Machine Learning for multimorbidity causal inference

Funded by Hoffman la Roche

Nicola Serra (University of Zurich /CERN)

Olivia Jullian Parra, Shrija Rajen Sheth (CERN)

Milo Alan Puhan, Maximilian Sebastian Janisch, Henock Yebyo, Giuseppe Genovese,

Thomas Lehéricy (Univeristy of Zurich)

CERN openlab Technical Workshop

17th March 2023

Outline

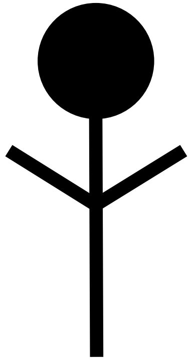
- Causal inference in multimorbidity: Introduction and Sate of the art
- Machine Learning for Multimorbidity causal inference: Introduction
- Machine Learning for Multimorbidity causal inference: Timeline

Causal inference in Multimorbidity

Causal inference challenge

Will statins reduce the risk of cardiovascular diseases?

Random Patient 1

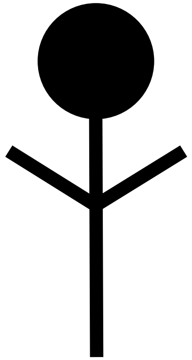


Statins taken

Sufficient to accept the treatment effect?

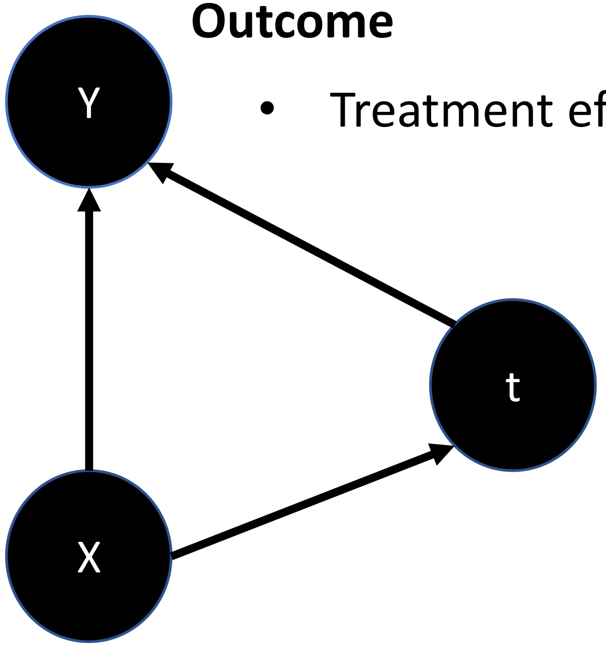
NO, other variables must be considered: age, other diseases, meds exposure, etc.

Random Patient 2



Statins non taken

Causal inference challenge



Outcome

- Treatment effect on patient variables

Treatment (yes/no)

- Treatment which effect is being studied

Confounders

- Patient variables that affect the outcome and the treatment effect

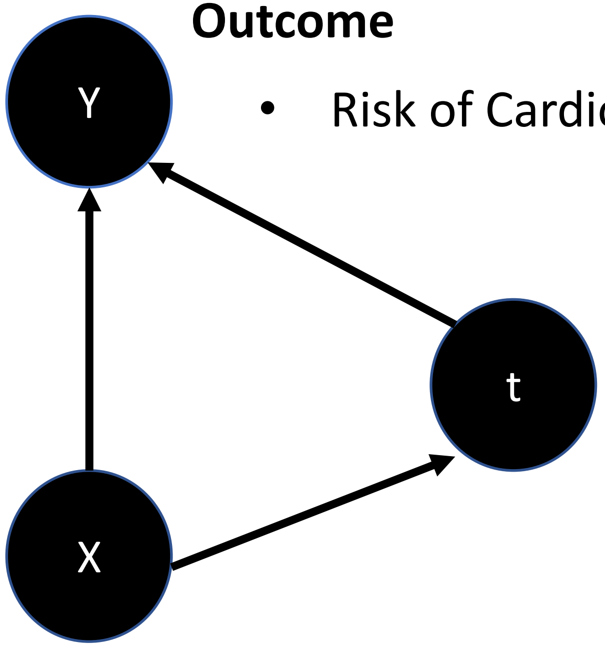
We study the treatment effect with the comparison of factual and counterfactual data

Causal inference challenge

Will statins reduce the risk of cardiovascular diseases?

Outcome

- Risk of Cardiovascular disease



Treatment (yes/no)

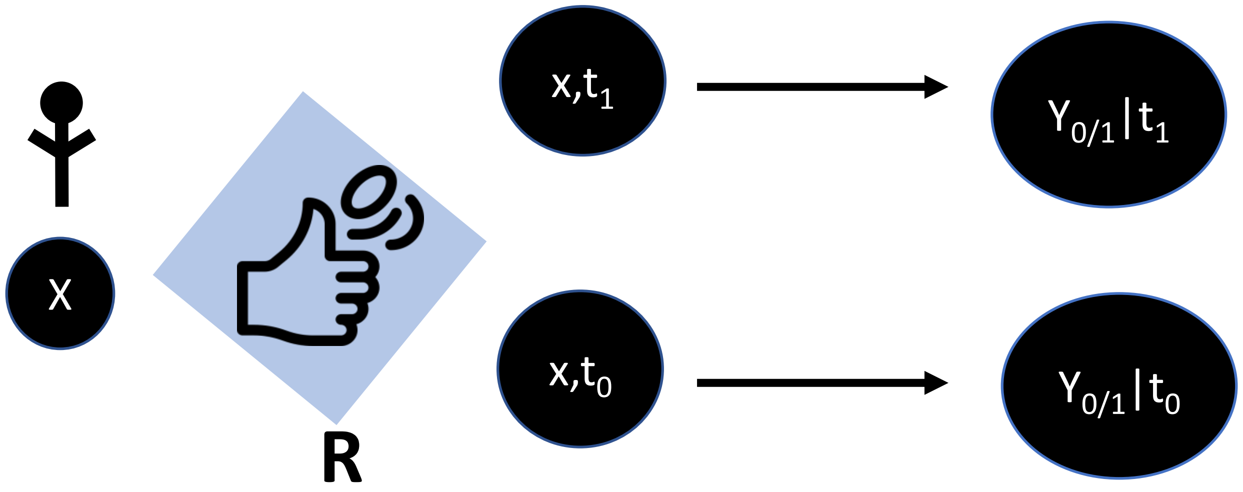
- Statins

Confounders

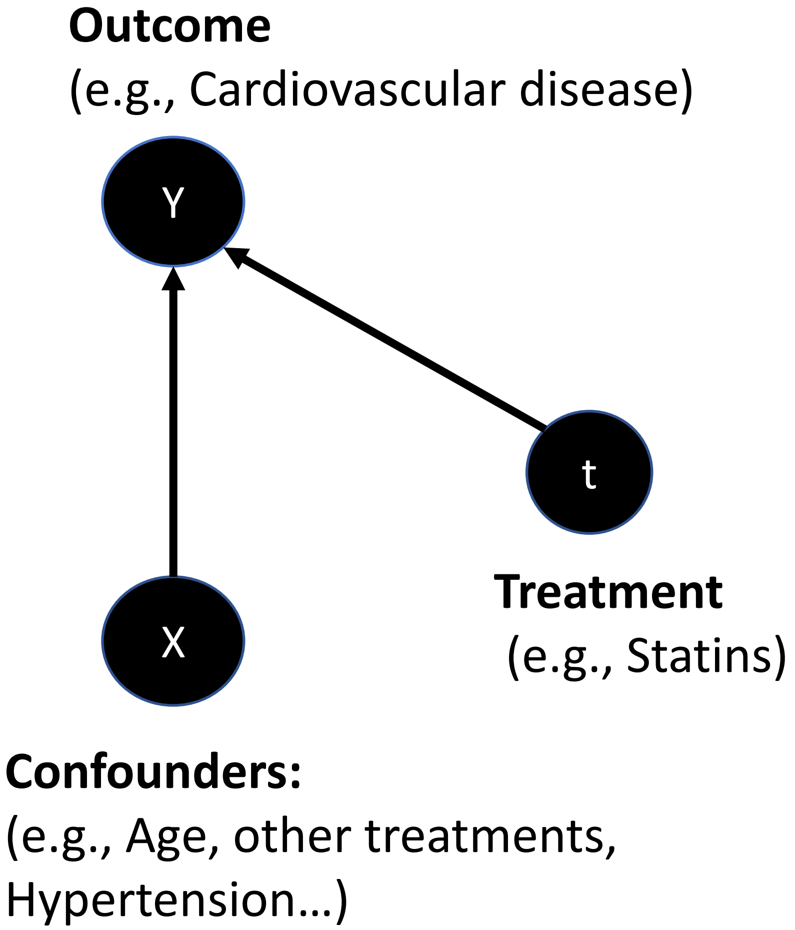
- Age
- Previous treatments
- Hypertension

Randomized Controlled Trials (RCTs)

- **Reductionist approach** to provide causal estimates for single/two treatments for single diseases



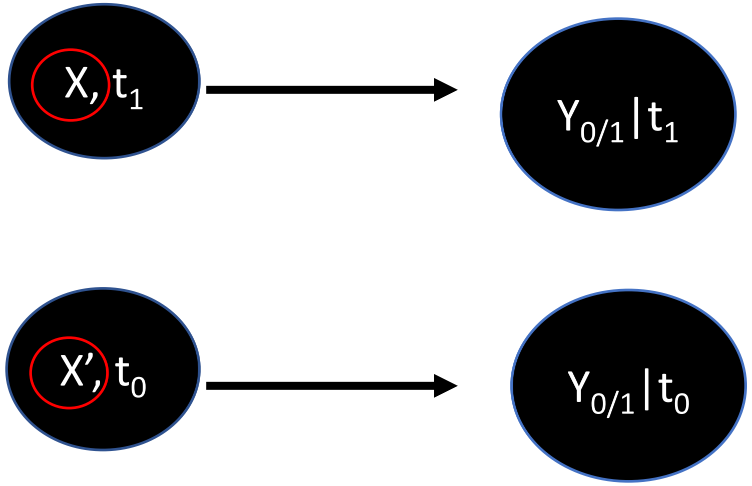
Design



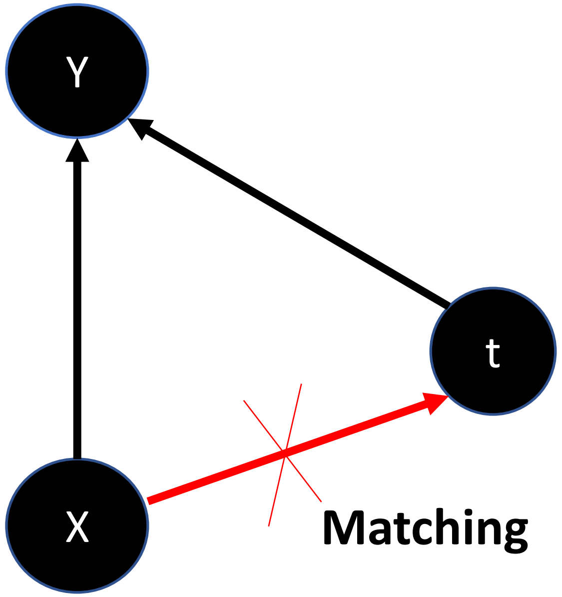
Causal structure in RCTs

What if no RCTs are available?

- **Observational data can be used to estimate causal effects**



Outcome
(e.g., Cardiovascular disease)

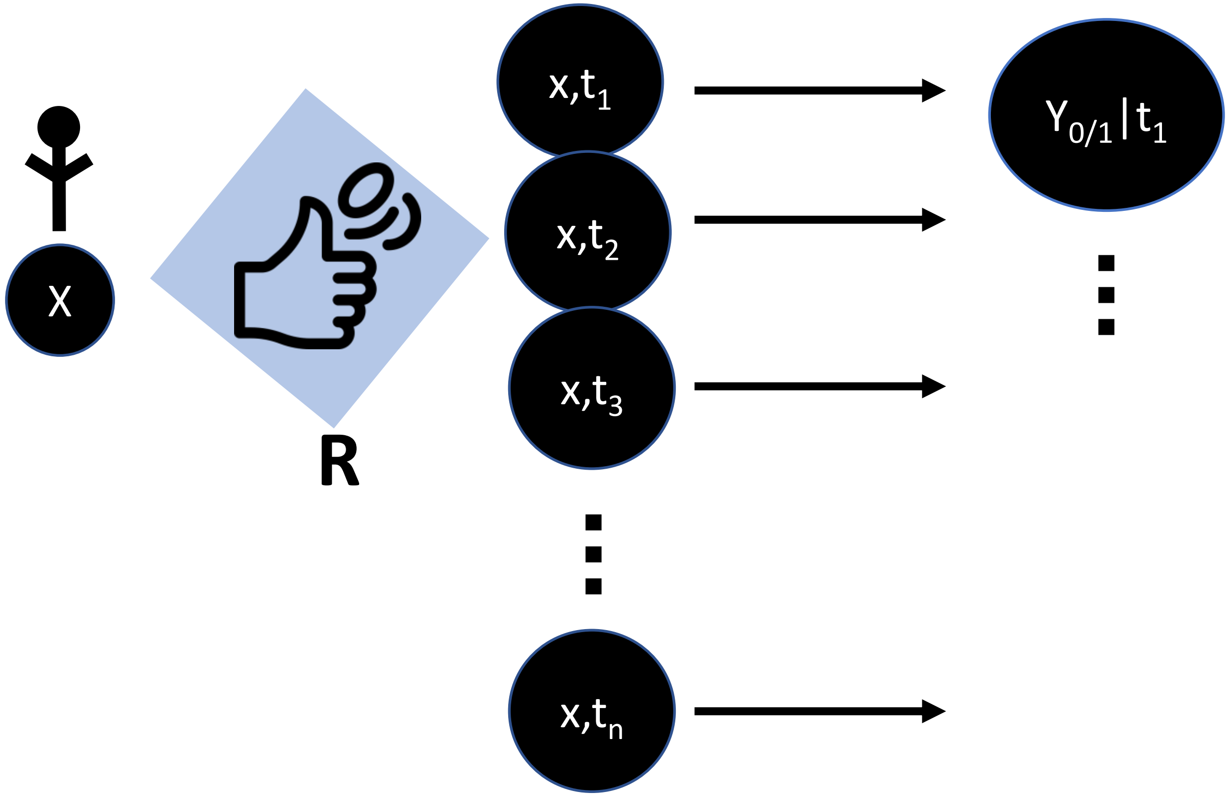


Treatment
(e.g., Statins)

Confounders:
(e.g., Age, other treatments, Hypertension…)

However, this is still reductionist approach

RCTs for multimorbidity study

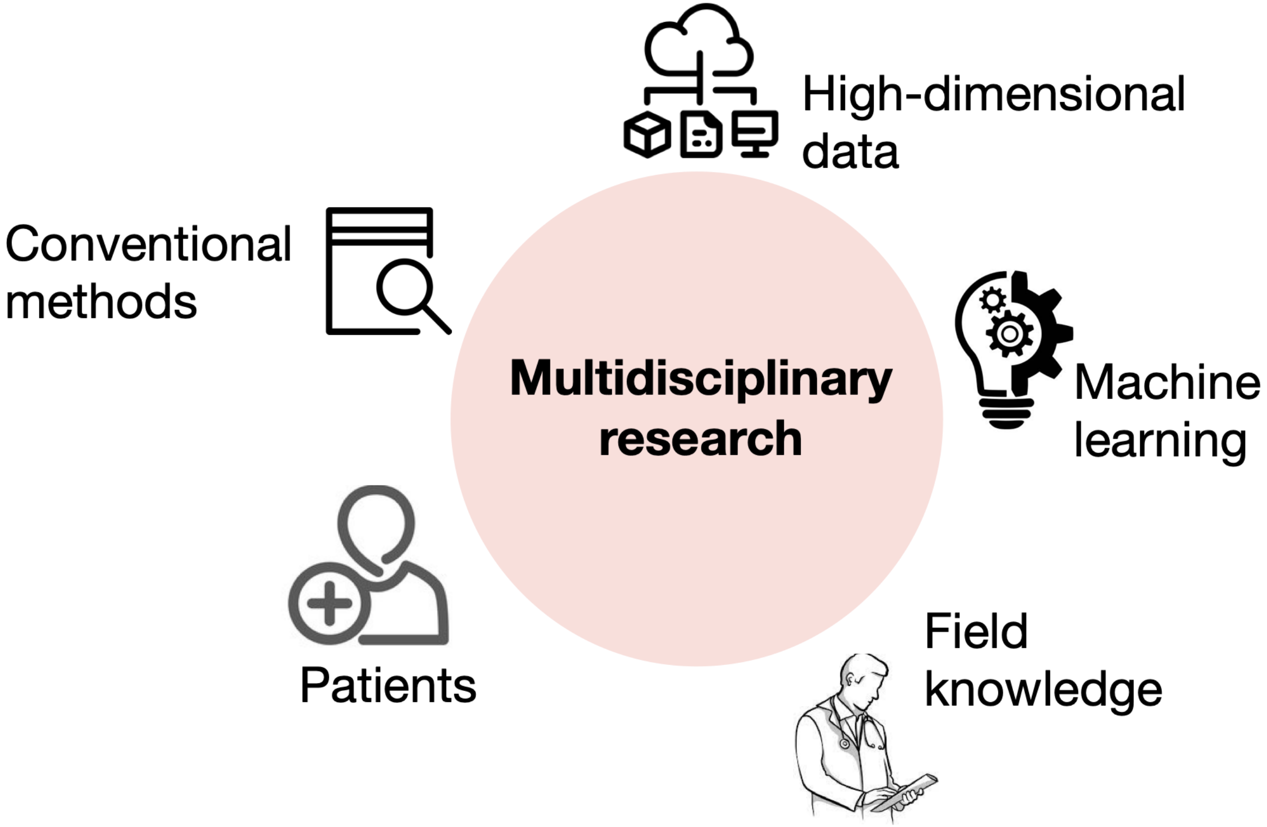


- **You need millions of people**
- **RCTs are not feasible nor fundable**

- Not good representation of all the multimorbid population (Selection Bias)
- The reductionist approach does not contemplate the relation confounder-treatment

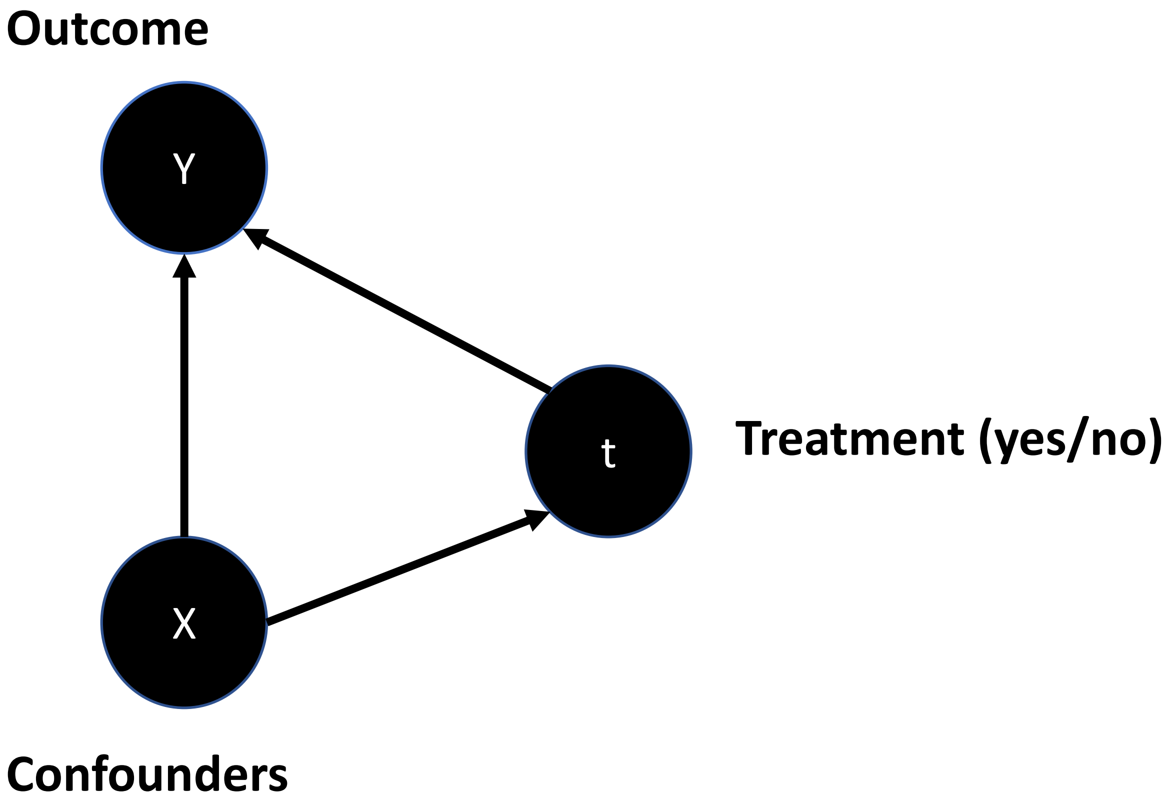
Solutions to the challenge

- ➢The high dimensional observatinal data are potential for generating causal inferences for **combinations** of treatments for multimobidity
- ➢Convergent efforts are needed



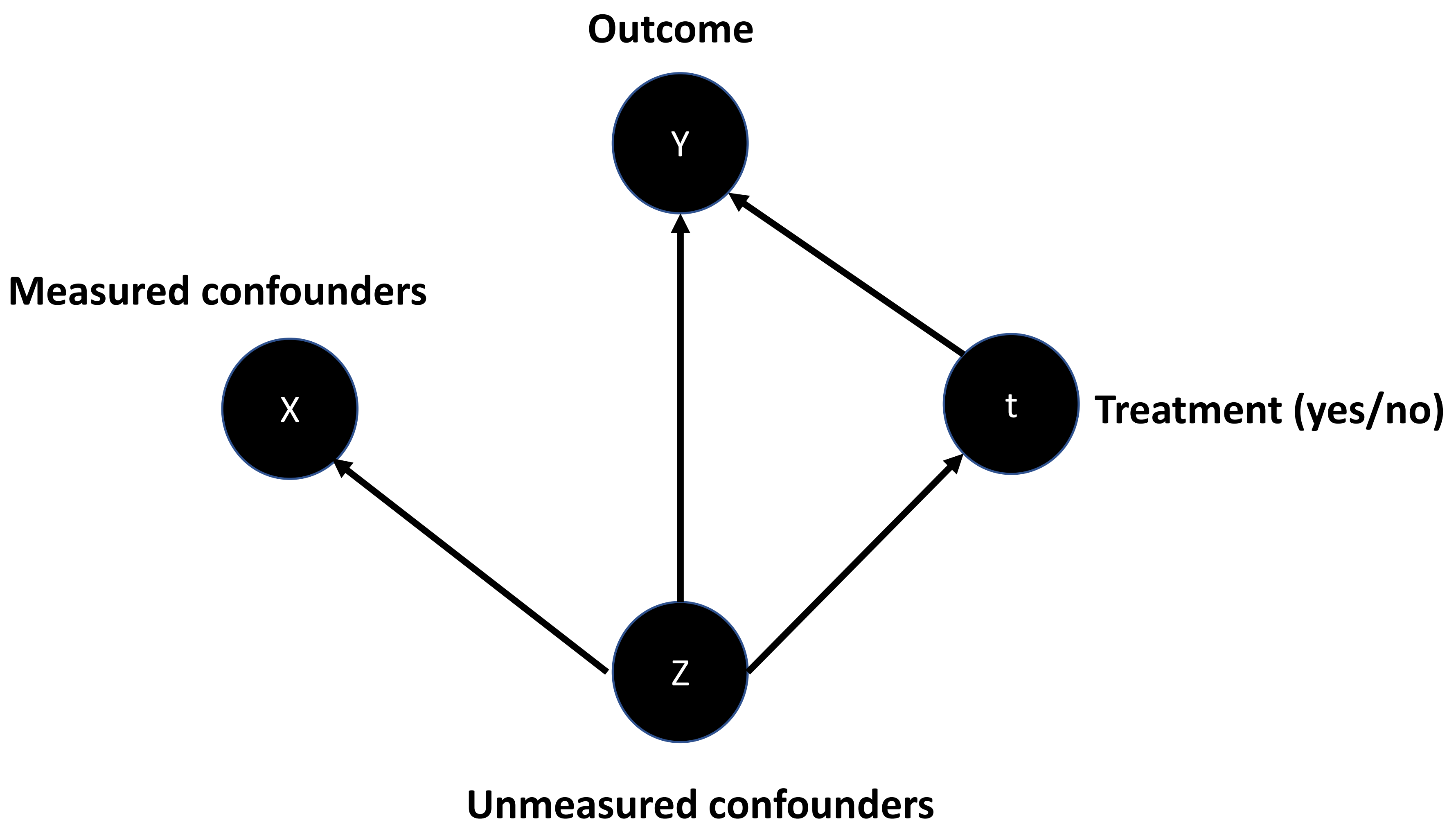
Machine Learning for Multimorbidity Causal Inference

Introduction to Causal Effect VAE



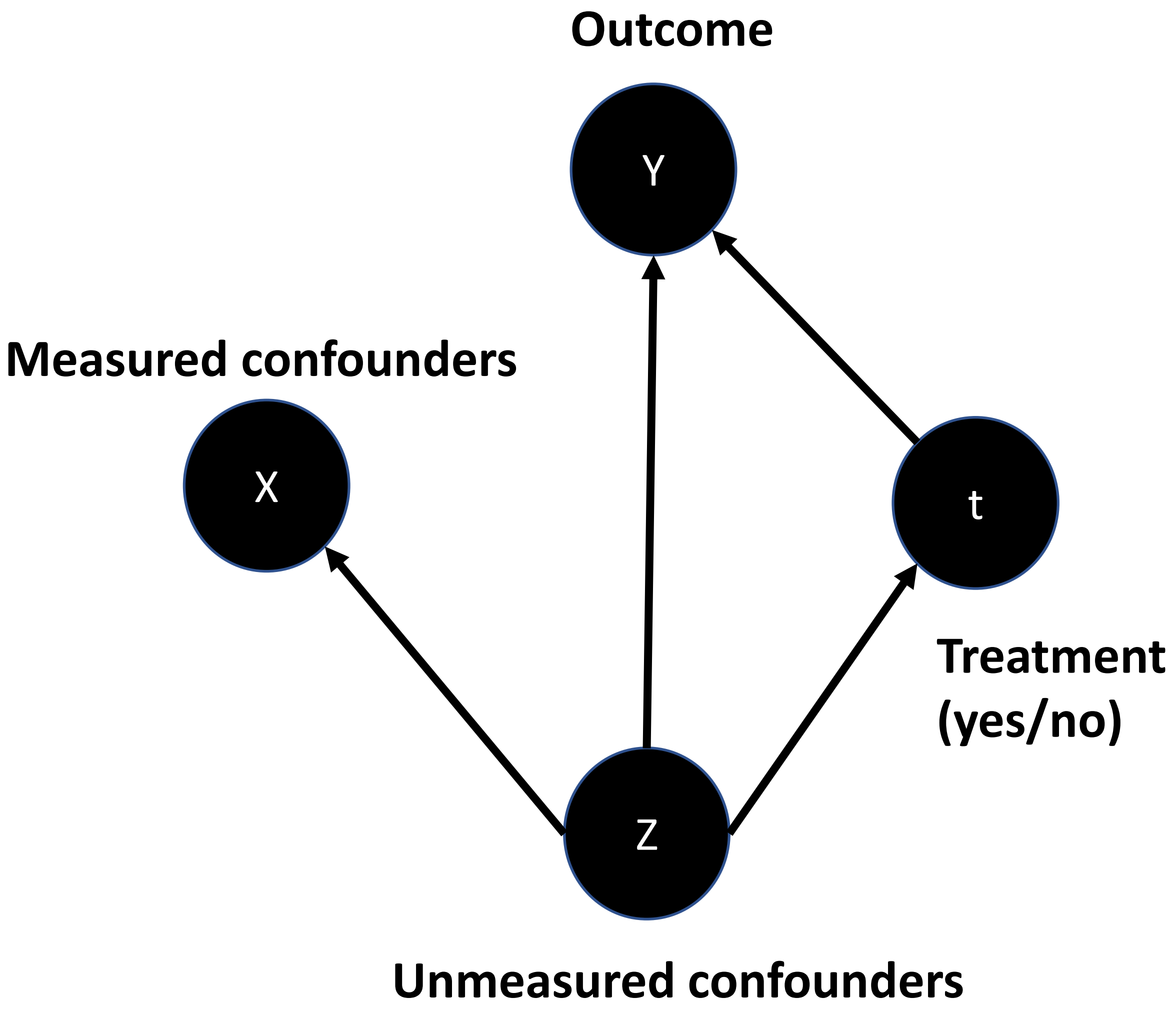
- The patient observable confounders represents partially the full reality
- Selection Bias since x is a noisy version of the confounders
- Heterogeneous groups of patients

Introduction to Causal Effect VAE



Introducing a full picture of the reality reduces biased models!

Objectives of the Causal Effect VAE

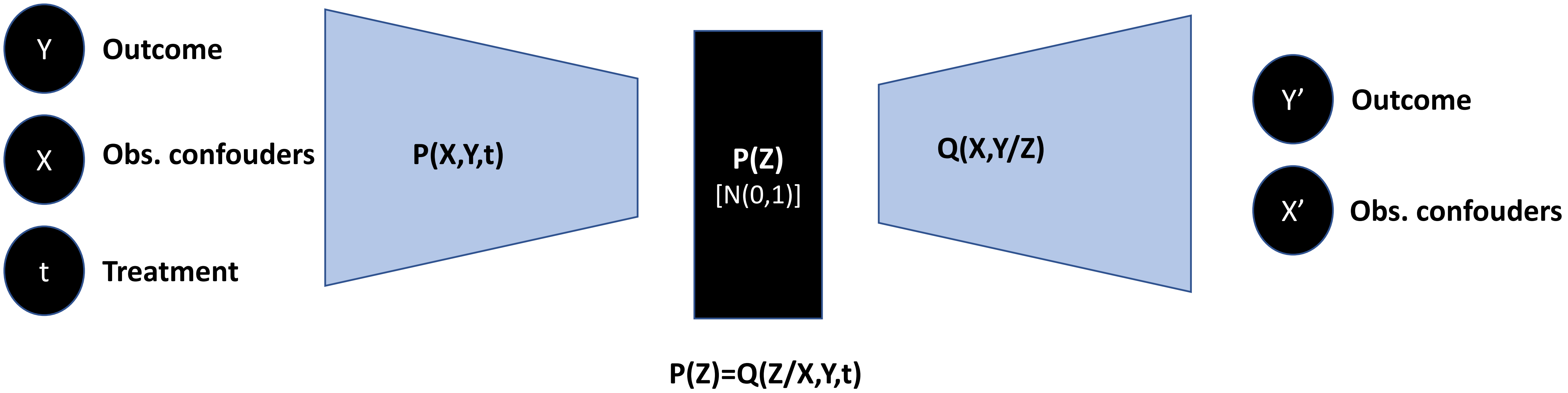


Objectives

- **P(X,Y,t):** adquire all the combinations of X,Y,t
- **Q(Y/Z,t):** Predict factual & conterfactual data to understand the causal effect of t

Idea behind Causal Effect VAE

1) Understand the causality and map it into a gaussian distribution

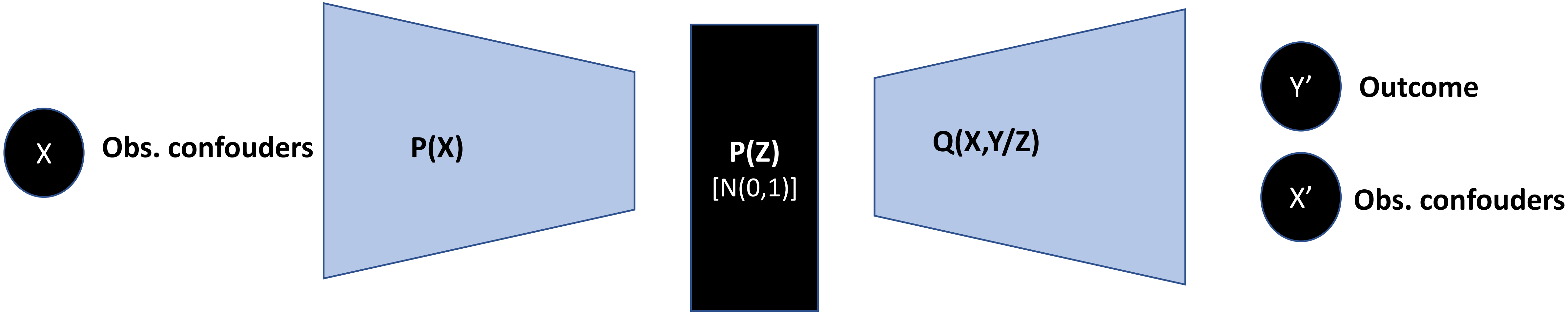


What a VAE offers:

- Extract Z from X,Y,t (p(Z/X,Y,t)): acquire all the data responsible for the outcome Y
- Predict Y (q(Y/Z,t)): have conterfactual data to understand how t affects the outcome Y

Idea behind Causal Effect VAE

2) Map it into the same gaussian distribution only samples of observational data



P(Z)=Q(Z/X) =Q(Z/X,Y,t)

What a CEVAE offers:

- Extract Z from X (p(Z/X,Y,t)): acquire all the data responsible for the outcome Y with only X
- Predict Y (q(Y/Z,t)): generate factual & conterfactual data for X to predict the causal effect for a single patient

Benefits of Causal Effect VAE

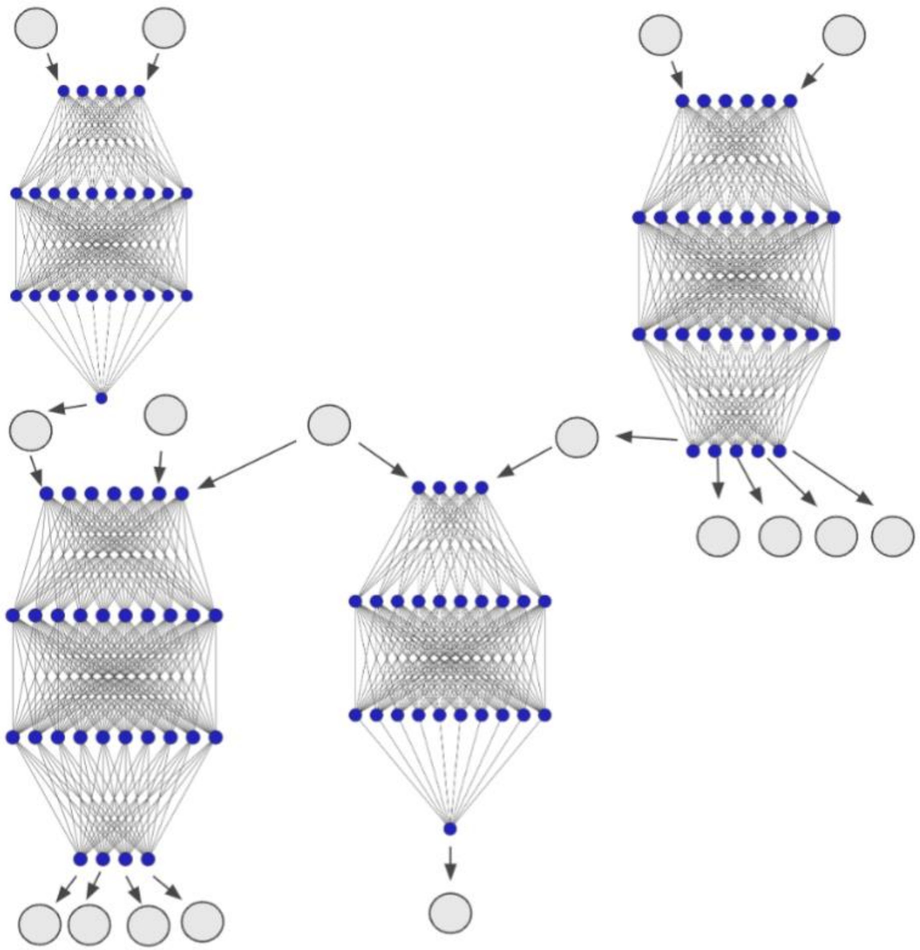
- Fully representation of the confounders
- Works well with heterogenous groups
- Very good for multidimensional problems (for more than one treatment, for time dependent treatments, etc)
- Useful for benefit-harm studies where with drugs interaction through time
- Interpretability

Perfect tool to select treatment combinations with the best possible outcome for multimorbidity patients

Machine Learning for Multimorbidity Causal Inference: Timeline

Plans and status

- Generated synthetic data with a toy model for patients (controled environment)
- We are now in the process of understanding the CEVAE with the toy data
- Test the model in a low dimensional case and compare it to traditional methods
- Scale to the fully multimorbidity case



CERN