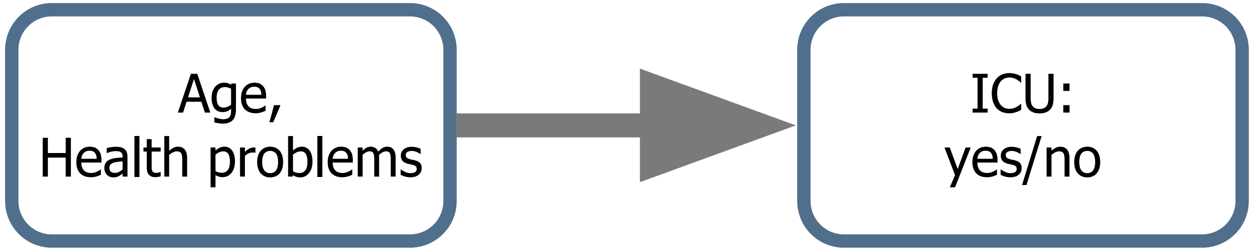
Introduction to Causal Inference

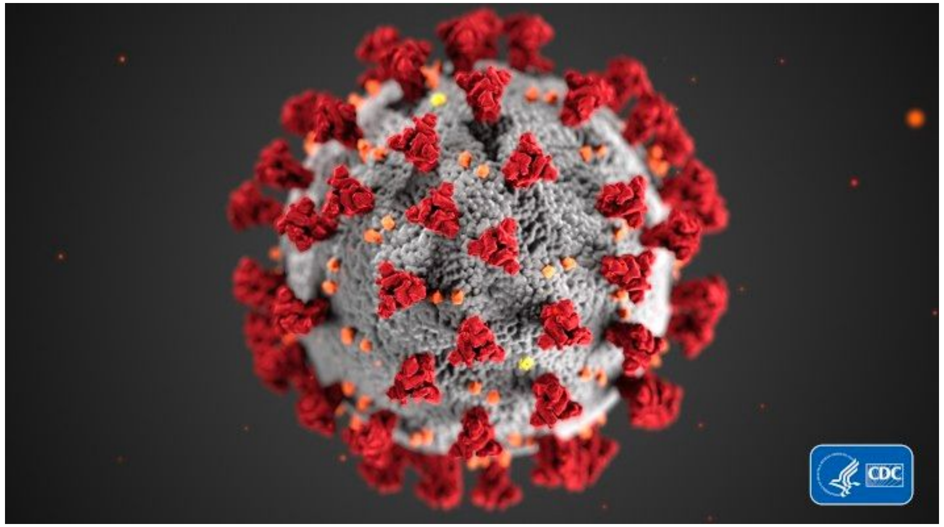
Sesiones de Estadística DAP-Cat

Aleix Ruiz de Villa

Main Objective of these 2 sessions

ICU ingress model due to covid





https://www.cancer.org/es/noticias-recientes/preguntas-comunes-acerca-del-brote-del-nuevo-coronavirus.html

STATISTICS / CAUSAL INFERENCE

Objective: ¿do age and healthcare affect ICU ingress?

Decision: design public health strategies

Model: finding the correct model

MACHINE LEARNING

Objective: Predict the risk

Decision: patient ranking

Model: finding the most accurate

Machine Learning main assumption

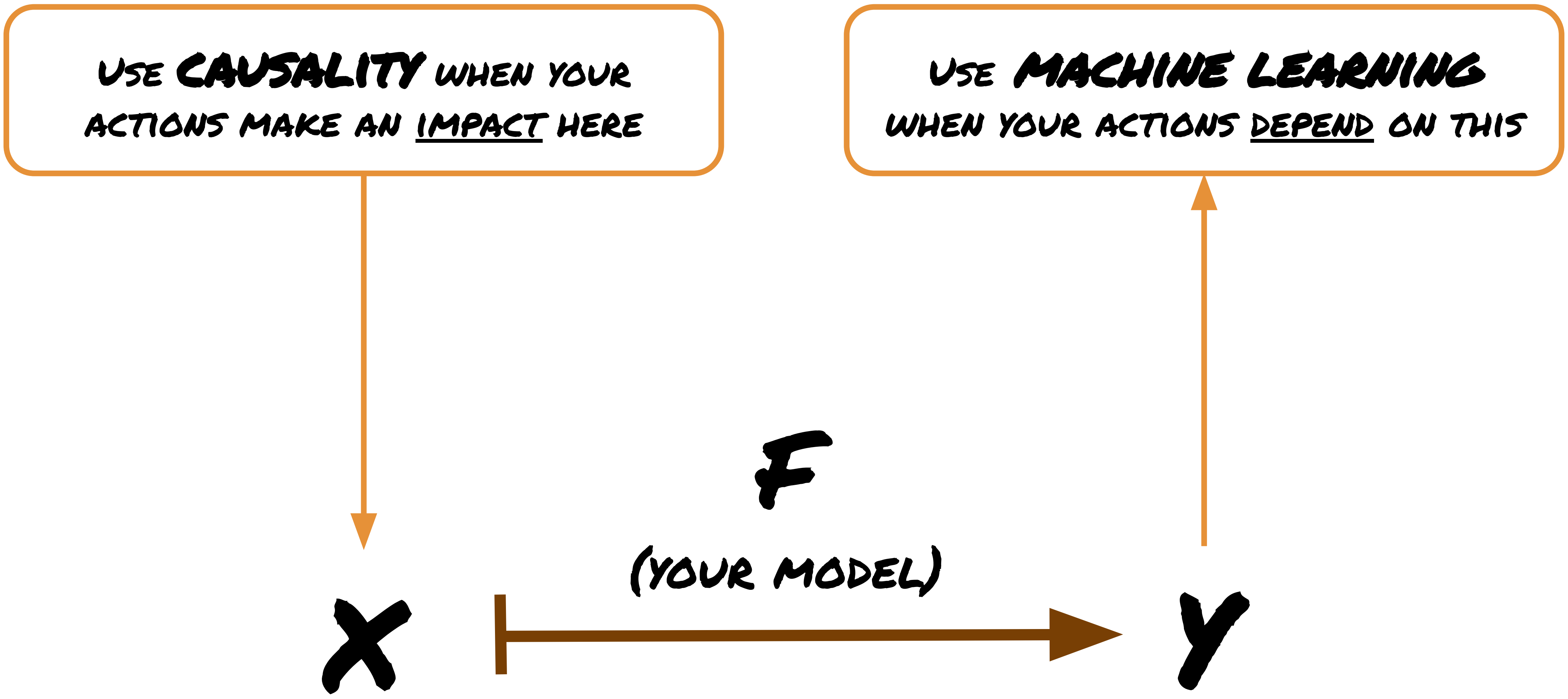
Past and Future behave the same

https://www.cancer.org/es/noticias-recientes/preguntas-comunes-acerca-del-brote-del-nuevo-coronavirus.html

Machine Learning limitation

What if we behave in a different manner?

https://www.alpinerecoverylodge.com/intervention-assistance/



Causality = RCTs + Causal Inference

Introduction to Causal Inference

Observational Examples

NEWS • LIVE TV INDIA TODAY APP

HOME | MY FEED | INDIA | WORLD | BUSINESS | TECH | MOVIES | SPORTS | SCIENCE | HEALTH | VIDEOS

News / Lifestyle / Health / Moderate drinking during pregnancy doesn't harm baby

Moderate drinking during pregnancy doesn't harm baby

Observational Examples

Mail Online

Home | News | U.S. | Sport | TV&Showbiz | Australia | Femail | Health | Science | Mone

Latest Headlines | Covid-19 | Royal Family | Prince Harry | Meghan Markle | World News | Headlines | Mos

Processed meat 'is to blame for one in 30 deaths': Scientists say a rasher of cheap bacon a day is harmful

Frameworks

- Potential Outcomes: Biostatistics & Econometrics
- Directed Acyclic Graphs: Computer Science

Both are equivalent, but each one suitable for different things

Simpson’s paradox

Simpson's Paradox

Treatment	Size	Number	Recovered
A	Small	87	81
B	Small	270	234
A	Large	263	192
B	Large	80	55

Hospital's Main Problem

Which of the two treatments should they take?

Simpson’s Paradox

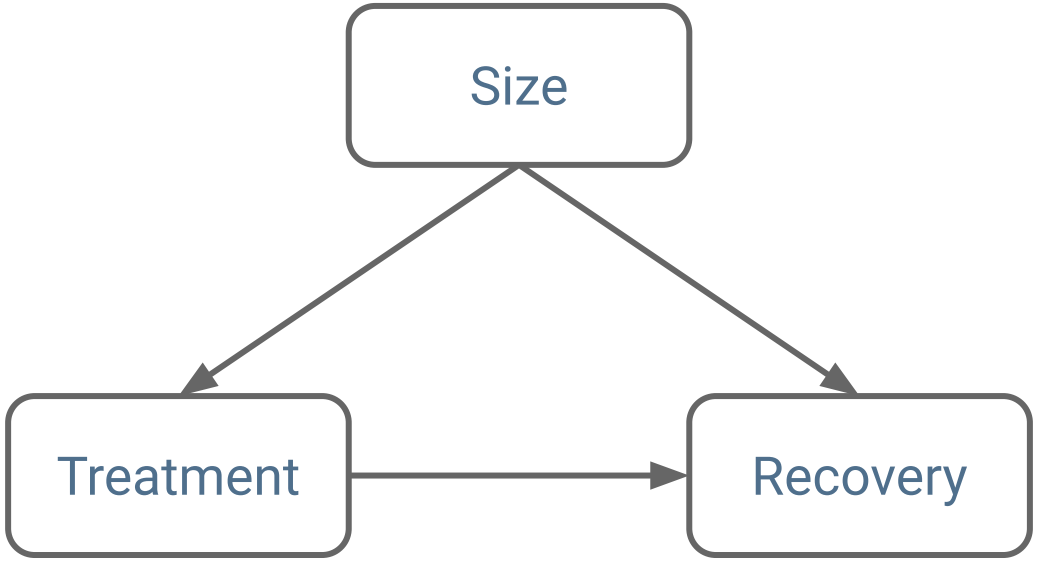
Size	Treatment A	Treatment B
Recovery	78% (273/350)	**83%** (289/350)

Simpson's Paradox

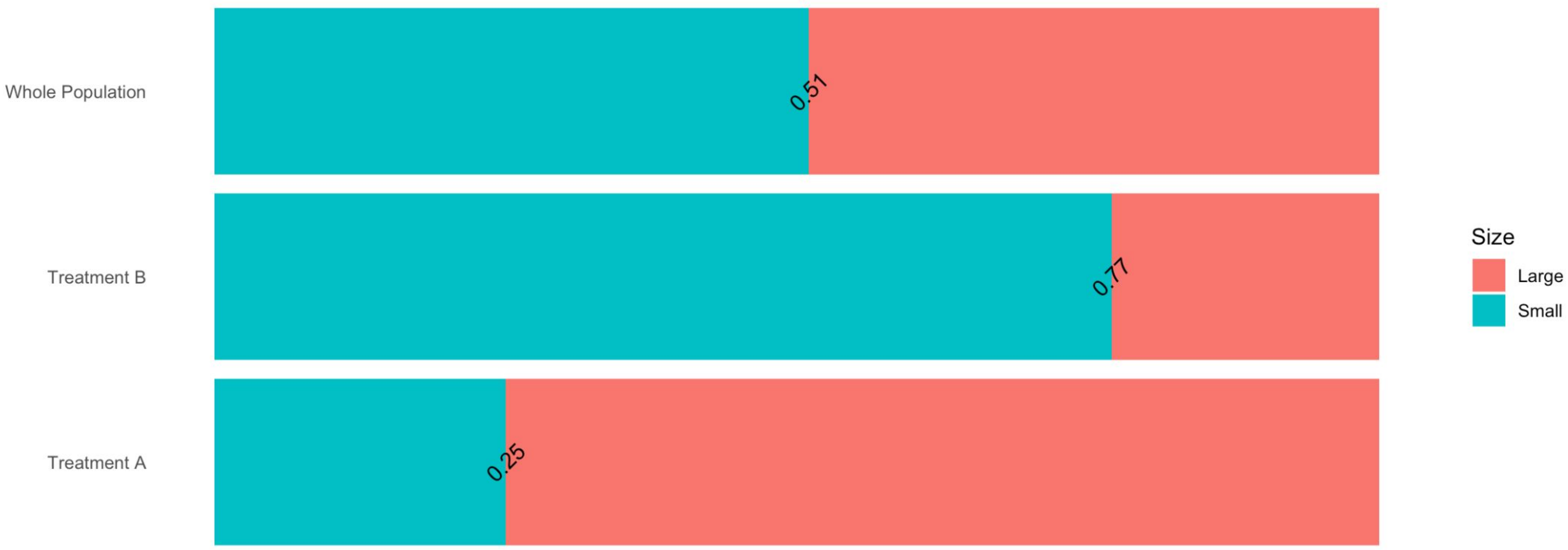
	Treatment A	Treatment B
Large	**93%** (81/87)	87% (234/270)
Small	**73%** (192/263)	69% (55/80)

Analysis of Simpson’s paradox

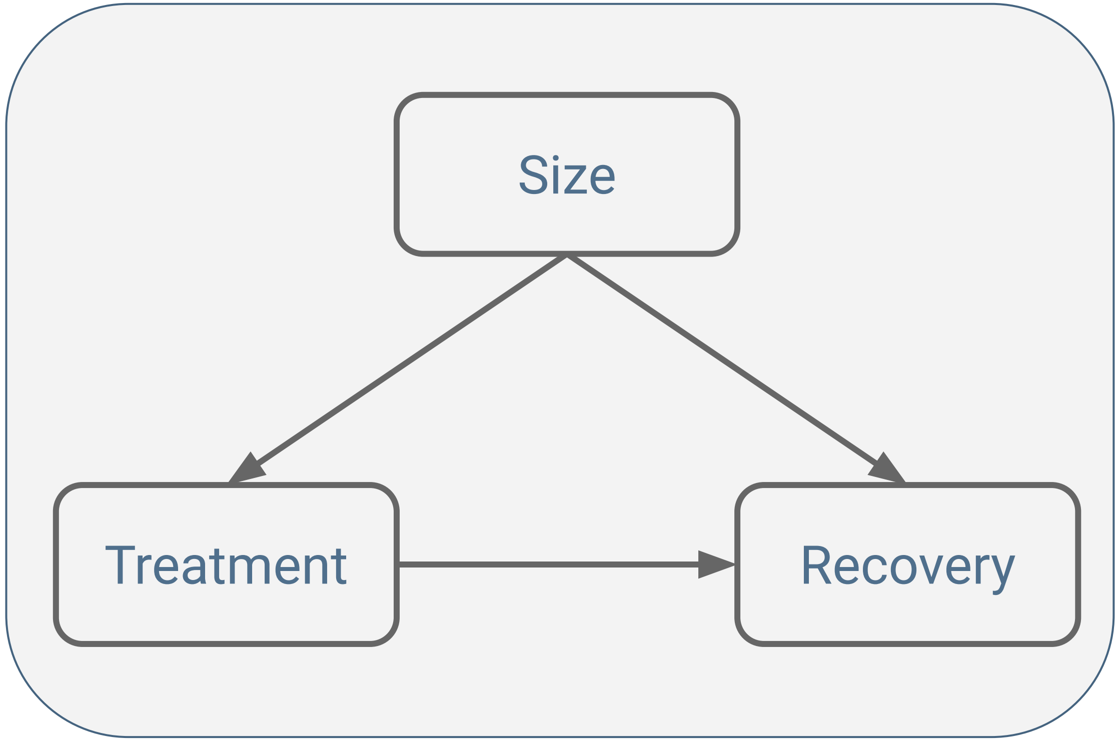
Graph

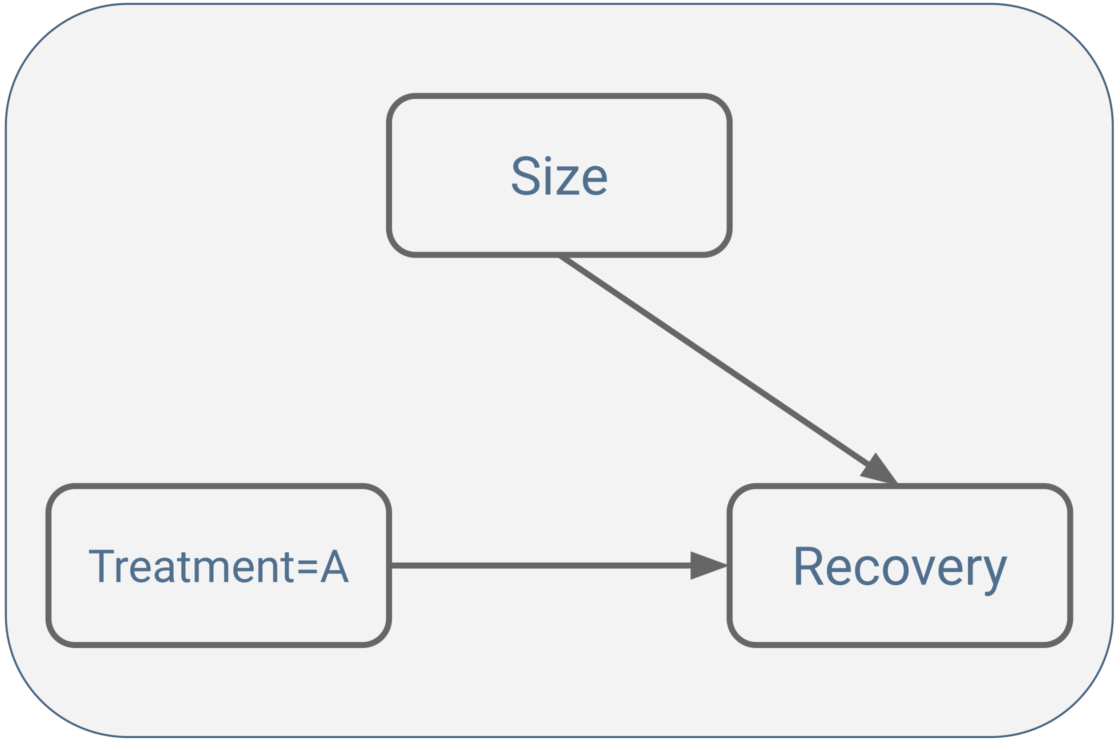


Size Distribution



What is an intervention?





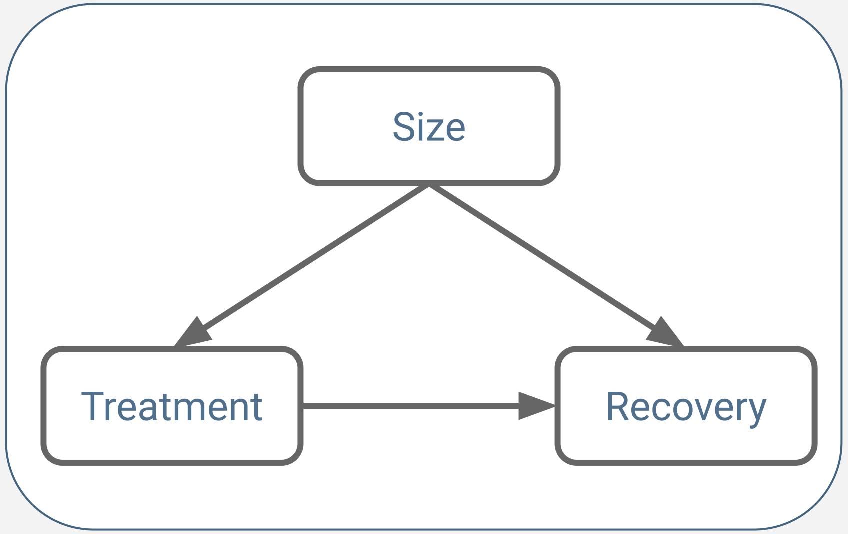
The data you have

Distribution P

The data you would like to have

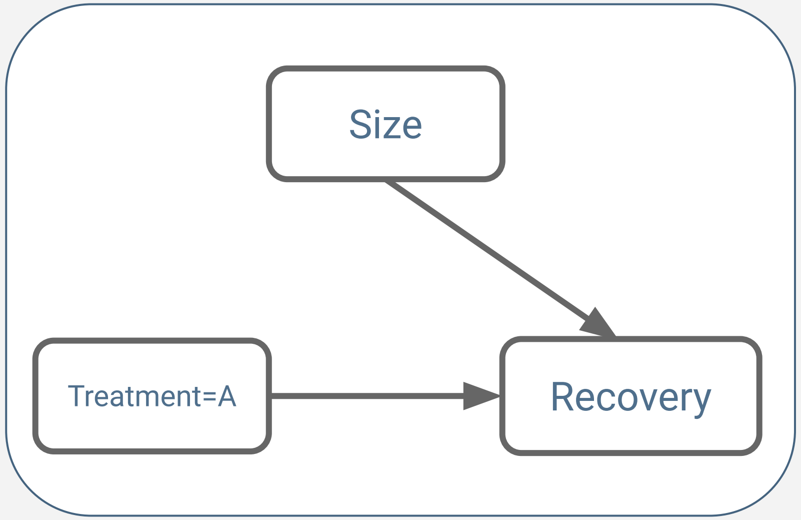
Distribution $P^{A}=P^{do(T:=A)}$

Main objective of causal inference



Use observational data

To infere about interventional data



Average Treatment Effect

Conditional Probability

P(R=1| T=A) = P(R=1| T=A, S=Small) * **P(S=Small|T=A)** + P(R=1| T=A, S=Large) * **P(S=Large|T=A)**

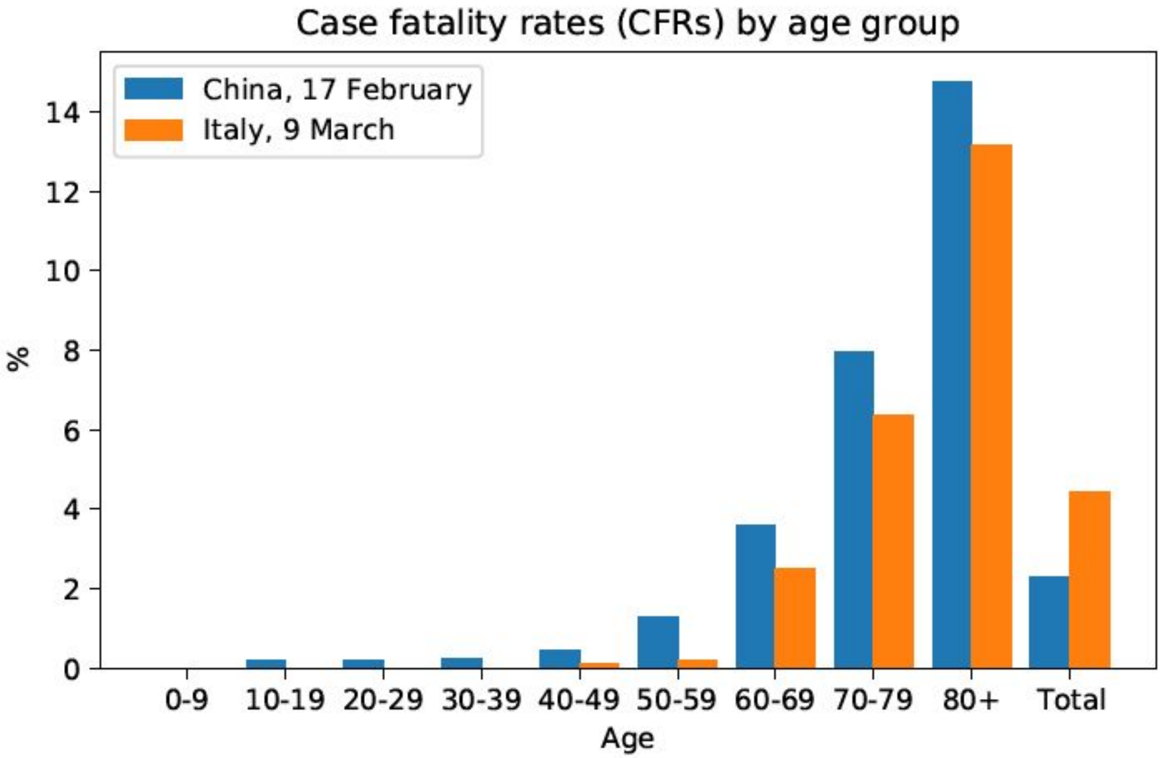
= 93% * **25%** + 73% * **75%** = 78%

Adjustment

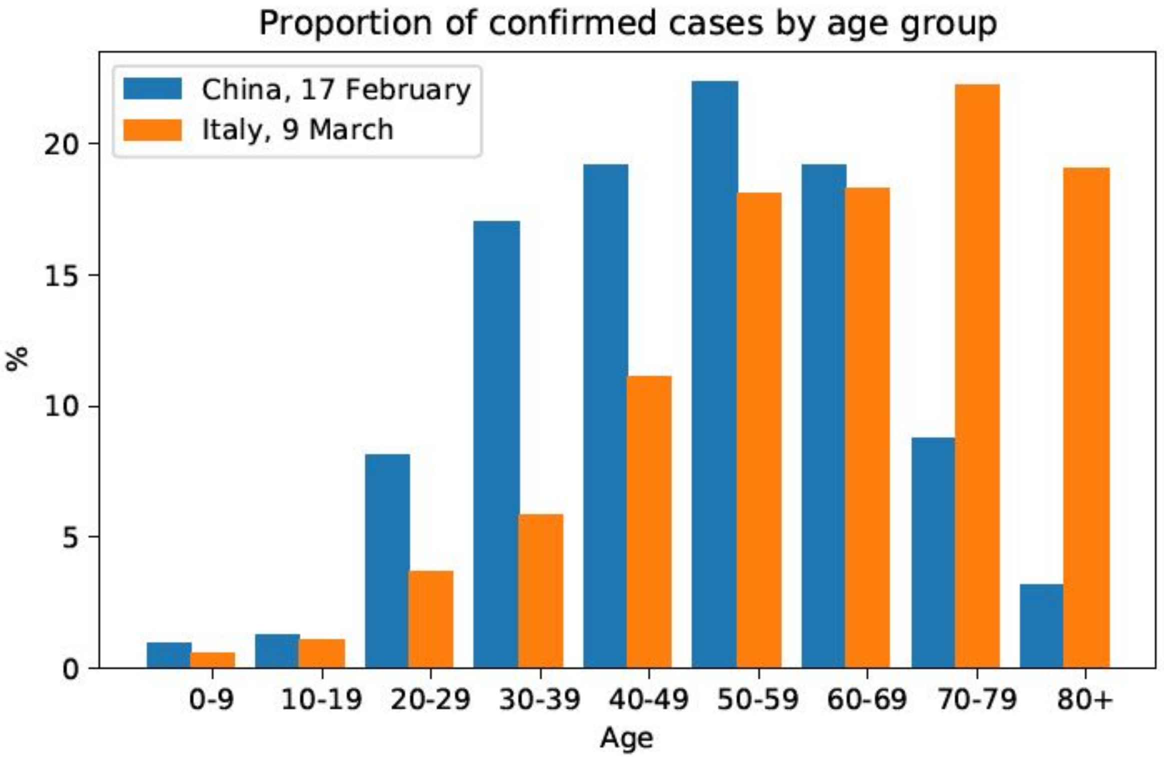
P(R=1|do(T=A)) = P(R=1| T=A, S=Small) * **P(S=Small)** + P(R=1| T=A, S=Large) * **P(S=Large)**

= 93% * **51%** + 73% * **49%** = 83%

Covid Example



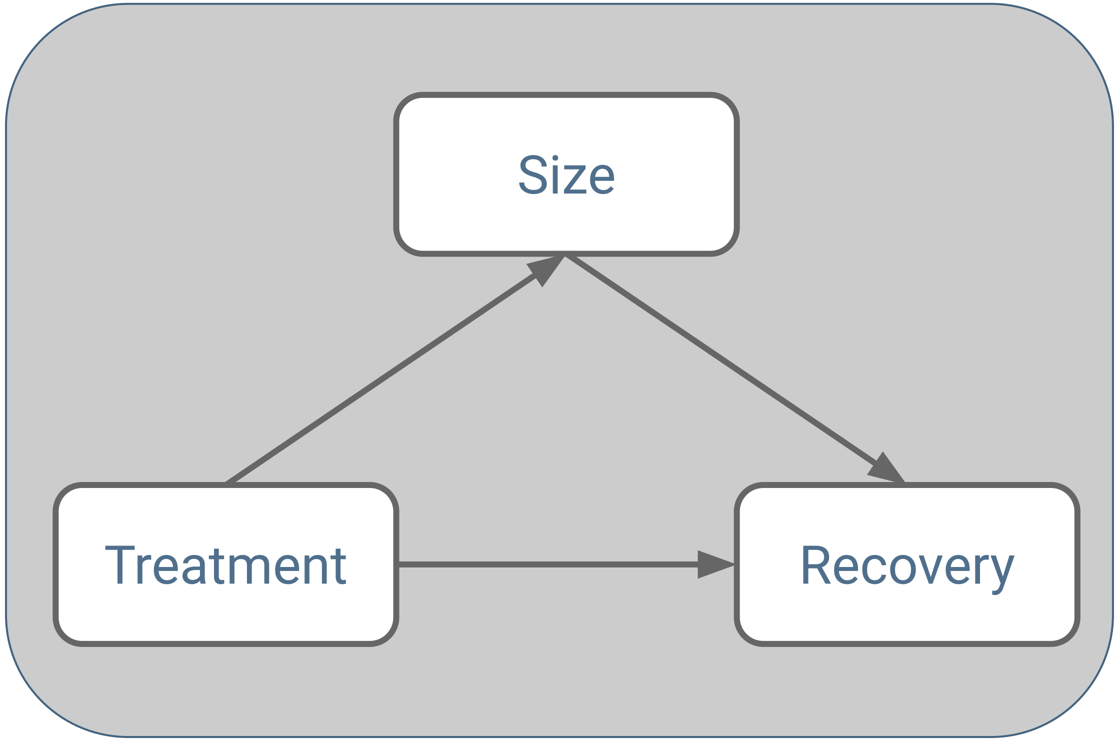
(a)



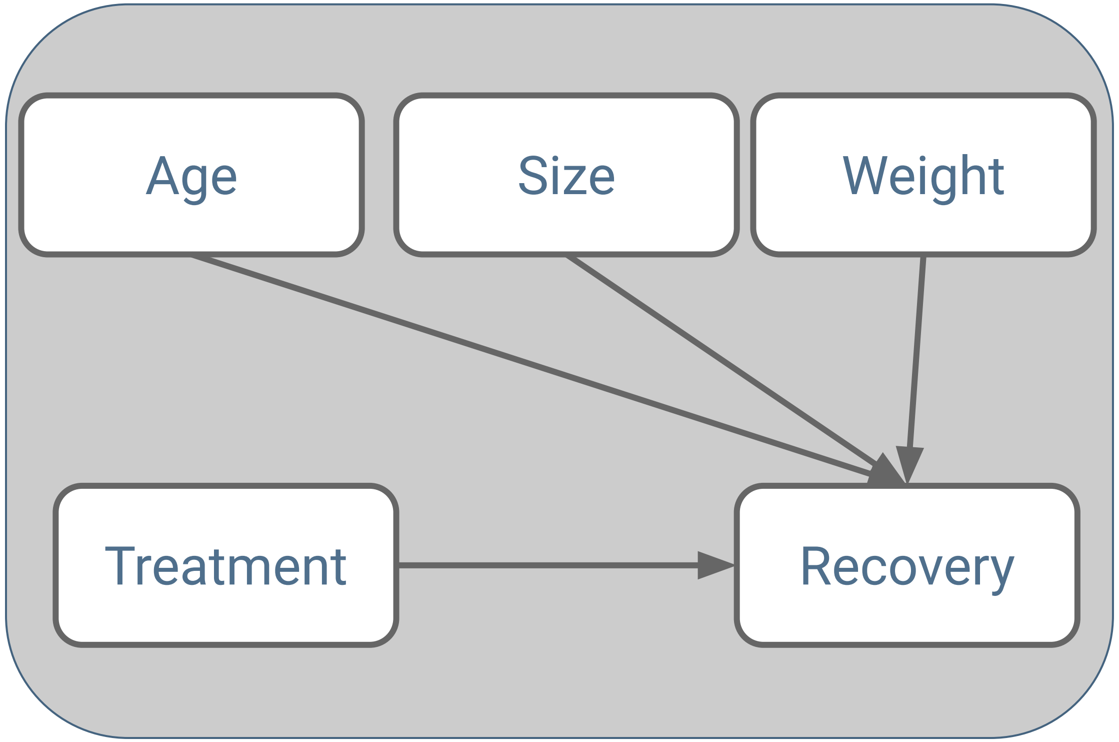
(b)

When to adjust?

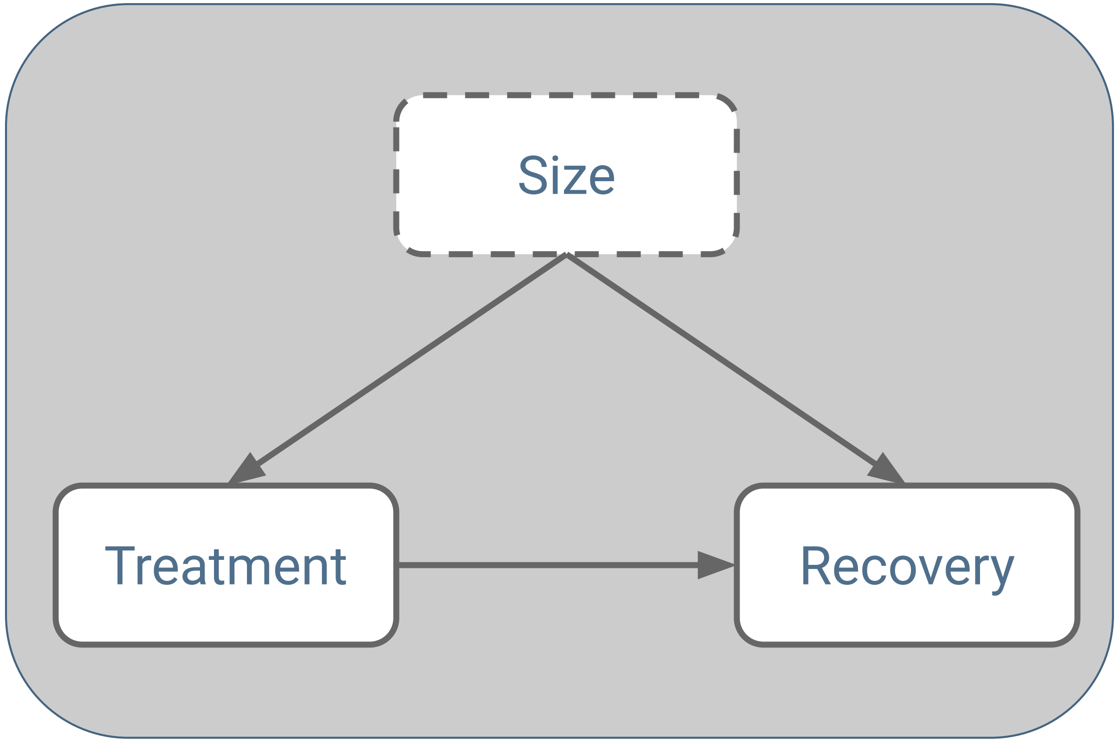
Do we need to adjust?



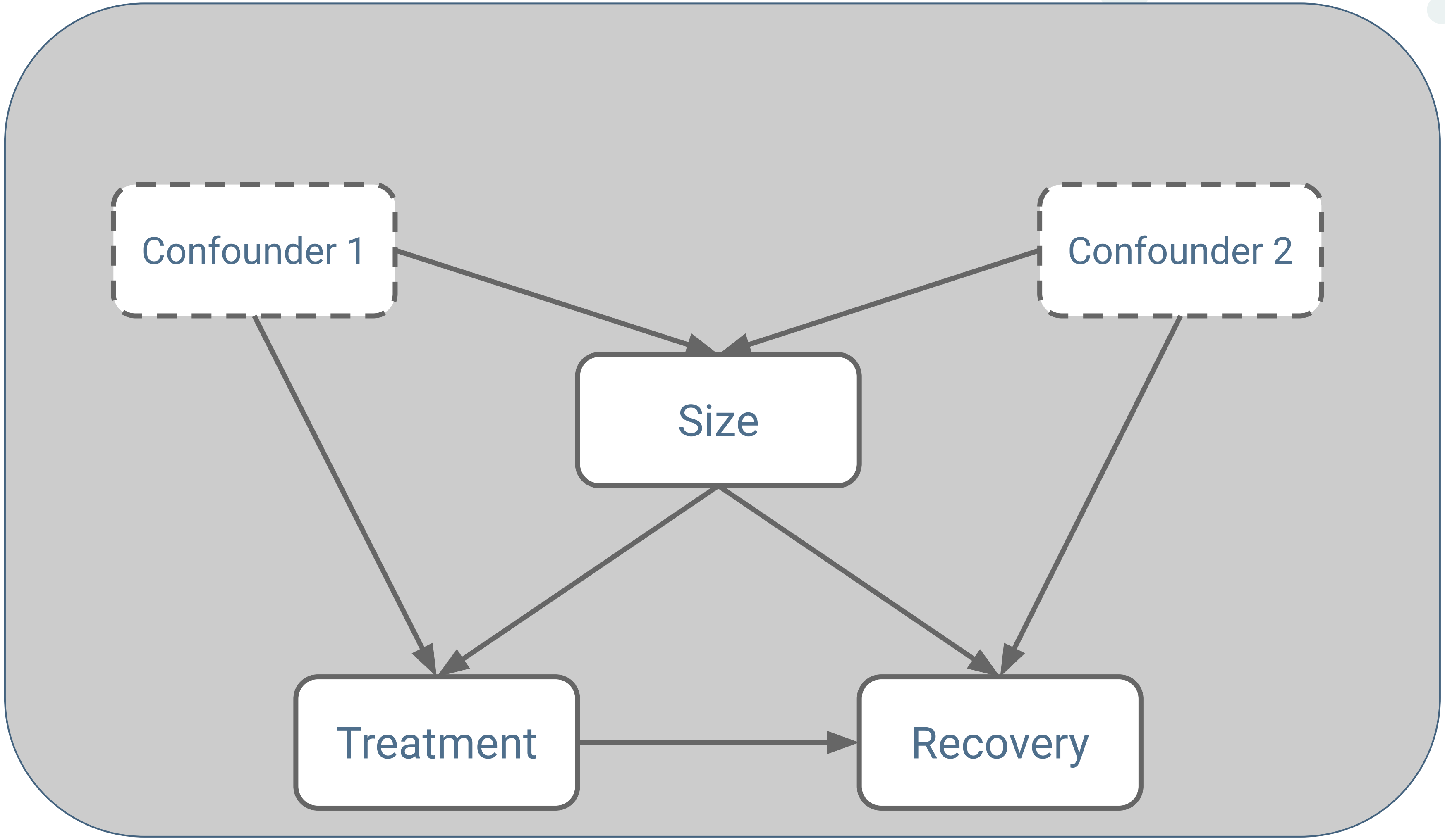
Do we need to adjust?



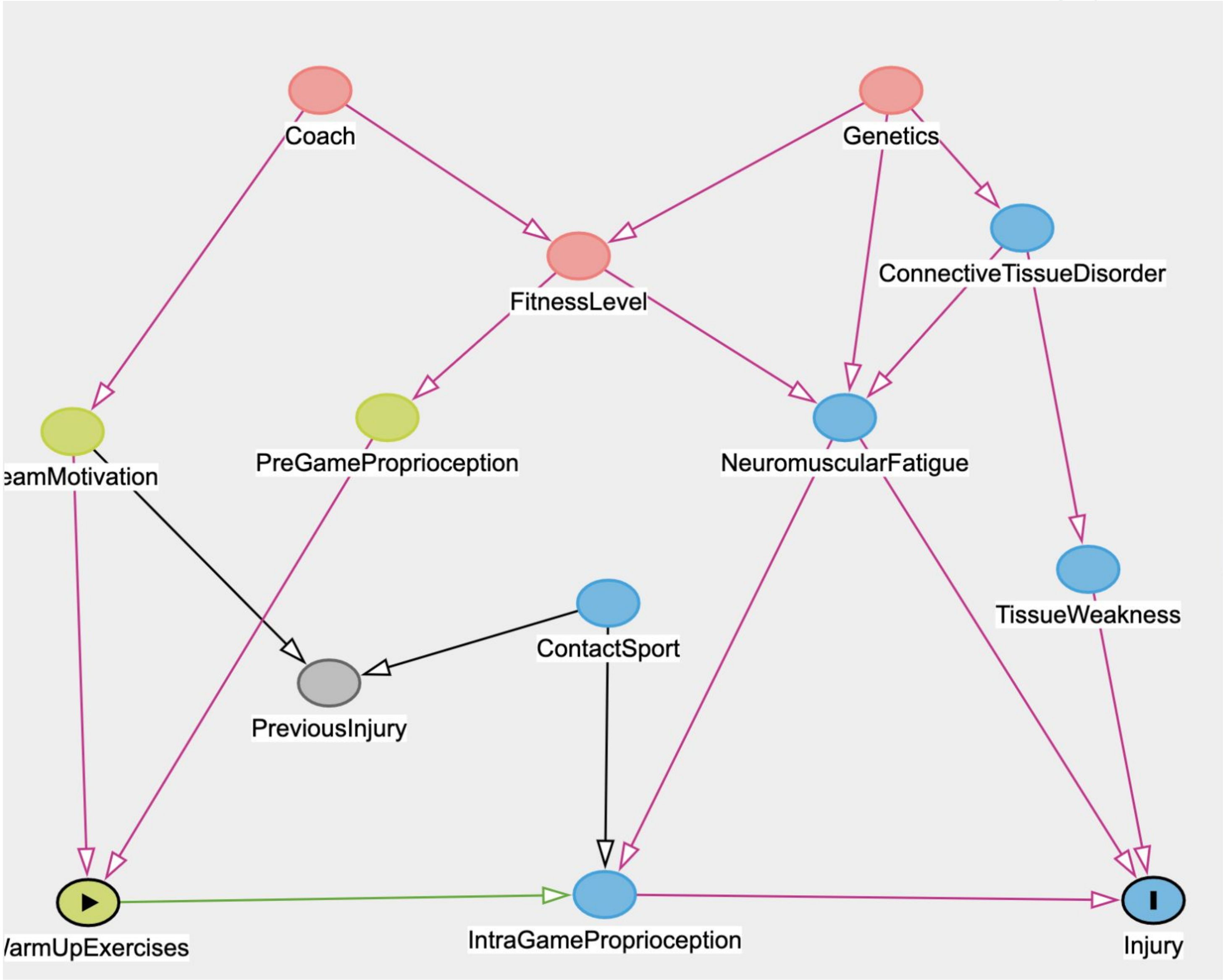
Do we need to adjust?



Do we need to adjust?



Example



Adjustment (total effect)

Minimal sufficient adjustment sets for estimating the total effect of WarmUpExercises on Injury:

- Coach, FitnessLevel
- Coach, PreGameProprioception
- ConnectiveTissueDisorder, NeuromuscularFatigue
- FitnessLevel, Genetics
- FitnessLevel, TeamMotivation
- NeuromuscularFatigue, TissueWeakness
- PreGameProprioception, TeamMotivation

Testable implications

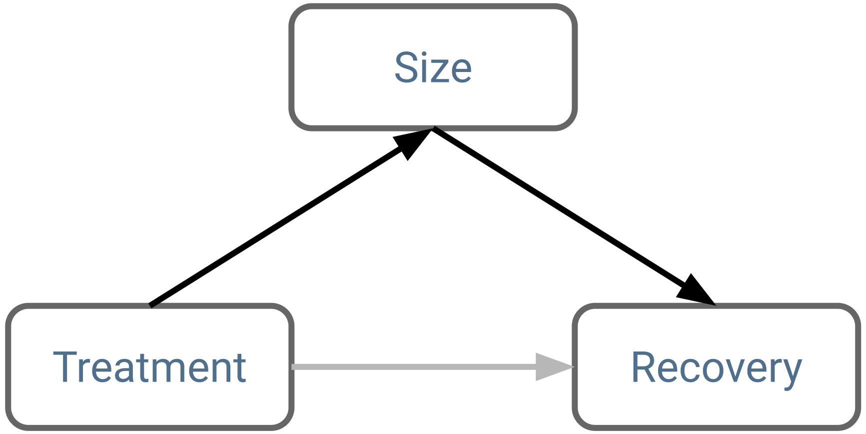
The model implies the following conditional independences:

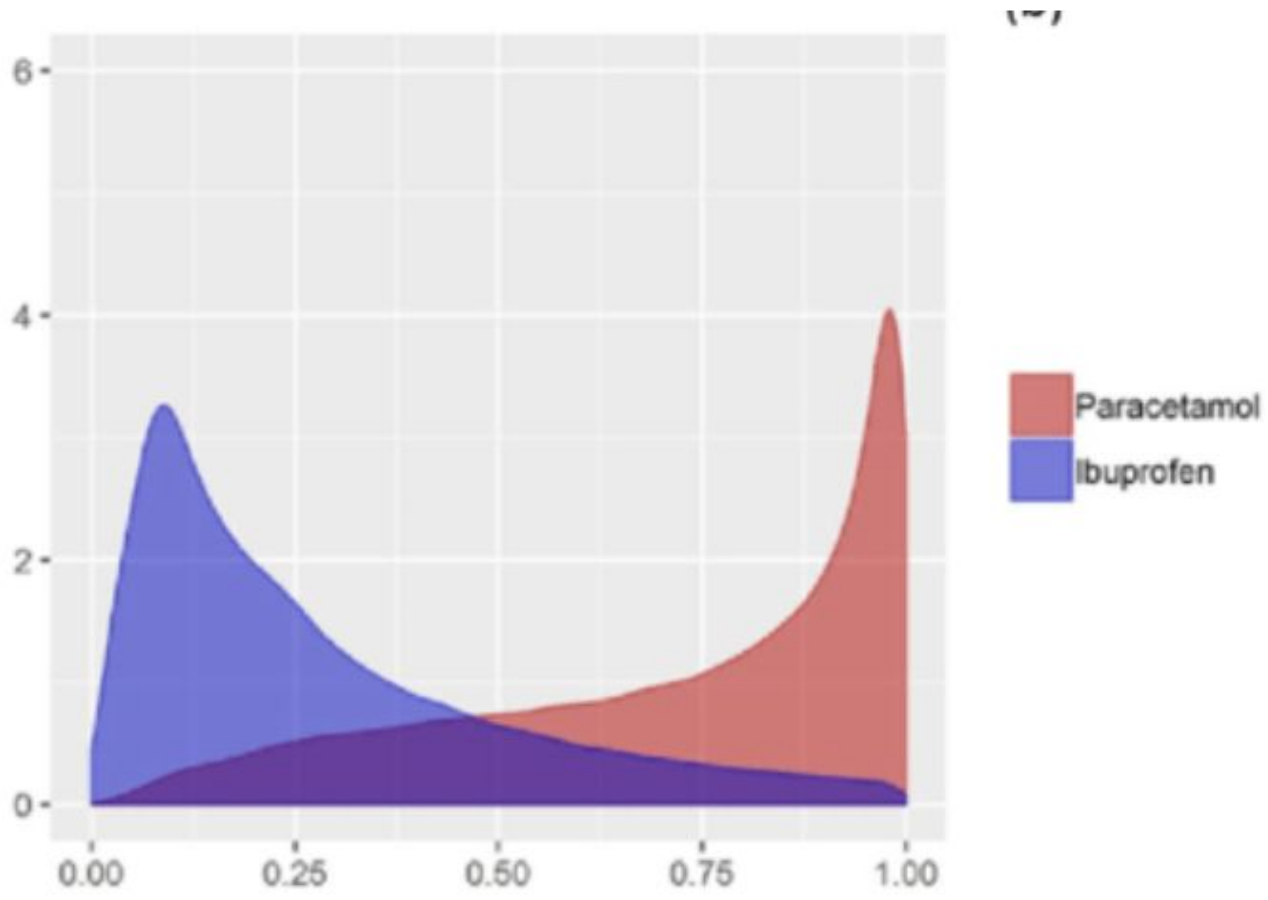
- WarmUpExercises ⊥ Injury | IntraGameProprioception, NeuromuscularFatigue, TissueWeakness
- WarmUpExercises ⊥ Injury | ConnectiveTissueDisorder, IntraGameProprioception, NeuromuscularFatigue
- WarmUpExercises ⊥ Injury | FitnessLevel, Genetics, IntraGameProprioception, NeuromuscularFatigue

Applications

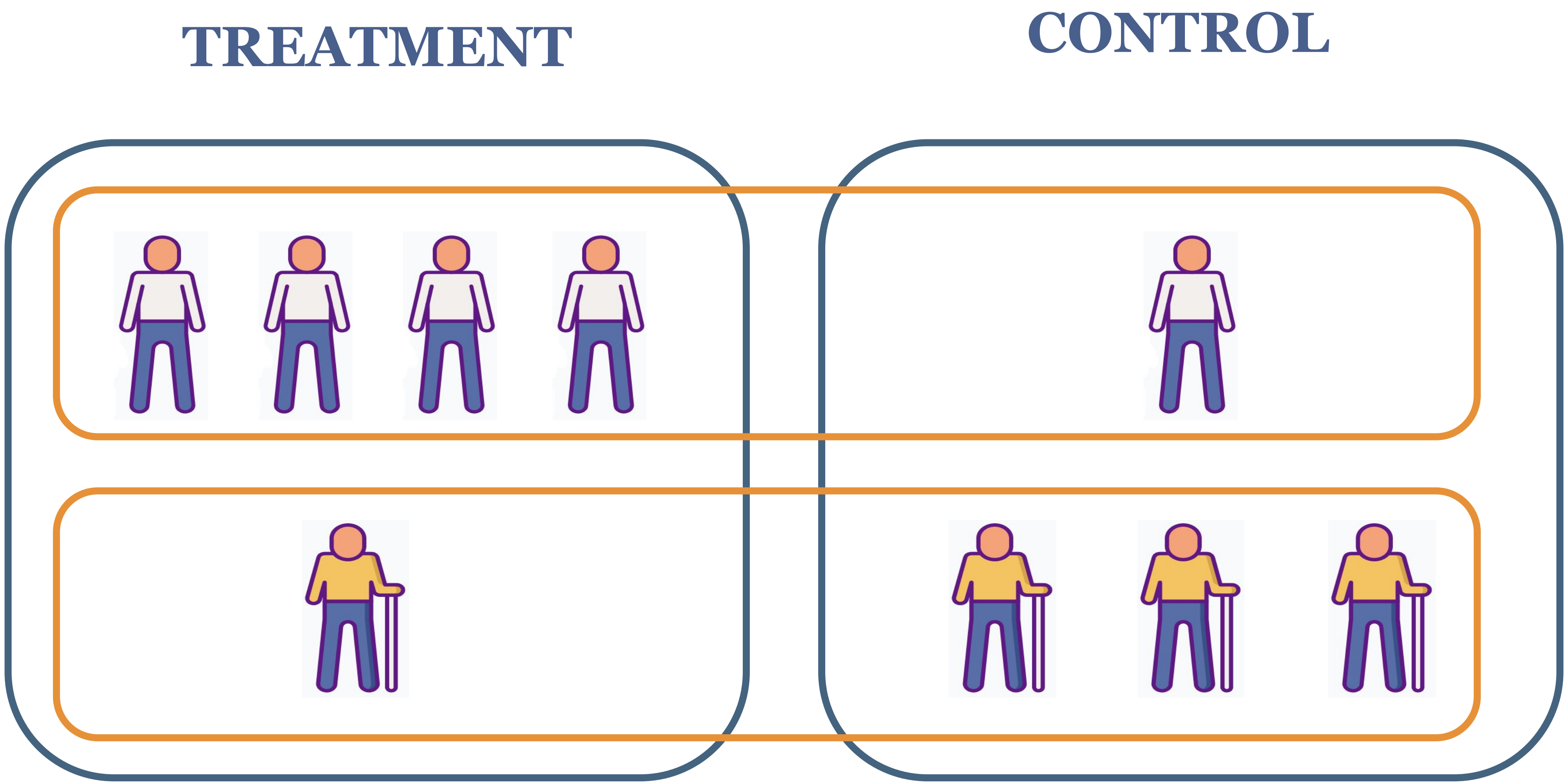
Variations of the adjustment formula

1. Propensity Scores
2. In linear models: controlling for some variables
3. Mediation Analysis





Propensity Score



We can compare if they have same attributes

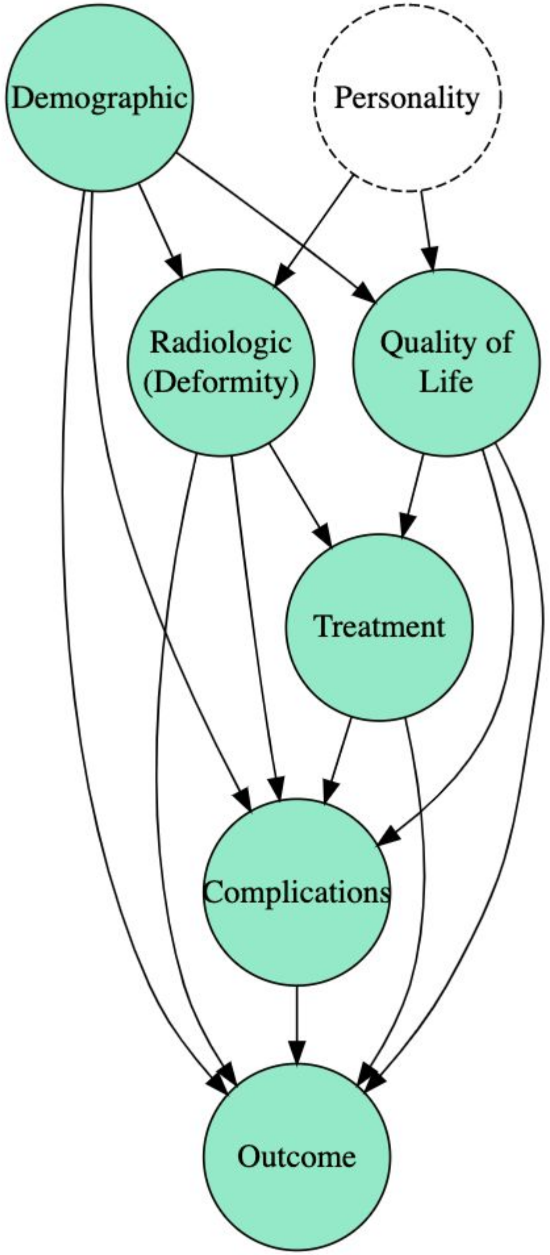
<u>Propensity Score:</u>

We can compare if they have the same chances to be treated

Example

Differences between

- What is the impact of a particular type of treatment
- What is the difference between treating the patient or not



Why RCTs are so important

Why RCTs are so important

- Which are the confounders of a RCT?
- In general, how are we sure that are considering all possible confounders?

RCT

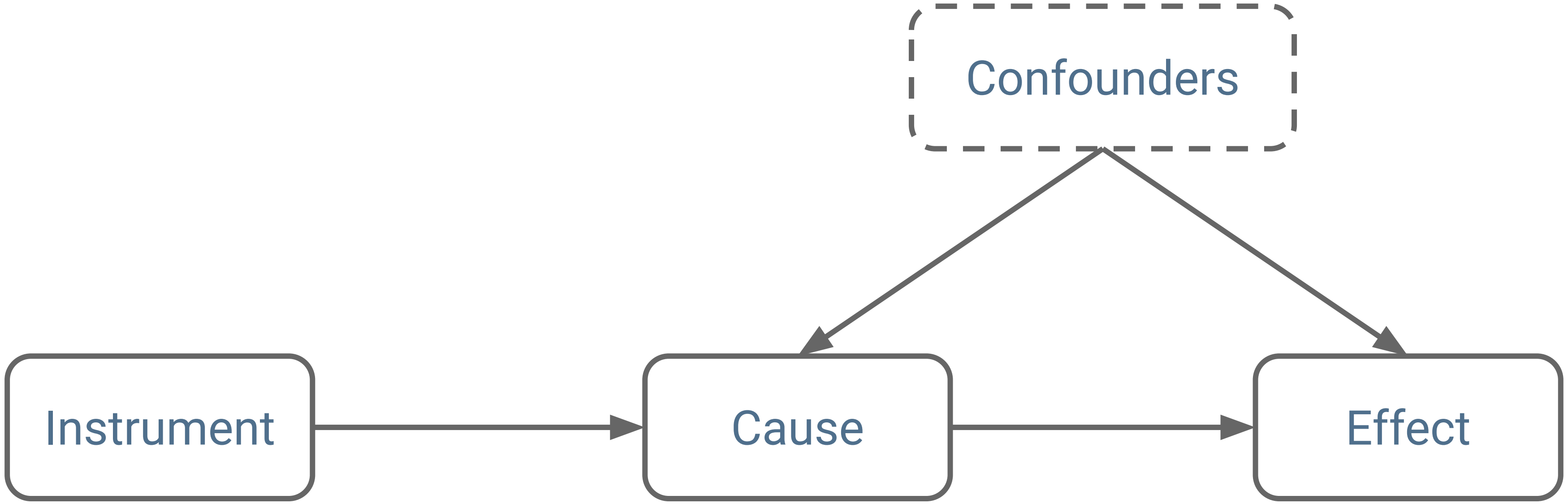
- Measuring an outcome.
- Mean risk: uncertainty

Causal Modeling

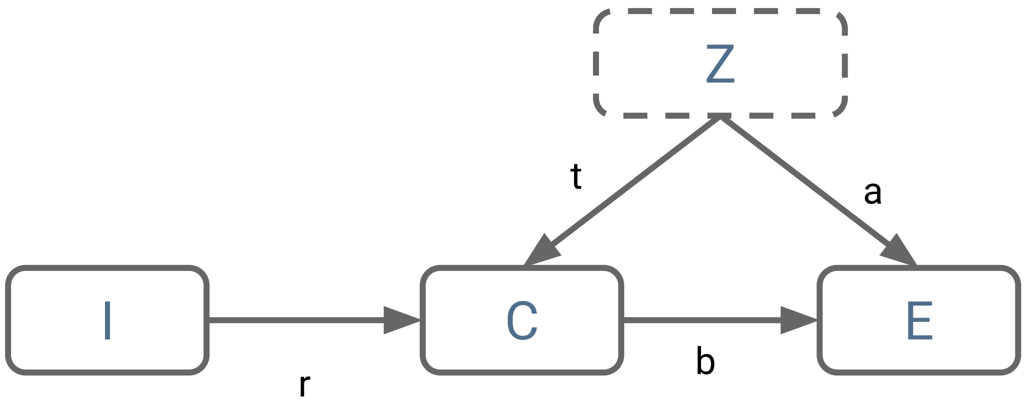
- Modeling Causes + Measuring outcome
- Main risk: errors in modeling + uncertainty

Instrumental Variables

Graph



Estimation



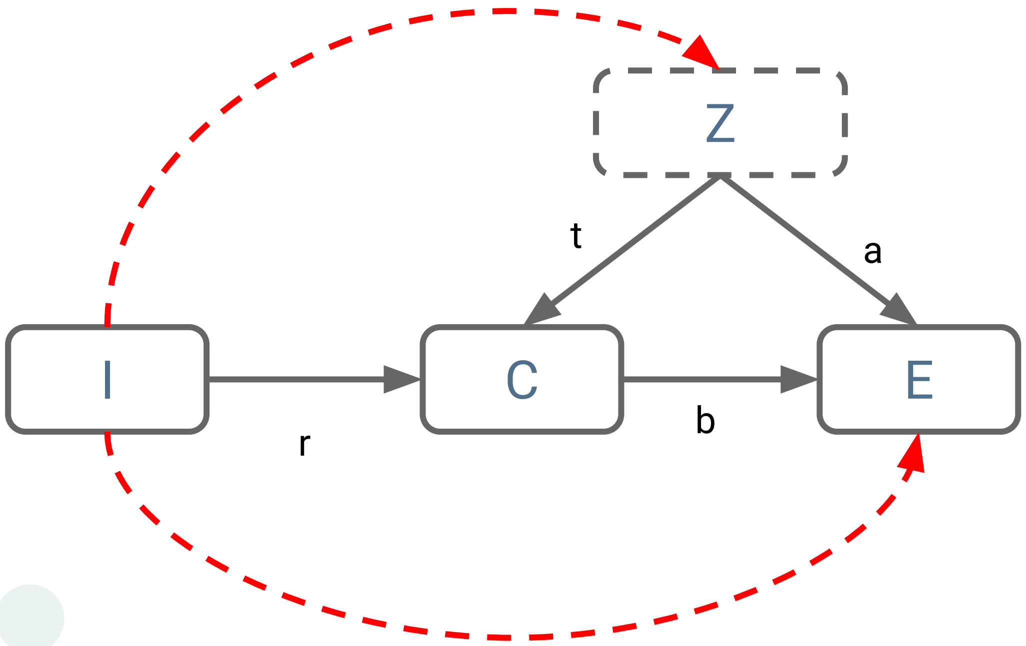
<u>Model</u>

E = a Z + b C
C = r I + t Z

<u>Formulation</u>

E = a Z + b (r I + t Z) =
= (a + bt) Z + br I

Assumptions

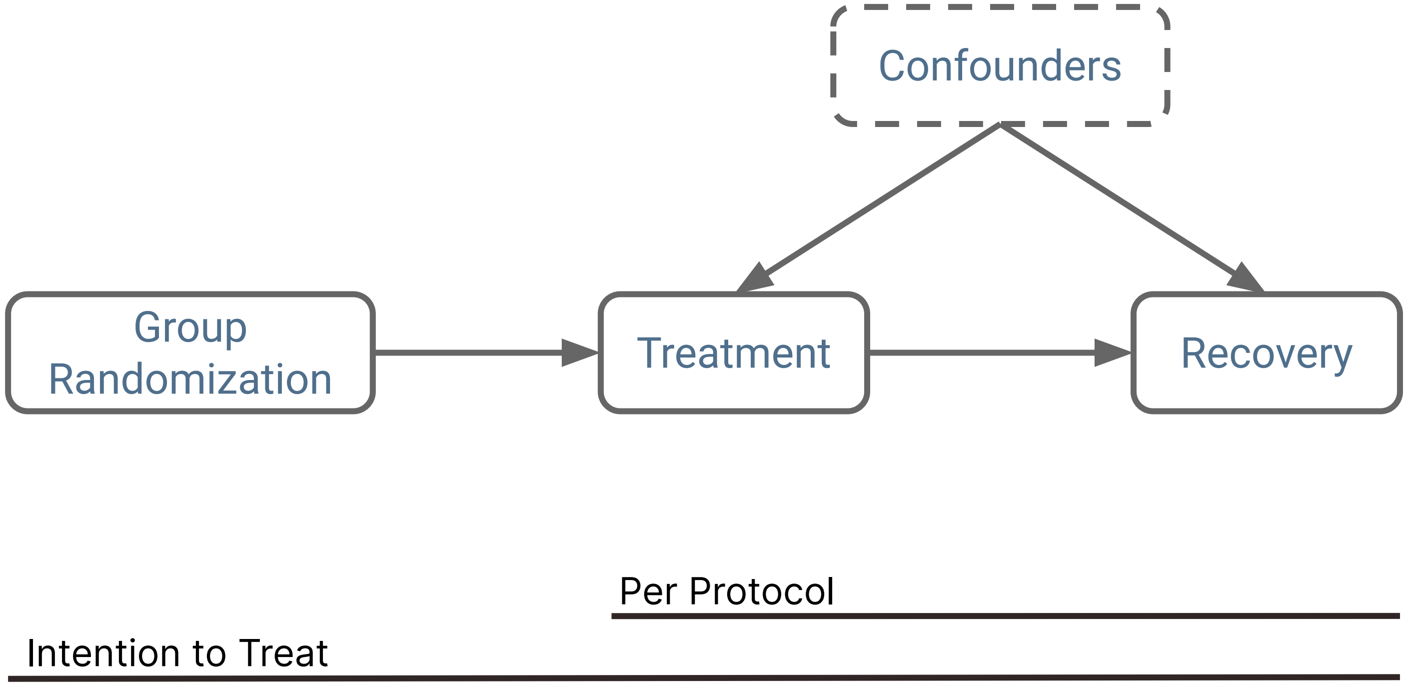


Exclusion restriction:

There is no other path from I to E

Non compliant RCTs

Graph



Intention to Treat

- Unbiased
- Diluted

Per Protocol

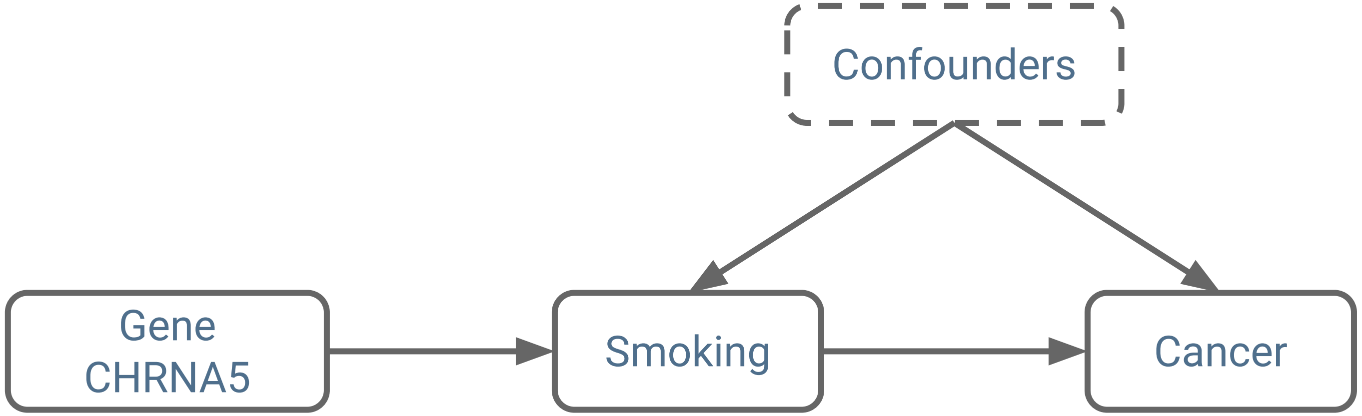
- Potentially biased

Instrumental Variables:

- Unbiased
- Not diluted

Mendelian Randomization

Mendelian Randomization



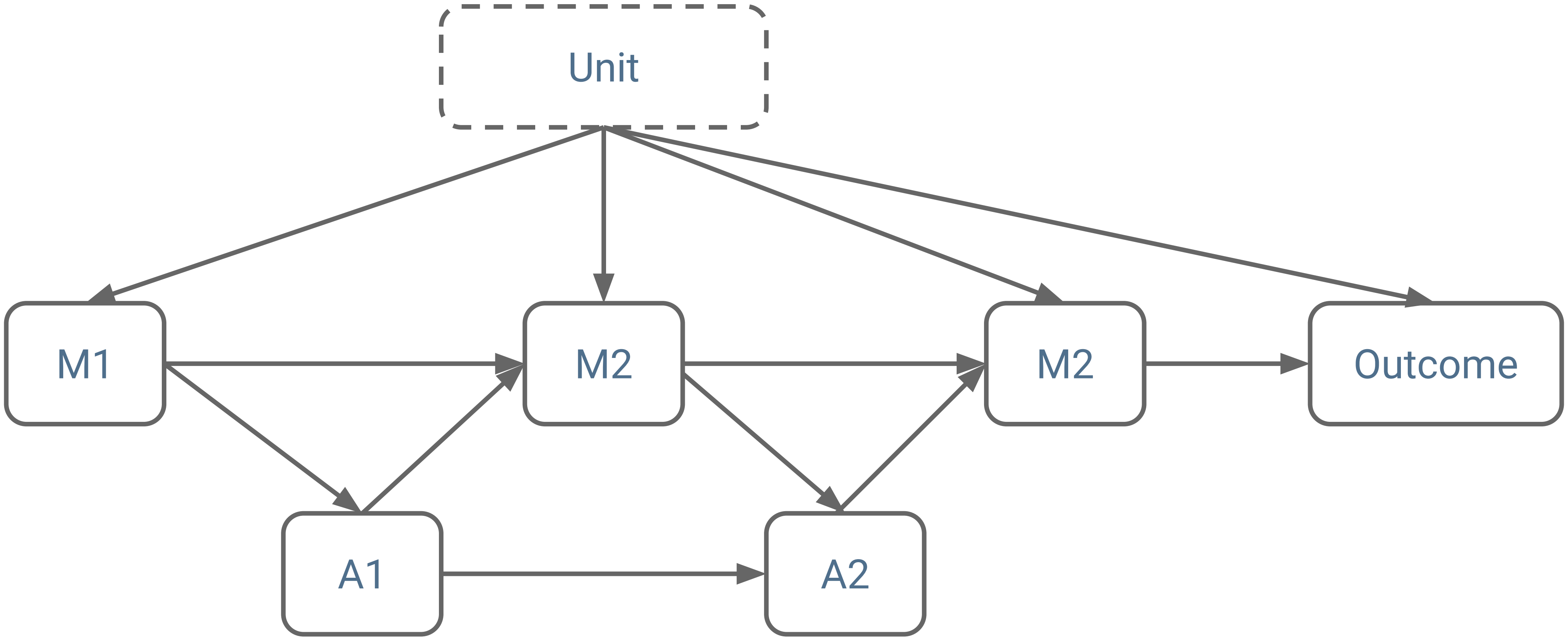
- CHRNA 4 highly correlated with Smoking (well estimated)
- Exclusion restriction: Out of the smoking group there seems not be any correlation between CHRNA5 and Cancer

Advanced Stuff

Statistical Estimation of Average Treatment Effects

- Using Machine Learning to calculate Propensity Scores and Adjustment Formula
- Using Double Machine Learning to Estimate Adjustment Formula and Time-Varying treatment effects
- Confidence intervals and p-values

Time-Varying Treatment Effects



External Validity and Transportability

Example 1 *We conduct a randomized trial in Los Angeles (LA) and estimate the causal effect of treatment X on outcome Y for every age group $Z = z$ as depicted in Fig. 1(a). We now wish to generalize the results to the population of New York City (NYC), but we find that the distribution $P(x, y, z)$ in LA is different from the one in NYC (call the latter $P^*(x, y, z)$). In particular, the average age in NYC is significantly higher than that in LA. How are we to estimate the causal effect of X on Y in NYC, denoted $P^*(y|do(x))$.*[1]

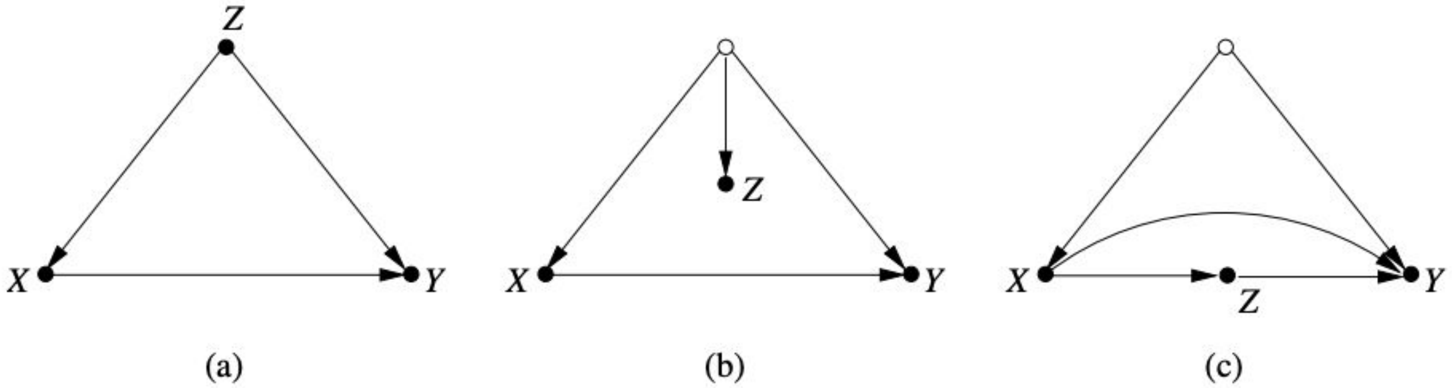
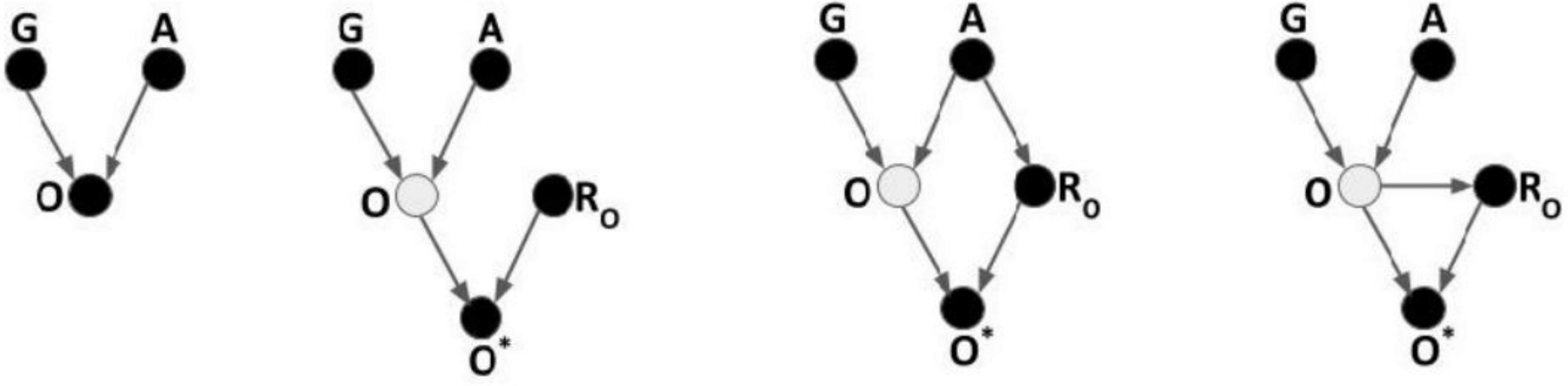
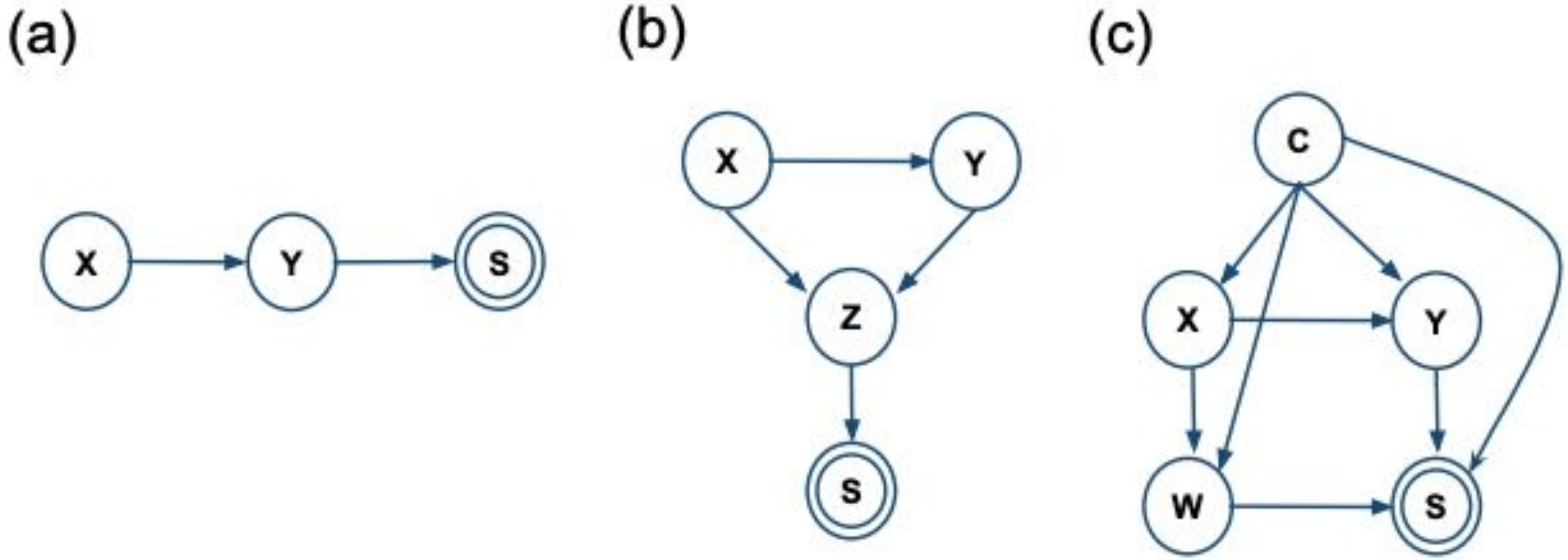


Figure 1: Causal diagrams depicting Examples 1–3. In (a) Z represents "age." In (b) Z represents "linguistic skills" while age (hollow circle) is unmeasured. In (c) Z represents a biological marker situated between the treatment (X) and a disease (Y).

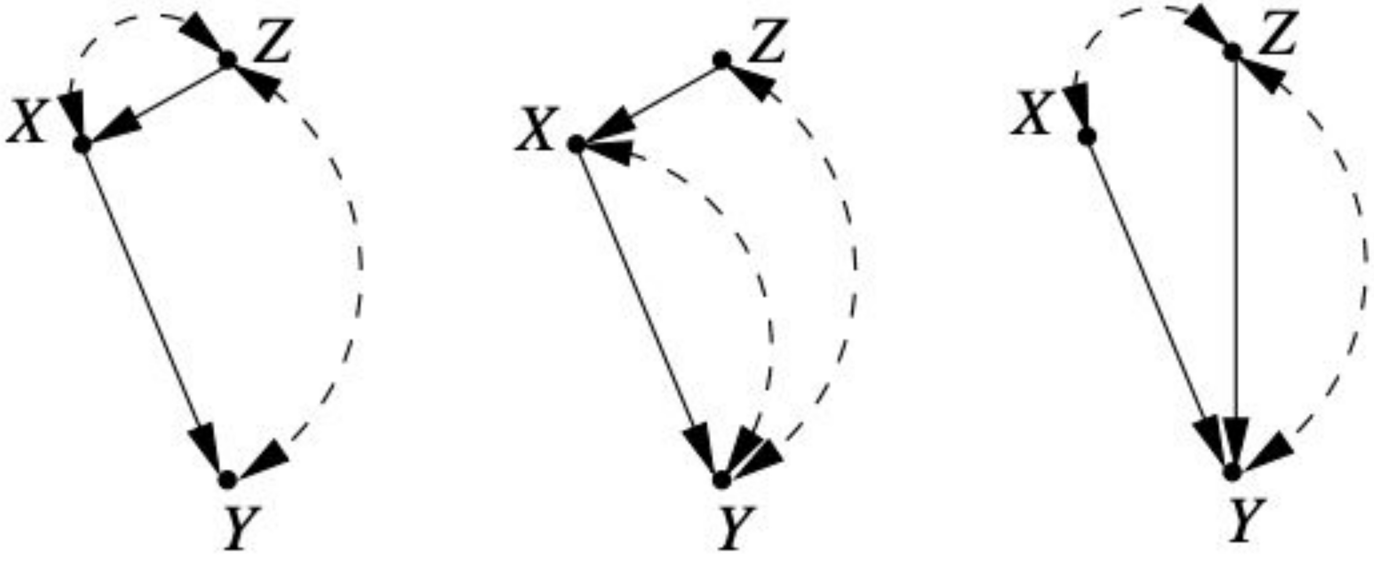
Recoverability of missing values



Recovery from selection Bias



Z-identifiability (difficult): identification through auxiliary experiments



(a) (b) (c)

Conclusions

Conclusions

- CI are more flexible than RCTs, but come at a price: making assumptions with its associated risk
- Causal Inference builds on top of classical statistics, where causality plays a central role
- Causal **modeling** is about (formal) modeling

Applications

- CI when there is no alternative, prioritizing RCTs, noncompliant RCT, mendelian randomization, ...