Causal Inferences

Mentor: Kenny Zhang
Mentee: Qiguang Yan

Ladder of Causation

1. Association
What does fever and pneumonia tell me about the likelihood of COVID-19
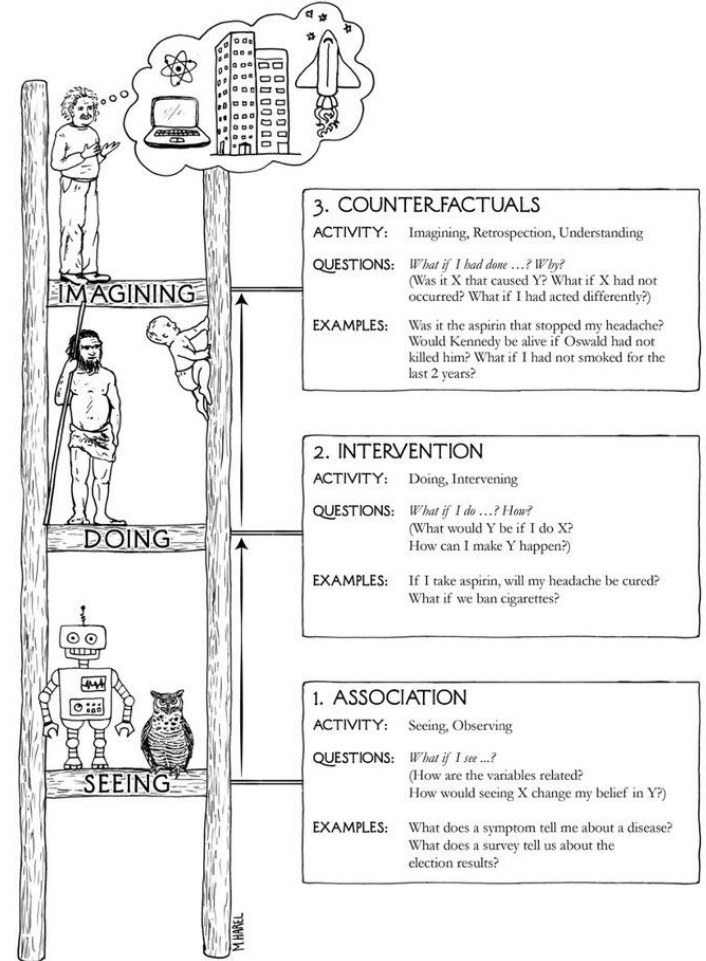
2. Intervention
If I get the vaccine, will I be protected from a disease

If I study, will I do better on the final exam

3. Counterfactuals
Was it the aspirin that stopped my headache (if I did not take the aspirin, would I still have a headache.

If I didn't smoke for the past decades, would I still be alive today

3. COUNTERFACTUALS
ACTIVITY: Imagining, Retrospection, Understanding
QUESTIONS: *What if I had done …? Why?* (Was it X that caused Y? What if X had not occurred? What if I had acted differently?)
EXAMPLES: Was it the aspirin that stopped my headache? Would Kennedy be alive if Oswald had not killed him? What if I had not smoked for the last 2 years?

2. INTERVENTION
ACTIVITY: Doing, Intervening
QUESTIONS: *What if I do …? How?* (What would Y be if I do X? How can I make Y happen?)
EXAMPLES: If I take aspirin, will my headache be cured? What if we ban cigarettes?

1. ASSOCIATION
ACTIVITY: Seeing, Observing
QUESTIONS: *What if I see …?* (How are the variables related? How would seeing X change my belief in Y?)
EXAMPLES: What does a symptom tell me about a disease? What does a survey tell us about the election results?

Causal Diagrams

A ——▶ B

A causes B
B listens to A
Assumption: Directed Acyclic Graph

Types of Junctions

Mediator:

$$X \longrightarrow A \longrightarrow Y$$

fire -> smoke -> alarm

Confounder:

$$X \longleftarrow A \longrightarrow Y$$

shoe size <- age -> reading ability

Collider:

$$X \longrightarrow A \longleftarrow Y$$

Talent -> success <- hard work

Markov equivalence

We cannot tell the difference between these junctions from observation data alone, if we control for A, X and Y are independent

Mediator:

$$X \rightarrow A \rightarrow Y$$

Mediator (reverse order):

$$X \leftarrow A \leftarrow Y$$

Confounder:

$$X \leftarrow A \rightarrow Y$$

We can find colliders from data, X and Z are correlated with Y but X and Z are independent from each other

Intervention: Randomized Control Trials

Removes the arrow from confounders to X, even if we do not know about the confounders or the confounders that cannot be measured

Ex. testing the effect of a new fertilizer on crop yields
- Soil fertility, soil texture, drainage, and many other confounders
- Randomizing eliminates selection bias and balance treatment and control with respect to confounders

Randomized Control Trials
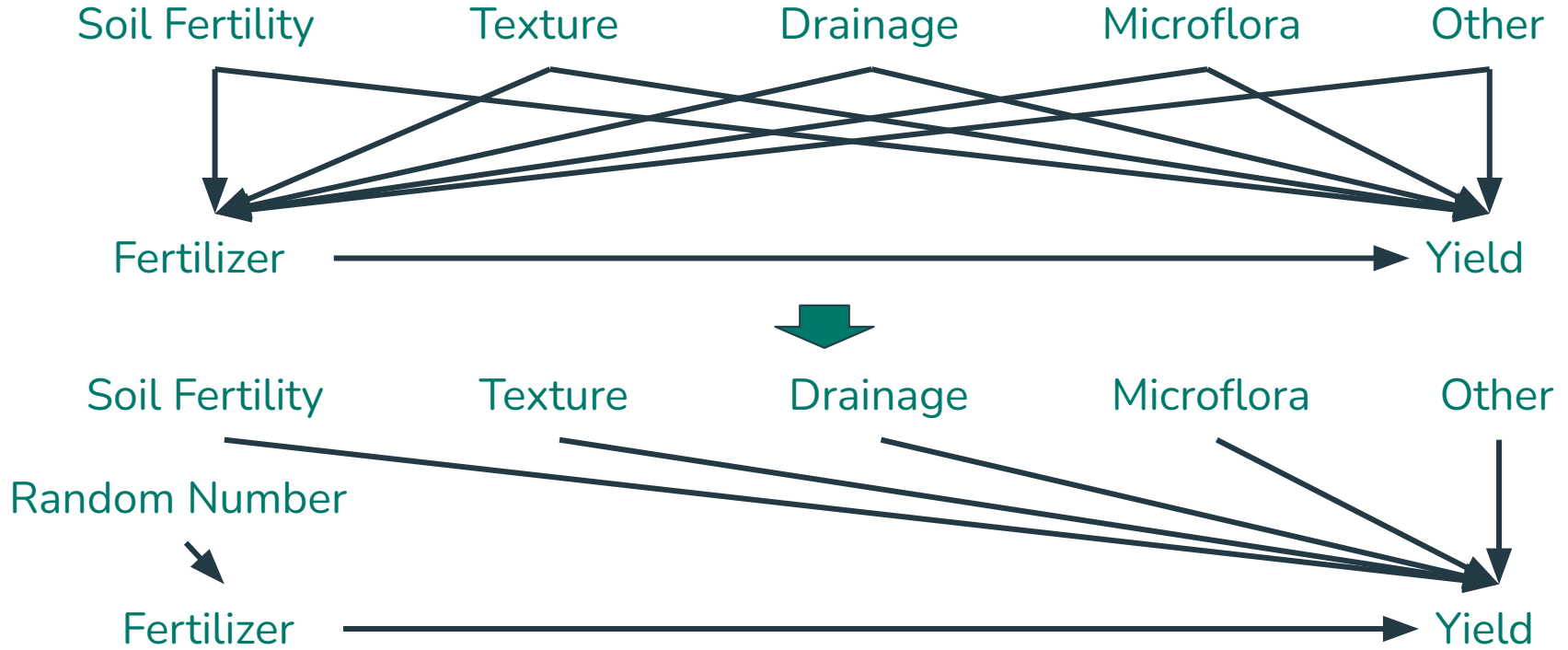
Limitations to Randomized Control Trials

- Intervention may be physically impossible
  - It is difficult to assign someone to be obese or not
- Unethical
  - forcing the treatment group to smoke is not recommended
- Difficult to recruit subjects that are not volunteers
- Expensive
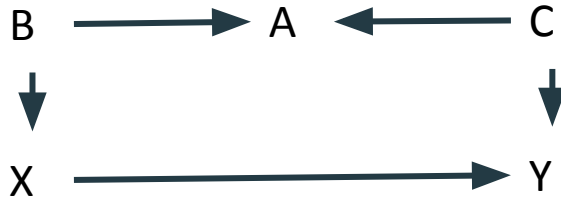
Intervention: Back-Door Adjustment
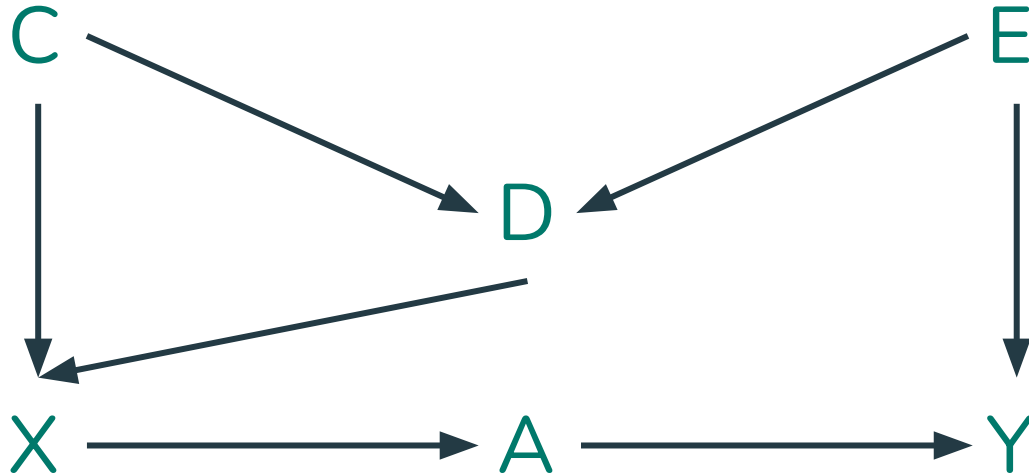
- What does control mean
- Control confounders

$$A \rightarrow X, \quad A \rightarrow Y, \quad X \rightarrow Y$$
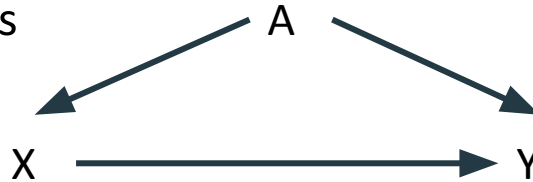
- Don't control mediators

$$X \rightarrow A \rightarrow Y$$

- Don't control colliders

$$B \rightarrow A \leftarrow C$$
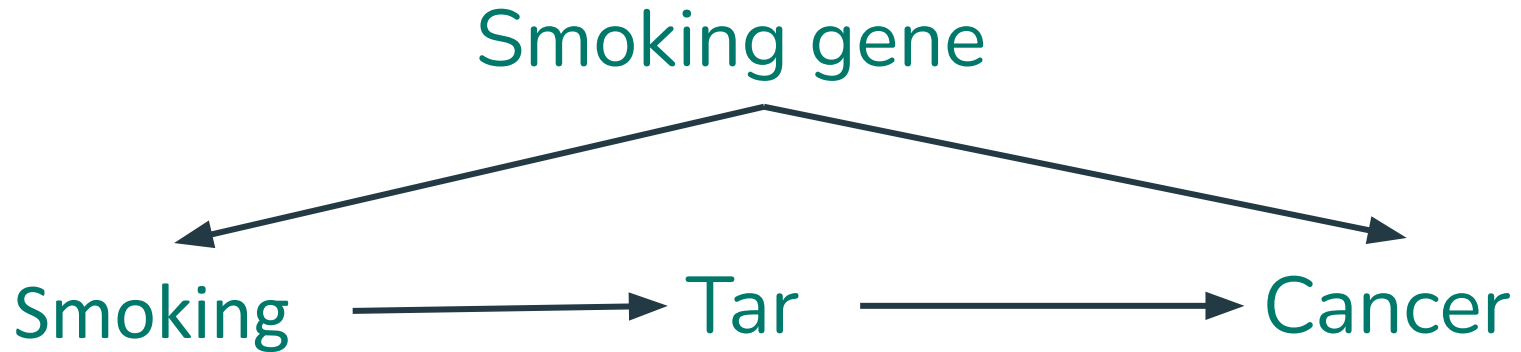$$B \downarrow \qquad C \downarrow$$
$$X \rightarrow Y$$

Back-Door Adjustment Example

Intervention: Front-Door Adjustment

When the confounders are not observed
   hypothetical gene, motivation, lifestyle, etc.

Measure the average causal effect between smoking and tar, then measure the average causal effect between tar and cancer

Intervention: Instrumental Variables

There are no confounder between water company and cholera, any observable association must be causal
Since the effect on cholera by water company must go through water purity, the association must be causal, too

Poverty, foul air, etc.

Water company → Water Purity → Cholera
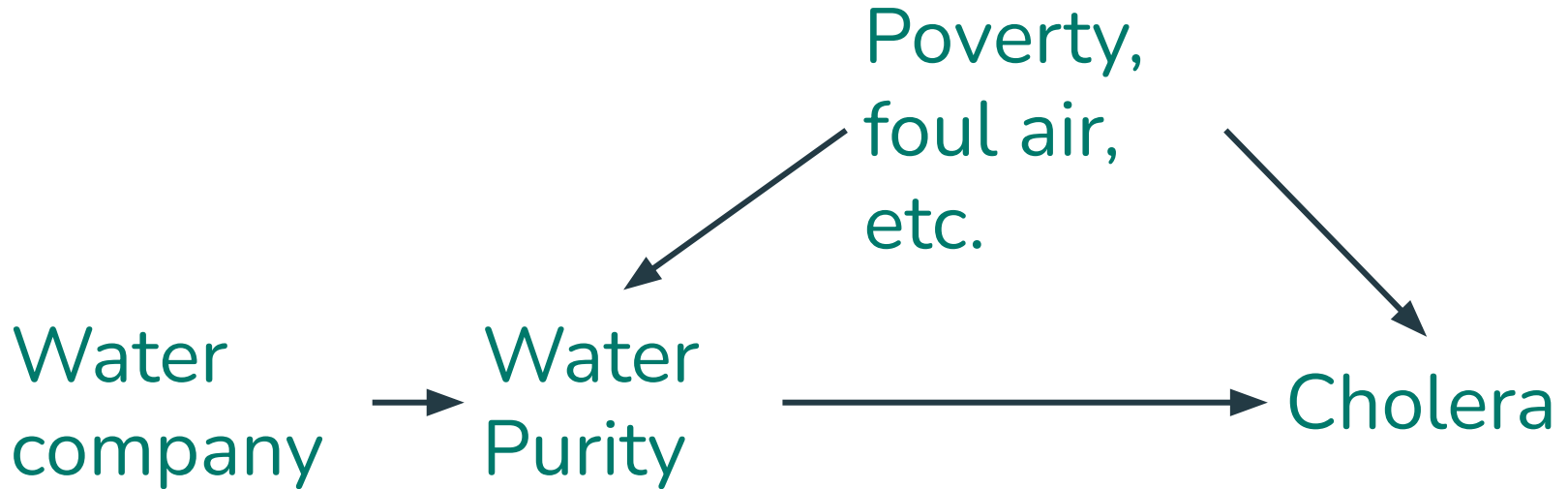
Work Cited

Pearl, J., & Mackenzie, D. (2020). The book of why: The new science of cause and effect. Hachette Books Group.