Causal Data Science:

Estimating Identifiable Causal Effects

Yonghan Jung

Department of Computer Science

Purdue University

www.yonghanjung.me

KAIST | Apr. 25, 2025

Causal Data Science:

Estimating Identifiable Causal Effects

Yonghan Jung

Department of Computer Science

Purdue University

www.yonghanjung.me

- **Overview**
- **Less Technical**
- **Introduction at a broad & intuitive level**

KAIST | Apr. 25, 2025

“ *Remdesivir use is associated with lower mortality in patients with COVID*

Clinical Infectious Diseases, 2019

“ *Remdesivir use is associated with lower mortality in patients with COVID*

Clinical Infectious Diseases, 2019

“ *Remdesivir becomes first Covid-19 treatment to receive FDA approval*

CNN, 2020

“Remdesivir use is associated with lower mortality in patients with COVID

Clinical Infectious Diseases, 2019

“Remdesivir becomes first Covid-19 treatment to receive FDA approval

CNN, 2020

“WHO recommends against use of Remdesivir for COVID patients

CNN, 2020

What’s going on?

Story Behind the Data

Observational Study (FDA)

	Mortality Rate
Remdesivir	11%
Non Remdesivir	20%

Positive Correlation with Lower Mortality

vs.

Randomized Trial (WHO)

	Mortality Rate
Remdesivir	15%
Non Remdesivir	15%

No Causal Effect to Lower Mortality

Story Behind the Data

Since Remdesivir costs over $2000, wealthier patients are more likely to receive it.

Observational Study (FDA)

	Mortality Rate
Remdesivir	11%
Non Remdesivir	20%

Positive Correlation with Lower Mortality

vs.

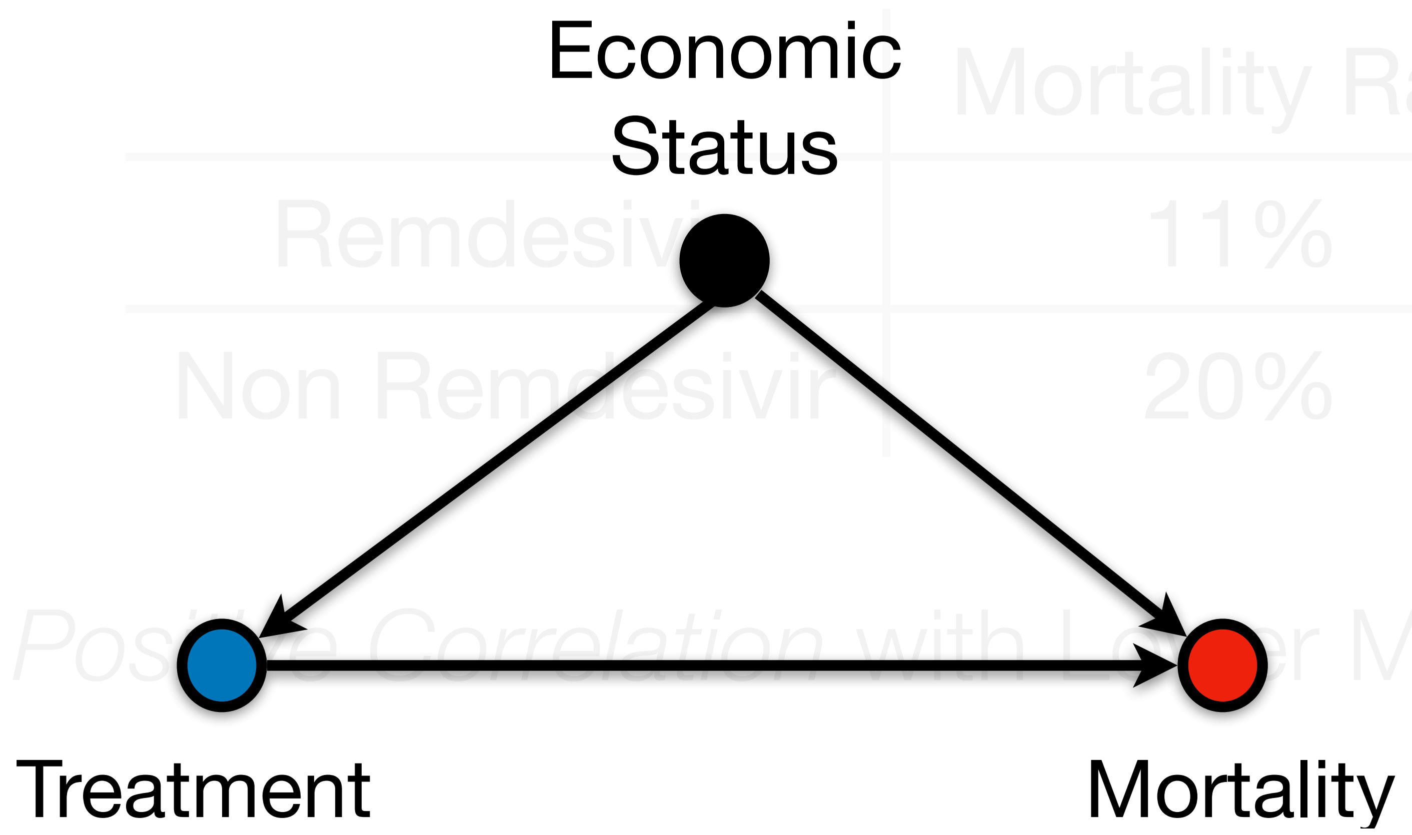
Randomized Trial (WHO)

	Mortality Rate
Remdesivir	15%
Non Remdesivir	15%

No Causal Effect to Lower Mortality

Story Behind the Data

Observational Study (FDA)



vs.

Randomized Trial (WHO)

	Mortality Rate
Remdesivir	15%
Non Remdesivir	15%

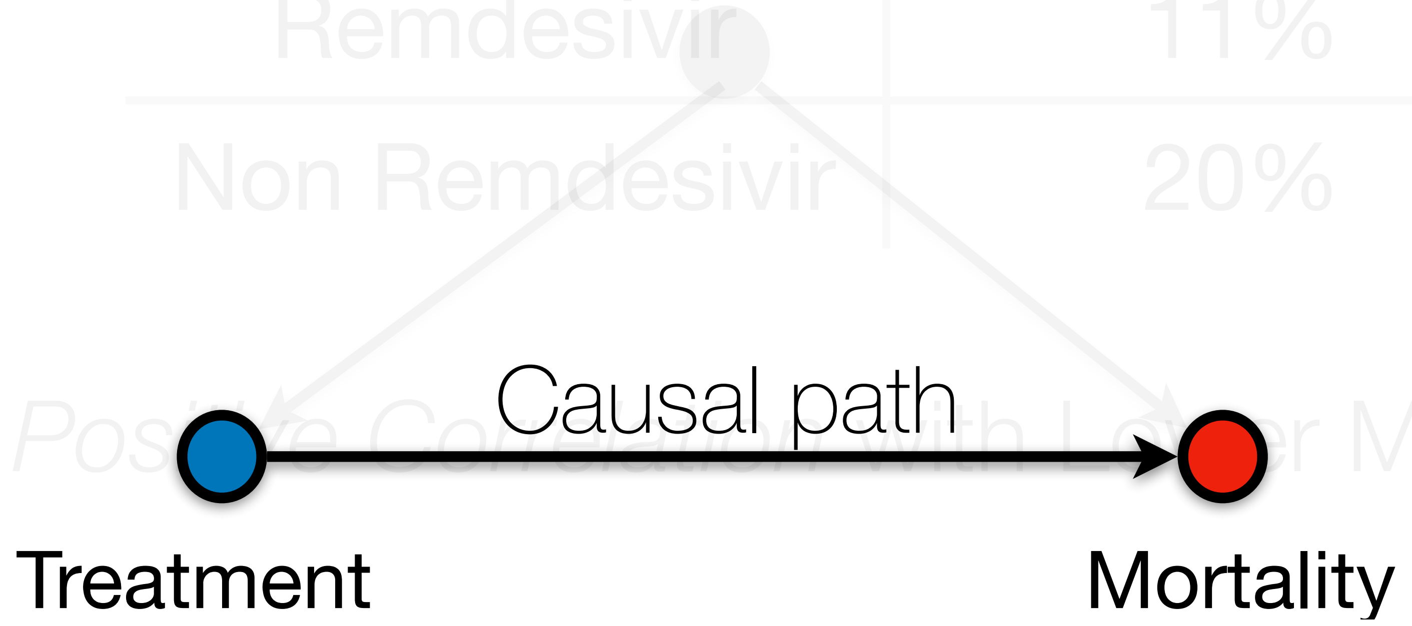
No Causal Effect to Lower Mortality

Story Behind the Data

Observational Study (FDA)

	Mortality Rate
Remdesivir	11%
Non Remdesivir	20%

Positive Correlation with Lower Mortality



vs.

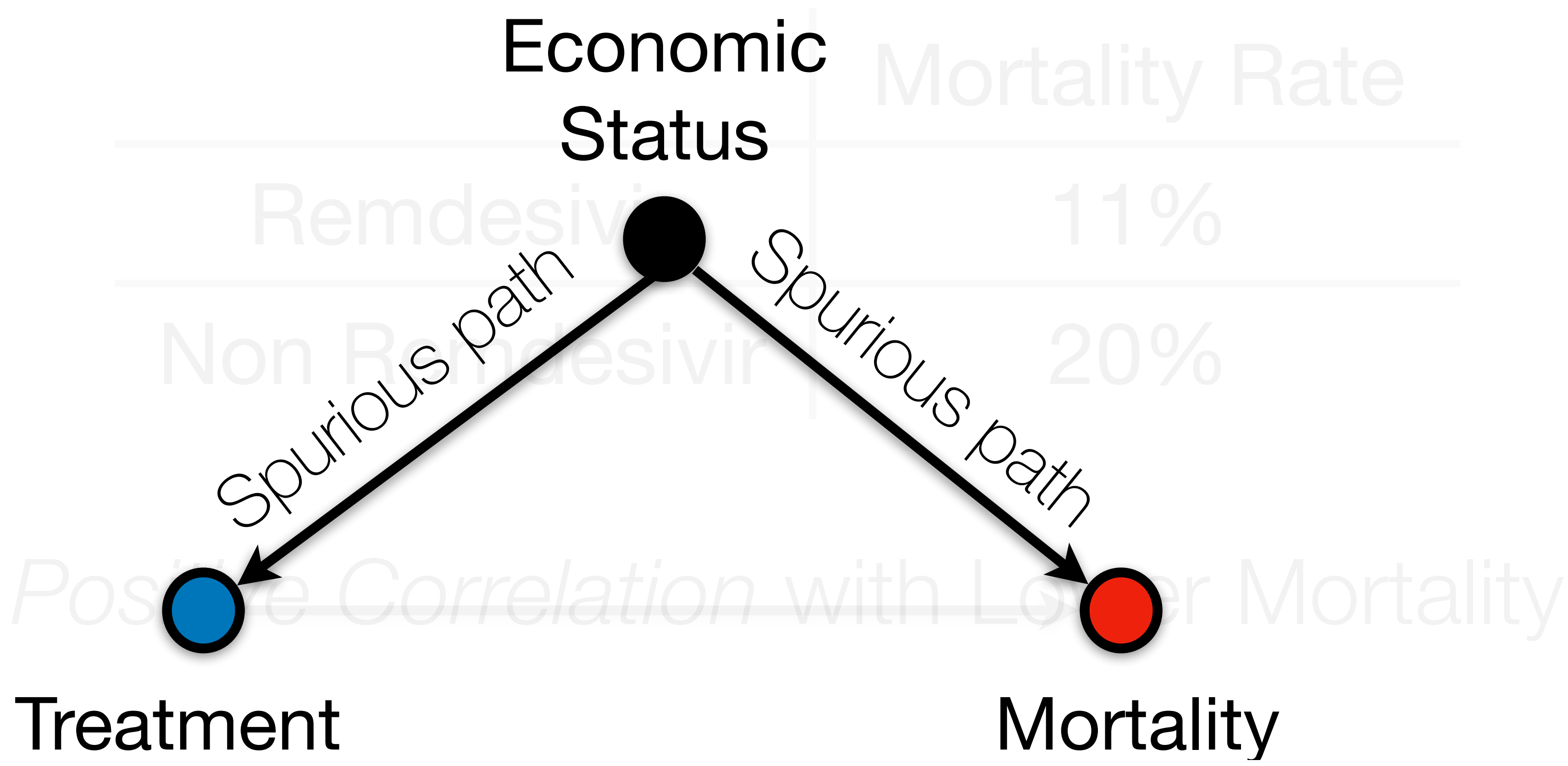
Randomized Trial (WHO)

	Mortality Rate
Remdesivir	15%
Non Remdesivir	15%

No Causal Effect to Lower Mortality

Story Behind the Data

Observational Study (FDA)



vs.

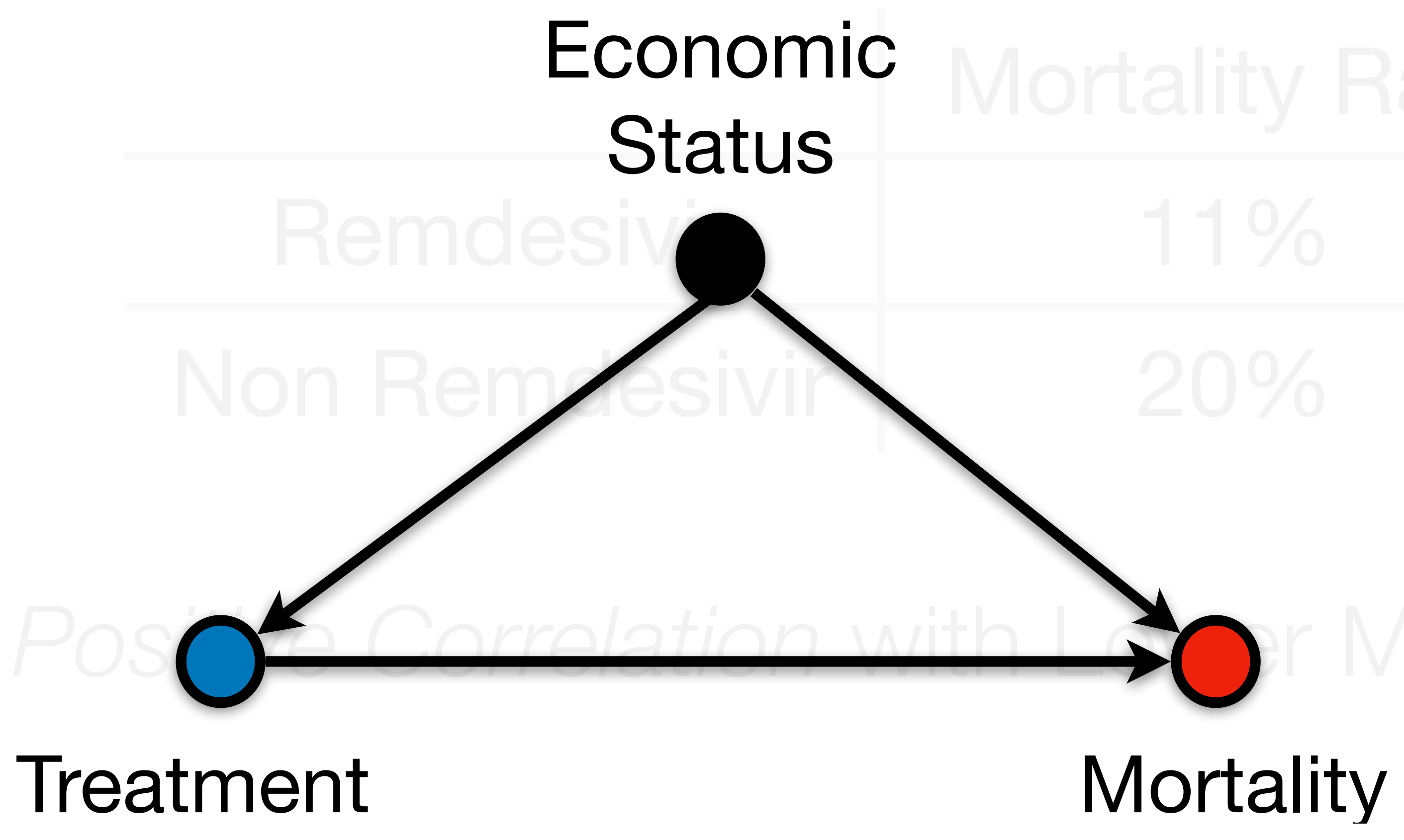
Randomized Trial (WHO)

	Mortality Rate
Remdesivir	15%
Non Remdesivir	15%

No Causal Effect to Lower Mortality

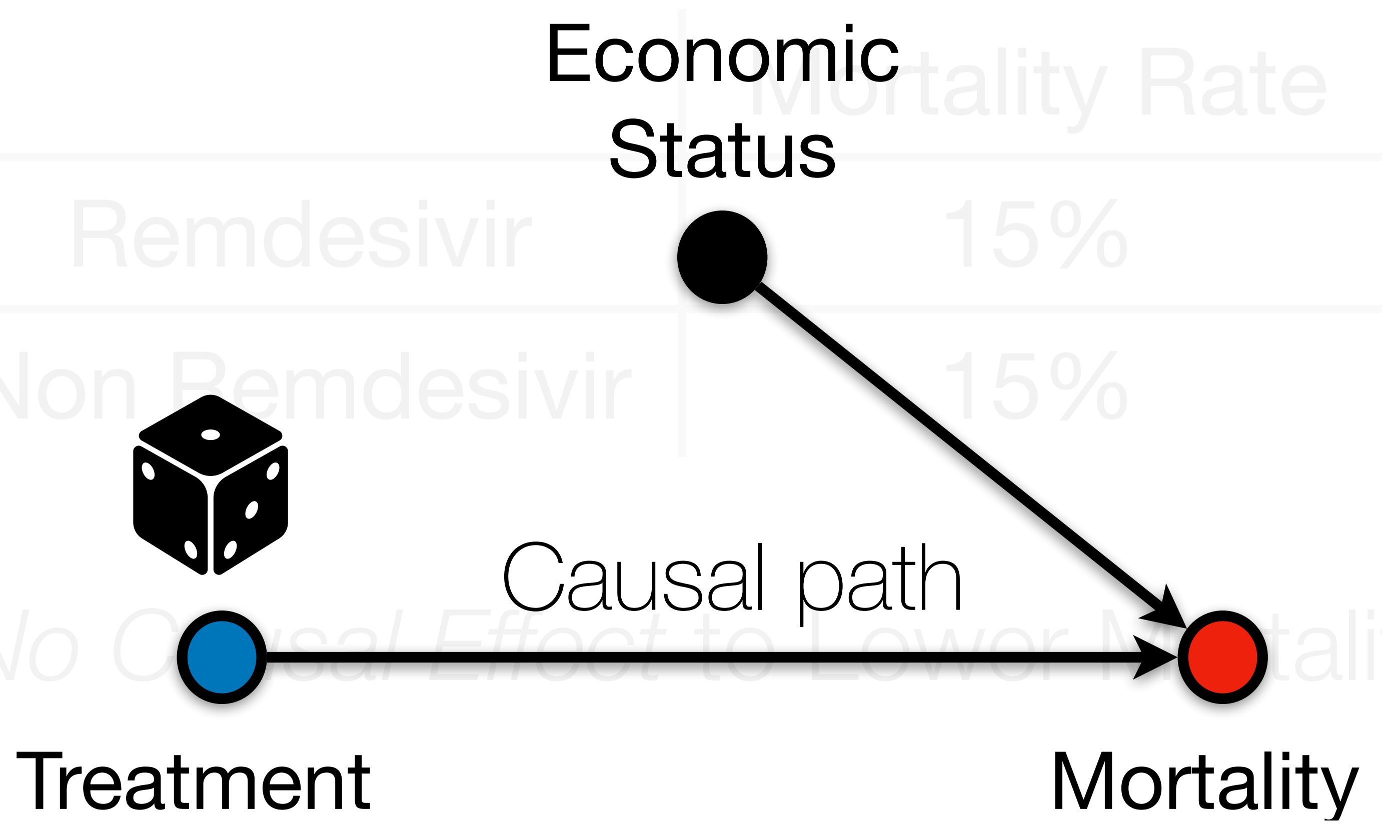
Story Behind the Data

Observational Study (FDA)



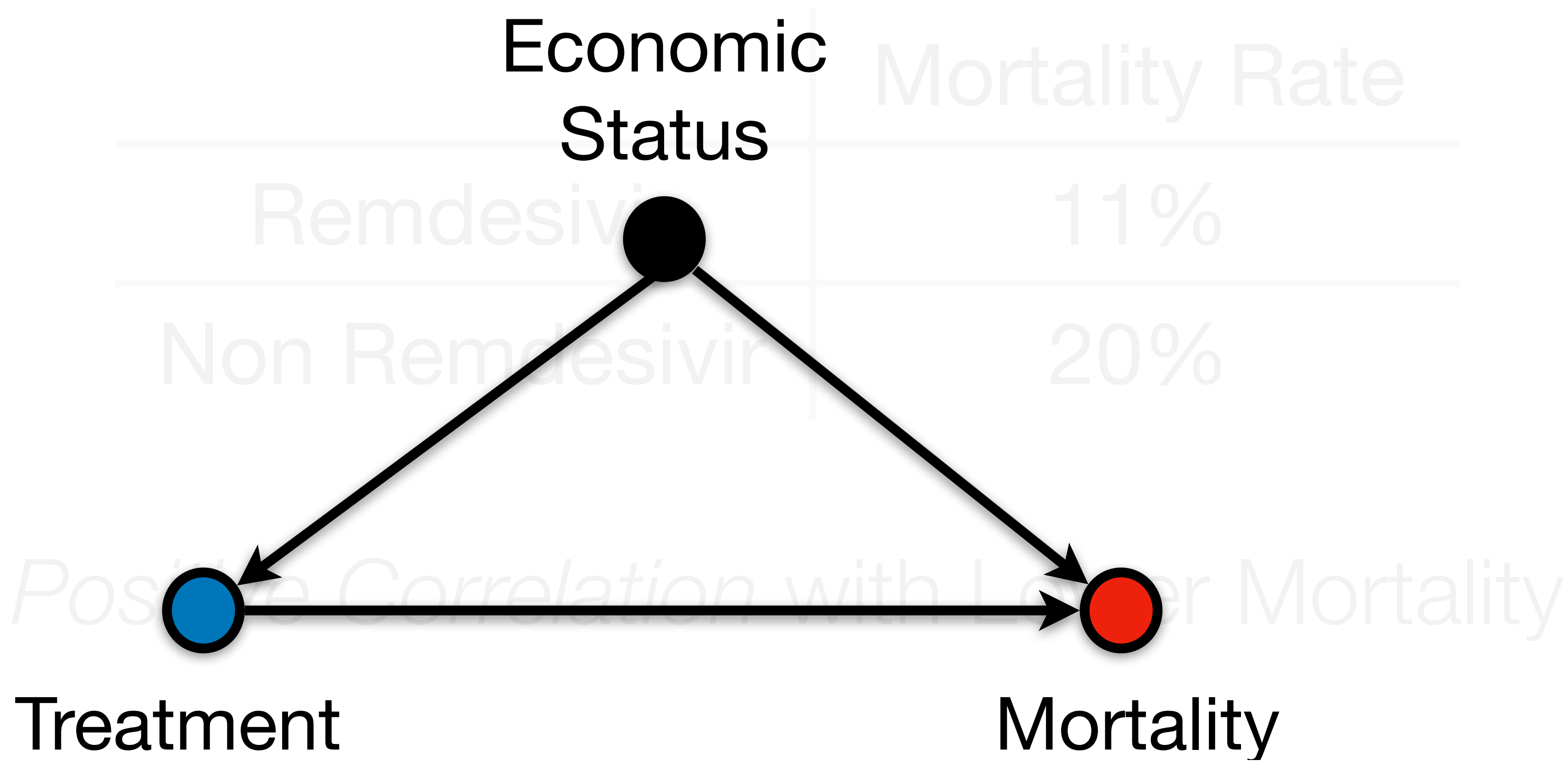
vs.

Randomized Trial (WHO)



Story Behind the Data

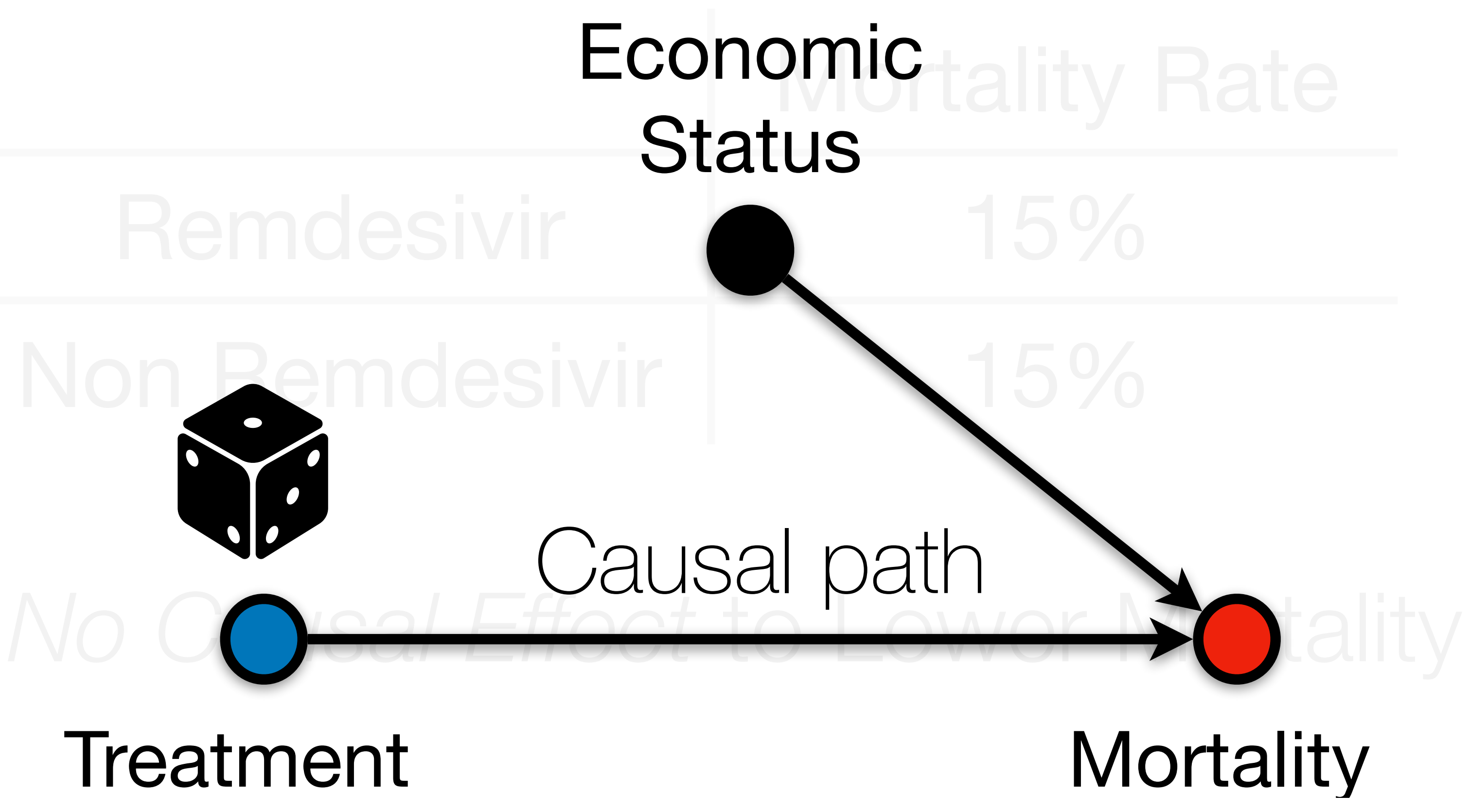
Observational Study (FDA)



- Feasible, Accessible
- Confounding bias

vs.

Randomized Trial (WHO)

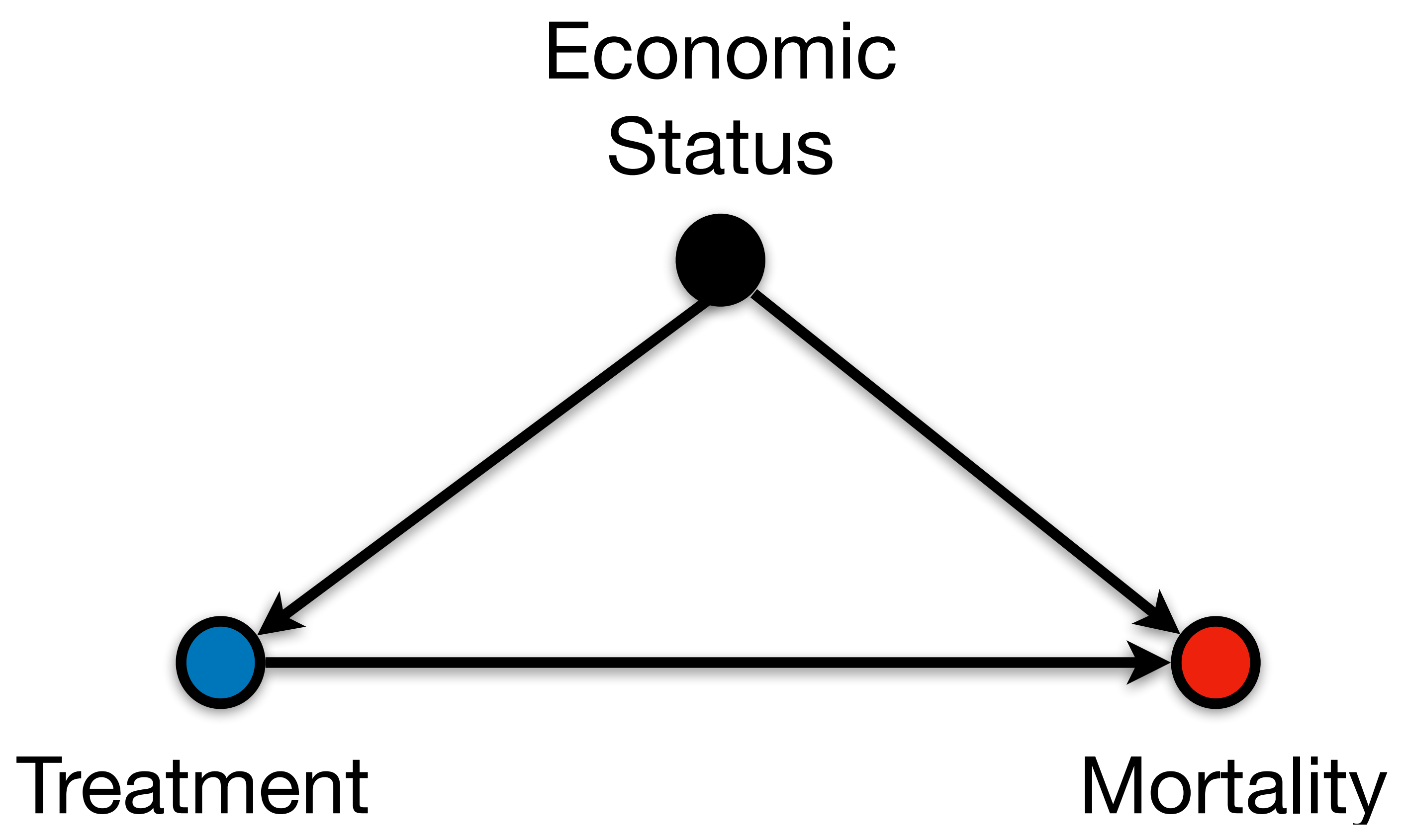


- Gold standard in causal inference
- Expensive, Infeasible

Story Behind the Data

Observational Study (FDA)

	Mortality Rate
Remdesivir	11%
Non Remdesivir	20%



"Causal Inference Pipeline"

Causal Effect

	Mortality Rate
Remdesivir	15%
Non Remdesivir	15%

Standard Causal Inference Pipeline

Standard Causal Inference Pipeline

Input

Effect (Q)

$\mathbb{E}[Y \mid \mathrm{do}(x)]$

Graph

Samples

D from a distribution P

Standard Causal Inference Pipeline

Input

Effect (Q)

$\mathbb{E}[Y \mid \mathrm{do}(x)]$

expected outcome on intervening/fixing treatment X to x

$$\mathbb{E}[Y \mid \mathrm{do}(X{=}x)]$$

Graph

Samples

D from a distribution P

Standard Causal Inference Pipeline

Input

Effect (Q)

$\mathbb{E}[Y \mid \mathrm{do}(x)]$

Graph

Encode a story (or assumptions) behind the dataset

Samples

D from a distribution P

Standard Causal Inference Pipeline

Input

Effect (Q)

$\mathbb{E}[Y \mid \mathrm{do}(x)]$

Graph

Samples

D from a distribution P

Identification

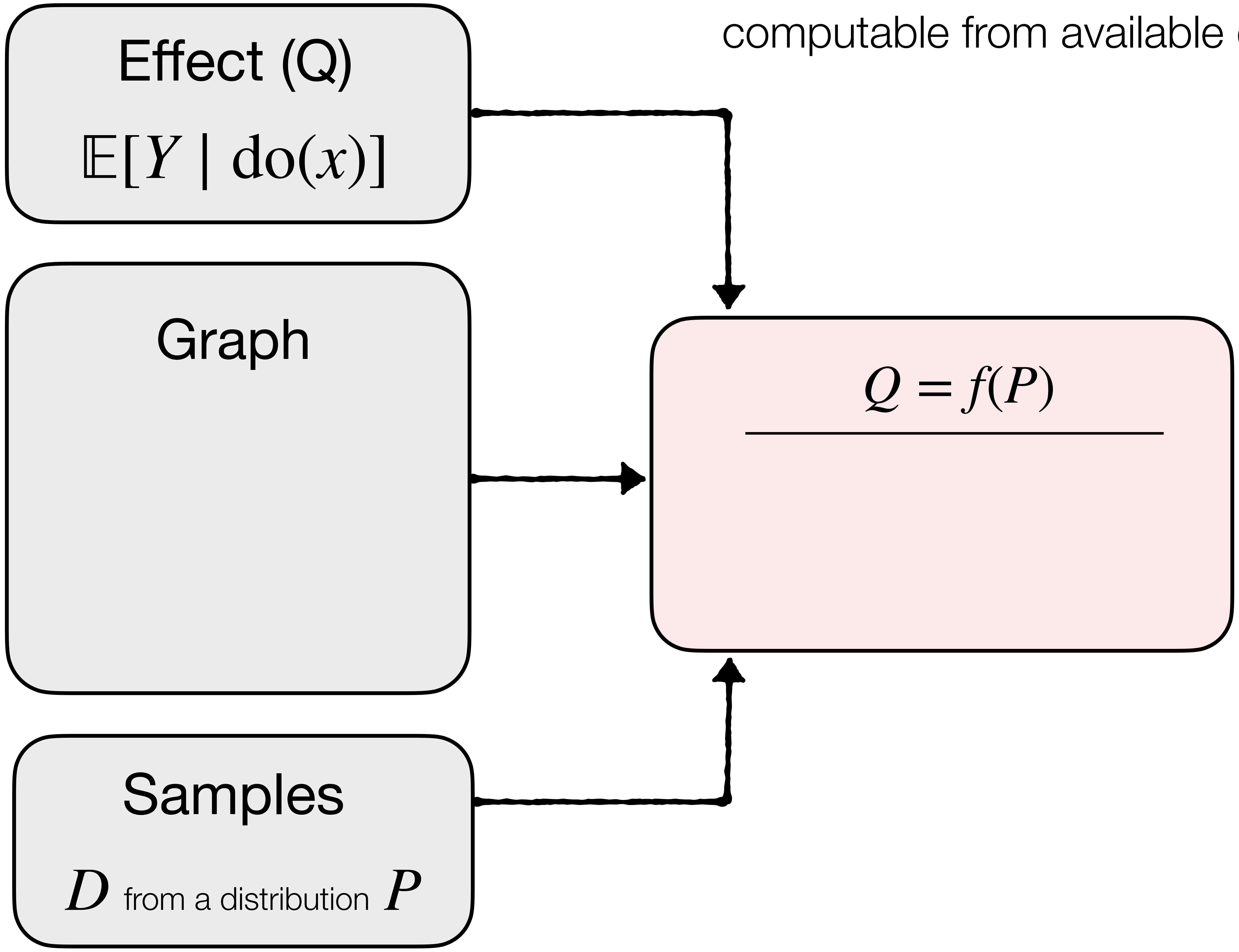
"When is the causal effect computable from available data?"

Standard Causal Inference Pipeline

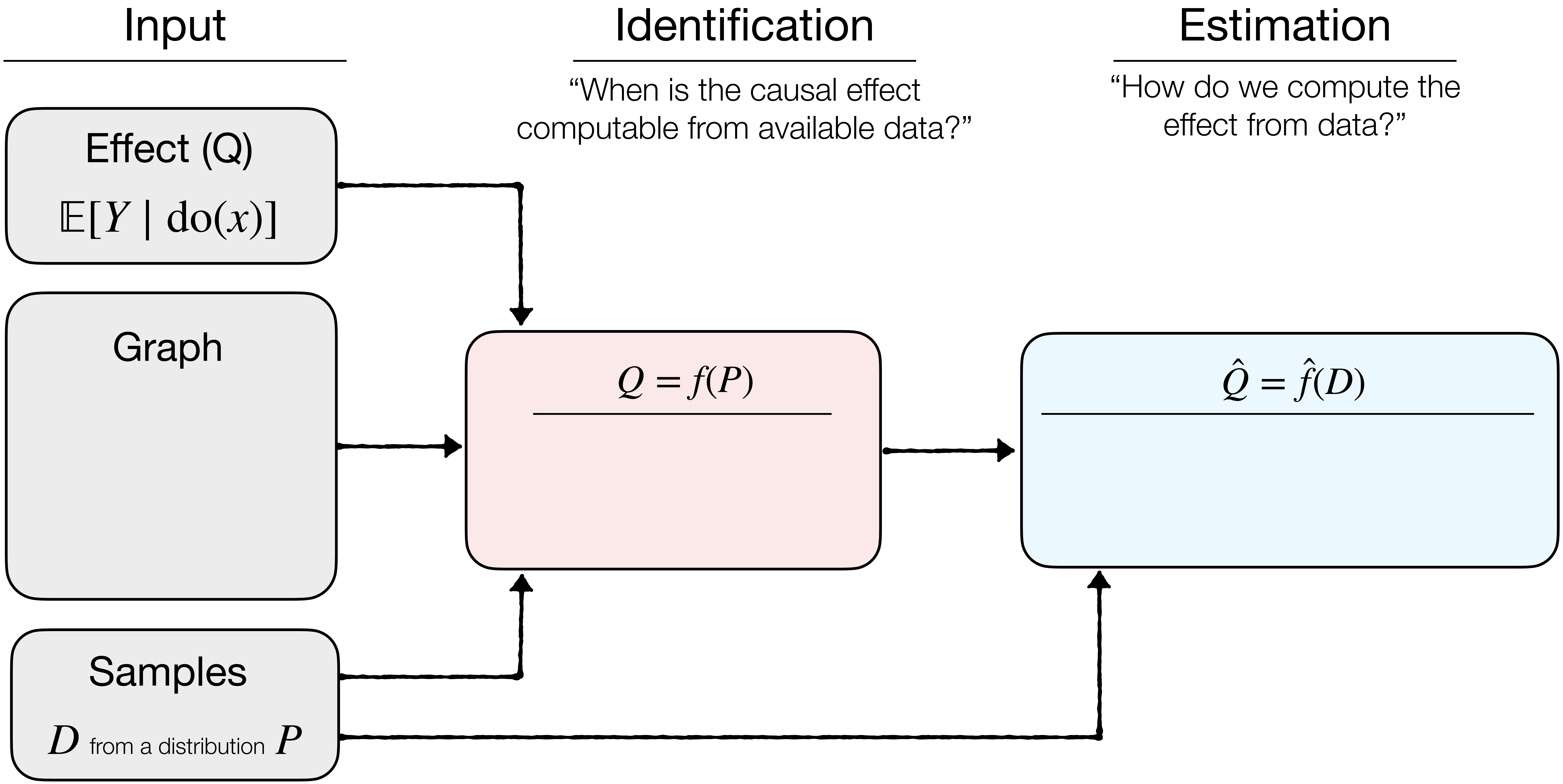
Input

Identification

"When is the causal effect computable from available data?"



Standard Causal Inference Pipeline



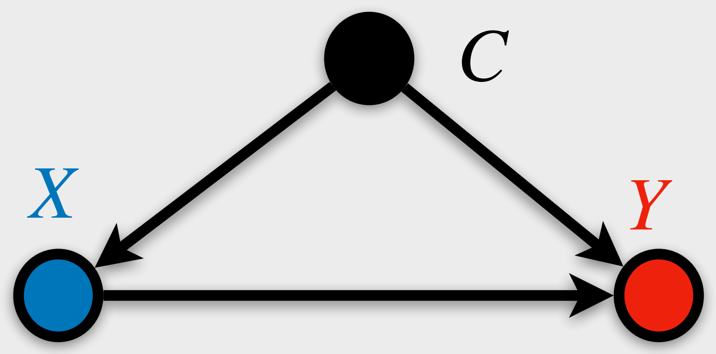
Standard Causal Inference Pipeline

Input

Effect (Q)

$\mathbb{E}[Y \mid \mathrm{do}(x)]$

Graph



"Back-Door"

Samples

D from a distribution P

Identification

"When is the causal effect computable from available data?"

$Q = f(P)$

Estimation

"How do we compute the effect from data?"

$\hat{Q} = \hat{f}(D)$

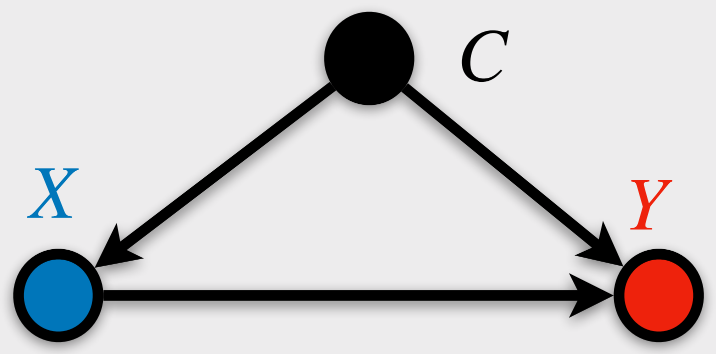
Standard Causal Inference Pipeline

Input

Effect (Q)

$\mathbb{E}[Y \mid \mathrm{do}(x)]$

Graph



"Back-Door"

Samples

D from a distribution P

Identification

"When is the causal effect computable from available data?"

$$Q = f(P)$$

$$Q = \sum_c \mathbb{E}[Y \mid x, c] P(c)$$

"Back-Door Adjustment"

Pearl, 95

Estimation

"How do we compute the effect from data?"

$$\hat{Q} = \hat{f}(D)$$

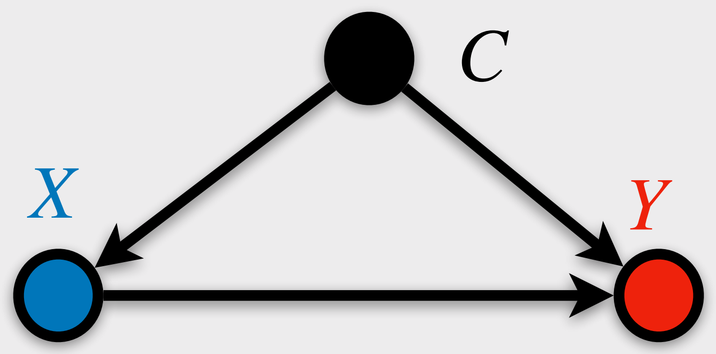
Standard Causal Inference Pipeline

Input

Effect (Q)

$\mathbb{E}[Y \mid \mathrm{do}(x)]$

Graph



"Back-Door"

Samples

D from a distribution P

Identification

"When is the causal effect computable from available data?"

$$Q = f(P)$$

$$Q = \sum_c \mathbb{E}[Y \mid x, c] P(c)$$

"Back-Door Adjustment"

Pearl, 95

Estimation

"How do we compute the effect from data?"

$$\hat{Q} = \hat{f}(D)$$

Double Machine Learning (DML)

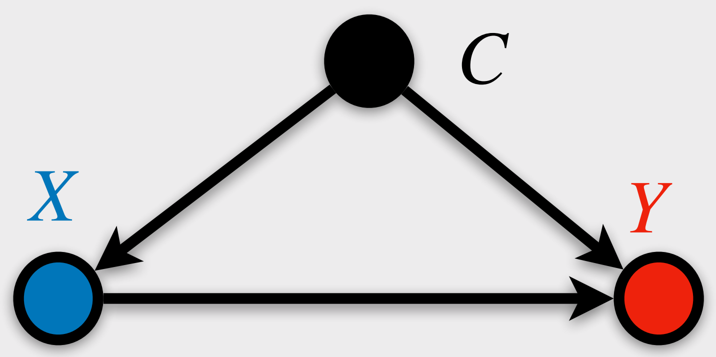
Chernozhukov et al, 2018

Challenges in Causal Inference

Effect (Q)

$\mathbb{E}[Y \mid \mathrm{do}(x)]$

Graph



"Back-Door"

Samples

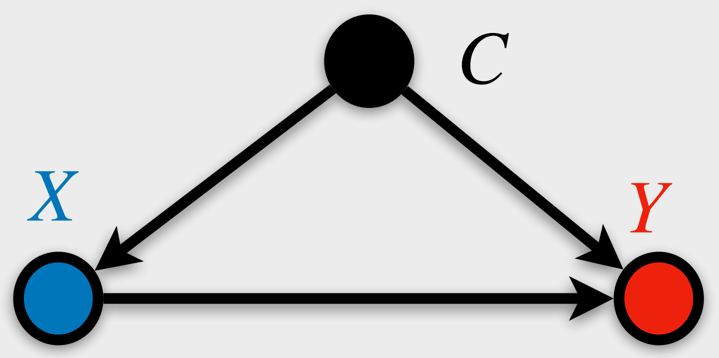
D from P

Challenges in Causal Inference

Effect (Q)

$\mathbb{E}[Y \mid \mathrm{do}(x)]$

Graph



"Back-Door"

Samples

D from P

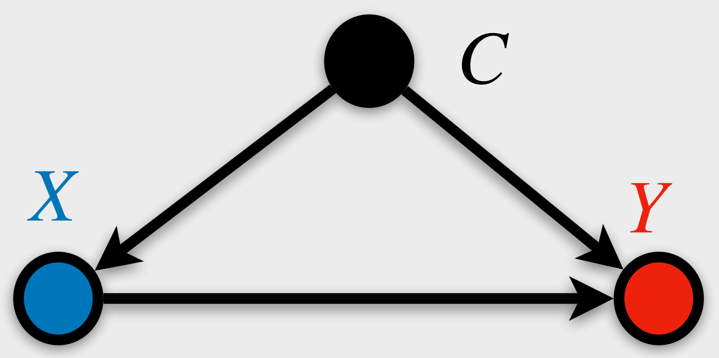
1 Complex dependences

Challenges in Causal Inference

Effect (Q)

$\mathbb{E}[Y \mid \mathrm{do}(x)]$

Graph

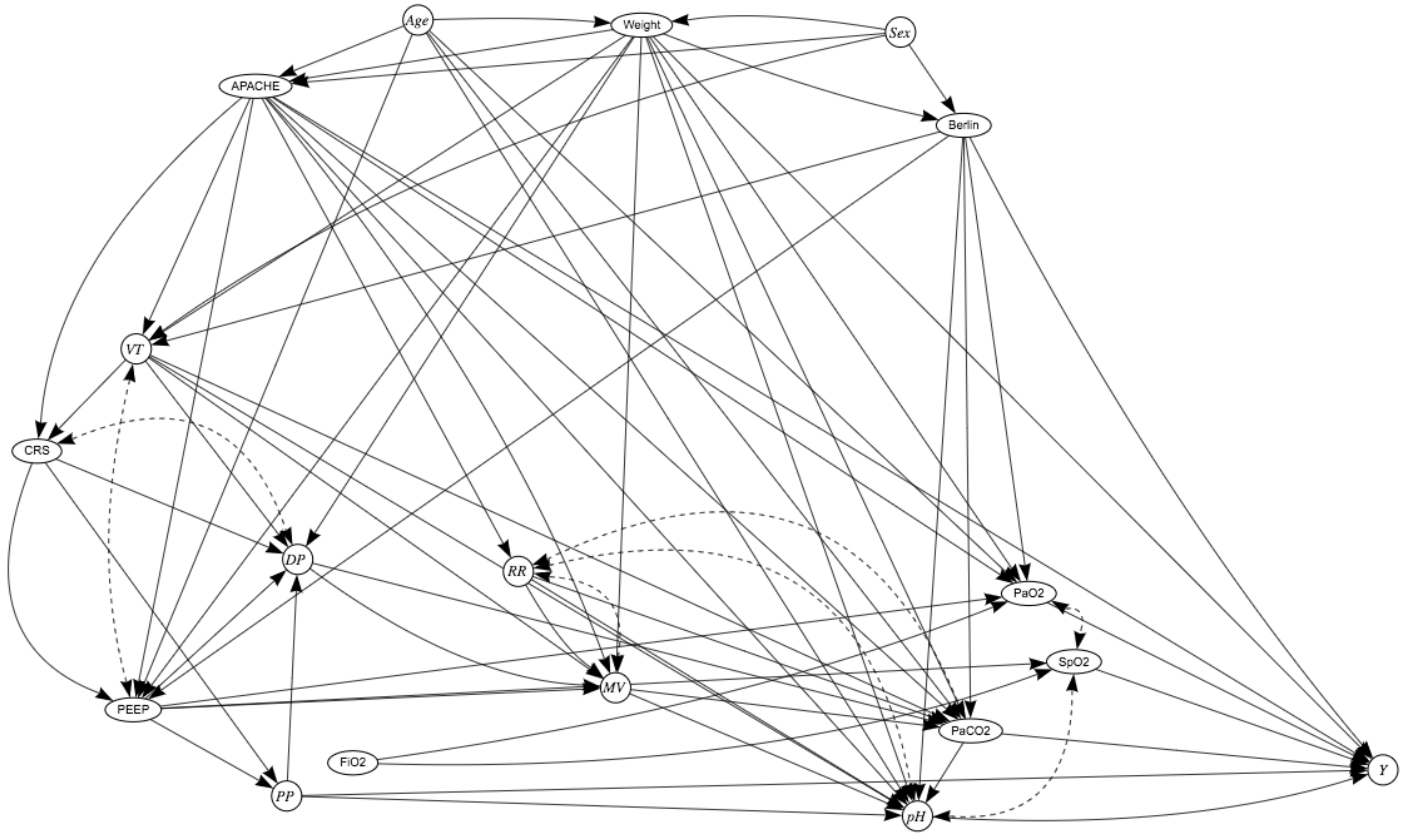


"Back-Door"

Samples

D from P

1 Complex dependences



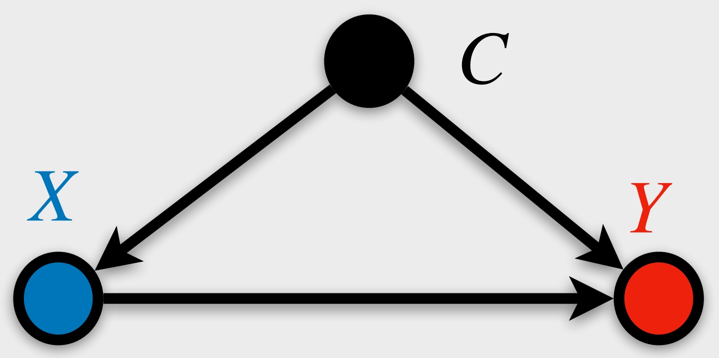
Causal graph on acute respiratory distress syndrome (ARDS)

Challenges in Causal Inference

Effect (Q)

$\mathbb{E}[Y \mid \mathrm{do}(x)]$

Graph



“Back-Door”

Samples

D from P

1 Complex dependences

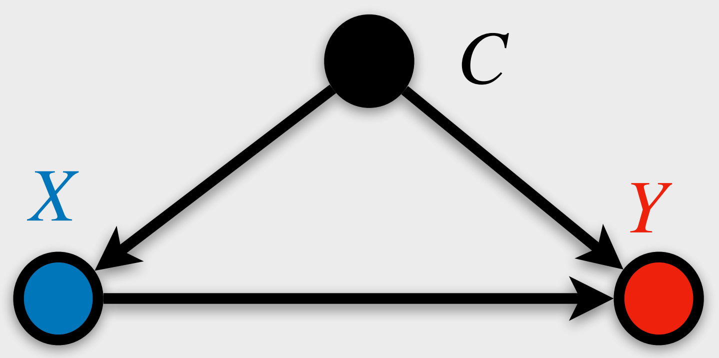
2 Data fusion (observations & experiments)

Challenges in Causal Inference

Effect (Q)

$\mathbb{E}[Y \mid do(x)]$

Graph

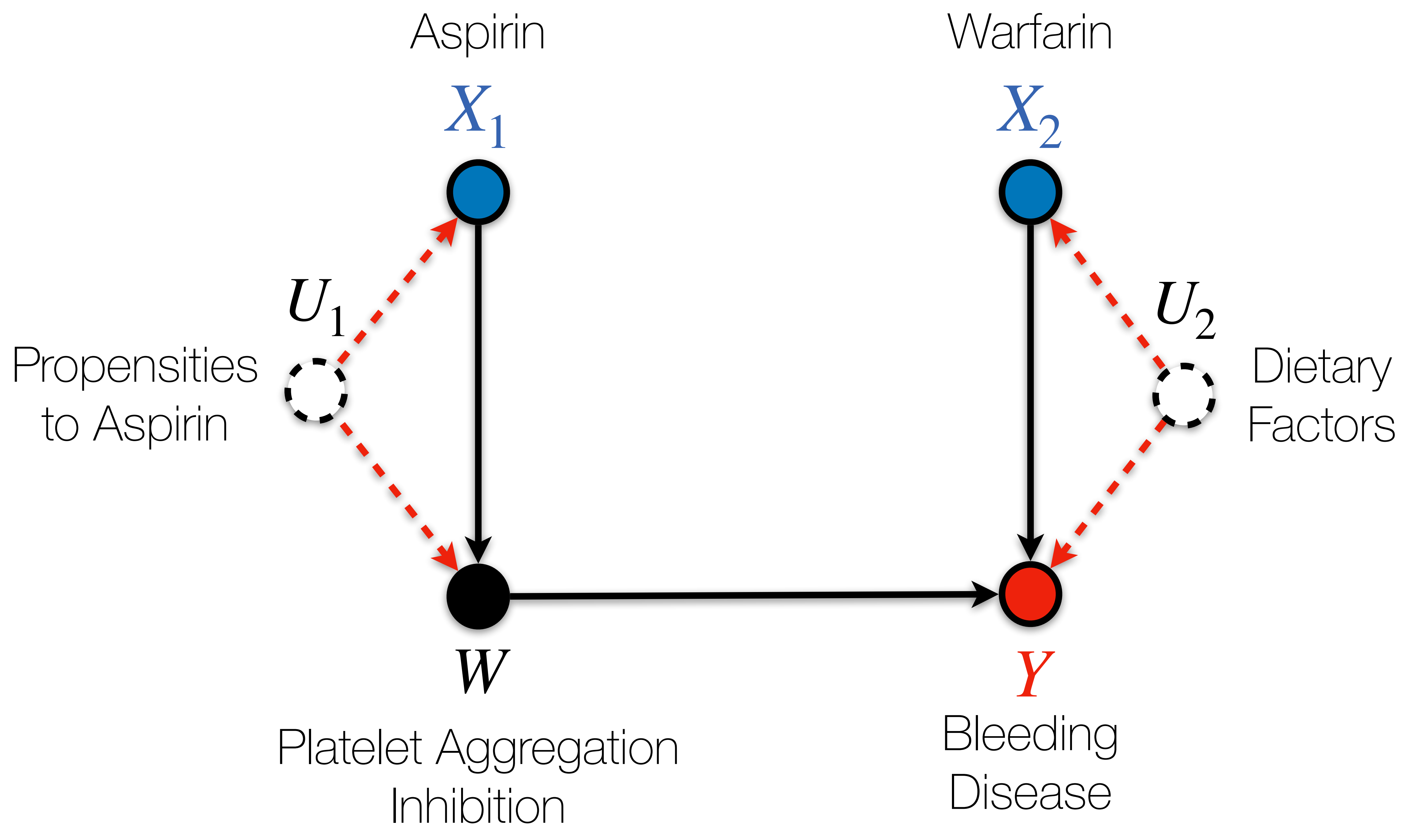


"Back-Door"

Samples

D from P

1. Complex dependences
2. Data fusion (observations & experiments)



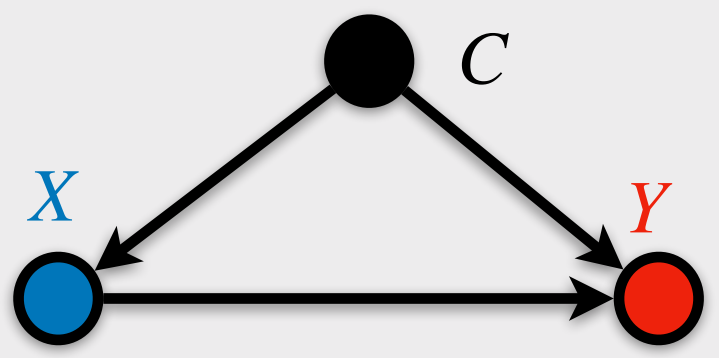
- Goal: Estimate $\mathbb{E}[Y \mid do(x_1, x_2)]$ from single interventions $do(x_1)$ and $do(x_2)$.

Challenges in Causal Inference

Effect (Q)

$\mathbb{E}[Y \mid do(x)]$

Graph

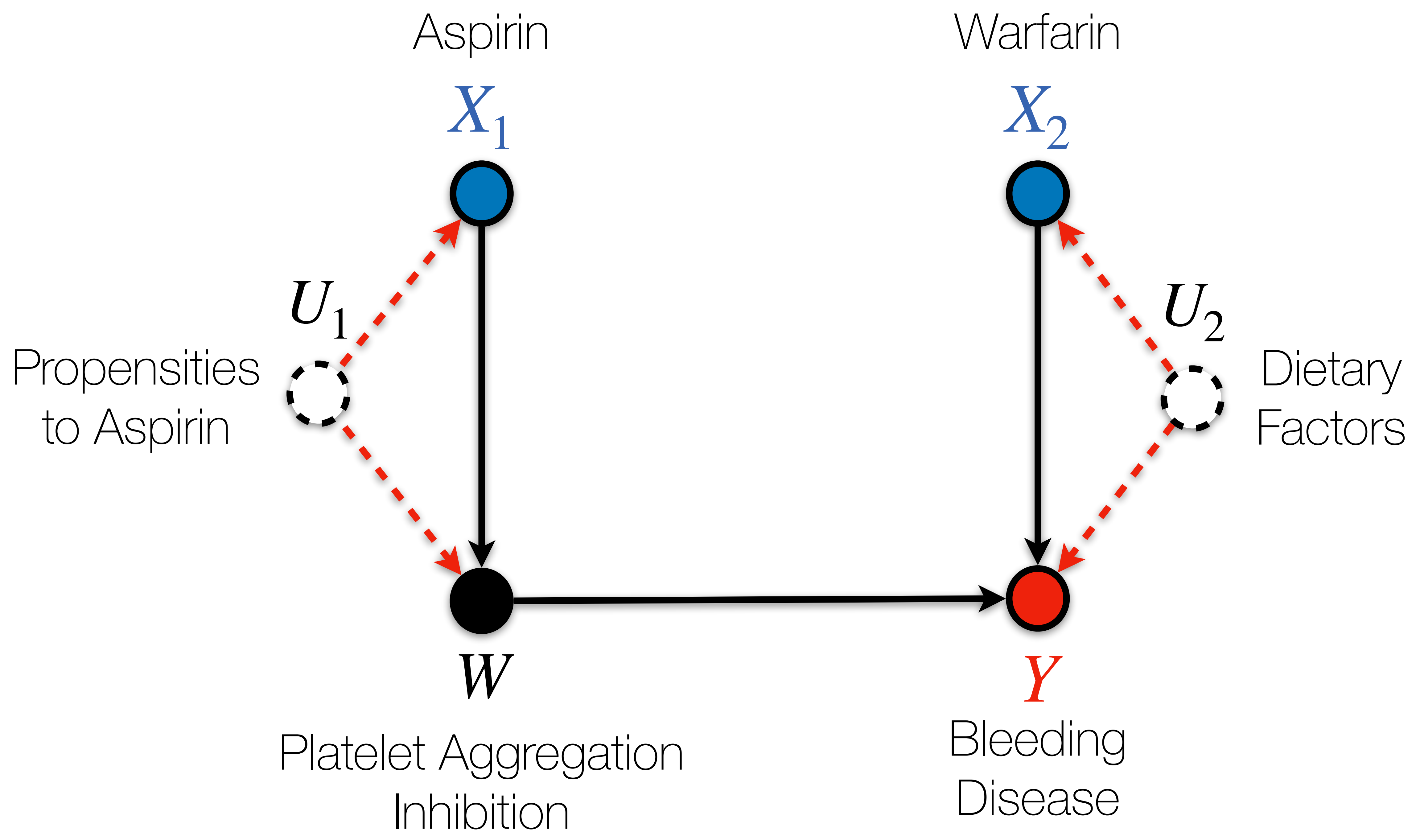


"Back-Door"

Samples

D from P

1. Complex dependences
2. Data fusion (observations & experiments)



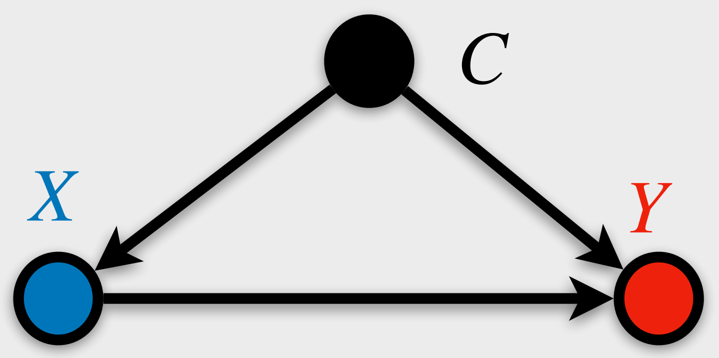
- Goal: Estimate $\mathbb{E}[Y \mid do(x_1, x_2)]$ from single interventions $do(x_1)$ and $do(x_2)$.
- Drug interactions between X_1 and X_2

Challenges in Causal Inference

Effect (Q)

$\mathbb{E}[Y \mid do(x)]$

Graph



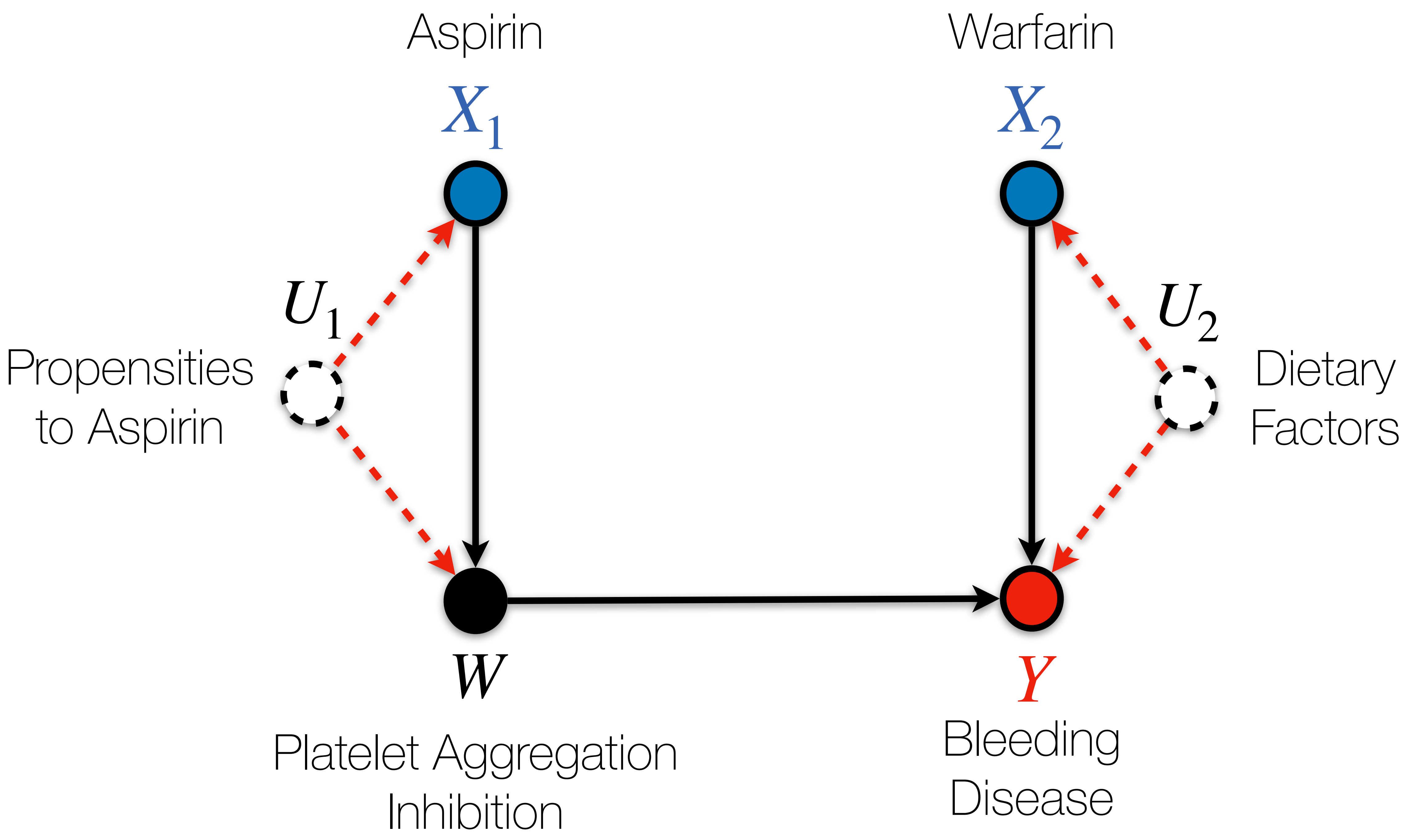
"Back-Door"

Samples

D from P

1. Complex dependences

2. Data fusion (observations & experiments)



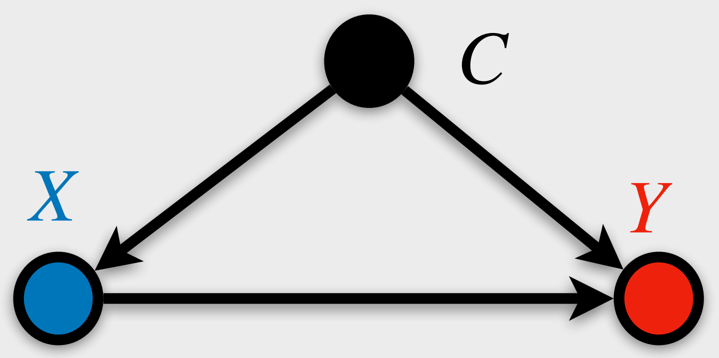
- Goal: Estimate $\mathbb{E}[Y \mid do(x_1, x_2)]$ from single interventions $do(x_1)$ and $do(x_2)$.
- Drug interactions between X_1 and X_2
- Not identifiable from observations

Challenges in Causal Inference

Effect (Q)

$\mathbb{E}[Y \mid \mathrm{do}(x)]$

Graph



"Back-Door"

Samples

D from P

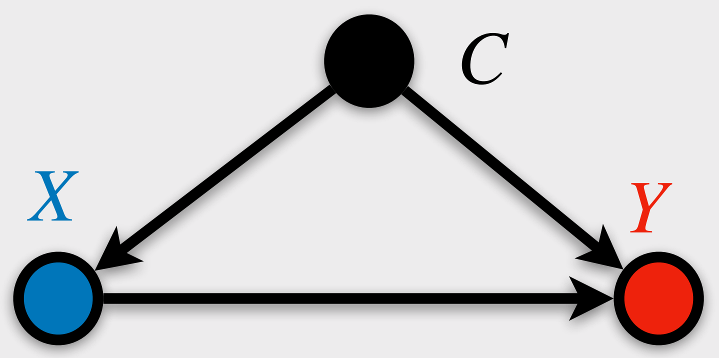
1. Complex dependences
2. Data fusion (observations & experiments)
3. More general scenarios

Challenges in Causal Inference

Effect (Q)

$\mathbb{E}[Y \mid \mathrm{do}(x)]$

Graph



"Back-Door"

Samples

D from P

1 Complex dependences

2 Data fusion (observations & experiments)

3 More general scenarios

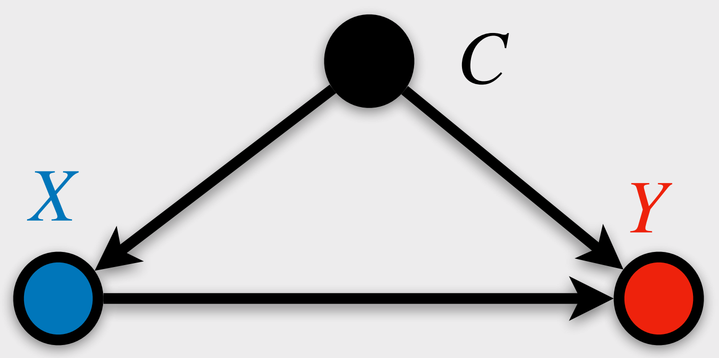
(Fairness) Salary (Y) a man ($X = x$) would earn if he is given the opportunities (M) that other genders ($X \neq x$) had received

Challenges in Causal Inference

Effect (Q)

$\mathbb{E}[Y \mid \mathrm{do}(x)]$

Graph



"Back-Door"

Samples

D from P

1. Complex dependences
2. Data fusion (observations & experiments)
3. More general scenarios

(Fairness) Salary (Y) a man ($X = x$) would earn if he is given the opportunities (M) that other genders ($X \neq x$) had received

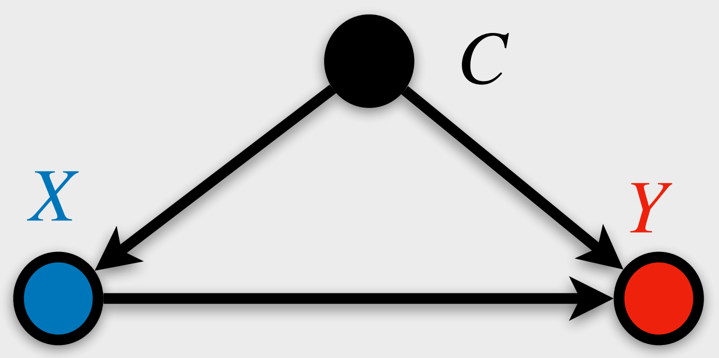
$$\mathbb{E}[\, Y_{x, M_{\neg x}} \,]$$

Challenges in Causal Inference

Effect (Q)

$\mathbb{E}[Y \mid \mathrm{do}(x)]$

Graph



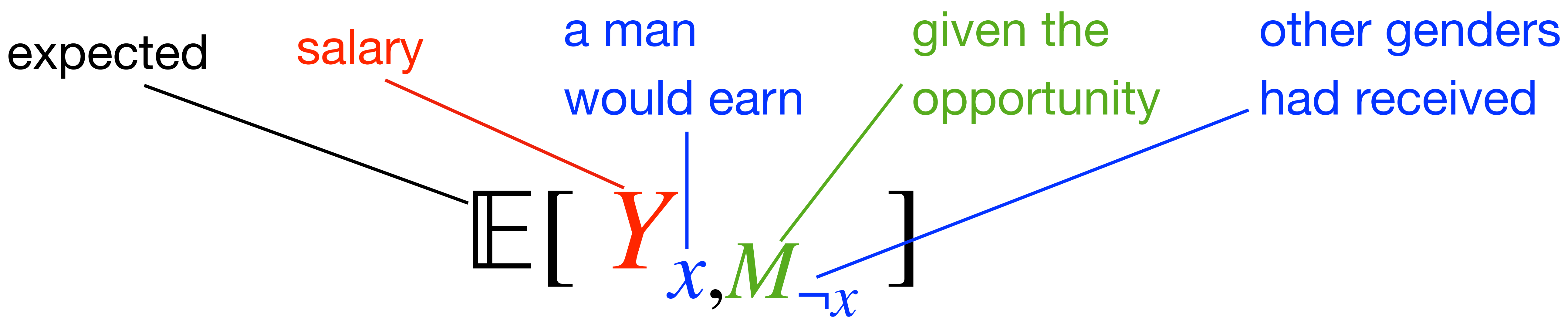
"Back-Door"

Samples

D from P

1. Complex dependences
2. Data fusion (observations & experiments)
3. More general scenarios

(Fairness) Salary (Y) a man ($X = x$) would earn if he is given the opportunities (M) that other genders ($X \neq x$) had received



Estimating Identifiable Causal Effects

Tasks

Challenges

1 Complicated dependences

2 Data fusion
(observations + experiments)

3 More general scenarios

Estimating Identifiable Causal Effects

Tasks

Challenges

1 Estimating causal effects from observations ✔

2 Data fusion (observations + experiments)

3 More general scenarios

Estimating Identifiable Causal Effects

Tasks		Challenges
1 Estimating causal effects from observations	→	✔
2 Estimating causal effects from data fusion	→	✔

3 More general scenarios

Estimating Identifiable Causal Effects

Tasks		Challenges
1 Estimating causal effects from observations	→	✔
2 Estimating causal effects from data fusion	→	✔
3 Unified causal effect estimation method	→	✔

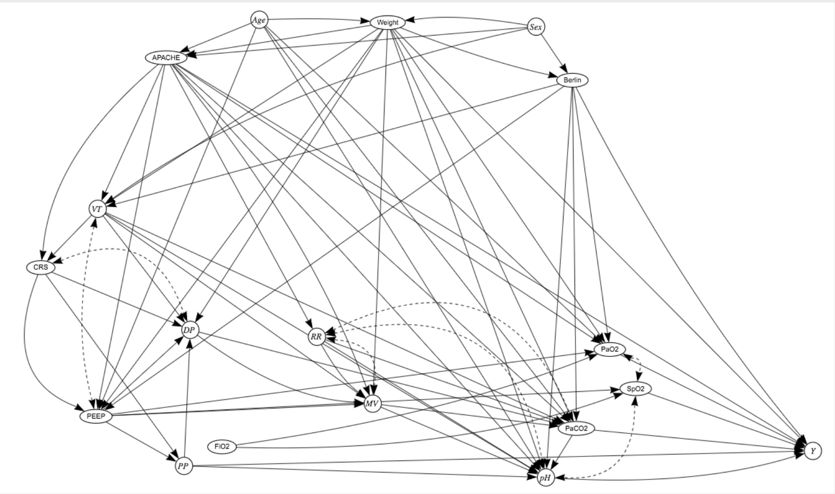
Task 1: Estimating from Observations

Input

Effect (Q)

$\mathbb{E}[Y \mid \mathrm{do}(x)]$

Graph



Samples

$D \sim P$

Identification

"When is the causal effect computable from available data?"

$Q = f(P)$

Estimation

"How do we compute the effect from available data?"

$\hat{Q} = \hat{f}(D)$

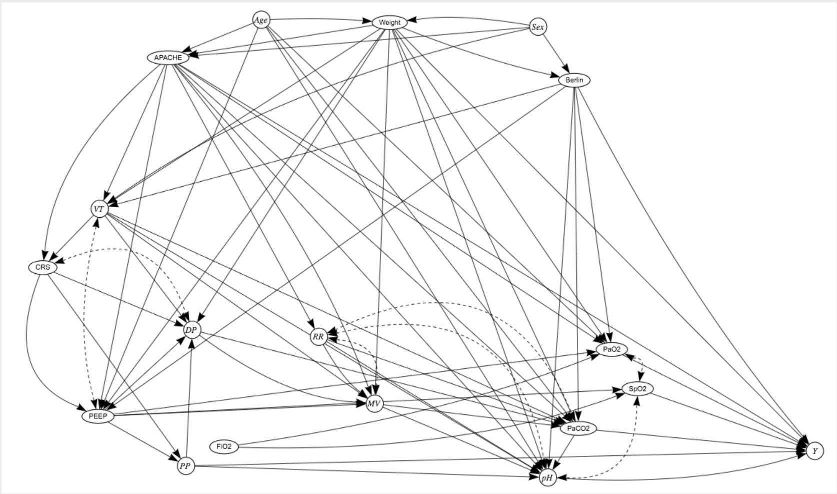
Task 1: Estimating from Observations

Input

Effect (Q)

$\mathbb{E}[Y \mid \mathrm{do}(x)]$

Graph



Samples

$D \sim P$

Identification

"When is the causal effect computable from available data?"

$Q = f(P)$

✔

Estimation

"How do we compute the effect from available data?"

$\hat{Q} = \hat{f}(D)$

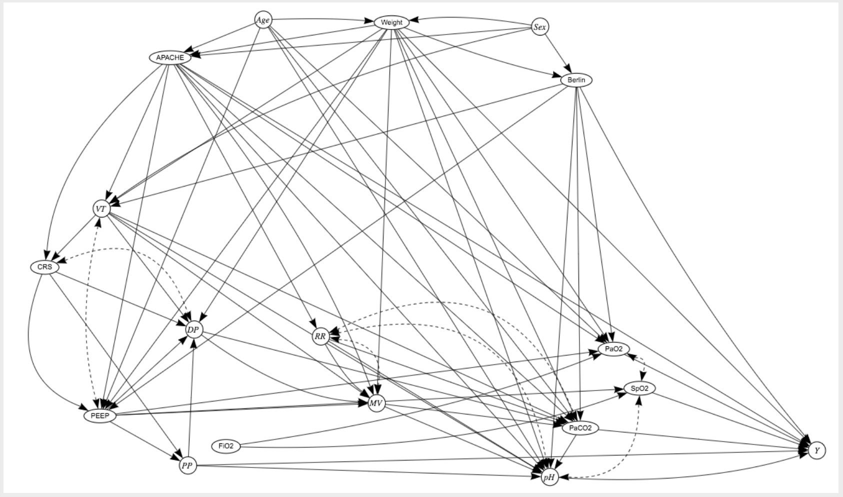
Task 1: Estimating from Observations

Input

Effect (Q)

$\mathbb{E}[Y \mid \mathrm{do}(x)]$

Graph



Samples

$D \sim P$

Identification

"When is the causal effect computable from available data?"

$Q = f(P)$

✔

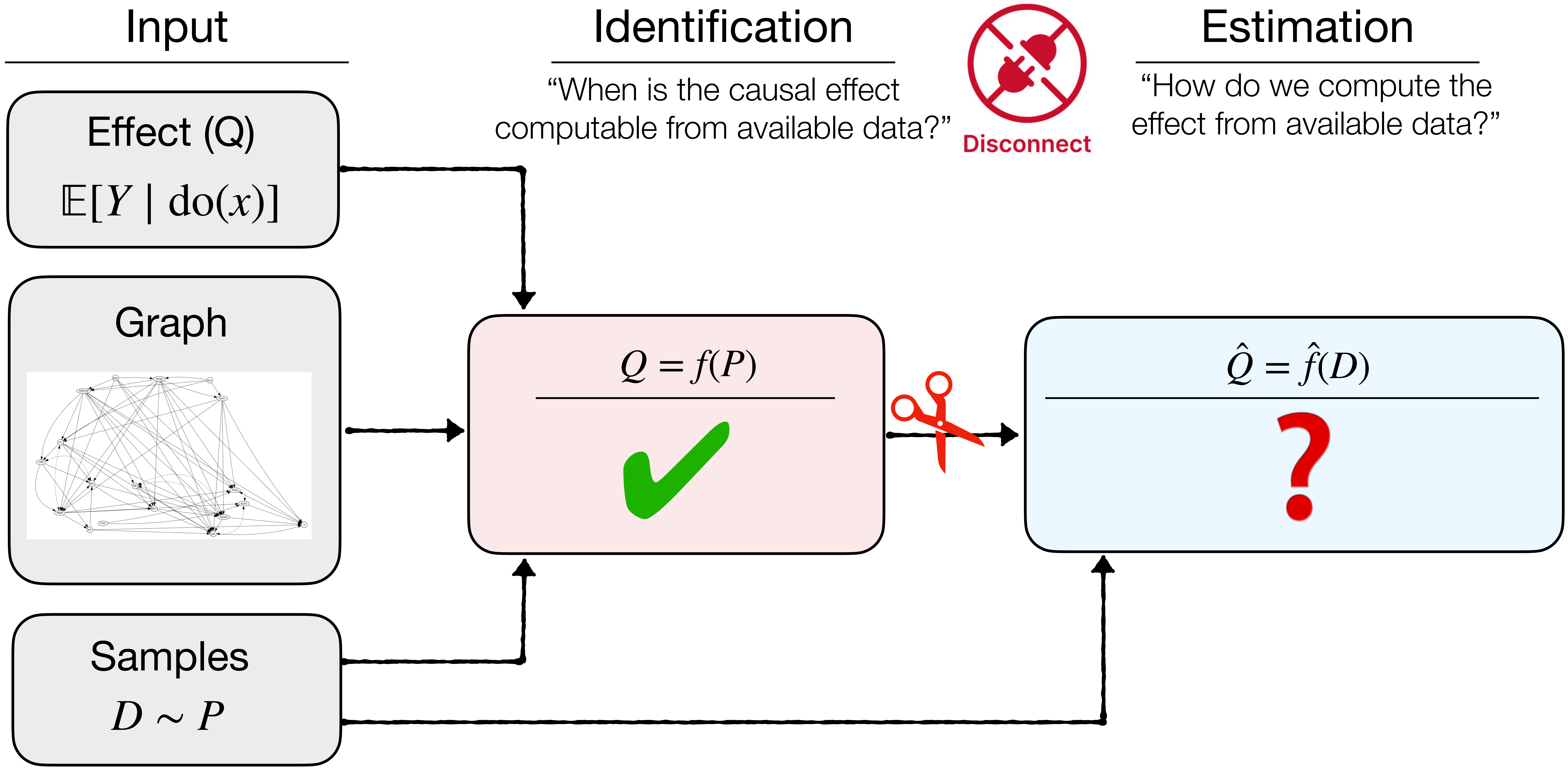
Estimation

"How do we compute the effect from available data?"

$\hat{Q} = \hat{f}(D)$

?

Task 1: Estimating from Observations



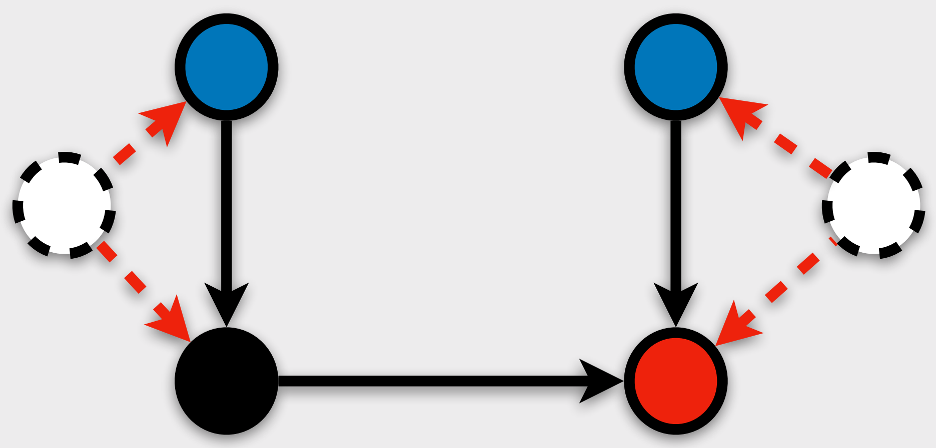
Task 2: Estimating from Data Fusion

Input

Effect (Q)

$\mathbb{E}[Y \mid \mathrm{do}(x)]$

Graph



Data Fusion

$\{D_Z\} \sim \{P_{\mathrm{do}(Z)}\}$

Identification

"When is the causal effect computable from available data?"

$Q = f(\{P_{\mathrm{do}(Z)}\})$

Estimation

"How do we compute the effect from available data?"

$\hat{Q} = \hat{f}(\{D_Z\})$

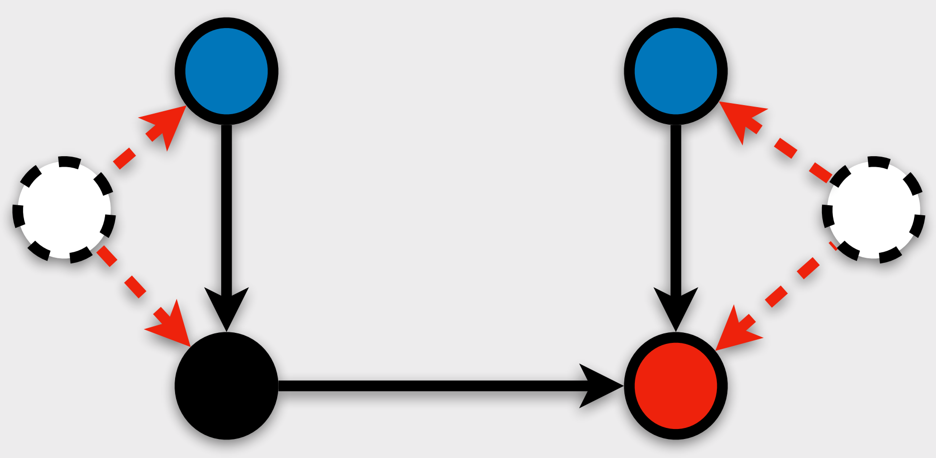
Task 2: Estimating from Data Fusion

Input

Effect (Q)

$\mathbb{E}[Y \mid \mathrm{do}(x)]$

Graph



Data Fusion

$\{D_Z\} \sim \{P_{\mathrm{do}(Z)}\}$

Identification

"When is the causal effect computable from available data?"

$Q = f(\{P_{\mathrm{do}(Z)}\})$

✔

Estimation

"How do we compute the effect from available data?"

$\hat{Q} = \hat{f}(\{D_Z\})$

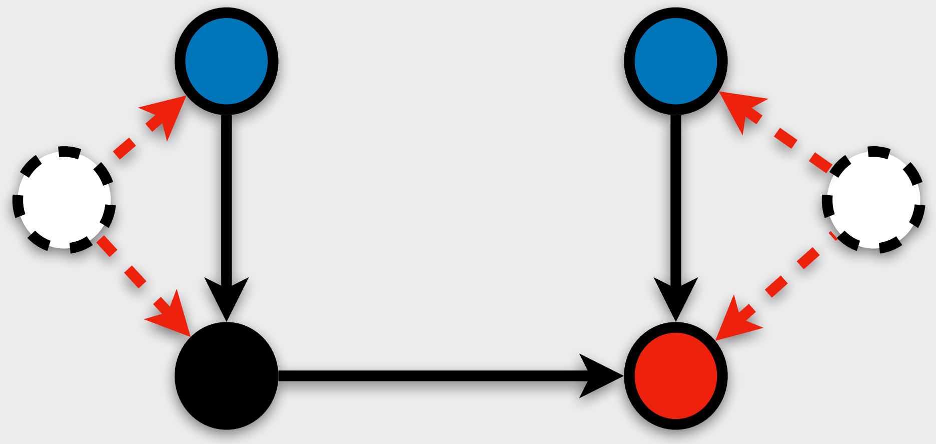
Task 2: Estimating from Data Fusion

Input

Effect (Q)

$\mathbb{E}[Y \mid \mathrm{do}(x)]$

Graph



Data Fusion

$\{D_Z\} \sim \{P_{\mathrm{do}(Z)}\}$

Identification

"When is the causal effect computable from available data?"

$Q = f(\{P_{\mathrm{do}(Z)}\})$

✔

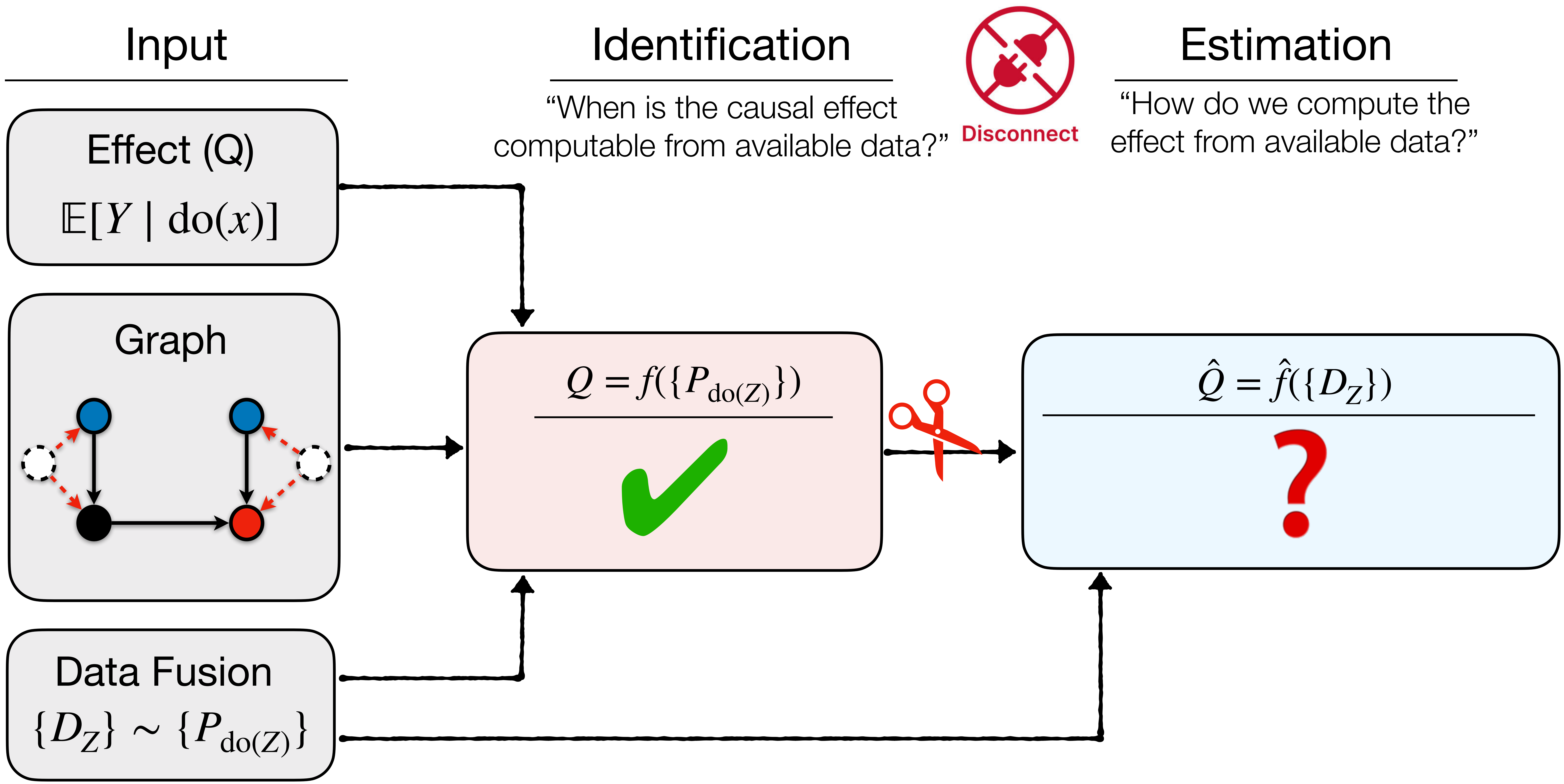
Estimation

"How do we compute the effect from available data?"

$\hat{Q} = \hat{f}(\{D_Z\})$

?

Task 2: Estimating from Data Fusion



Task 3: Unified Estimation Methods

Fairness Analysis

$$\mathbb{E}[Y_{x,M_{\neg x}}]$$

Salary a man would earn if given the opportunities that other genders would receive

Offline Policy Evaluation

$$\mathbb{E}[Y_{\tau(X|C)}]$$

Recovery rate of a drug dosage policy given baseline characteristics

. . .

Retrospection

$$\mathbb{E}[Y_x | \neg x]$$

The headache intensity for patients who took aspirin, had they not taken it

Domain Transfer

$$\mathbb{E}[Y \mid \mathrm{do}(x), S{=}\mathrm{NY}]$$

The effect of a treatment in NY identifiable from trials in Chicago

. . .

Task 3: Unified Estimation Methods

Fairness Analysis

Salary a man would earn if given the ... other genders would receive

Estimator 1

Offline Policy Evaluation

Recove... of a drug dosage policy given baseline characteristics

Estimator 2

Retrospection

The headache intensity for patient... had they not taken it

Estimator 3

Domain Transfer

The effect of a treatment in NY identifi... in Chicago

Estimator 4

Task 3: Unified Estimation Methods

Unified causal effect estimation method

Talk Outline

Talk Outline

1. Estimating causal effects from observations

   + its application in healthcare & explainable AI

Talk Outline

1. Estimating causal effects from observations

   + its application in healthcare & explainable AI

2. Estimating causal effects from data fusion

Talk Outline

1. Estimating causal effects from observations

   + its application in healthcare & explainable AI

2. Estimating causal effects from data fusion

3. Unified causal effect estimation method

Talk Outline

1. Estimating causal effects from observations

   + its application in healthcare & explainable AI

2. Estimating causal effects from data fusion

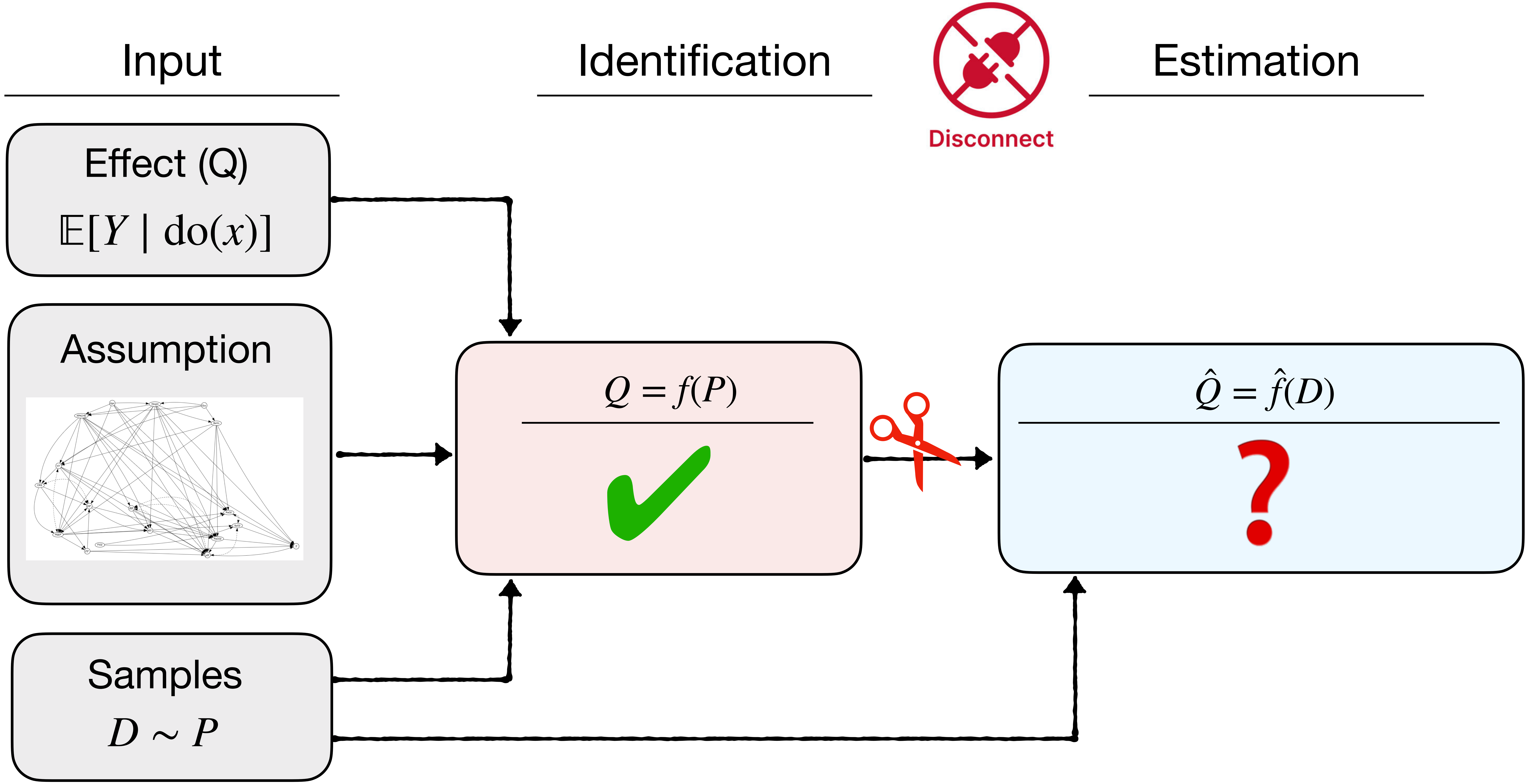
3. Unified causal effect estimation method

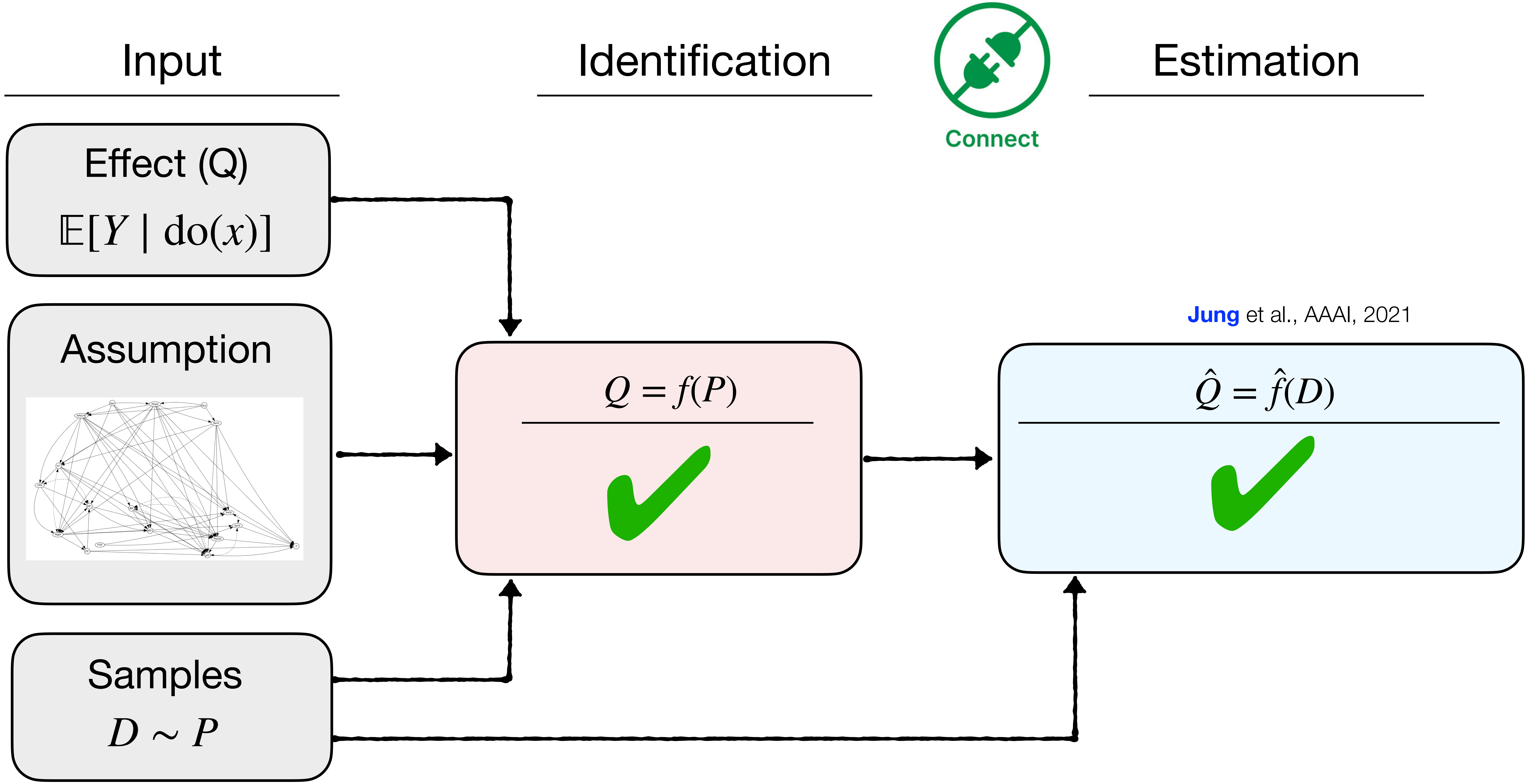
4. Summary & Future direction

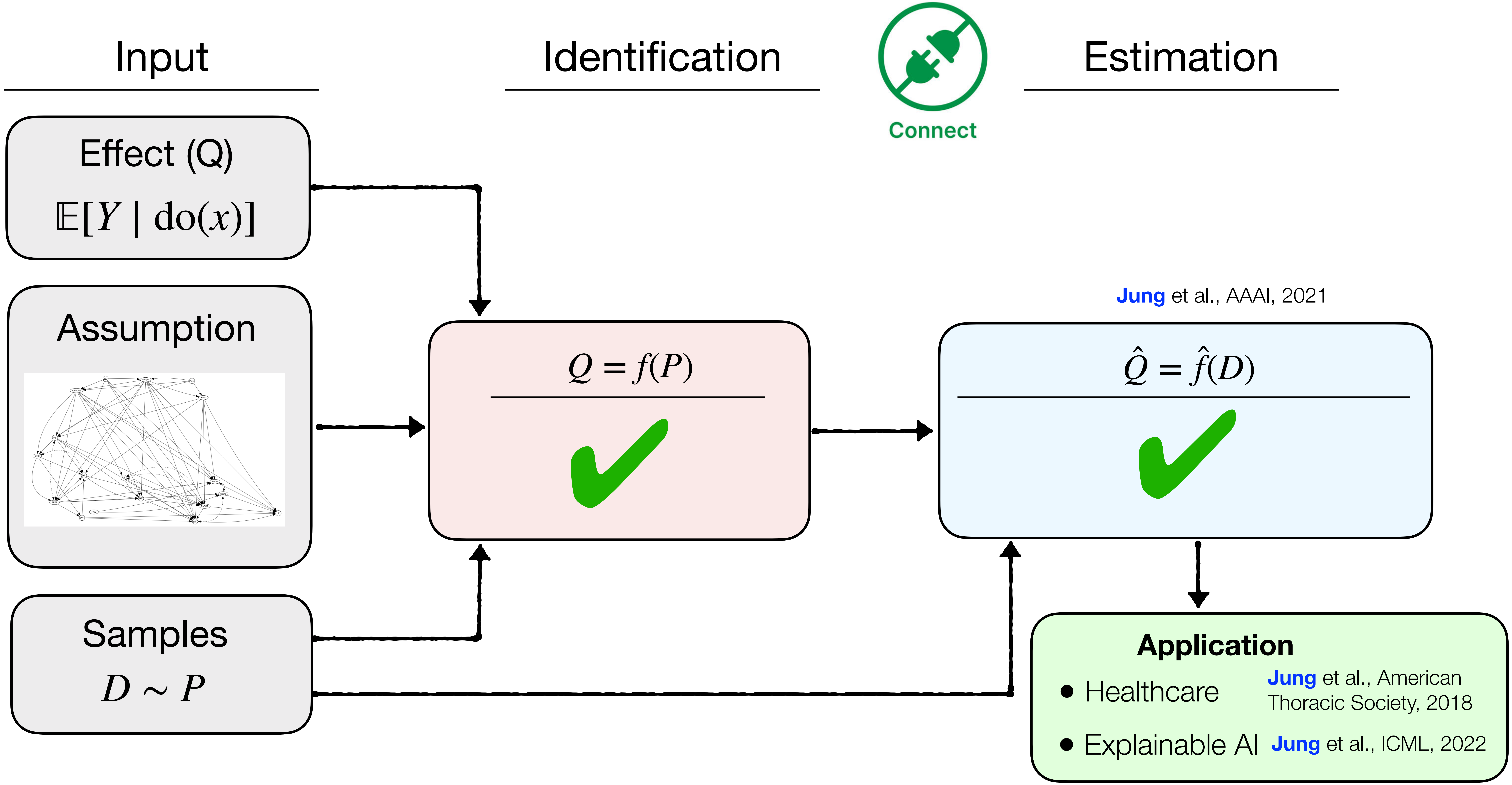
Talk Outline

1 Estimating causal effects from observations

+ its application in healthcare & explainable AI

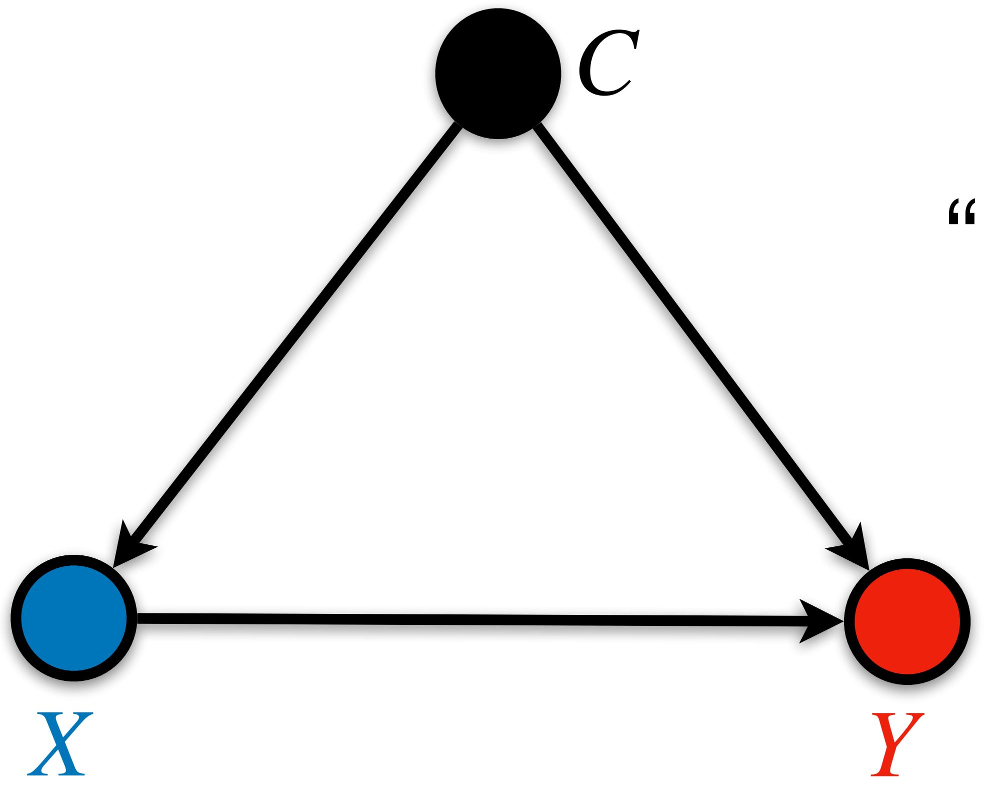

Input
Identification
Disconnect
Estimation
Effect (Q)
$\mathbb{E}[Y \mid \mathrm{do}(x)]$
Assumption
$Q = f(P)$
$\hat{Q} = \hat{f}(D)$
?
Samples
$D \sim P$


Input
Identification
Connect
Estimation
Effect (Q)
$\mathbb{E}[Y \mid \mathrm{do}(x)]$
Assumption
Samples
$D \sim P$
$Q = f(P)$
Jung et al., AAAI, 2021
$\hat{Q} = \hat{f}(D)$


Input
Identification
Connect
Estimation
Effect (Q)
$\mathbb{E}[Y \mid \mathrm{do}(x)]$
Assumption
Samples
$D \sim P$
$Q = f(P)$
Jung et al., AAAI, 2021
$\hat{Q} = \hat{f}(D)$
Application
Healthcare
Jung et al., American Thoracic Society, 2018
Explainable AI
Jung et al., ICML, 2022

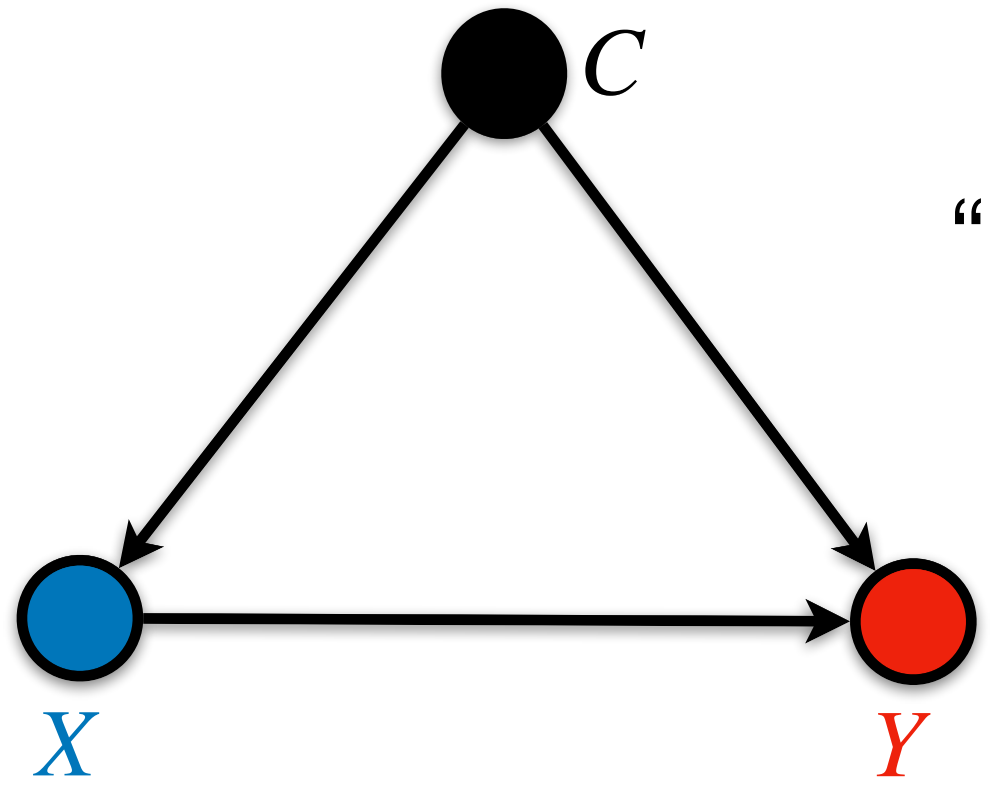
Background: Back-door Adjustment (BD)

Background: Back-door Adjustment (BD)



"Back-door graph"

Background: Back-door Adjustment (BD)



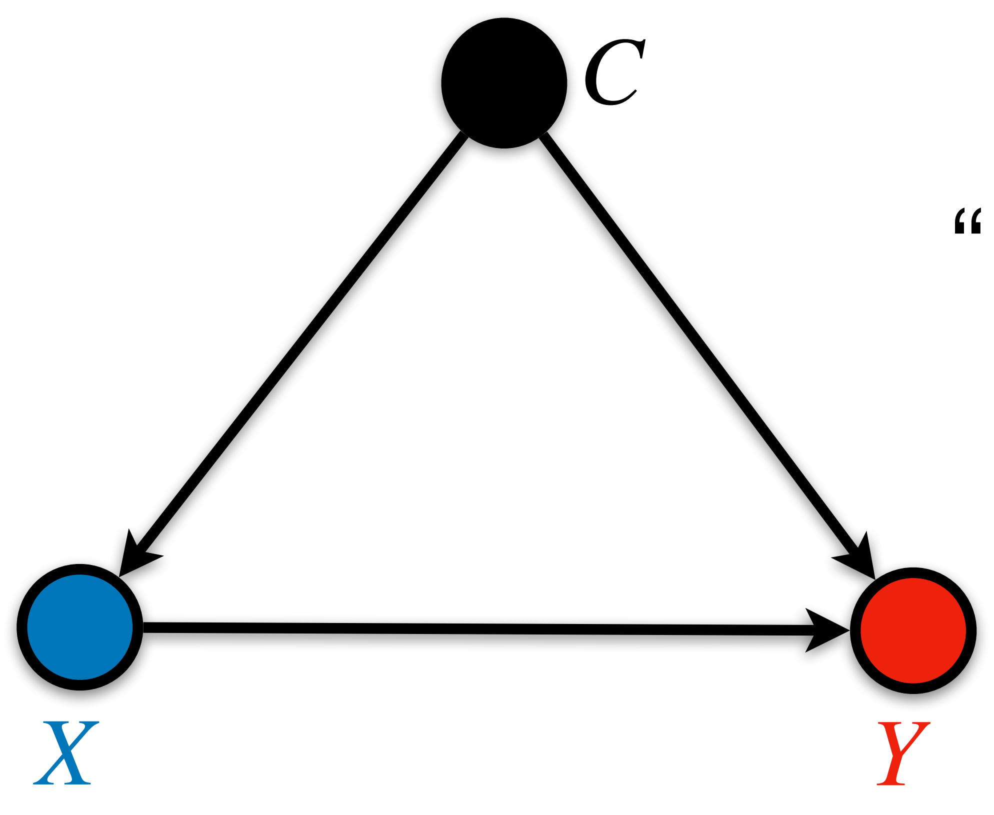
"Back-door graph"

Back-door Criterion

(Rubin 74,, Robins 86, Pearl, 95)

Spurious paths between (treatments, outcome) are blocked by observed variables (i.e., *no unmeasured confounders*)

Background: Back-door Adjustment (BD)



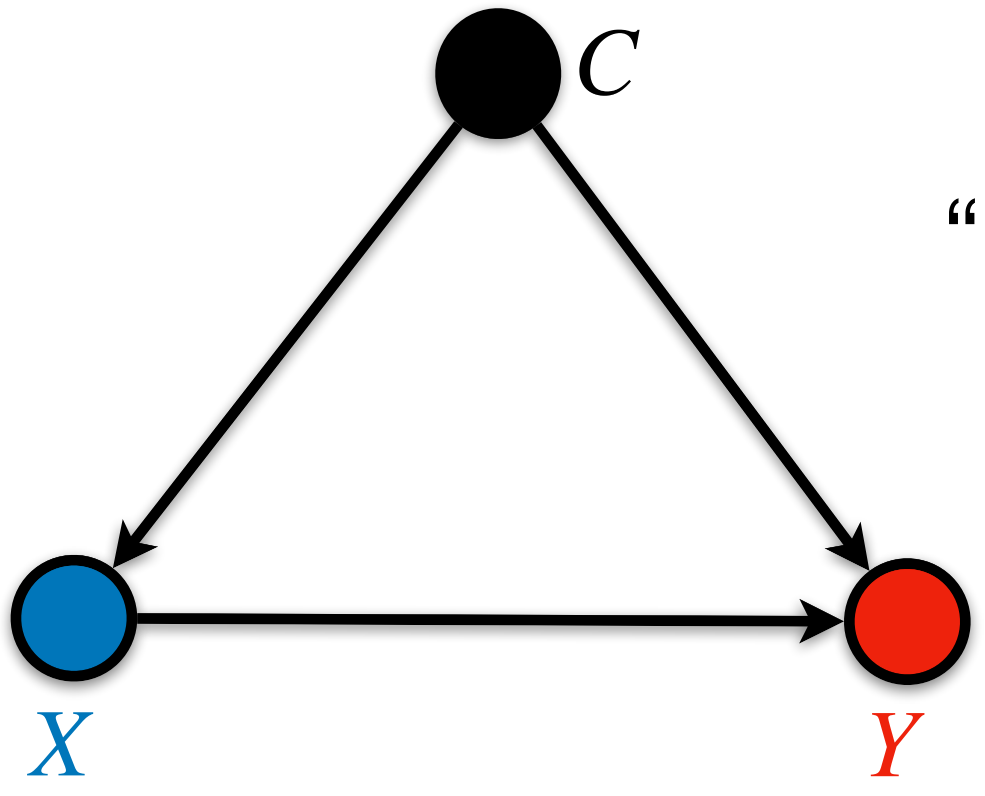
"Back-door graph"

Back-door Criterion

(Rubin 74,, Robins 86, Pearl, 95)

Spurious paths between (treatments, outcome) are blocked by observed variables (i.e., *no unmeasured confounders*)

Background: Back-door Adjustment (BD)



"Back-door graph"

Back-door Criterion

(Rubin 74,, Robins 86, Pearl, 95)

Spurious paths between (treatments, outcome) are blocked by observed variables (i.e., *no unmeasured confounders*)

"Back-door adjustment (BD)"

$$\mathbb{E}[Y \mid \mathrm{do}(x)] = \text{BD} \triangleq \sum_{c} \mathbb{E}[Y \mid x, c] P(c)$$

DML-BD: Robust Estimator for BD

DML-BD: Robust Estimator for BD

❶ $\text{BD}(\mu, \pi) = \mathbb{E}[\mu \times \pi]$, where $\mu(XC) \triangleq \mathbb{E}[Y \mid X, C]$ and $\pi(XC) \triangleq \dfrac{\mathbb{I}_x(X)}{P(X \mid C)}$

DML-BD: Robust Estimator for BD

"Double Machine Learning Estimator for Back-door Adjustment" (Chernozhukov et al., 2018)

2 DML-BD($\hat{\mu}$,$\hat{\pi}$) is a robust estimator:

DML-BD: Robust Estimator for BD

"Double Machine Learning Estimator for Back-door Adjustment" (Chernozhukov et al., 2018)

2 DML-BD($\hat{\mu}$,$\hat{\pi}$) is a robust estimator:

$$\text{Error}(\text{DML-BD}(\hat{\mu},\hat{\pi}),\ \text{BD}(\mu,\pi)) = \text{Error}(\hat{\mu},\mu) \times \text{Error}(\hat{\pi},\pi)$$

DML-BD: Robust Estimator for BD

2 DML-BD($\hat{\mu}$,$\hat{\pi}$) is a robust estimator:

$$\text{Error}(\text{DML-BD}(\hat{\mu},\hat{\pi}),\ \text{BD}(\mu,\pi)) = \text{Error}(\hat{\mu},\mu) \times \text{Error}(\hat{\pi},\pi)$$

- **Double Robustness:** Error $= 0$ if either $\hat{\mu} = \mu$ or $\hat{\pi} = \pi$

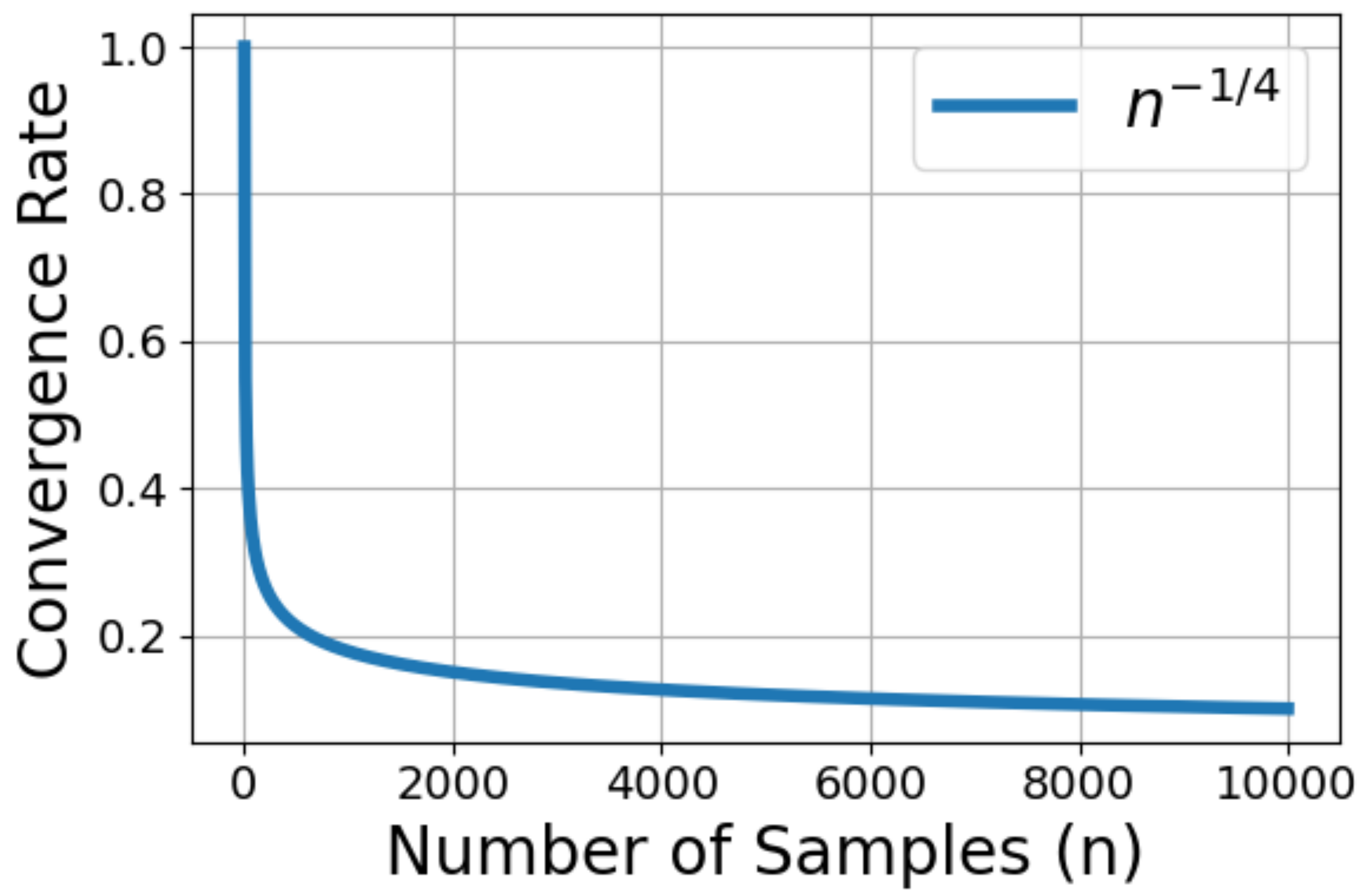
DML-BD: Robust Estimator for BD

❷ DML-BD($\hat{\mu}$,$\hat{\pi}$) is a robust estimator:

$$\text{Error}(\text{DML-BD}(\hat{\mu},\hat{\pi}),\ \text{BD}(\mu,\pi)) = \text{Error}(\hat{\mu},\mu) \times \text{Error}(\hat{\pi},\pi)$$

- **Double Robustness:** Error $= 0$ if either $\hat{\mu} = \mu$ or $\hat{\pi} = \pi$
- **Fast Convergence:** Error $\rightarrow 0$ *fast* even when $\hat{\mu} \rightarrow \mu$ and $\hat{\pi} \rightarrow \pi$ *slowly*.

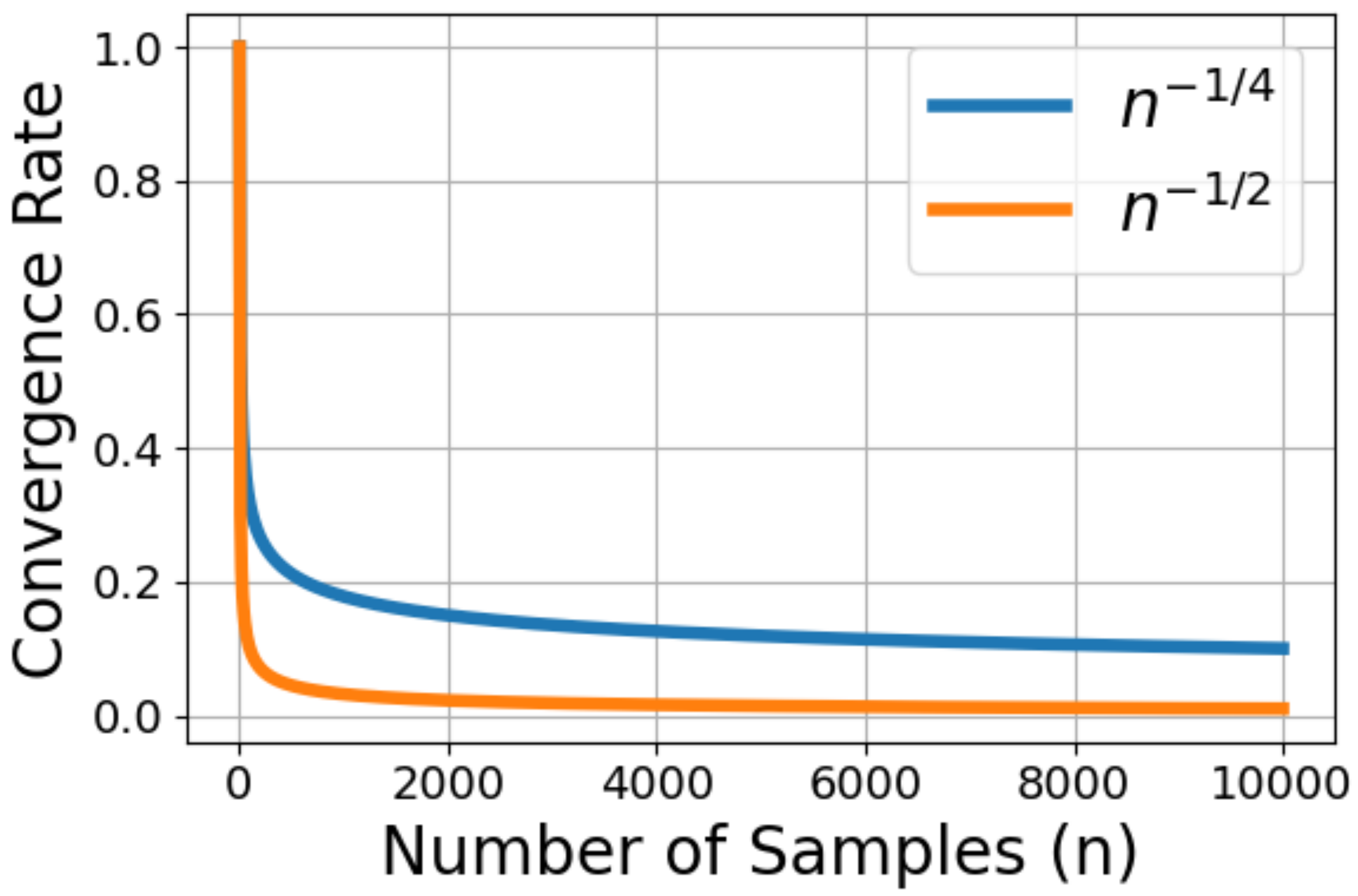
DML-BD: Robust Estimator for BD

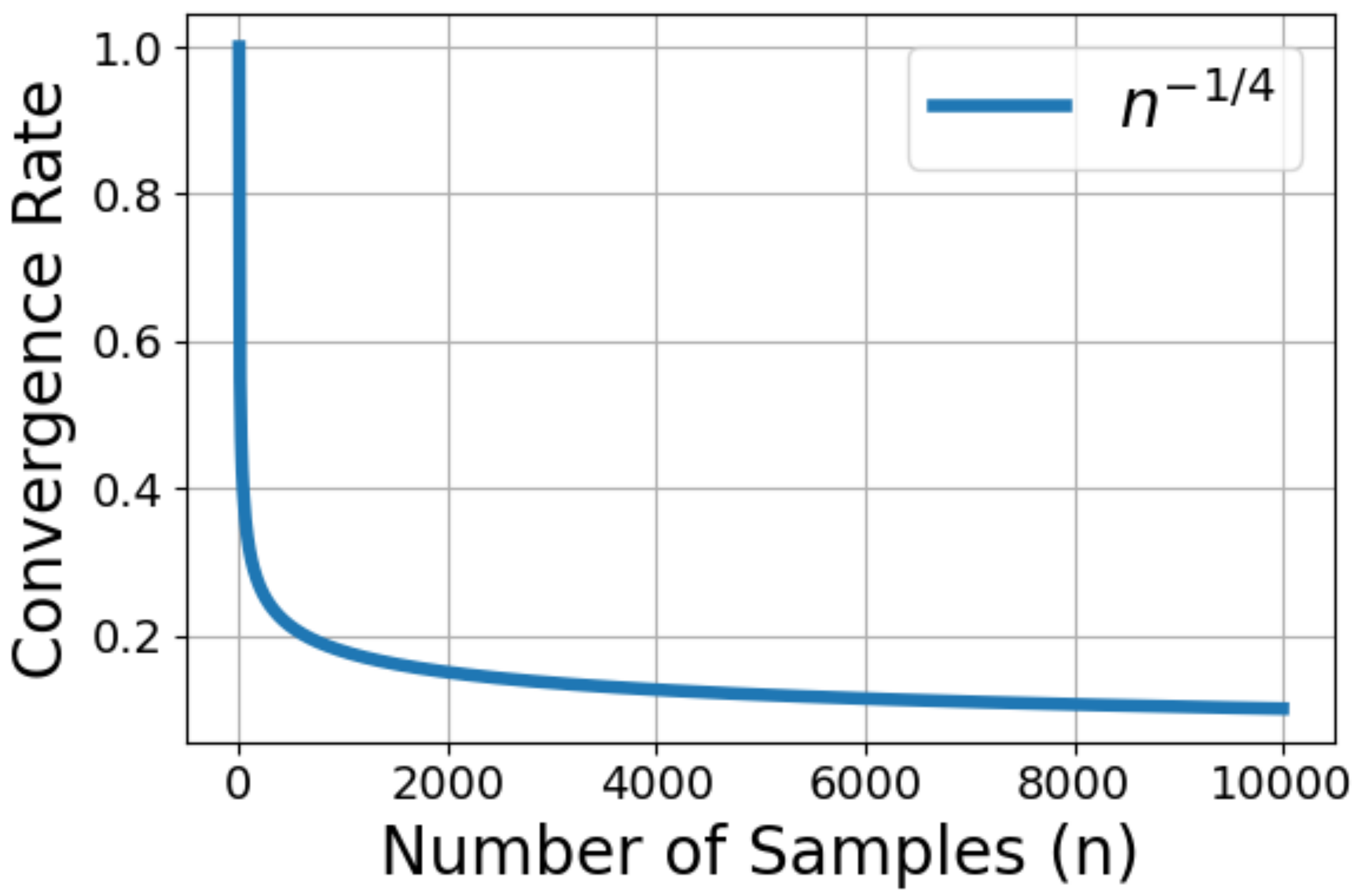


$$\text{Error}(\text{DML-BD}(\hat{\mu},\hat{\pi}),\ \text{BD}(\mu,\pi)) = \underbrace{\text{Error}(\hat{\mu},\mu)}_{n^{-1/4}} \times \underbrace{\text{Error}(\hat{\pi},\pi)}_{n^{-1/4}}$$

- **Fast Convergence:** Error $\to 0$ *fast* even when $\hat{\mu} \to \mu$ and $\hat{\pi} \to \pi$ *slowly*.

DML-BD: Robust Estimator for BD



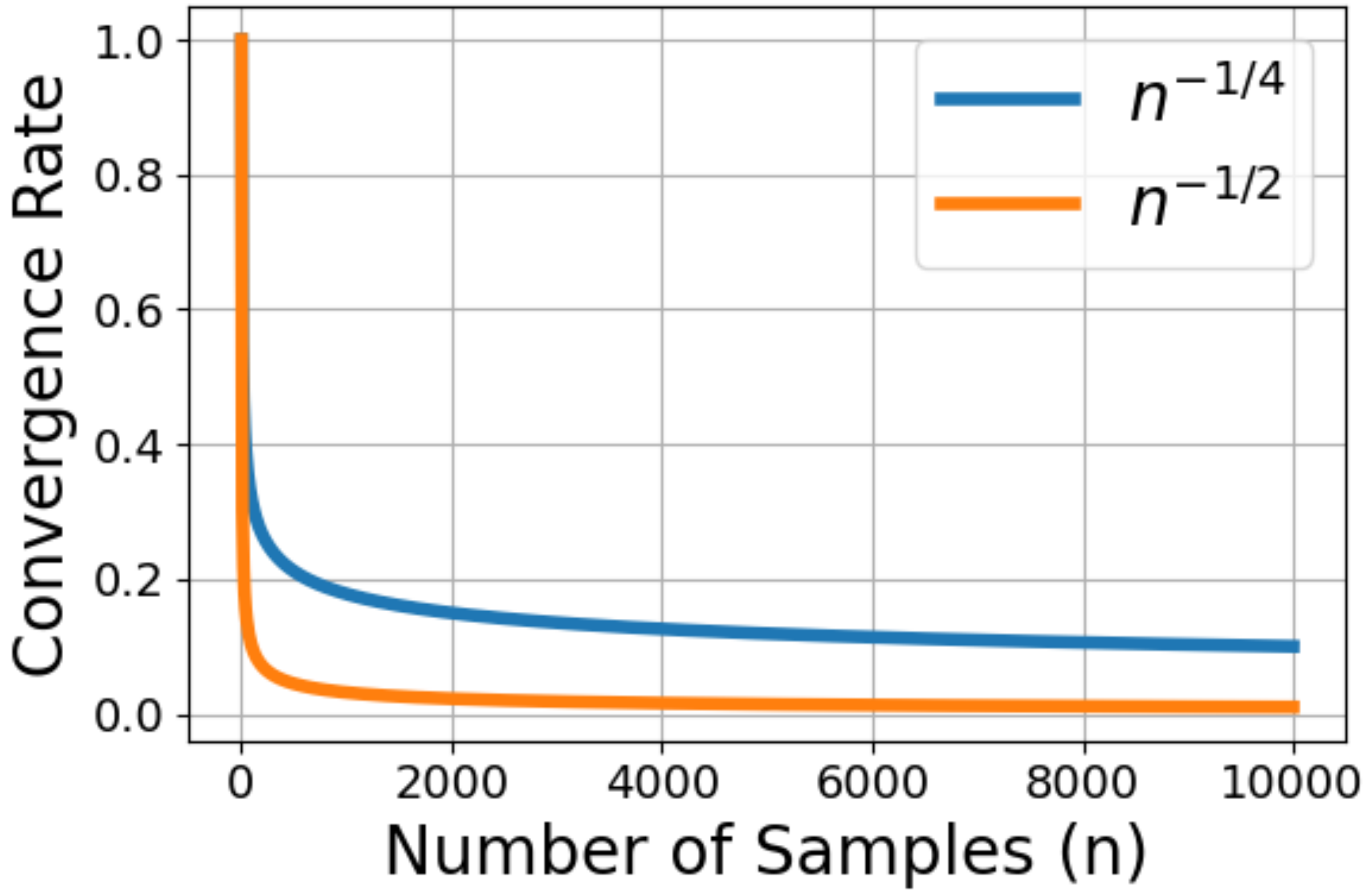


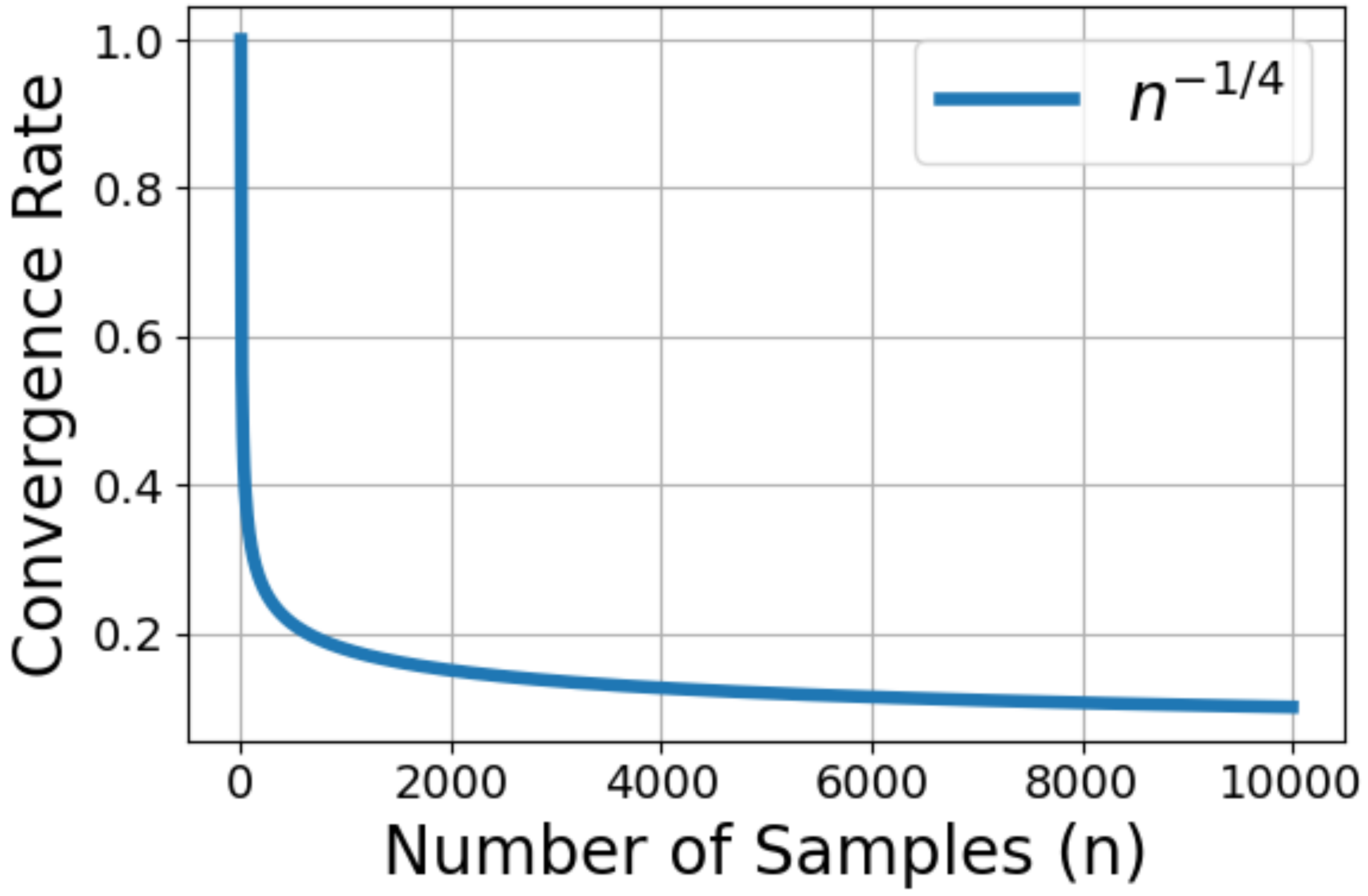
$$\text{Error}(\text{DML-BD}(\hat{\mu},\hat{\pi}),\ \text{BD}(\mu,\pi)) = \text{Error}(\hat{\mu},\mu) \times \text{Error}(\hat{\pi},\pi)$$

$n^{-1/2}$ $\quad$ $n^{-1/4}$ $\quad$ $n^{-1/4}$

- **Fast Convergence:** Error $\rightarrow 0$ *fast* even when $\hat{\mu} \rightarrow \mu$ and $\hat{\pi} \rightarrow \pi$ *slowly*.

DML-BD: Robust Estimator for BD





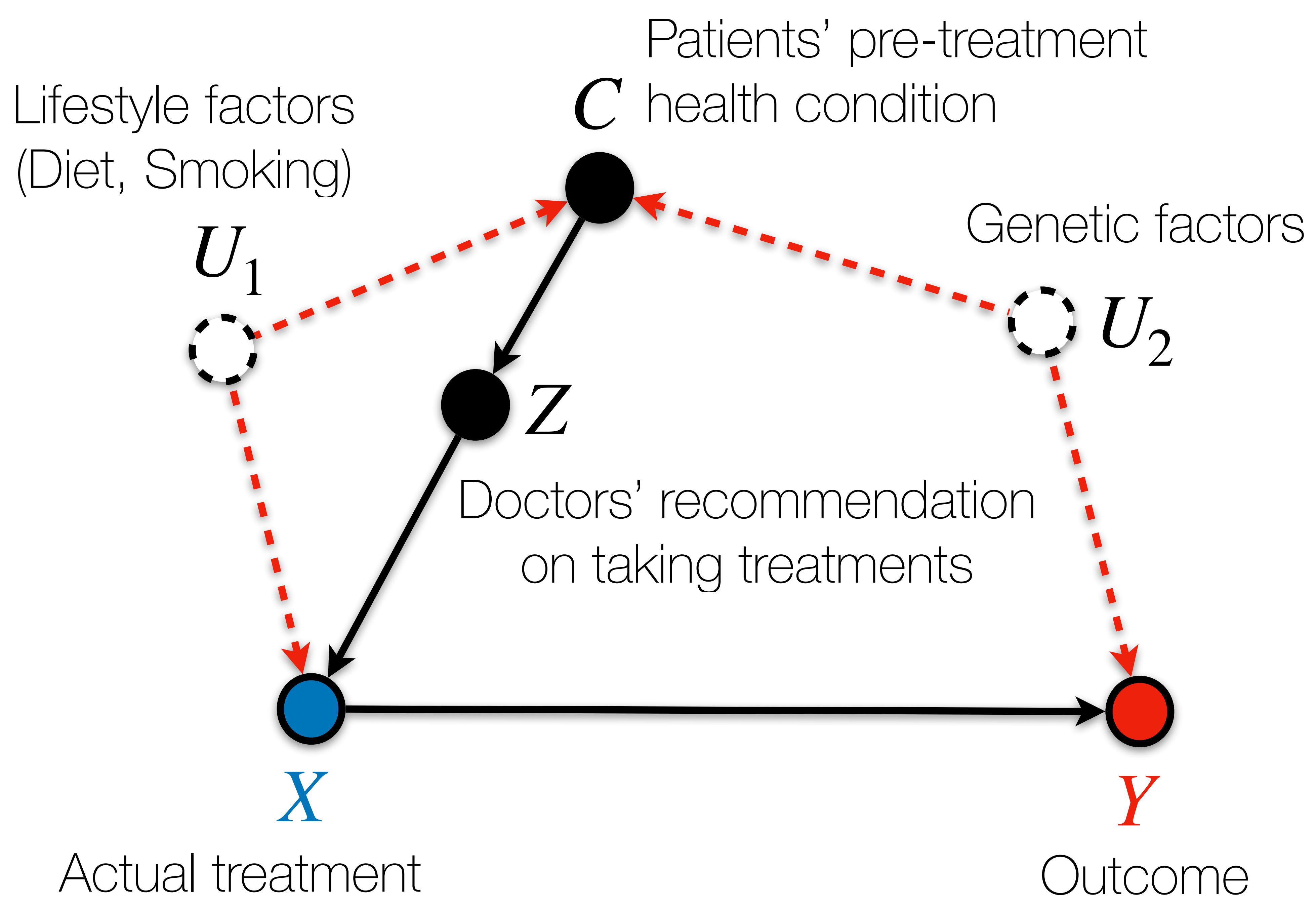
$$\text{Error}(\text{DML-BD}(\hat{\mu},\hat{\pi}), \text{BD}(\mu,\pi)) = \text{Error}(\hat{\mu},\mu) \times \text{Error}(\hat{\pi},\pi)$$

$n^{-1/2}$ $n^{-1/4}$ $n^{-1/4}$

- **Fast Convergence:** Error $\rightarrow 0$ *fast* even when $\hat{\mu} \rightarrow \mu$ and $\hat{\pi} \rightarrow \pi$ *slowly*.

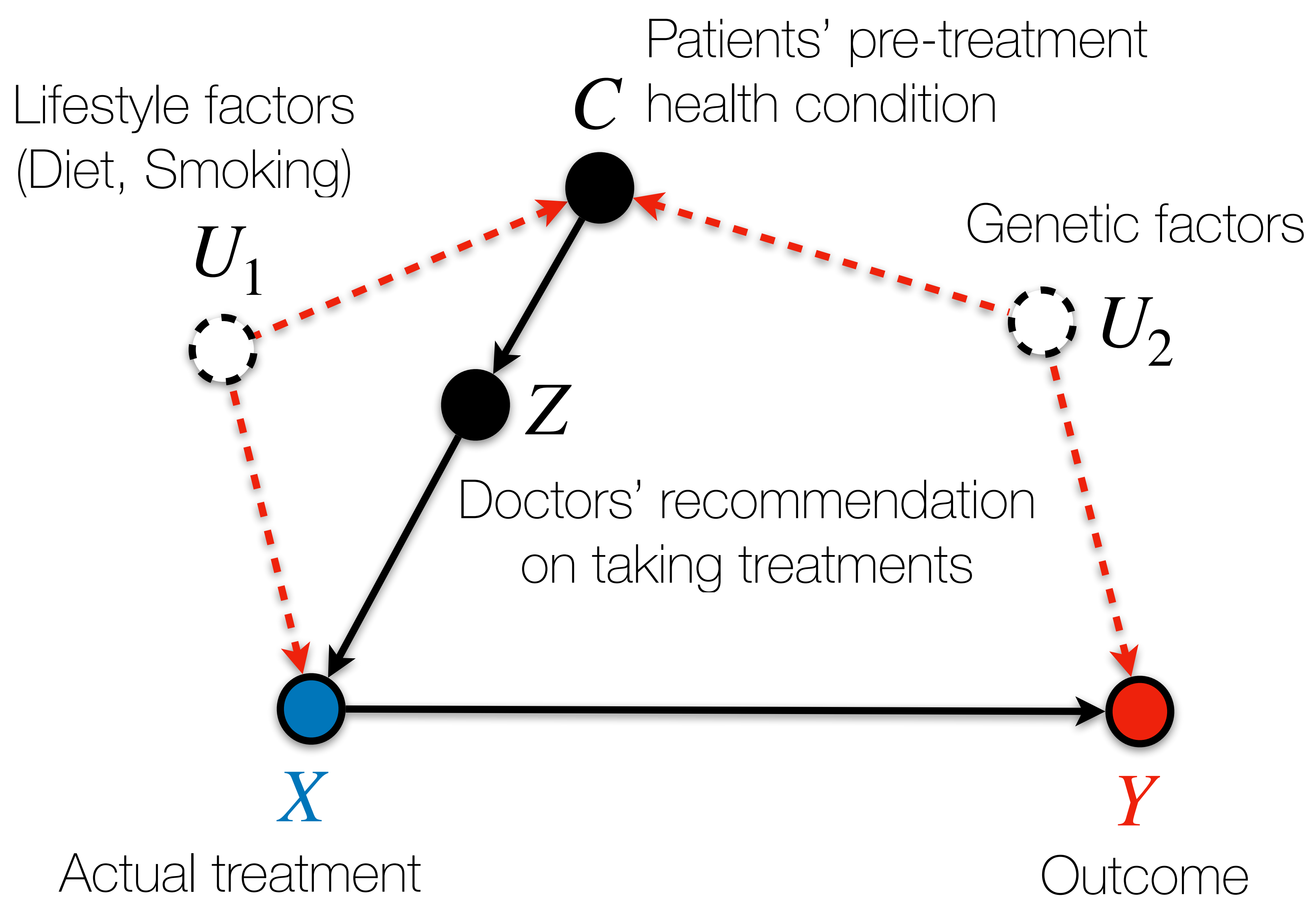
Property of modern ML models

Working example



$$\mathbb{E}[Y \mid \mathrm{do}(x)]$$

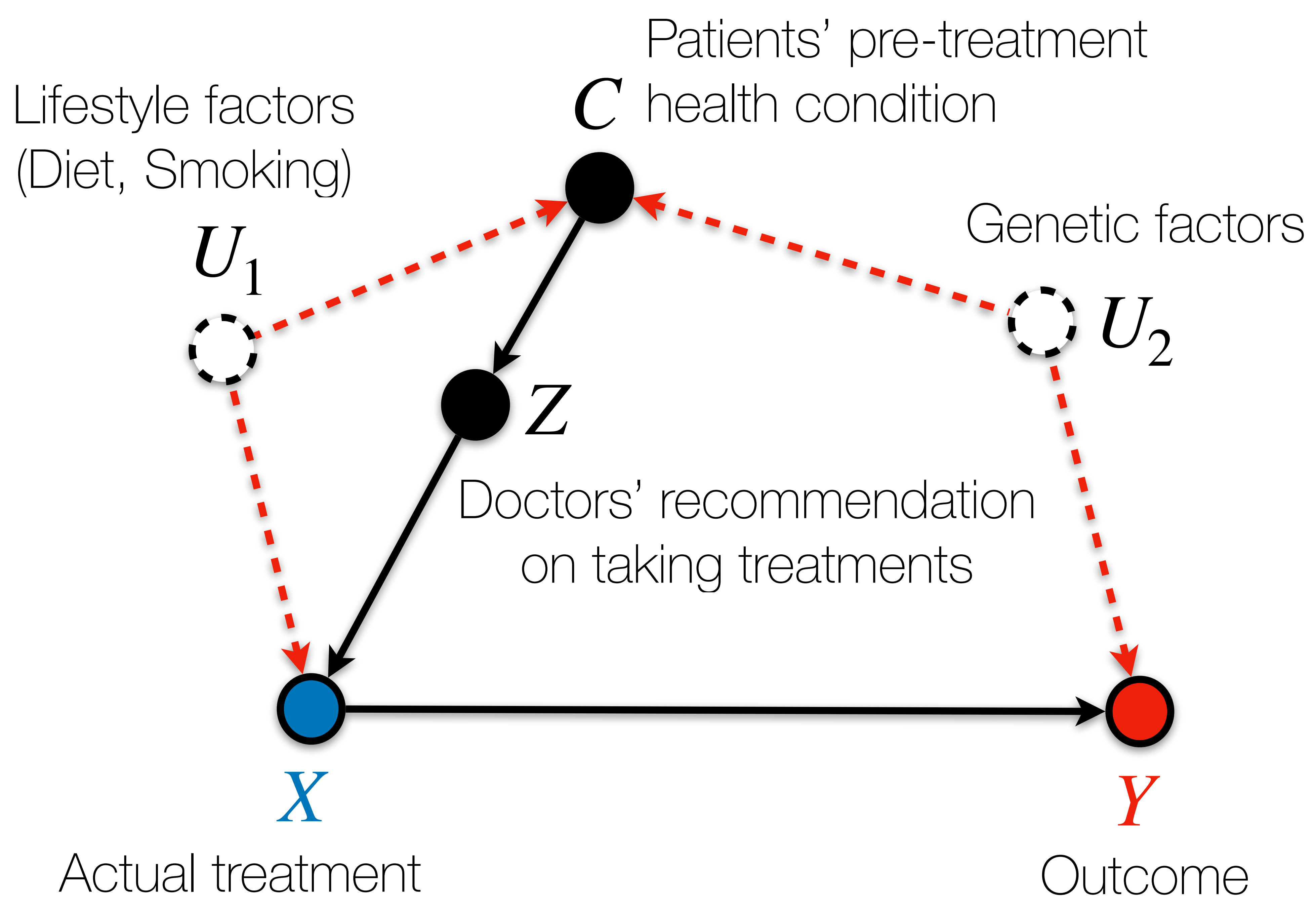
Working example



$$\mathbb{E}[Y \mid do(x)]$$

Non-BD: ∃ spurious paths between treatments and outcomes

Working example

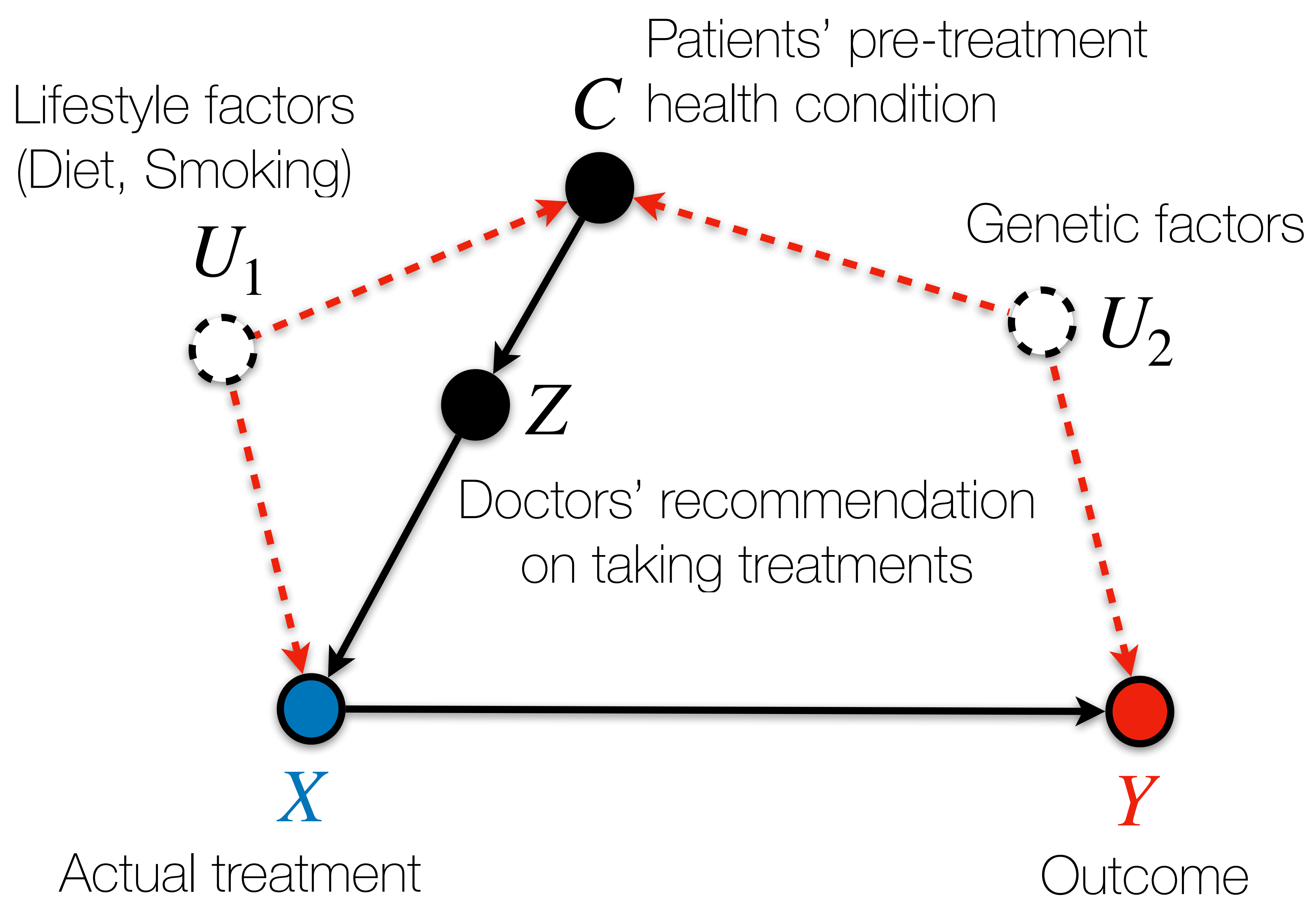


$$\mathbb{E}[Y \mid \mathrm{do}(x)]$$

Identification

$$\mathbb{E}[Y \mid \mathrm{do}(x)] = \frac{\sum_c \mathbb{E}[Y \mid x, z, c]P(x \mid z, c)P(c)}{\sum_c P(x \mid z, c)P(c)}$$

Working example



$$\mathbb{E}[Y \mid do(x)]$$

Identification

$$\mathbb{E}[Y \mid do(x)] = \frac{\sum_c \mathbb{E}[Y \mid x, z, c] P(x \mid z, c) P(c)}{\sum_c P(x \mid z, c) P(c)}$$

Estimation

$$\widehat{\mathbb{E}[Y \mid do(x)]} = \quad ?$$

Identification vs. Estimation

Data	Scenario	Identification	Estimation
$D \sim P$ Observational	Back-door (BD)		
	Non-BD		

Identification vs. Estimation

Data	Scenario	Identification	Estimation
$D \sim P$ Observational	Back-door (BD)	✔	✔
	Non-BD		

Identification vs. Estimation

Data	Scenario	Identification	Estimation
$D \sim P$ Observational	Back-door (BD)	✔	✔
	Non-BD	✔	

Identification vs. Estimation

Data	Scenario	Identification	Estimation
$D \sim P$ Observational	Back-door (BD)	✔	✔
	Non-BD	✔	?

Idea for connecting BD and Identification

Idea for connecting BD and Identification

If $\mathbb{E}[Y \mid \mathrm{do}(x)]$ is expressible as **a function of BDs** (i.e., $\mathbb{E}[Y \mid \mathrm{do}(x)] = g(\{\text{BD}\})$),

Idea for connecting BD and Identification

If $\mathbb{E}[Y \mid \mathrm{do}(x)]$ is expressible as **a function of BDs** (i.e., $\mathbb{E}[Y \mid \mathrm{do}(x)] = g(\{\text{BD}\})$),

then, a general estimator for $\mathbb{E}[Y \mid \mathrm{do}(x)]$ can be constructed

Idea for connecting BD and Identification

If $\mathbb{E}[Y \mid \mathrm{do}(x)]$ is expressible as **a function of BDs** (i.e., $\mathbb{E}[Y \mid \mathrm{do}(x)] = g(\{\text{BD}\})$),

then, a general estimator for $\mathbb{E}[Y \mid \mathrm{do}(x)]$ can be constructed

by strategically **combining DML-BD estimators.**

Background: Causal Effect Identification

Background: Causal Effect Identification

Identification

- spanning a *tree* from $P(\mathbf{V})$
- to reach to causal distribution $P(Y \mid \mathrm{do}(X))$
- through factorization & marginalization of distributions

Background: Causal Effect Identification

Identification

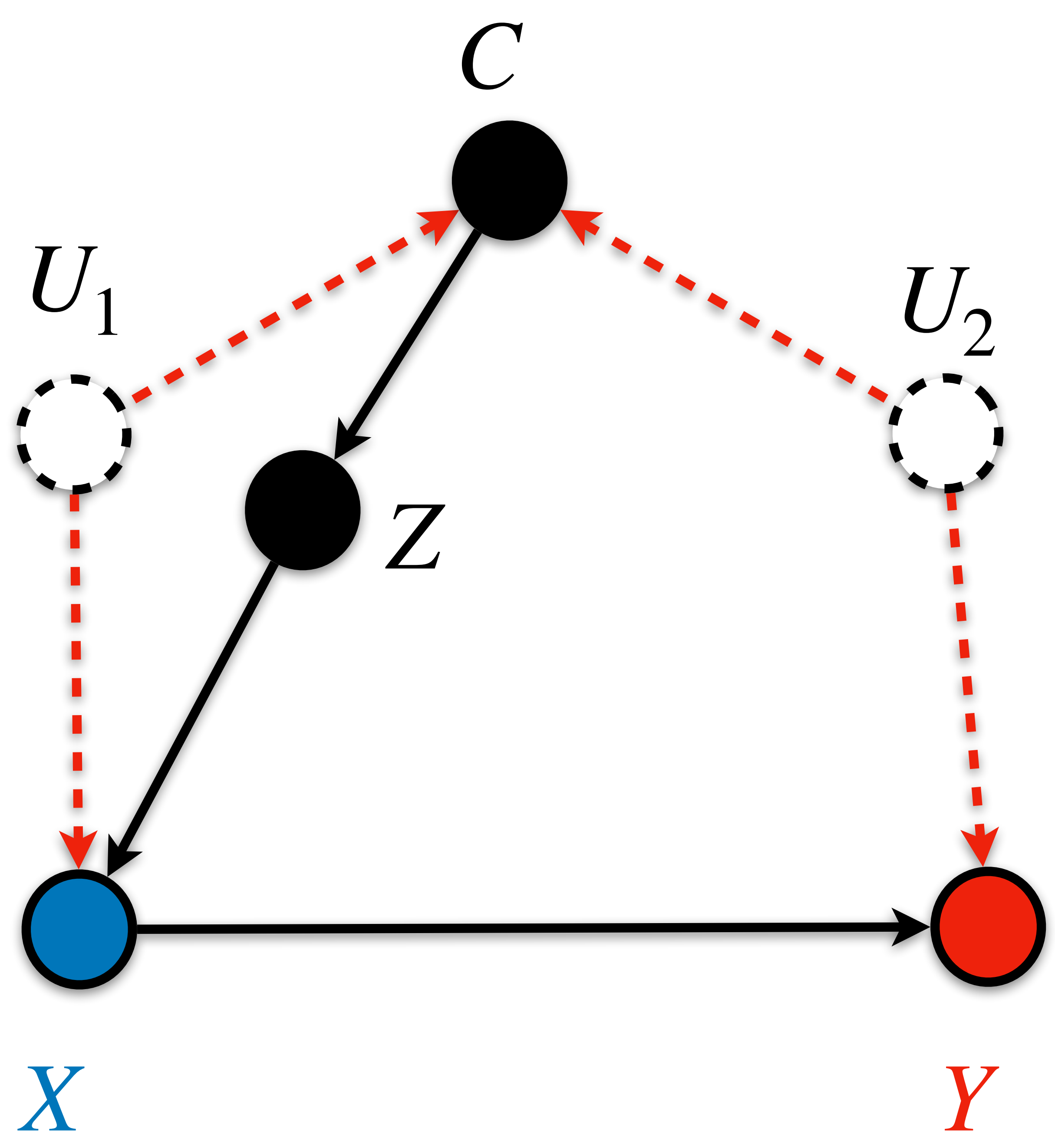
- spanning a *tree* from $P(\mathbf{V})$
- to reach to causal distribution $P(Y \mid \mathrm{do}(X))$
- through factorization & marginalization of distributions

"$P(Y \mid \mathrm{do}(X))$ is a function of $P(\mathbf{V})$ via factorizations & marginalizations"

Background: Causal Effect Identification

Identification

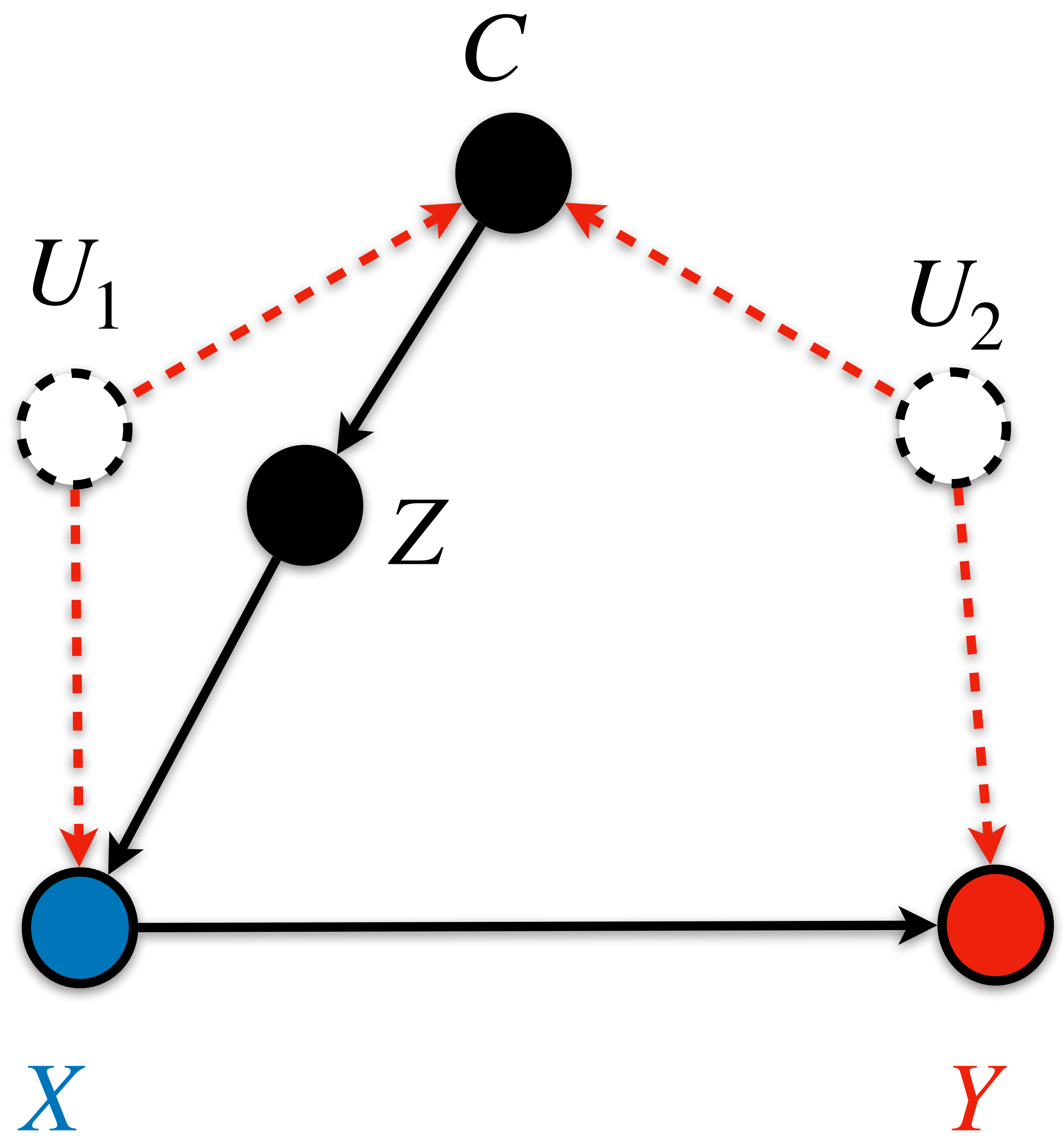
- spanning a *tree* from $P(\mathbf{V})$
- to reach to causal distribution $P(Y \mid \mathrm{do}(X))$
- through factorization & marginalization of distributions



Background: Causal Effect Identification

Identification

- spanning a *tree* from $P(\mathbf{V})$
- to reach to causal distribution $P(Y \mid \mathrm{do}(X))$
- through factorization & marginalization of distributions

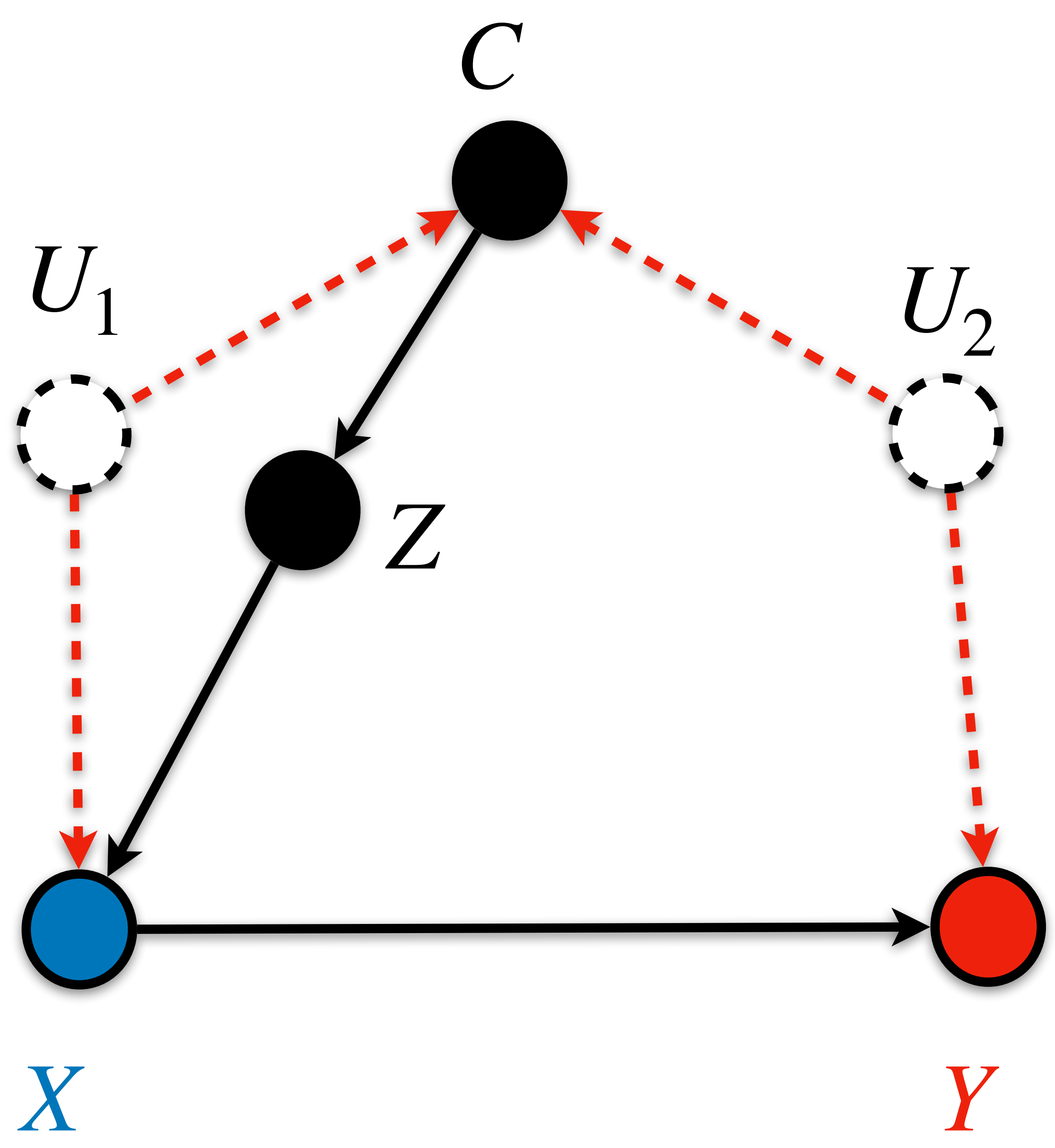


$P(CZXY)$

Background: Causal Effect Identification

Identification

- spanning a *tree* from $P(\mathbf{V})$
- to reach to causal distribution $P(Y \mid \mathrm{do}(X))$
- through factorization & marginalization of distributions



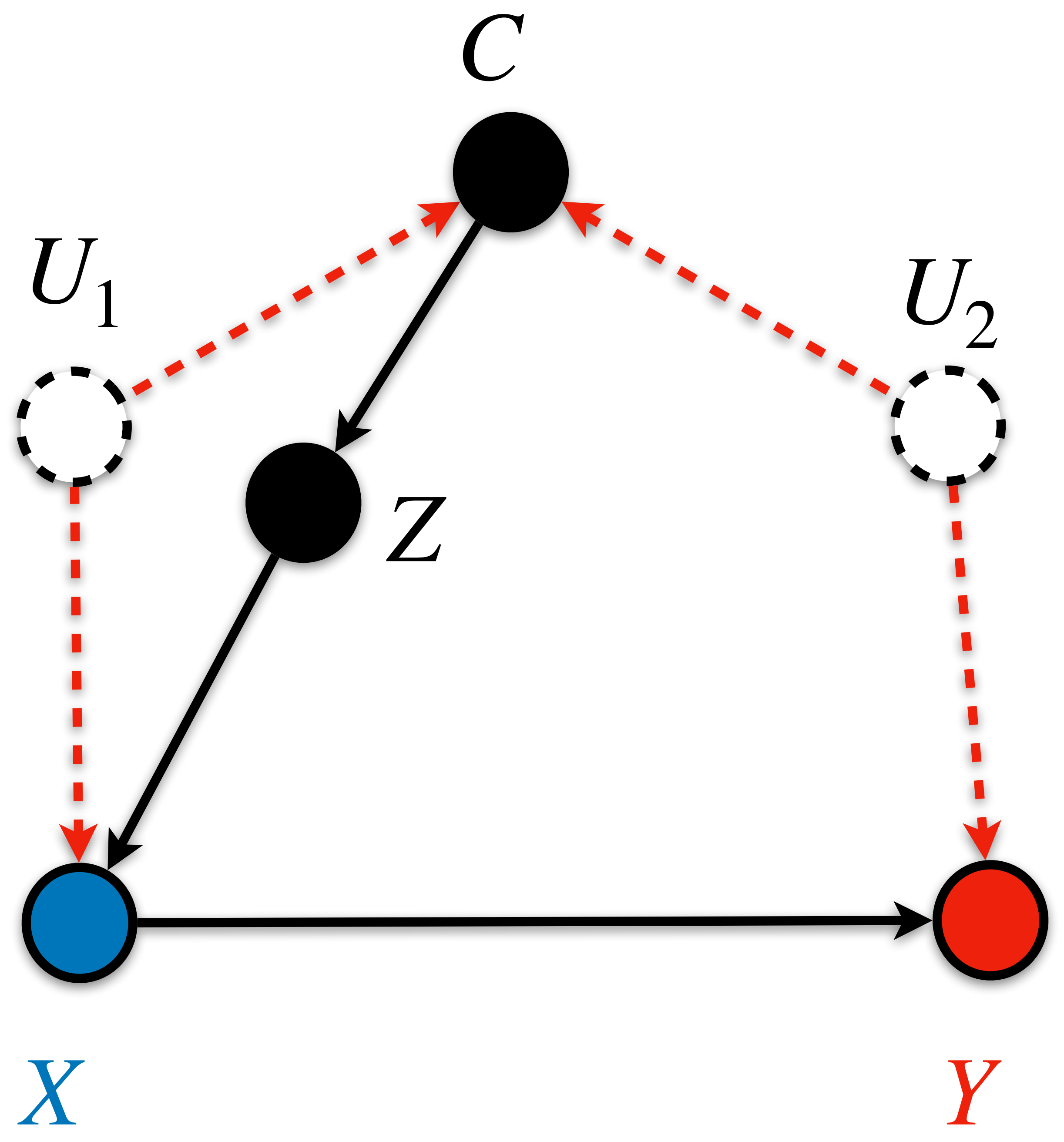
$$P(CZXY) \xrightarrow{\text{Factorization}} P_{\mathrm{do}(Z)}(CXY)$$

$P(C)P(XY \mid ZC)$

Background: Causal Effect Identification

Identification

- spanning a *tree* from $P(\mathbf{V})$
- to reach to causal distribution $P(Y \mid \mathrm{do}(X))$
- through factorization & marginalization of distributions



$$P(CZXY) \xrightarrow{\text{Factorization}} P_{\mathrm{do}(Z)}(CXY) \xrightarrow{\text{Marginalization}} P_{\mathrm{do}(Z)}(XY)$$

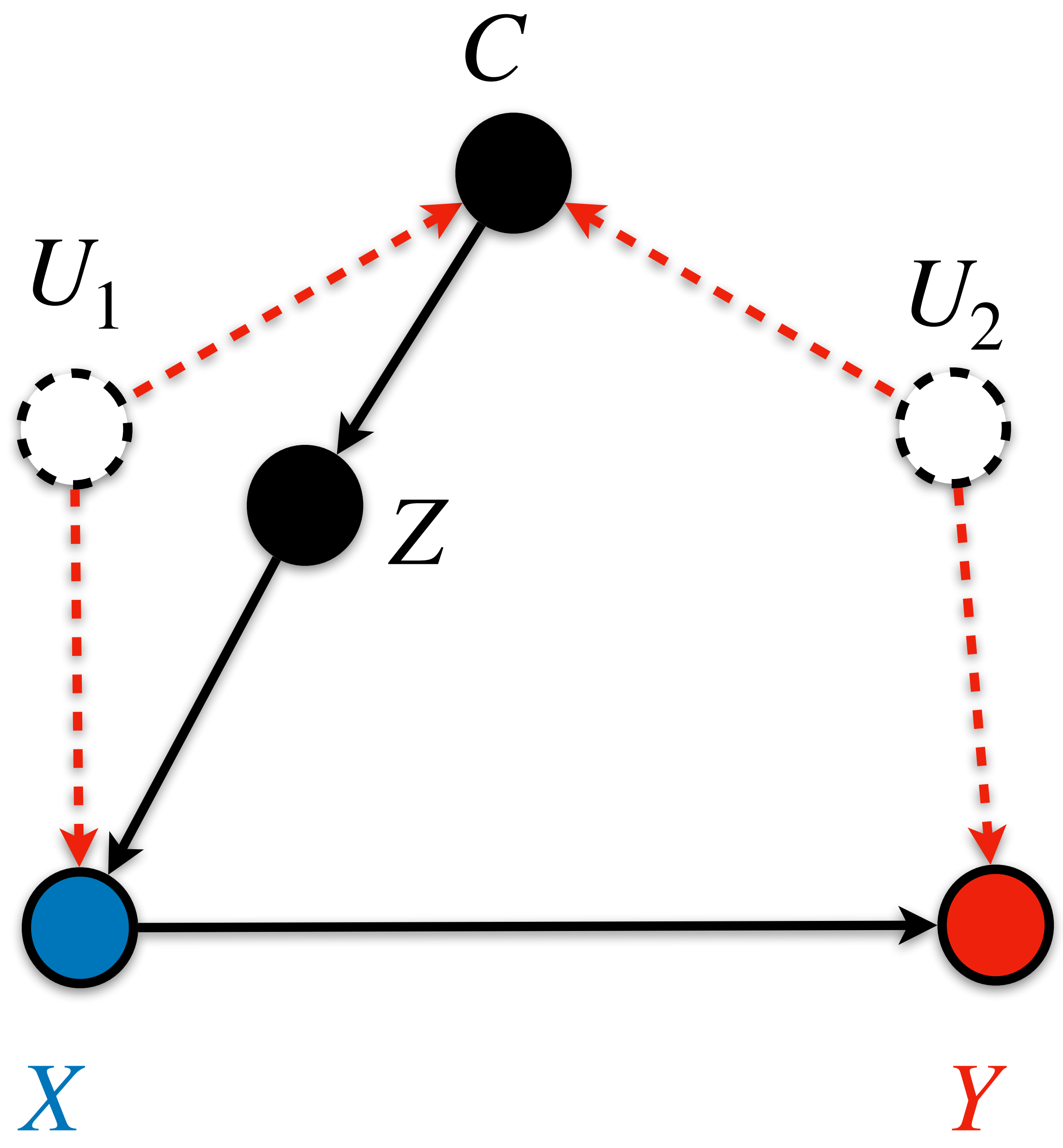
$P(C)P(XY \mid ZC)$

$\sum_c P(c)P(XY \mid Zc)$

Background: Causal Effect Identification

Identification

- spanning a *tree* from $P(\mathbf{V})$
- to reach to causal distribution $P(Y \mid \mathrm{do}(X))$
- through factorization & marginalization of distributions



$$P(CZXY) \xrightarrow{\text{Factorization}} P_{\mathrm{do}(Z)}(CXY) \xrightarrow{\text{Marginalization}} P_{\mathrm{do}(Z)}(XY) \xrightarrow{\text{Factorization}} P(Y \mid \mathrm{do}(X))$$

$P(C)P(XY \mid ZC)$

$\sum_c P(c)P(XY \mid Zc)$

$$P_{\mathrm{do}(Z)}(Y \mid X) = \frac{\sum_c P(c)P(XY \mid Zc)}{\sum_c P(c)P(X \mid Zc)}$$

My Approach: 3-Step

My Approach: 3-Step

So far,

- *Back-door adjustment (*BD*)* can be computed through DML-BD
- The computation tree for *causal effect identification* composes of interventional distributions.

My Approach: 3-Step

To connect BD & Identification,

My Approach: 3-Step

To connect BD & Identification,

1. **Check** if each interventional distribution on the tree is expressible as BD

My Approach: 3-Step

To connect BD & Identification,

1. **Check** if each interventional distribution on the tree is expressible as BD
2. **Express** causal effects as a function of BD

My Approach: 3-Step

To connect BD & Identification,

1. **Check** if each interventional distribution on the tree is expressible as BD
2. **Express** causal effects as a function of BD
3. **Construct** robust estimators by combining DML-BD

Estimating Causal Effects in 3-Steps

1 **Check** if each interventional distribution on the tree is expressible as BD

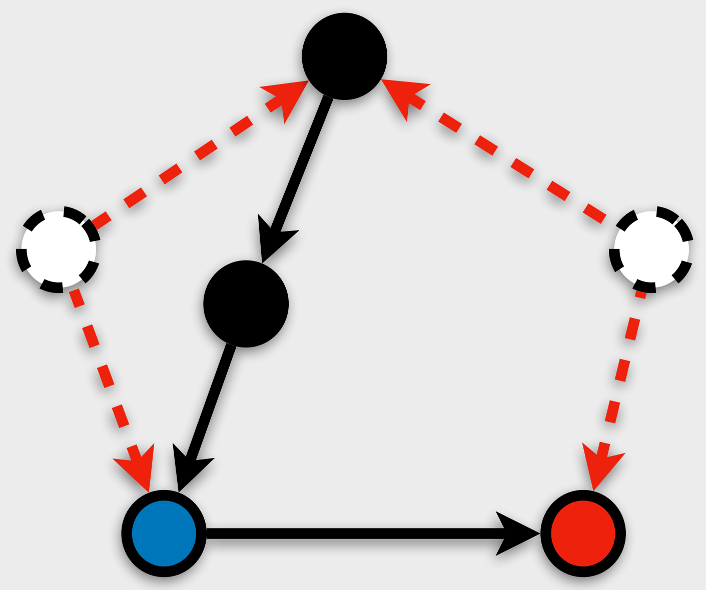
Estimating Causal Effects in 3-Steps

1. **Check** if each interventional distribution on the tree is expressible as BD

Estimating Causal Effects in 3-Steps

1 **Check** if each interventional distribution on the tree is expressible as BD

Graph



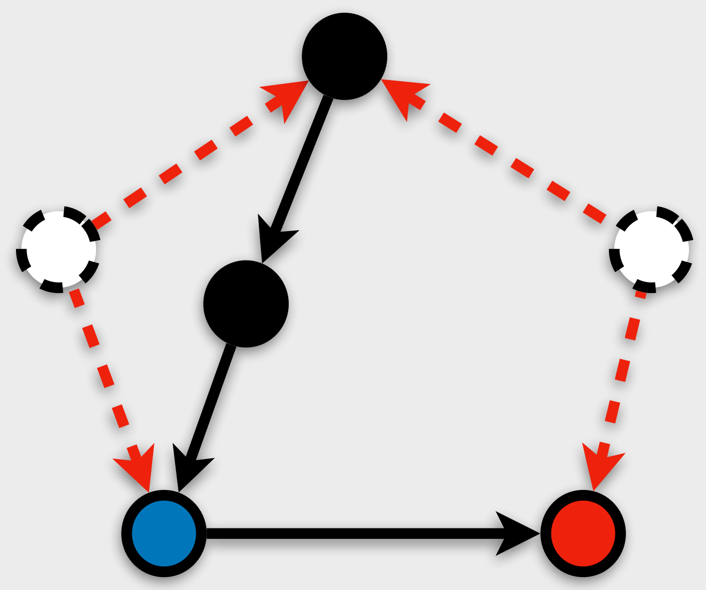
Interventional Distribution

$P_{\mathrm{do}(a)}(B)$

Estimating Causal Effects in 3-Steps

1 **Check** if each interventional distribution on the tree is expressible as BD

Graph



Jung et al., AAAI 2021

Jung et al., NeurIPS 2024

CheckBD

Check if $P_{\mathrm{do}(a)}(B)$ is expressed as BD

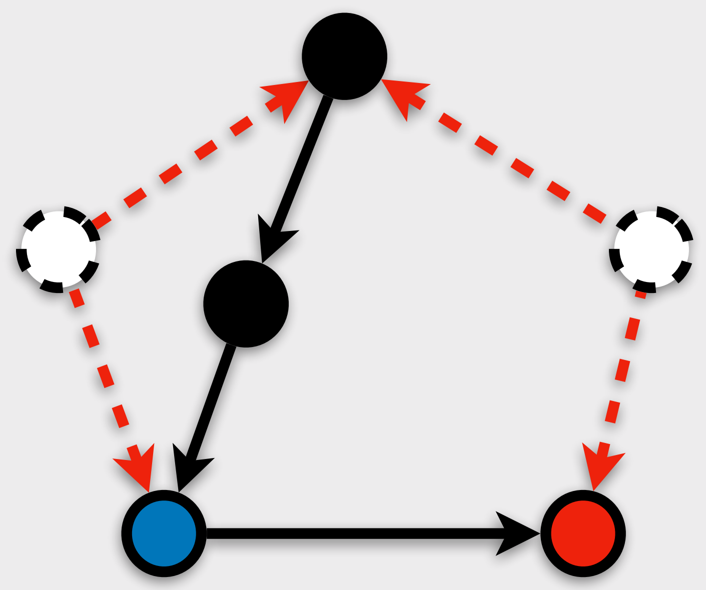
Interventional Distribution

$P_{\mathrm{do}(a)}(B)$

Estimating Causal Effects in 3-Steps

1 **Check** if each interventional distribution on the tree is expressible as BD

Graph



Jung et al., AAAI 2021

Jung et al., NeurIPS 2024

CheckBD

Check if $P_{\mathrm{do}(a)}(B)$ is expressed as BD

$P_{\mathrm{do}(a)}(B) = \mathrm{BD}(\mu, \pi)$ if BD-expressible

Interventional Distribution

$P_{\mathrm{do}(a)}(B)$

Estimating Causal Effects in 3-Steps

1 **Check** if each interventional distribution on the tree is expressible as BD

Theorem

$P_{\mathrm{do}(a)}(B)$ is expressible through BD

if and only if

$P_{\mathrm{do}(a)}(B)$ passes CheckBD

$= \mathrm{BD}(\mu, \pi)$ if

xpressible

Interventional Distribution

$P_{\mathrm{do}(a)}(B)$

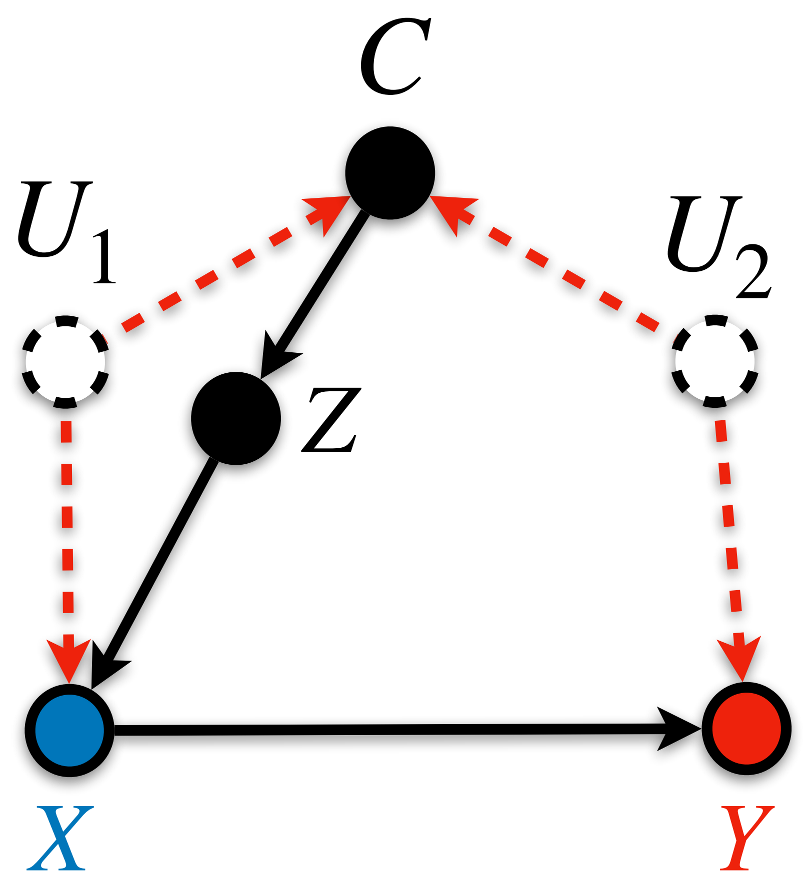
Estimating Causal Effects in 3-Steps

Estimating Causal Effects in 3-Steps

2 **Express** causal effects as a function of BD

Estimating Causal Effects in 3-Steps

2 **Express** causal effects as a function of BD



$$P(C)P(XY \mid ZC)$$

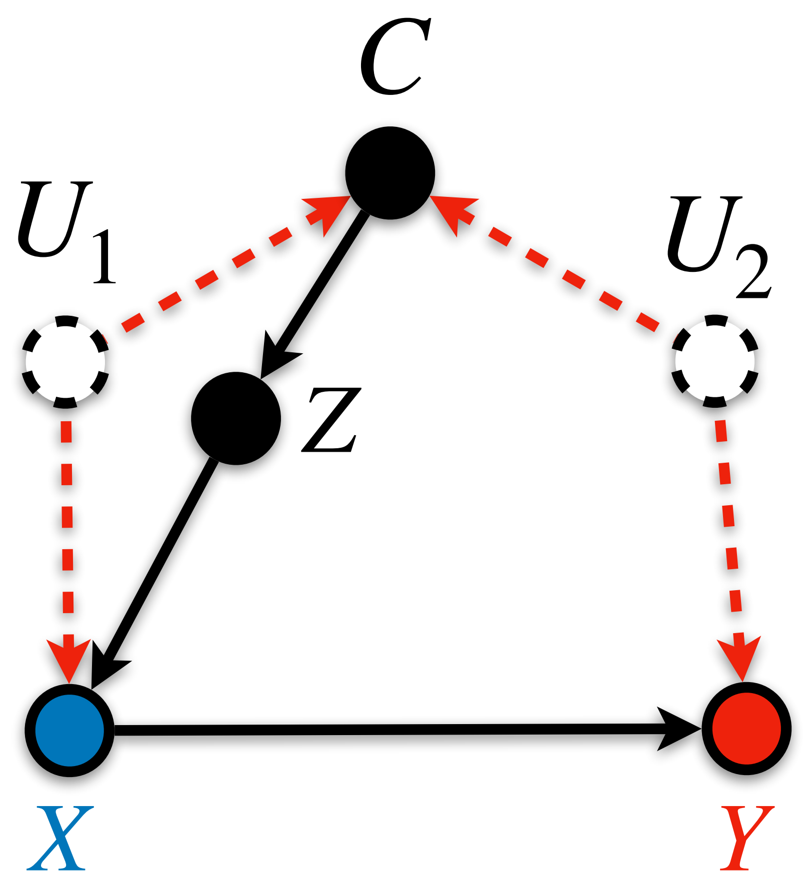
$$\sum_c P(c)P(XY \mid Zc)$$

$$\frac{\sum_c P(c)P(XY \mid Zc)}{\sum_c P(c)P(X \mid Zc)}$$

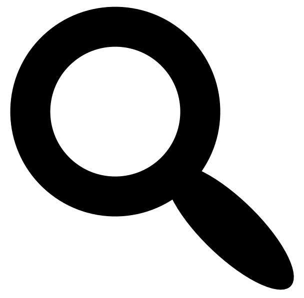
$$P(CZXY) \xrightarrow{\text{Factorization}} P_{\text{do}(Z)}(CXY) \xrightarrow{\text{Marginalization}} P_{\text{do}(Z)}(XY) \xrightarrow{\text{Factorization}} P(Y \mid \text{do}(X))$$

Estimating Causal Effects in 3-Steps

2 **Express** causal effects as a function of BD



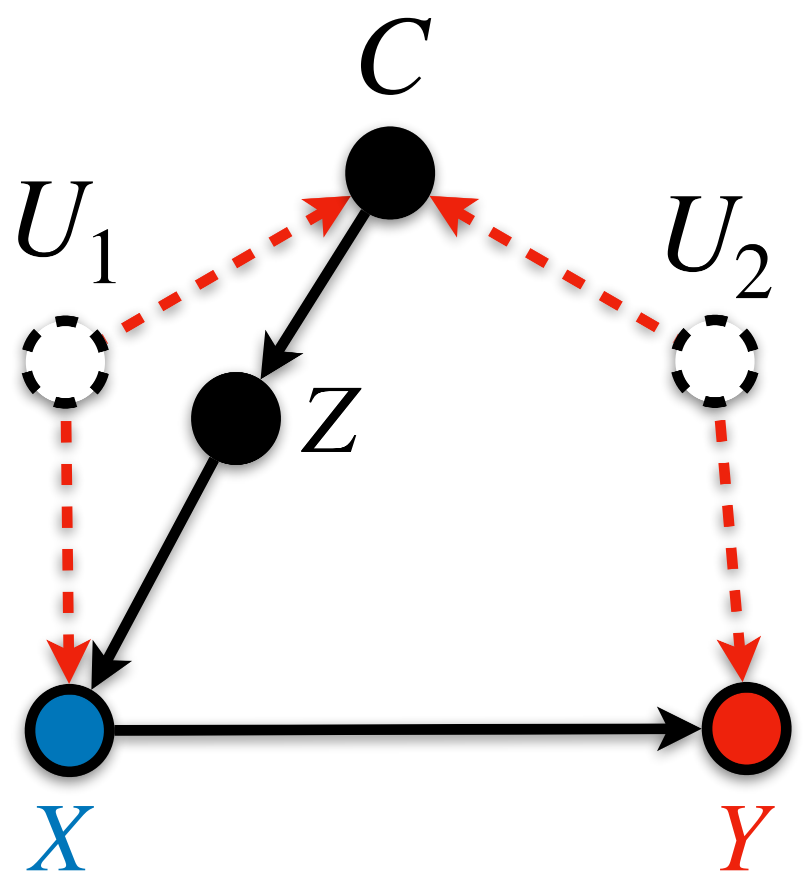
$$P(CZXY) \xrightarrow{\text{Factorization}} P_{\text{do}(Z)}(CXY) \xrightarrow{\text{Marginalization}} P_{\text{do}(Z)}(XY) \xrightarrow{\text{Factorization}} P(Y \mid \text{do}(X))$$



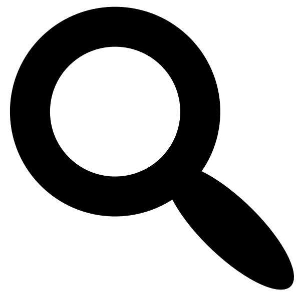
CheckBD

Estimating Causal Effects in 3-Steps

2 **Express** causal effects as a function of BD



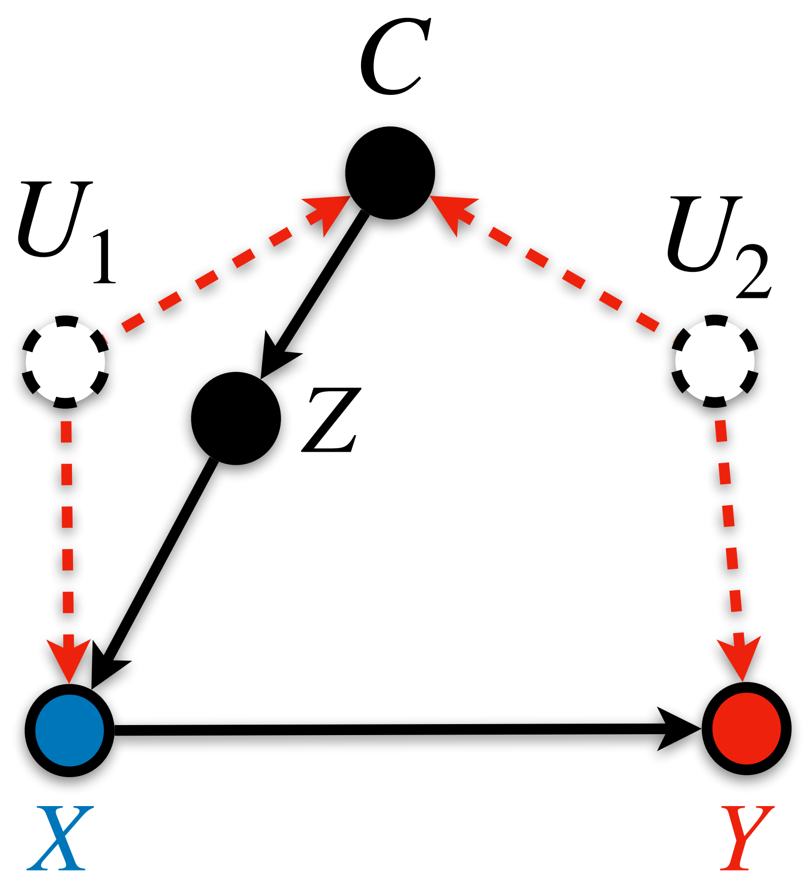
$$P(CZXY) \xrightarrow{\text{Factorization}} P_{\text{do}(Z)}(CXY) \xrightarrow{\text{Marginalization}} P_{\text{do}(Z)}(XY) \xrightarrow{\text{Factorization}} P(Y \mid \text{do}(X))$$



CheckBD

Estimating Causal Effects in 3-Steps

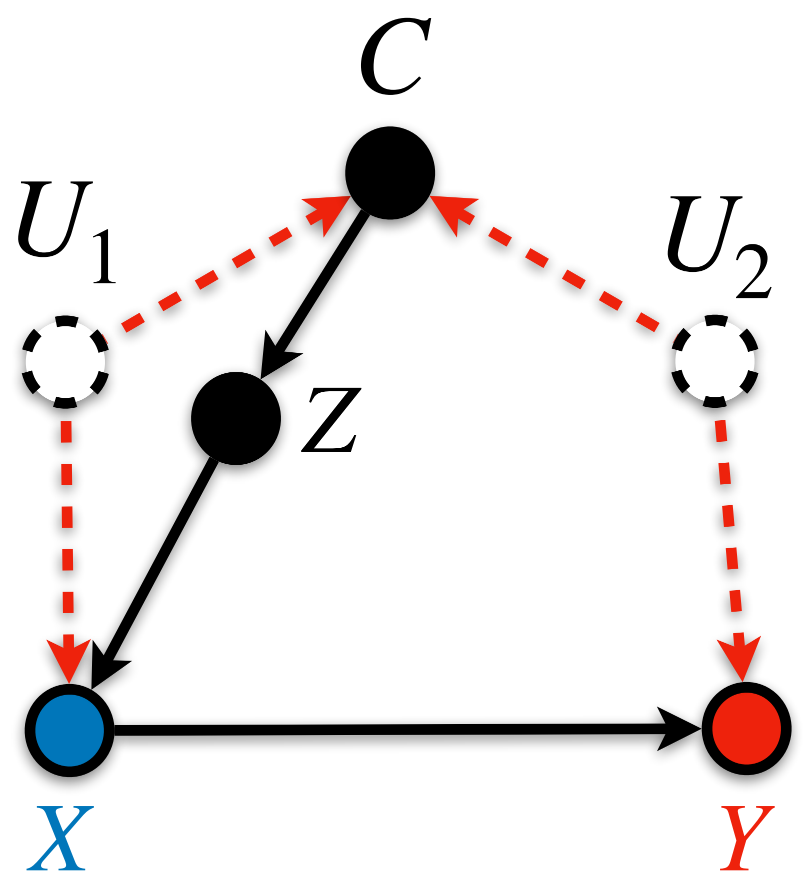
2 **Express** causal effects as a function of BD



$$\mathrm{BD}_1(\mu, \pi) \xrightarrow{\text{Factorization}} P(Y \mid \mathrm{do}(X))$$

Estimating Causal Effects in 3-Steps

2 **Express** causal effects as a function of BD



$$= \frac{\mathrm{BD}_1(\mu, \pi)}{\mathrm{BD}_2(\mu, \pi)}$$

$$\mathrm{BD}_1(\mu, \pi) \xrightarrow{\text{Factorization}} P(Y \mid \mathrm{do}(X))$$

Estimating Causal Effects in 3-Steps

2 **Express** causal effects as a function of BD

C

Theorem

Causal effect is identifiable

If and only if

It's expressible as a ***function of BD***

$$= \frac{\mathrm{BD}_1(\mu, \pi)}{\mathrm{BD}_2(\mu, \pi)}$$

$$\mathrm{BD}_1(\mu, \pi) \xrightarrow{\text{Factorization}} P(Y \mid \mathrm{do}(X))$$

DML-ID: Estimator for Identifiable Causal Effects

DML-ID: Estimator for Identifiable Causal Effects

3 **Construct** robust estimators by combining DML-BD

DML-ID: Estimator for Identifiable Causal Effects

3 **Construct** robust estimators by combining DML-BD

$$\mathbb{E}[Y \mid do(\mathbf{x})] = g(\{\mathrm{BD}(\mu_1, \pi_1), \mathrm{BD}(\mu_2, \pi_2), \cdots, \mathrm{BD}(\mu_m, \pi_m)\})$$

DML-ID: Estimator for Identifiable Causal Effects

3 **Construct** robust estimators by combining DML-BD

$$\mathbb{E}[Y \mid do(\mathbf{x})] = g(\{\text{BD}(\mu_1, \pi_1), \text{BD}(\mu_2, \pi_2), \cdots, \text{BD}(\mu_m, \pi_m)\})$$

"DML-ID"

$$\mathbb{E}[\widehat{Y \mid do(\mathbf{x}})]$$

DML-ID: Estimator for Identifiable Causal Effects

3 **Construct** robust estimators by combining DML-BD

$$\mathbb{E}[Y \mid \text{do}(\mathbf{x})] = g(\{\text{BD}(\mu_1, \pi_1), \text{BD}(\mu_2, \pi_2), \cdots, \text{BD}(\mu_m, \pi_m)\})$$

"DML-ID"

$$\widehat{\mathbb{E}[Y \mid \text{do}(\mathbf{x})]} \triangleq g(\{ \qquad \})$$

DML-ID: Estimator for Identifiable Causal Effects

❸ **Construct** robust estimators by combining DML-BD

$$\mathbb{E}[Y \mid do(\mathbf{x})] = g(\{\text{BD}(\mu_1,\pi_1), \text{BD}(\mu_2,\pi_2), \cdots, \text{BD}(\mu_m,\pi_m)\})$$

DML-BD   DML-BD   ⋯   DML-BD

"DML-ID"

$$\widehat{\mathbb{E}[Y \mid do(\mathbf{x})]} \triangleq g(\{\widehat{\text{BD}}(\mu_1,\pi_1), \widehat{\text{BD}}(\mu_2,\pi_2), \ldots, \widehat{\text{BD}}(\mu_m,\pi_m)\})$$

Robustness of DML-ID

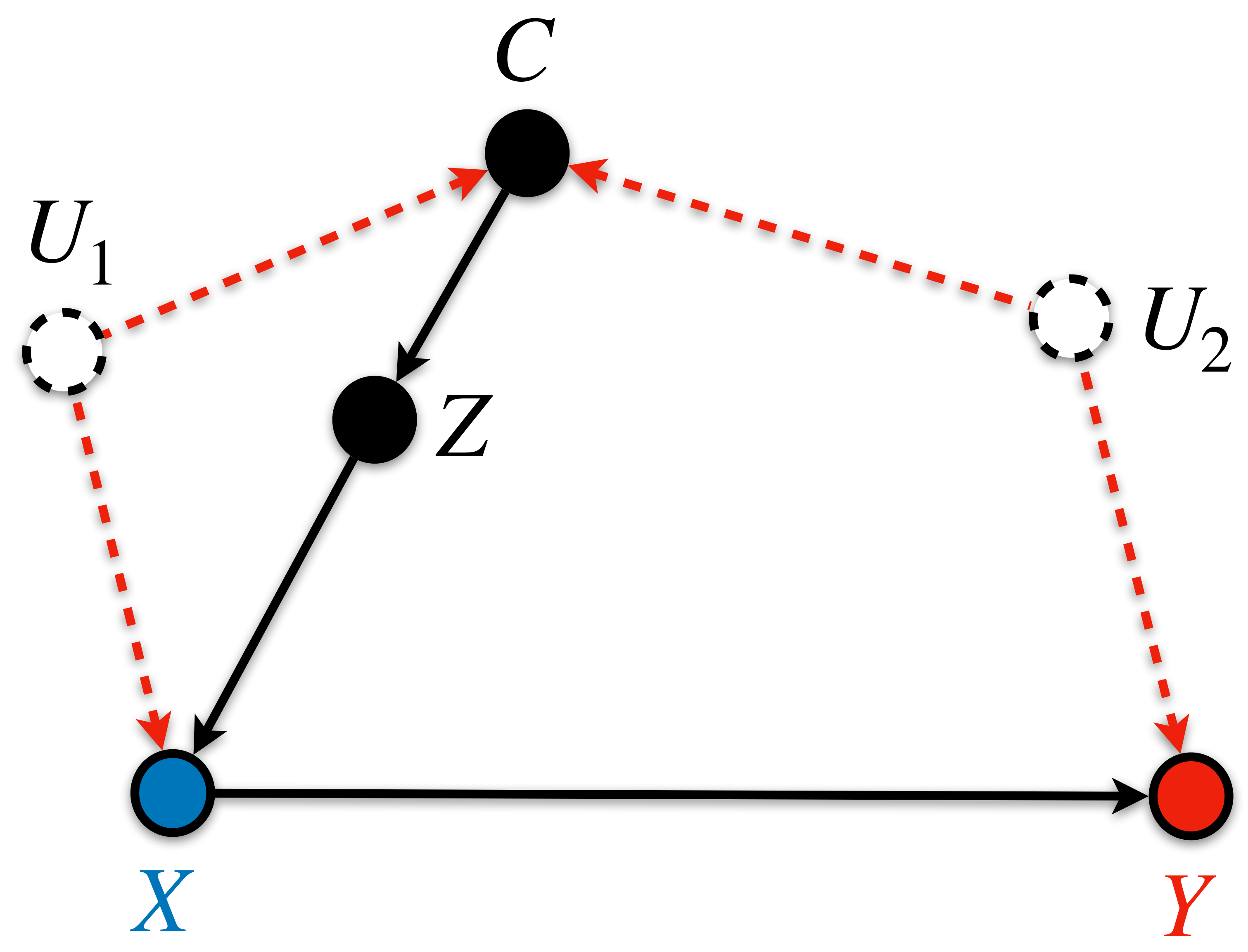
Theorem

$$\text{Error}(\text{DML-ID}, \mathbb{E}[Y \mid \text{do}(x)]) = \sum_{i=1}^{m} \text{Error}(\hat{\mu}_i, \mu_i) \times \text{Error}(\hat{\pi}_i, \pi_i)$$

- **Double Robustness:** Error $= 0$ if either $\hat{\mu}_i = \mu_i$ or $\hat{\pi}_i = \pi_i$ for all $i = 1, \cdots, m$.
- **Fast Convergence:** Error $\to 0$ *fast* even when $\hat{\mu}_i \to \mu_i$ and $\hat{\pi}_i \to \pi_i$ *slow*.

DML-ID - Simulation

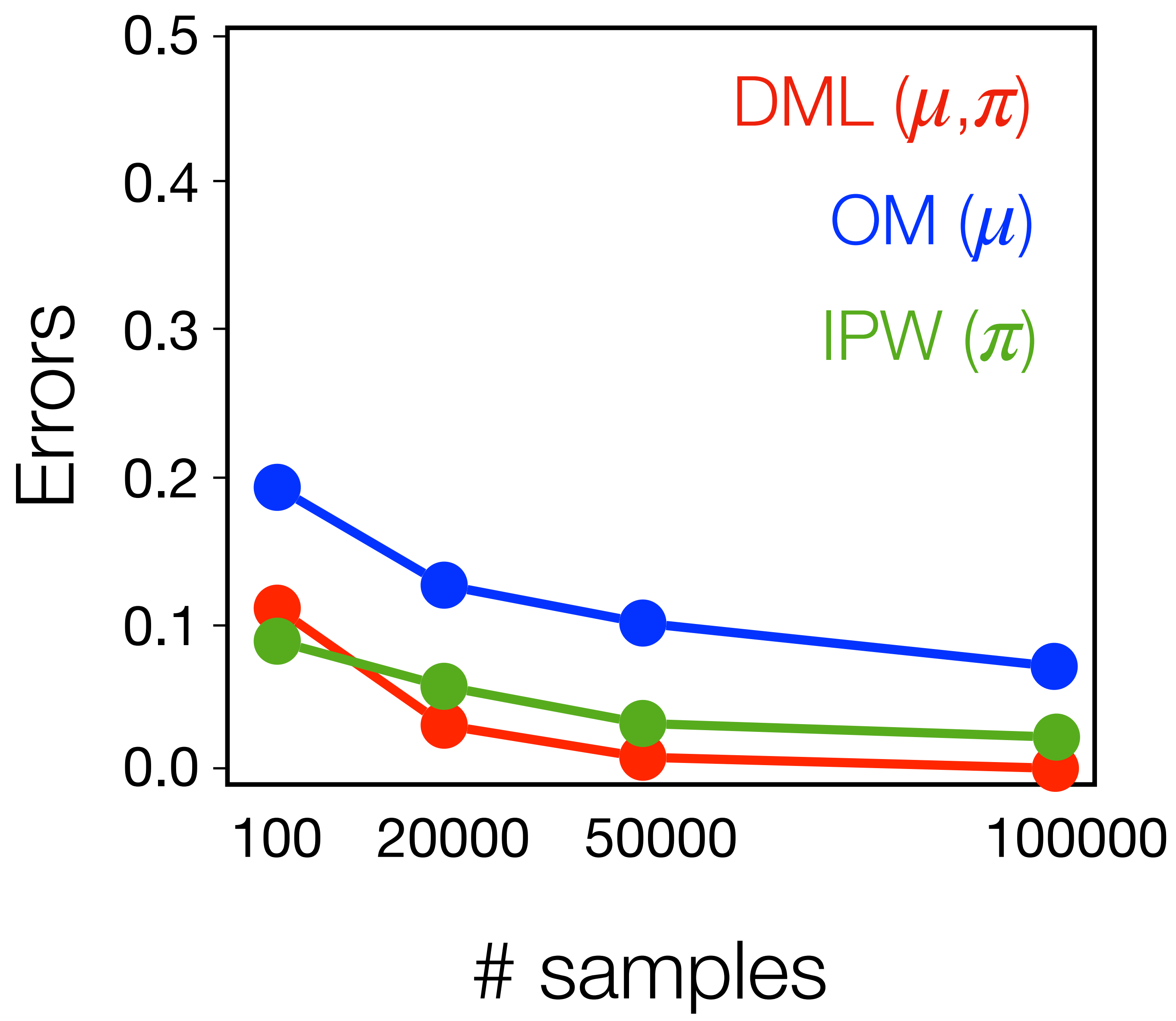
DML-ID - Simulation



DML-ID - Simulation

Fast Convergence

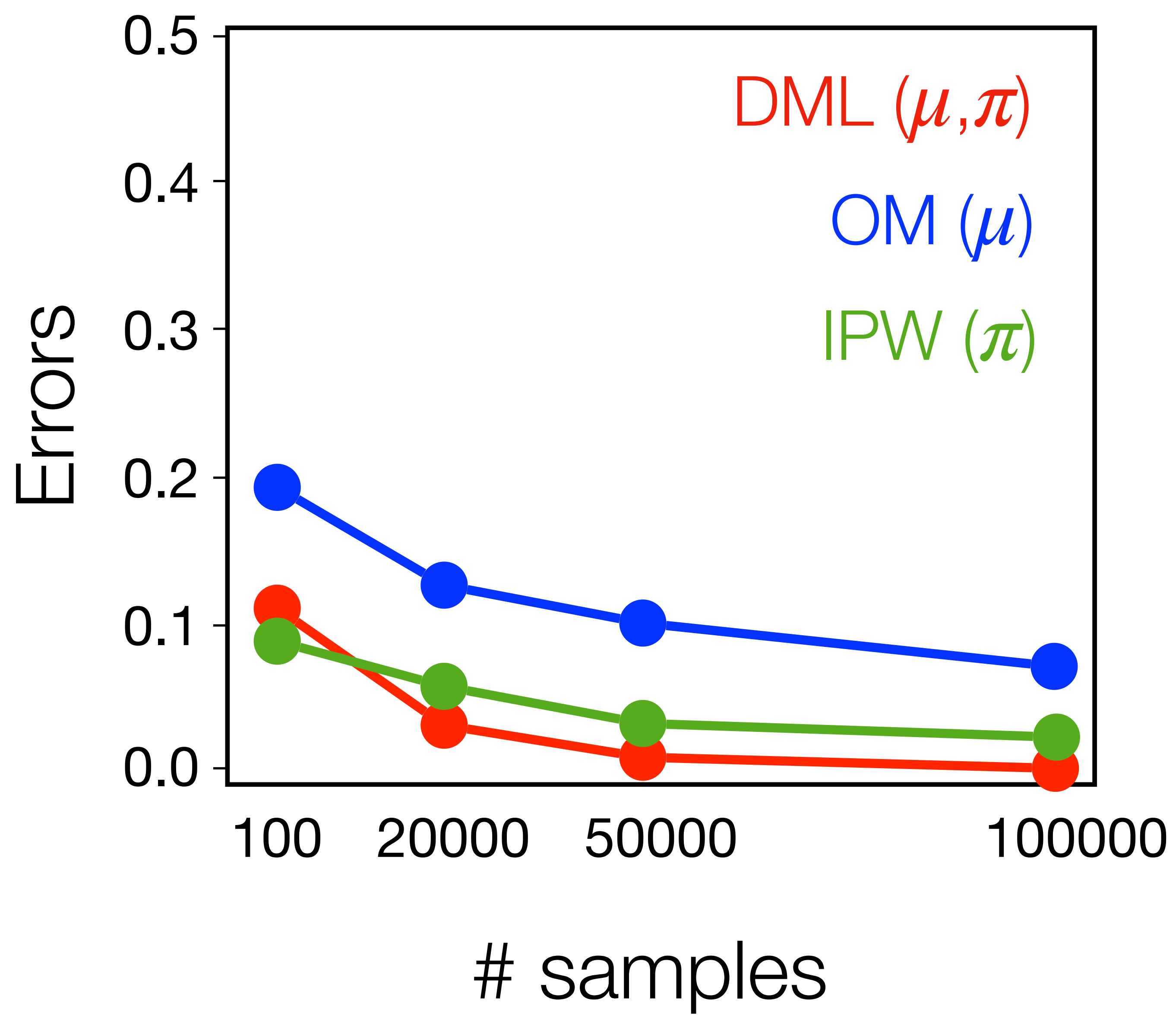
$(\hat{\boldsymbol{\mu}}, \hat{\boldsymbol{\pi}}) \to (\boldsymbol{\mu}_0, \boldsymbol{\pi}_0)$ slowly



DML-ID - Simulation

Fast Convergence

$(\hat{\boldsymbol{\mu}},\hat{\boldsymbol{\pi}}) \rightarrow (\boldsymbol{\mu}_0, \boldsymbol{\pi}_0)$ slowly

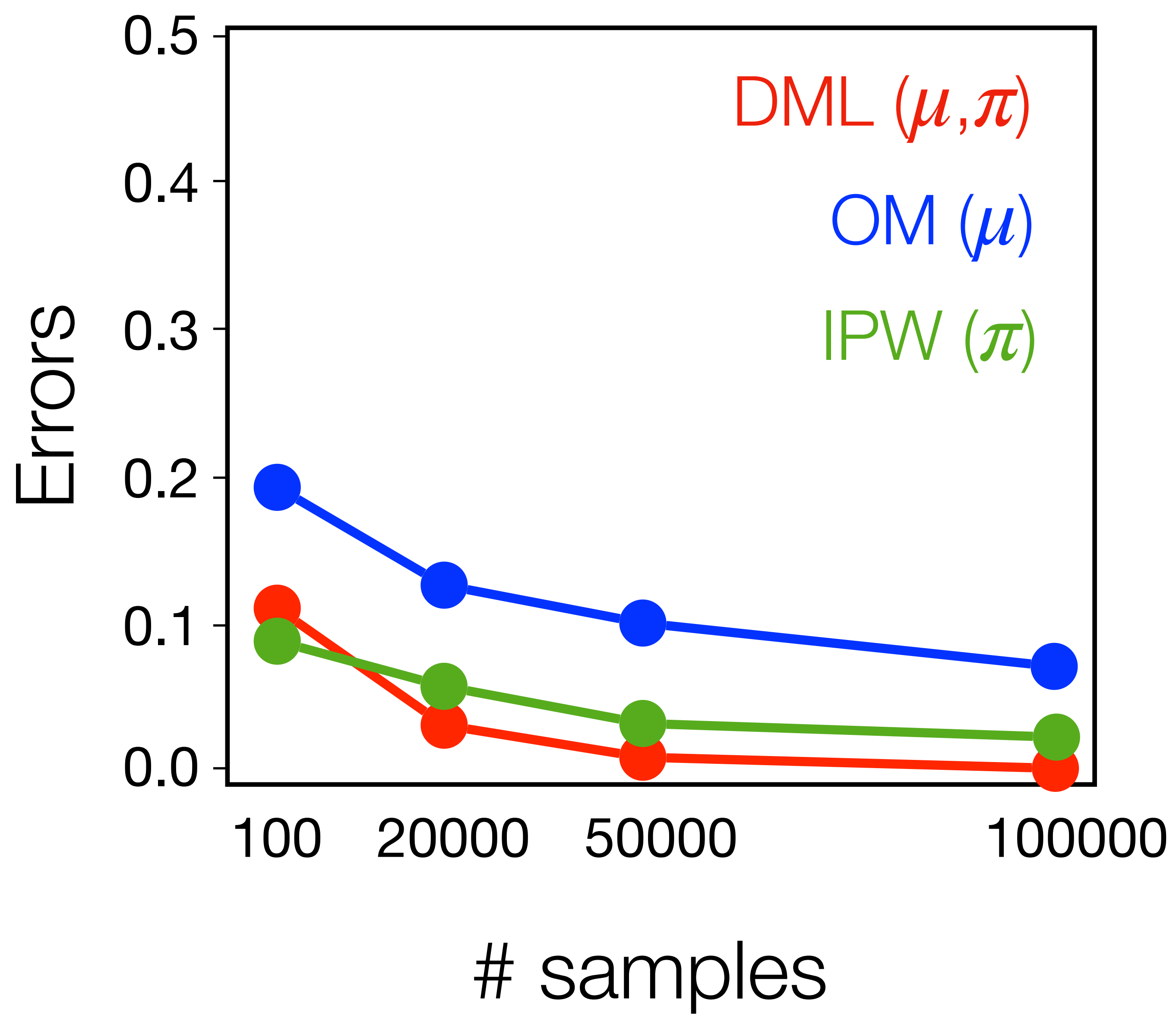


DML-ID converges fast, even when $(\hat{\boldsymbol{\mu}},\hat{\boldsymbol{\pi}})$ converge slowly

DML-ID - Simulation

Fast Convergence

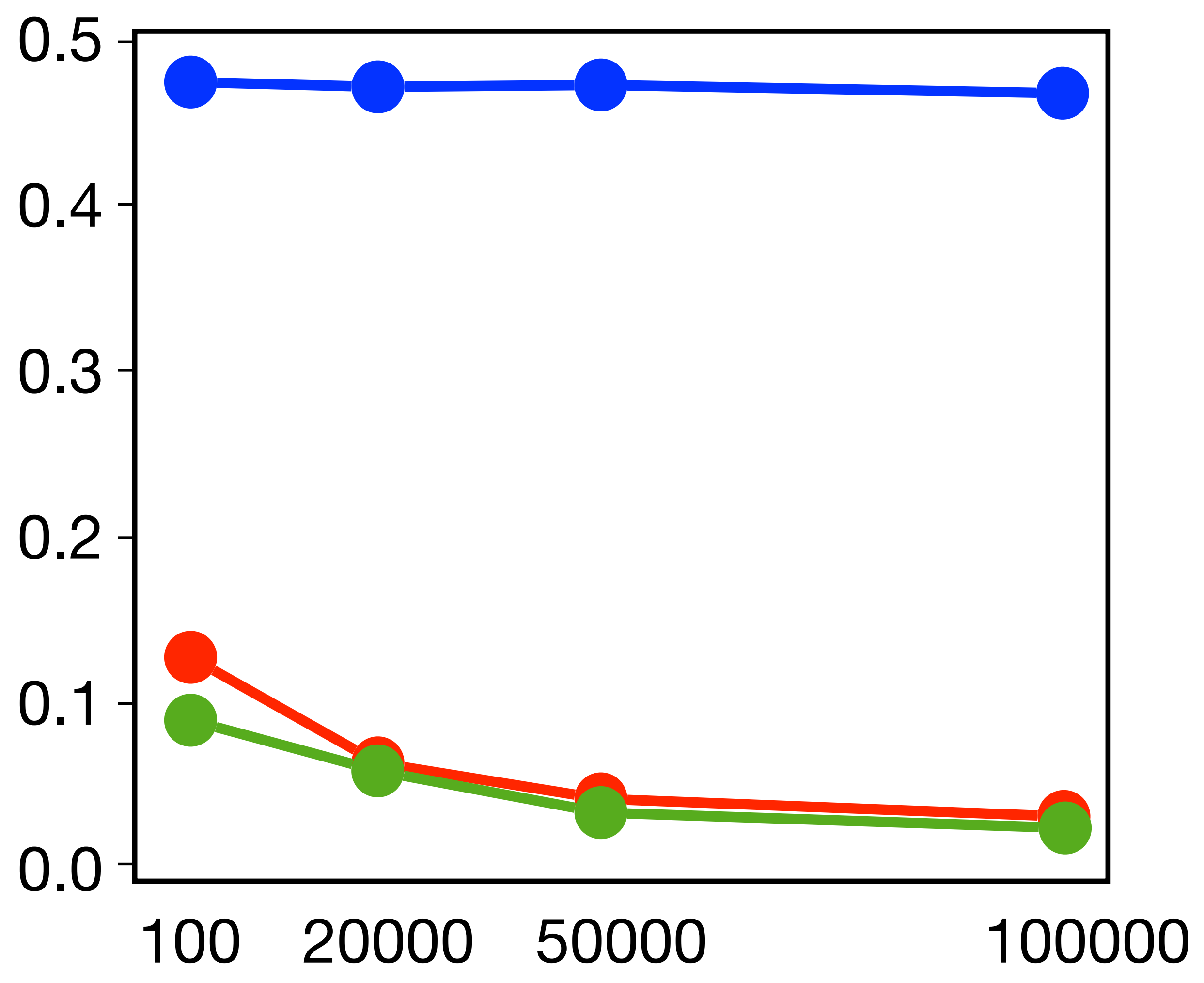
$(\hat{\mu},\hat{\pi}) \rightarrow (\mu_0, \pi_0)$ slowly



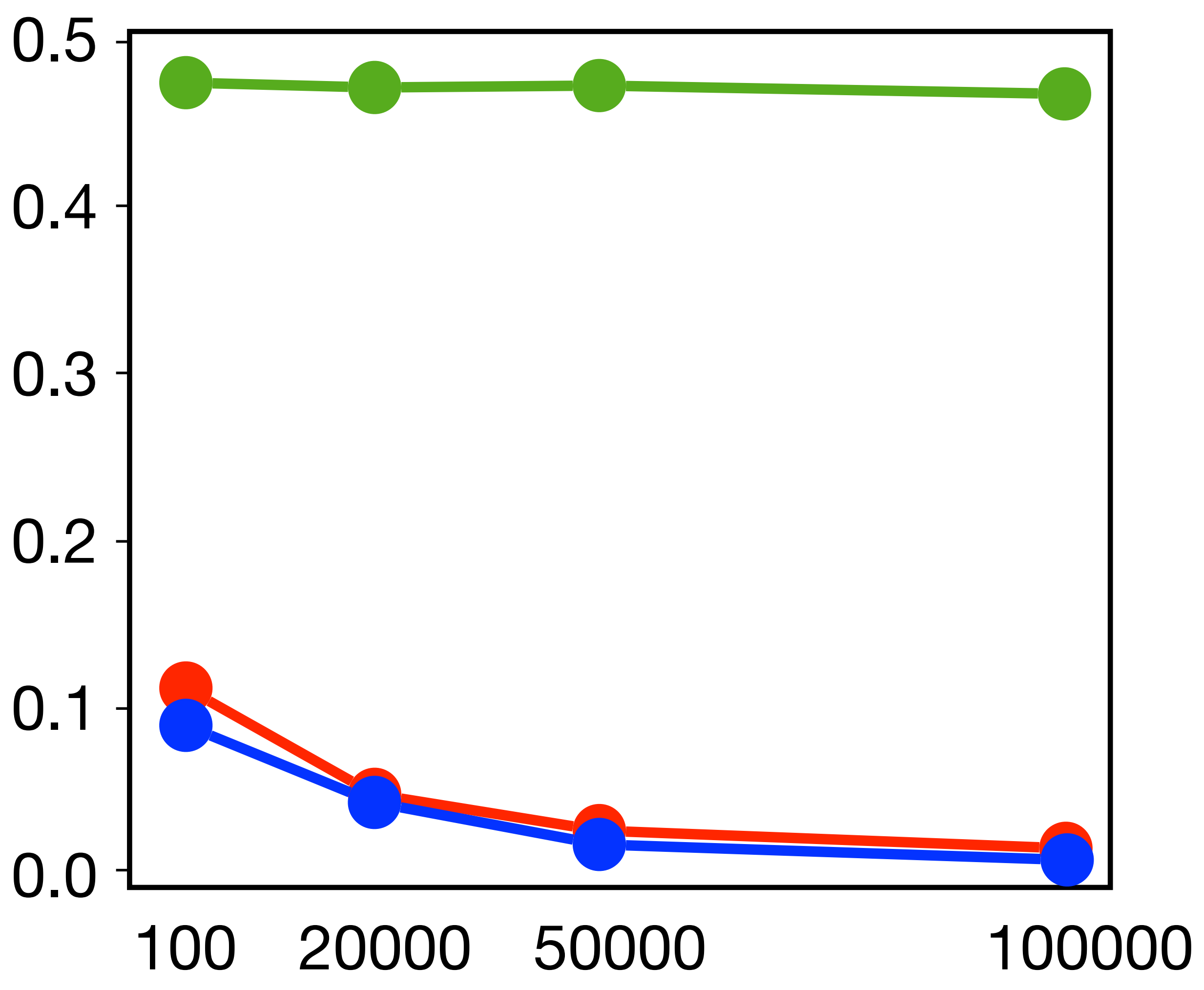
DML-ID converges fast, even when $(\hat{\mu},\hat{\pi})$ converge slowly

Double Robustness

$\hat{\mu}$ misspecified ($\hat{\mu}\neq\mu$)



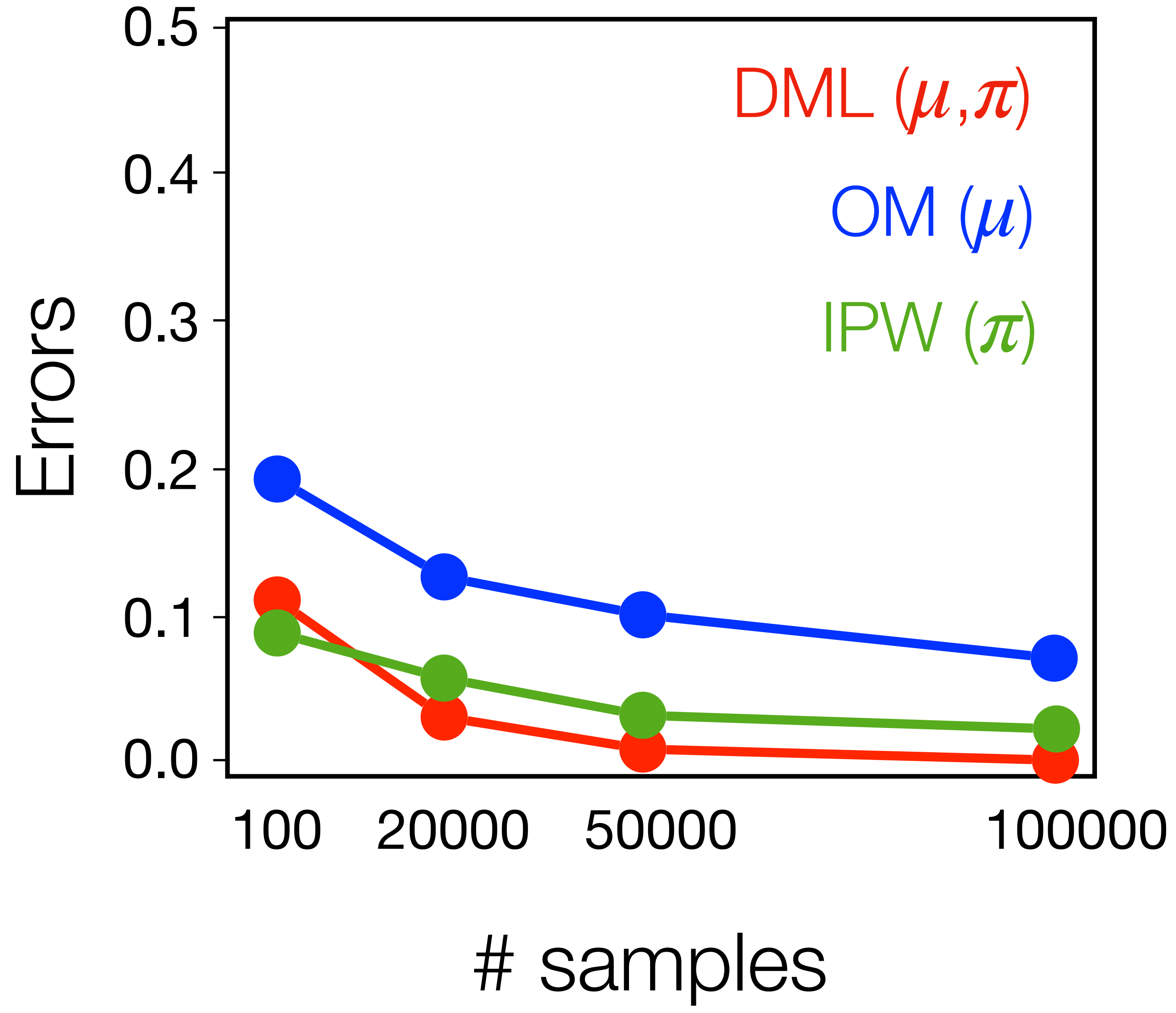
$\hat{\pi}$ misspecified ($\hat{\pi}\neq\pi$)



DML-ID - Simulation

Fast Convergence

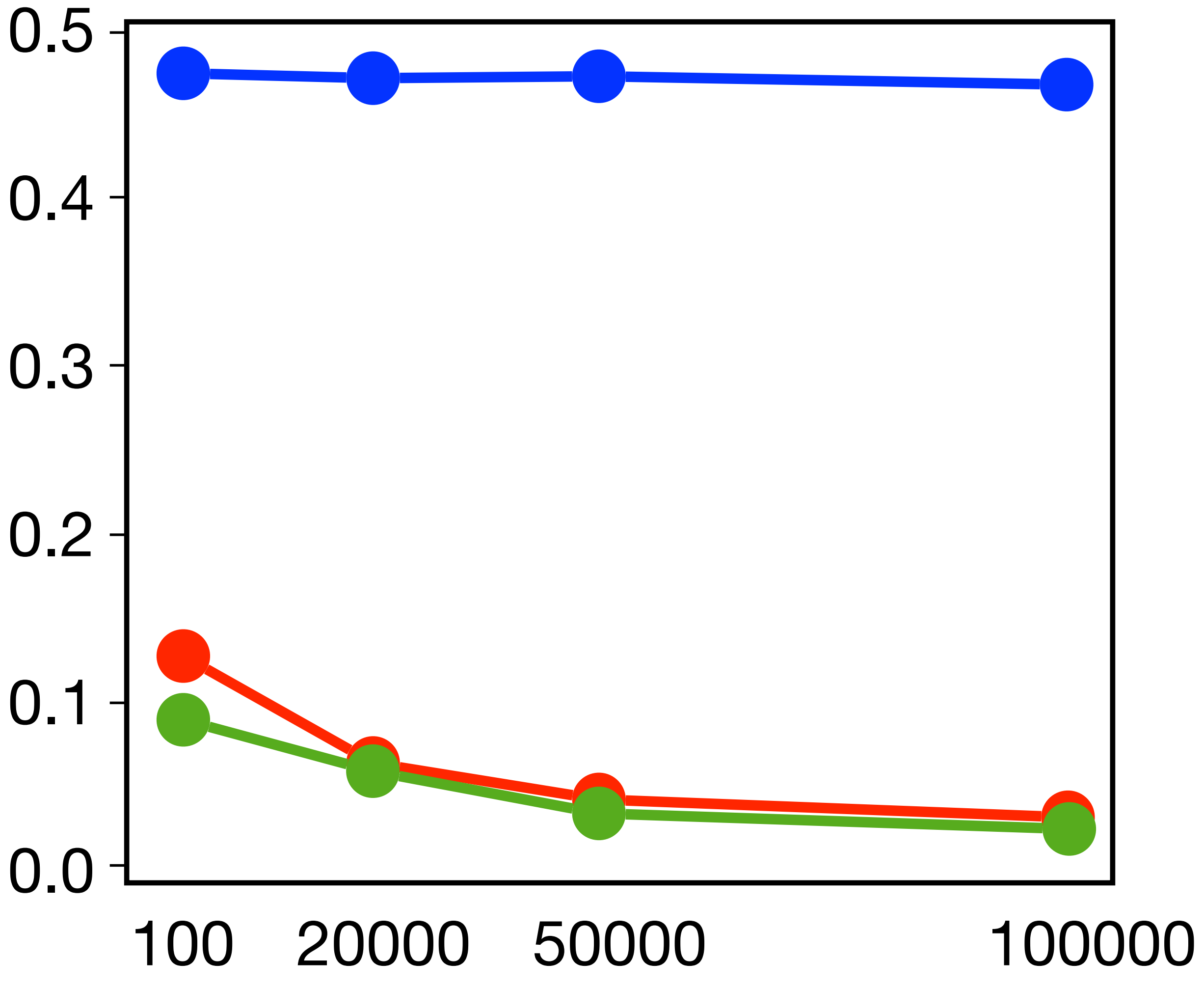
$(\hat{\boldsymbol{\mu}},\hat{\boldsymbol{\pi}}) \rightarrow (\boldsymbol{\mu}_0, \boldsymbol{\pi}_0)$ slowly



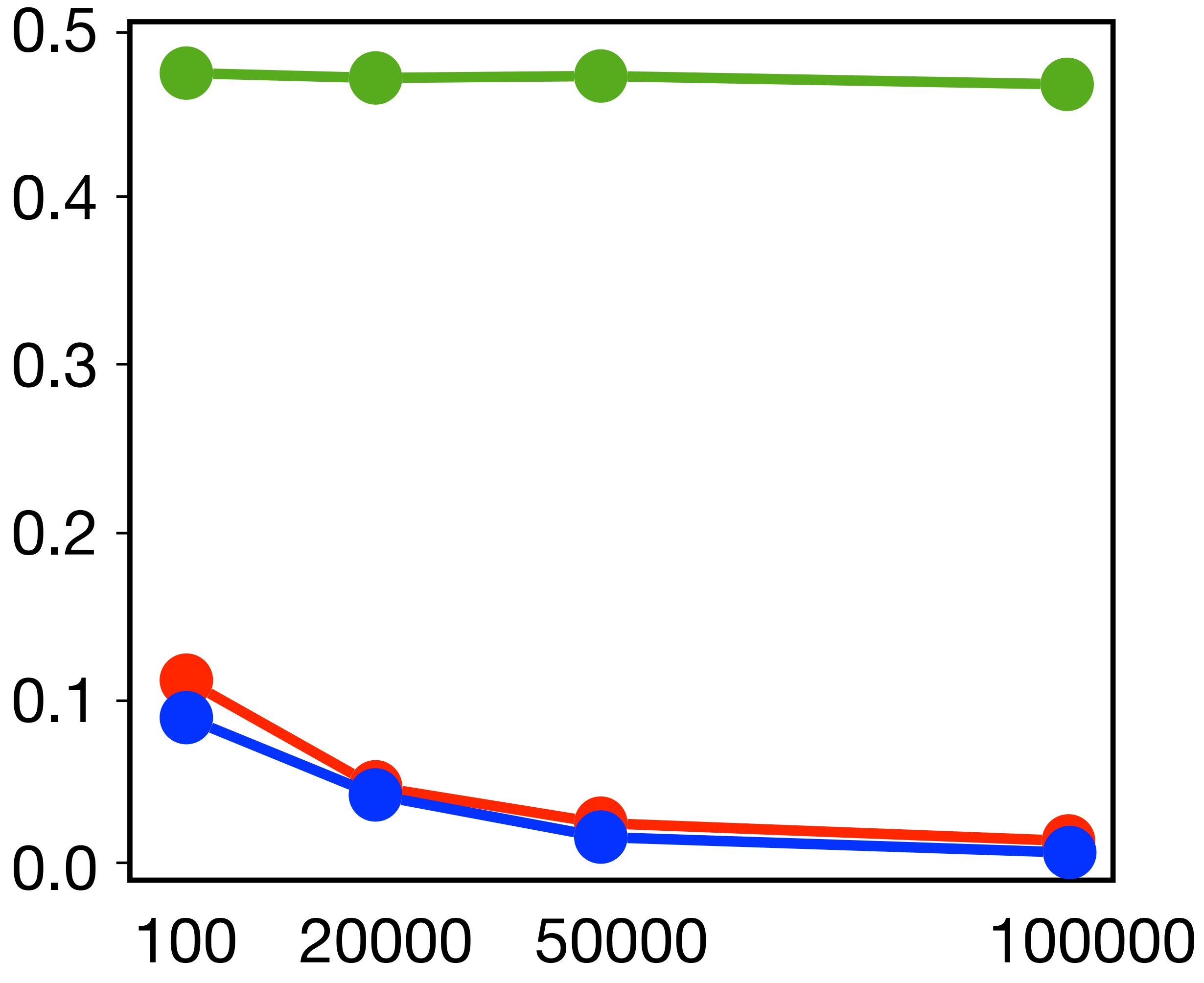
DML-ID converges fast, even when $(\hat{\boldsymbol{\mu}},\hat{\boldsymbol{\pi}})$ converge slowly

Double Robustness

$\hat{\boldsymbol{\mu}}$ misspecified ($\hat{\boldsymbol{\mu}} \neq \boldsymbol{\mu}$)



$\hat{\boldsymbol{\pi}}$ misspecified ($\hat{\boldsymbol{\pi}} \neq \boldsymbol{\pi}$)



DML-ID converges to the true causal effect even when $\hat{\boldsymbol{\mu}}$ or $\hat{\boldsymbol{\pi}}$ are misspecified.

Identification

"When is the causal effect computable from data?"

Estimation

“How do we compute the effect from available data?”

Identification

"When is the causal effect computable from data?”

Connect

Estimation

“How do we compute the effect from available data?”

> “ *Whenever computable from data,*
> *We can compute sample-efficiently.*

Identification

"When is the causal effect computable from data?"

Connect

Estimation

“How do we compute the effect from available data?”

> *Whenever computable from data, We can compute sample-efficiently.*

Econ Professor at MIT, who developed DML

Victor Chernozhukov
@VC31415
Follow

Replying to @YonghanJung @PHuenermund

This is really a fantastic work and is a major contribution.(Incidentally, our DML work was meant to be a service paper to help applied researchers use ML for causal inference, and I don't view our work as a major contribution, certainly not seminal :-) .)

Turing Award winner, pioneer of causal inference

Judea Pearl
@yudapearl
Follow

the do-calculus. The answer, surprisingly and pleasingly is YES. This recent paper causalai.net/r62.pdf shows that EVERY identifiable causal effect can be estimated by a "Weighted Empirical Risk Minimization" method, a fancy name for IPW-like estimation. Worth keeping in mind.

Talk Outline

1. Estimating causal effects from observations

   + its application in healthcare & explainable AI

2. Estimating causal effects from data fusion

3. Unified causal effect estimation method

4. Summary & Future direction

Talk Outline

\+ its application in healthcare & explainable AI

Application 1. Healthcare Science

Application 1. Healthcare Science

RCT

- ➕ Gold standard in causal inference
- ➖ Expensive
- ➖ Selection bias

Application 1. Healthcare Science

RCT	EHR MIMIC-IV, OpenMRS eICU, …
+ Gold standard in causal inference	− Confounding bias
− Expensive	+ Easy to collect
− Selection bias	+ Generalizable

Application 1. Healthcare Science

RCT	EHR MIMIC-IV, OpenMRS eICU, …
⊕ Gold standard in causal inference	⊖ Confounding bias
⊖ Expensive	⊕ Easy to collect
⊖ Selection bias	⊕ Generalizable

Best of Both Worlds

Emulating RCT from EHR

Application 1. Emulating RCT from EHR

Application 1. Emulating RCT from EHR

Input

Effect (Q)
$\mathbb{E}[Y \mid \mathrm{do}(x)]$

EHR
D from P

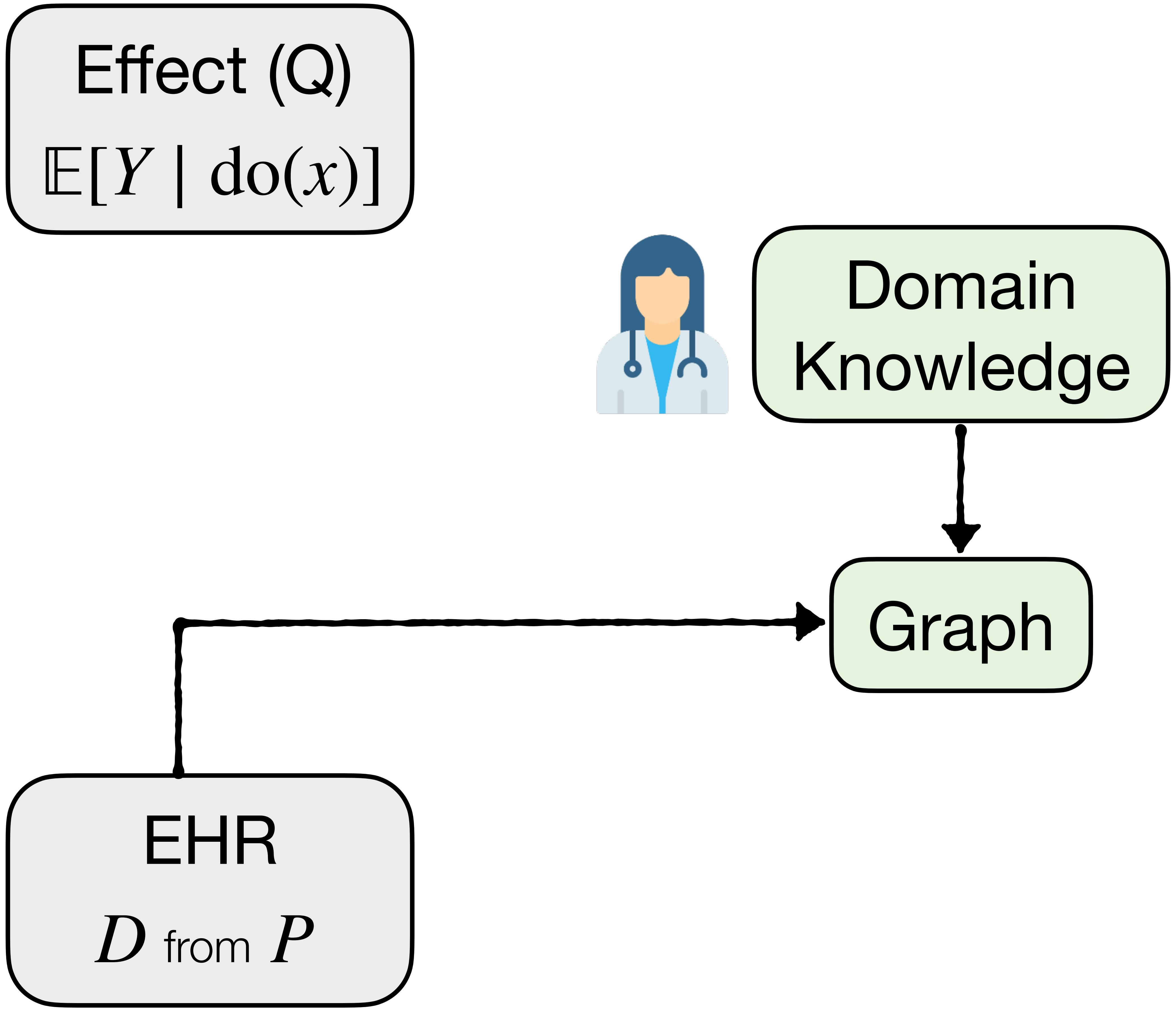
Application 1. Emulating RCT from EHR

Input	Graph Discovery
Effect (Q) $\mathbb{E}[Y \mid \mathrm{do}(x)]$	
EHR D from P	Graph

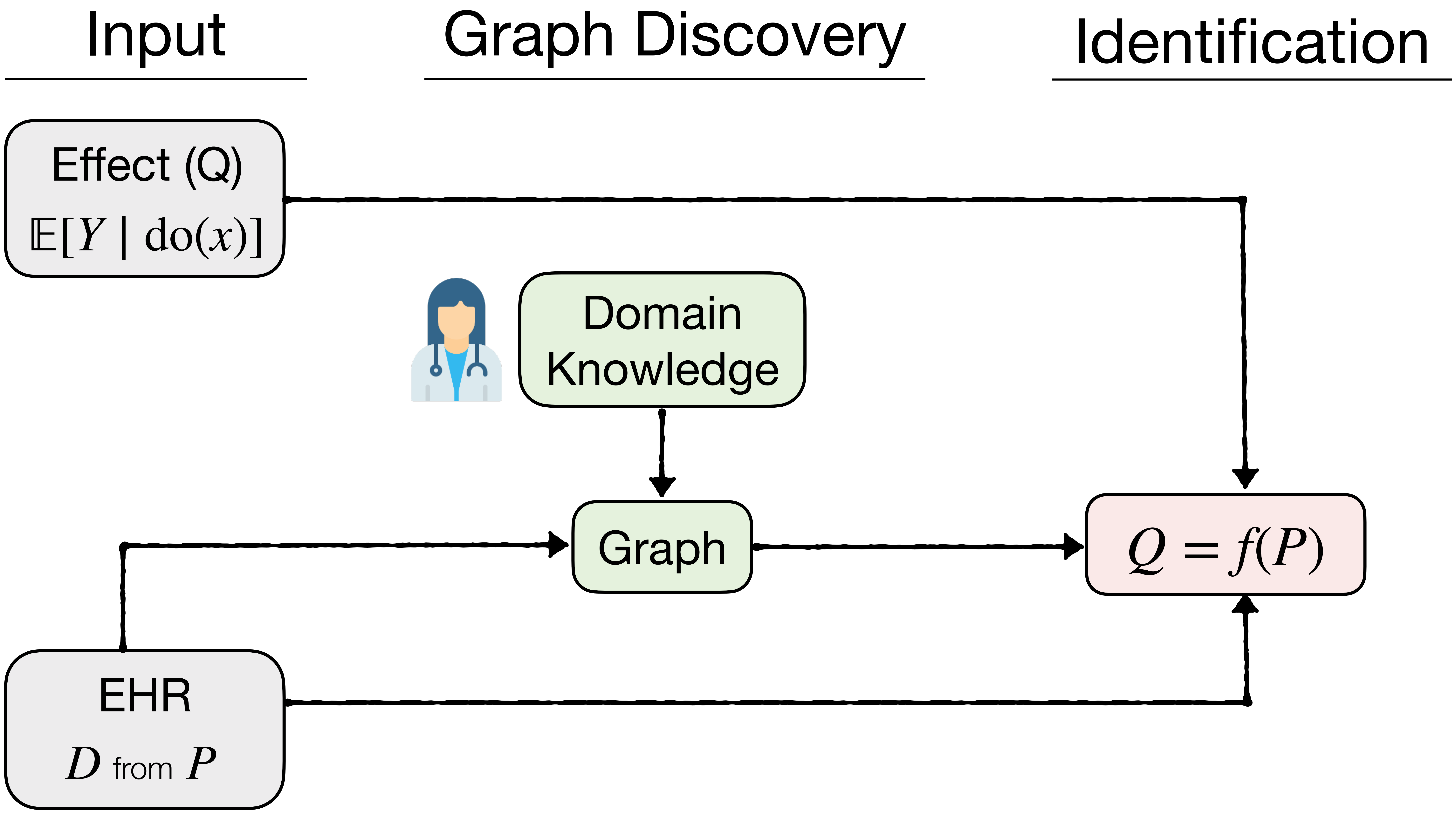
Application 1. Emulating RCT from EHR

Input

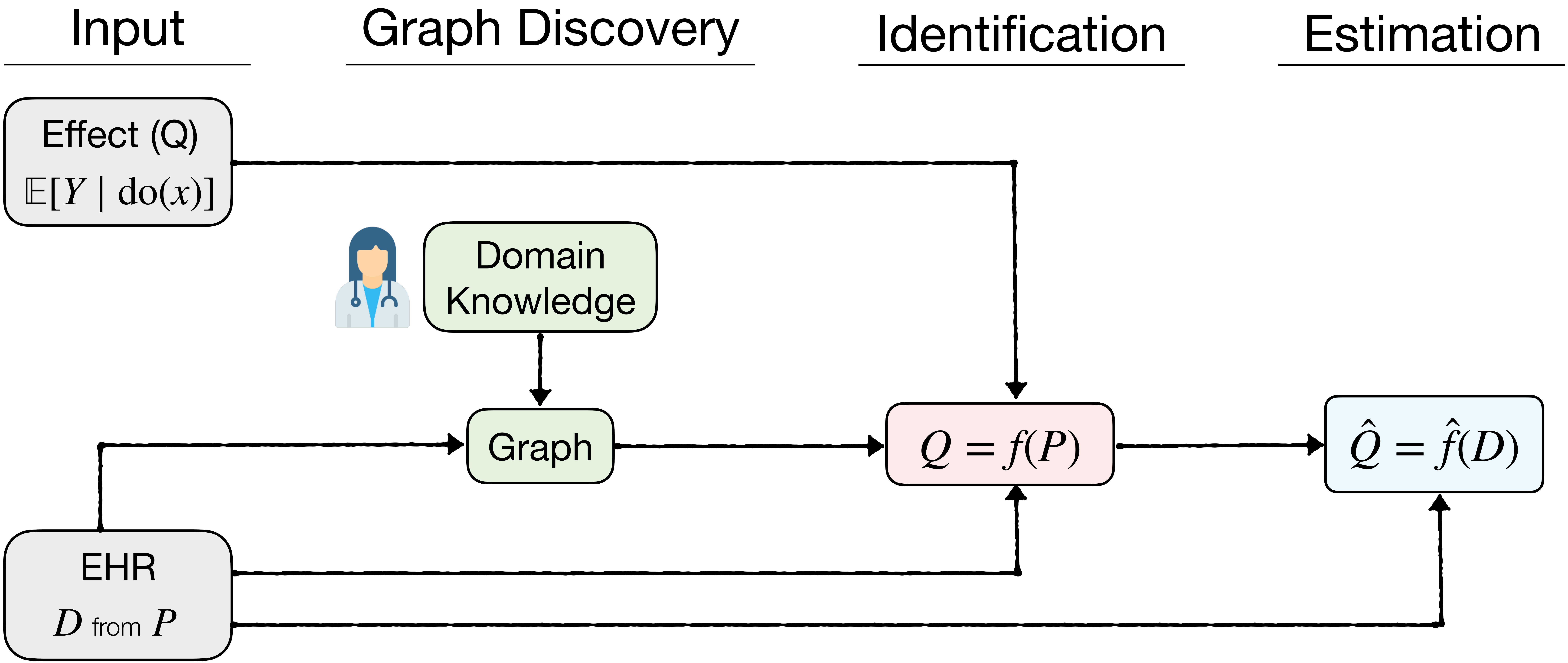
Graph Discovery



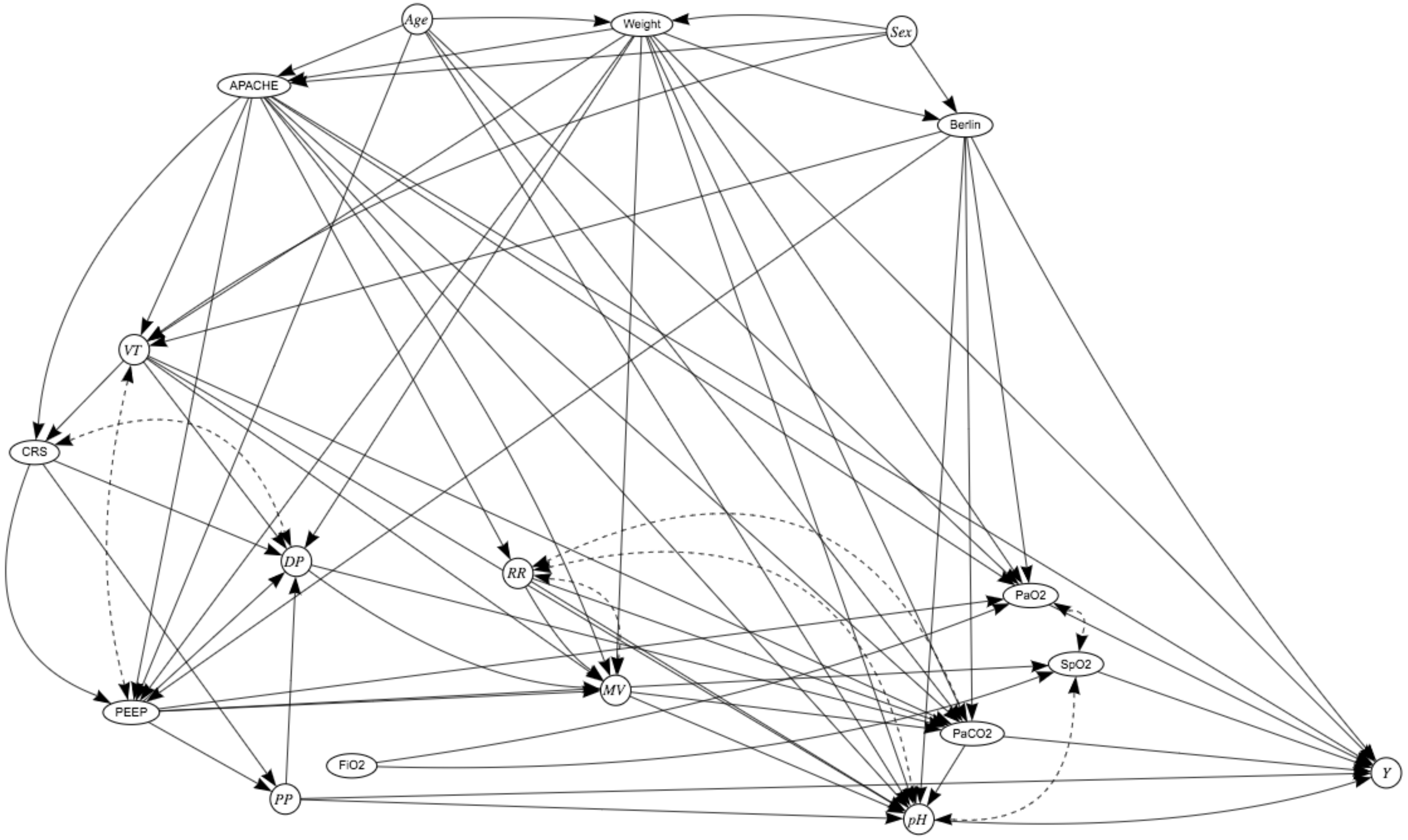
Application 1. Emulating RCT from EHR



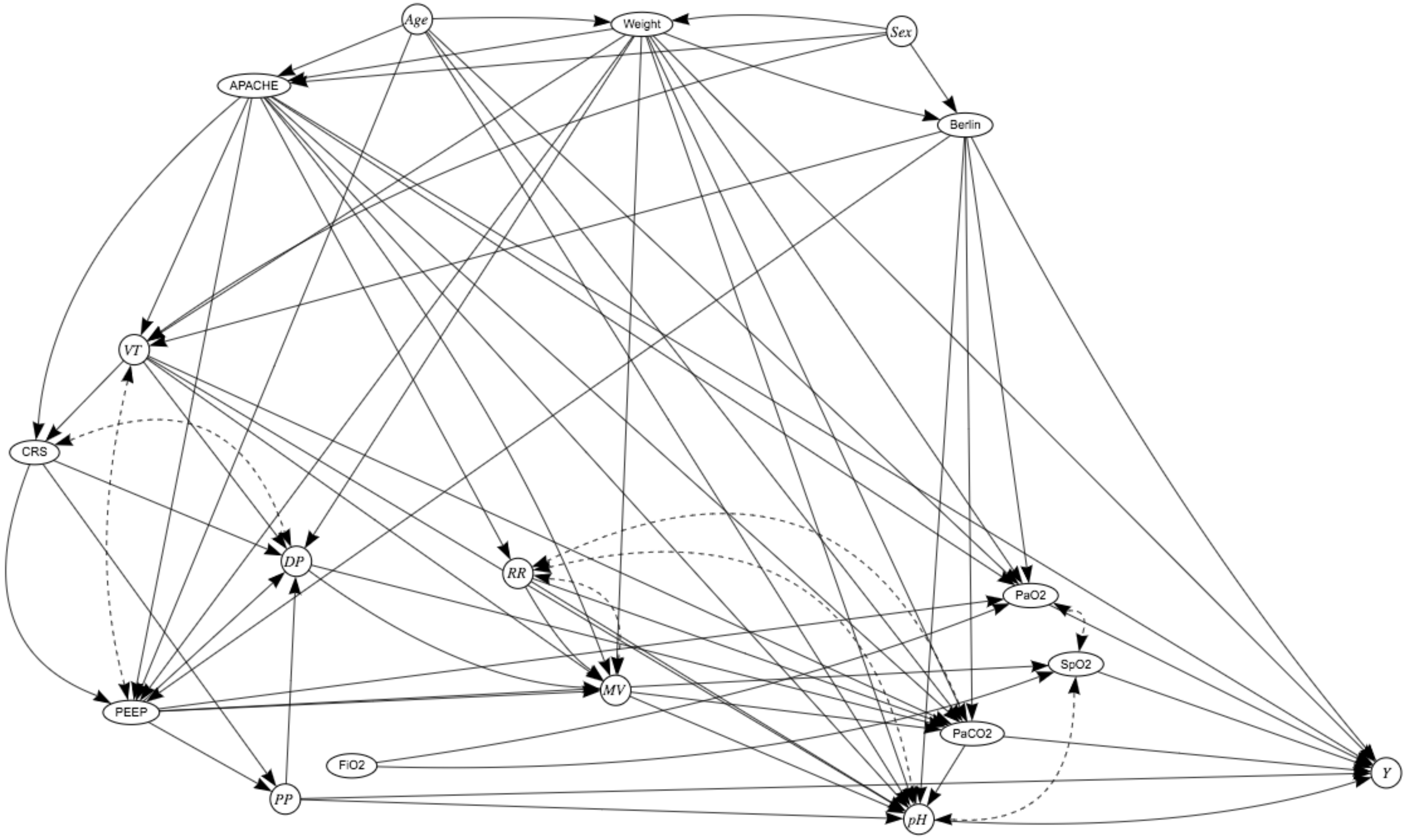
Application 1. Emulating RCT from EHR



Application 1. Emulating RCT from EHR



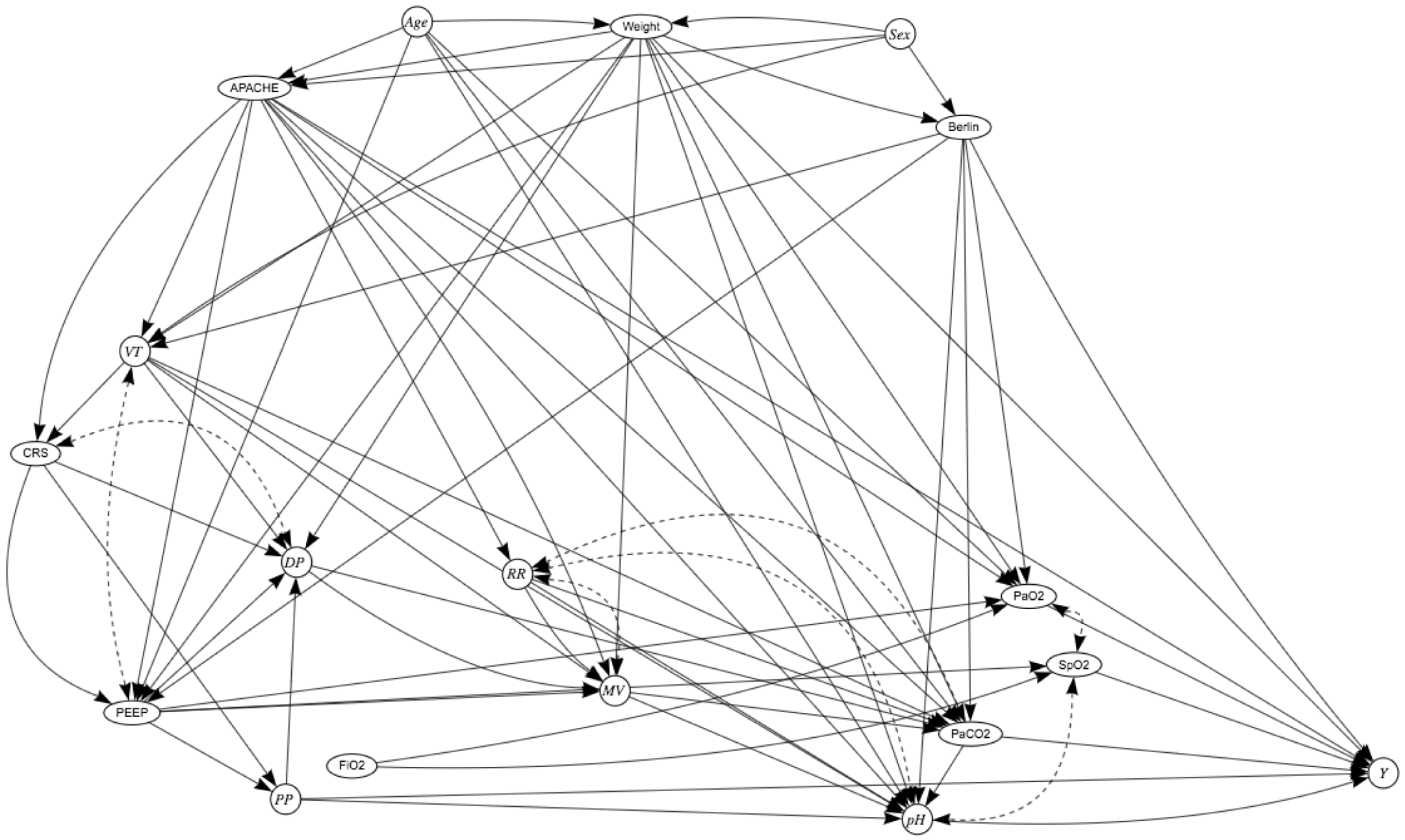
Application 1. Emulating RCT from EHR



Causal graph on Acute Respiratory Distress Syndrome (ARDS)

Application 1. Emulating RCT from EHR

Jung et al., American Thoracic Society, 2018



Causal graph on Acute Respiratory Distress Syndrome (ARDS)

Result

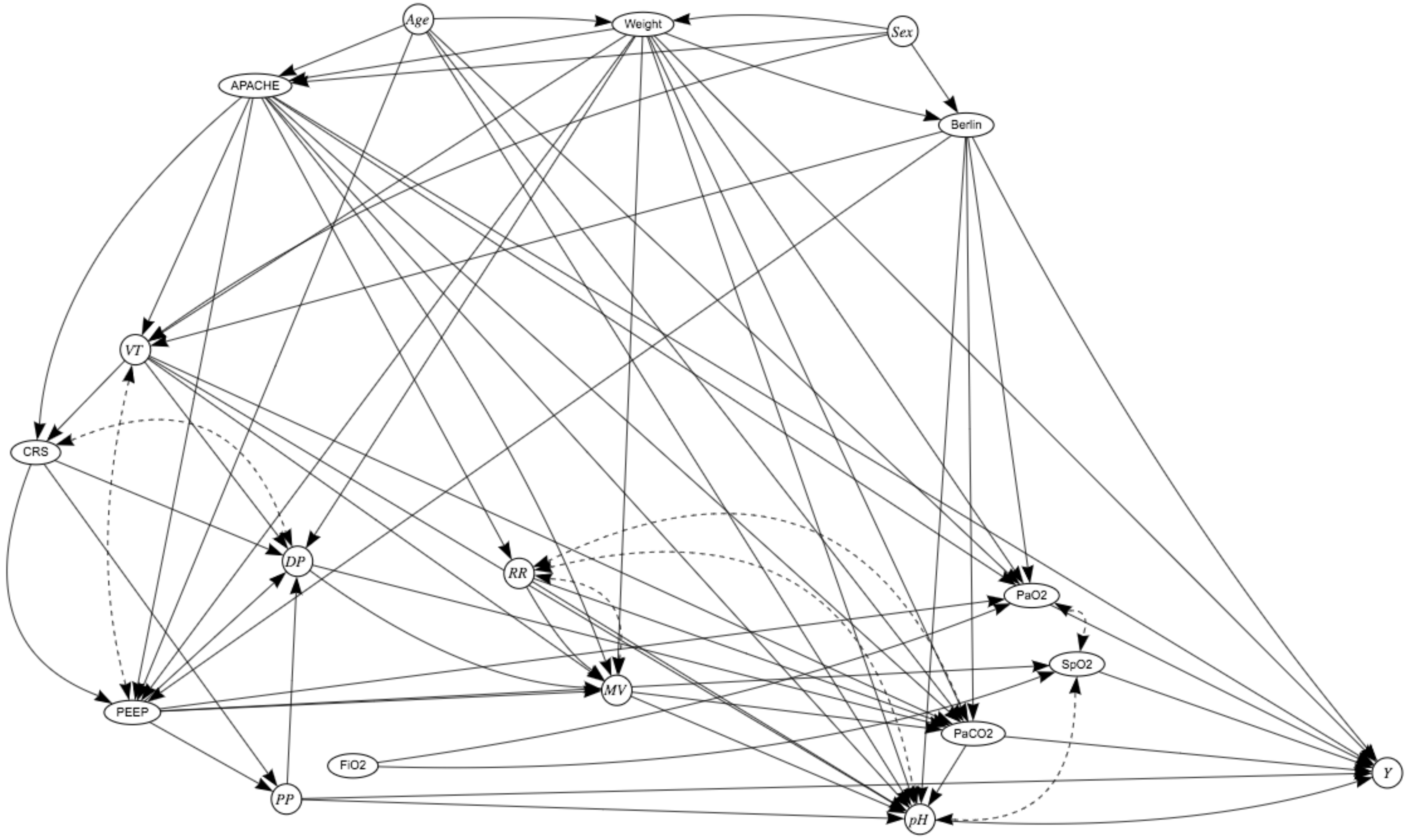
For seminal RCTs,

Our treatment recommendation

= Trials' treatment recommendation

Application 1. Emulating RCT from EHR

Jung et al., American Thoracic Society, 2018



Causal graph on Acute Respiratory Distress Syndrome (ARDS)

Result

For seminal RCTs,

Our treatment recommendation

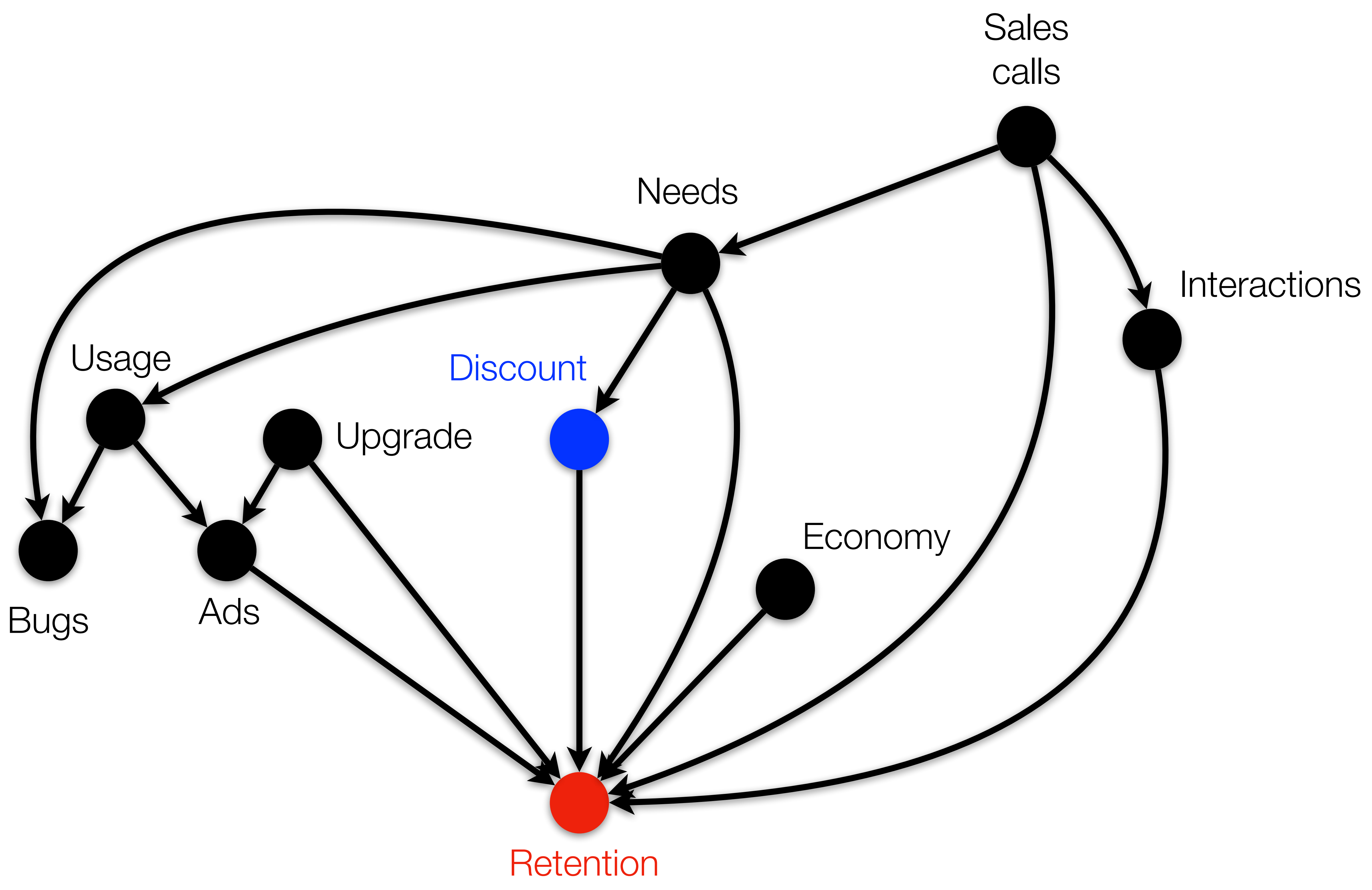
= Trials' treatment recommendation

Impact

Our method can be used to construct an initial hypothesis before conducting trials.

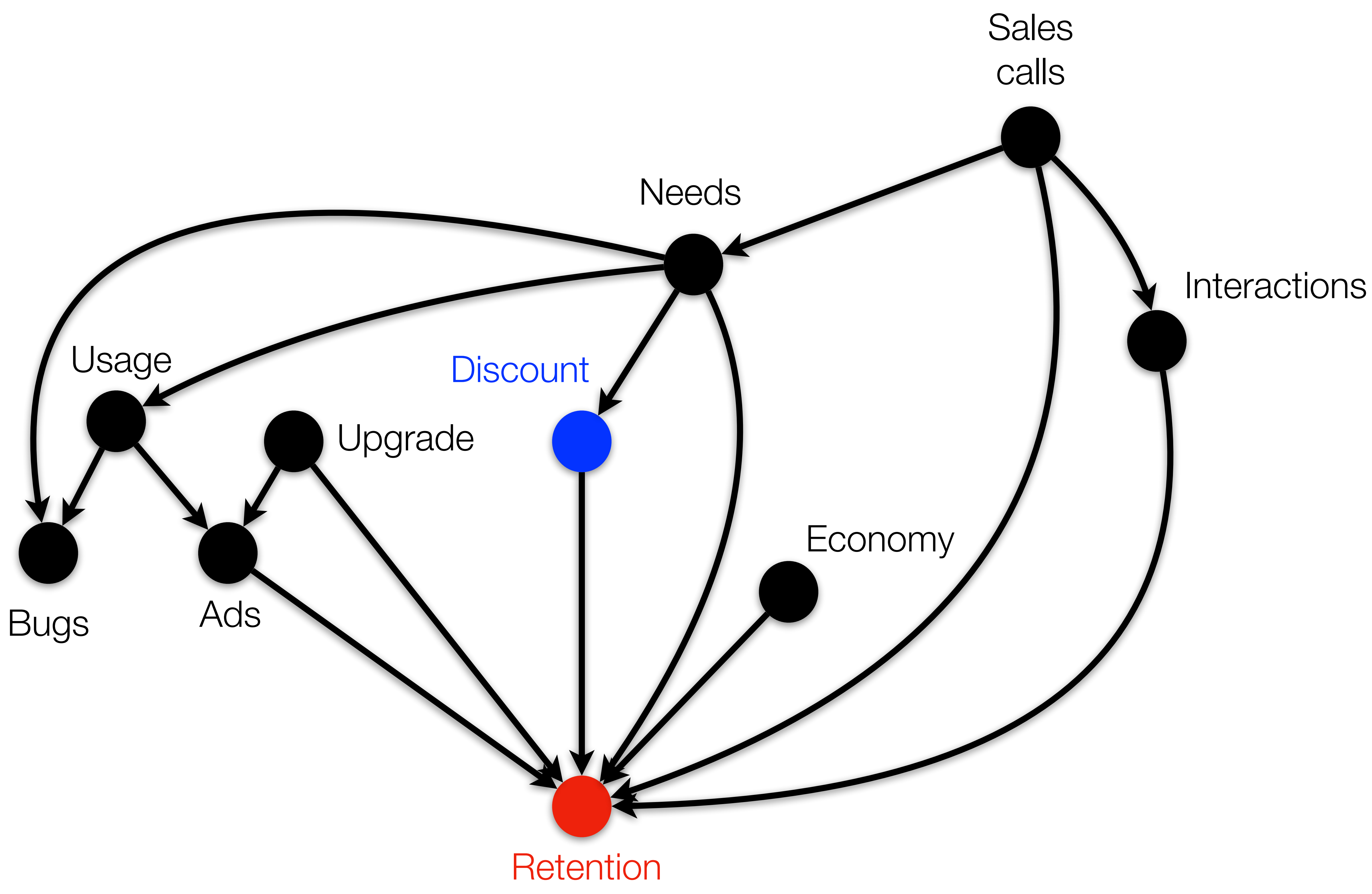
Application 2. Explainable AI

Application 2. Explainable AI



Contribution of Discount to the Retention?

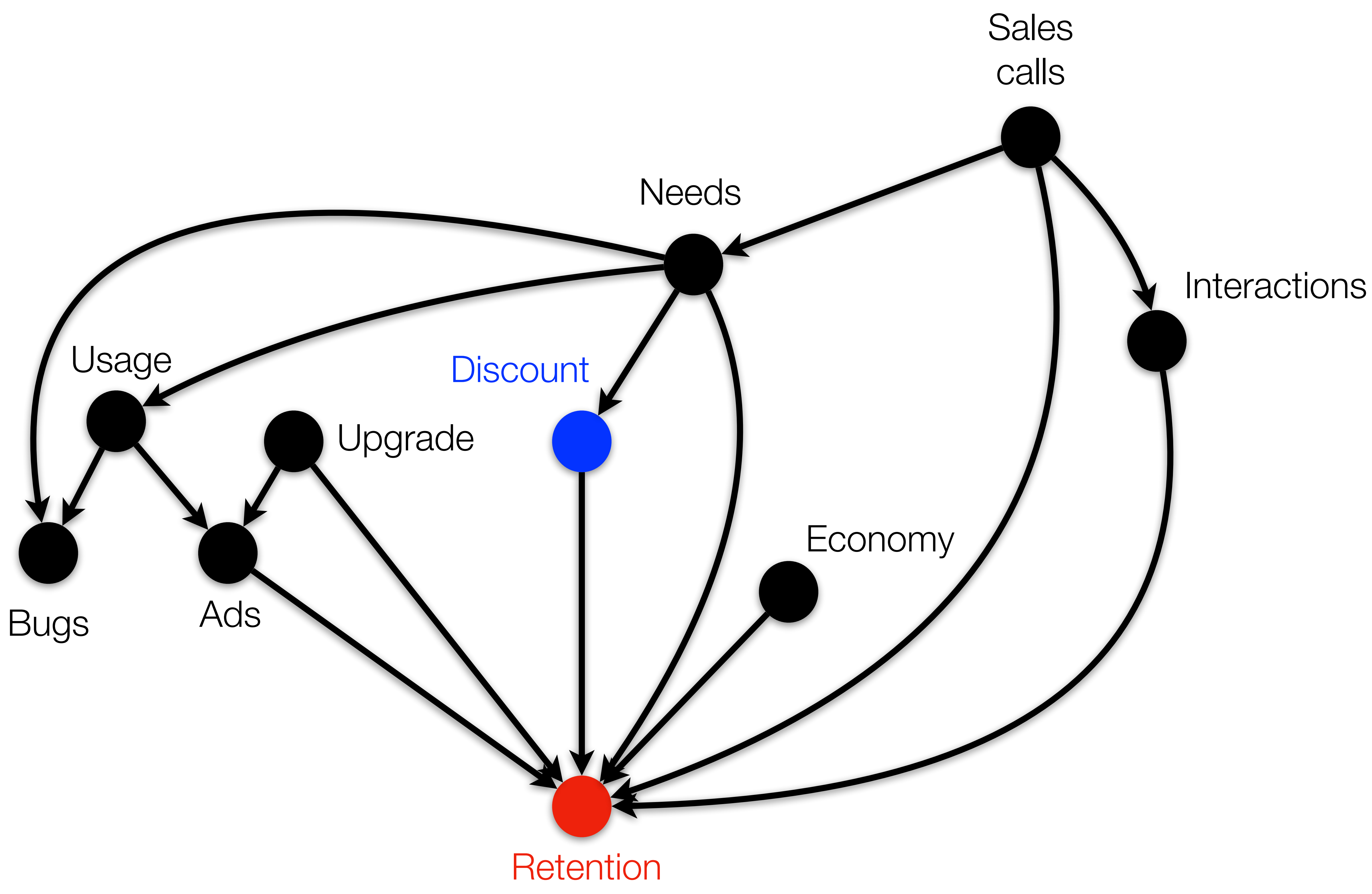
Application 2. Explainable AI



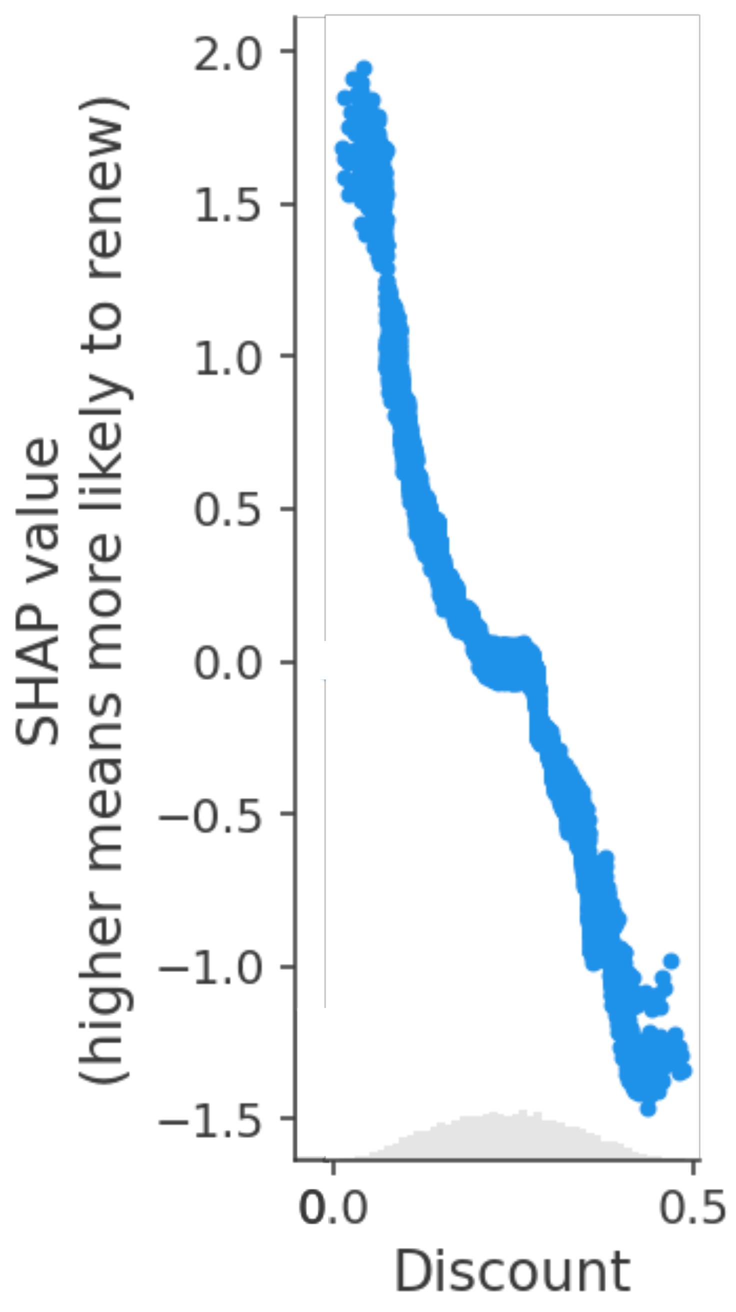
Contribution of Discount to the Retention?

- SHAP value: one of the most cited measure for the feature importance

Application 2. Explainable AI

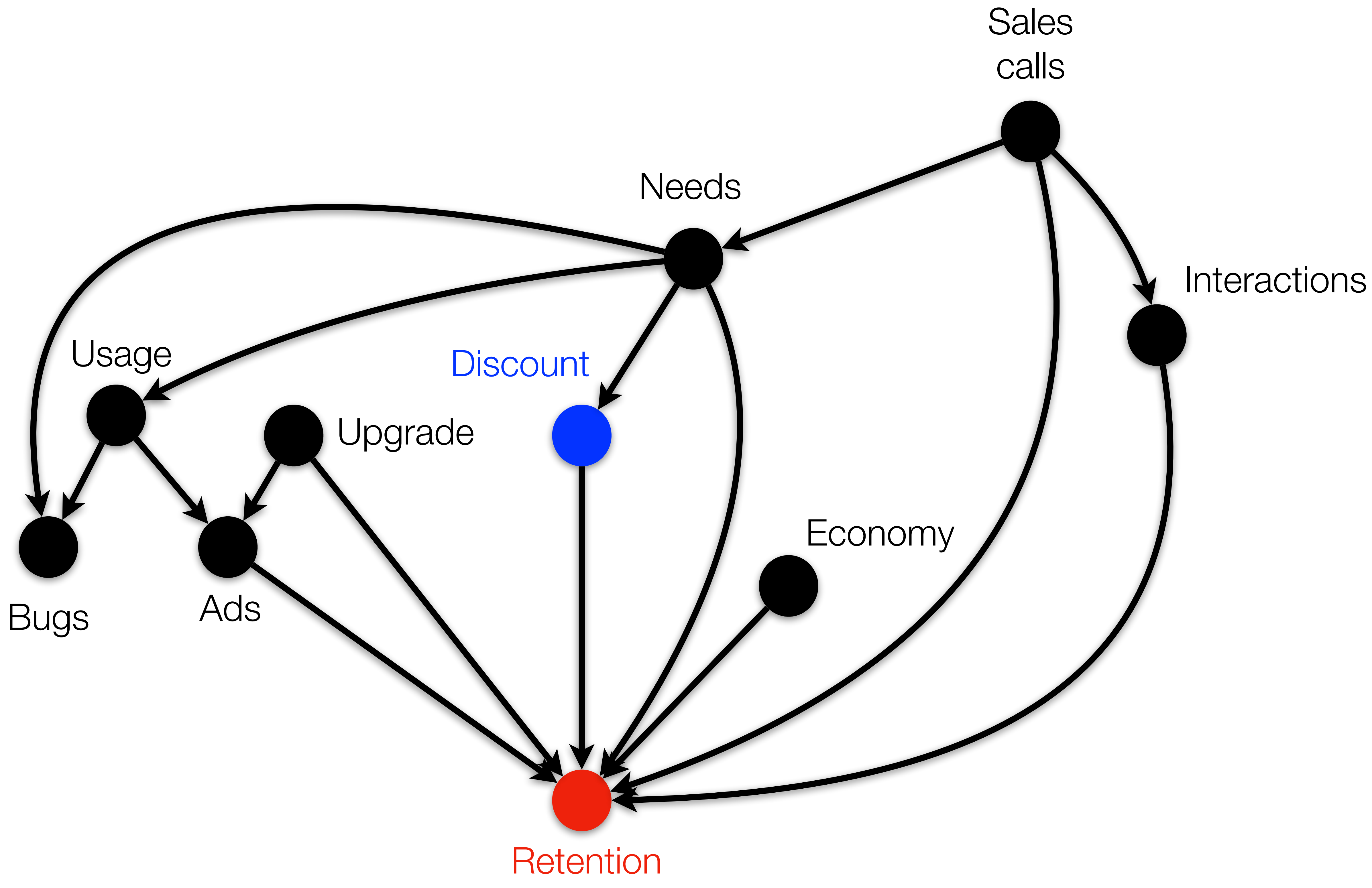


Contribution of Discount to the Retention?

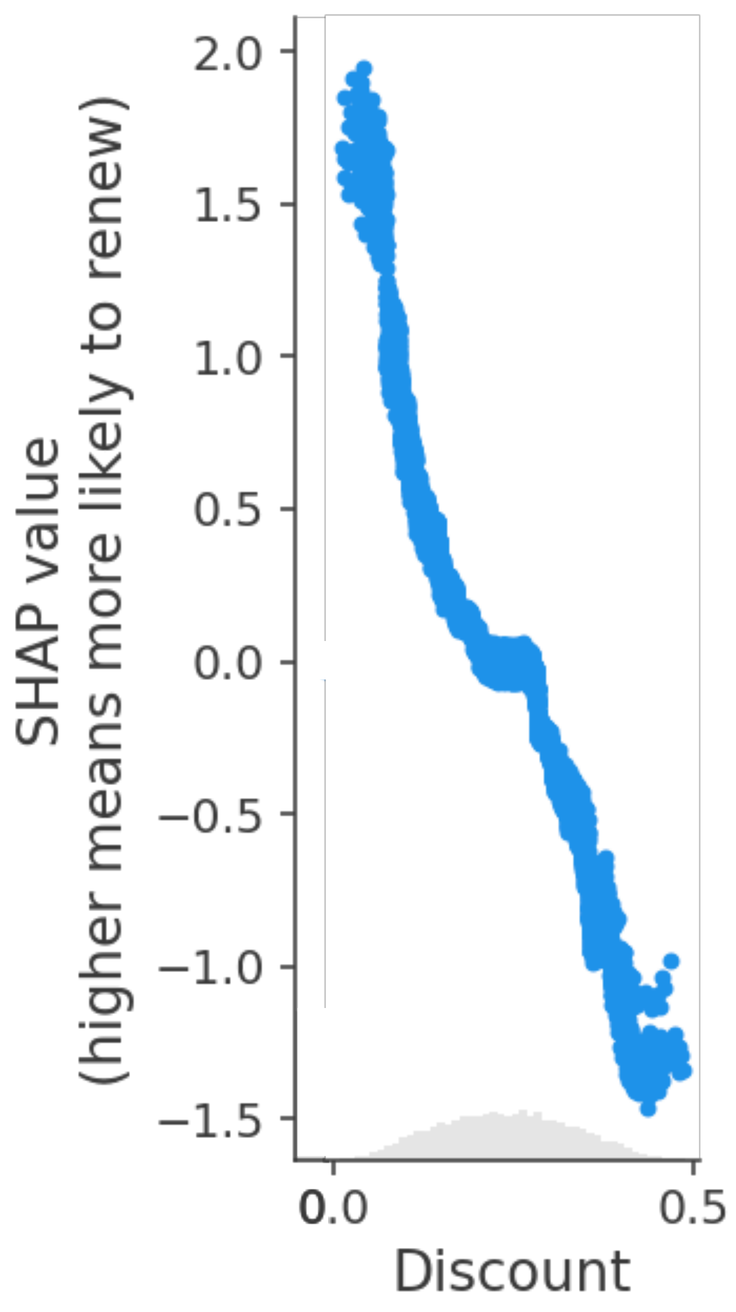


- SHAP value: one of the most cited measure for the feature importance

Application 2. Explainable AI

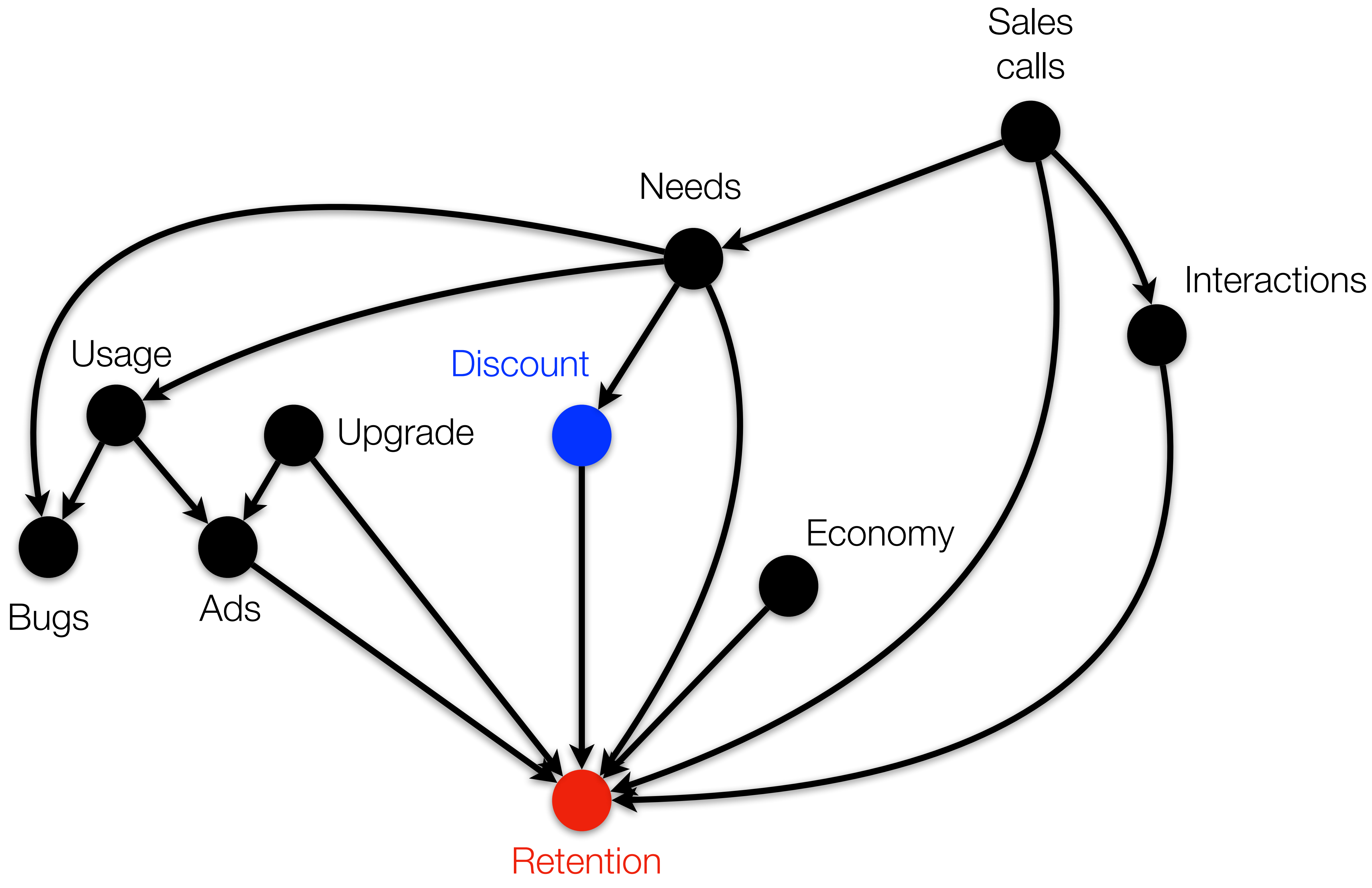


Contribution of Discount to the Retention?

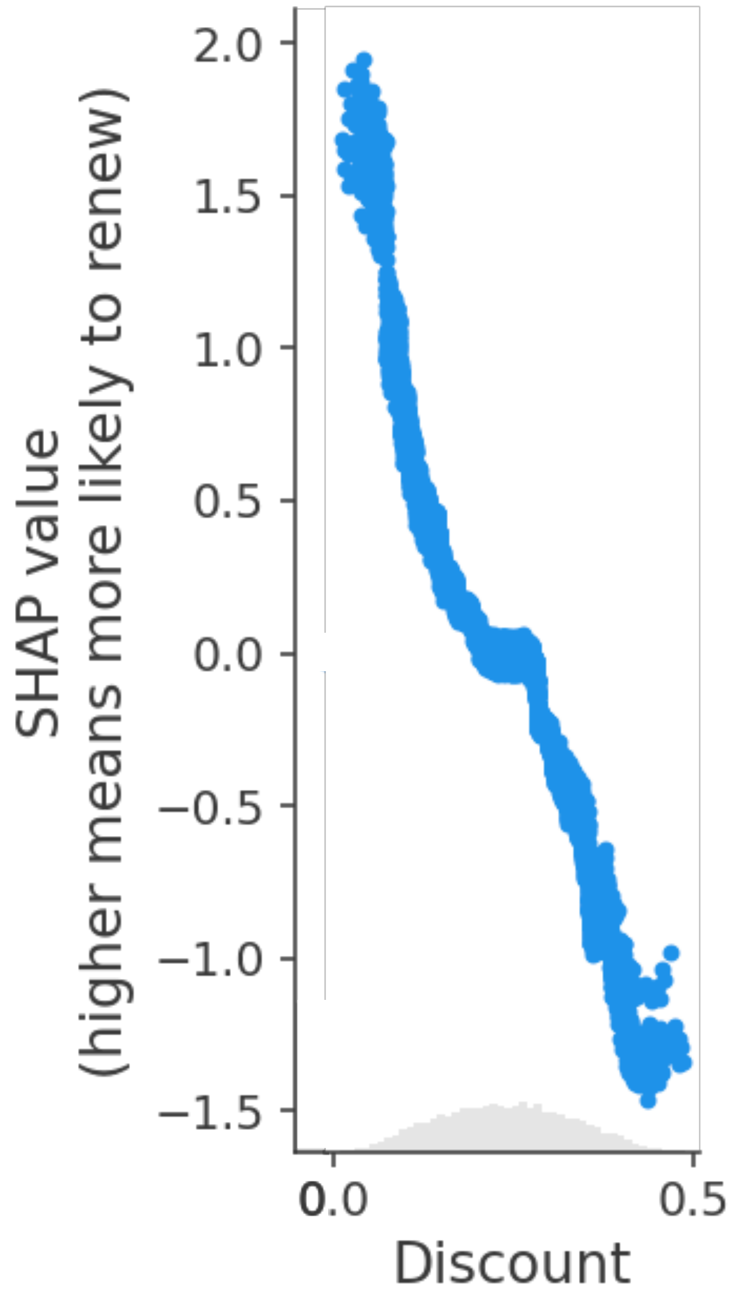


- SHAP value: one of the most cited measure for the feature importance
- *Larger* discounts contribute *less* to retention?

Application 2. Explainable AI

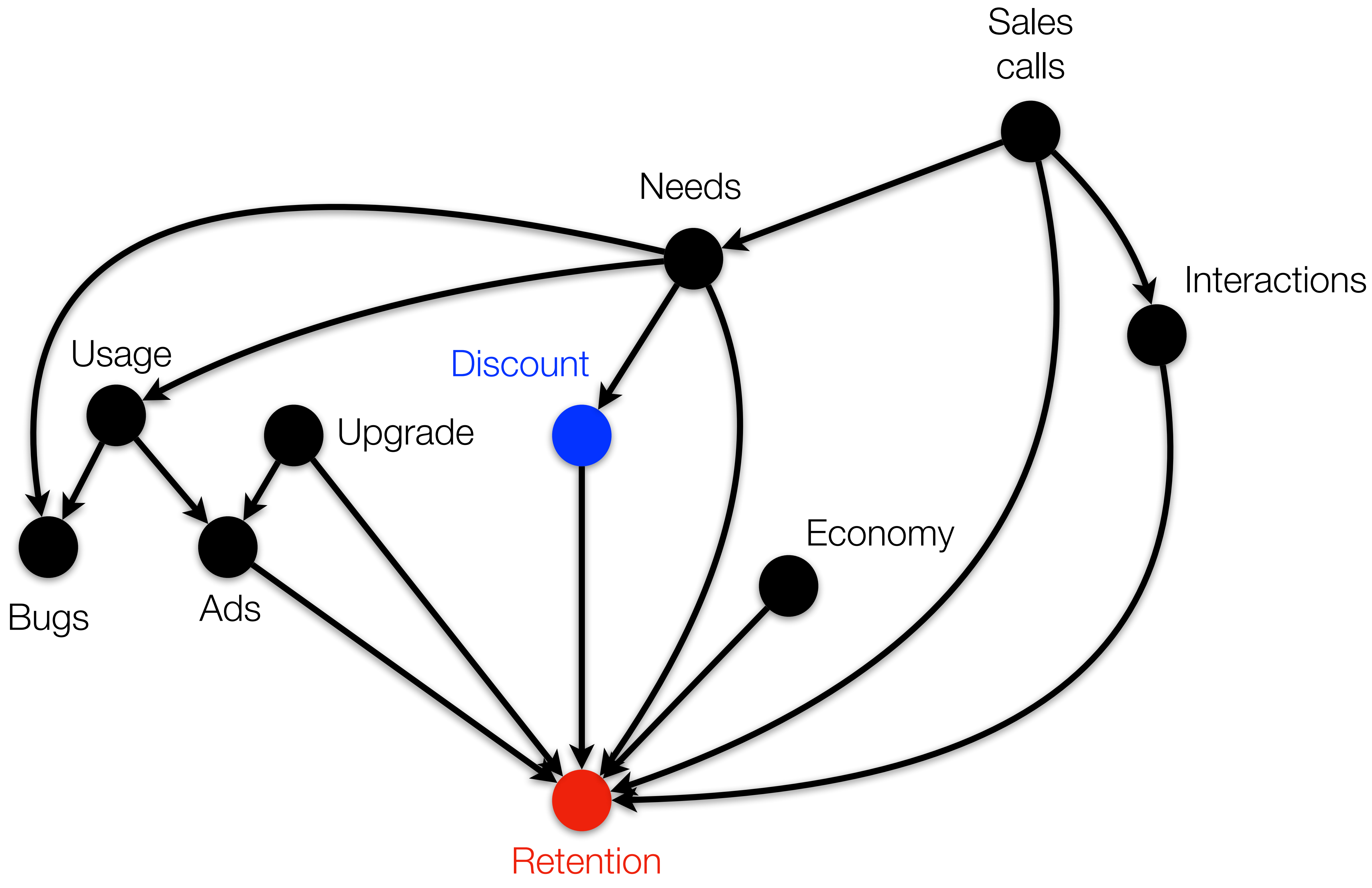


Contribution of Discount to the Retention?

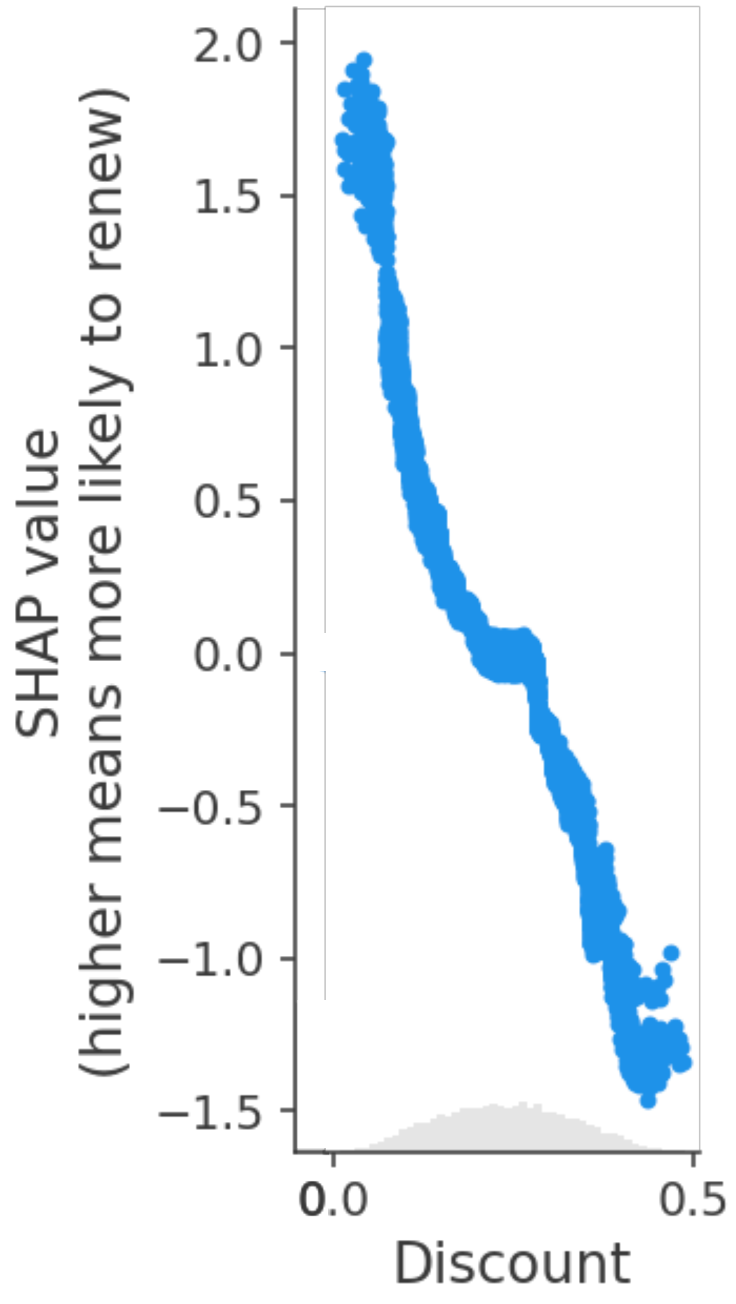


- SHAP value: one of the most cited measure for the feature importance
- *Larger* discounts contribute *less* to retention?
- Mismatch with human intuition is due to computing the importance based on correlation (e.g. $\mathbb{E}[\text{retention}|\text{discount}]$)

Application 2. Explainable AI



Contribution of Discount to the Retention?



- SHAP value: one of the most cited measure for the feature importance
- *Larger* discounts contribute *less* to retention?
- Mismatch with human intuition is due to computing the importance based on correlation (e.g. $\mathbb{E}[\text{retention}|\text{discount}]$)

Causality-based feature importance measure is essential

do-Shapley: Causality-based Feature Attribution

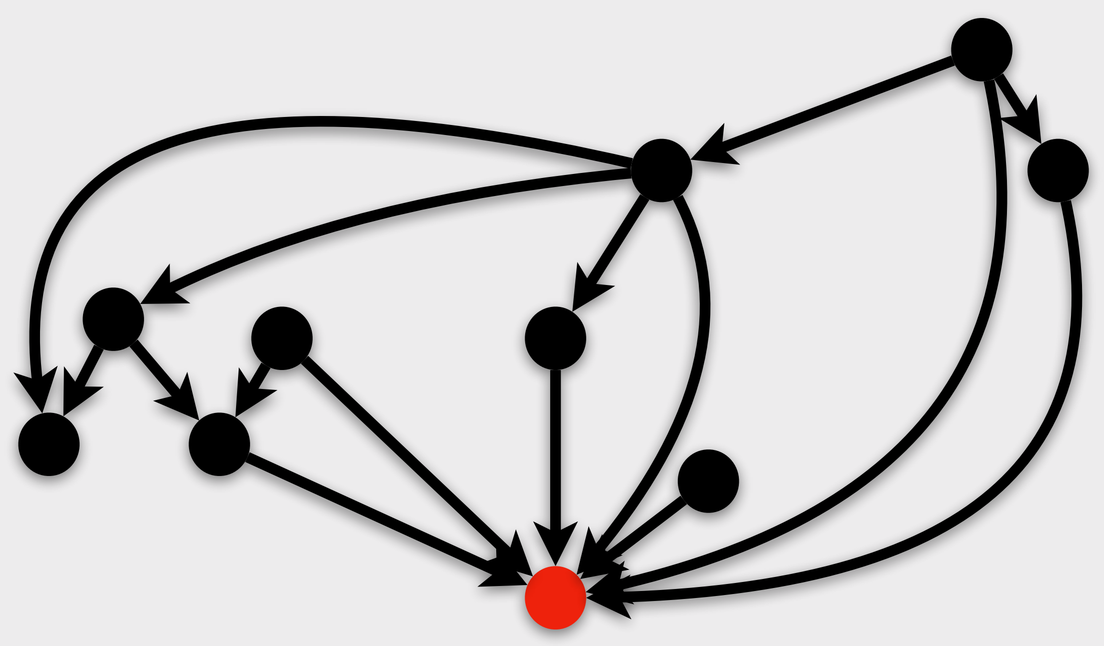
Jung et al.,ICML 2022

do-Shapley: Causality-based Feature Attribution

Jung et al.,ICML 2022

Input

Graph



Data

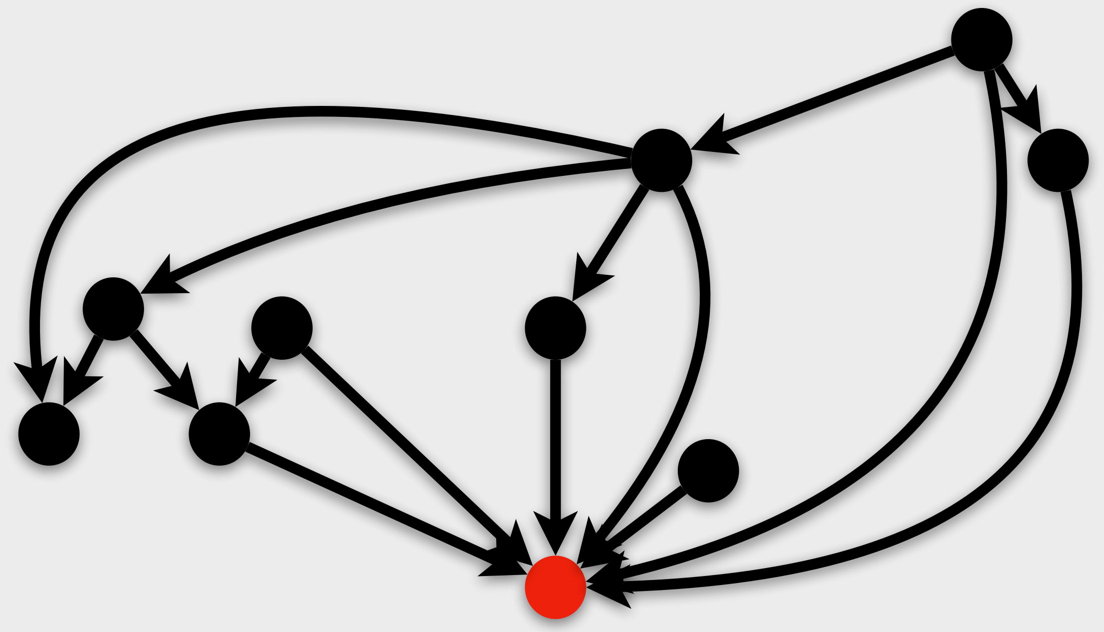
- Input: $(X_1, \cdots, X_m)$
- output: $f(X_1, \cdots, X_m)$

do-Shapley: Causality-based Feature Attribution

Jung et al.,ICML 2022

Input

Graph



Data

- Input: $(X_1, \cdots, X_m)$
- output: $f(X_1, \cdots, X_m)$

Attribution

do-Shapley

$(\phi_1, \cdots, \phi_m)$

Causality-based feature attribution

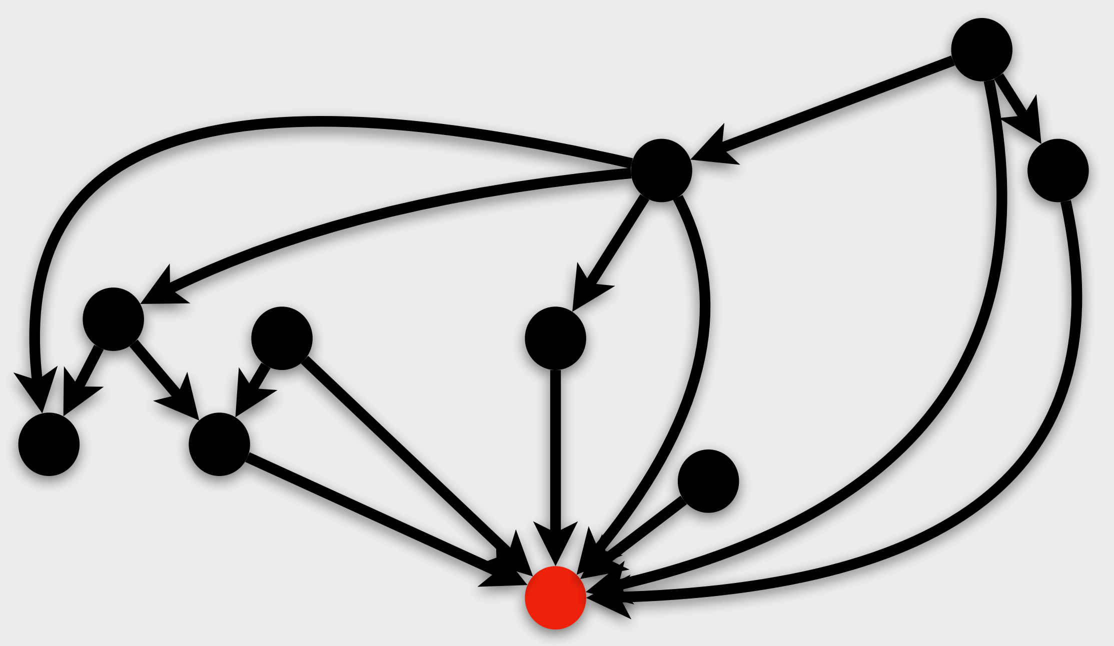
do-Shapley: Causality-based Feature Attribution

Jung et al.,ICML 2022

Input

Attribution

Graph



Data

- Input: $(X_1, \cdots, X_m)$
- output: $f(X_1, \cdots, X_m)$

do-Shapley

$(\phi_1, \cdots, \phi_m)$

Causality-based feature attribution

$$\phi_i = \frac{1}{n} \sum_{S \subseteq [n] \setminus \{i\}} \binom{n-1}{|S|}^{-1} \{\mathbb{E}[Y \mid do(\mathbf{x}_S, x_i)] - \mathbb{E}[Y \mid do(\mathbf{x}_S)]\}$$

do-Shapley: Causality-based Feature Attribution

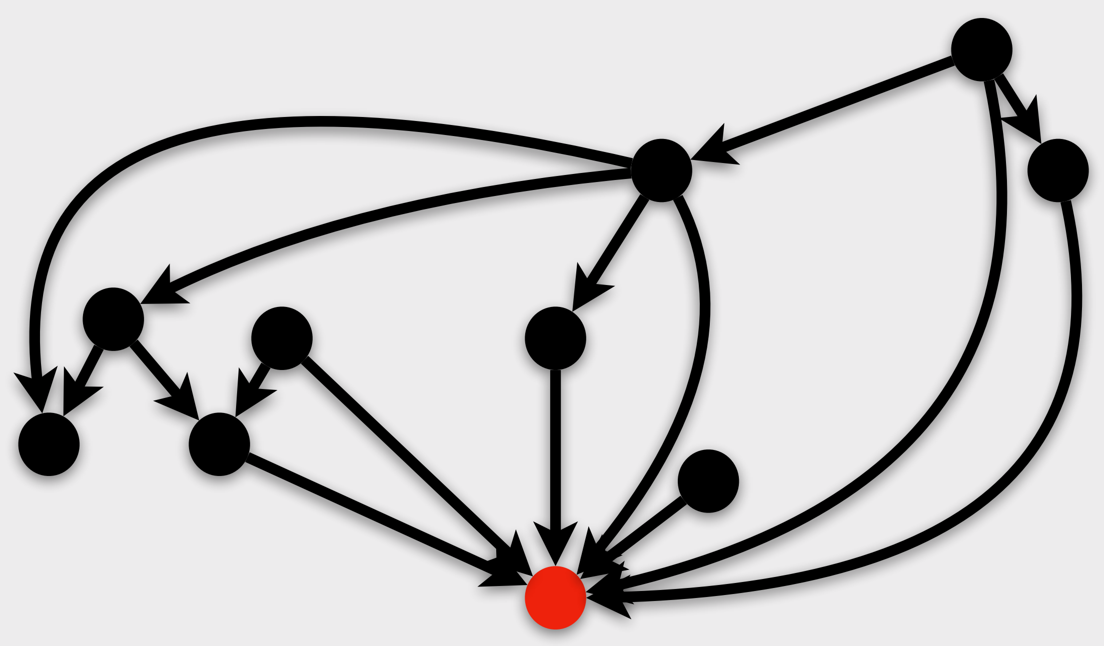
Jung et al.,ICML 2022

Input

Attribution

Identification

Graph



do-Shapley

$(\phi_1, \cdots, \phi_m)$

Causality-based feature attribution

$\phi_i = f(P)$

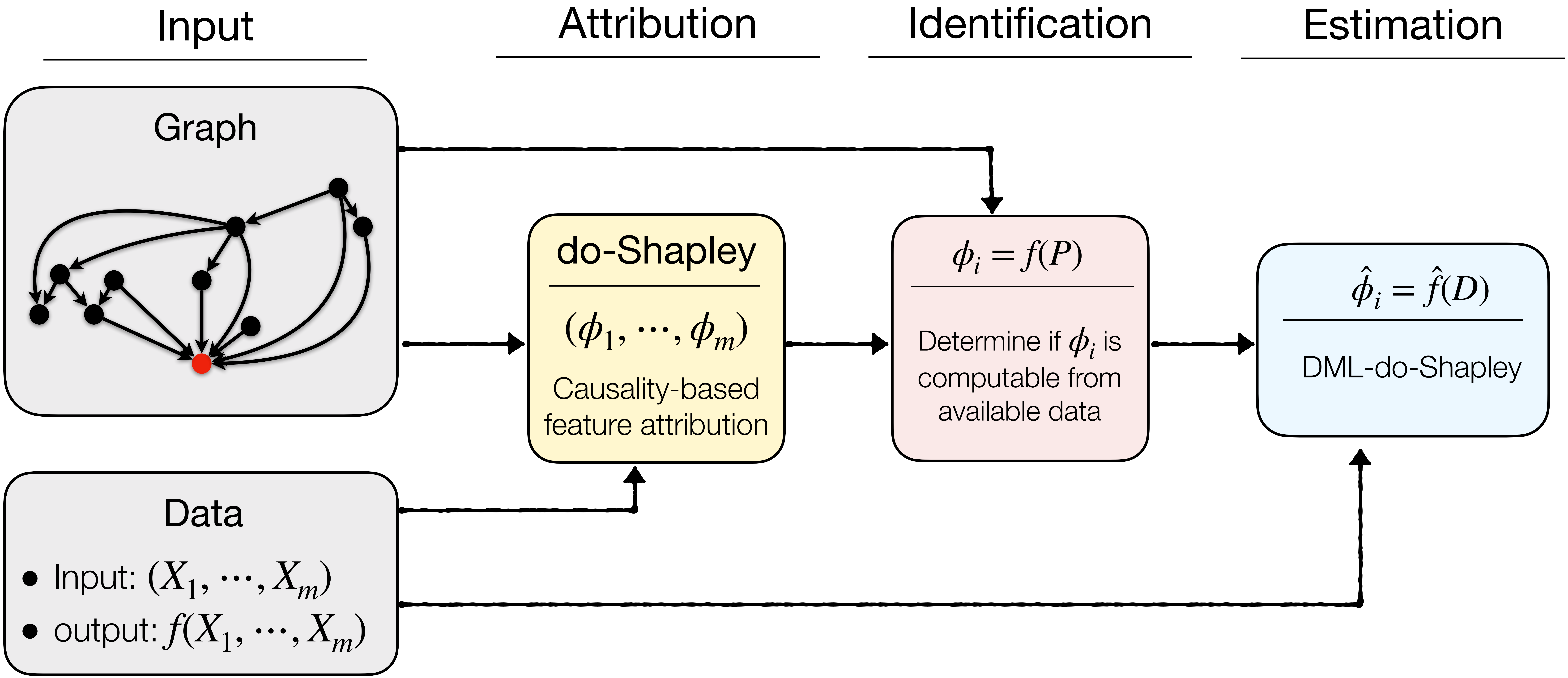
Determine if ϕ_i is computable from available data

Data

- Input: $(X_1, \cdots, X_m)$
- output: $f(X_1, \cdots, X_m)$

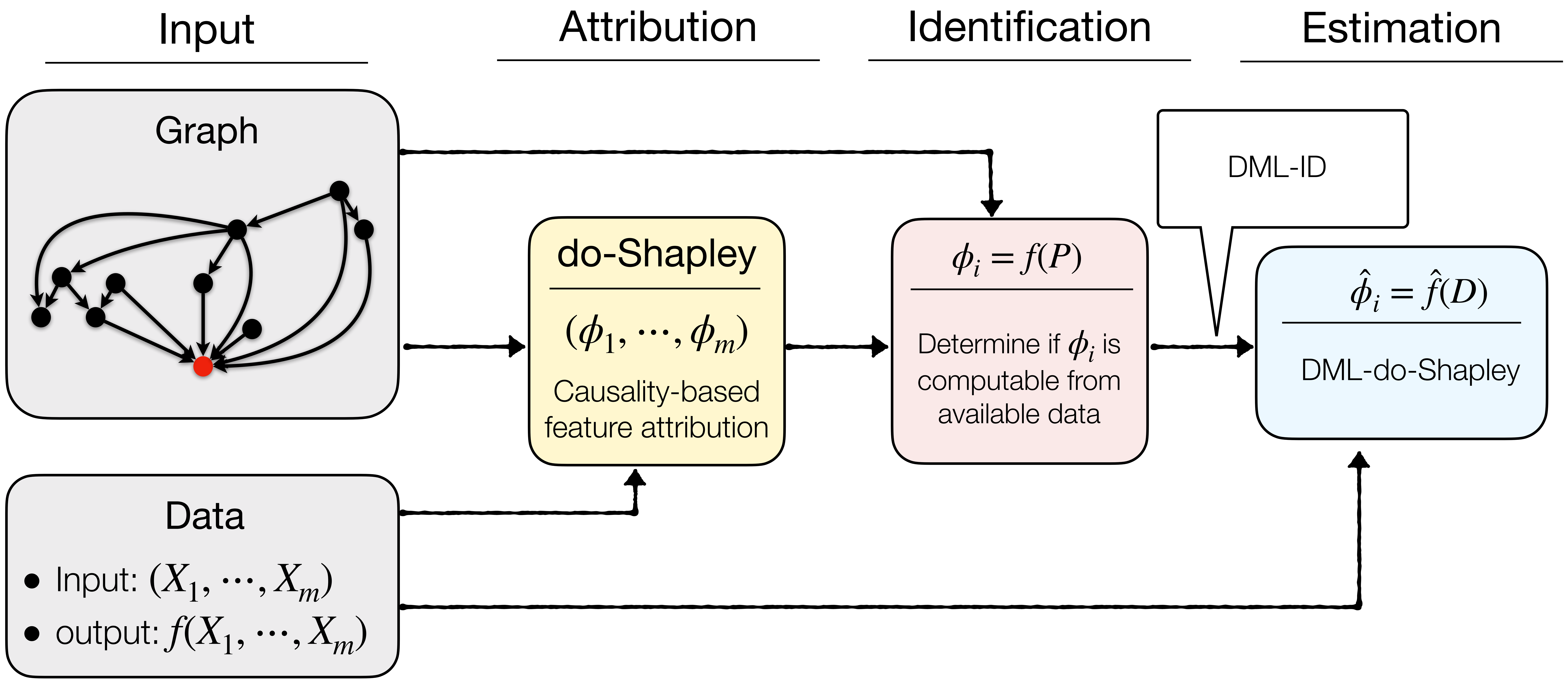
do-Shapley: Causality-based Feature Attribution

Jung et al.,ICML 2022



do-Shapley: Causality-based Feature Attribution

Jung et al.,ICML 2022



Simulation: Better Interpretability

Estimator	Rank Correlation with True Importances	Implication
DML-do-Shapley	**1.0**	
SHAP	**-0.28**	

Simulation: Better Interpretability

Estimator	Rank Correlation with True Importances	Implication
DML-do-Shapley	**1.0**	Estimated feature importance ranking = True ranking of feature importance
SHAP	**-0.28**	

Simulation: Better Interpretability

Estimator	Rank Correlation with True Importances	Implication
DML-do-Shapley	**1.0**	Estimated feature importance ranking = True ranking of feature importance
SHAP	**-0.28**	High true importance ranking = Low estimated ranks

Impact on Explainable AI

Impact on Explainable AI

Unique causality-based feature importance measure that aligns with human intuition:

Impact on Explainable AI

Unique causality-based feature importance measure that aligns with human intuition:

- Two features receive equal contributions whenever their causal effects are the same.

Impact on Explainable AI

Unique causality-based feature importance measure that aligns with human intuition:

- Two features receive equal contributions whenever their causal effects are the same.
- Feature's contribution = 0 if it has no causal effect

Impact on Explainable AI

Unique causality-based feature importance measure that aligns with human intuition:

- Two features receive equal contributions whenever their causal effects are the same.
- Feature's contribution = 0 if it has no causal effect
- Feature contributions closely approximate their causal effects on the outcome

Impact on Explainable AI

Unique causality-based feature importance measure that aligns with human intuition:

- Two features receive equal contributions whenever their causal effects are the same.
- Feature's contribution = 0 if it has no causal effect
- Feature contributions closely approximate their causal effects on the outcome
- The sum of feature contributions = The outcome $f(X_1, \cdots, X_m)$

Talk Outline

1. Estimating causal effects from observations

   + its application in healthcare & explainable AI

2. Estimating causal effects from data fusion

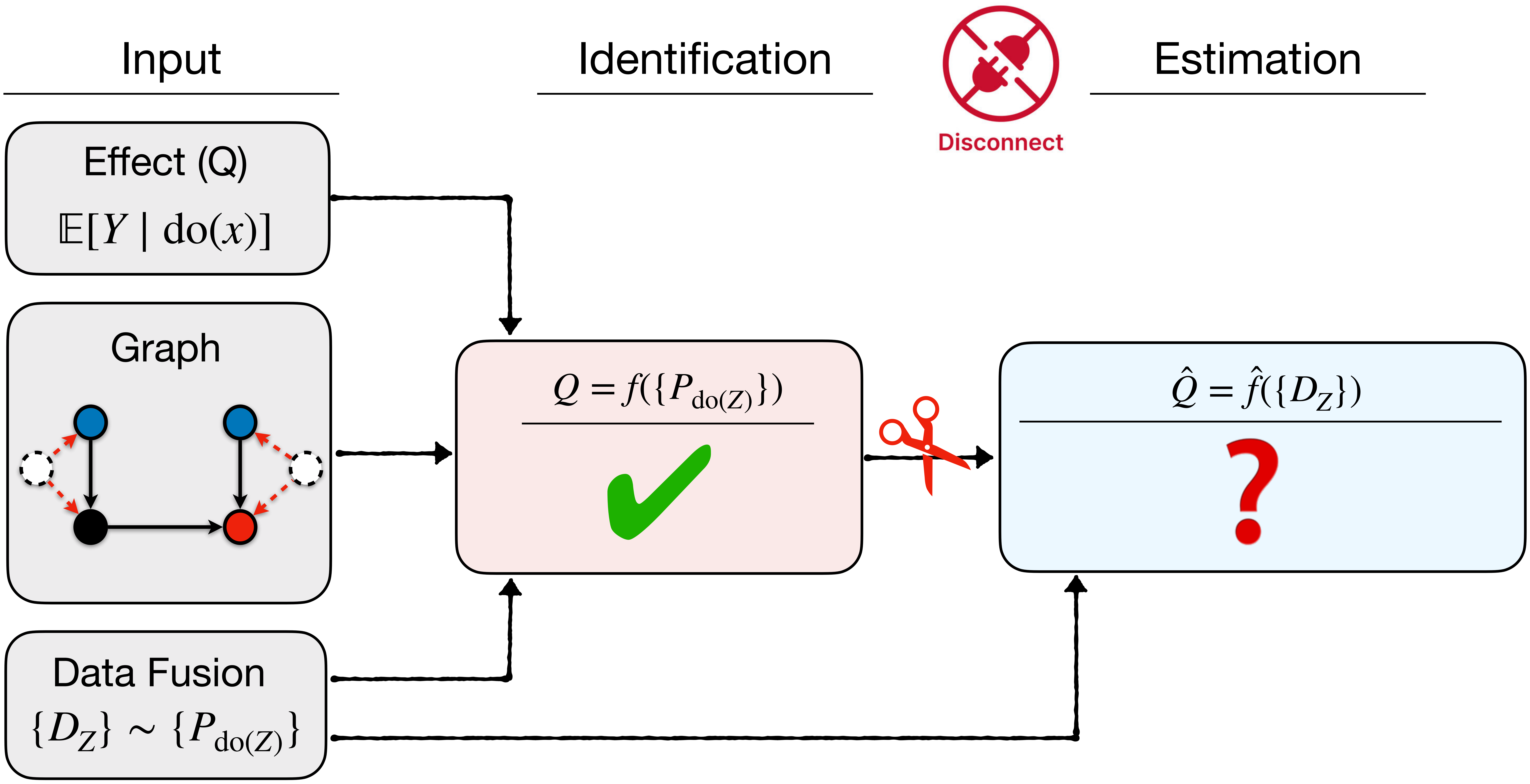
3. Unified causal effect estimation method

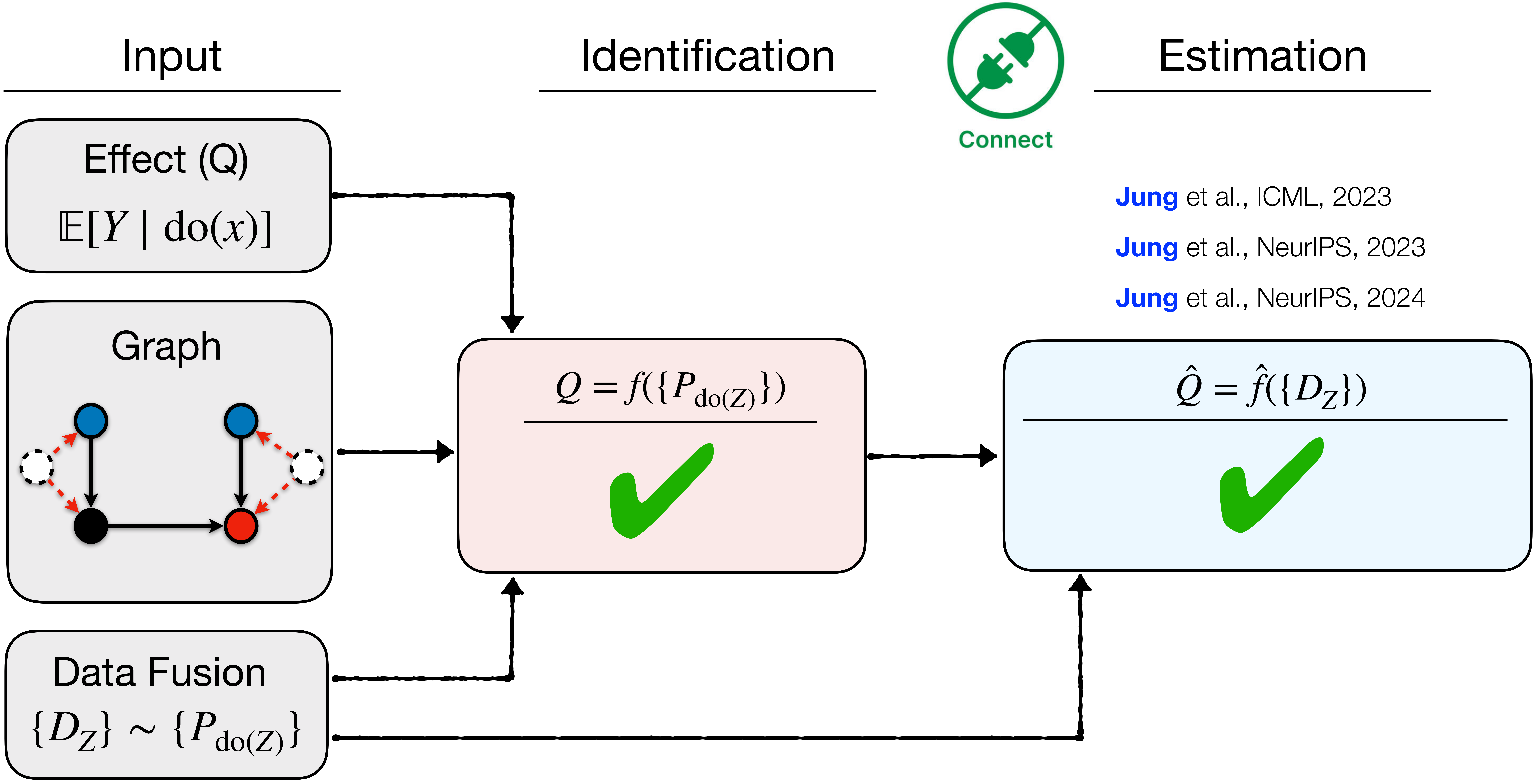
4. Summary & Future direction

Talk Outline

✔

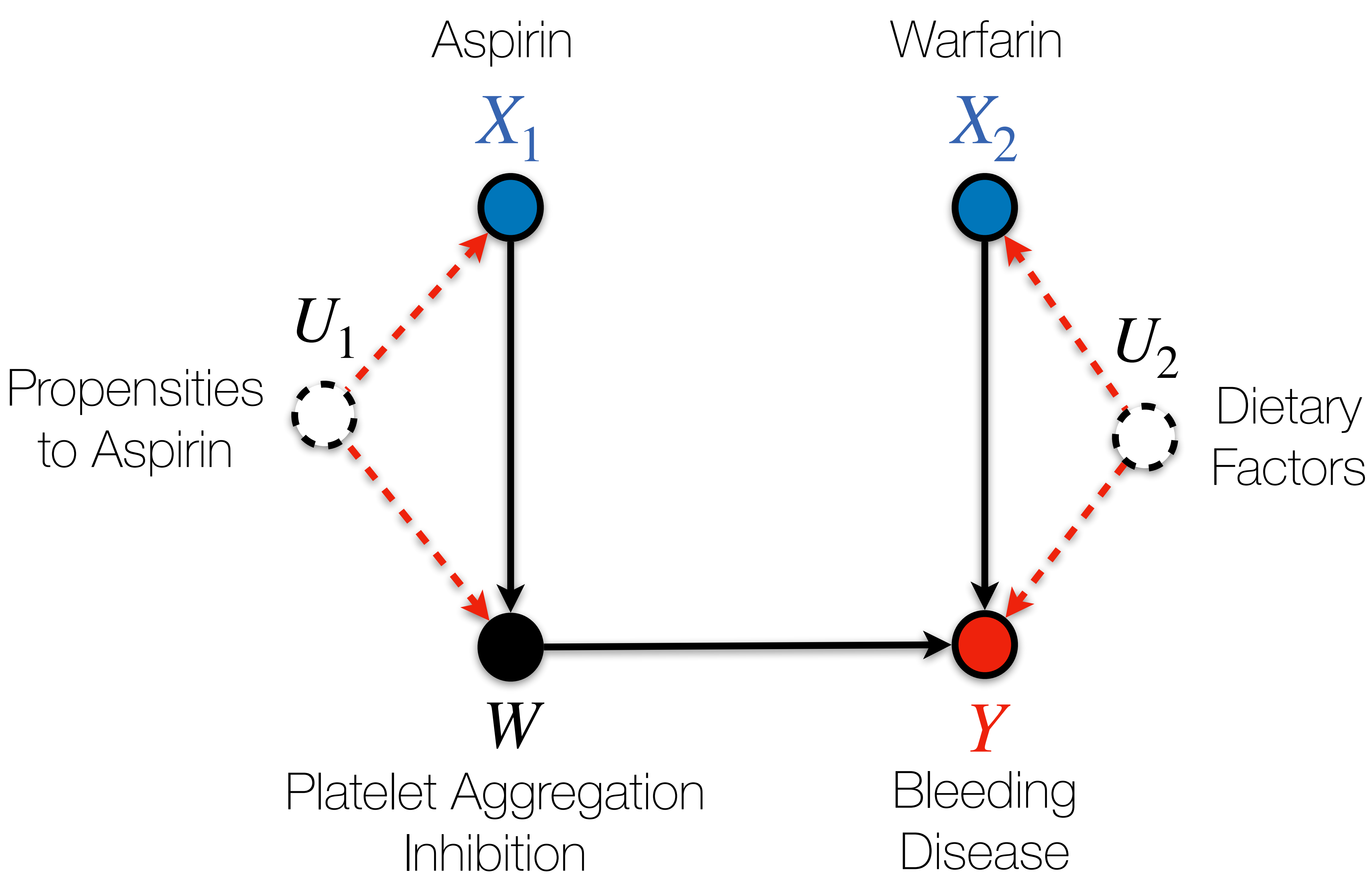
➋ Estimating causal effects from data fusion


Input
Identification
Disconnect
Estimation
Effect (Q)
$\mathbb{E}[Y \mid \mathrm{do}(x)]$
Graph
Data Fusion
$\{D_Z\} \sim \{P_{\mathrm{do}(Z)}\}$
$Q = f(\{P_{\mathrm{do}(Z)}\})$
$\hat{Q} = \hat{f}(\{D_Z\})$
?

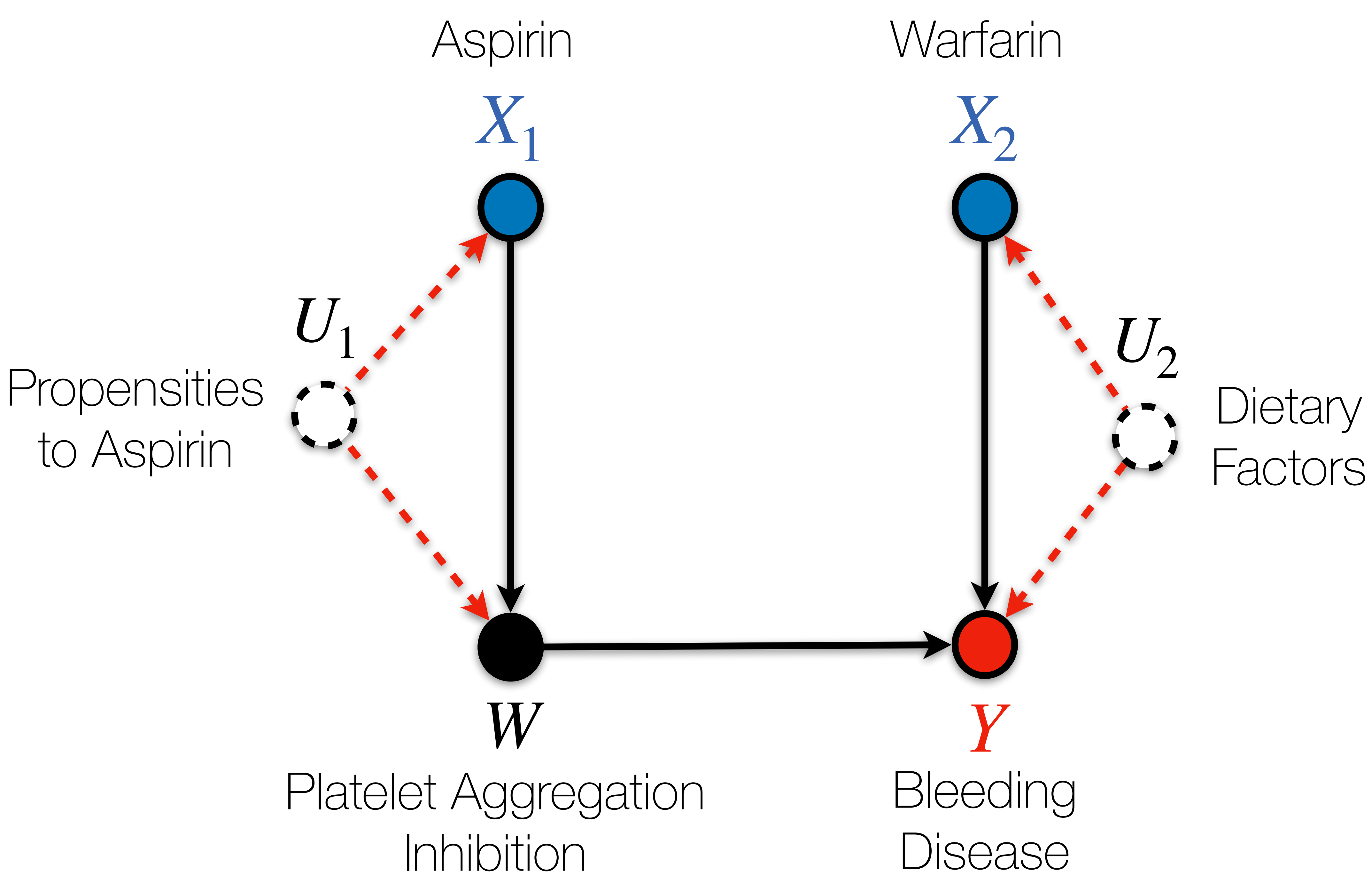


Input
Identification
Connect
Estimation
Effect (Q)
$\mathbb{E}[Y \mid do(x)]$
Graph
Data Fusion
$\{D_Z\} \sim \{P_{do(Z)}\}$
$Q = f(\{P_{do(Z)}\})$
$\hat{Q} = \hat{f}(\{D_Z\})$
Jung et al., ICML, 2023
Jung et al., NeurIPS, 2023
Jung et al., NeurIPS, 2024

Motivation: Joint Treatment Effect Estimation

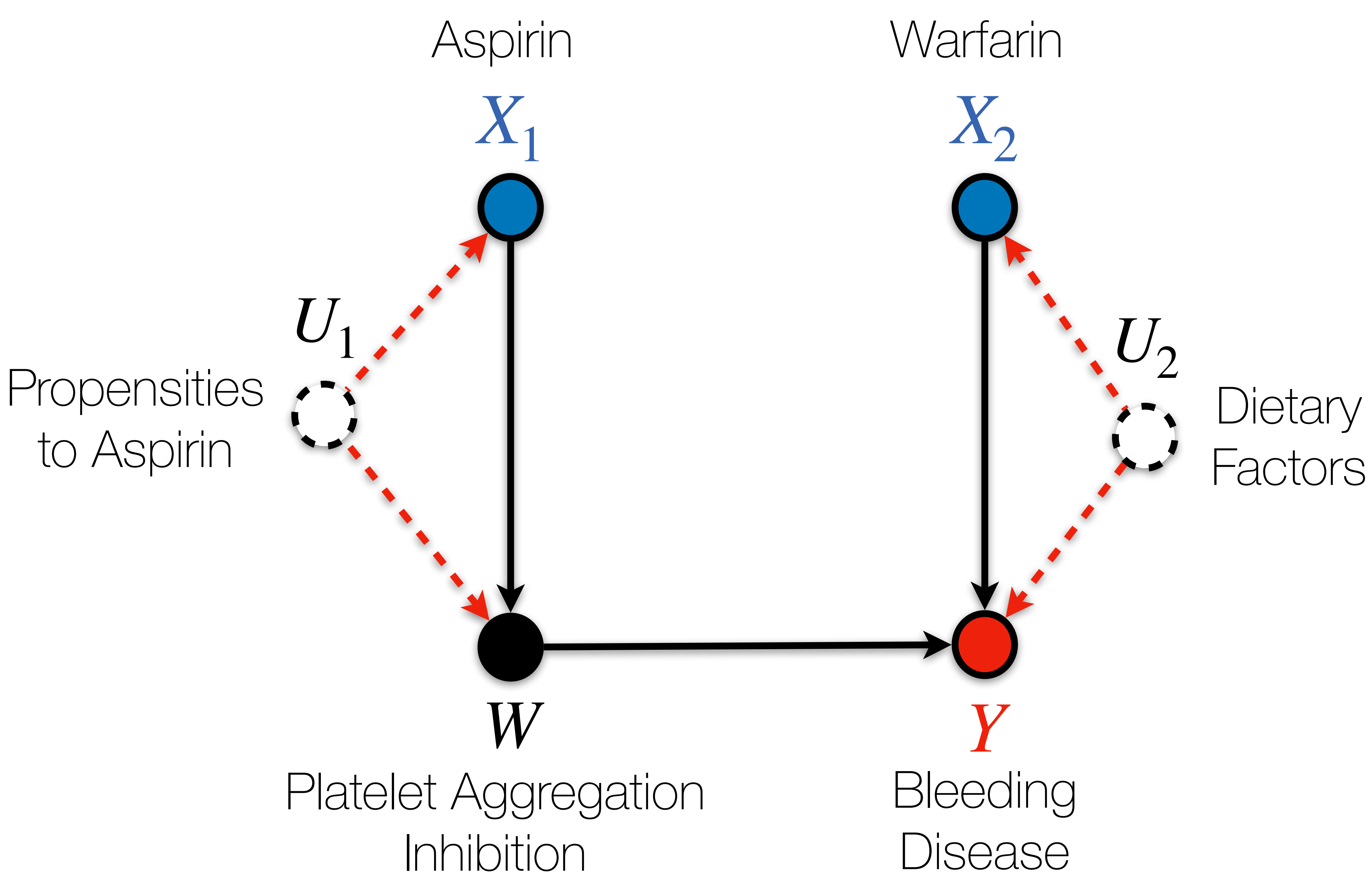


Motivation: Joint Treatment Effect Estimation



Challenges for Estimating $\mathbb{E}[Y \mid \mathrm{do}(x_1, x_2)]$

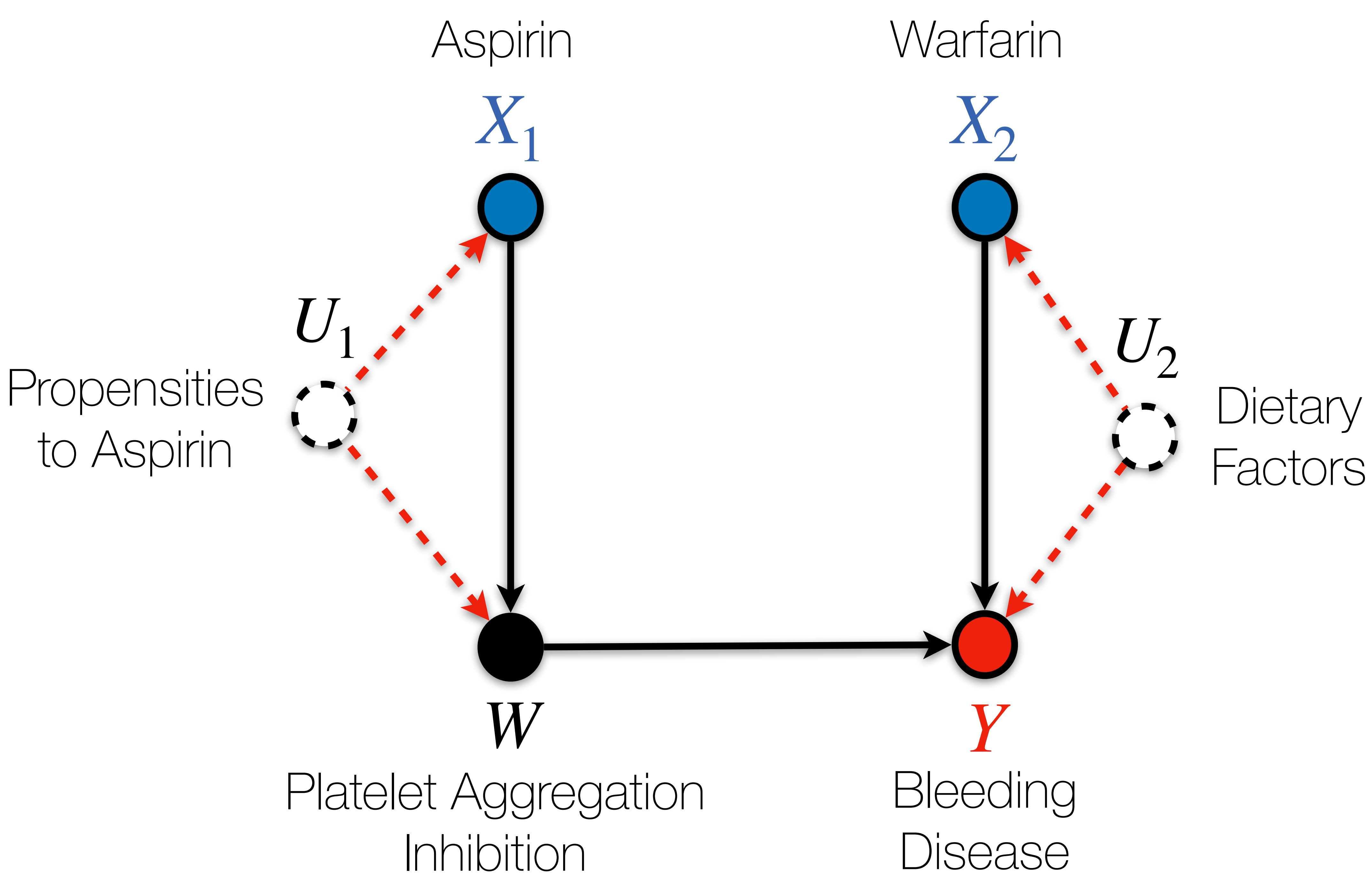
Motivation: Joint Treatment Effect Estimation



Challenges for Estimating $\mathbb{E}[Y \mid \mathrm{do}(x_1, x_2)]$

- BD is not applicable

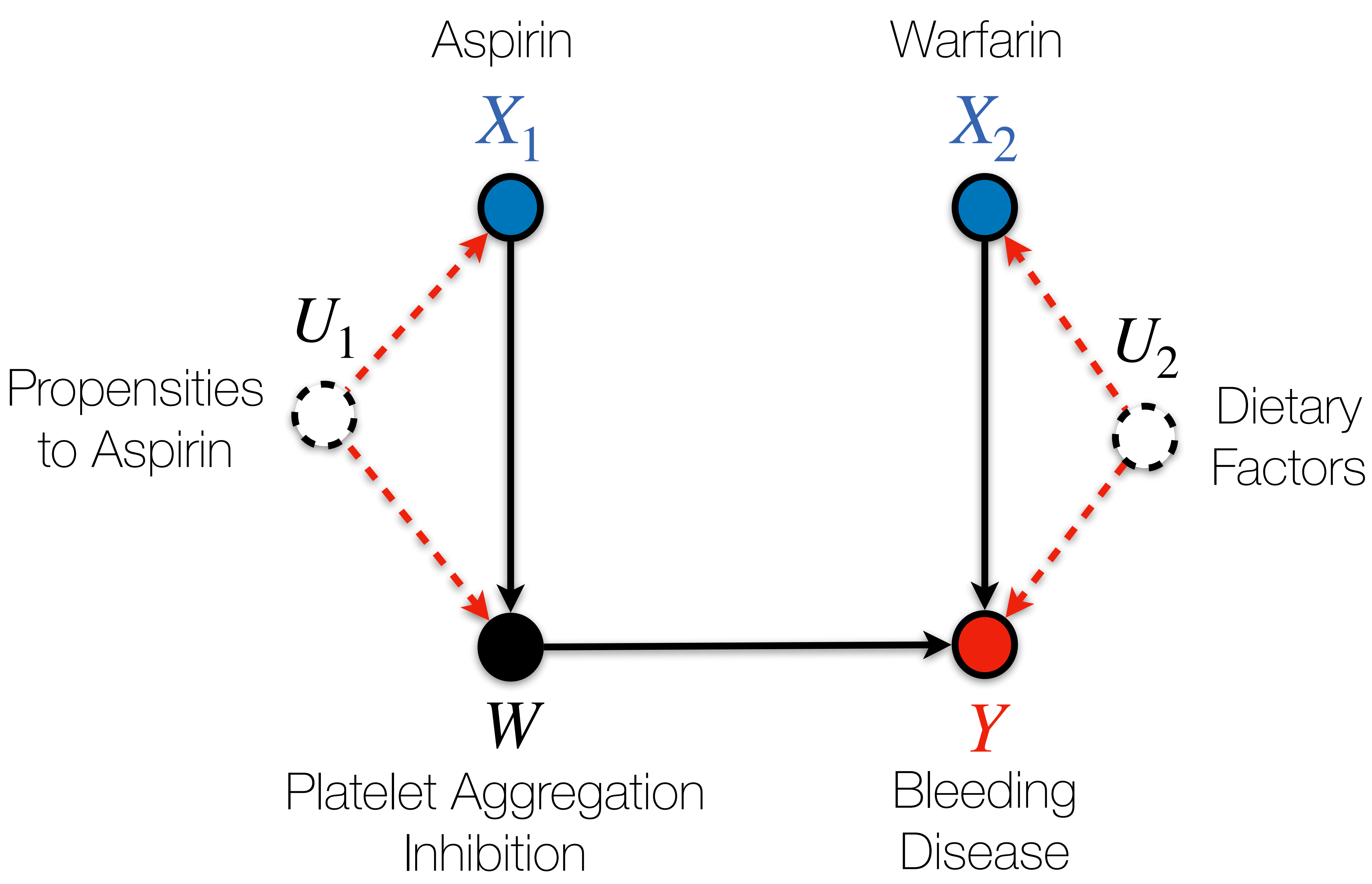
Motivation: Joint Treatment Effect Estimation



Challenges for Estimating $\mathbb{E}[Y \mid \mathrm{do}(x_1, x_2)]$

- BD is not applicable
- Not identifiable from observations $P(\mathbf{V})$.

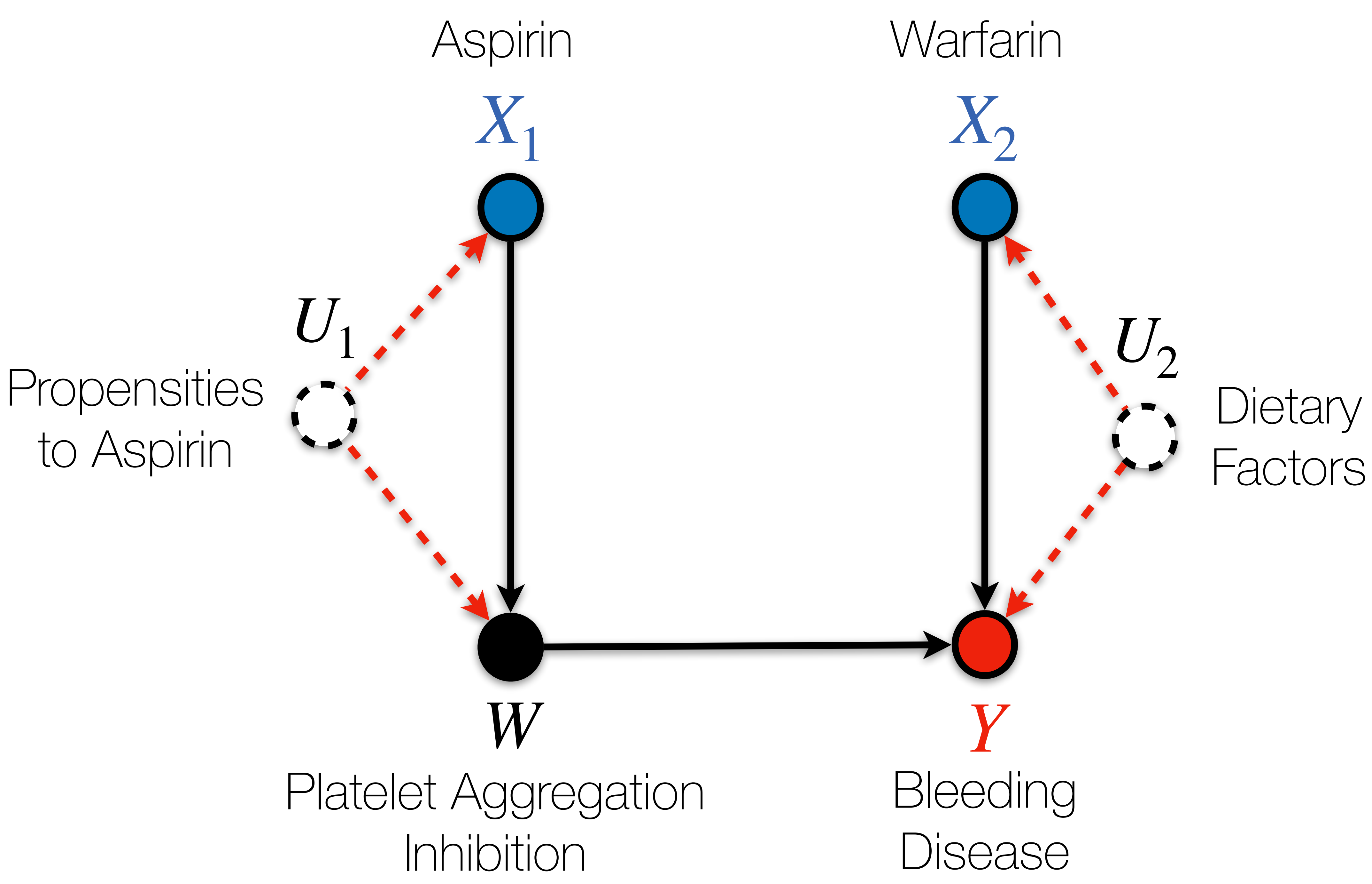
Motivation: Joint Treatment Effect Estimation



Challenges for Estimating $\mathbb{E}[Y \mid \mathrm{do}(x_1, x_2)]$

- BD is not applicable
- Not identifiable from observations $P(\mathbf{V})$.
- Can't run experiments $\mathrm{do}(x_1, x_2)$ due to drug-interactions

Motivation: Joint Treatment Effect Estimation

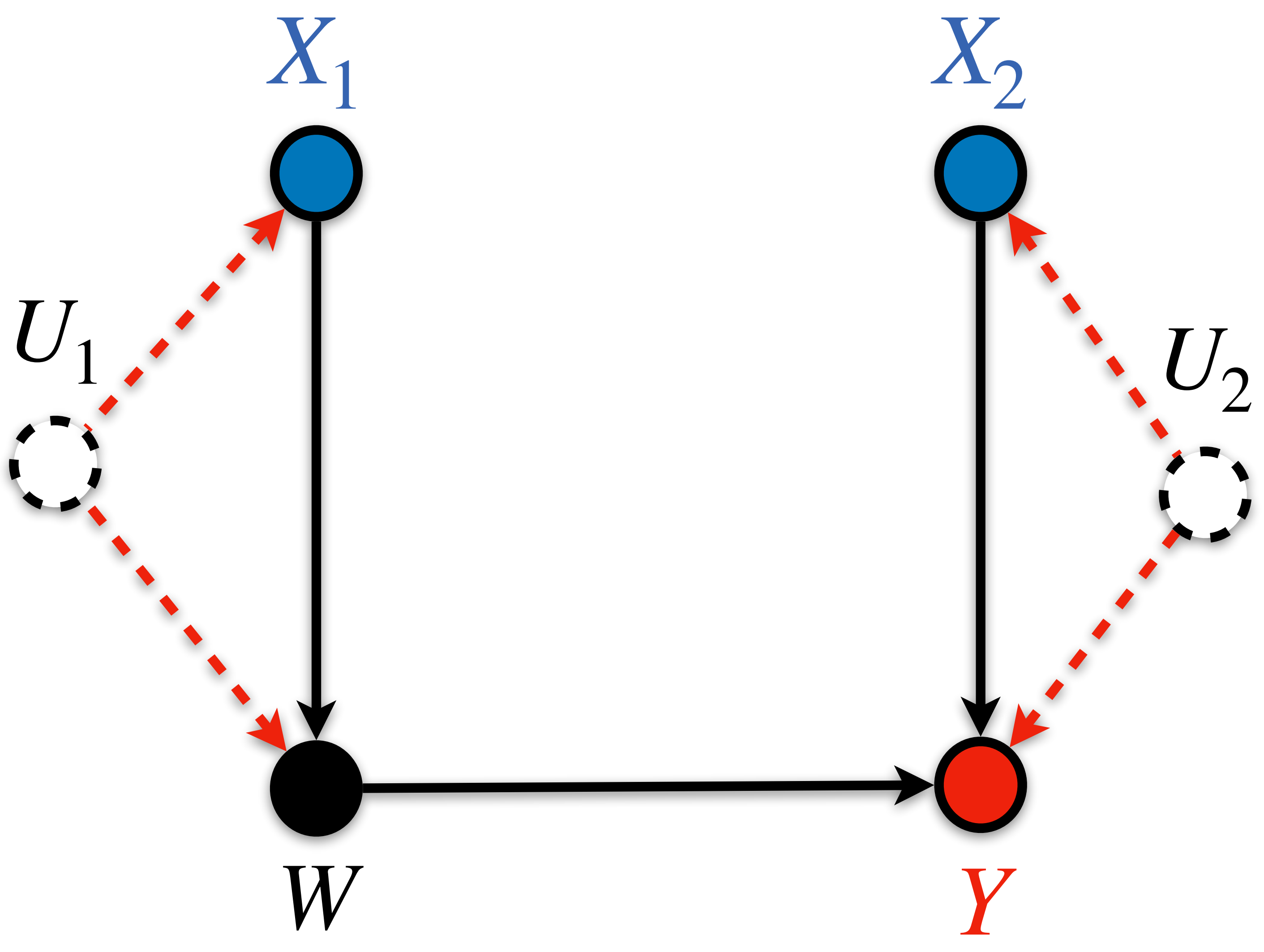


Challenges for Estimating $\mathbb{E}[Y \mid \mathrm{do}(x_1, x_2)]$

- BD is not applicable
- Not identifiable from observations $P(\mathbf{V})$.
- Can't run experiments $\mathrm{do}(x_1, x_2)$ due to drug-interactions

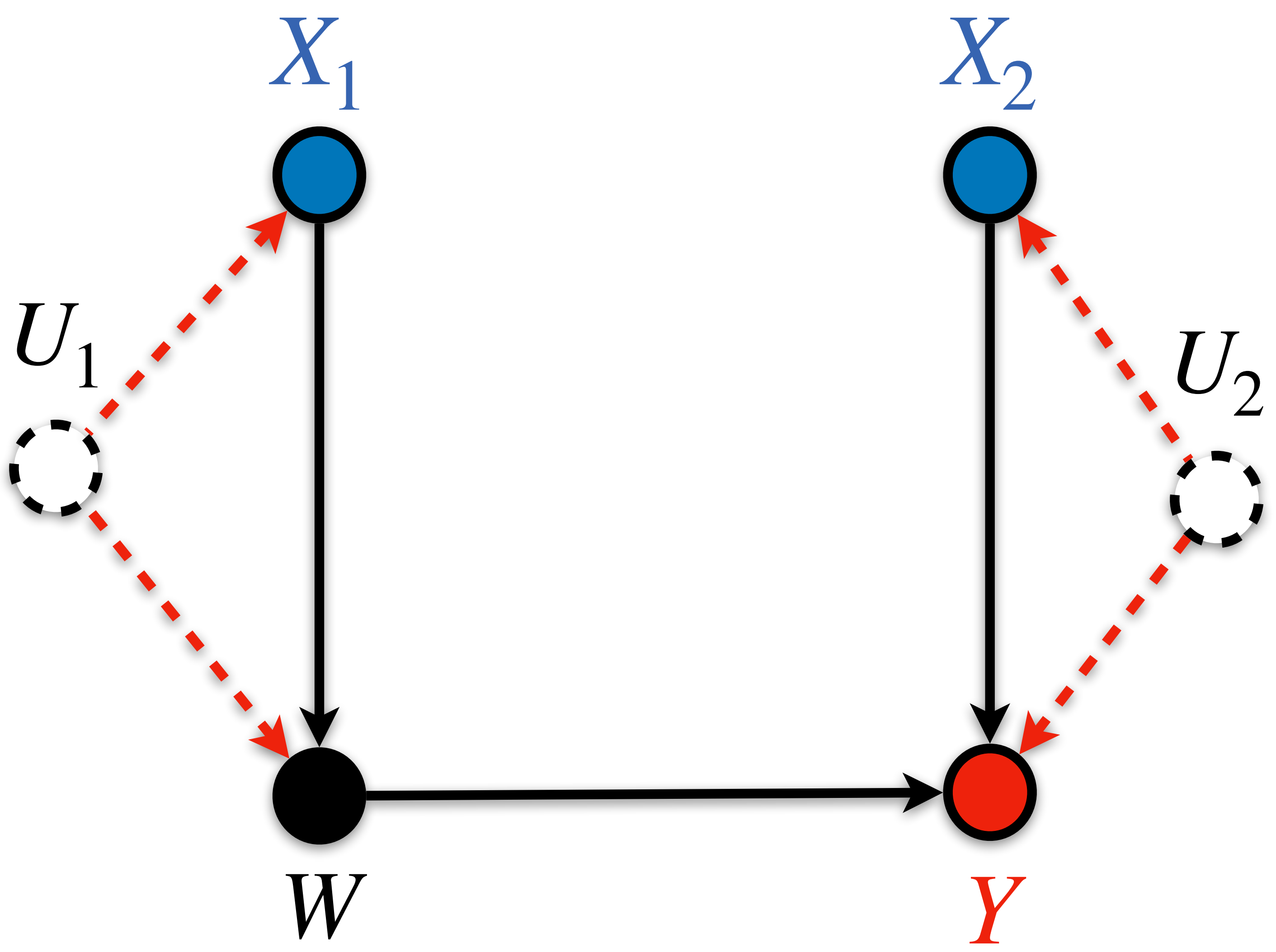
Can $\mathbb{E}[Y \mid \mathrm{do}(x_1, x_2)]$ be estimated from two trials $P_{\mathrm{do}(x_1)}(\mathbf{V})$ and $P_{\mathrm{do}(x_2)}(\mathbf{V})$?

BD+: BD for Data Fusion



$$\mathbb{E}[Y \mid \mathrm{do}(x_1, x_2)]$$

BD+: BD for Data Fusion

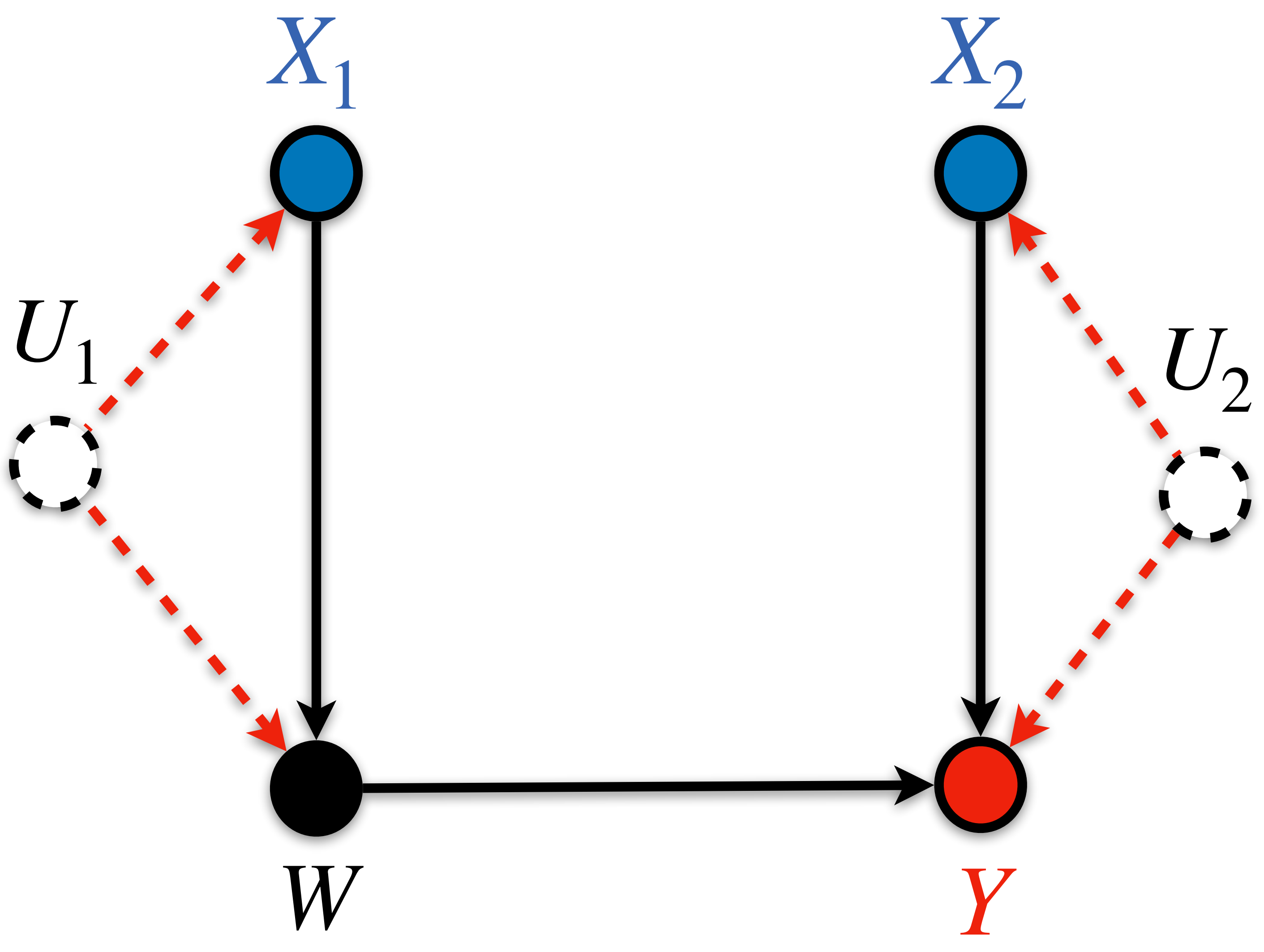


$$\mathbb{E}[Y \mid do(x_1, x_2)]$$

Back-door (BD) [Pearl, 95]

Spurious paths between (treatments, outcome) are blocked by observed variables

BD+: BD for Data Fusion

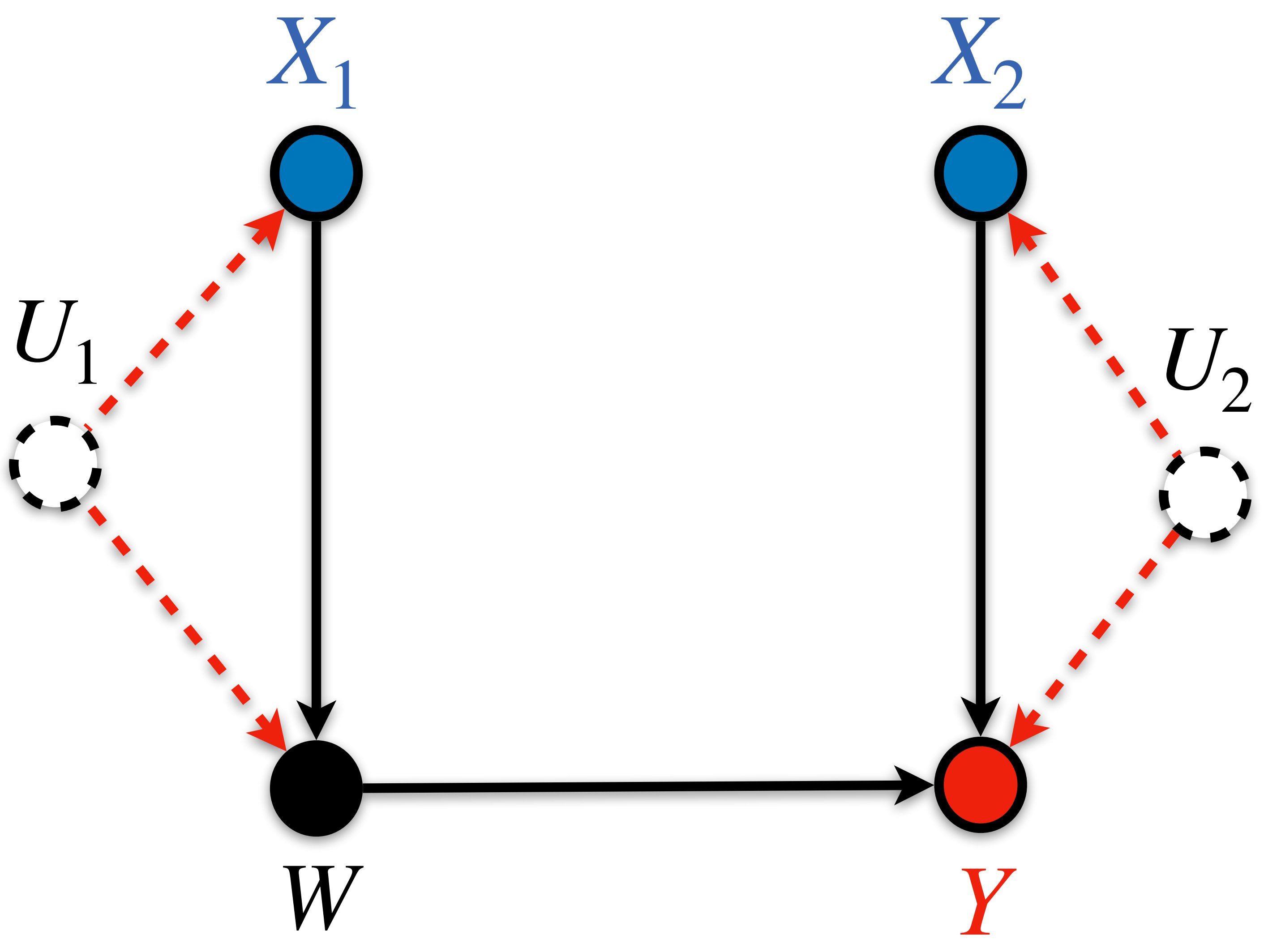


$$\mathbb{E}[Y \mid \mathrm{do}(x_1, x_2)]$$

Back-door (BD) [Pearl, 95]

Spurious paths between (treatments, outcome) are blocked by observed variables

BD+: BD for Data Fusion



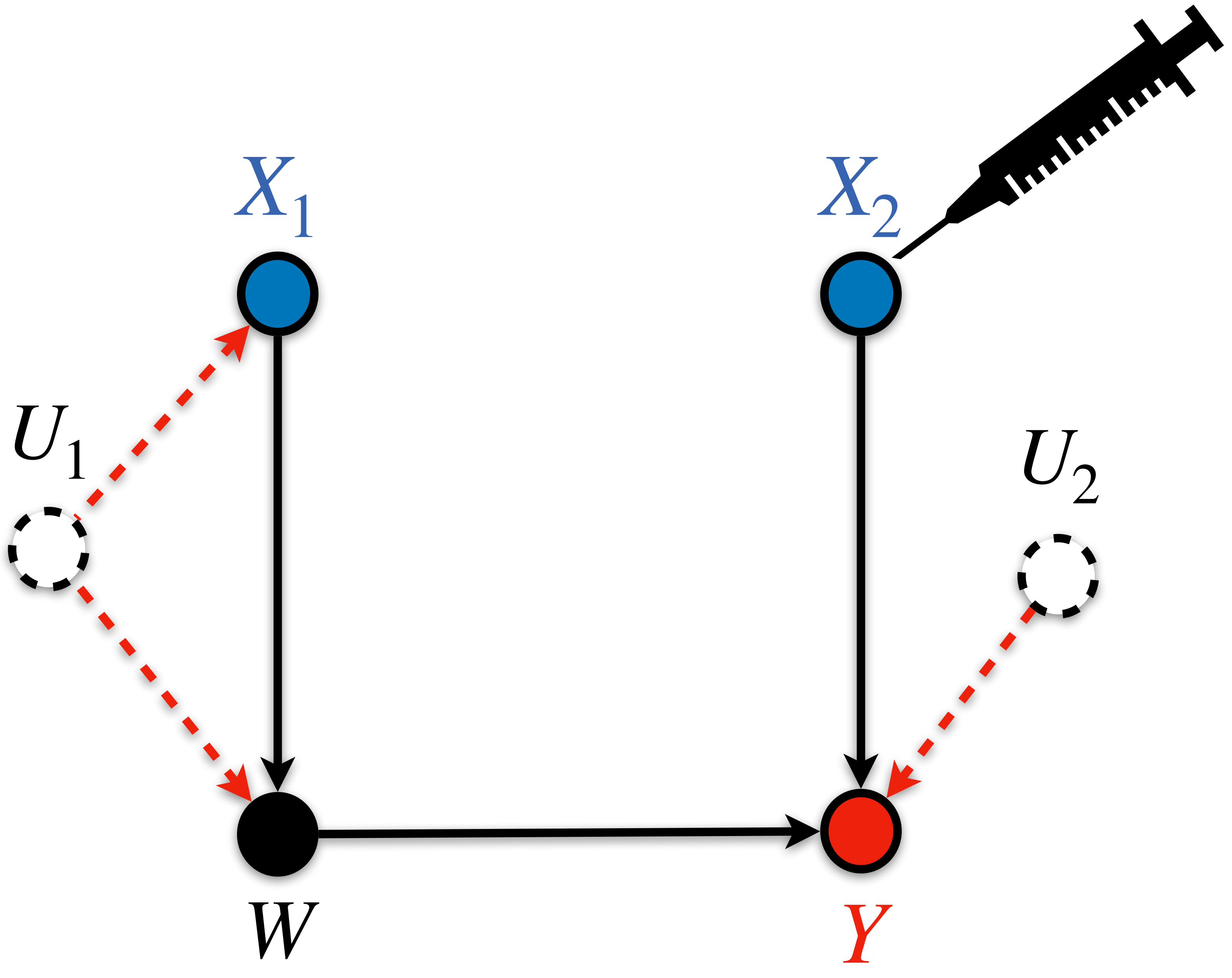
$$\mathbb{E}[Y \mid \mathrm{do}(x_1, x_2)]$$

Jung et al., ICML 2023

BD for Fusion (BD+)

- For (treatment_1, treatment_2) partitioning treatments,
- spurious paths between (treatment_1, outcome) are blocked by observed variables
- in the experiments for $\mathrm{do}(\text{treatment}_2)$.

BD+: BD for Data Fusion



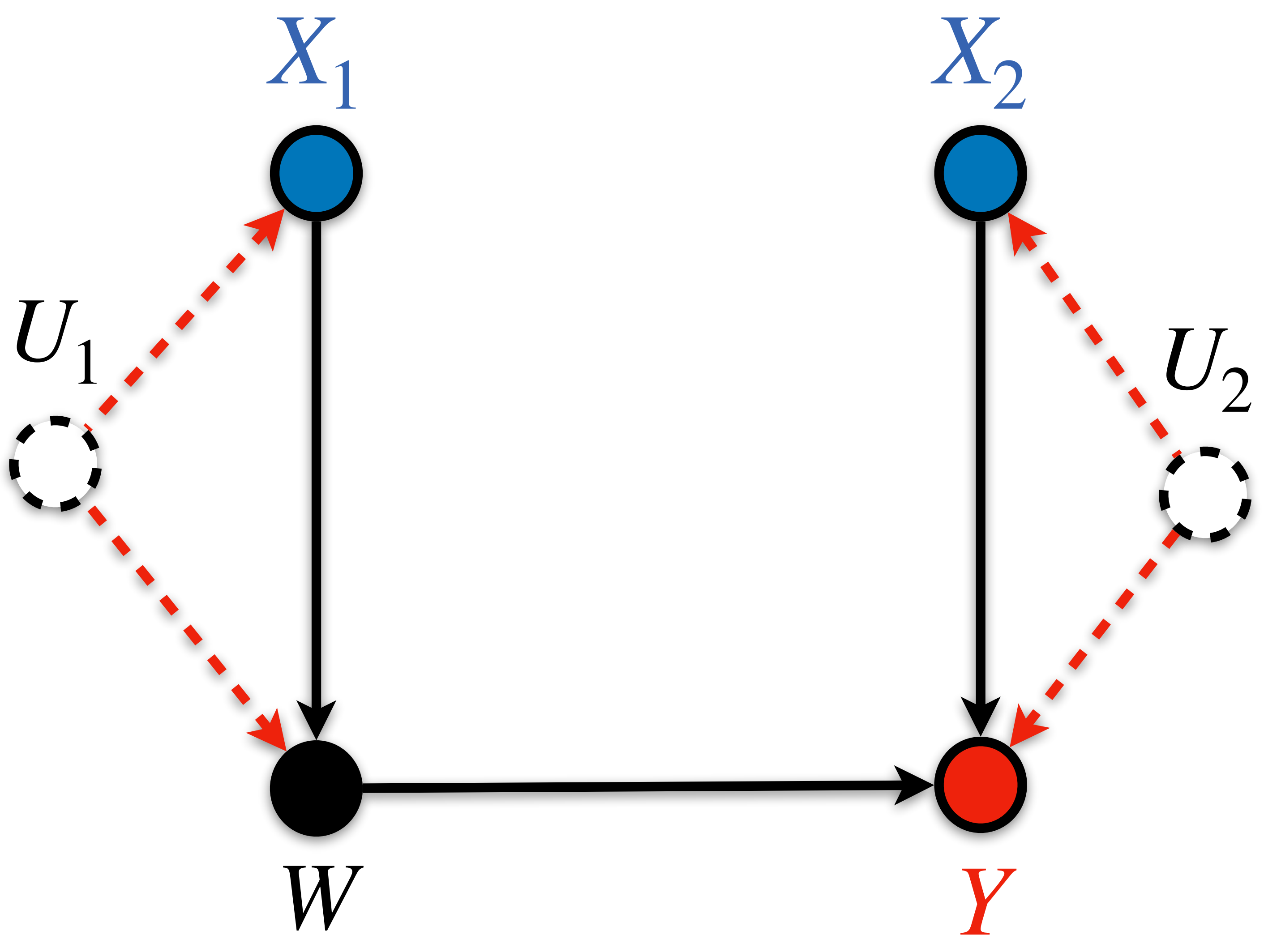
$$\mathbb{E}[Y \mid \mathrm{do}(x_1, x_2)]$$

Jung et al., ICML 2023

BD for Fusion (BD+)

- For (treatment$_1$, treatment$_2$) partitioning treatments,
- spurious paths between (treatment$_1$, outcome) are blocked by observed variables
- in the experiments for do(treatment$_2$).

BD+: BD for Data Fusion



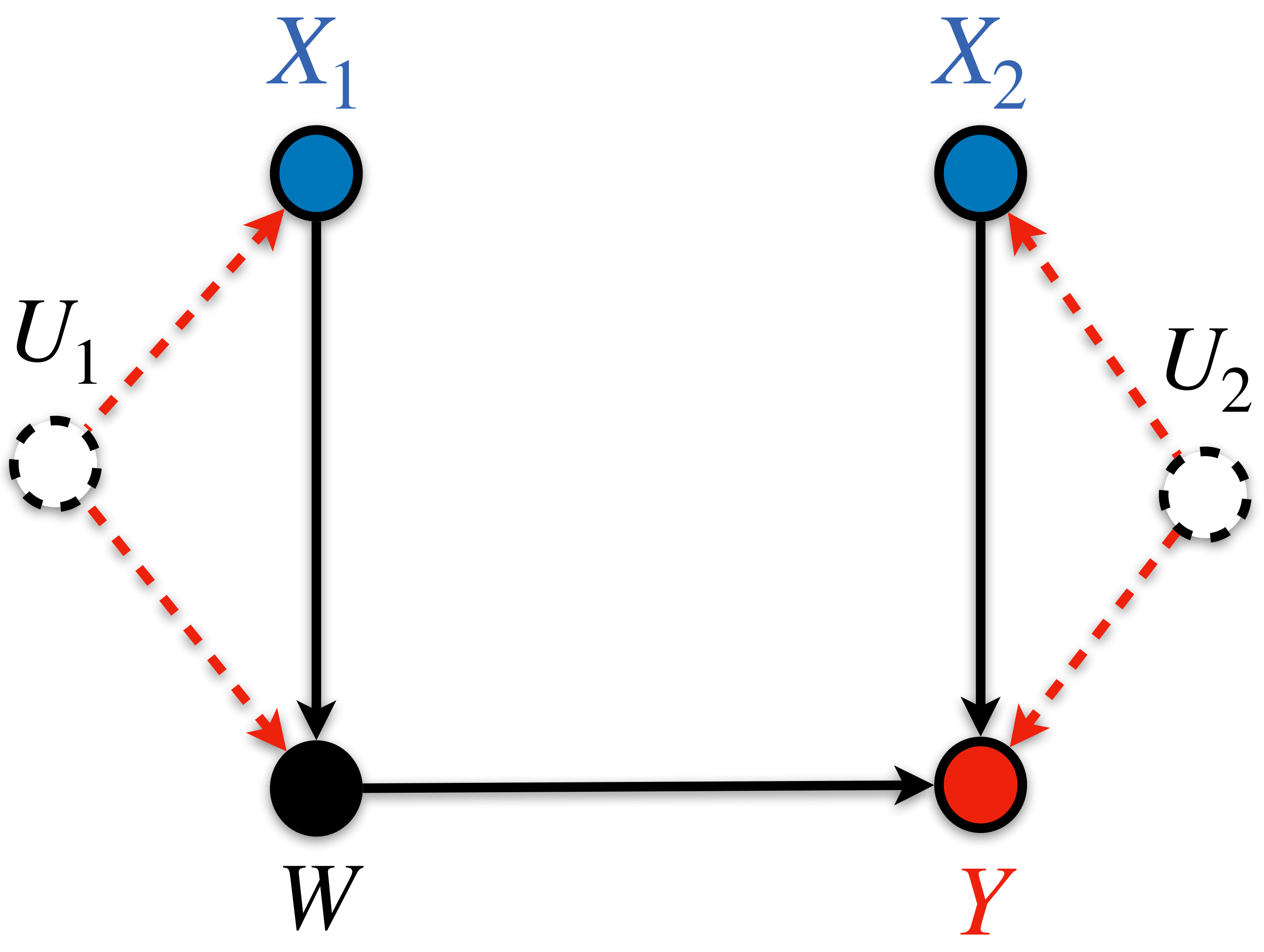
$$\mathbb{E}[Y \mid \mathrm{do}(x_1, x_2)]$$

Jung et al., ICML 2023

BD for Fusion (BD+)

- For (treatment$_1$, treatment$_2$) partitioning treatments,
- spurious paths between (treatment$_1$, outcome) are blocked by observed variables
- in the experiments for do(treatment$_2$).

BD+: BD for Data Fusion



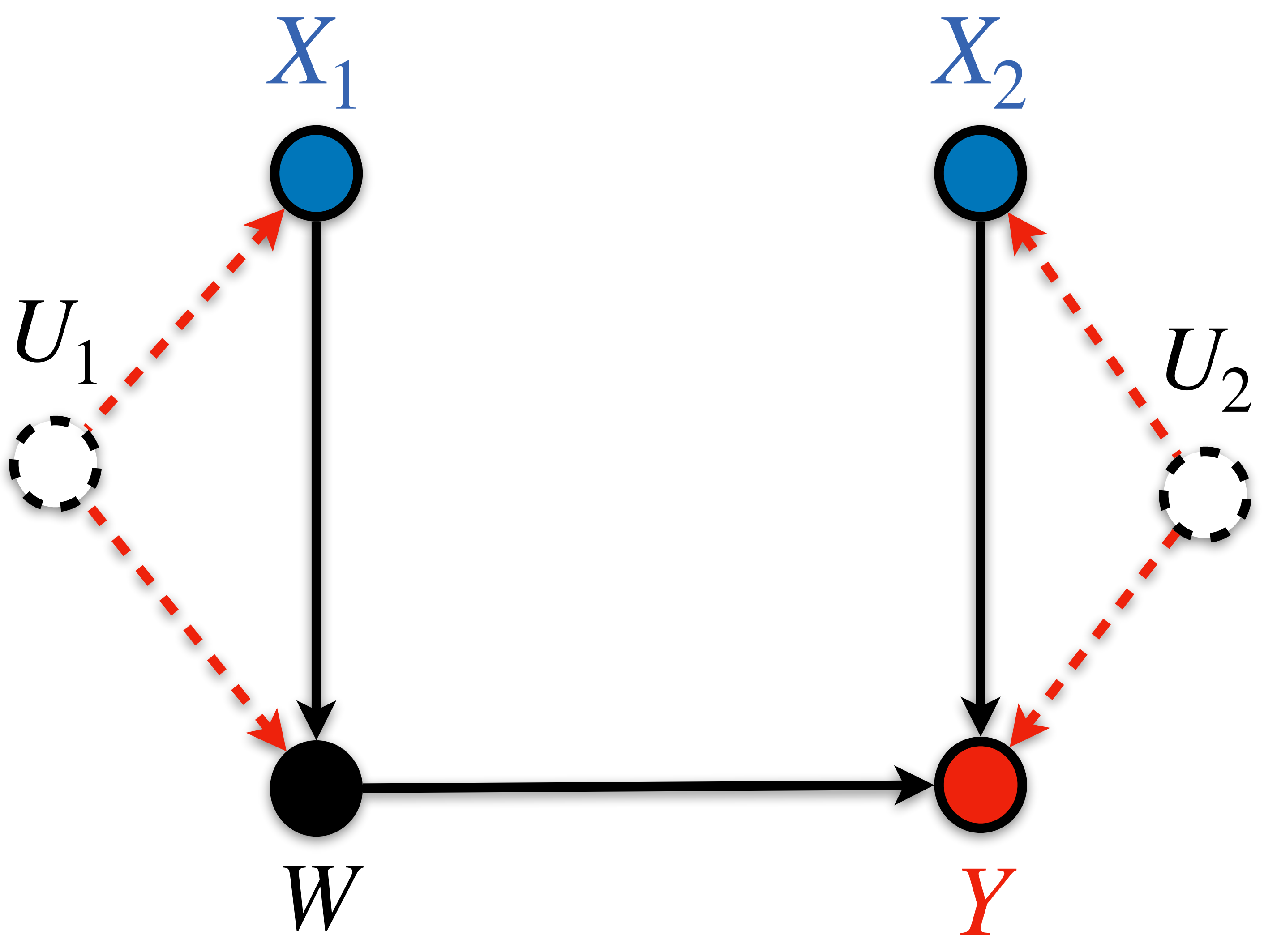
Jung et al., ICML 2023

BD for Fusion (BD+)

- For (treatment₁, treatment₂) partitioning treatments,
- spurious paths between (treatment₁, outcome) are blocked by observed variables
- in the experiments for do(treatment₂).

$$\mathbb{E}[Y \mid \mathrm{do}(x_1, x_2)] = \sum_w \mathbb{E}_{\mathrm{do}(x_2)}[Y \mid x_1, w] P_{\mathrm{do}(x_1)}(w)$$

BD+: BD for Data Fusion



Jung et al., ICML 2023

BD for Fusion (BD+)

- For (treatment_1, treatment_2) partitioning treatments,
- spurious paths between (treatment_1, outcome) are blocked by observed variables
- in the experiments for $\text{do}(\text{treatment}_2)$.

$$\mathbb{E}[Y \mid \text{do}(x_1, x_2)] = \sum_w \underbrace{\mathbb{E}_{\text{do}(x_2)}[Y \mid x_1, w]}_{\text{Trial on } X_2} \underbrace{P_{\text{do}(x_1)}(w)}_{\text{Trial on } X_1}$$

Doubly Robust Estimator for BD$^+$

Doubly Robust Estimator for BD+

BD+ Parametrization

$$\text{BD}^{+}(\mu,\pi) \triangleq \mathbb{E}_{P_{\text{do}(x_2)}}[\mu \times \pi]$$

where

- $\mu(X_1 W) \triangleq \mathbb{E}_{P_{\text{do}(x_2)}}[Y \mid X_1, W]$
- $\pi(X_1 W) \triangleq \dfrac{\mathbb{I}_{x_1}(X_1)}{P_{\text{do}(x_2)}(X_1 \mid W)} \dfrac{P_{\text{do}(x_1)}(W)}{P_{\text{do}(x_2)}(W)}$

Doubly Robust Estimator for BD+

BD+ Parametrization

$$\text{BD}^+(\mu,\pi) \triangleq \mathbb{E}_{P_{\text{do}(x_2)}}[\mu \times \pi]$$

where

- $\mu(X_1 W) \triangleq \mathbb{E}_{P_{\text{do}(x_2)}}[Y \mid X_1, W]$
- $\pi(X_1 W) \triangleq \dfrac{\mathbb{I}_{x_1}(X_1)}{P_{\text{do}(x_2)}(X_1 \mid W)} \dfrac{P_{\text{do}(x_1)}(W)}{P_{\text{do}(x_2)}(W)}$

Theorem

DML-BD+$(\hat{\mu},\hat{\pi})$ achieves the followings:

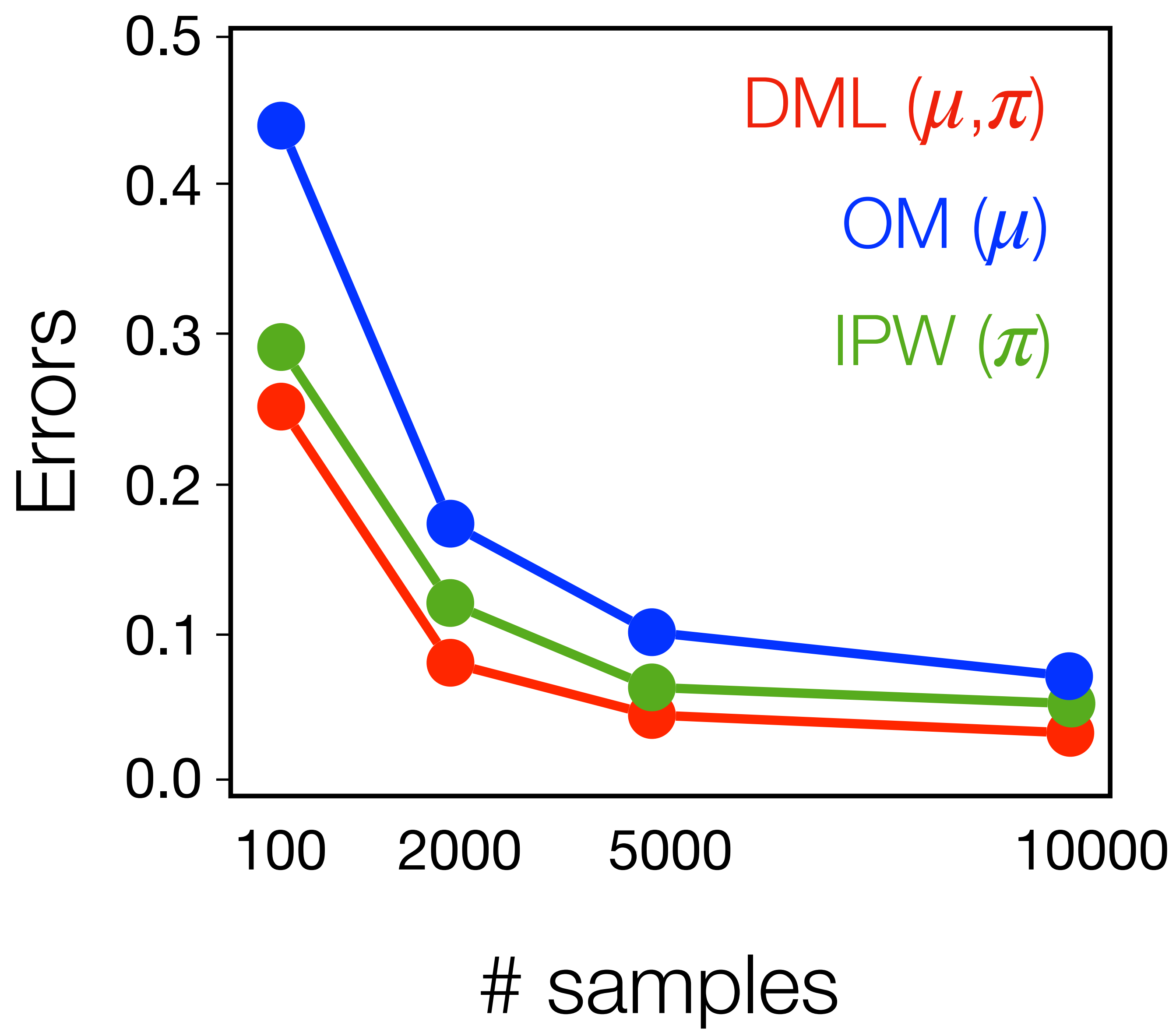
- **Double Robustness:** Error $= 0$ if either $\hat{\mu} = \mu$ or $\hat{\pi} = \pi$
- **Fast Convergence:** Error $\to 0$ fast even when $\hat{\mu} \to \mu$ and $\hat{\pi} \to \pi$ slowly

Simulation: DML-BD^{+}

Simulation: DML-BD+

Fast Convergence

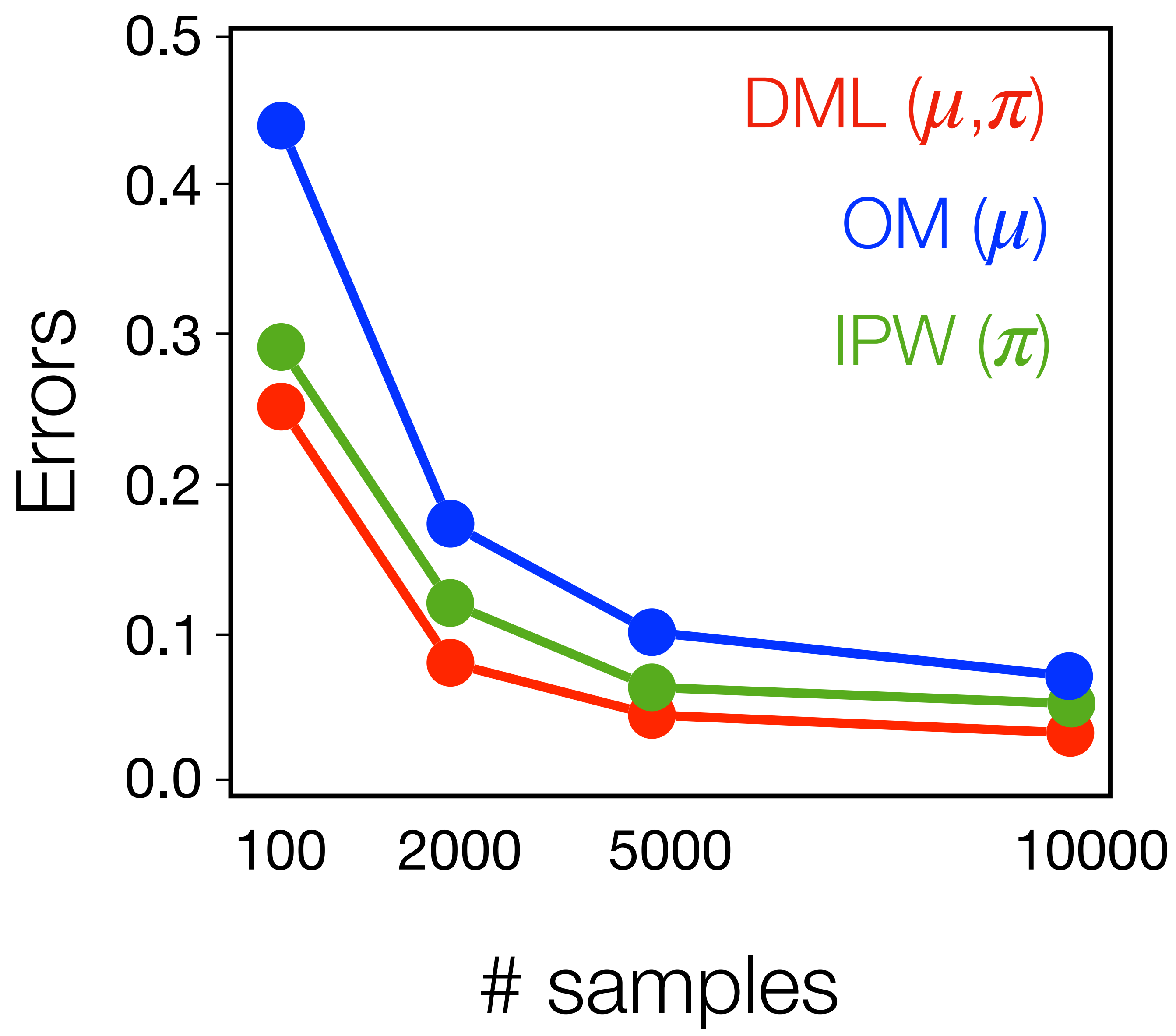
$(\hat{\boldsymbol{\mu}}, \hat{\boldsymbol{\pi}}) \rightarrow (\boldsymbol{\mu}_0, \boldsymbol{\pi}_0)$ slowly



Simulation: DML-BD+

Fast Convergence

$(\hat{\boldsymbol{\mu}}, \hat{\boldsymbol{\pi}}) \to (\boldsymbol{\mu}_0, \boldsymbol{\pi}_0)$ slowly

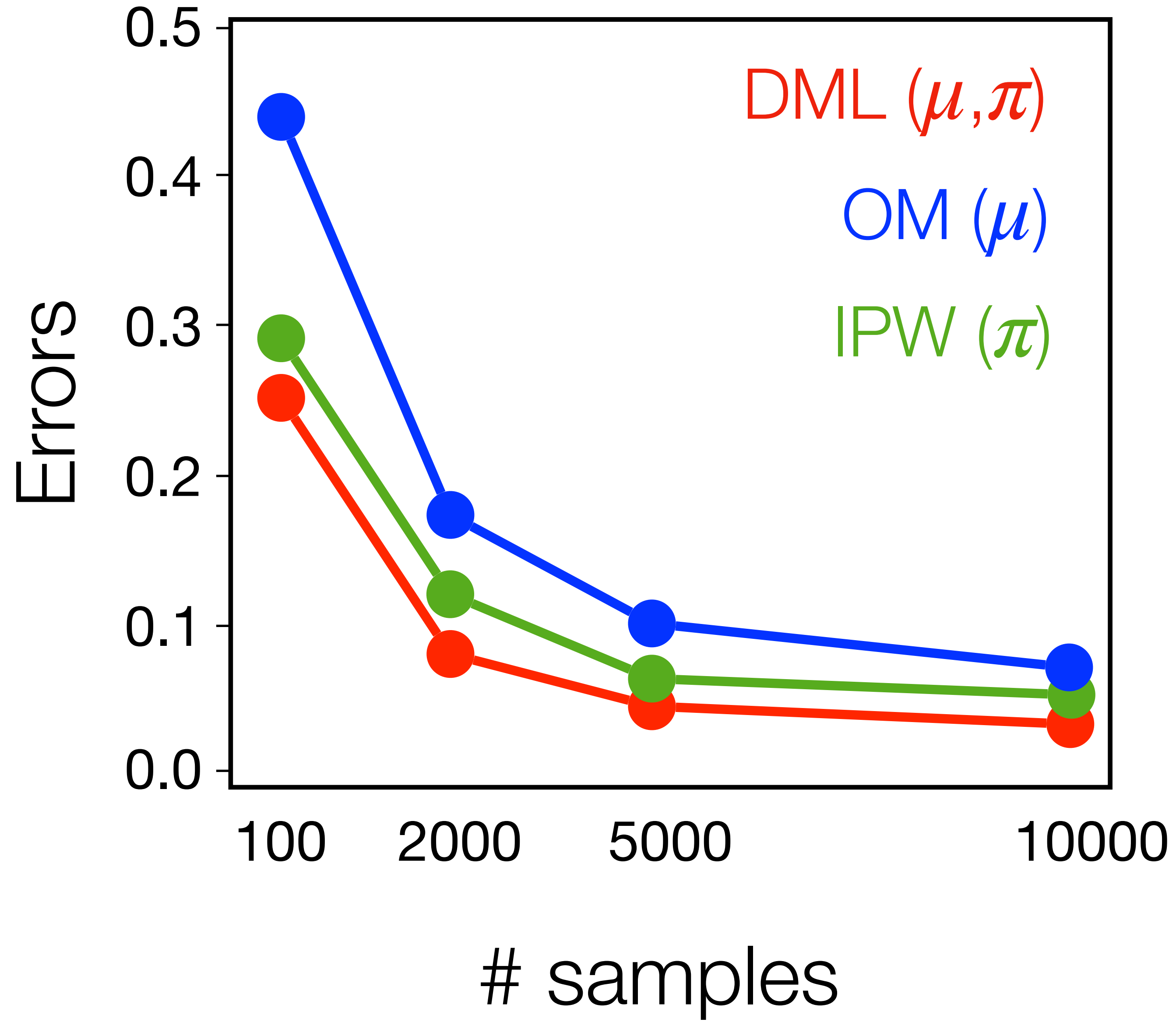


DML-BD+ converges fast, even when $(\hat{\boldsymbol{\mu}}, \hat{\boldsymbol{\pi}})$ converge slowly

Simulation: DML-BD+

Fast Convergence

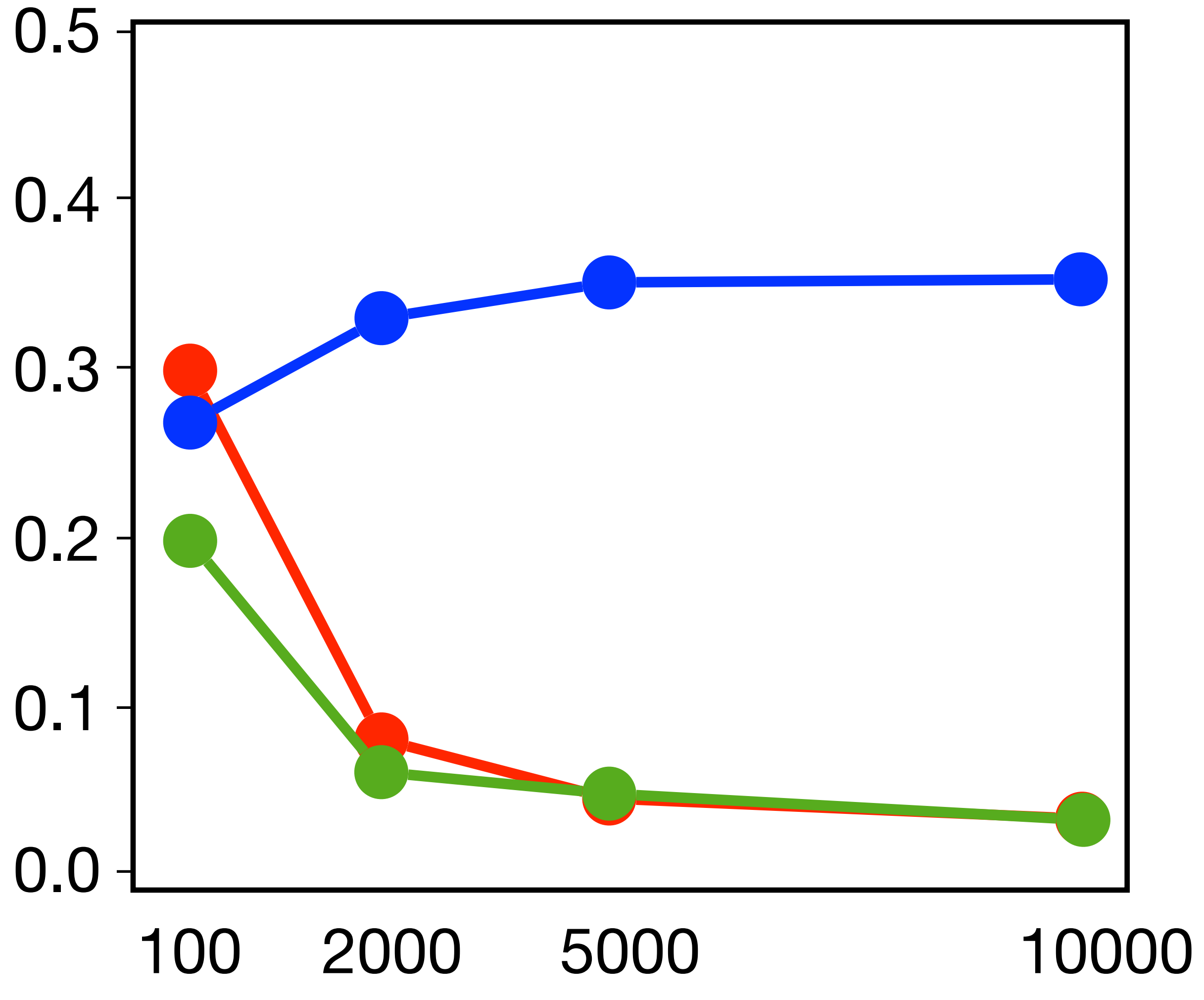
$(\hat{\boldsymbol{\mu}},\hat{\boldsymbol{\pi}}) \rightarrow (\boldsymbol{\mu}_0, \boldsymbol{\pi}_0)$ slowly



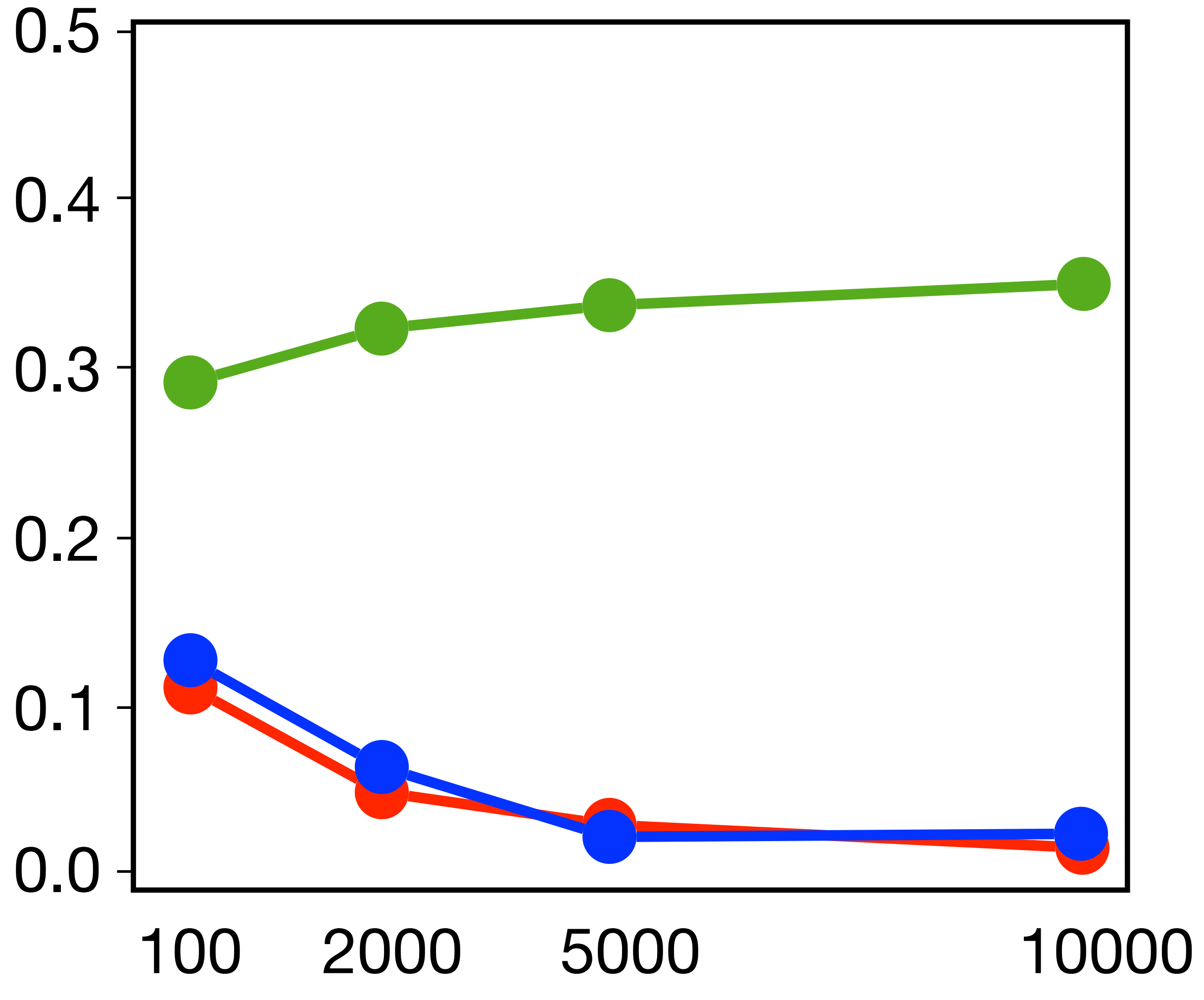
DML-BD+ converges fast, even when $(\hat{\boldsymbol{\mu}},\hat{\boldsymbol{\pi}})$ converge slowly

Double Robustness

$\hat{\boldsymbol{\mu}}$ misspecified $(\hat{\boldsymbol{\mu}} \neq \boldsymbol{\mu})$



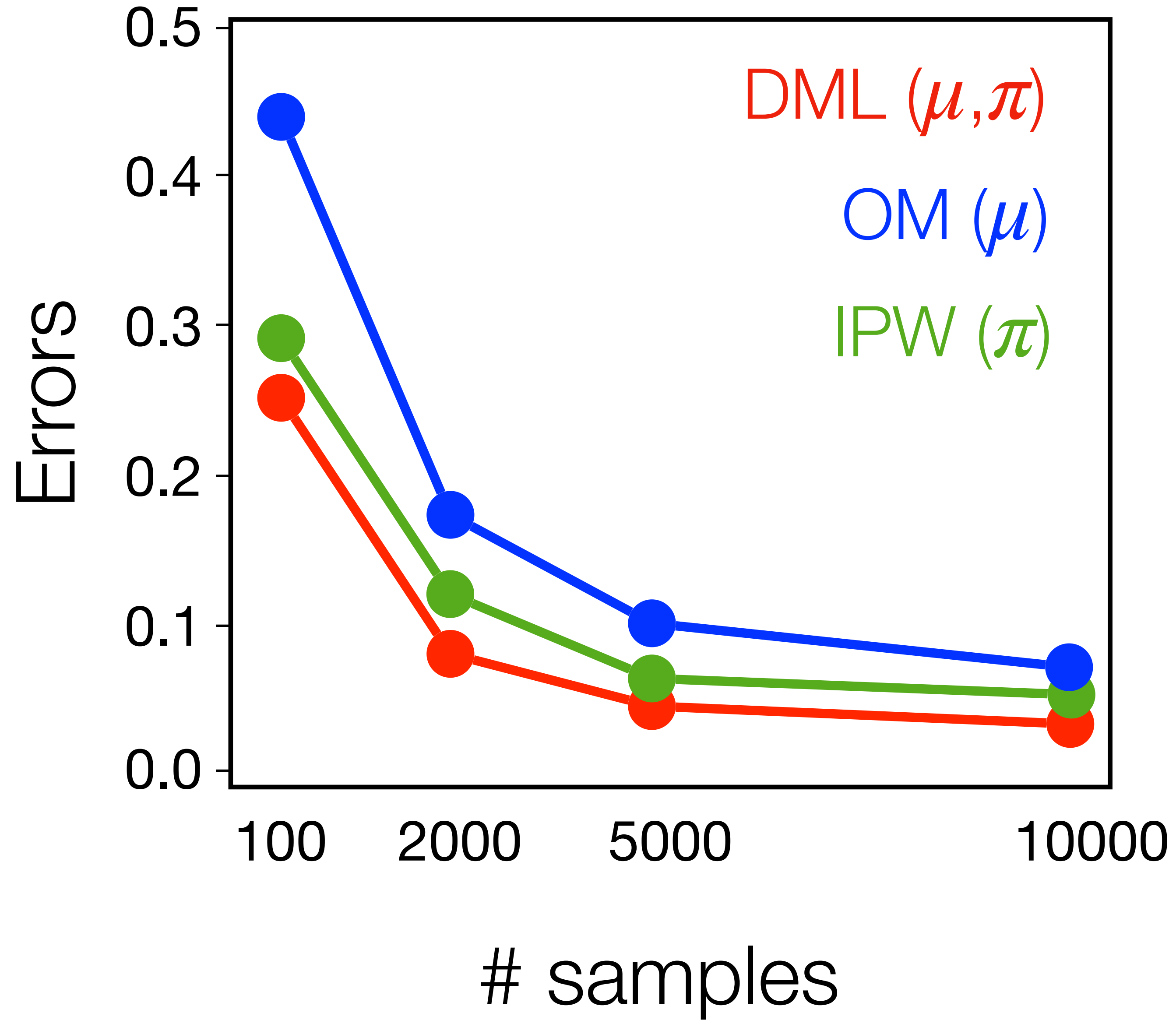
$\hat{\boldsymbol{\pi}}$ misspecified $(\hat{\boldsymbol{\pi}} \neq \boldsymbol{\pi})$



Simulation: DML-BD+

Fast Convergence

$(\hat{\boldsymbol{\mu}},\hat{\boldsymbol{\pi}}) \to (\boldsymbol{\mu}_0,\boldsymbol{\pi}_0)$ slowly

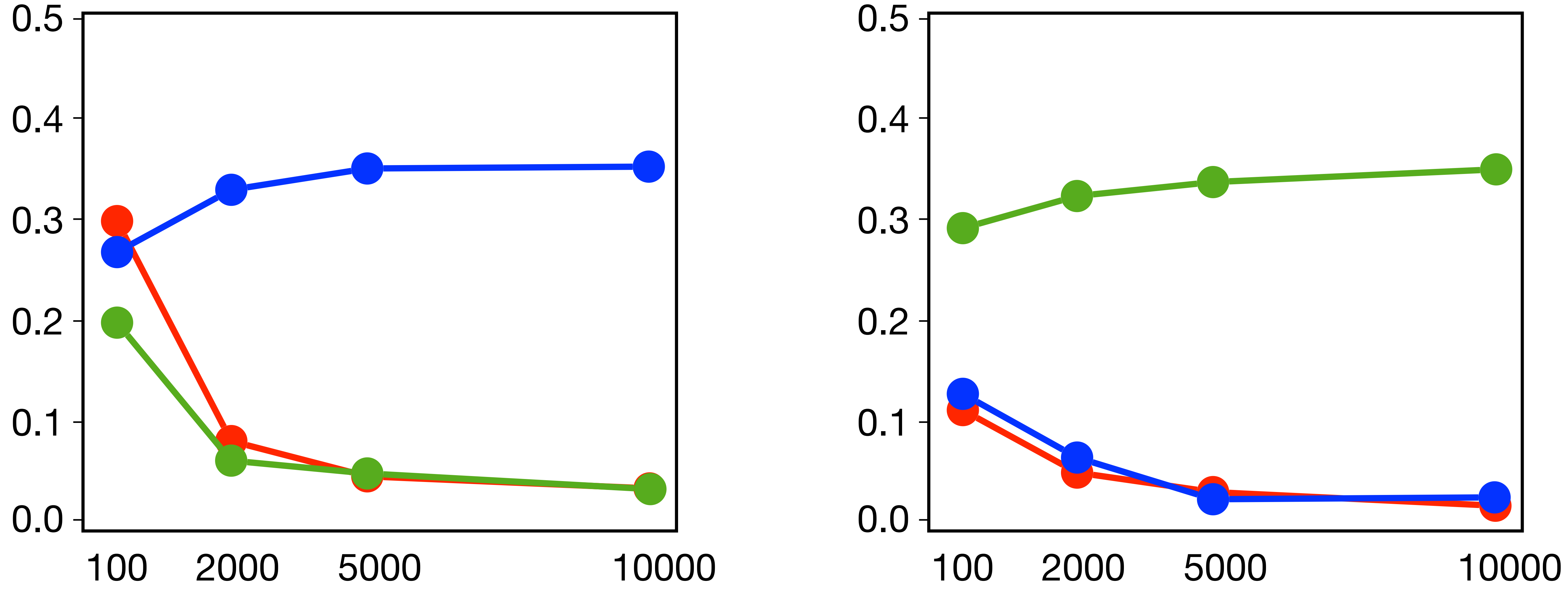


DML-BD+ converges fast, even when $(\hat{\boldsymbol{\mu}},\hat{\boldsymbol{\pi}})$ converge slowly

Double Robustness

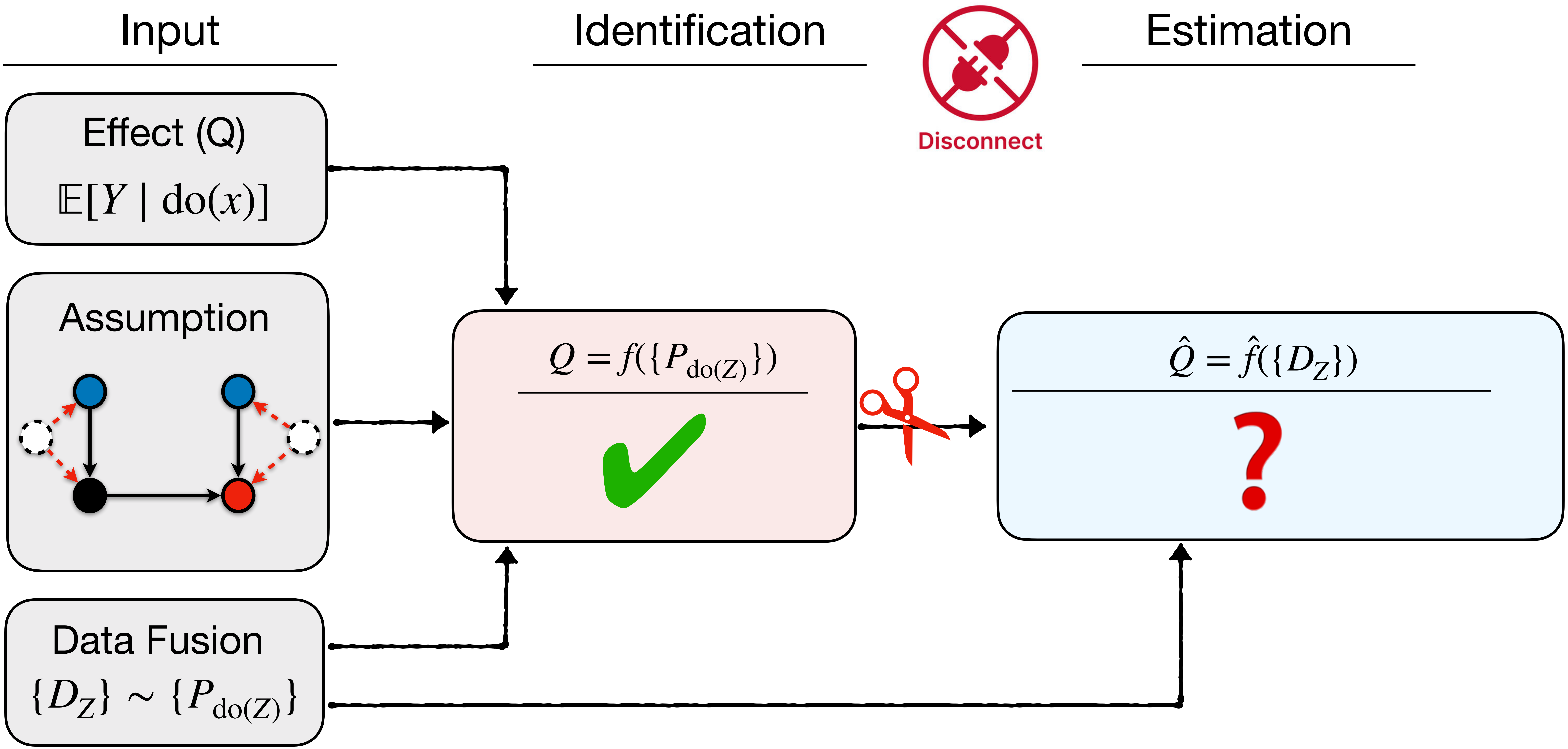
$\hat{\boldsymbol{\mu}}$ misspecified ($\hat{\boldsymbol{\mu}}\neq\boldsymbol{\mu}$)

$\hat{\boldsymbol{\pi}}$ misspecified ($\hat{\boldsymbol{\pi}}\neq\boldsymbol{\pi}$)

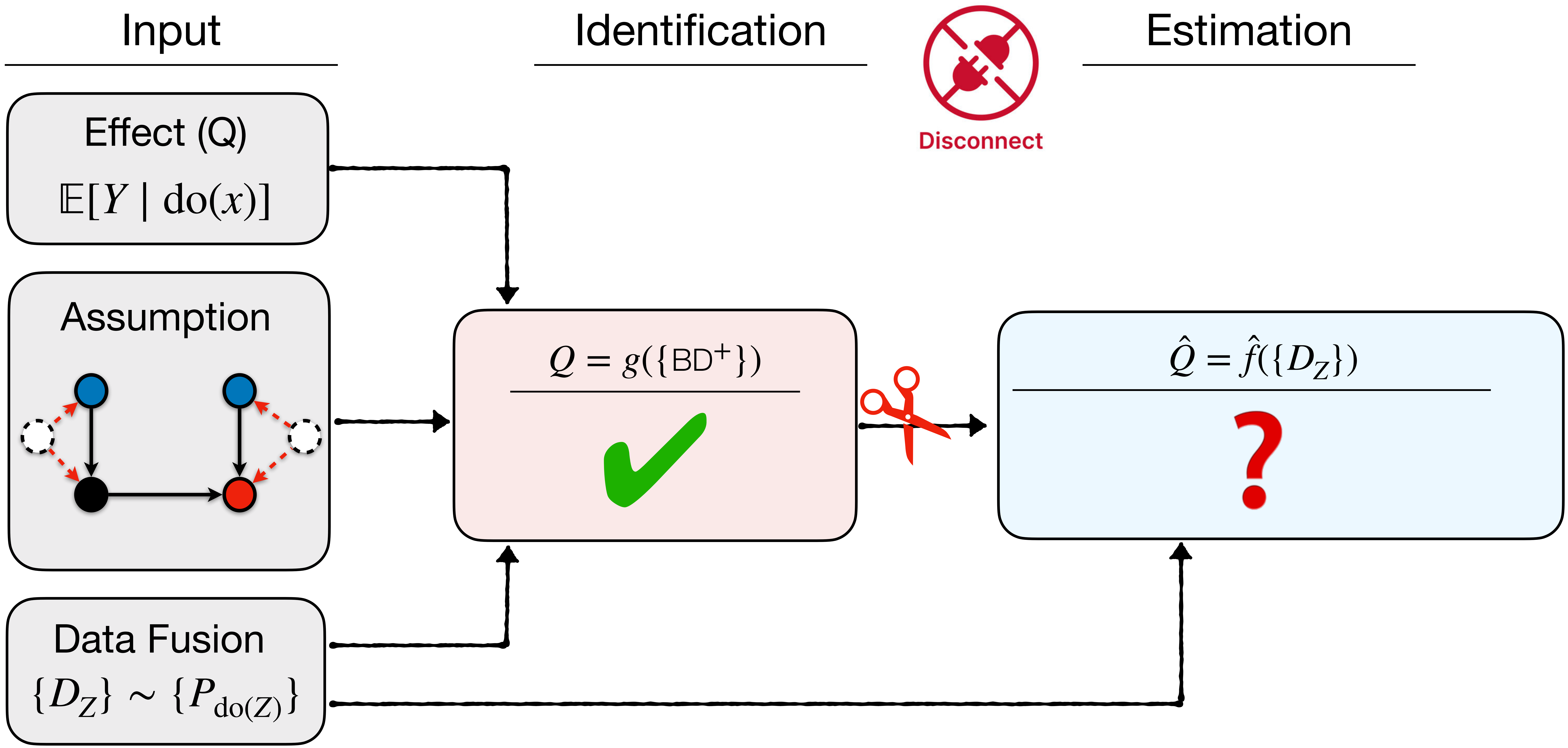


DML-BD+ converges to the true causal effect even when $\hat{\boldsymbol{\mu}}$ or $\hat{\boldsymbol{\pi}}$ are misspecified.

General Estimator for Data Fusion



General Estimator for Data Fusion



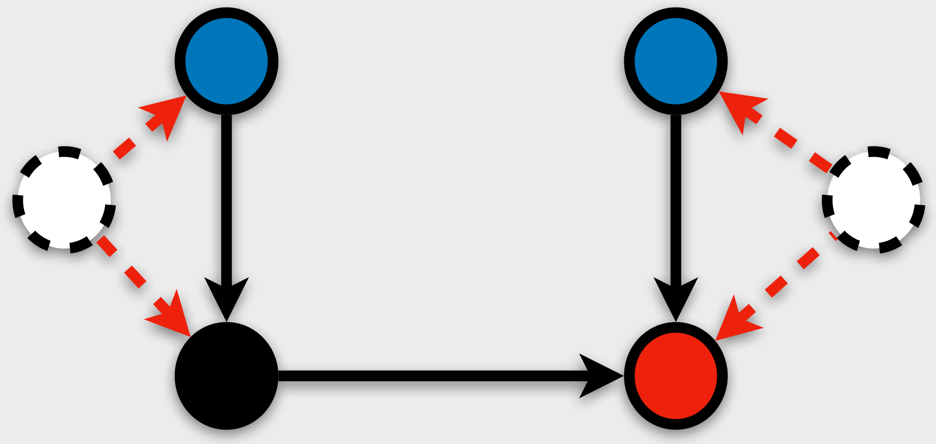
General Estimator for Data Fusion

Input | **Identification** | Connect | **Estimation**

Effect (Q)

$\mathbb{E}[Y \mid do(x)]$

Assumption



Data Fusion

$\{D_Z\} \sim \{P_{do(Z)}\}$

$Q = g(\{\mathrm{BD}^+\})$

✔

Jung et al., NeurIPS 2023

"DML-gID"

$\hat{Q} = g(\widehat{\{\mathrm{BD}^+\}})$

✔

General Estimator for Data Fusion

Theorem

1. Any causal effect identifiable from data-fusion can be expressed as a function of BD+.

2. DML-gID, which is an estimator for any identifiable causal effects from data fusion, achieves double robustness and fast convergence.

General Estimator for Data Fusion

Theorem

1. Any causal effect identifiable from data-fusion can be expressed as a function of BD^{+}.

2. DML-gID, which is an estimator for any identifiable causal effects from data fusion, achieves double robustness and fast convergence.

"*Whenever computable from data fusion,*
We can compute sample-efficiently."

Talk Outline

1. Estimating causal effects from observations

   + its application in healthcare & explainable AI

2. Estimating causal effects from data fusion

3. Unified causal effect estimation method

4. Summary & Future direction

Talk Outline

✔

3 Unified causal effect estimation method

Towards More General Causal Inference Queries

Estimating the interventional effects $\mathbb{E}[Y \mid \mathrm{do}(x)]$

Towards More General Causal Inference Queries

Fairness Analysis

$$\mathbb{E}[Y_{x,M_{\neg x}}]$$

Salary a man would earn if he had the opportunities that other genders would receive

Towards More General Causal Inference Queries

Offline Policy Evaluation

$$\mathbb{E}[Y_{\tau(X|C)}]$$

Recovery rate of a drug dosage policy given baseline characteristics

Towards More General Causal Inference Queries

Joint Treatment Effect

$$\mathbb{E}[Y \mid \mathrm{do}(x_1, x_2)]$$

Effect of drugs x_1 and x_2 from two trials $\mathrm{do}(x_1)$ and $\mathrm{do}(x_2)$, respectively

Towards More General Causal Inference Queries

Retrospection

$$\mathbb{E}[Y_x \mid \neg x]$$

The headache intensity for patients who took aspirin, had they not taken it

Towards More General Causal Inference Queries

Missing Data

$$\mathbb{E}[Y \mid \mathrm{do}(x), \mathrm{mis}{=}0]$$

The effect of a treatment identifiable from missing data

Towards More General Causal Inference Queries

Domain Transfer

$$\mathbb{E}[Y \mid \mathrm{do}(x), S{=}\mathrm{NY}]$$

The effect of a treatment in NY identifiable from trials in Chicago

Towards More General Causal Inference Queries

Towards More General Causal Inference Queries

Fairness Analysis

$$\sum_m \mathbb{E}[Y \mid m, x) P(m \mid \neg x)$$

Domain Transfer

$$\sum_c \mathbb{E}_{do(x)}[Y \mid c, S{=}\text{Chi}] P(c \mid S{=}\text{NY})$$

Offline Policy Evaluation

$$\sum_c \mathbb{E}[Y \mid c, x) \pi(x \mid c) P(c)$$

Missing Data

$$\sum_c \mathbb{E}[Y \mid x, c, \text{mis}{=}1] P(c \mid \text{mis}{=}1)$$

Joint Treatment Effect

$$\sum_w \mathbb{E}_{do(x_2)}[Y \mid x_1, c] P_{do(x_2)}(w)$$

Retrospection

$$\sum_c \mathbb{E}[Y \mid c, x) P(c \mid \neg x)$$

Towards More General Causal Inference Queries

Fairness Analysis

Estimator 1

Domain Transfer

Estimator 6

Offline Policy Evaluation

Estimator 2

Missing Data

Estimator 5

Joint Treatment Effect

Estimator 3

Retrospection

Estimator 4

Towards More General Causal Inference Queries

Jung et al., NeurIPS 2024

Unified Covariate Adjustment (UCA)

Unified causal estimation for summation of the product of arbitrary conditional distributions

Unified Covariate Adjustment (UCA)

$$\sum_{x,c} \mathbb{E}_{P_2}[Y \mid x, c]\tau(x \mid c)P_1(c)$$

Unified Covariate Adjustment (UCA)

arbitrary distributions

arbitrary distributions

$$\sum_{x,c} \mathbb{E}_{P_2}[Y \mid x, c]\tau(x \mid c)P_1(c)$$

Unified Covariate Adjustment (UCA)

arbitrary distributions

outcome

arbitrary distributions

$$\sum_{x,c} \mathbb{E}_{P_2}[Y \mid x, c]\tau(x \mid c)P_1(c)$$

Unified Covariate Adjustment (UCA)

arbitrary distributions

outcome

arbitrary policy of treatments

arbitrary distributions

$$\sum_{x,c} \mathbb{E}_{P_2}[Y \mid x, c]\tau(x \mid c)P_1(c)$$

Unified Covariate Adjustment (UCA)

arbitrary distributions

outcome

arbitrary policy of treatments

arbitrary distributions

set of variables

$$\sum_{x,c} \mathbb{E}_{P_2}[Y \mid x, c]\tau(x \mid c)P_1(c)$$

Unified Covariate Adjustment (UCA)

$$\underbrace{\sum_{x,c}}_{\text{set of variables}}\mathbb{E}_{\underbrace{P_2}_{\text{arbitrary distributions}}}[\underbrace{Y}_{\text{outcome}} \mid x, c]\,\underbrace{\tau(x \mid c)}_{\text{arbitrary policy of treatments}}\,\underbrace{P_1}_{\text{arbitrary distributions}}(c)$$

UCA

$$\boxed{\sum_{x,c}\mathbb{E}_{P_2}[Y \mid x, c]\tau(x \mid c)P_1(c)}$$

Unified Covariate Adjustment (UCA)

arbitrary distributions

outcome

arbitrary policy of treatments

arbitrary distributions

set of variables

$$\sum_{x,c}\mathbb{E}_{P_2}[Y \mid x, c]\tau(x \mid c)P_1(c)$$

UCA

$$\sum_{x,c}\mathbb{E}_{P_2}[Y \mid x, c]\tau(x \mid c)P_1(c)$$

BD+

$$\sum_{w}\mathbb{E}_{P_{\mathrm{do}(x_2)}}[Y \mid x, w]P_{\mathrm{do}(x_1)}(w)$$

$P_1 \leftarrow P_{\mathrm{do}(x_1)}$

$P_2 \leftarrow P_{\mathrm{do}(x_2)}$

$\tau \leftarrow \mathbb{1}_x(X)$

$C \leftarrow W$

Unified Covariate Adjustment (UCA)

arbitrary distributions

outcome

arbitrary policy of treatments

arbitrary distributions

set of variables

$$\sum_{x,c}\mathbb{E}_{P_2}[Y \mid x, c]\tau(x \mid c)P_1(c)$$

Theorem

UCA can represent ***any*** causal effects expressible as a sum of products of arbitrary conditional distributions, by choosing $C, P_1, P_2, \tau(\cdot \mid \cdot)$ properly.

$C \leftarrow W$

Doubly Robust Estimator for UCA

Doubly Robust Estimator for UCA

UCA Parametrization

$$\text{UCA}(\mu, \pi) \triangleq \mathbb{E}_{P_2}[\mu \times \pi]$$

where

- $\mu(XC) \triangleq \mathbb{E}_{P_2}[Y \mid X, C]$
- $\pi(XC) \triangleq \dfrac{\tau(X \mid C)\, P_1(C)}{P_2(X \mid C)\, P_2(C)}$

Doubly Robust Estimator for UCA

UCA Parametrization

$$\text{UCA}(\mu,\pi) \triangleq \mathbb{E}_{P_2}[\mu \times \pi]$$

where

- $\mu(XC) \triangleq \mathbb{E}_{P_2}[Y \mid X, C]$
- $\pi(XC) \triangleq \dfrac{\tau(X \mid C)\ P_1(C)}{P_2(X \mid C) P_2(C)}$

Theorem

DML-UCA($\hat{\mu},\hat{\pi}$) achieves the followings:

- **Double Robustness:** Error $= 0$ if either $\hat{\mu} = \mu$ or $\hat{\pi} = \pi$.
- **Fast Convergence:** Error $\to 0$ *fast* even when $\hat{\mu} \to \mu$ and $\hat{\pi} \to \pi$ slow.

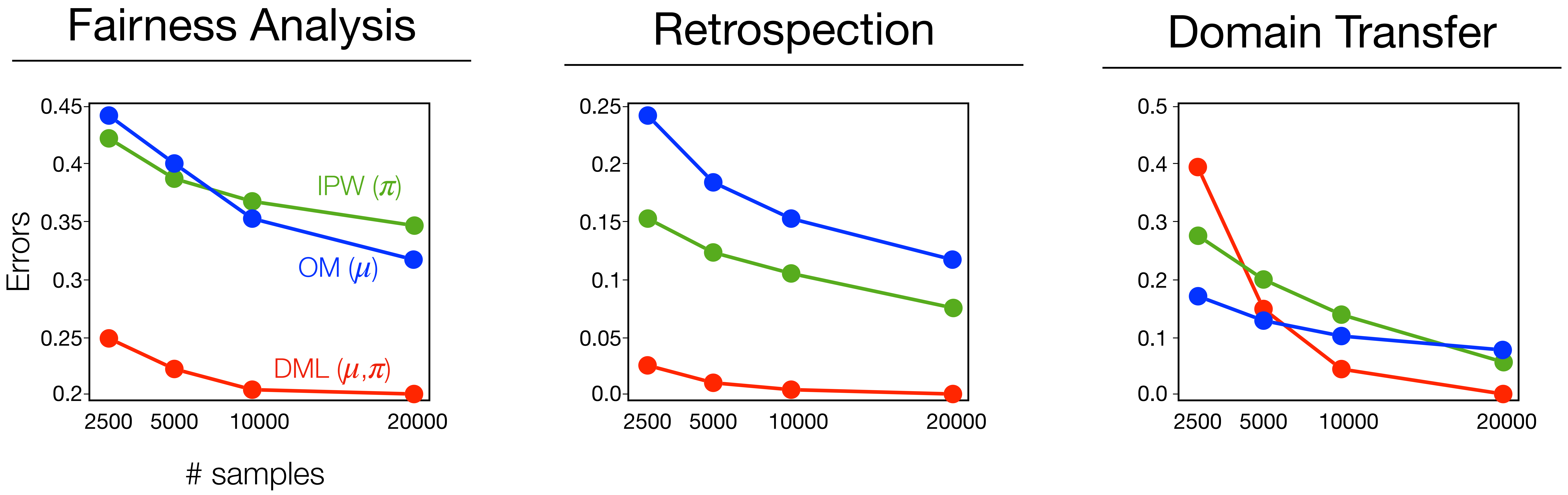
Simulation: DML-UCA

Simulation: DML-UCA

$(\hat{\boldsymbol{\mu}},\hat{\boldsymbol{\pi}}) \to (\boldsymbol{\mu}_0, \boldsymbol{\pi}_0)$ slowly

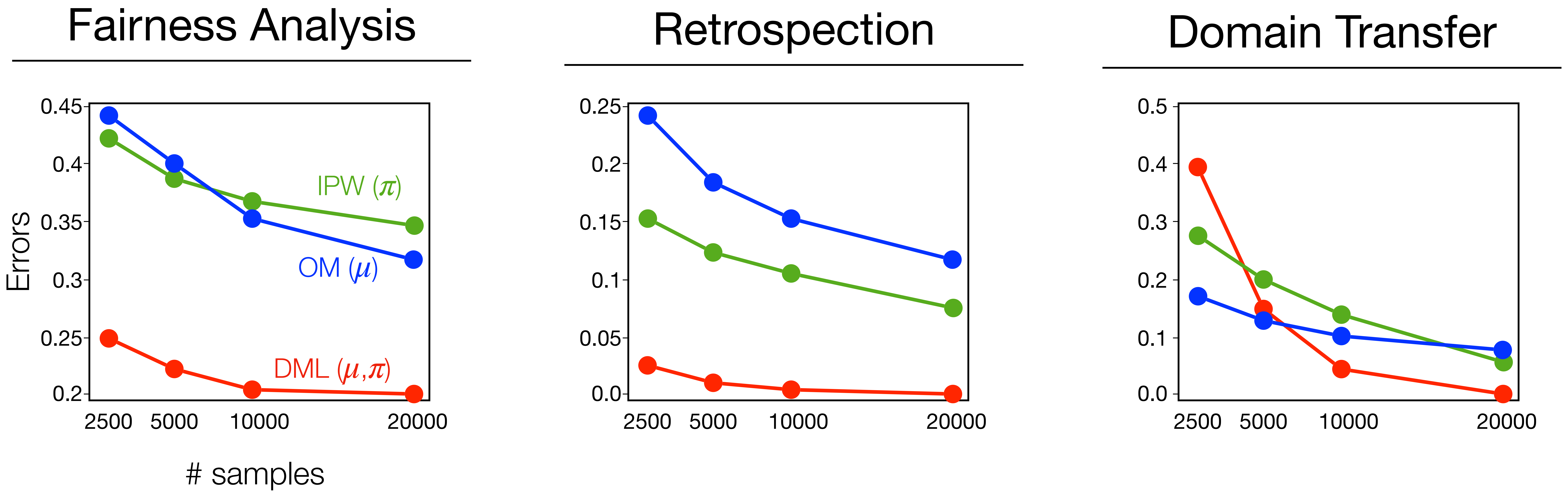
Simulation: DML-UCA

$$(\hat{\boldsymbol{\mu}}, \hat{\boldsymbol{\pi}}) \rightarrow (\boldsymbol{\mu}_0, \boldsymbol{\pi}_0) \text{ slowly}$$



Simulation: DML-UCA

$(\hat{\mu}, \hat{\pi}) \rightarrow (\mu_0, \pi_0)$ slowly



DML-UCA converges fast even when $(\hat{\mu}, \hat{\pi}) \rightarrow (\mu_0, \pi_0)$ slowly

Talk Outline

1. Estimating causal effects from observations

   + its application in healthcare & explainable AI

2. Estimating causal effects from data fusion

3. Unified causal effect estimation method

4. Summary & Future direction

Talk Outline

✔

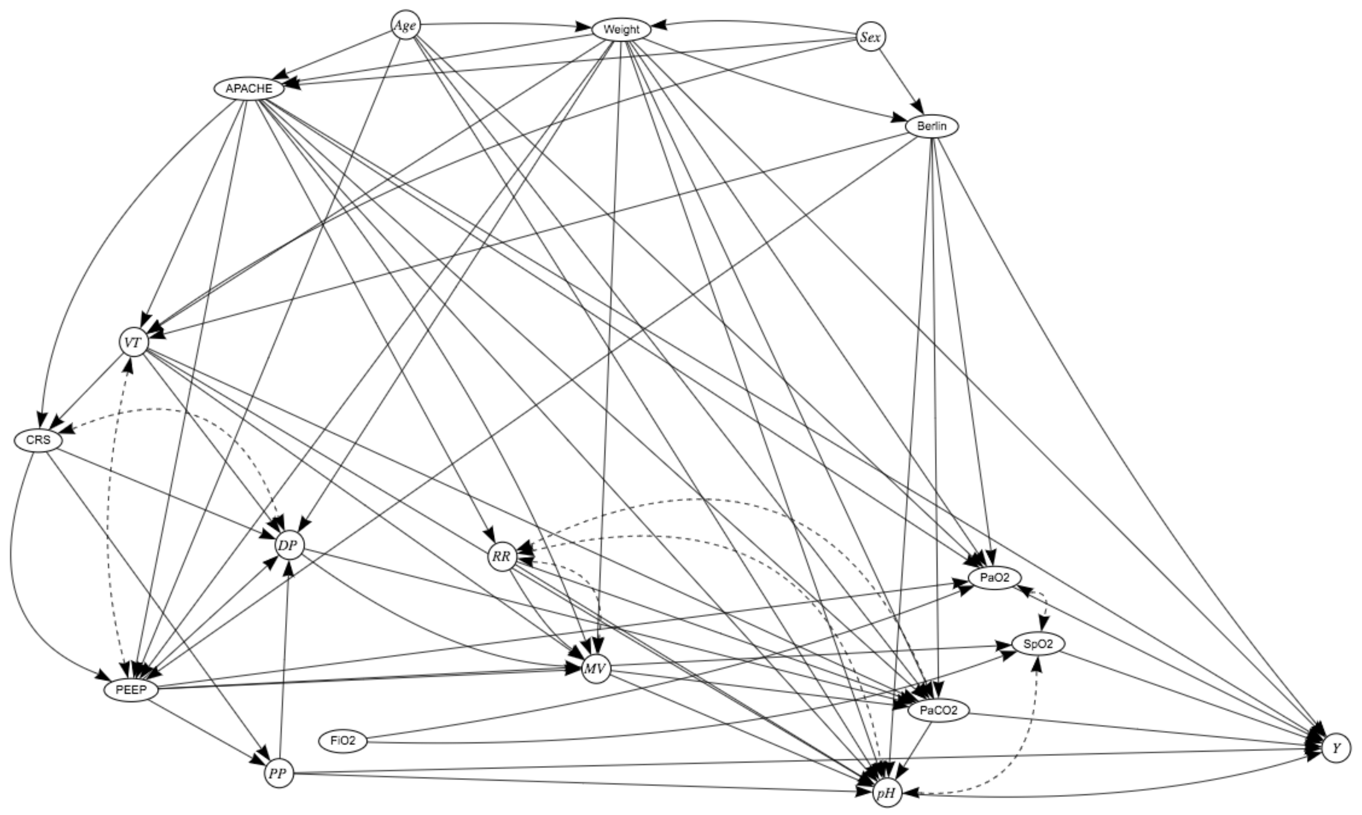
4 Summary & Future direction

This Talk: Estimating Causal Effects

This Talk: Estimating Causal Effects

Tasks

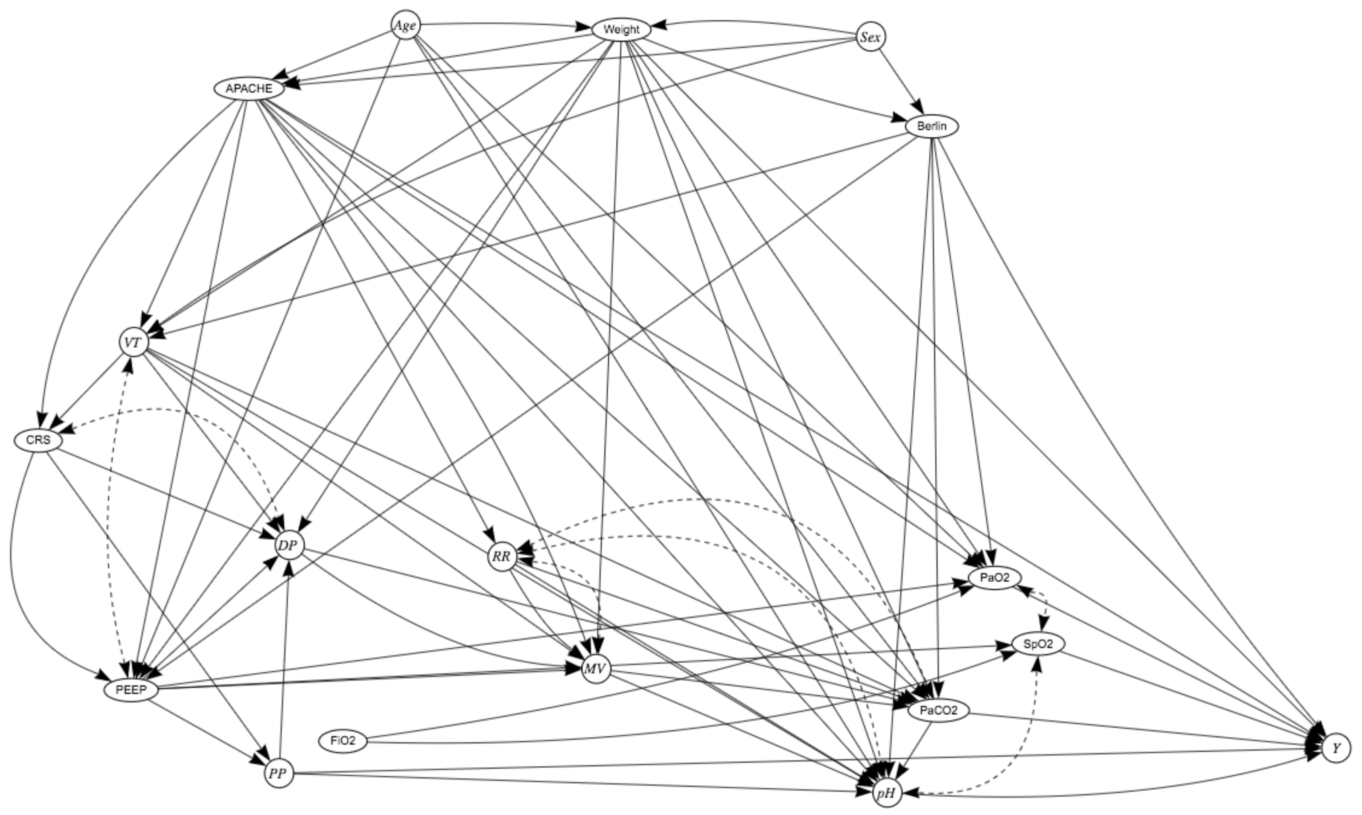
1. From Observation



This Talk: Estimating Causal Effects

Tasks

1. From Observation



Solution

DML-ID

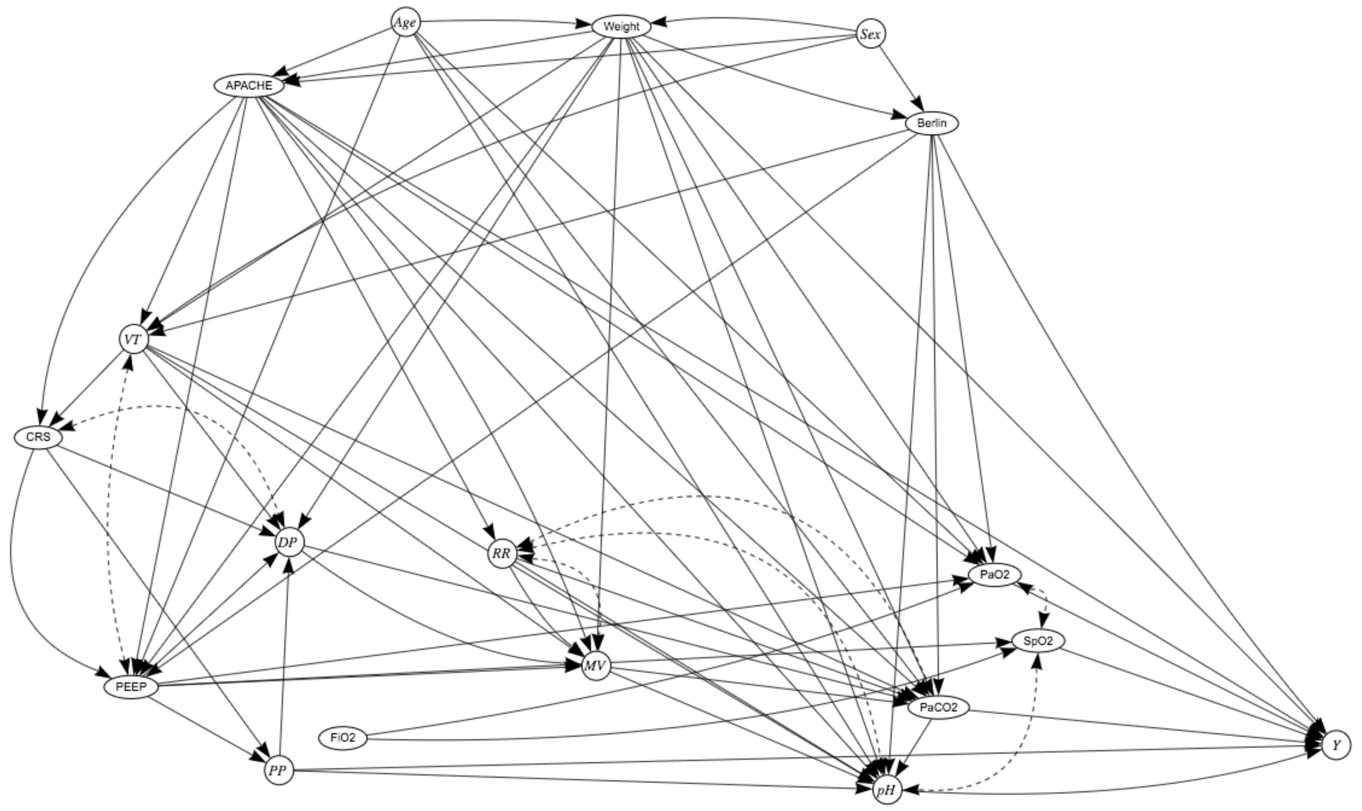
+ application to

- Healthcare
- Explainable AI

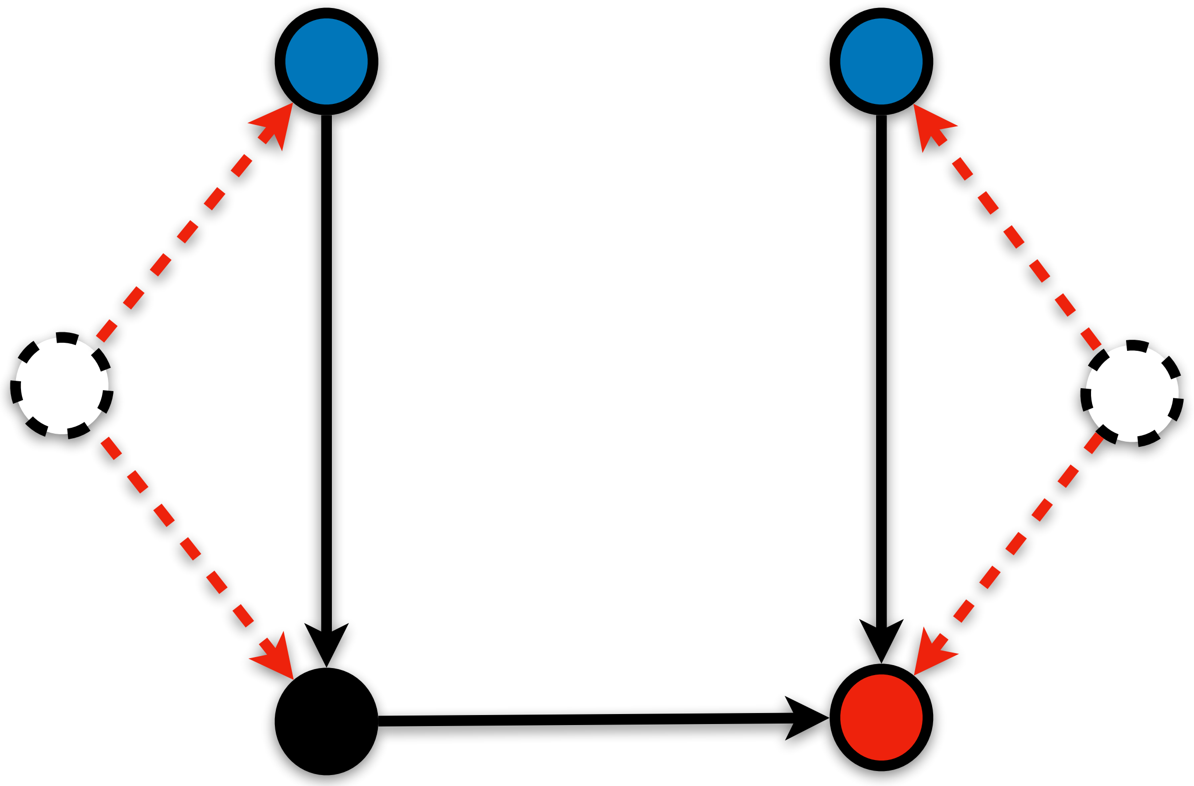
This Talk: Estimating Causal Effects

Tasks

1. From Observation



2. From Data Fusion



Solution

DML-ID

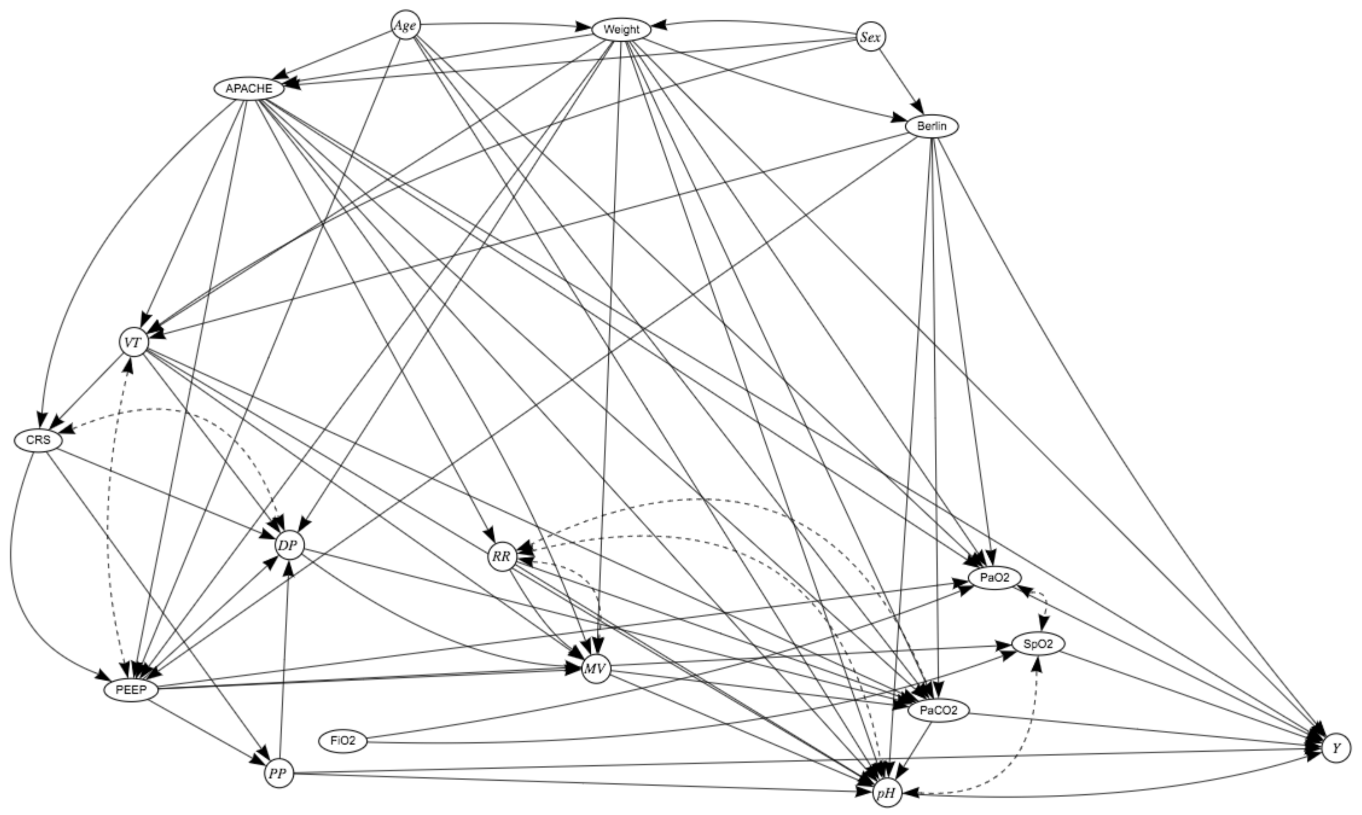
+ application to

- Healthcare
- Explainable AI

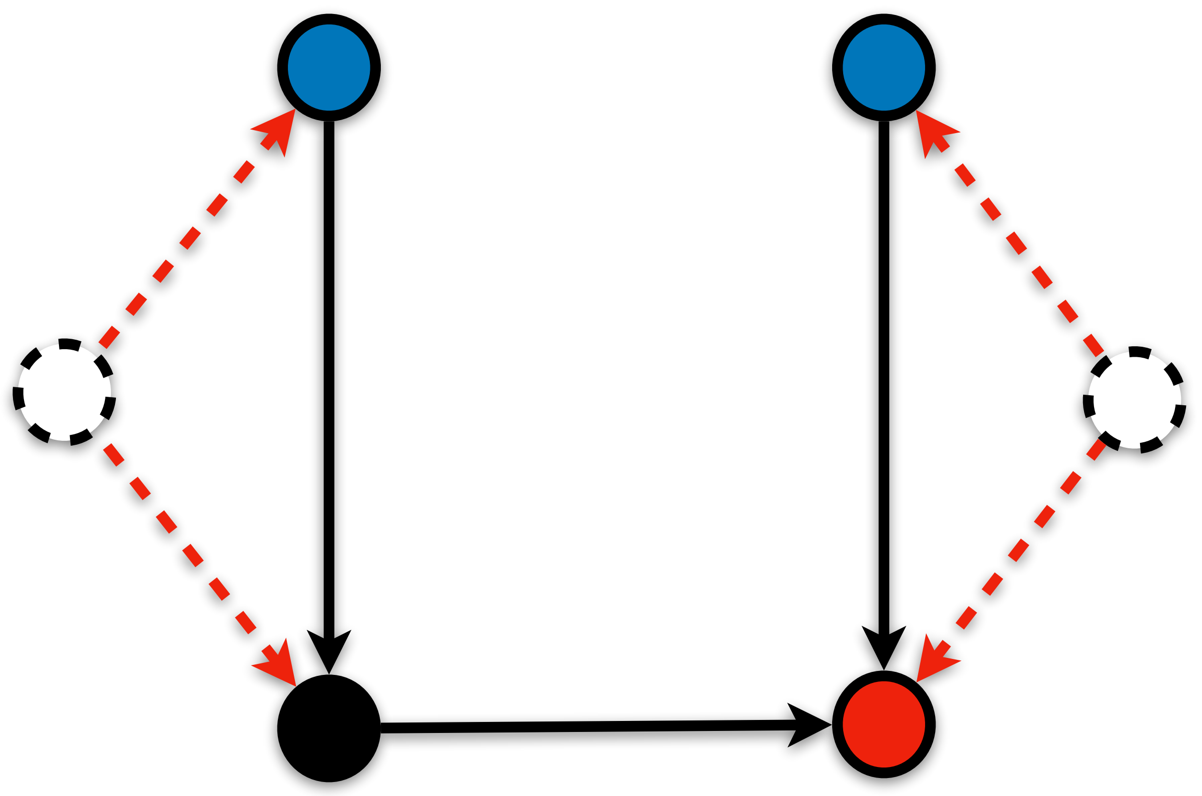
This Talk: Estimating Causal Effects

Tasks

1. From Observation



2. From Data Fusion



Solution

DML-ID

+ application to

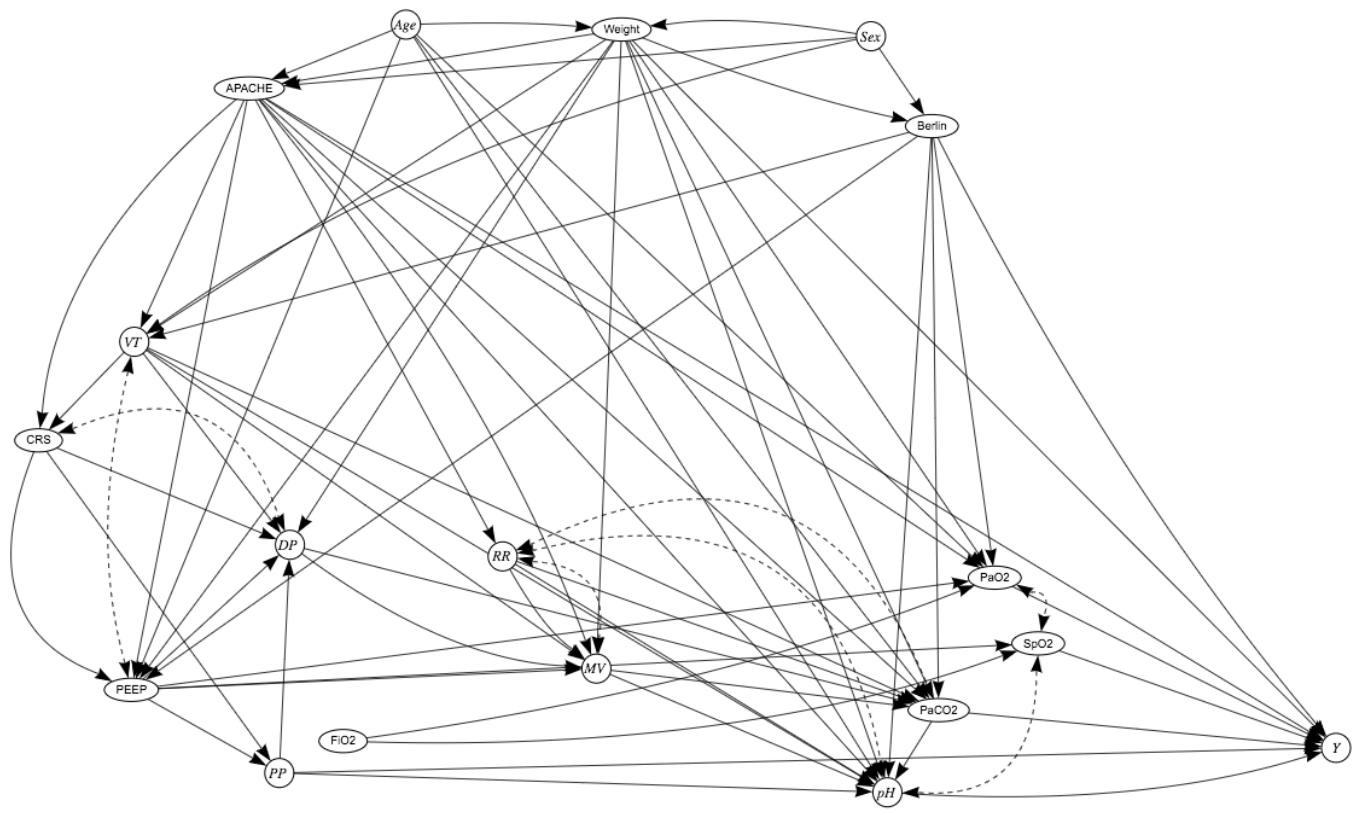
- Healthcare
- Explainable AI

- DML-BD^{+}
- DML-gID

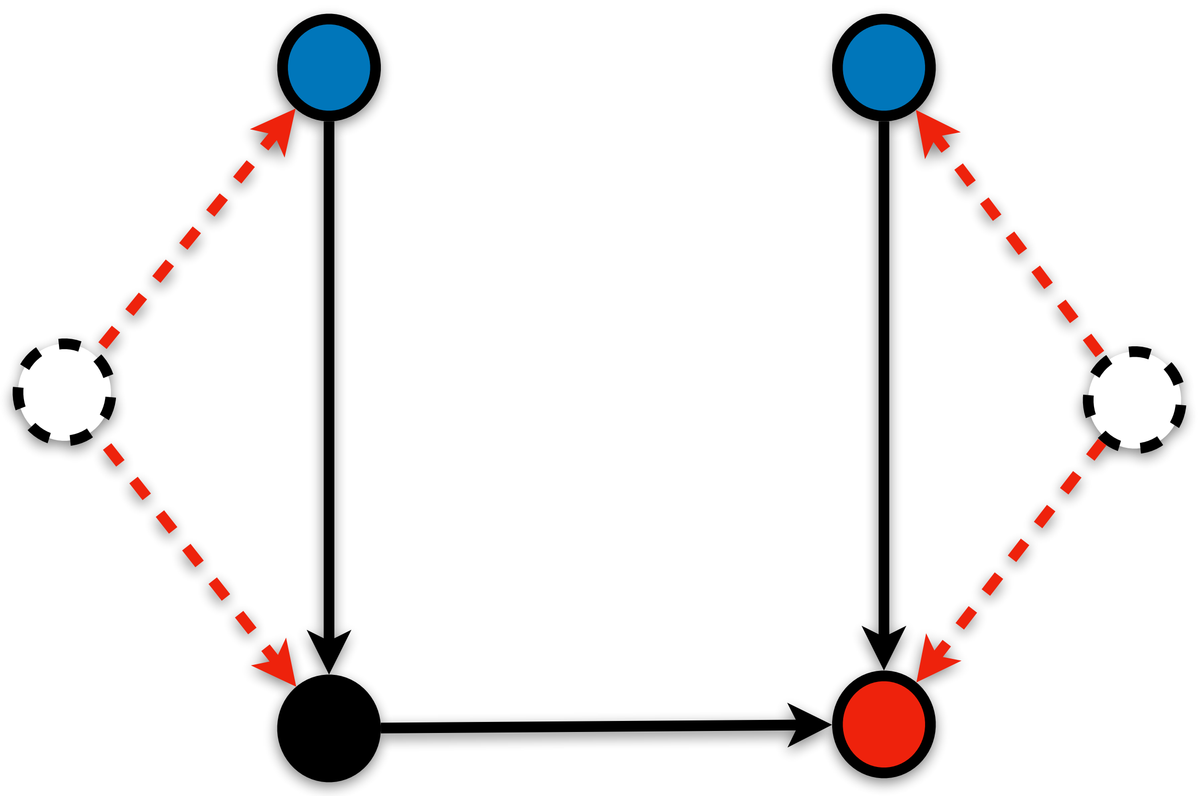
This Talk: Estimating Causal Effects

Tasks

1. From Observation



2. From Data Fusion



3. Unified Estimation

Fairness	$\mathbb{E}[Y_{x,M_{\neg x}}]$
Off-policy evaluation	$\mathbb{E}[Y_{\tau(X\|C)}]$
Counterfactuals	$\mathbb{E}[Y_x \| \neg x]$

...

Solution

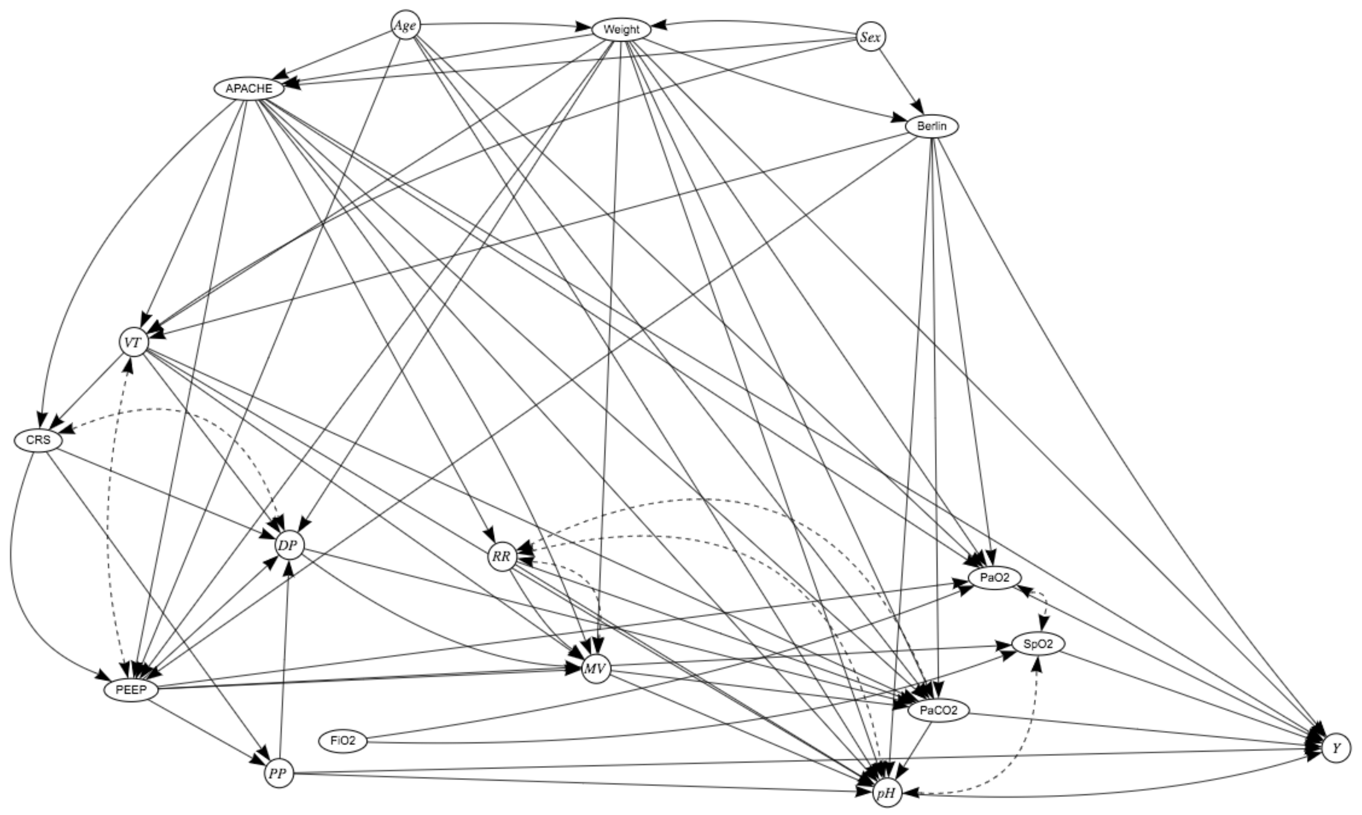
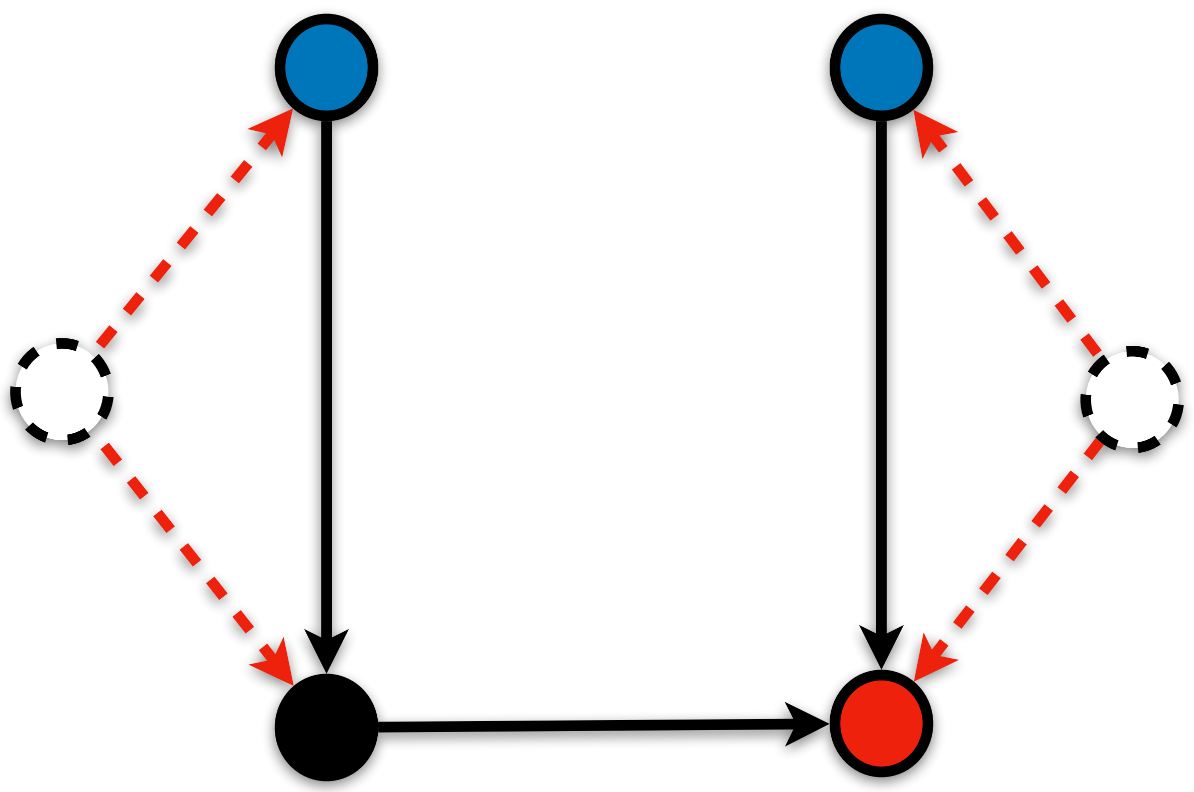
DML-ID

+ application to

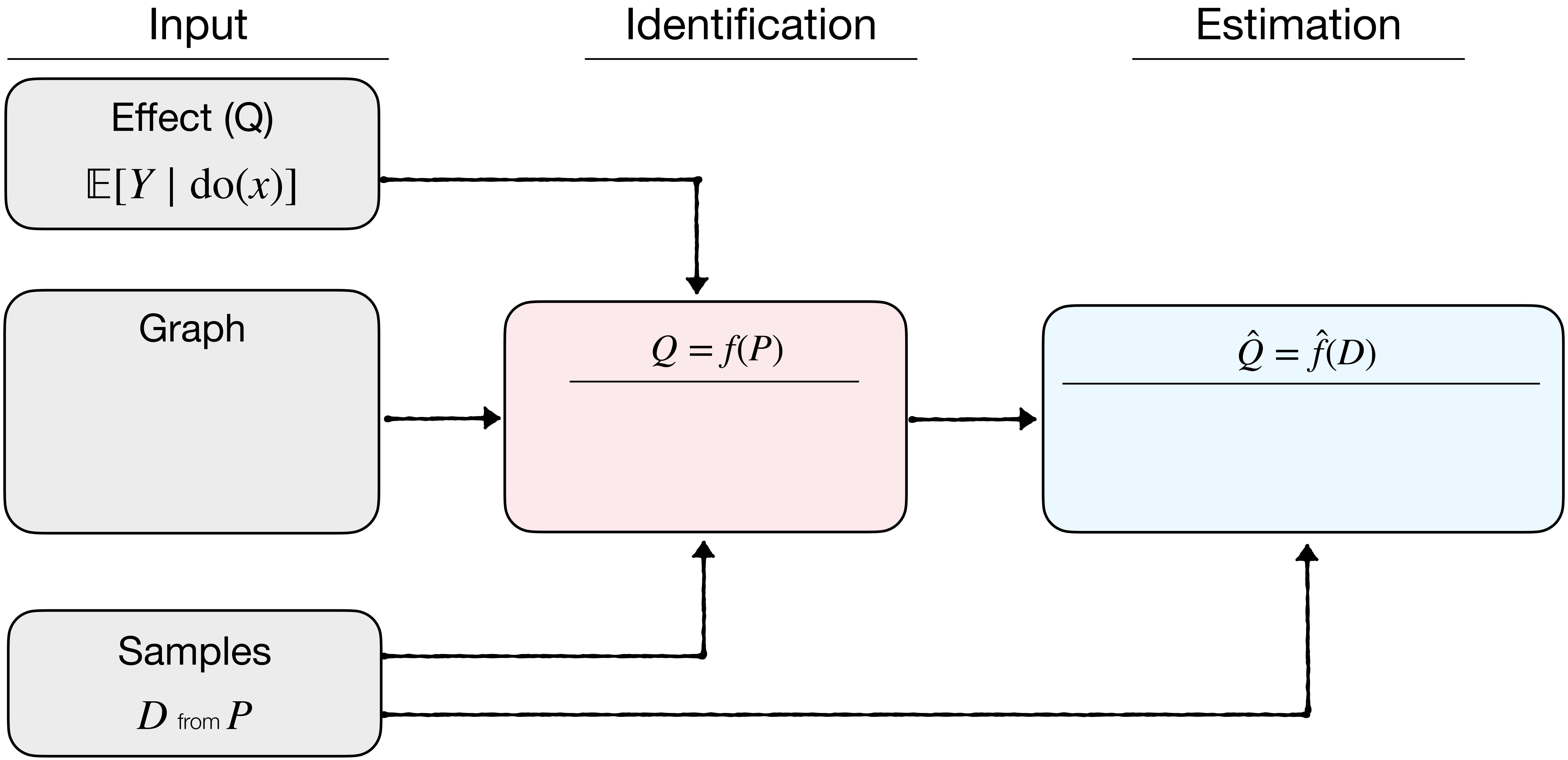
- Healthcare
- Explainable AI

- DML-BD+
- DML-gID

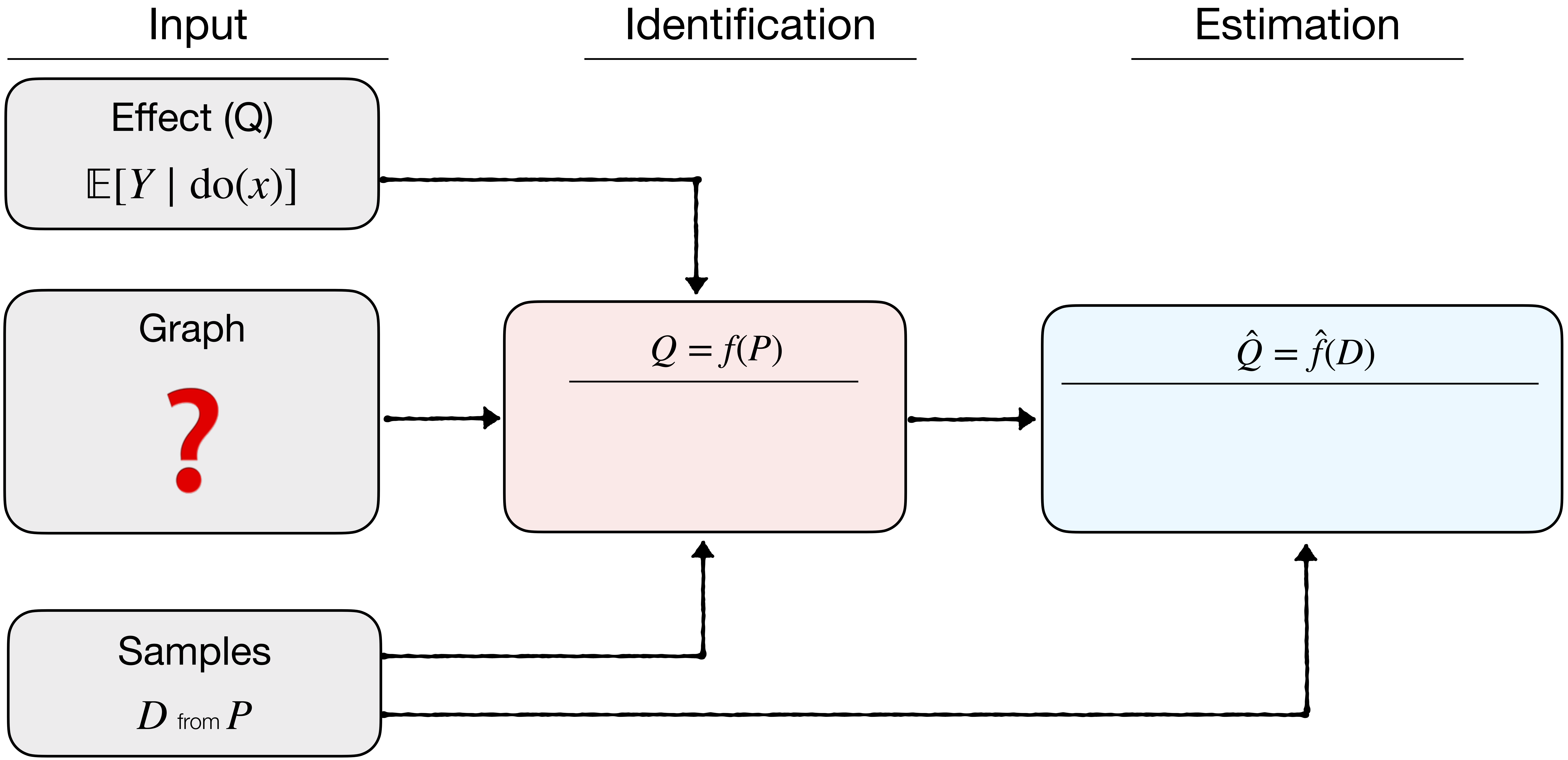
This Talk: Estimating Causal Effects

	1. From Observation	2. From Data Fusion	3. Unified Estimation
Tasks			Fairness $\mathbb{E}[Y_{x,M_{\neg x}}]$ Off-policy evaluation $\mathbb{E}[Y_{\tau(X\|C)}]$ Counterfactuals $\mathbb{E}[Y_x \| \neg x]$ …
Solution	DML-ID + application to • Healthcare • Explainable AI	• DML-BD+ • DML-gID	DML-UCA

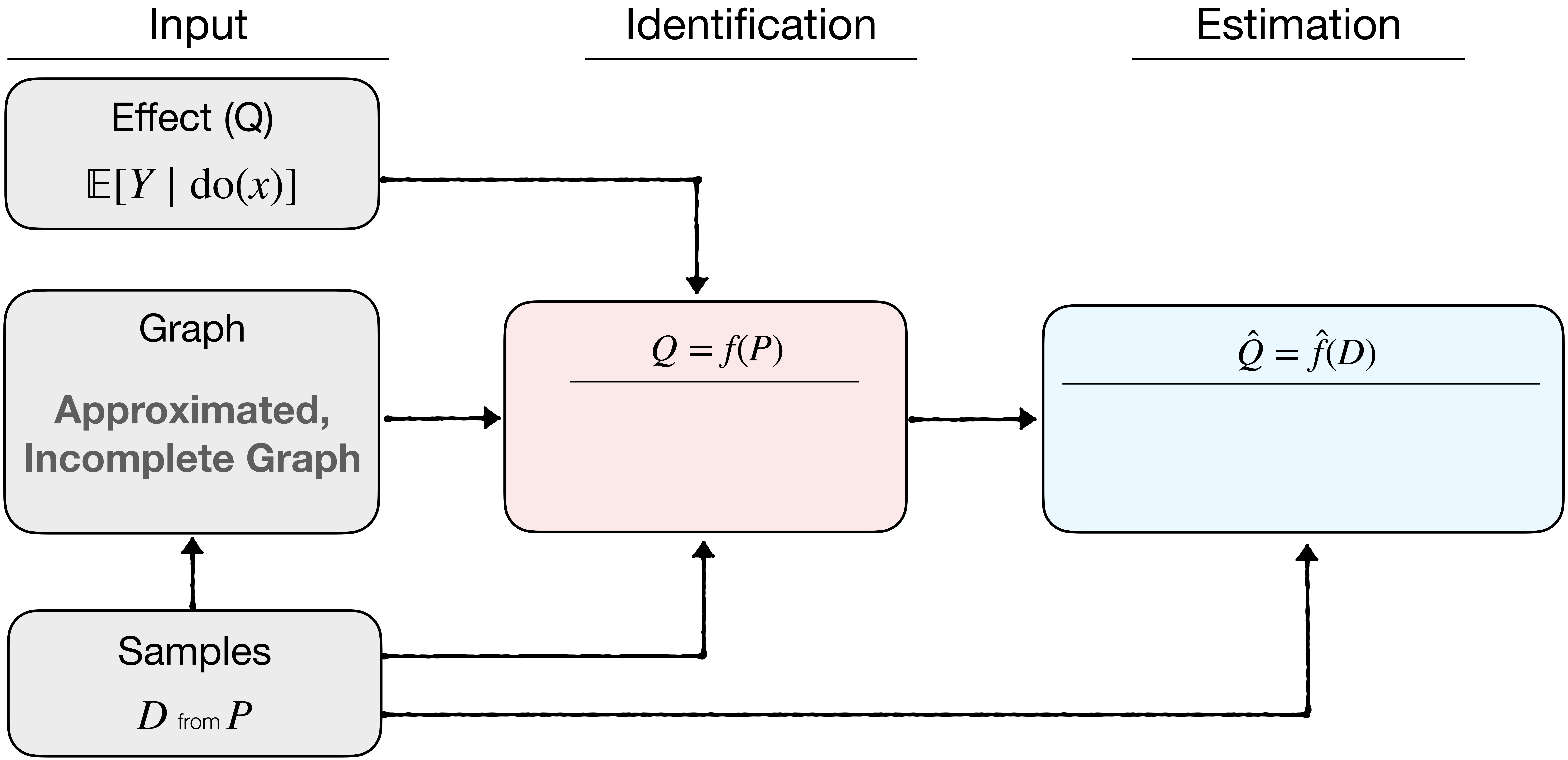
Other Work 1: Causal inference Without Graphs



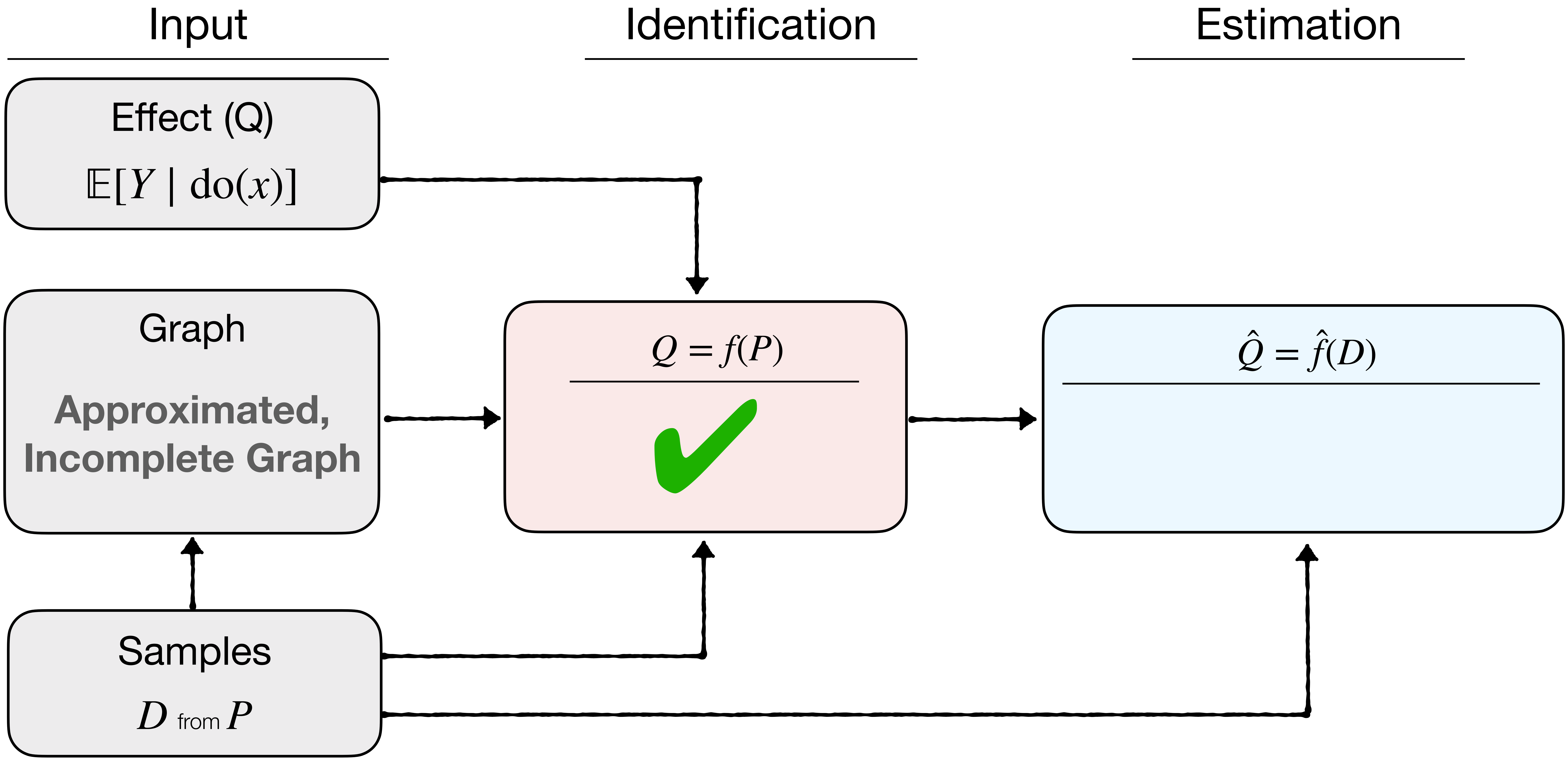
Other Work 1: Causal inference Without Graphs



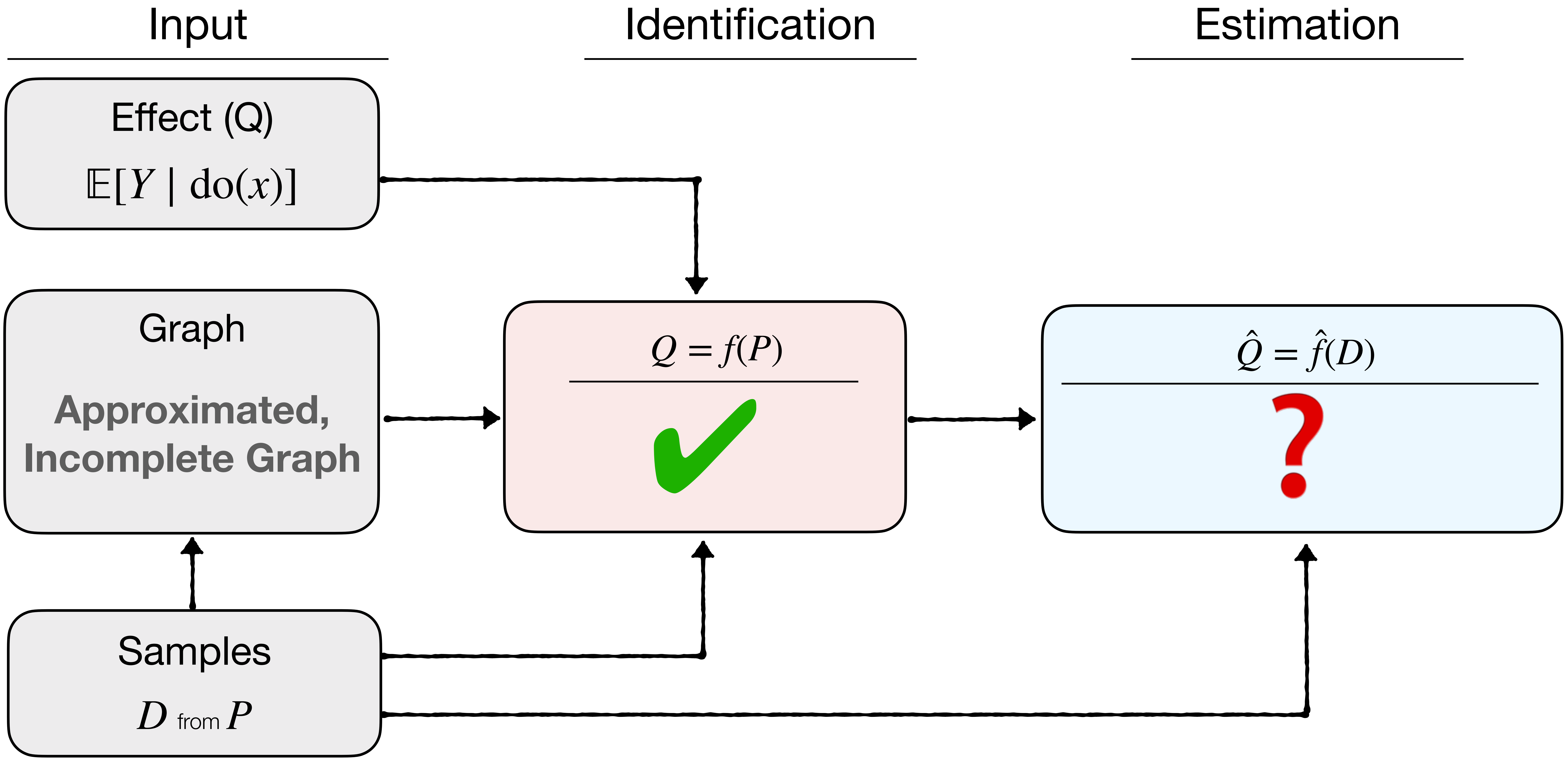
Other Work 1: Causal inference Without Graphs



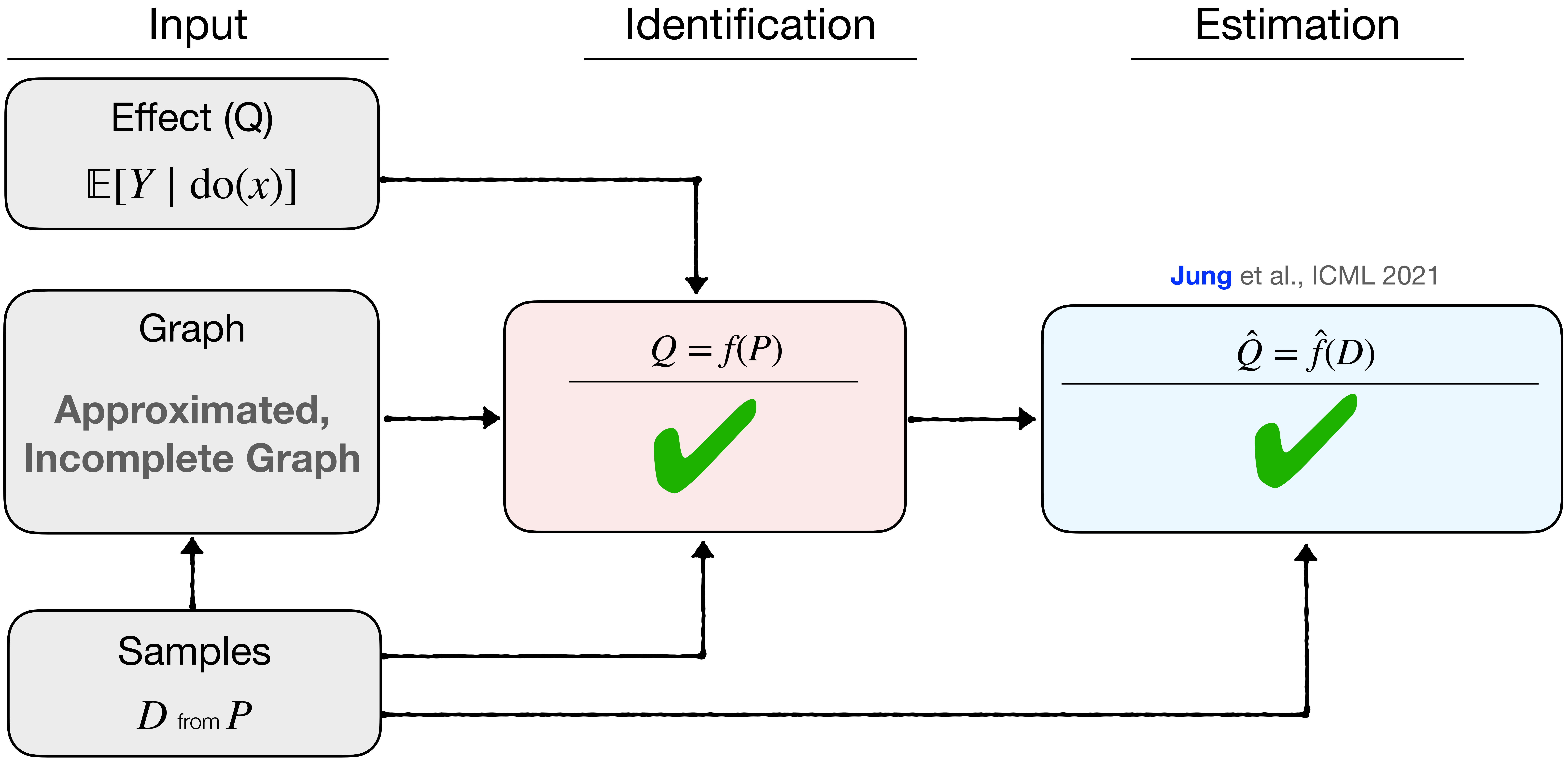
Other Work 1: Causal inference Without Graphs



Other Work 1: Causal inference Without Graphs



Other Work 1: Causal inference Without Graphs



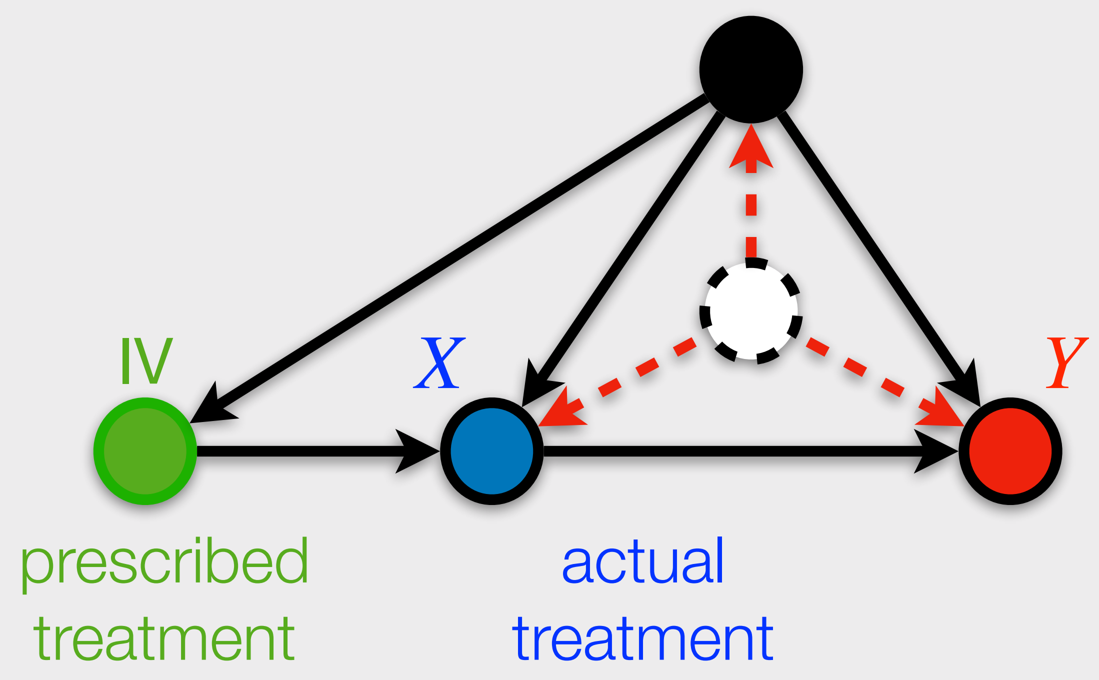
Other Work 2: Instrumental Variable (IV) Analysis

Input	Identification	Estimation

Effect (Q)

$\mathbb{E}[Y \mid \mathrm{do}(x), \mathrm{complier}]$

Graph



Samples

D from P

$Q = f(P)$

$\hat{Q} = \hat{f}(D)$

Other Work 2: Instrumental Variable (IV) Analysis

Input | **Identification** | **Estimation**

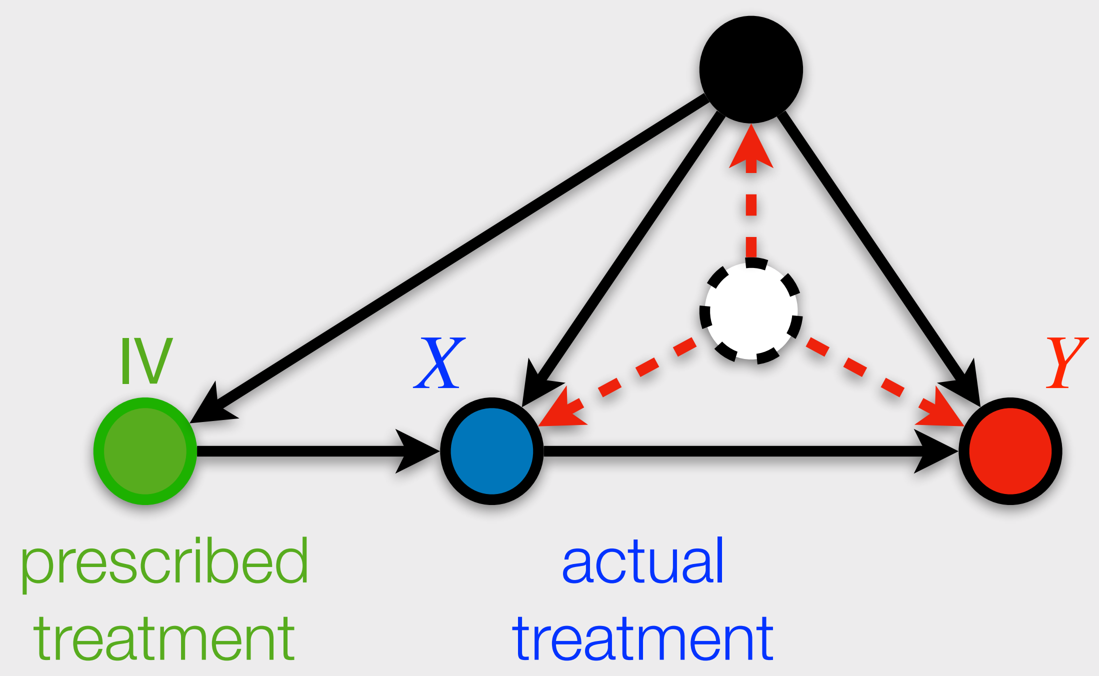
Input

Effect (Q)

$\mathbb{E}[Y \mid \mathrm{do}(x), \mathrm{complier}]$

Compliers: samples that prescribed treatments (IV) = actual treatment (X)

Graph



Samples

D from P

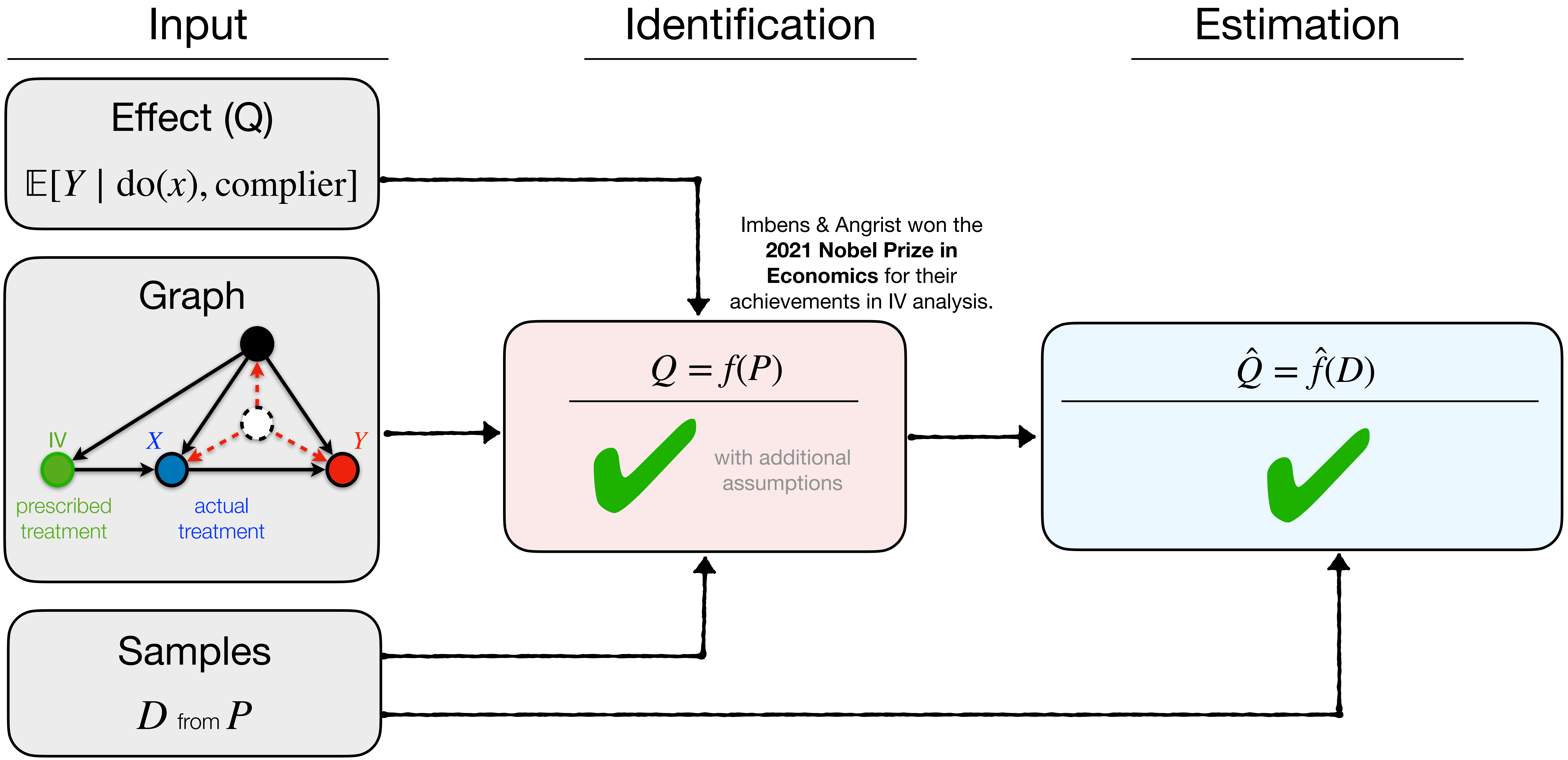
Identification

$Q = f(P)$

Estimation

$\hat{Q} = \hat{f}(D)$

Other Work 2: Instrumental Variable (IV) Analysis



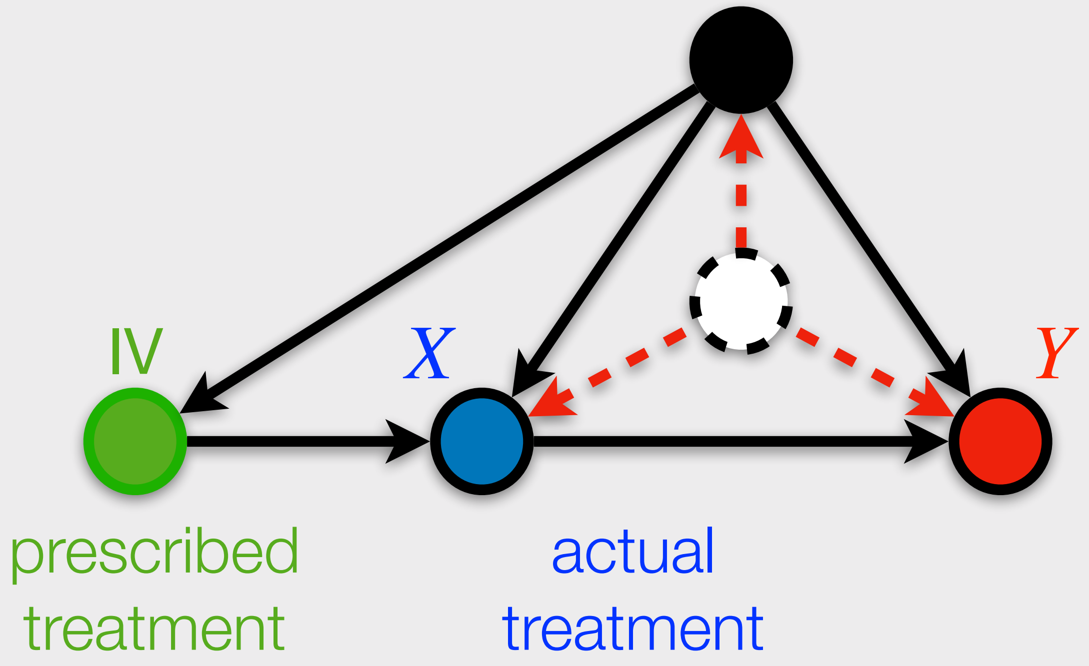
Other Work 2: Instrumental Variable (IV) Analysis

Input

Effect (Q)

$P(Y \mid \mathrm{do}(x), \mathrm{complier})$

Graph



Samples

D from P

Identification

Imbens & Angrist won the **2021 Nobel Prize in Economics** for their achievements in IV analysis.

$$Q = f(P)$$

✔ with additional assumptions

Estimation

$$\hat{Q} = \hat{f}(D)$$

?

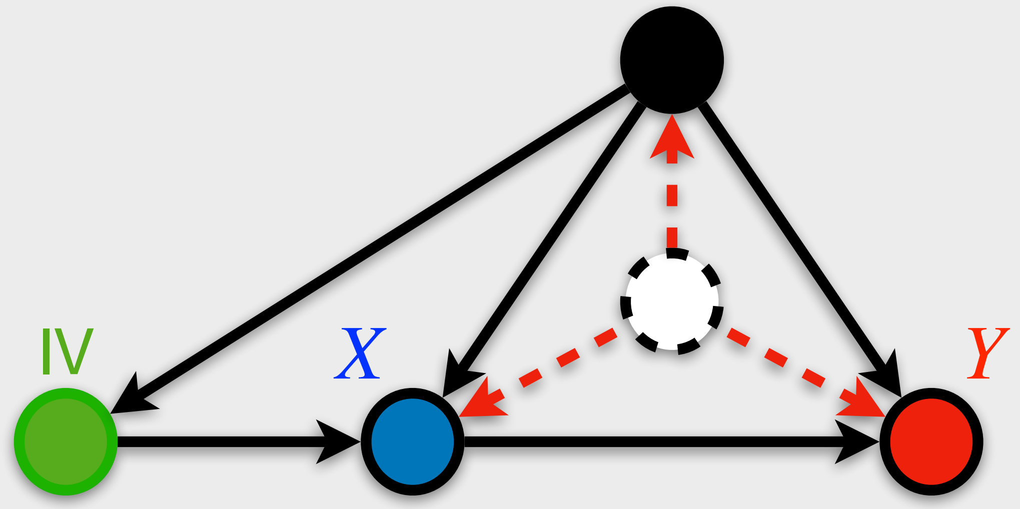
Other Work 2: Instrumental Variable (IV) Analysis

Input

Effect (Q)

$P(Y \mid \mathrm{do}(x), \text{complier})$

Graph



Samples

D from P

Identification

Imbens & Angrist won the **2021 Nobel Prize in Economics** for their achievements in IV analysis.

$Q = f(P)$

✔ with additional assumptions

Estimation

⭐ **Jung** et al., NeurIPS 2021 (Spotlight)

$\hat{Q} = \hat{f}(D)$

✔

Develop robust estimation methods for causal effects **across diverse scenarios**

Develop robust estimation methods for causal effects **across diverse scenarios**

Identification

"When is the causal effect computable from data?”

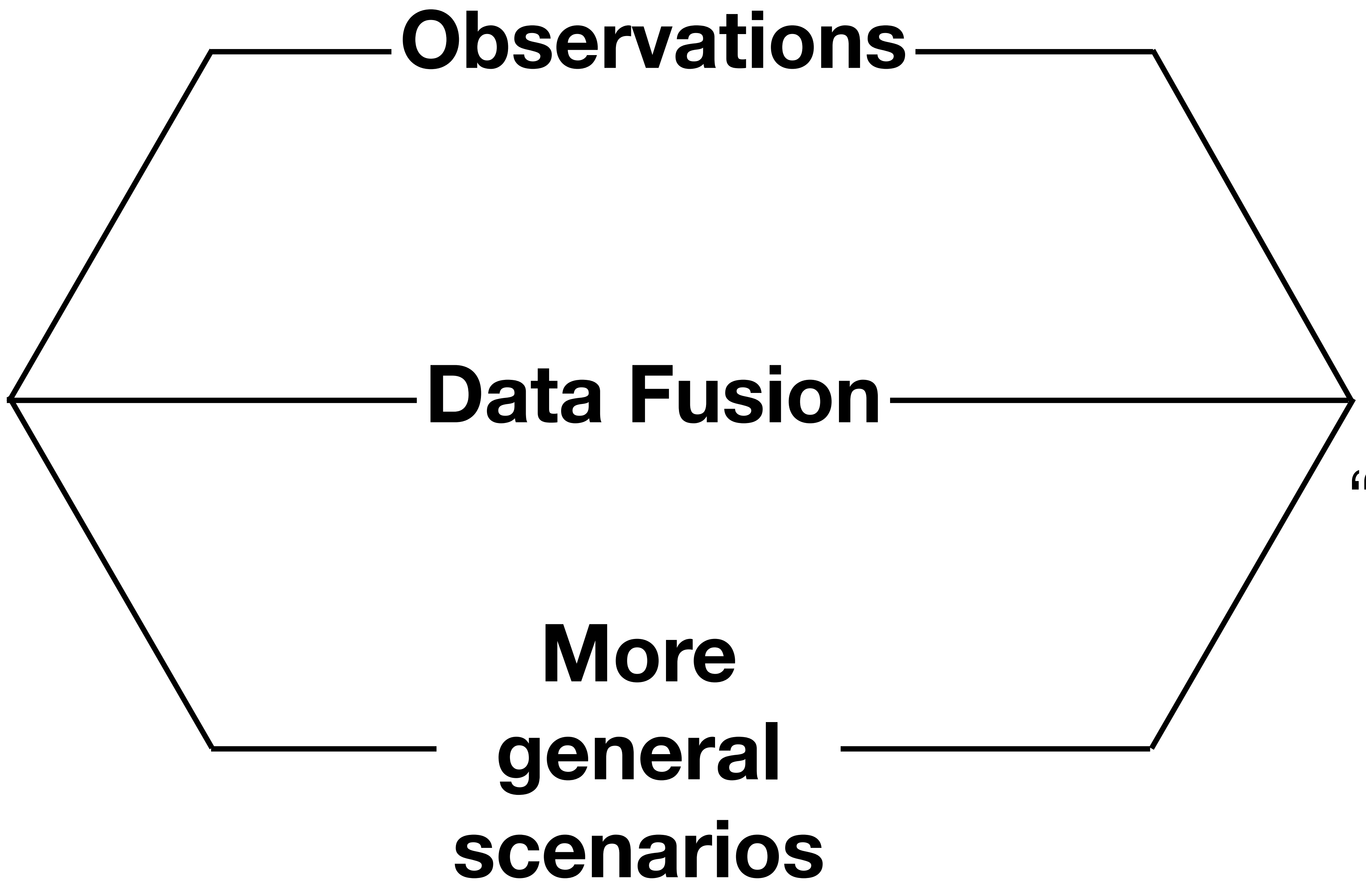
Estimation

“How do we compute the effect from data?”

Develop robust estimation methods for causal effects **across diverse scenarios**

Identification

"When is the causal effect computable from data?”



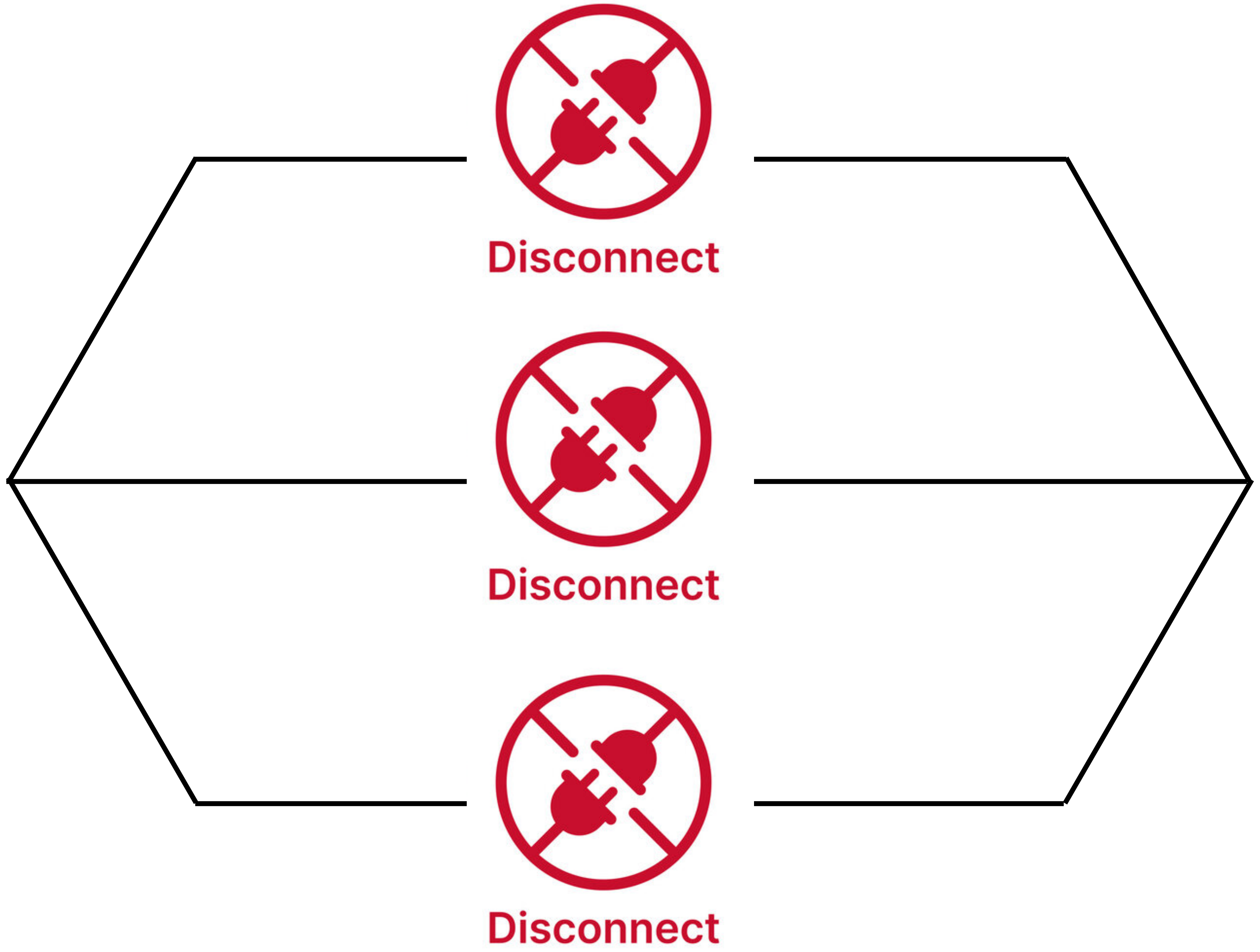
Estimation

“How do we compute the effect from data?”

Develop robust estimation methods for causal effects **across diverse scenarios**

Identification

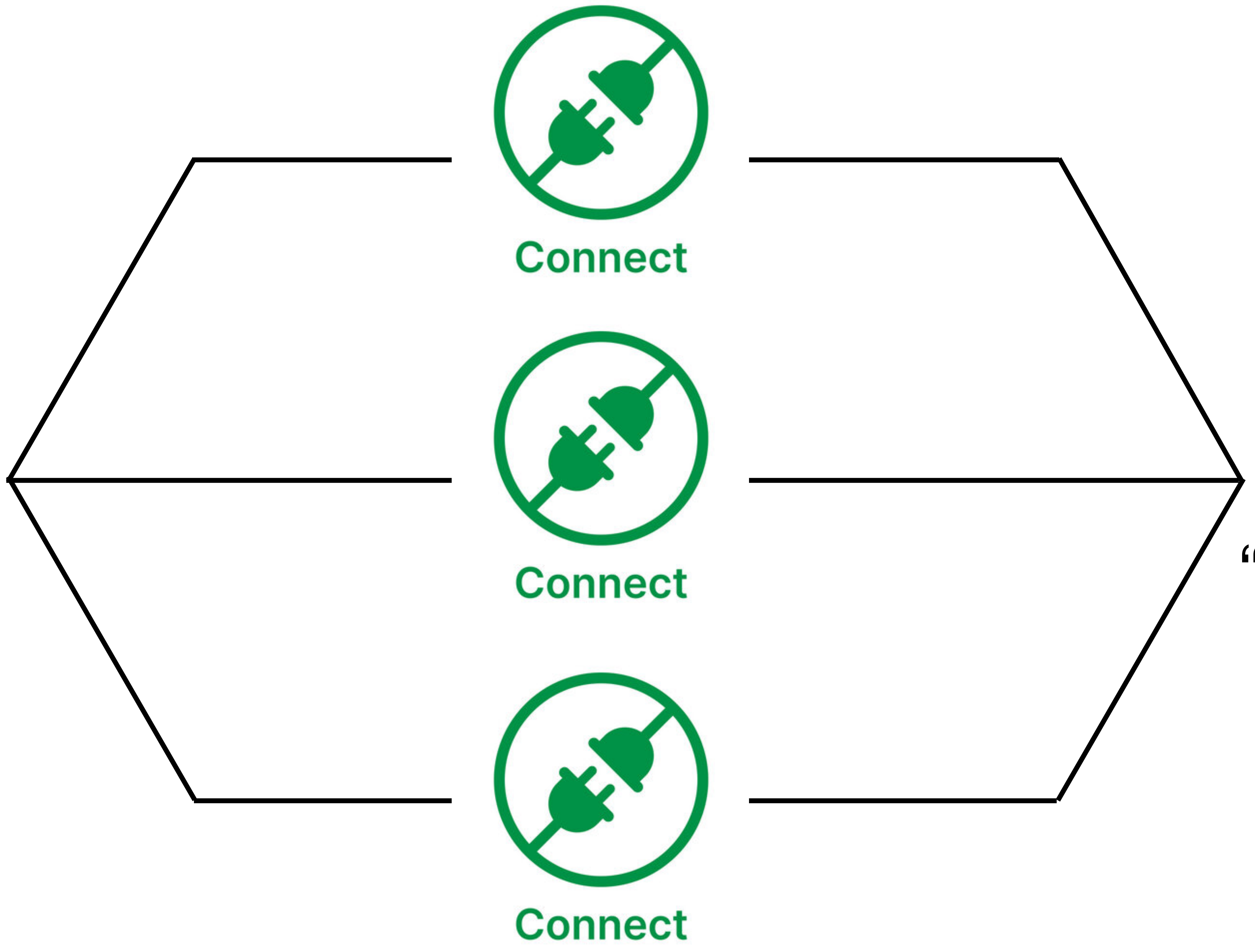
"When is the causal effect computable from data?"



Develop robust estimation methods for causal effects **across diverse scenarios**

Identification

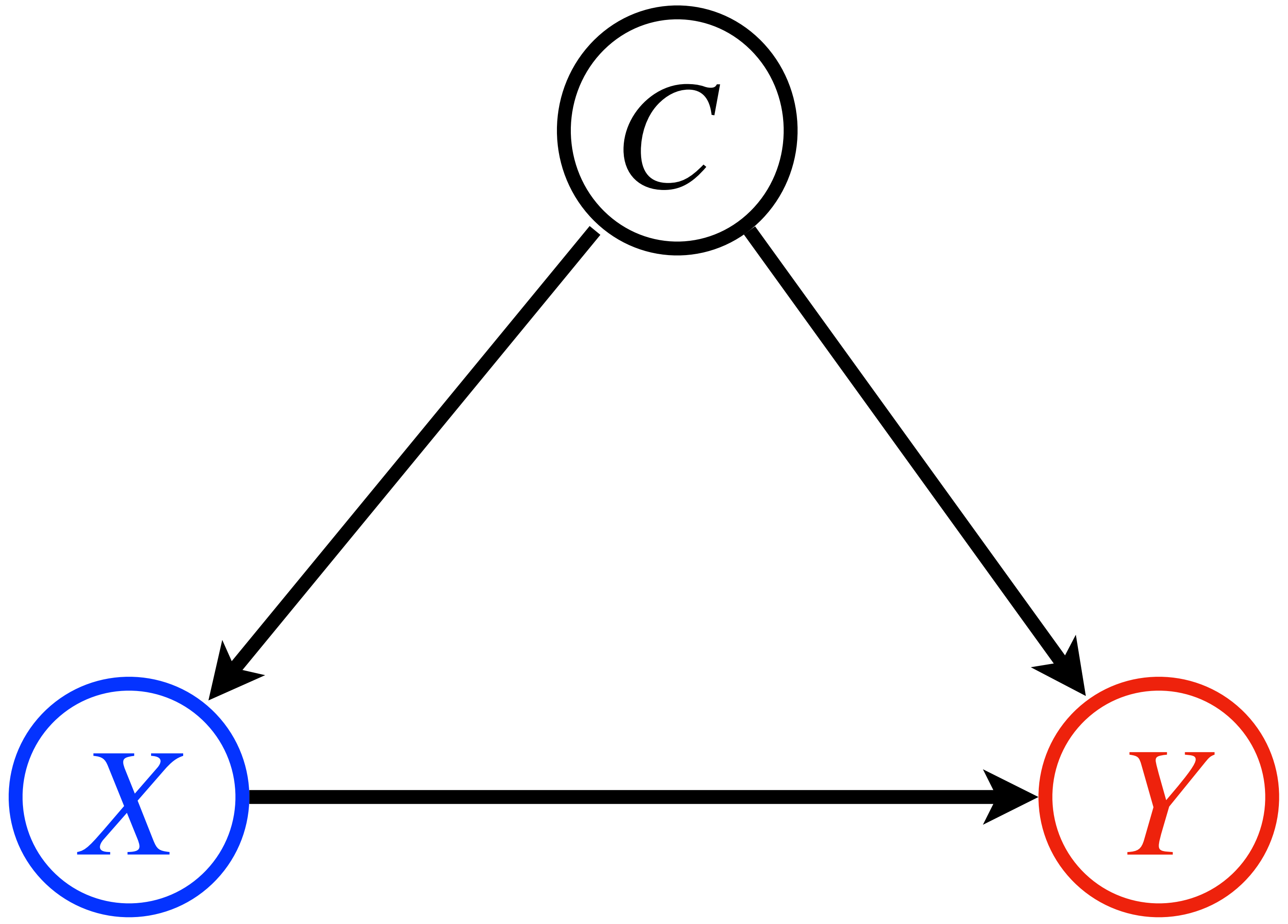
"When is the causal effect computable from data?"



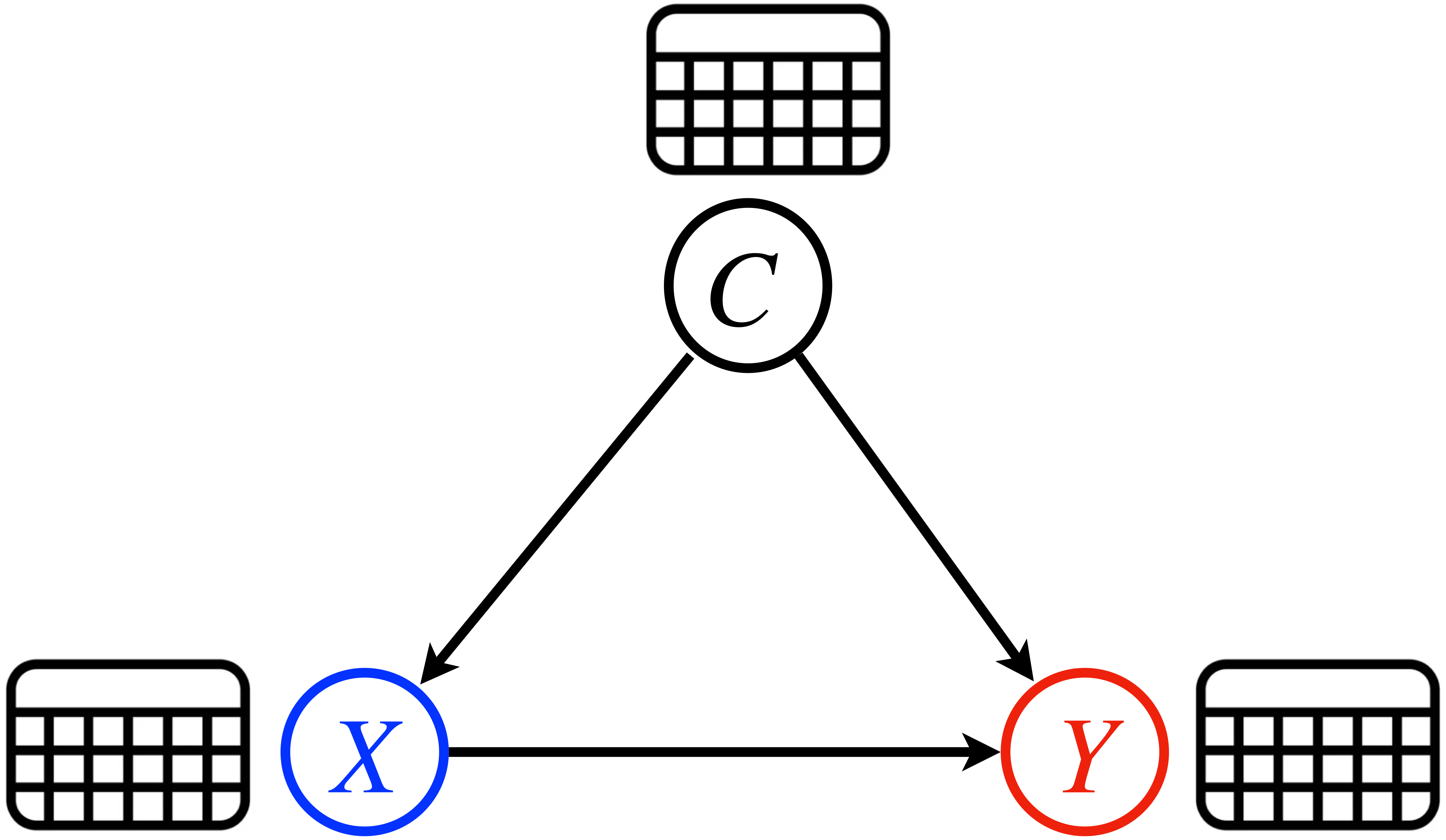
Estimation

“How do we compute the effect from data?”

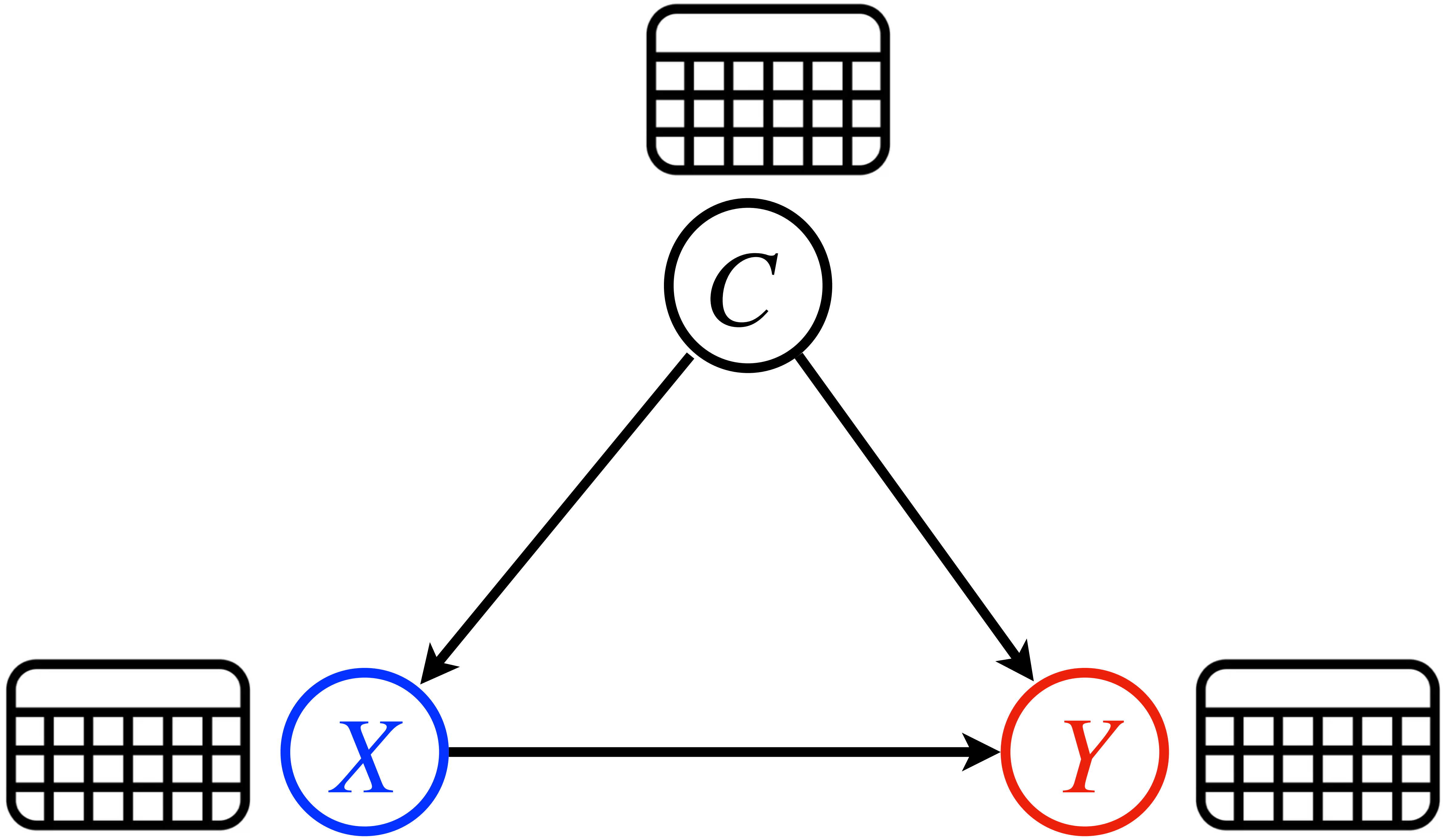
Future 1: Inference with Multi-modal Data



Future 1: Inference with Multi-modal Data

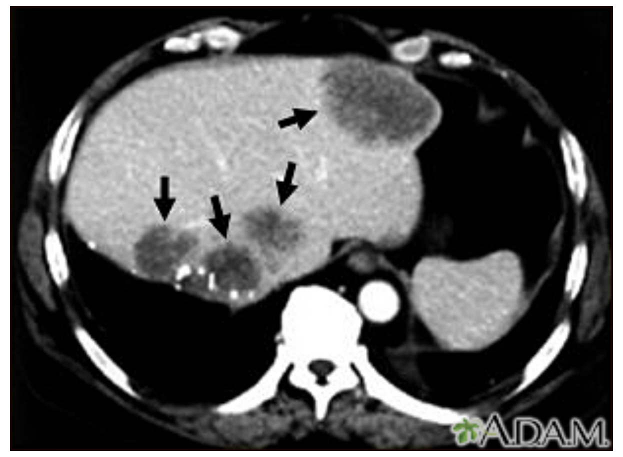


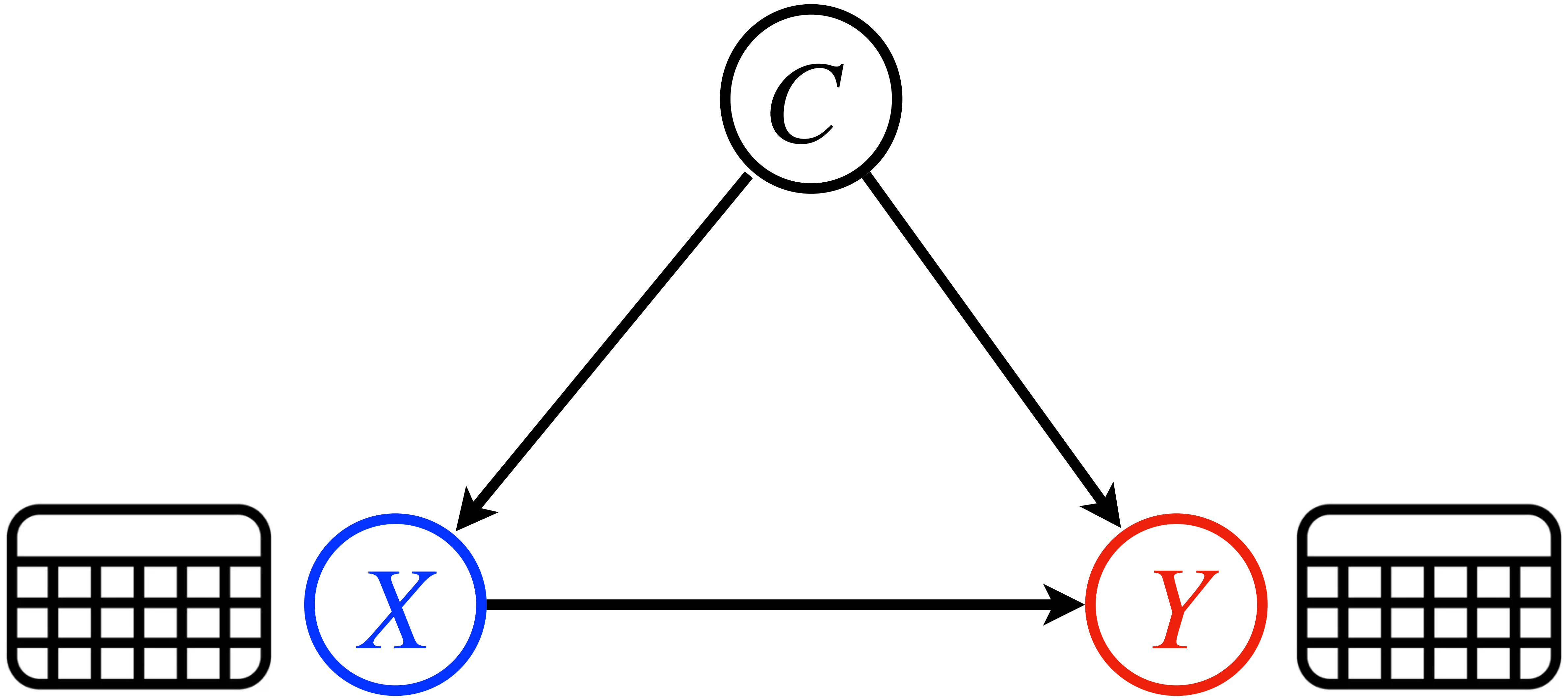
Future 1: Inference with Multi-modal Data



$$\mathbb{E}[Y \mid \mathrm{do}(x)] = \sum_c \mathbb{E}[Y \mid x, c]\, P(c)$$

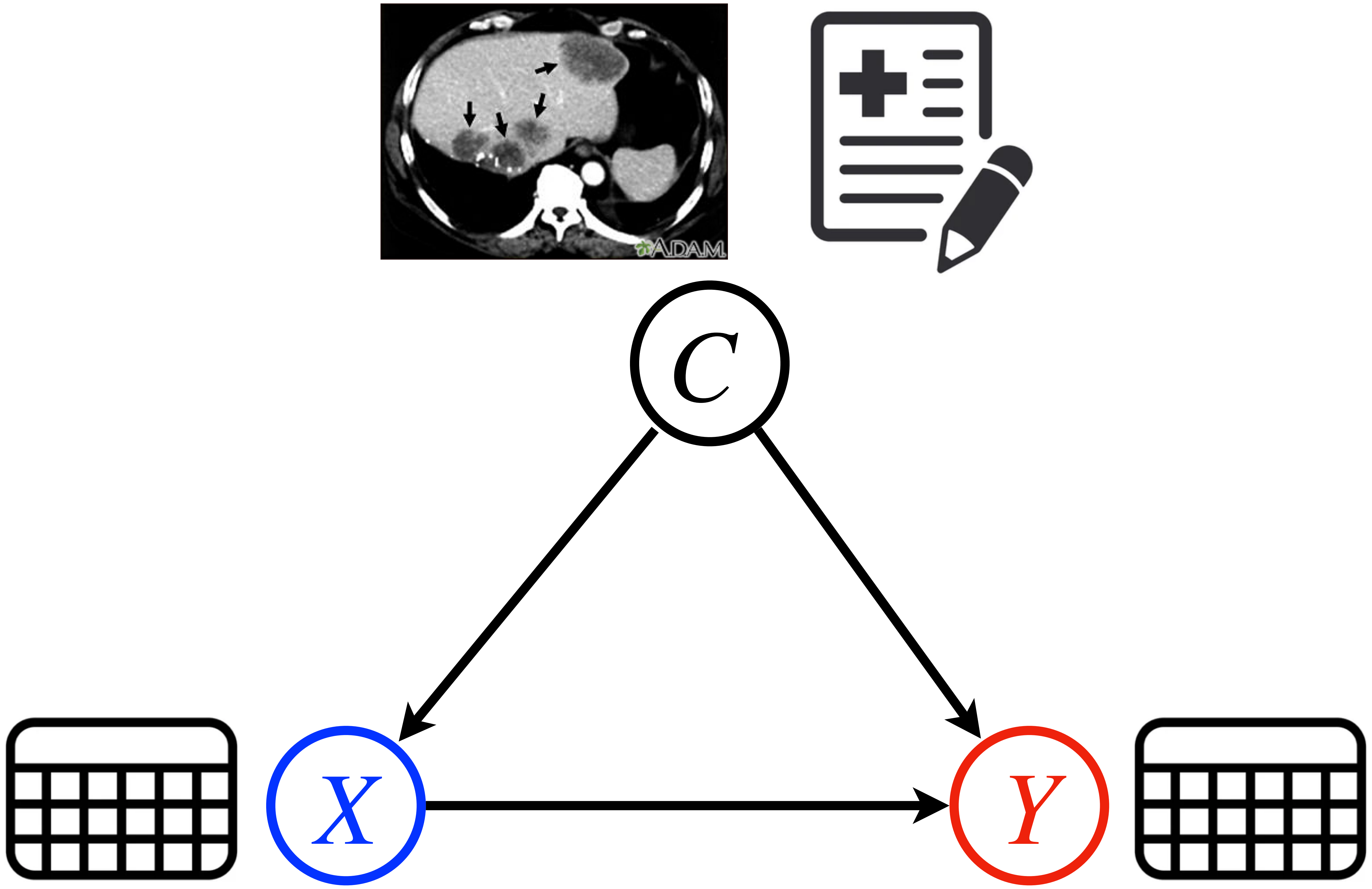
Future 1: Inference with Multi-modal Data





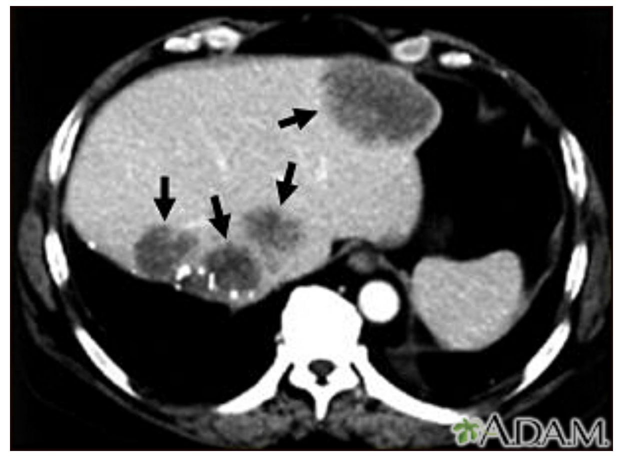
$$\mathbb{E}[Y \mid \mathrm{do}(x)] = \sum_c \mathbb{E}[Y \mid x, c]\, P(c)$$

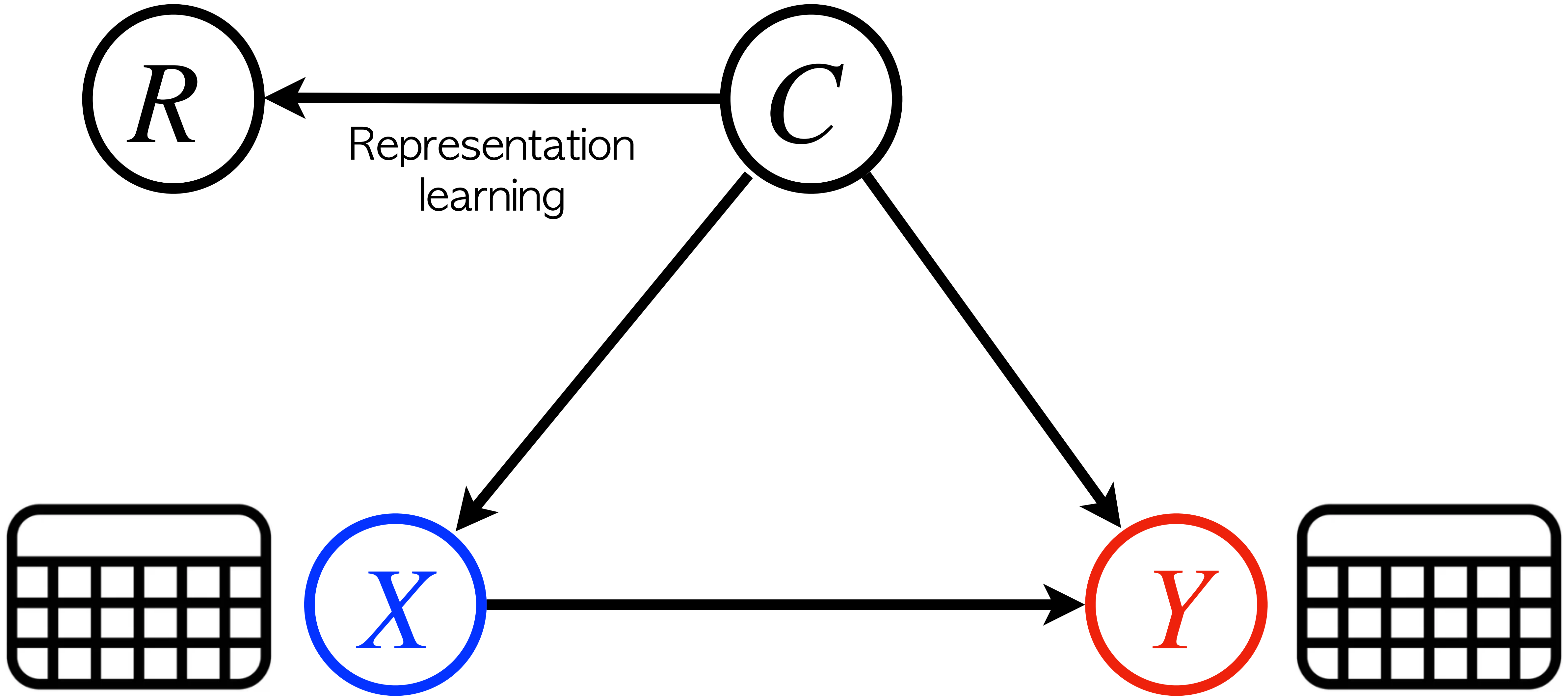
Future 1: Inference with Multi-modal Data



$$\mathbb{E}[Y \mid \mathrm{do}(x)] = \quad ?$$

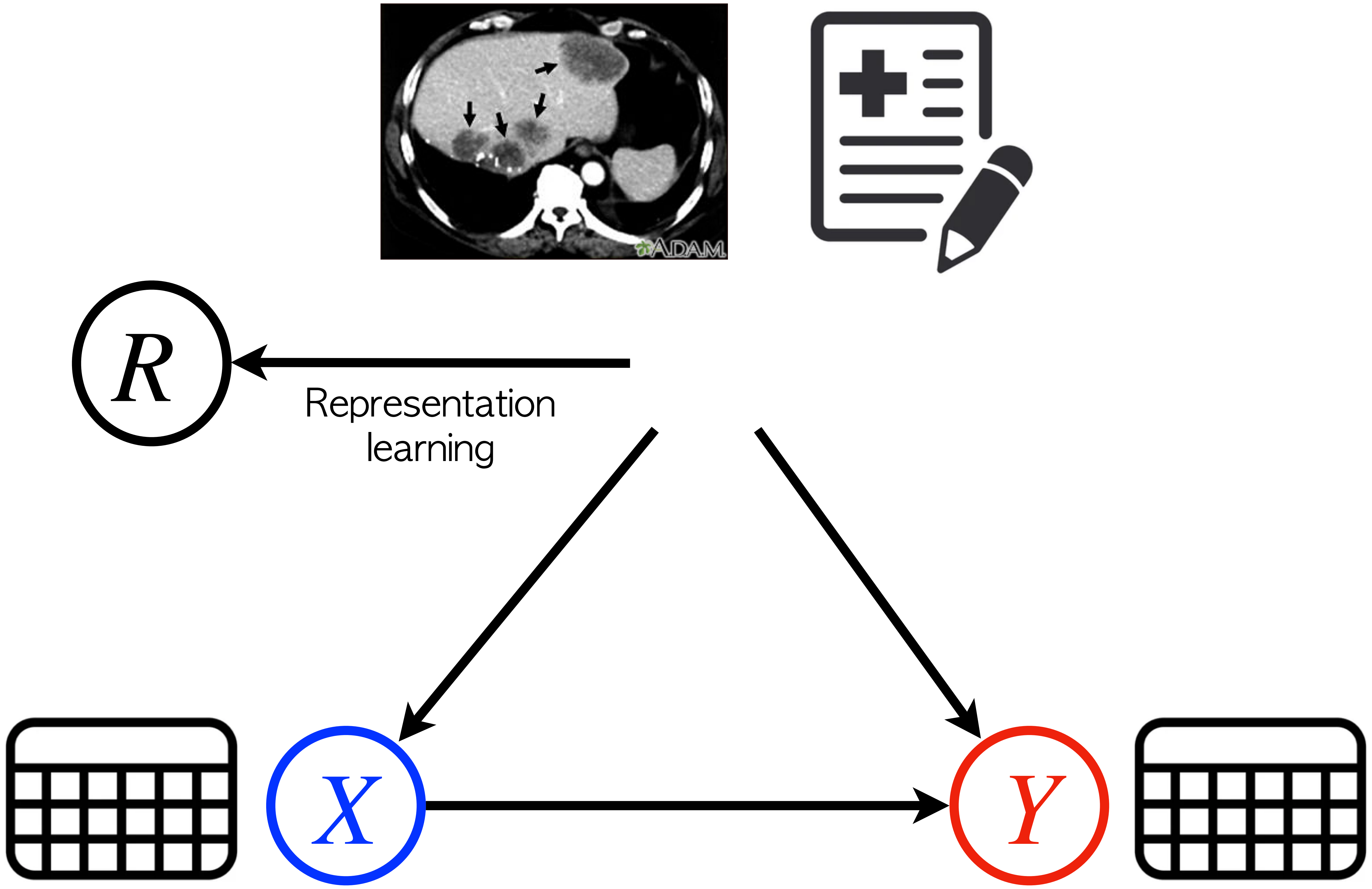
Future 1: Inference with Multi-modal Data





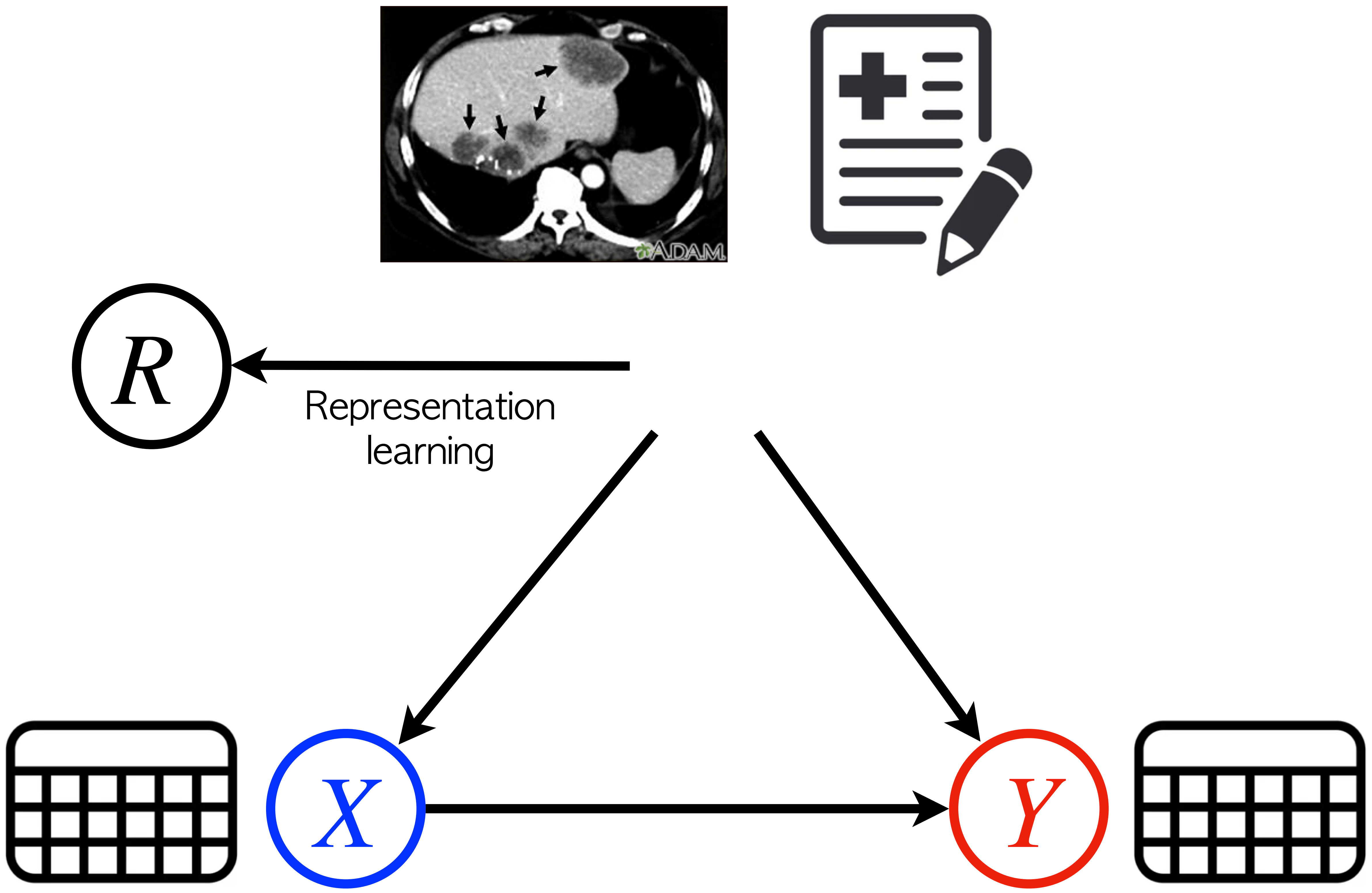
$$\mathbb{E}[Y \mid \mathrm{do}(x)] = \quad ?$$

Future 1: Inference with Multi-modal Data



$$\mathbb{E}[Y \mid \mathrm{do}(x)] = \sum_r \mathbb{E}[Y \mid x, r] \; P(r)$$

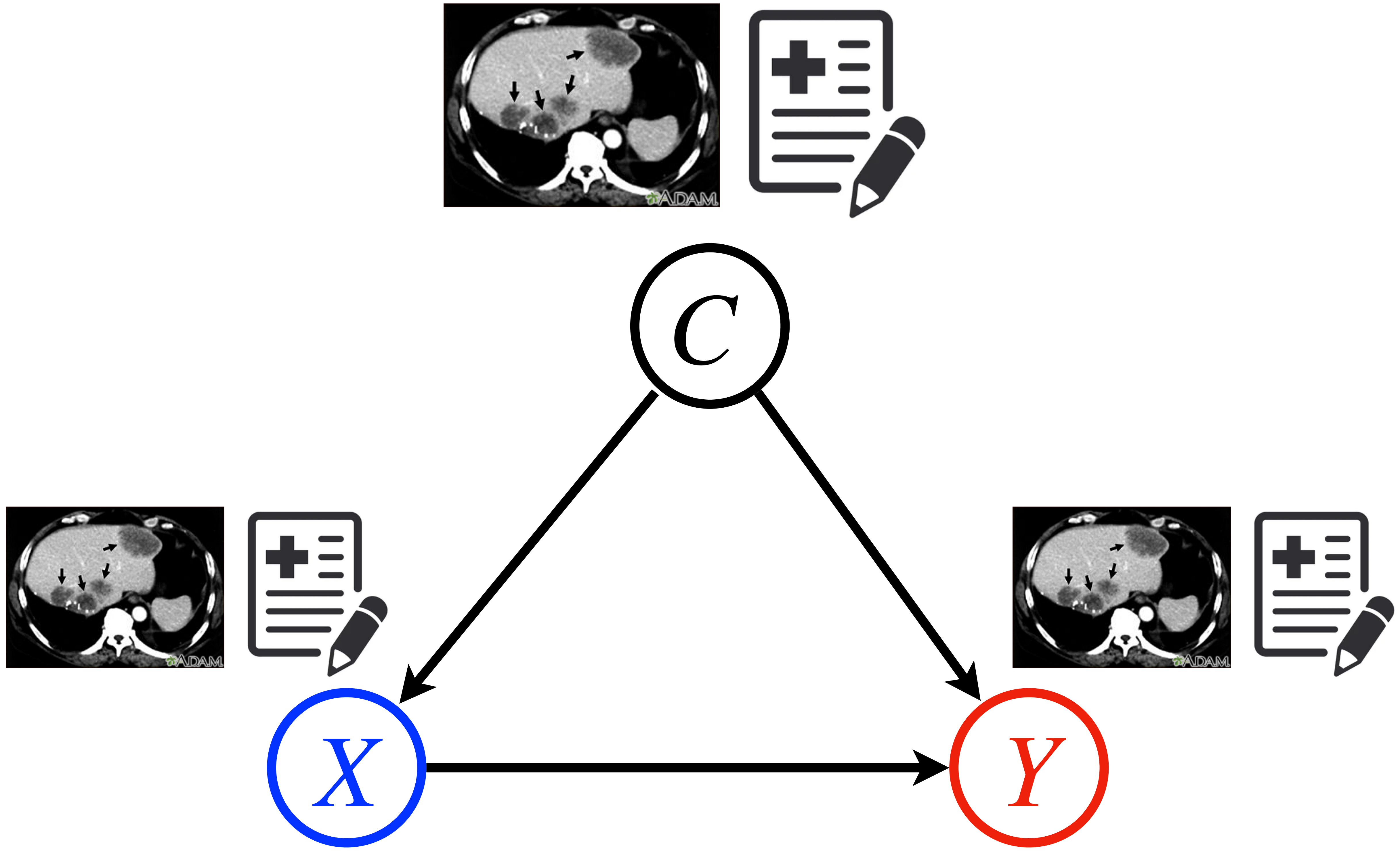
Future 1: Inference with Multi-modal Data



$$\mathbb{E}[Y \mid \mathrm{do}(x)] \neq \sum_r \mathbb{E}[Y \mid x, r] \; P(r)$$

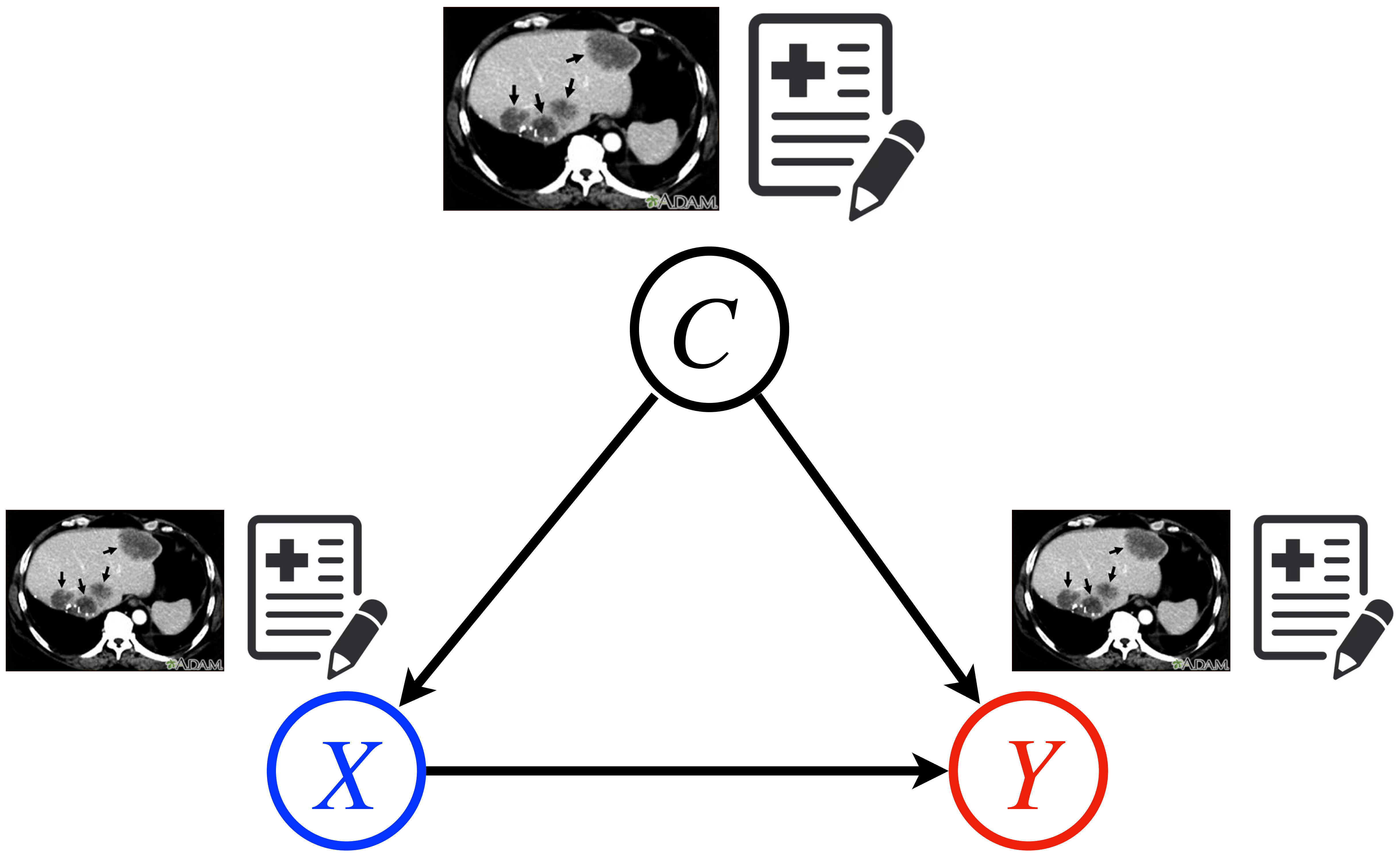
R doesn't satisfy the BD criterion

Future 1: Inference with Multi-modal Data



$$\mathbb{E}[Y \mid \mathrm{do}(x)] = \text{?}$$

Future 1: Inference with Multi-modal Data



$$\mathbb{E}[Y \mid \mathrm{do}(x)] = \text{?}$$

Approach

- Representation learning taking account of causal dependencies
- New causal inference methods that allows us to use existing representation learning models

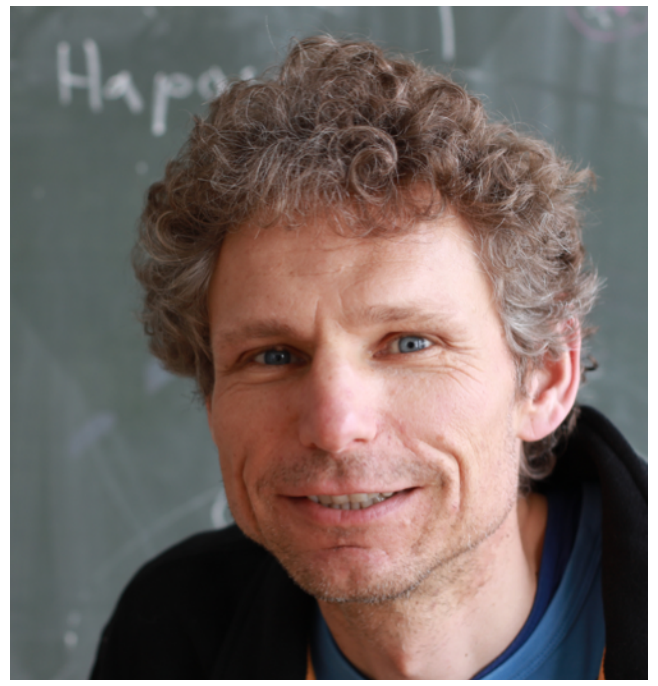
Collaborators

Elias Bareinboim
(Columbia University)

Jin Tian
(MBZUAI)

Ivan Diaz
(NYU Biostatistics)

Shiva Kasiviswanathan
(Amazon)

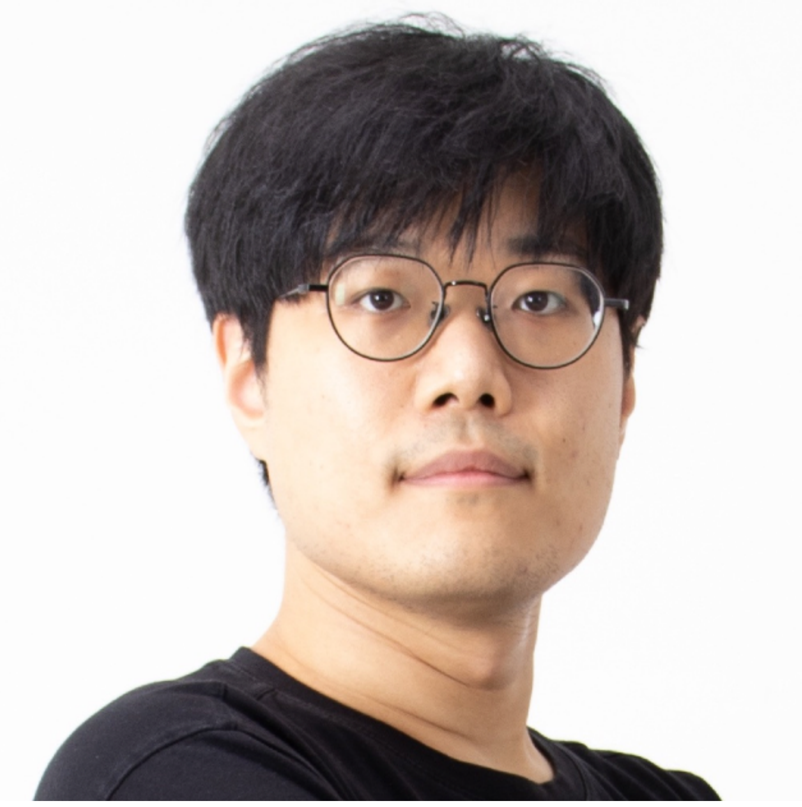


Dominik Janzing
(Amazon)

Alexis Bellot
(Google DeepMind)

Sanghack Lee
(Seoul National University)

Kyungwoo Song
(Yonsei University)



Sanghyuk Chun
(Naver AI)

Shamali Joshi
(Columbia Univ. DBMI)

Thank you

Current: Developing robust estimators for causal effects across diverse scenarios
Future: Advancing causal inference for complex, real-world benefits

www.yonghanjung.me/

PhD Student Recruitment

I currently recruiting PhD students to work with me starting in Spring or Fall 2026. My research focuses on ***causal inference with AI/ML, trustworthy AI, and applications to public health***. If you're interested in these areas, please feel free to reach out. You can find more details on my website.

www.yonghanjung.me/