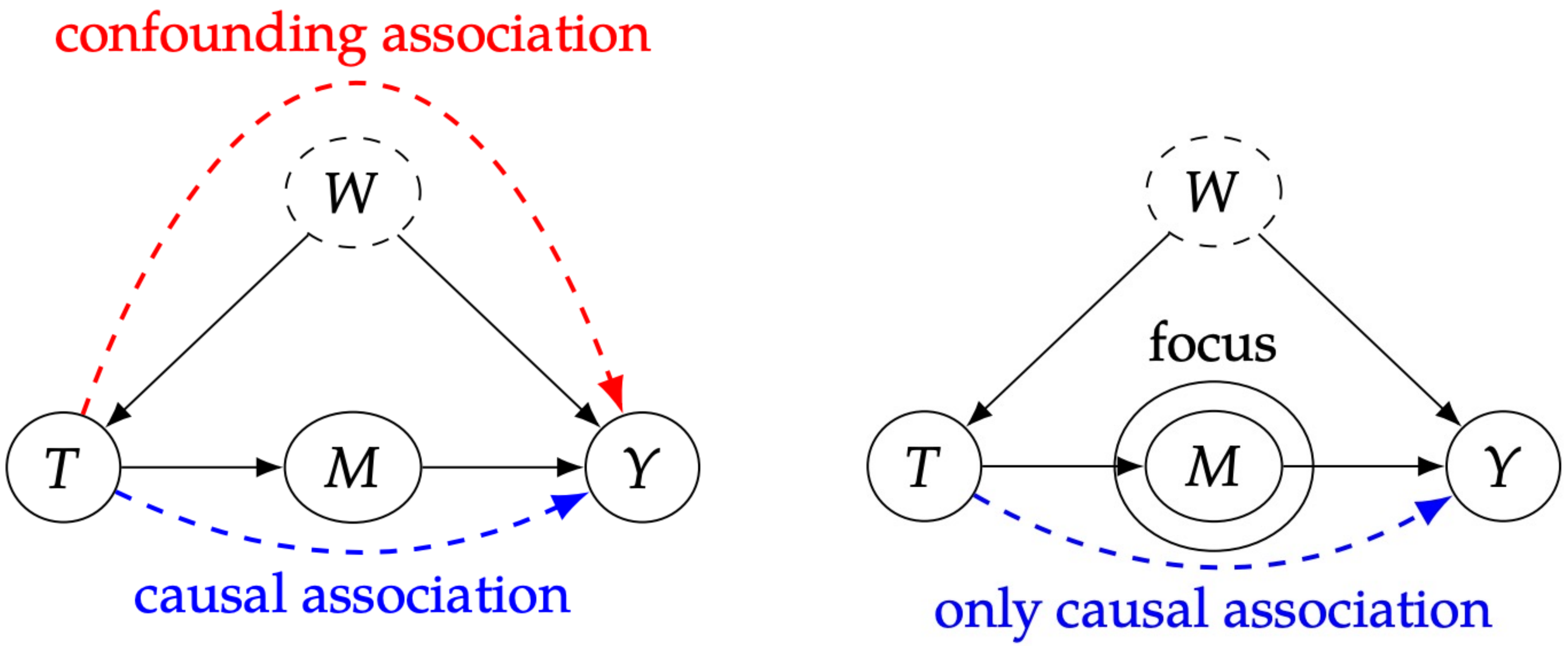


We can see the fundamental problem of causal inference here

Controlling confounding factors with causal graphs

The front-door adjustment: Alternative to RCTs (Finally!)

Problem: Can we calculate the causal association between *T* and *Y* by only focusing on *M*?



Problem setting

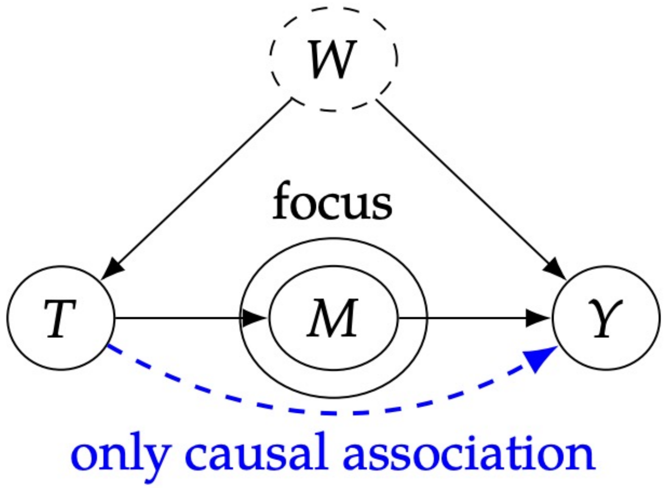
Want to do something like…

Controlling confounding factors with causal graphs

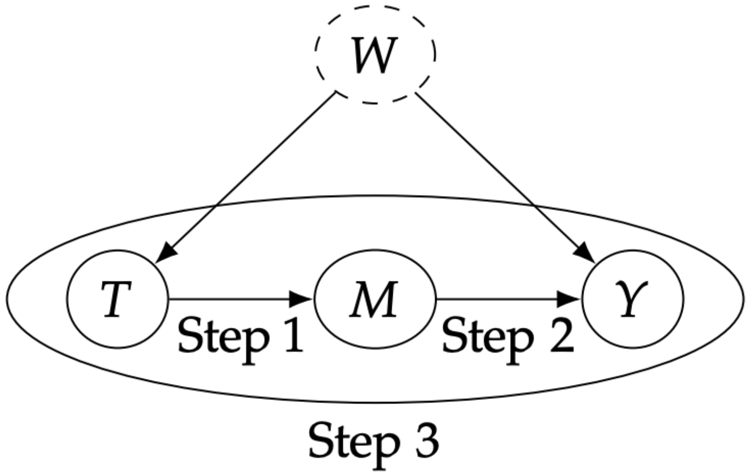
The front-door adjustment: Alternative to RCTs (Finally!)

Problem: Can we calculate the causal association between *T* and *Y* by only focusing on *M*?

"Front-door adjustment"



Intuition: Instead of trying to remove all confounding associations, ignore *W* and just focus on the mediating variable *M*.



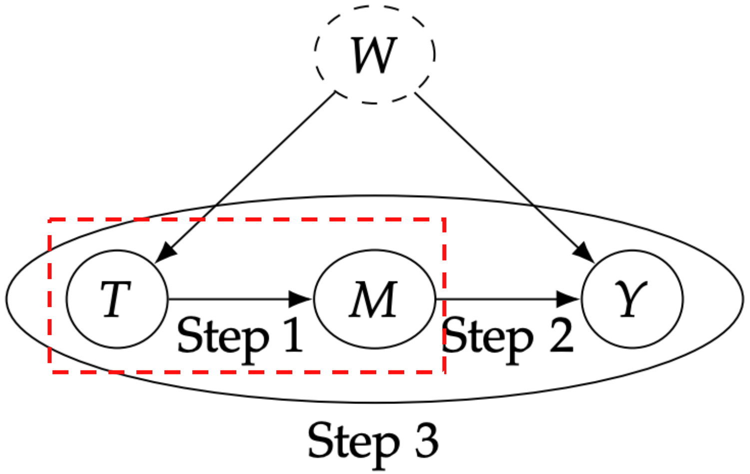
<3-steps for front-door adjustment>

1. Identify the causal effect of *T* on *M*.
2. Identify the causal effect of *M* on *Y*.
3. Combine step 1 & 2 to get the causal effect of *T* on *Y*.

Controlling confounding factors with causal graphs

The front-door adjustment: Alternative to RCTs (Finally!)

Problem: Can we calculate the causal association between *T* and *Y* by only focusing on *M*?

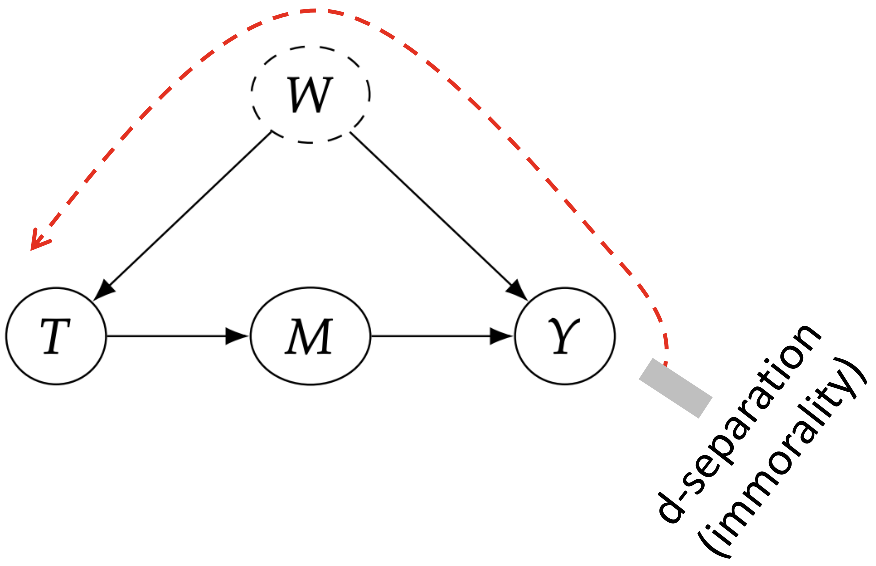


<3-steps for front-door adjustment>

1. Identify the causal effect of *T* on *M*.
2. Identify the causal effect of *M* on *Y*.
3. Combine step 1 & 2 to get the causal effect of *T* on *Y*.

It is safe to say:

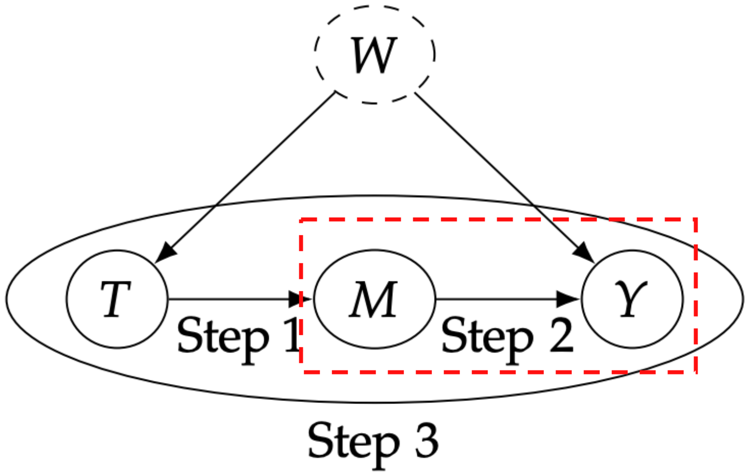
$$P(m \mid do(t)) = P(m \mid t)$$



Controlling confounding factors with causal graphs

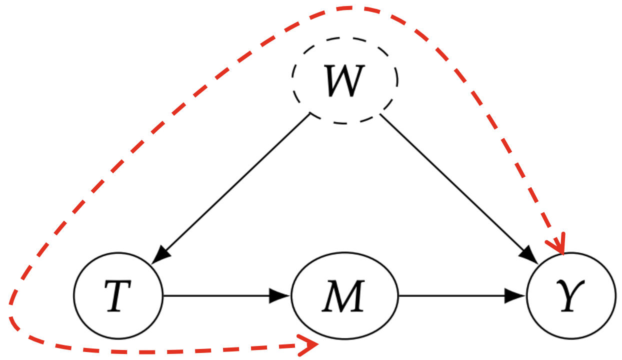
The front-door adjustment: Alternative to RCTs (Finally!)

Problem: Can we calculate the causal association between *T* and *Y* by only focusing on *M*?



<3-steps for front-door adjustment>

1. Identify the causal effect of *T* on *M*.
2. Identify the causal effect of *M* on *Y*.
3. Combine step 1 & 2 to get the causal effect of *T* on *Y*.



Unfortunatelly, *M* and *Y* are confounded (M ← T ← W → Y), so we use the *back-door adjustment formula:

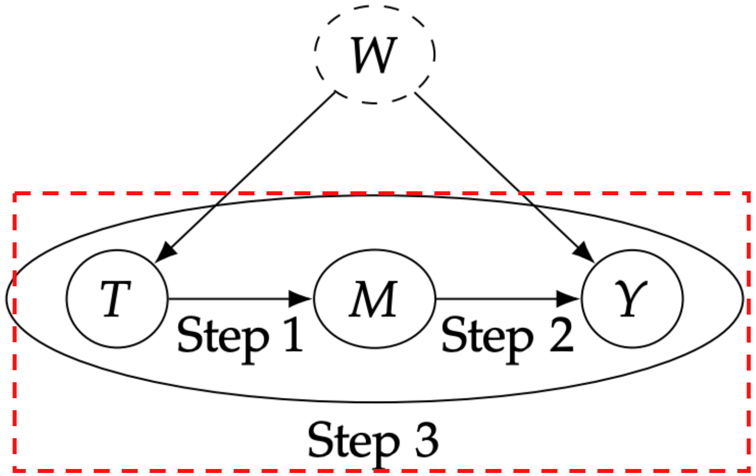
$$P(y \mid do(m)) = \sum_t P(y \mid m, t) P(t)$$

**Omitted for brevity*

Controlling confounding factors with causal graphs

The front-door adjustment: Alternative to RCTs (Finally!)

Problem: Can we calculate the causal association between *T* and *Y* by only focusing on *M*?



<3-steps for front-door adjustment>

1. Identify the causal effect of *T* on *M*.
2. Identify the causal effect of *M* on *Y*.
3. Combine step 1 & 2 to get the causal effect of *T* on *Y*.

$$P(y, |do(t)) = \sum_{m} \underbrace{P(m|t)}_{\text{(Step 1) } T \rightarrow M} \sum_{t'} \underbrace{P(y|m, t')}_{\text{(Step 2) } M \rightarrow Y} P(t')$$

*Intuition: https://probablymarcus.com/blocks/2021/11/04/some-causal-inference-intuition.html

Takeaway messages

- *Causal structure is a framework for a deeper understanding of the underlying problem*

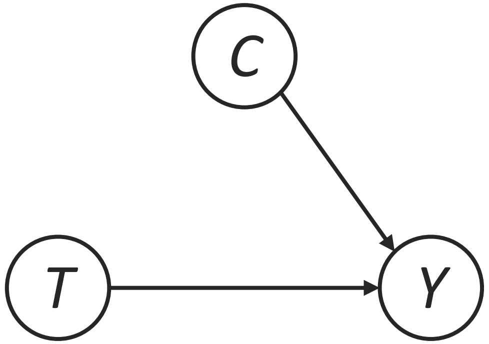
- ***Correlation does not imply Causation**, and **Randomized Control Trials (RCTs)** is the gold standard to achieve causal inference.*

- *There are other ways to identify causal effect: Back-door adjustment & **Front-door adjustment***

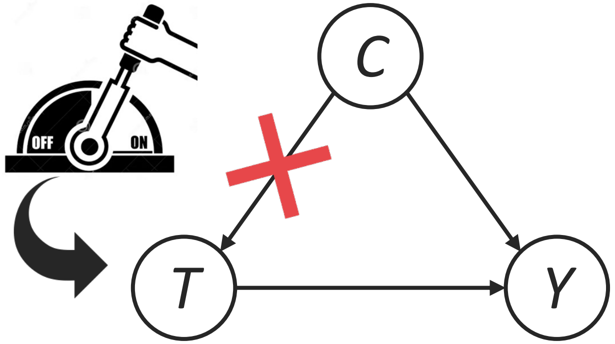
Thank you!

Appendix: Intuition for Front-door adjustment

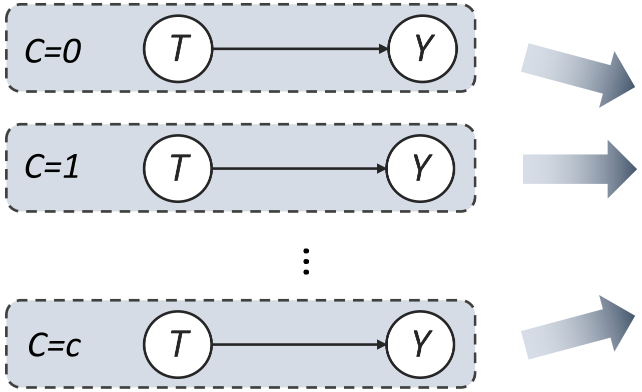
Intervention



Intervention ***cuts***
the causal link



Back-door adjustment

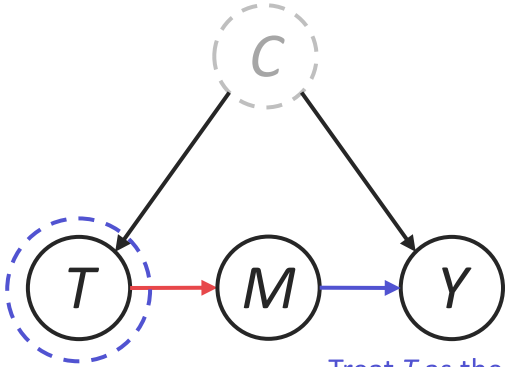


Weighted sum
(weighted by P(*C*))

Each *T* → *Y* is causal when
conditioned on *C*
**Caveat: All confounders need to be controlled this way*

$$P(y|do(t)) = \sum_c P(c)P(y|t,c)$$

Front-door adjustment



Direct causal link
(no confounders)

Treat *T* as the confounder, perform
back-door adjustment

Reference: https://probablymarcus.com/blocks/2021/11/04/some-causal-inference-intuition.html