Causal Data Science:

Estimating Identifiable Causal Effects

Yonghan Jung

School of Information Science (iSchool)

UIUC

www.yonghanjung.me

2025 Fall UIUC Statistics Seminar

> “*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.

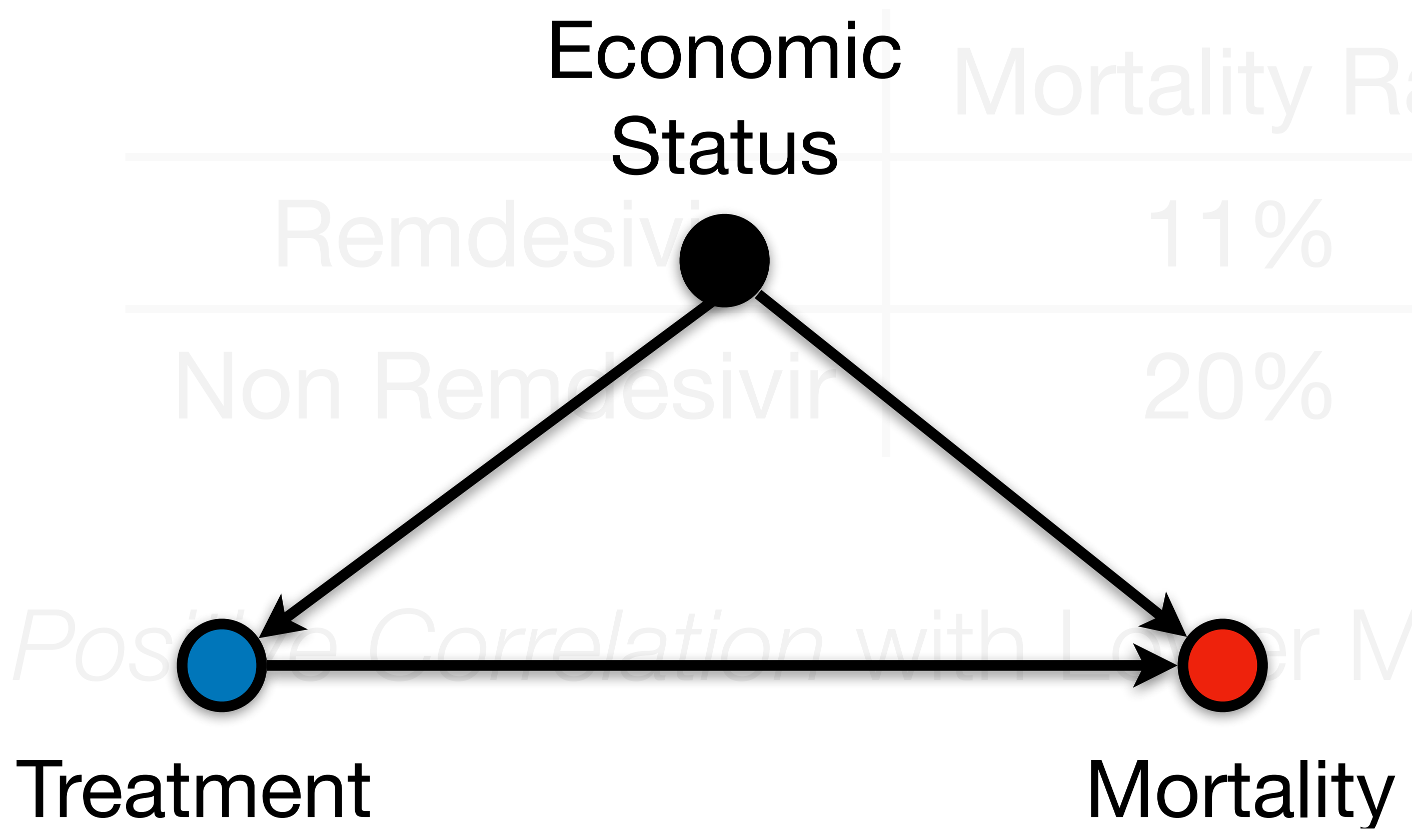
Randomized Trial (WHO)

	Mortality Rate
Remdesivir	15%
Non Remdesivir	15%

No Causal Effect to Lower Mortality

Story Behind the Data

Observational Study (FDA)



	Mortality Rate
Remdesivir	11%
Non Remdesivir	20%

Positive Correlation with Lower Mortality

vs.

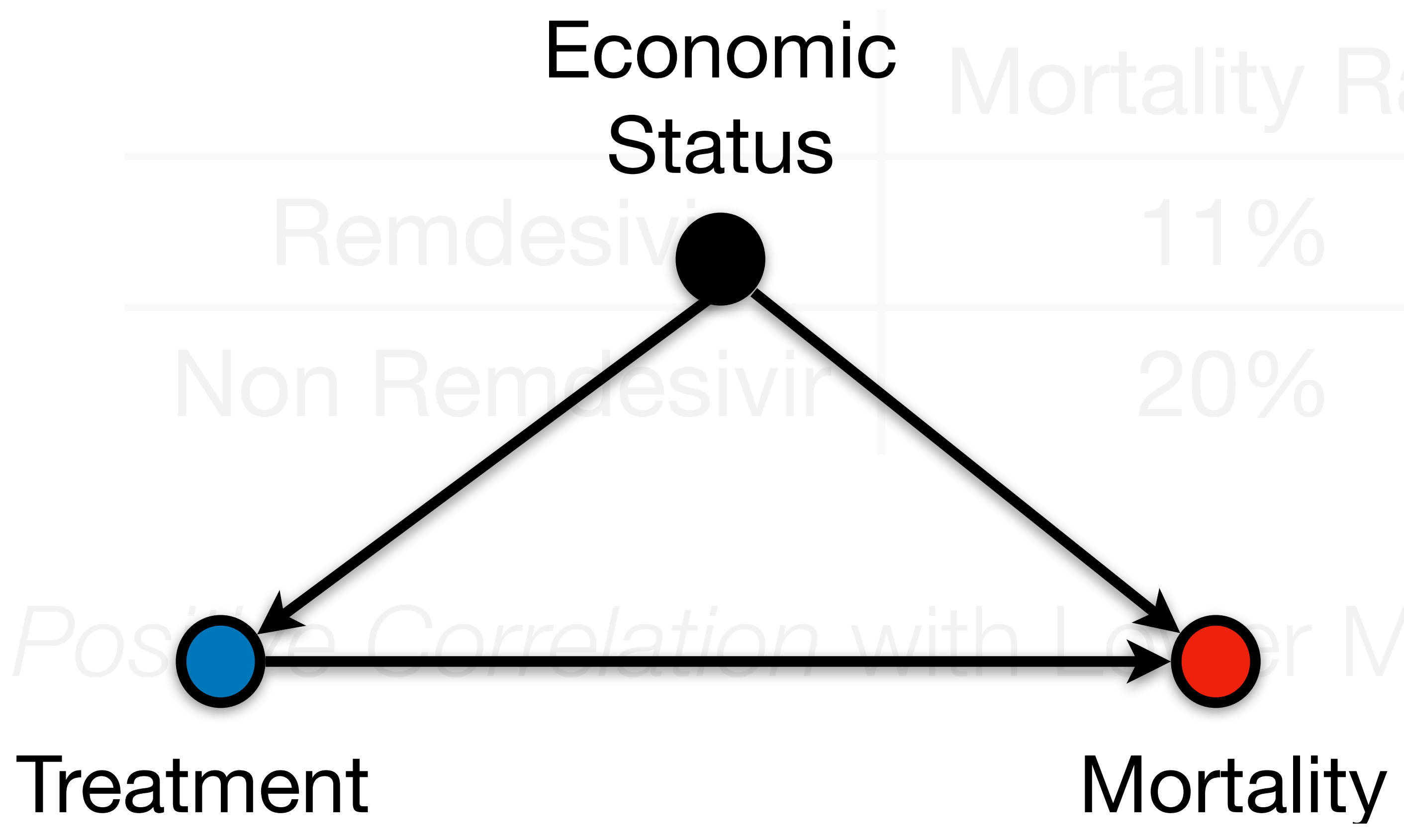
Randomized Trial (WHO)

	Mortality Rate
Remdesivir	15%
Non Remdesivir	15%

No Causal Effect to Lower Mortality

Story Behind the Data

Observational Study (FDA)

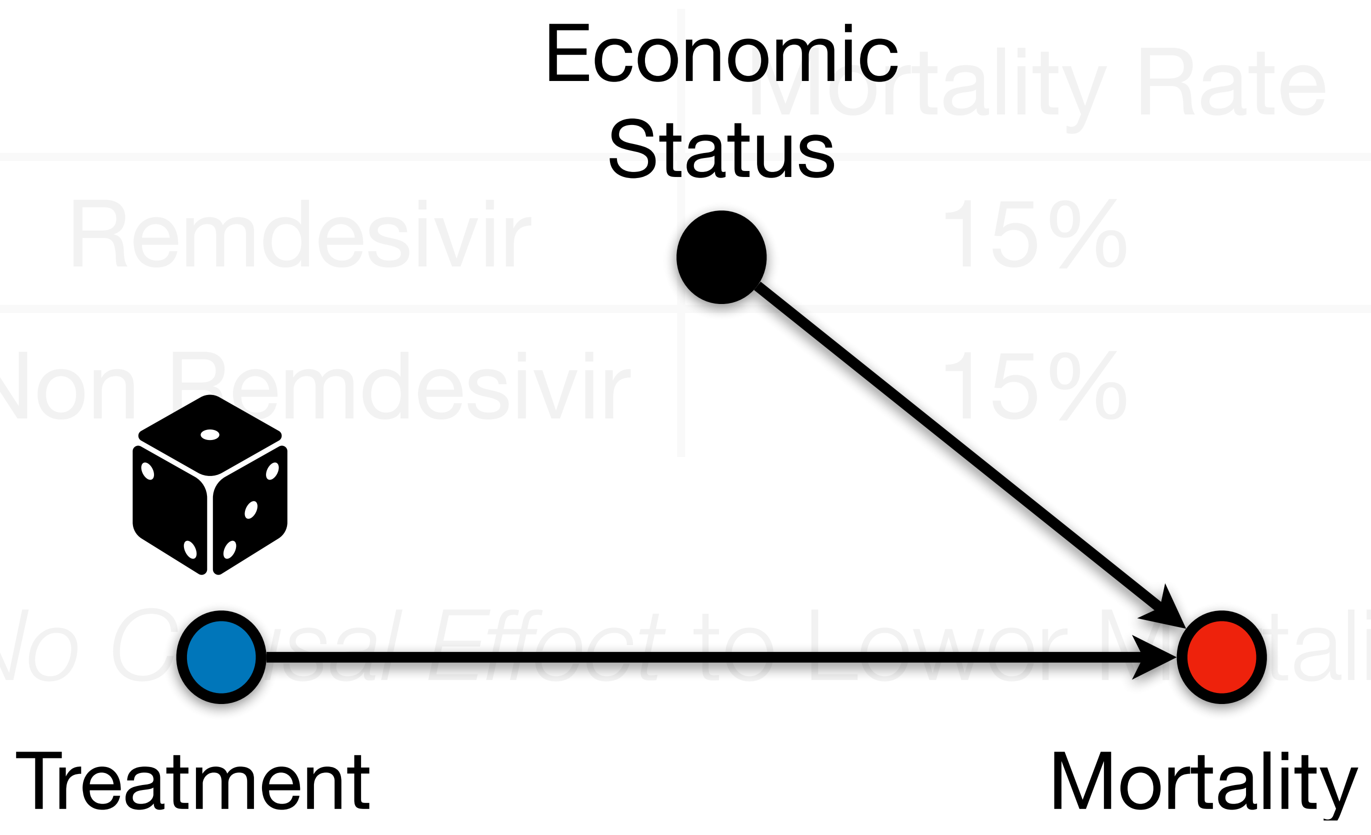


	Mortality Rate
Remdesivir	11%
Non Remdesivir	20%

Positive Correlation with Lower Mortality

vs.

Randomized Trial (WHO)



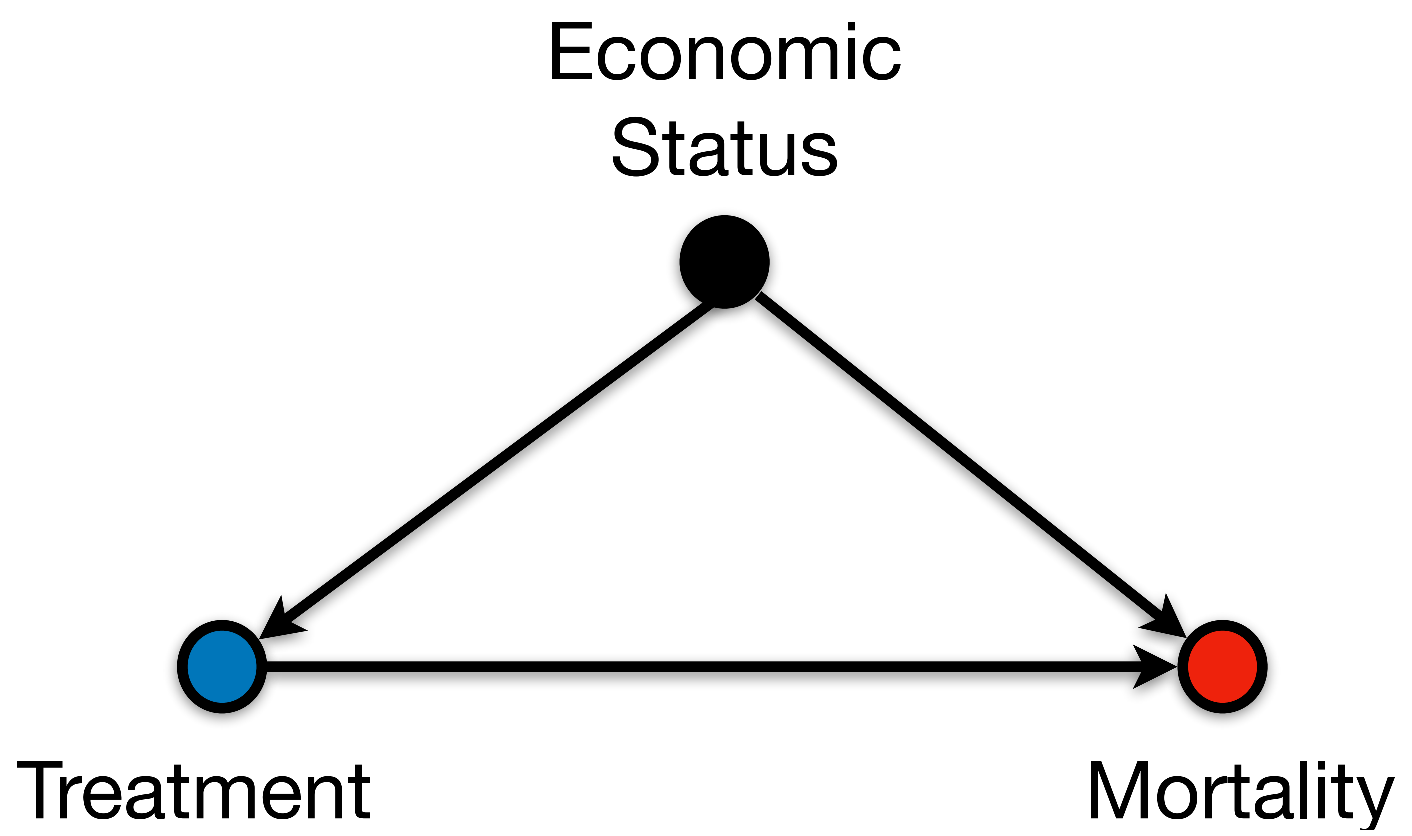
	Mortality Rate
Remdesivir	15%
Non Remdesivir	15%

No Causal Effect to Lower Mortality

Story Behind the Data

Observational Study (FDA)

	Mortality Rate
Remdesivir	11%
Non Remdesivir	20%



"Causal Inference Engine"

Causal Effect

	Mortality Rate
Remdesivir	15%
Non Remdesivir	15%

Standard Causal Inference Engine

Standard Causal Inference Engine

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 Engine

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?"

Estimation

"How do we compute the effect from data?"

Standard Causal Inference Engine

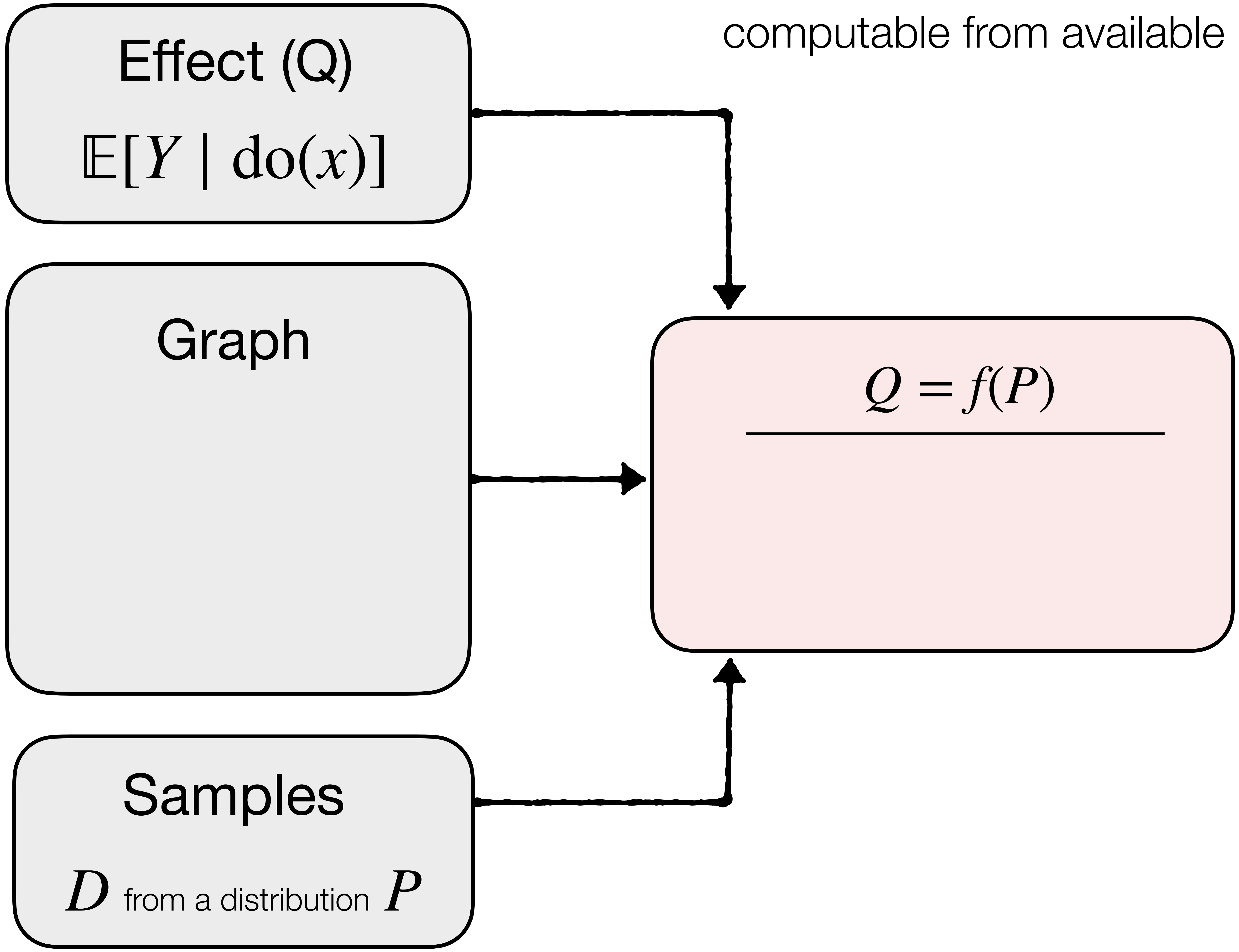
Input

Identification

“When is the causal effect computable from available data?”

Estimation

“How do we compute the effect from data?”



Effect (Q)

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

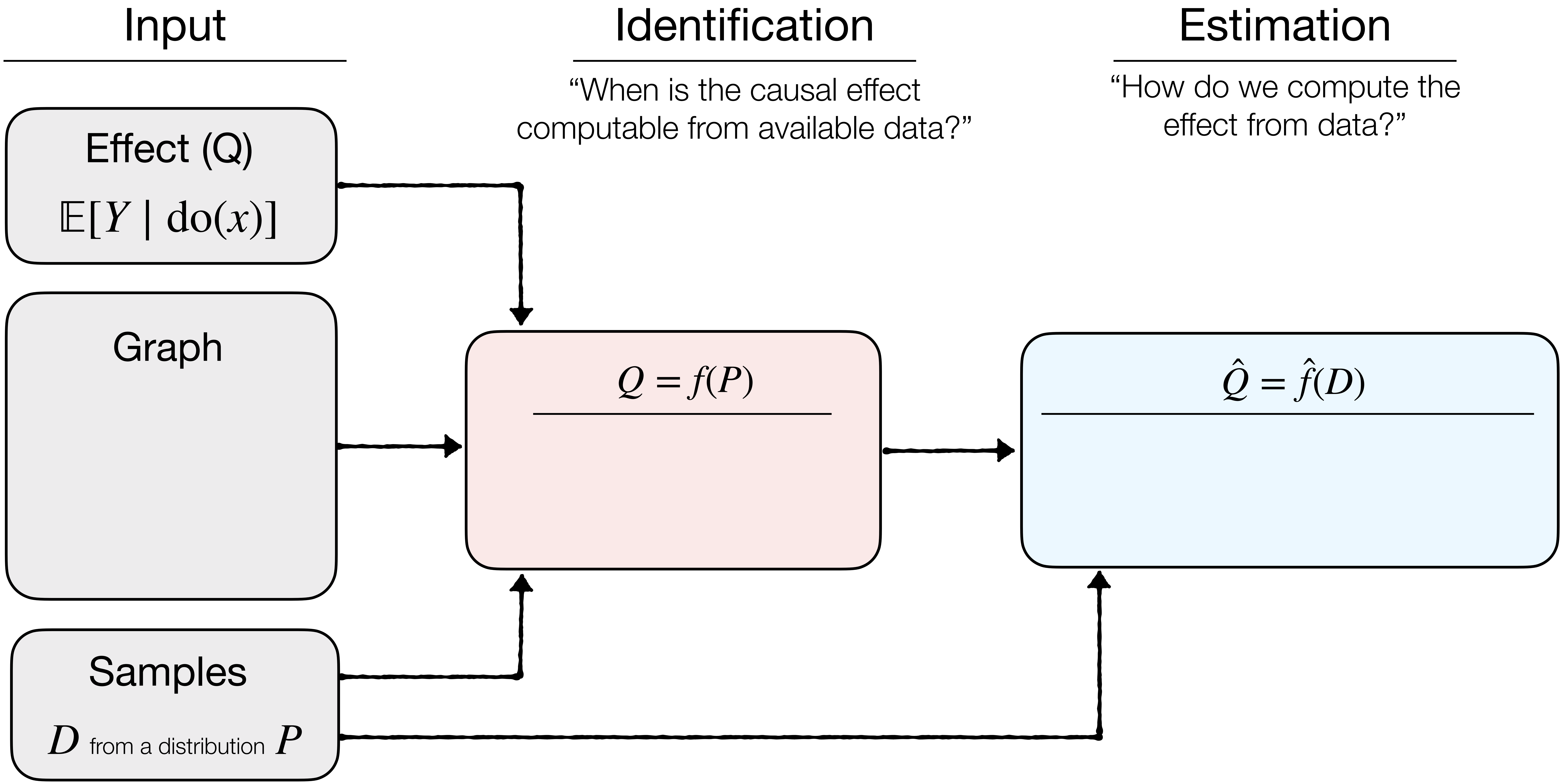
Graph

Samples

D from a distribution P

$Q = f(P)$

Standard Causal Inference Engine



Input

Identification

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

Estimation

"How do we compute the effect from data?"

Effect (Q)

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

Graph

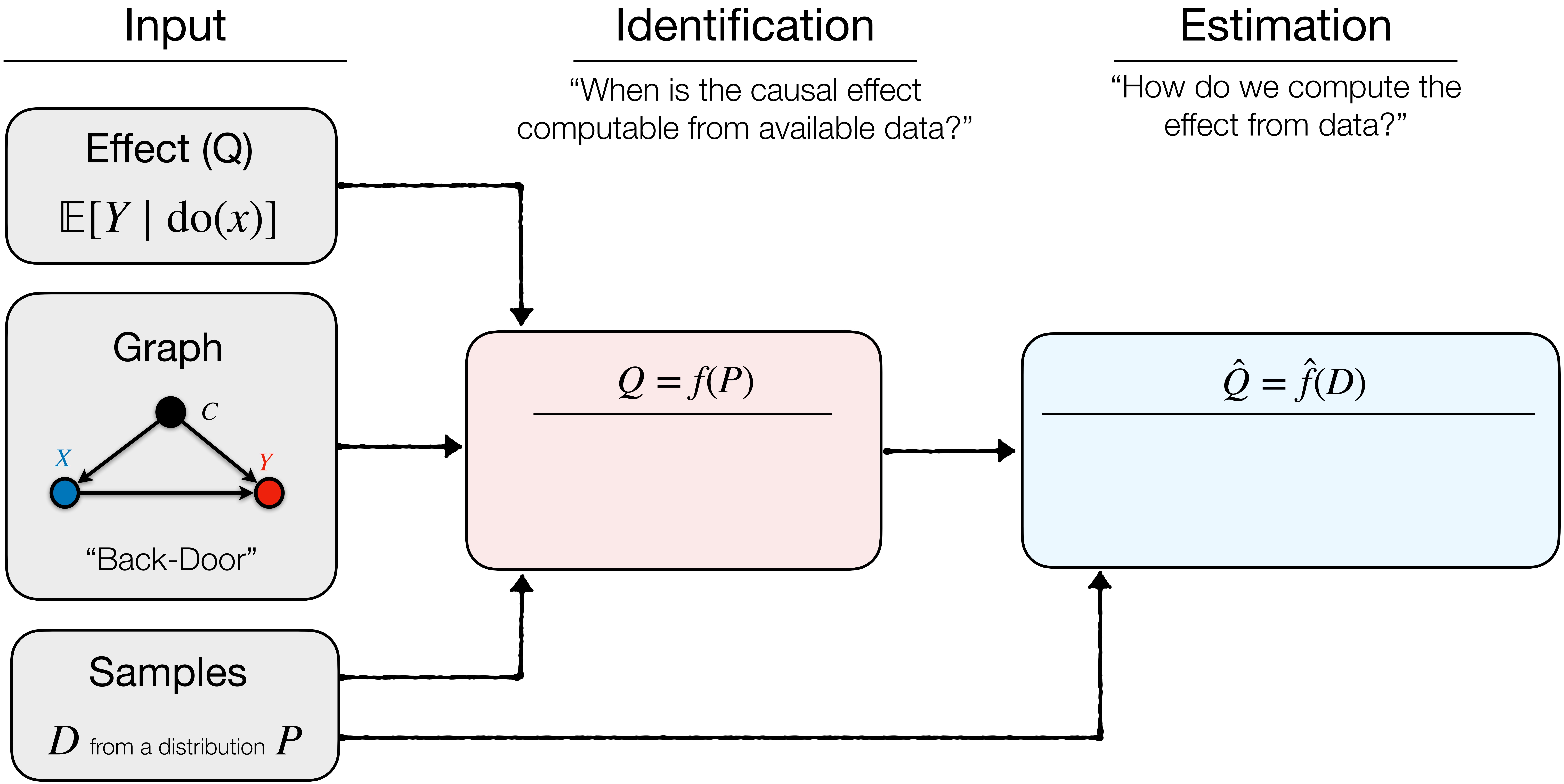
Samples

D from a distribution P

$Q = f(P)$

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

Standard Causal Inference Engine



Input

Identification

“When is the causal effect computable from available data?”

Estimation

“How do we compute the effect from data?”

Effect (Q)

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

Graph

“Back-Door”

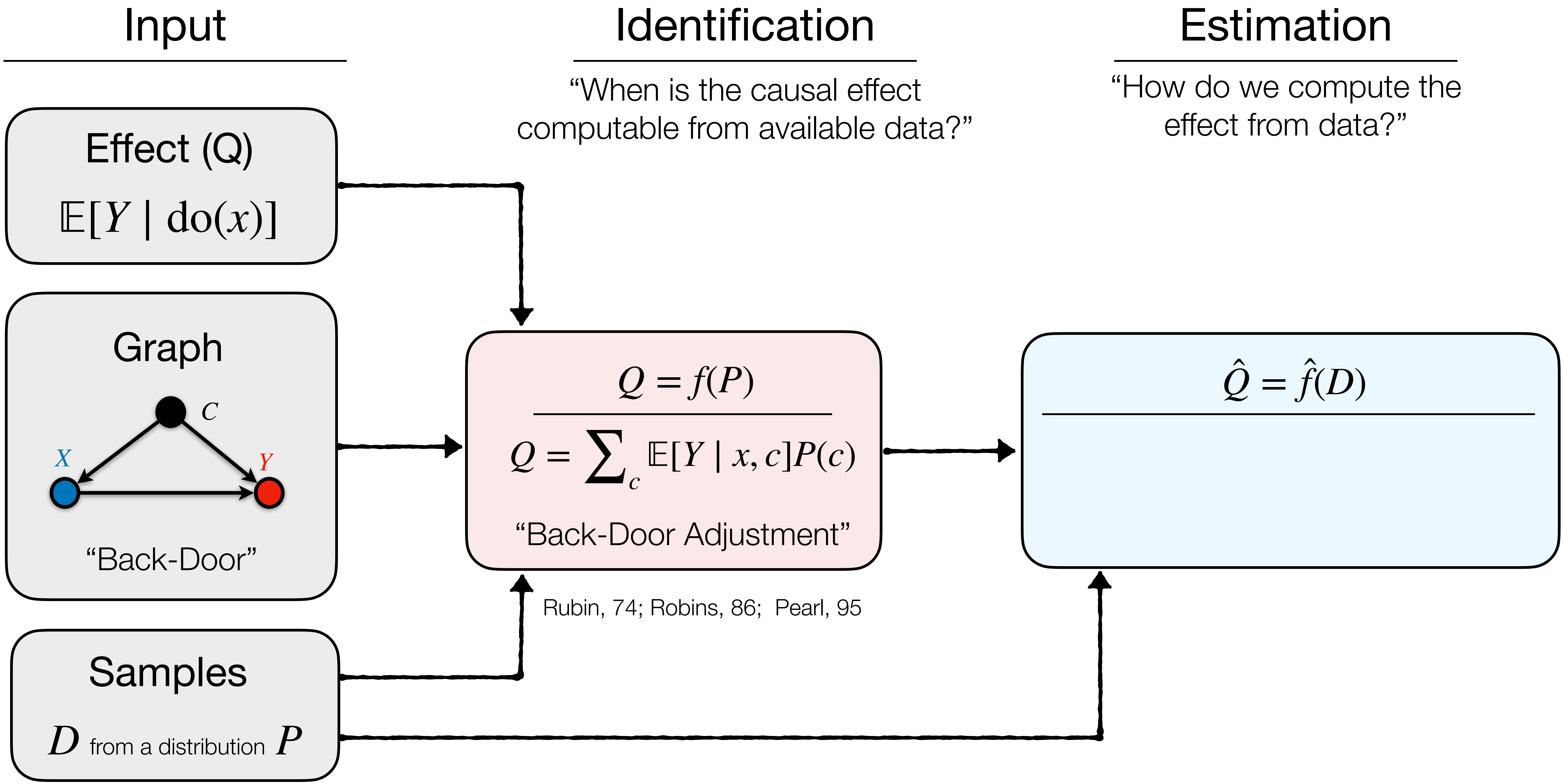
Samples

D from a distribution P

$Q = f(P)$

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

Standard Causal Inference Engine



Input

Effect (Q)

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

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"

Rubin, 74; Robins, 86; Pearl, 95

Estimation

"How do we compute the effect from data?"

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

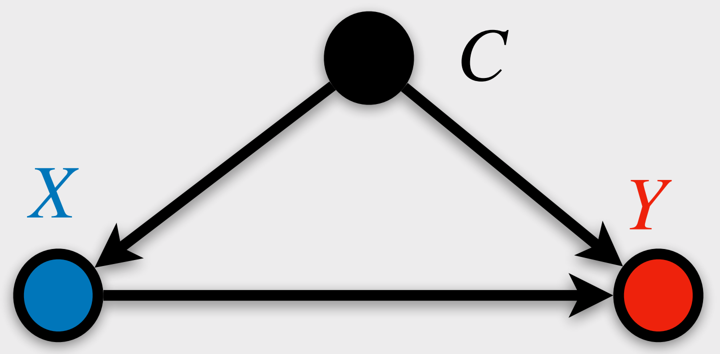
Standard Causal Inference Engine

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"

Rubin, 74; Robins, 86; Pearl, 95

Estimation

"How do we compute the effect from data?"

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

One-Step Estimator, TMLE, DML

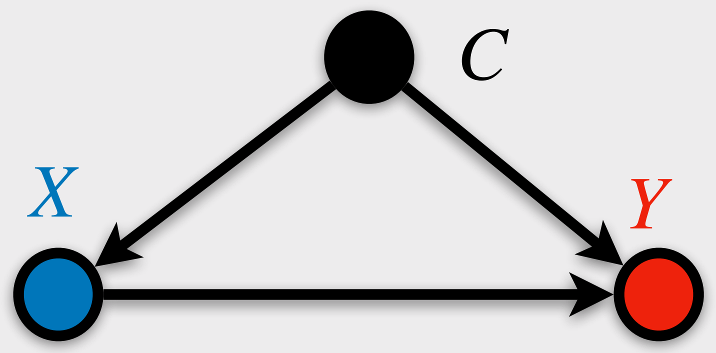
Bickel et al., 93; van der Laan and Rubin 06, van der Laan and Gruber12; Chernozhukov et al, 2018

Challenge 1: Complex dependences

Effect (Q)

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

Graph



“Back-Door”

Samples

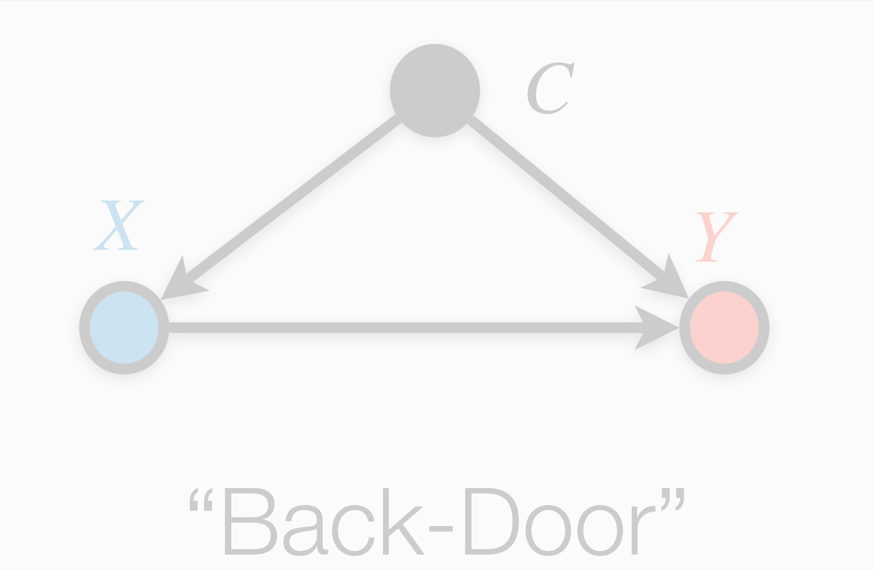
D from P

Challenge 1: Complex dependences

Effect (Q)

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

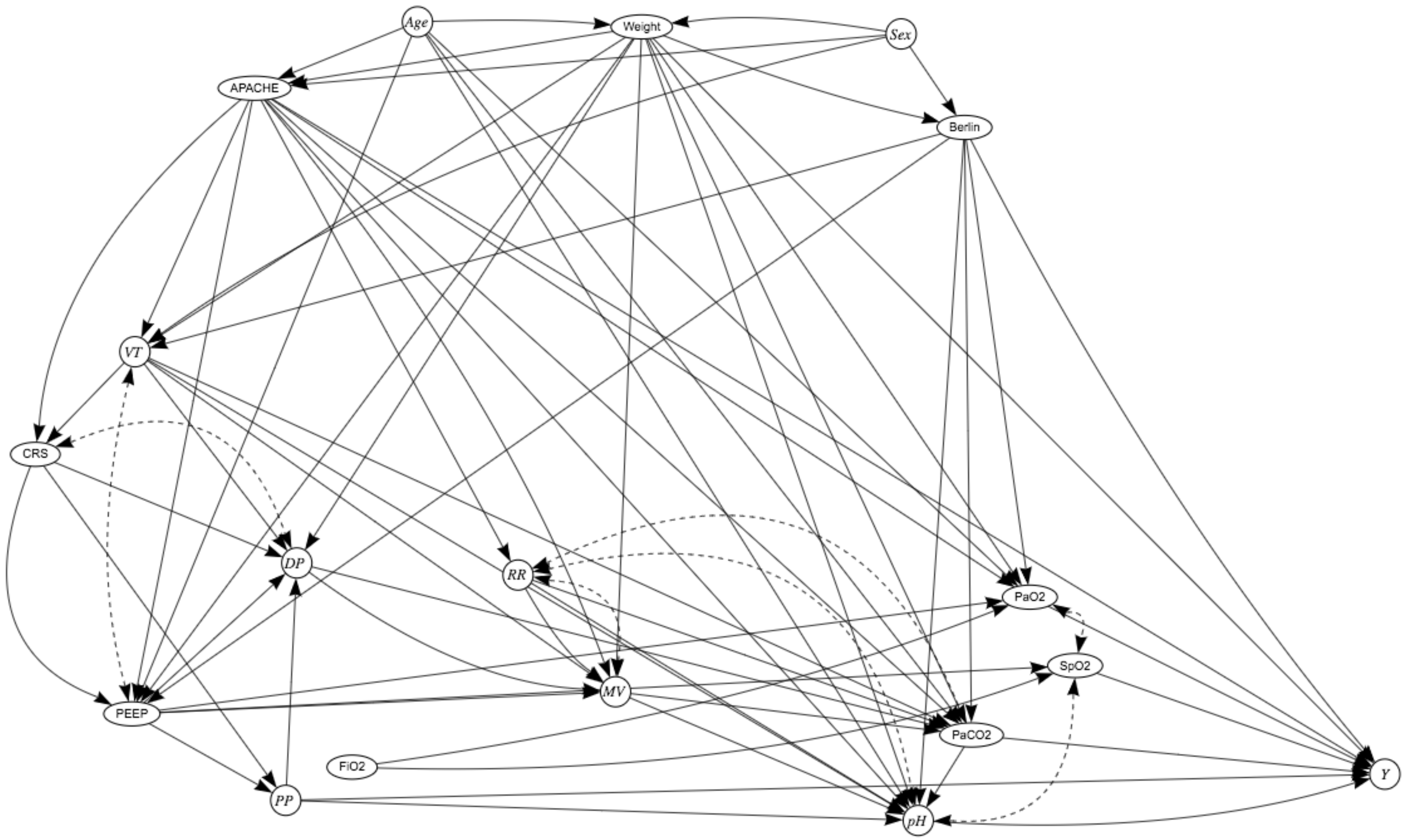
Graph



Samples

D from P

Complex dependences



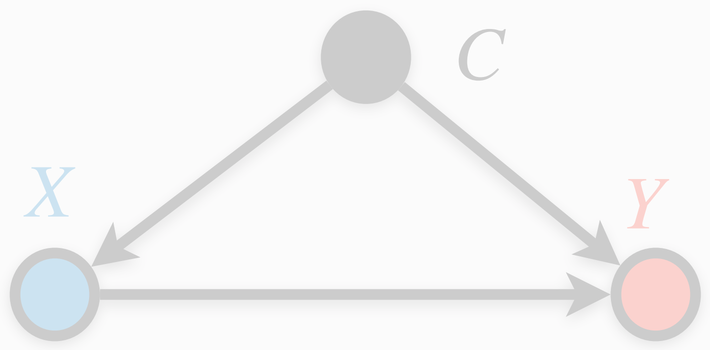
Causal graph on acute respiratory distress syndrome (ARDS)

Challenge 2: Data Fusion

Effect (Q)

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

Graph



“Back-Door”

Samples

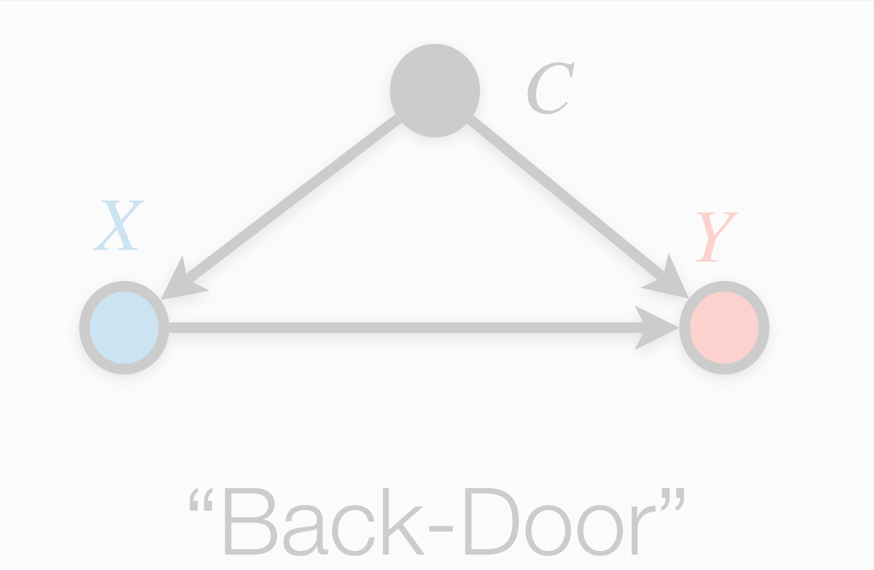
D from P

Challenge 2: Data Fusion

Effect (Q)

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

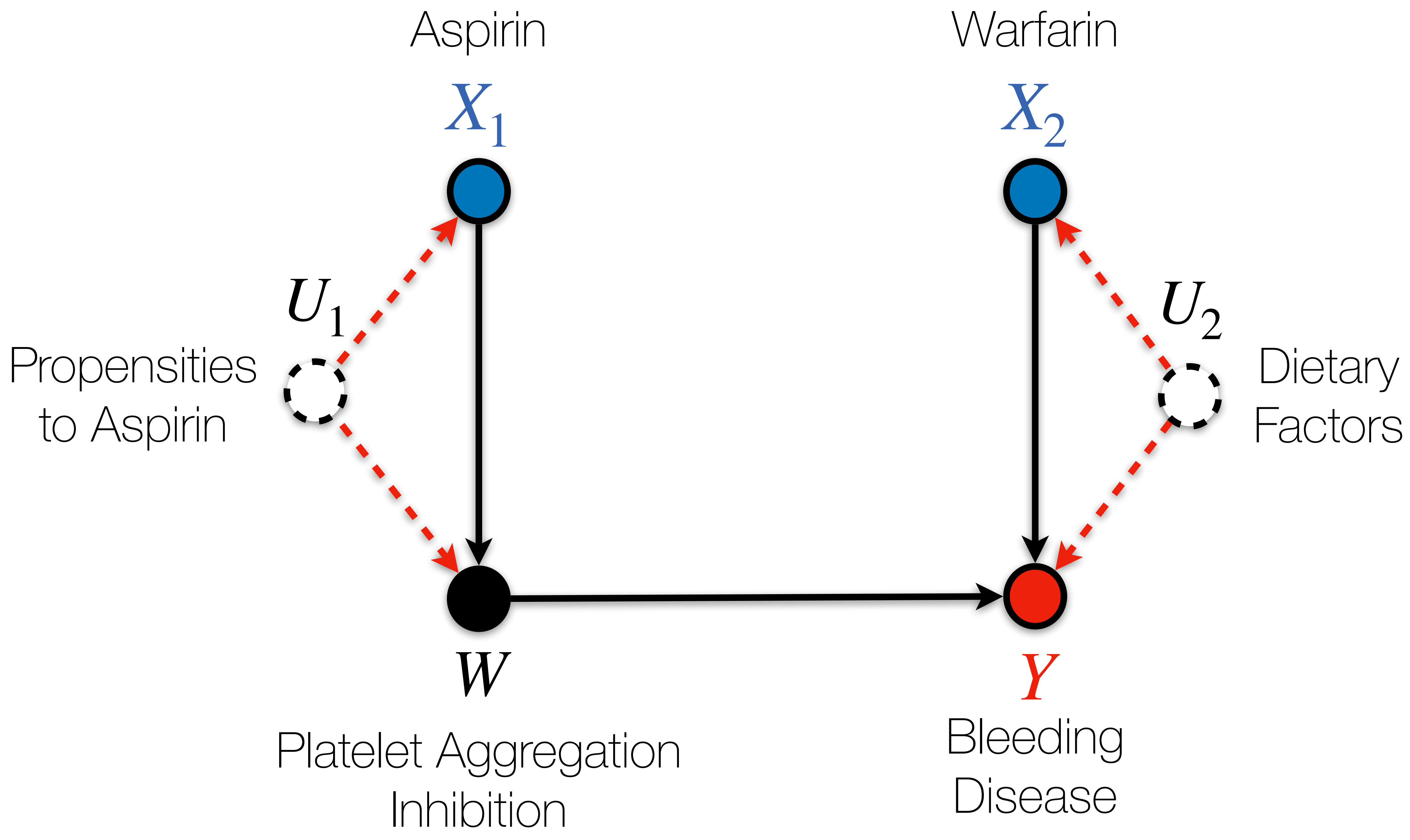
Graph



“Back-Door”

Samples

D from P



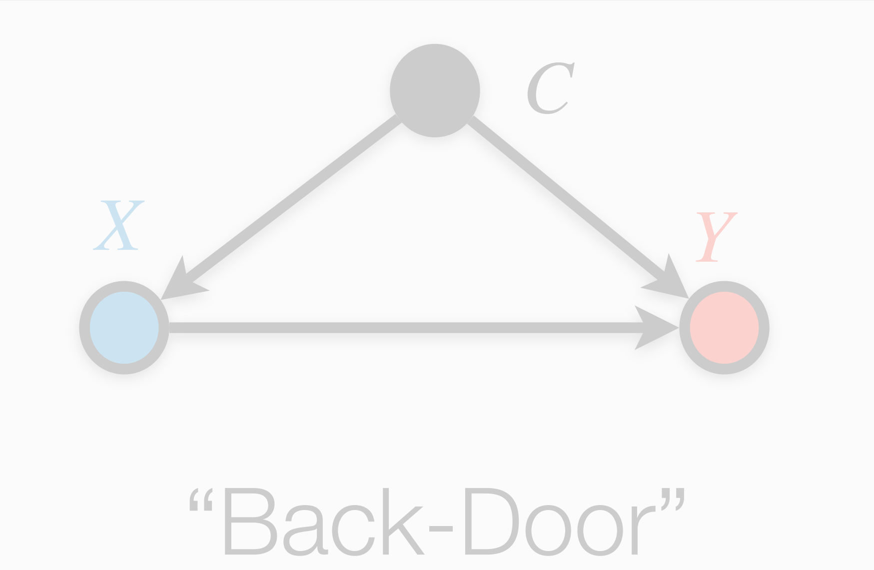
- Goal: Estimate $\mathbb{E}[Y \mid \mathrm{do}(x_1, x_2)]$ from single interventions $\mathrm{do}(x_1)$ and $\mathrm{do}(x_2)$.

Challenge 2: Data Fusion

Effect (Q)

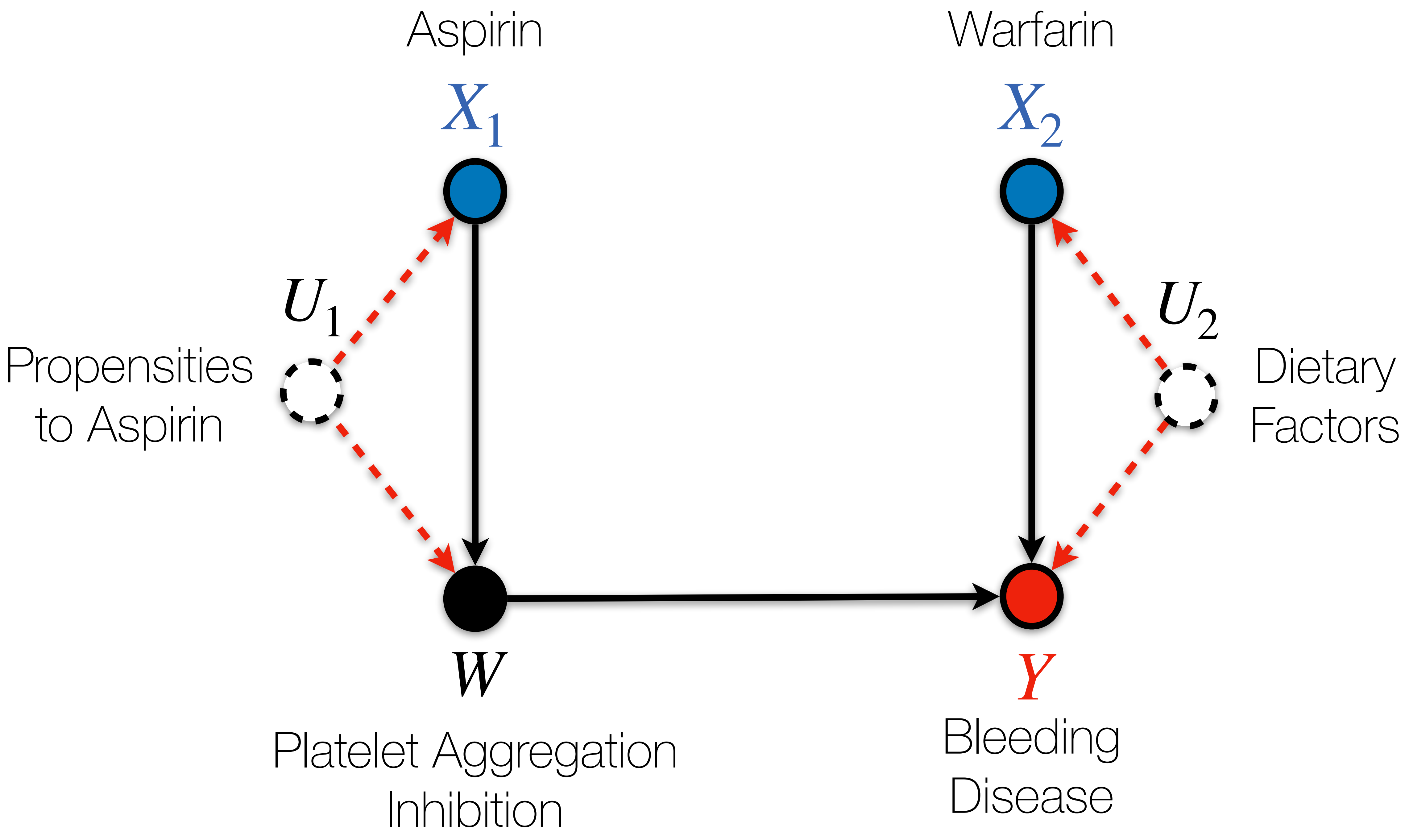
$\mathbb{E}[Y \mid \text{do}(x)]$

Graph



Samples

D from P



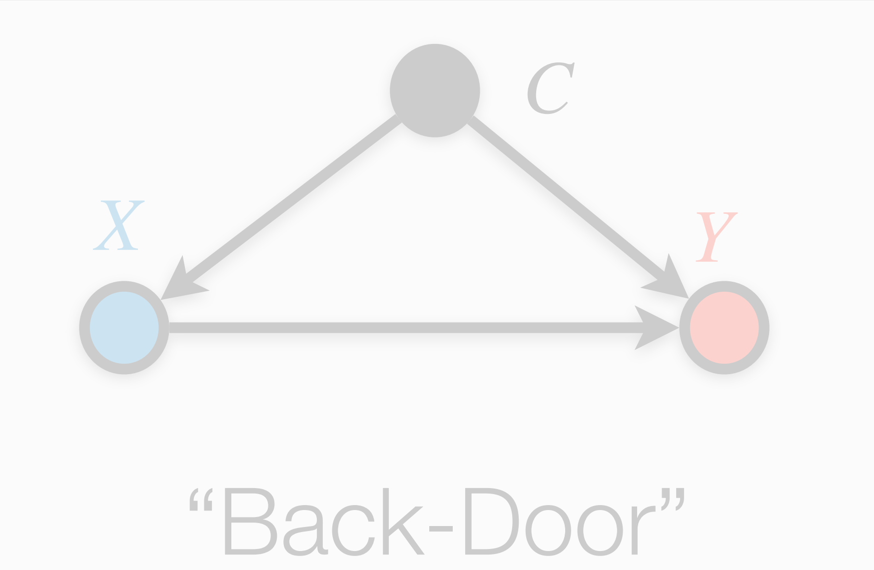
- Goal: Estimate $\mathbb{E}[Y \mid \text{do}(x_1, x_2)]$ from single interventions $\text{do}(x_1)$ and $\text{do}(x_2)$.
- Drug interactions between X_1 and X_2

Challenge 2: Data Fusion

Effect (Q)

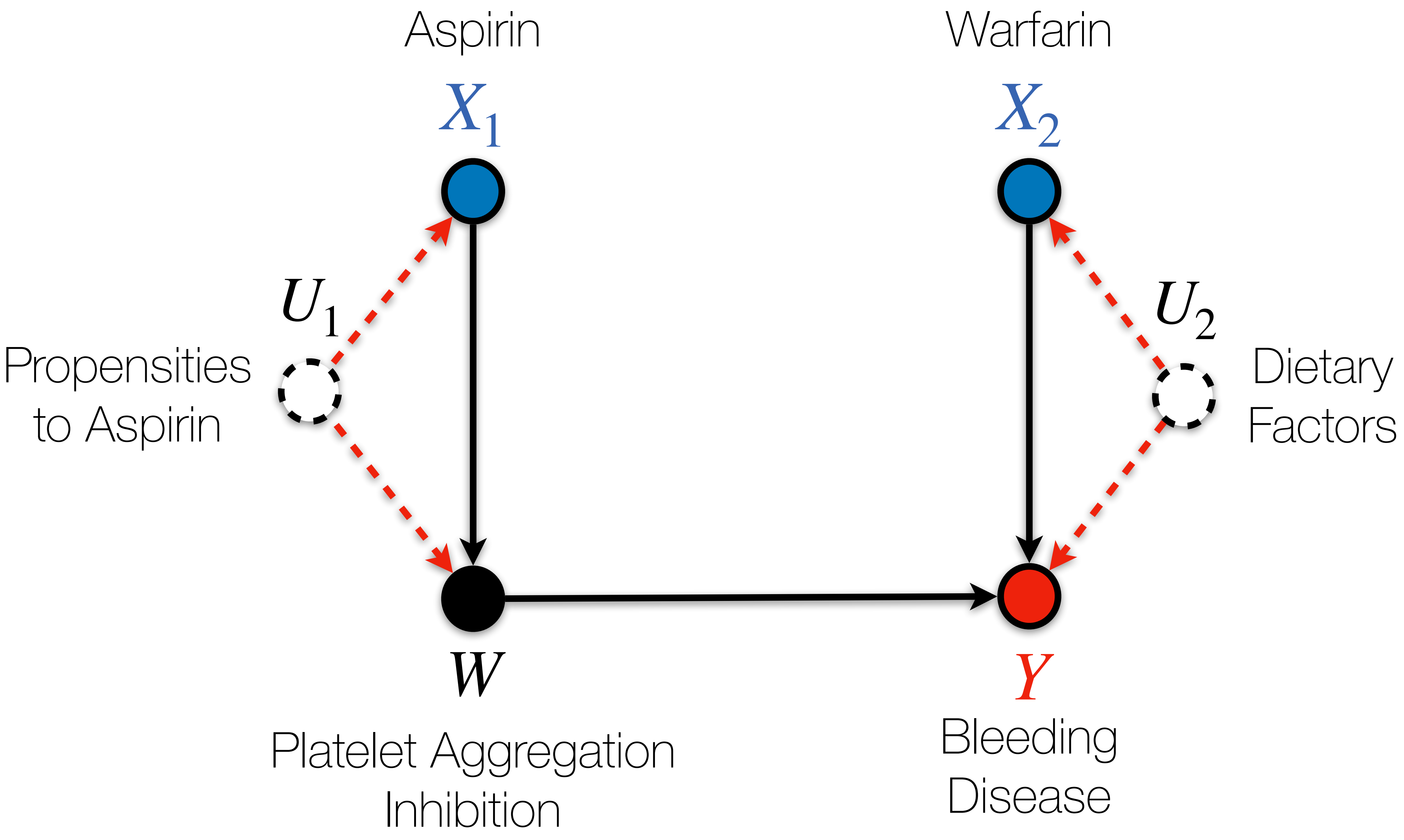
$\mathbb{E}[Y \mid \mathrm{do}(x)]$

Graph



Samples

D from P

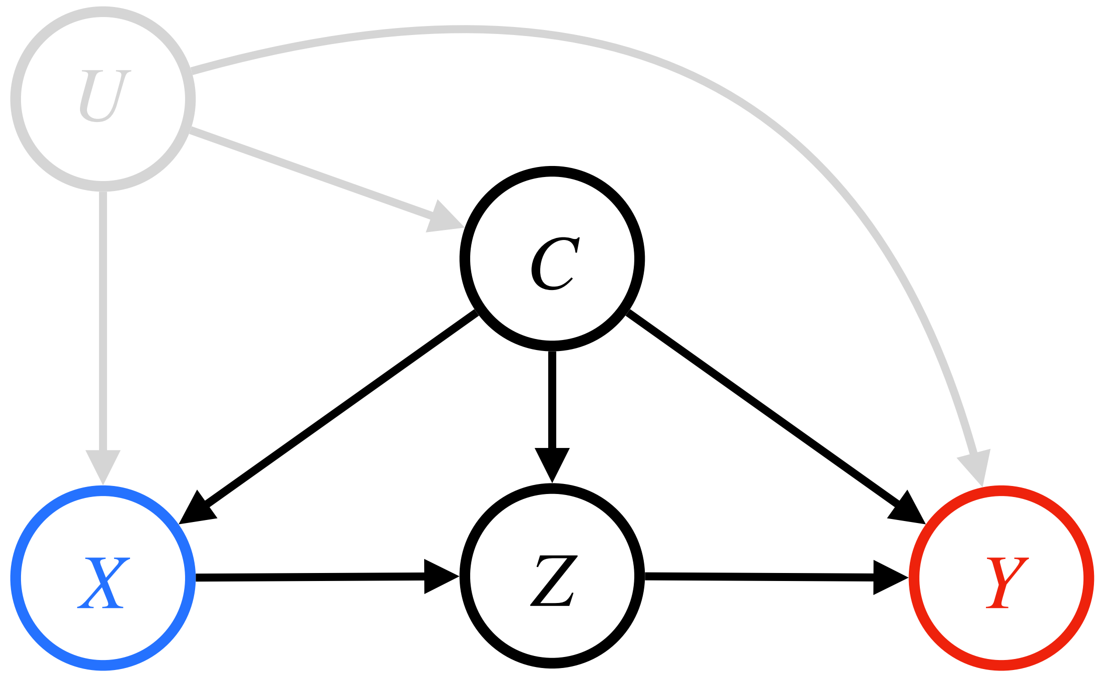


- Goal: Estimate $\mathbb{E}[Y \mid \mathrm{do}(x_1, x_2)]$ from single interventions $\mathrm{do}(x_1)$ and $\mathrm{do}(x_2)$.
- Drug interactions between X_1 and X_2
- Not identifiable from observations

Challenge 3: Computational Inefficiency

Challenge 3: Computational Inefficiency

Front-door Graph

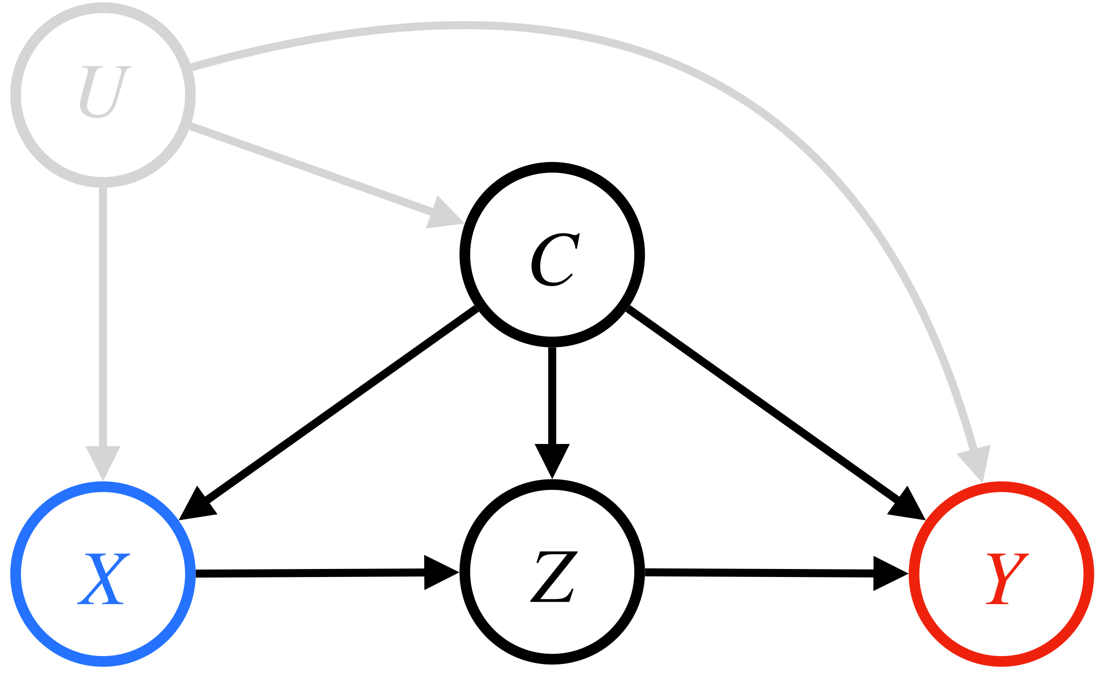


$$\mathbb{E}[Y \mid \mathrm{do}(x)]$$

$$= \sum_{zx'c} \mathbb{E}[Y \mid z, x', c] P(z \mid x, c) P(x'c)$$

Challenge 3: Computational Inefficiency

Front-door Graph



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

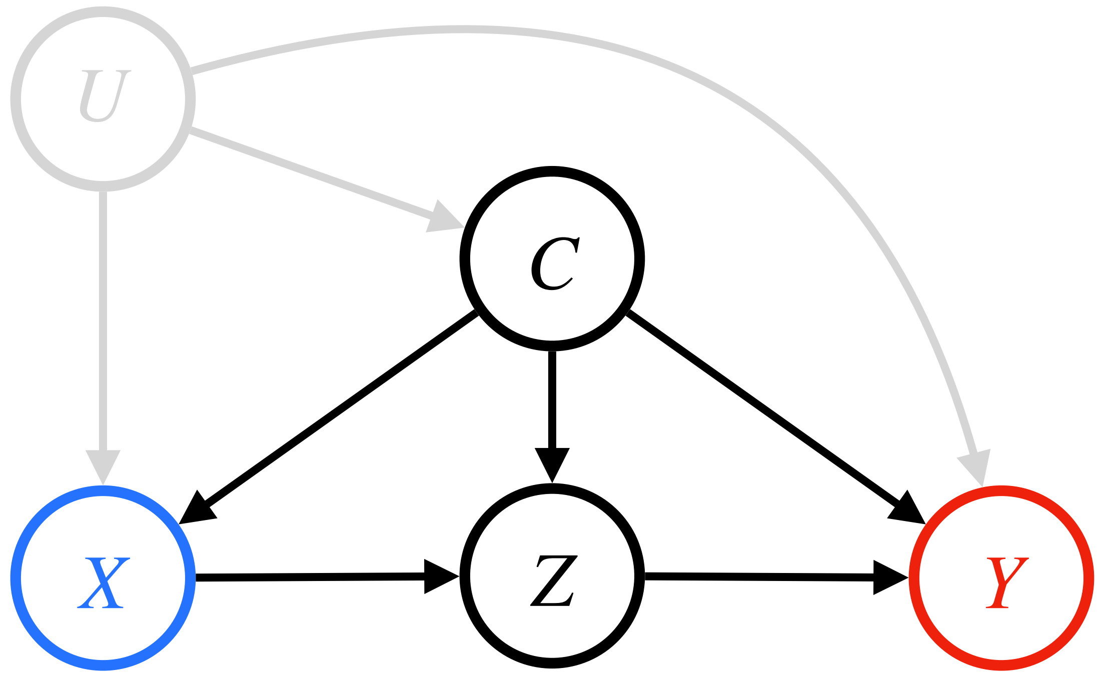
$$= \sum_{zx'c} \mathbb{E}[Y \mid z, x', c] P(z \mid x, c) P(x'c)$$

Treatments $\mathbf{X}$ fixed to $\mathbf{x}$ and marginalized $\mathbf{x}'$ simultaneously.

➡ Difficult to compute, because a standard g-computation (nested expectation) doesn't work.

Challenge 3: Computational Inefficiency

Front-door Graph



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

$$= \sum_{zx'c} \mathbb{E}[Y \mid z,x',c] P(z \mid x,c) P(x'c)$$

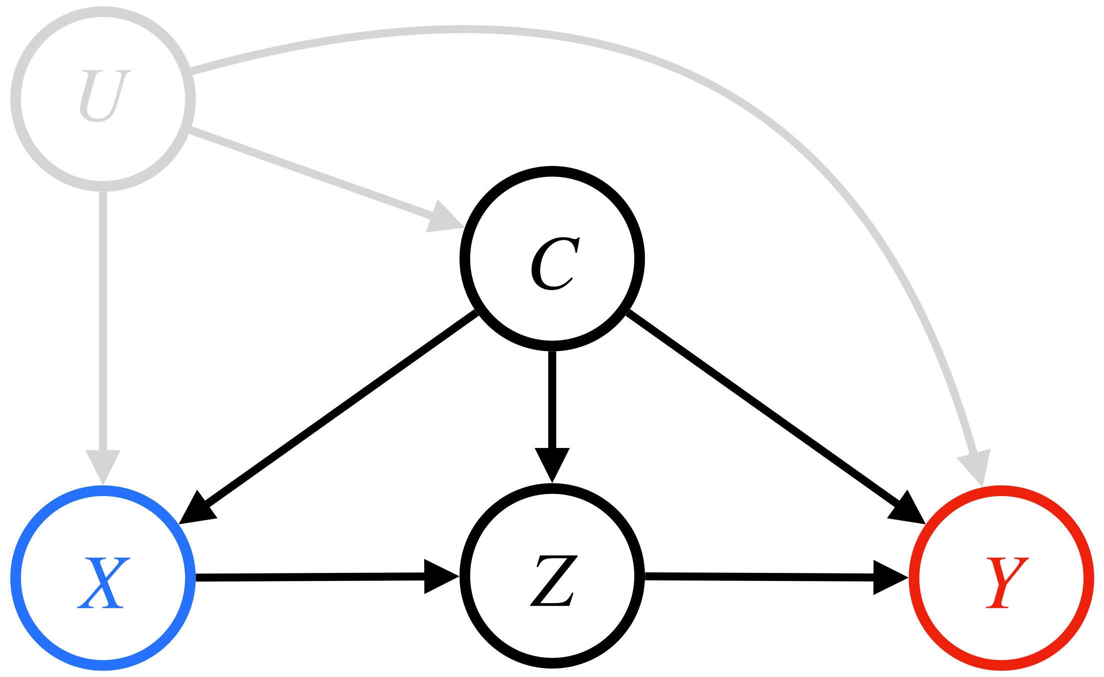
Treatments $\mathbf{X}$ fixed to $\mathbf{x}$ and marginalized $\mathbf{x}'$ simultaneously.

➡ Difficult to compute, because a standard g-computation (nested expectation) doesn't work.

G-computation doesn't work

Challenge 3: Computational Inefficiency

Front-door Graph



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

$$= \sum_{zx'c} \mathbb{E}[Y | z, x', c] P(z \mid x, c) P(x'c)$$

Treatments $\mathbf{X}$ fixed to $\mathbf{x}$ and marginalized $\mathbf{x}'$ simultaneously.

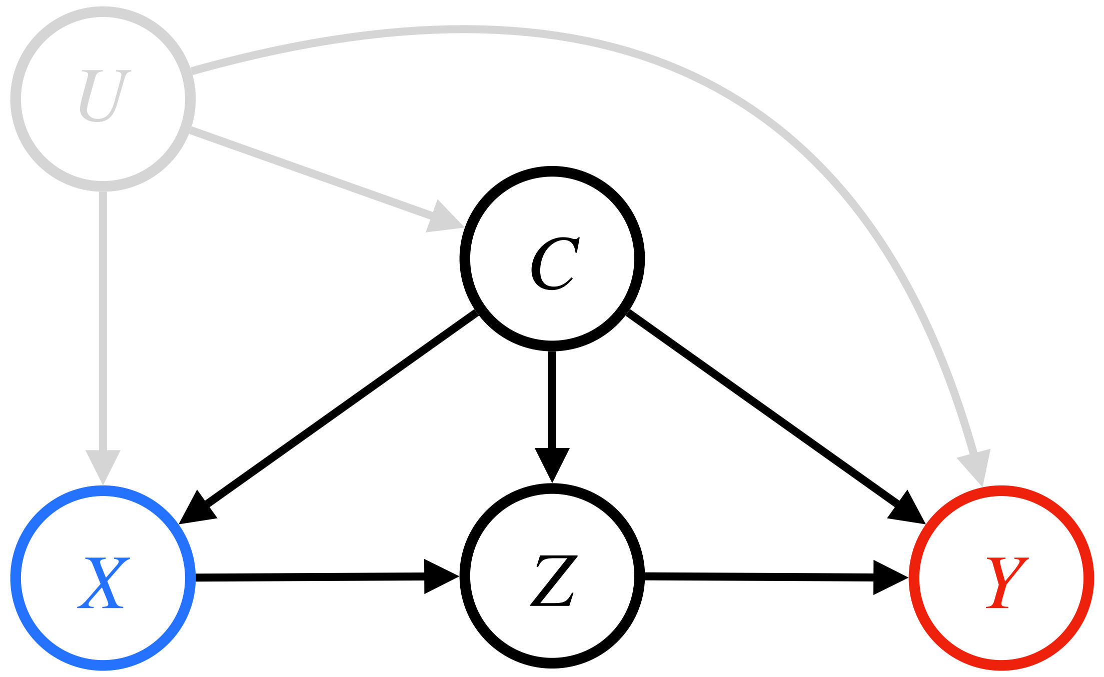
➡ Difficult to compute, because a standard g-computation (nested expectation) doesn't work.

G-computation doesn't work

❶ $\mu_2(Z, X, C) \triangleq \mathbb{E}[Y \mid Z, X, C]$

Challenge 3: Computational Inefficiency

Front-door Graph



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

$$= \sum_{zx'c} \mathbb{E}[Y \,|z,x',c]P(z \mid x,c)P(x'c)$$

Treatments $\mathbf{X}$ fixed to $\mathbf{x}$ and marginalized $\mathbf{x}'$ simultaneously.

➡ Difficult to compute, because a standard g-computation (nested expectation) doesn't work.

G-computation doesn't work

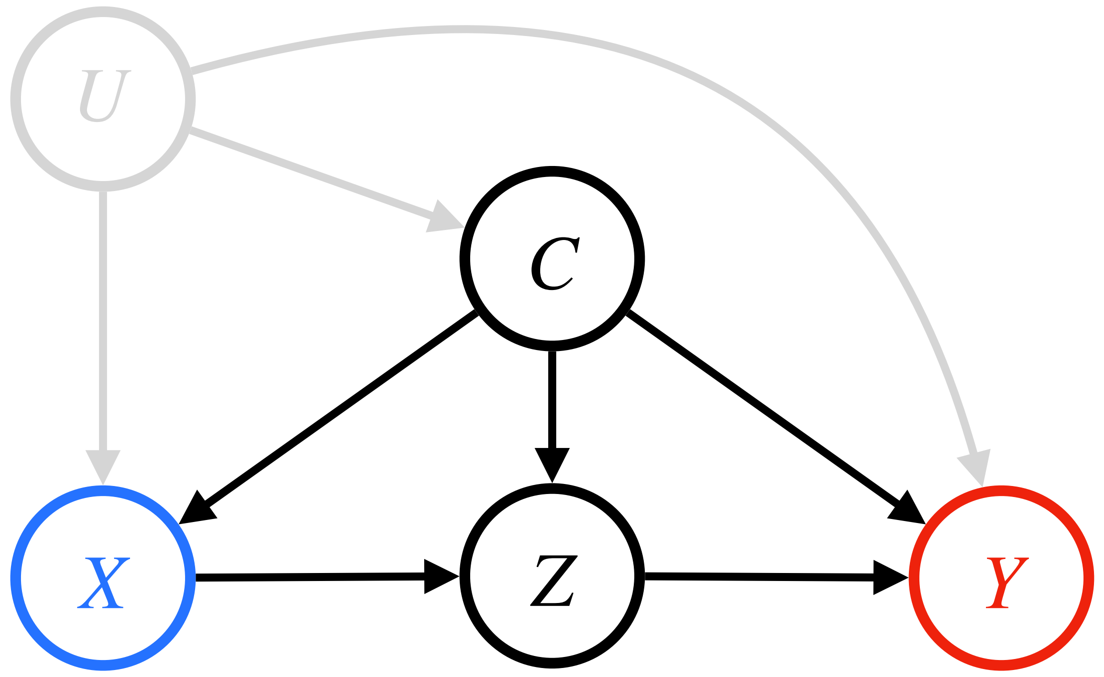
❶ $\mu_2(Z, X, C) \triangleq \mathbb{E}[Y \mid Z, X, C]$

❷ $\mu_1(X,C) \triangleq \mathbb{E}[\mu_2(Z, X, C) \,|X,C]$

$\sum_z \mathbb{E}[Y \,|z,X,C]P(z \mid X,C)$

Challenge 3: Computational Inefficiency

Front-door Graph



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

$$= \sum_{zx'c} \mathbb{E}[Y \mid z,x',c] P(z \mid x,c) P(x'c)$$

Treatments $\mathbf{X}$ fixed to $\mathbf{x}$ and marginalized $\mathbf{x}'$ simultaneously.

➡ Difficult to compute, because a standard g-computation (nested expectation) doesn't work.

G-computation doesn't work

❶ $\mu_2(Z, X, C) \triangleq \mathbb{E}[Y \mid Z, X, C]$

❷ $\mu_1(X,C) \triangleq \mathbb{E}[\mu_2(Z, X, C) \mid X,C]$

$$\sum_z \mathbb{E}[Y \mid z,X,C] P(z \mid X,C)$$

❸ $\mathbb{E}[\mu_1(x, C)] \neq \mathbb{E}[Y \mid \mathrm{do}(x)]$

$$\sum_{z,c} \mathbb{E}[Y \mid z,x,c] P(z \mid x,c) P(c)$$

Estimating Identifiable Causal Effects

Tasks

Challenges

1. Complicated dependences
2. Data fusion
(observations + experiments)
3. Computational Inefficiency

Estimating Identifiable Causal Effects

Tasks

Challenges

1. Estimating causal effects from observations → ✓

2. Data fusion (observations + experiments)

3. Computational Inefficiency

Estimating Identifiable Causal Effects

Tasks		Challenges
① Estimating causal effects from observations	→	✓
② Estimating causal effects from data fusion	→	✓
		③ Computational Inefficiency

Estimating Identifiable Causal Effects

Tasks	Challenges
❶ Estimating causal effects from observations	✓
❷ Estimating causal effects from data fusion	✓
❸ Unified and scalable estimators	✓

Talk Outline

❶ Estimating causal effects from observations

❷ Estimating causal effects from data fusion

❸ Unified and scalable estimation method

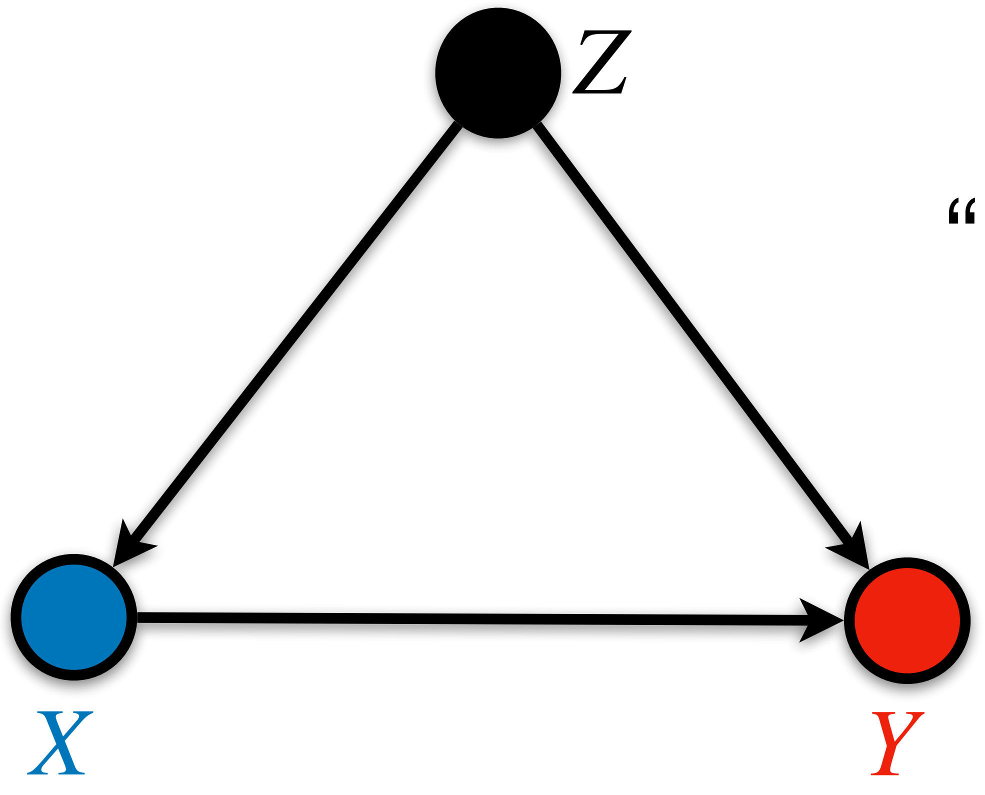
❹ Conclusion

Talk Outline

➡ ❶ Estimating causal effects from observations

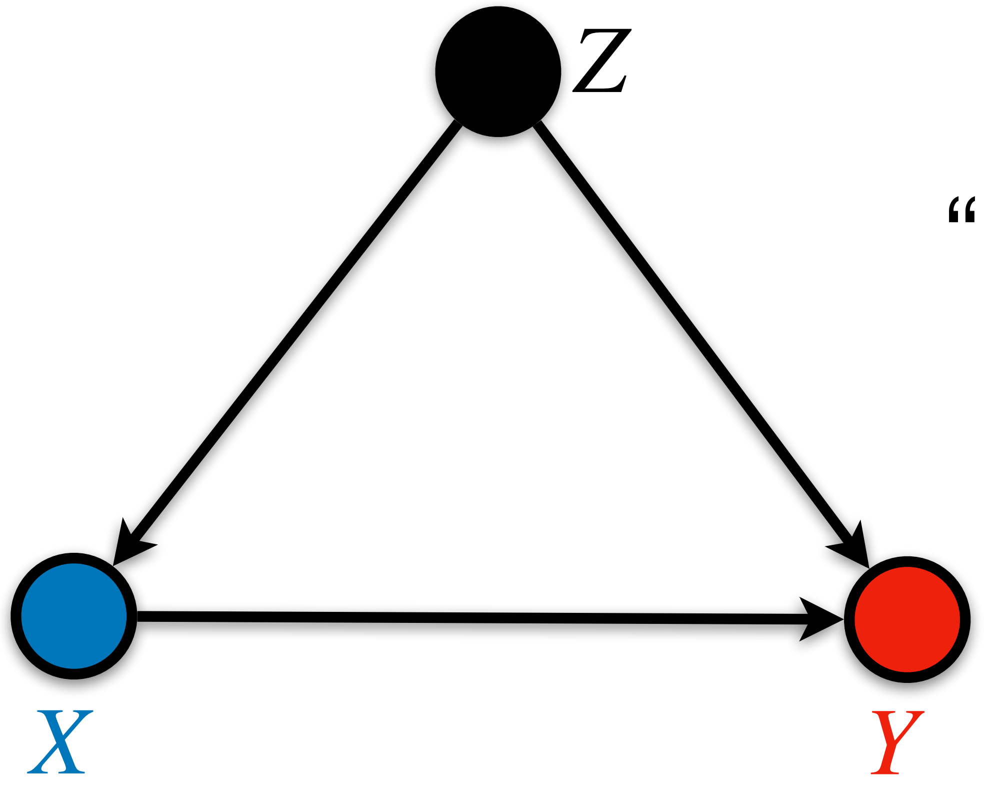
Background: Back-door Adjustment (BD)

Background: Back-door Adjustment (BD)



“Back-door graph”

Background: Back-door Adjustment (BD)



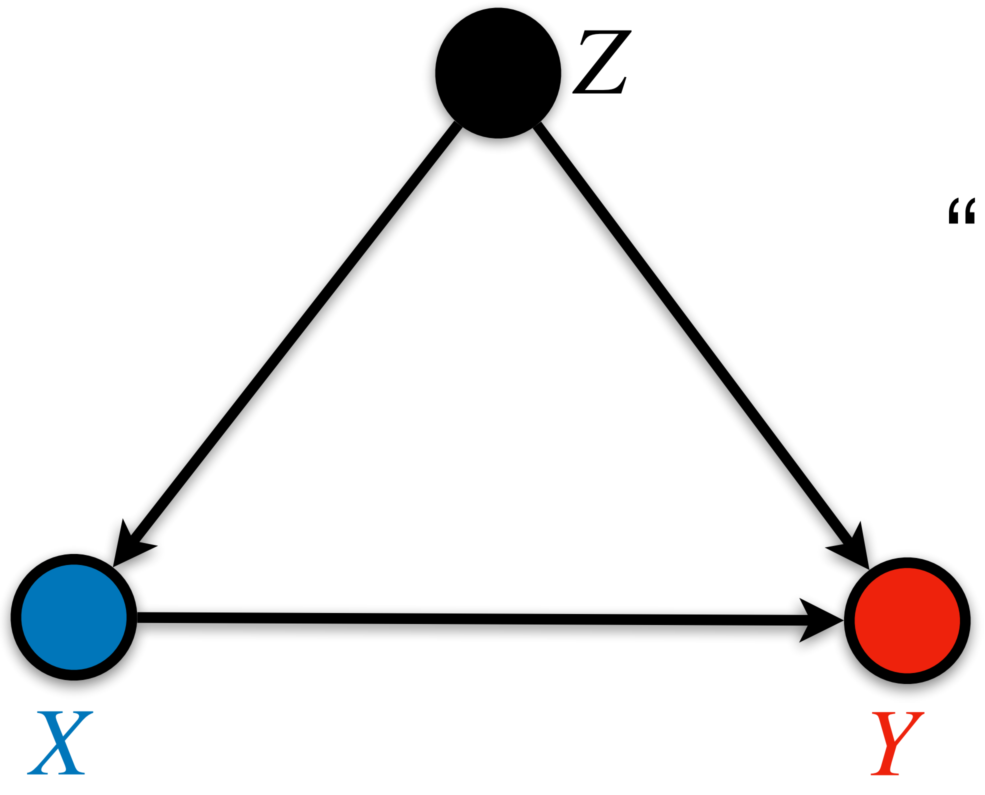
"Back-door graph"

Back-door Criterion

(Pearl 95)

1. $\mathbf{Z}$ is not a descendent of treatment;
2. $\mathbf{Z}$ blocks spurious paths between (treatments, outcome)

Background: Back-door Adjustment (BD)



"Back-door graph"

Back-door Criterion

(Pearl 95)

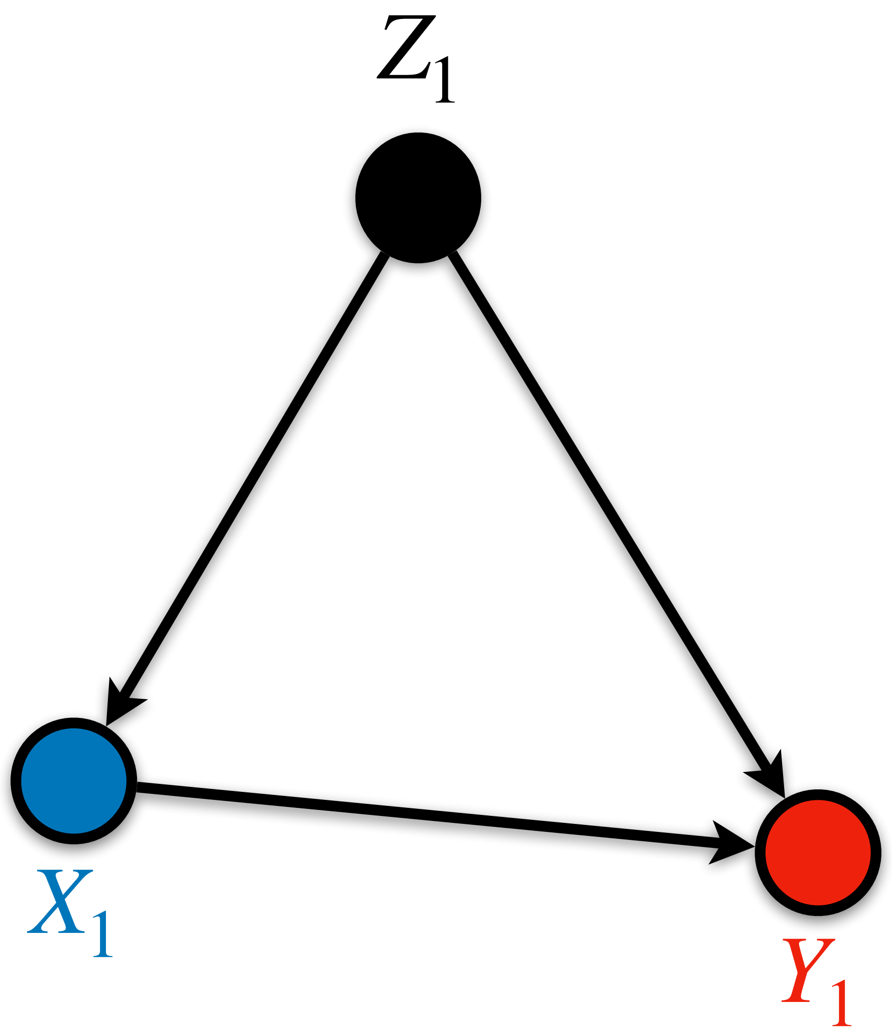
1. $\mathbf{Z}$ is not a descendent of treatment;
2. $\mathbf{Z}$ blocks spurious paths between (treatments, outcome)

"Back-door adjustment (BD)"

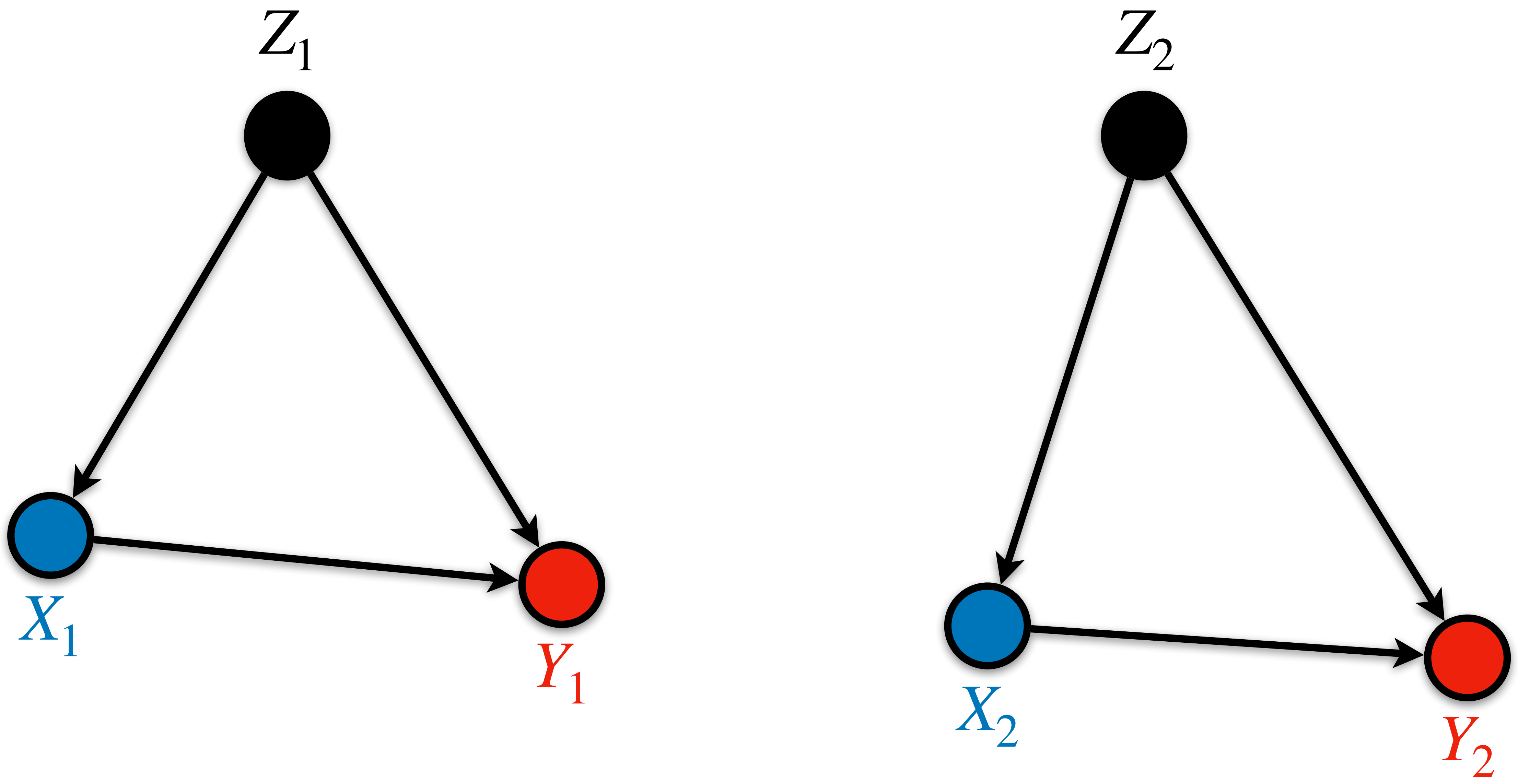
$$P(y \mid \mathrm{do}(x)) = \text{BD} \triangleq \sum_z P(y \mid x, z)P(z)$$

Background: Multi-outcome sequential BD

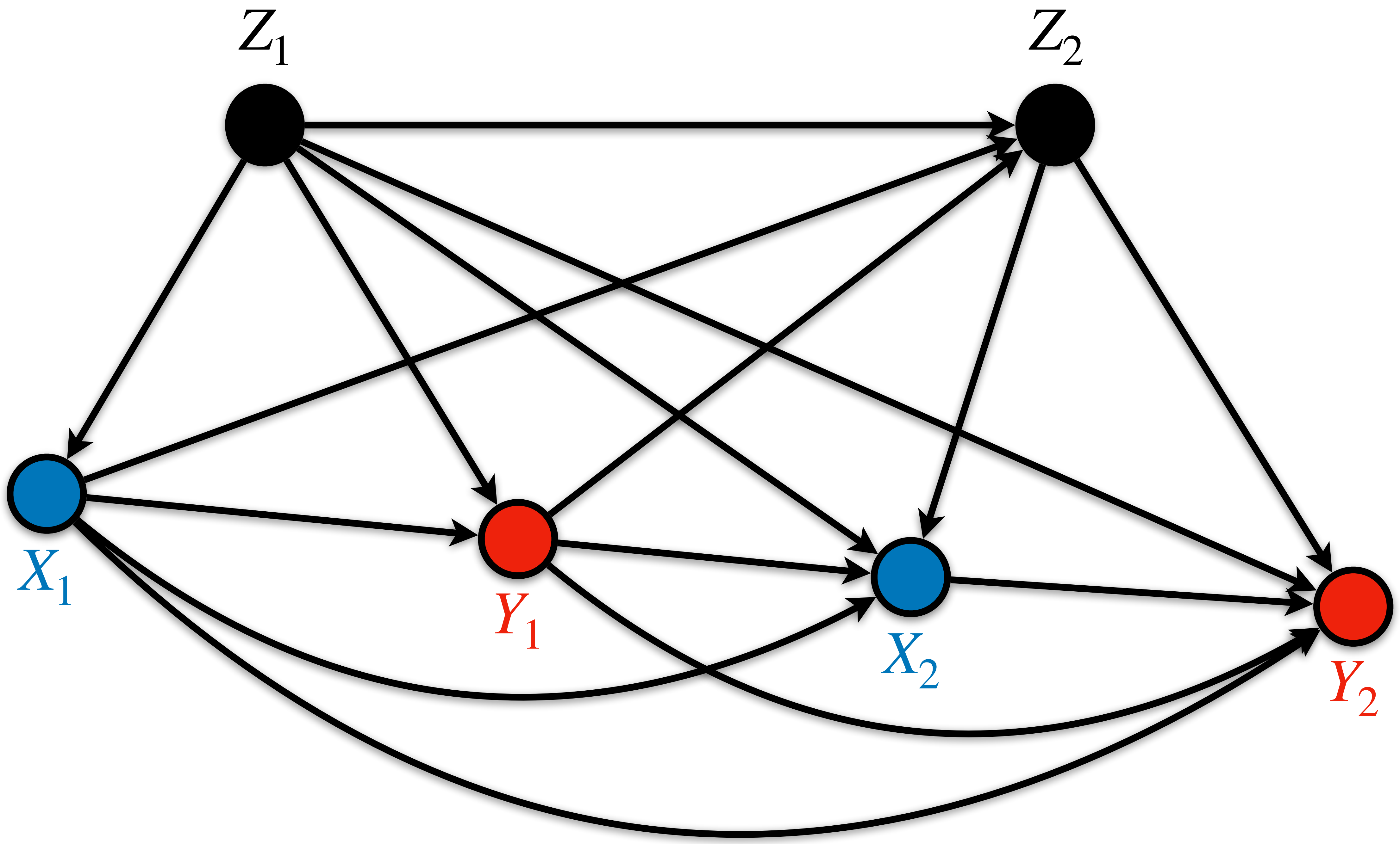
Background: Multi-outcome sequential BD



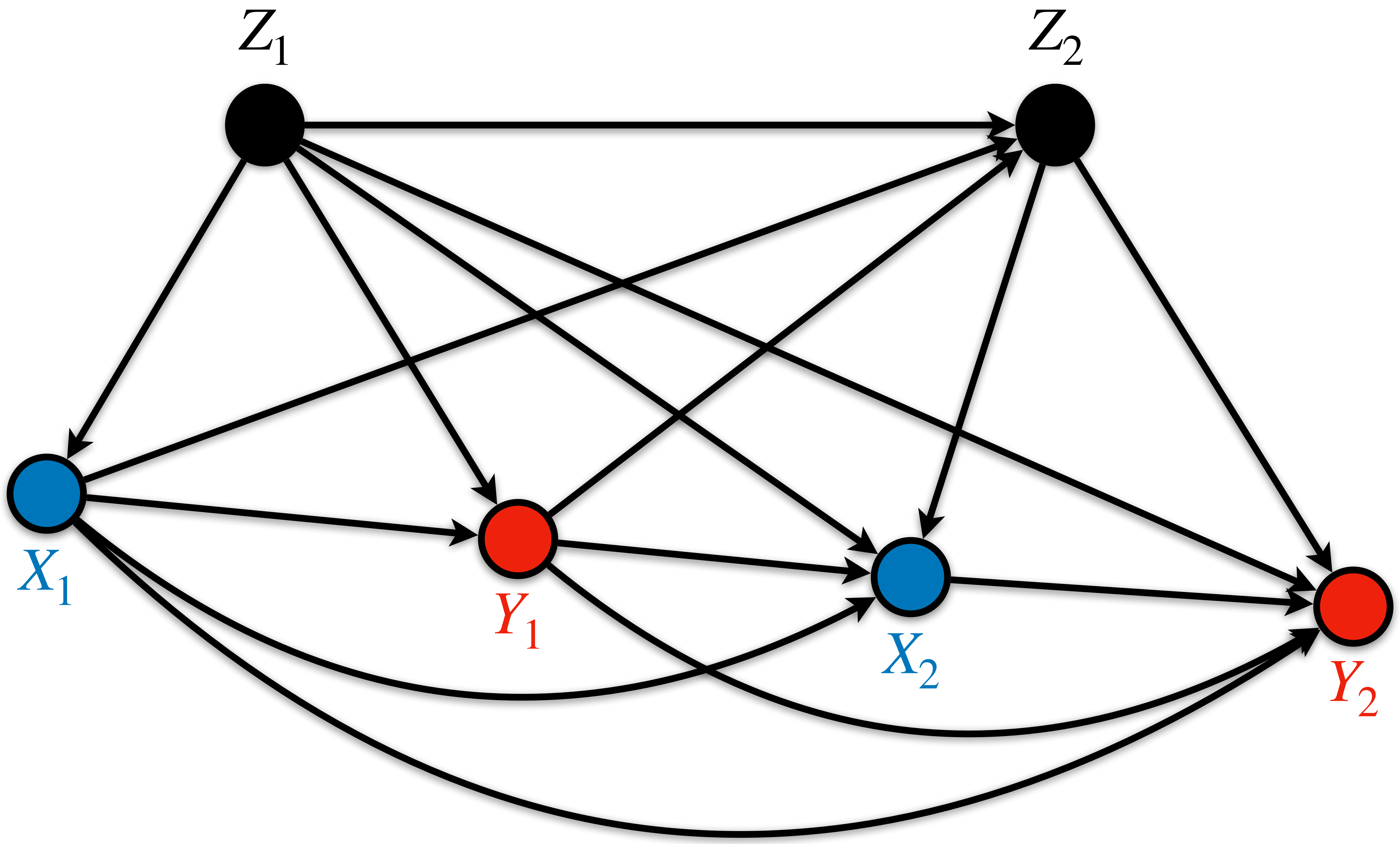
Background: Multi-outcome sequential BD



Background: Multi-outcome sequential BD



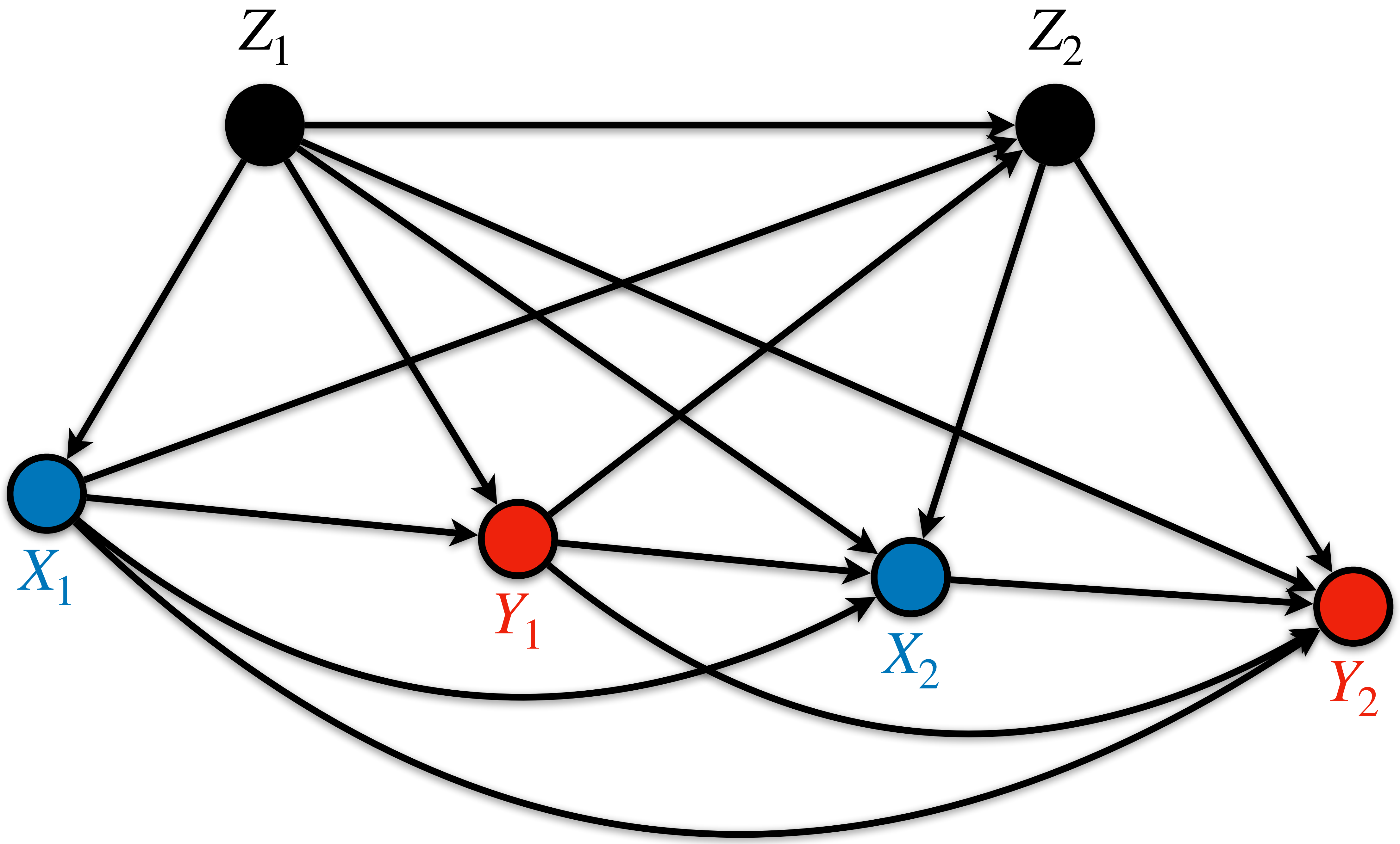
Background: Multi-outcome sequential BD



Multi-outcome Sequential BD (mSBD)

A seq. $\mathbf{Z} = (\mathbf{Z}_1, \cdots, \mathbf{Z}_m)$ satisfies the mSBD if, for $i = 1, \cdots, m$, $\mathbf{Z}_i$ satisfies the BD relative to $(\mathbf{X}_i, \mathbf{Y}^{\geq i})$ conditioning on prev. vectors.

Background: Multi-outcome sequential BD



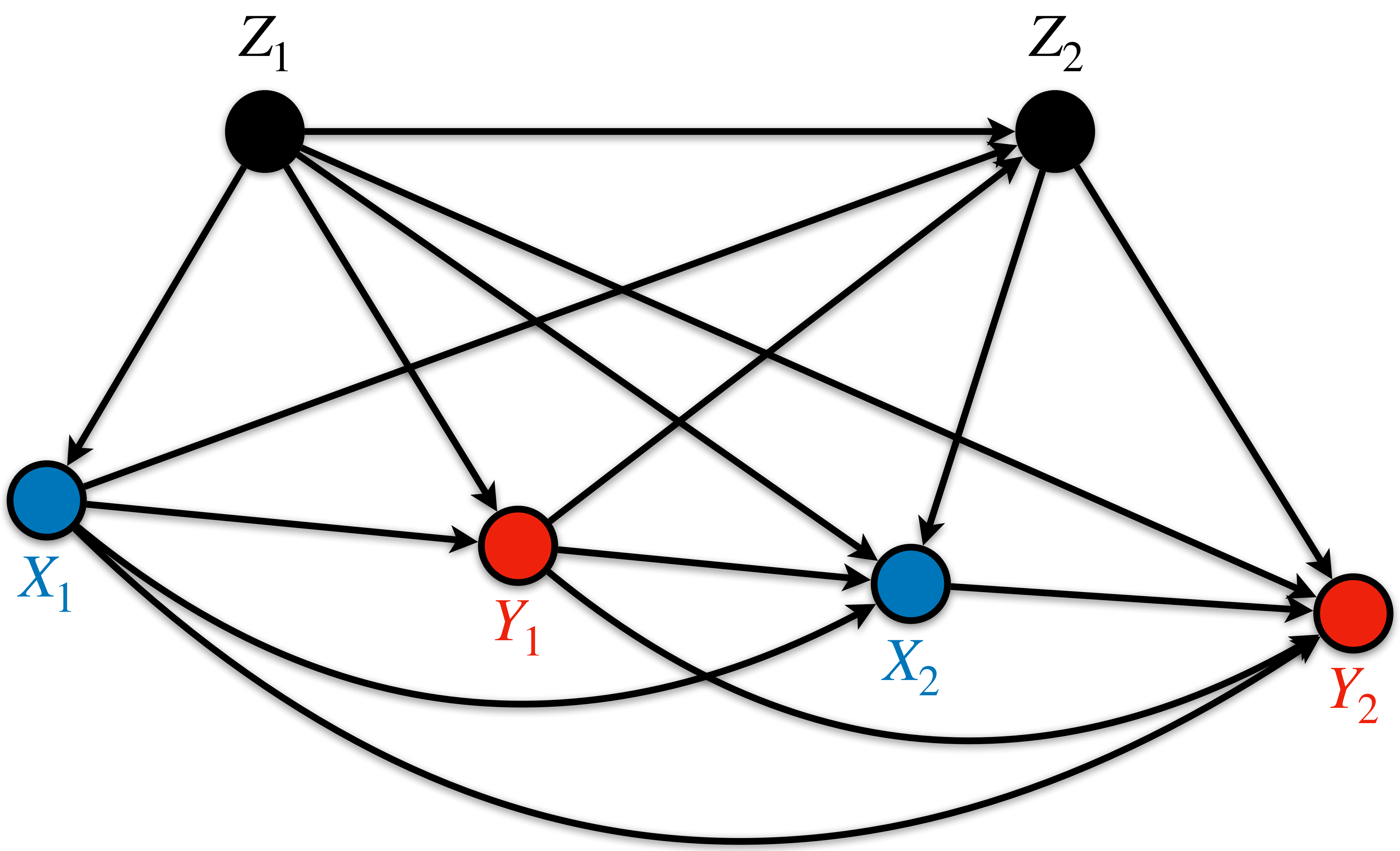
Multi-outcome Sequential BD (mSBD)

A seq. $\mathbf{Z} = (\mathbf{Z}_1, \cdots, \mathbf{Z}_m)$ satisfies the mSBD if, for $i = 1,\cdots, m$, $\mathbf{Z}_i$ satisfies the BD relative to $(\mathbf{X}_i, \mathbf{Y}^{\geq i})$ conditioning on prev. vectors.

$$P(\mathbf{y} \mid \mathrm{do}(\mathbf{x})) = \sum_{\mathbf{z}} \prod_{i=0}^{m+1} P(\mathbf{z}_{i+1}, \mathbf{y}_i \mid \mathrm{prev}_{i-1}, \mathbf{x}_i, \mathbf{z}_i)$$

“mSBD adjustment”

Background: Multi-outcome sequential BD



Multi-outcome Sequential BD (mSBD)

A seq. $\mathbf{Z} = (\mathbf{Z}_1, \cdots, \mathbf{Z}_m)$ satisfies the mSBD if, for $i = 1, \cdots, m$, $\mathbf{Z}_i$ satisfies the BD relative to $(\mathbf{X}_i, \mathbf{Y}^{\geq i})$ conditioning on prev. vectors.

$$P(\mathbf{y} \mid \mathrm{do}(\mathbf{x})) = \sum_{\mathbf{z}} \prod_{i=0}^{m+1} P(\mathbf{z}_{i+1}, \mathbf{y}_i \mid \mathrm{prev}_{i-1}, \mathbf{x}_i, \mathbf{z}_i)$$

“mSBD adjustment”

* I’ll use “BD” for simplicity, but all results extend to mSBD.

Background: Robust Estimator for BD

Background: 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 $\pi(XC) \triangleq \dfrac{\mathbb{I}_x(X)}{P(X \mid C)}$

Background: Robust Estimator for BD

One-step/Debiased ML estimator (Robins and Rotnitzk, 95; Band and Robins; 2005, van der Laan and Rubin 2006, van der Laan and Gruber 2012, Chernozhukov et al., 2018))

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

(i.e., $\text{avg}(\ \hat{\pi}(XC)(Y - \hat{\mu}(XC)) + \hat{\mu}(xC)$) with sample-splitting)

Background: Robust Estimator for BD

One-step/Debiased ML estimator (Robins and Rotnitzk, 95; Band and Robins; 2005, van der Laan and Rubin 2006, van der Laan and Gruber 2012, Chernozhukov et al., 2018))

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

(i.e., $\text{avg}(\ \hat{\pi}(XC)(Y-\hat{\mu}(XC))+\hat{\mu}(xC)$) with sample-splitting)

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

Background: Robust Estimator for BD

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

(i.e., $\text{avg}(\ \hat{\pi}(XC)(Y - \hat{\mu}(XC)) + \hat{\mu}(xC)\ )$ with sample-splitting)

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

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

Background: Robust Estimator for BD

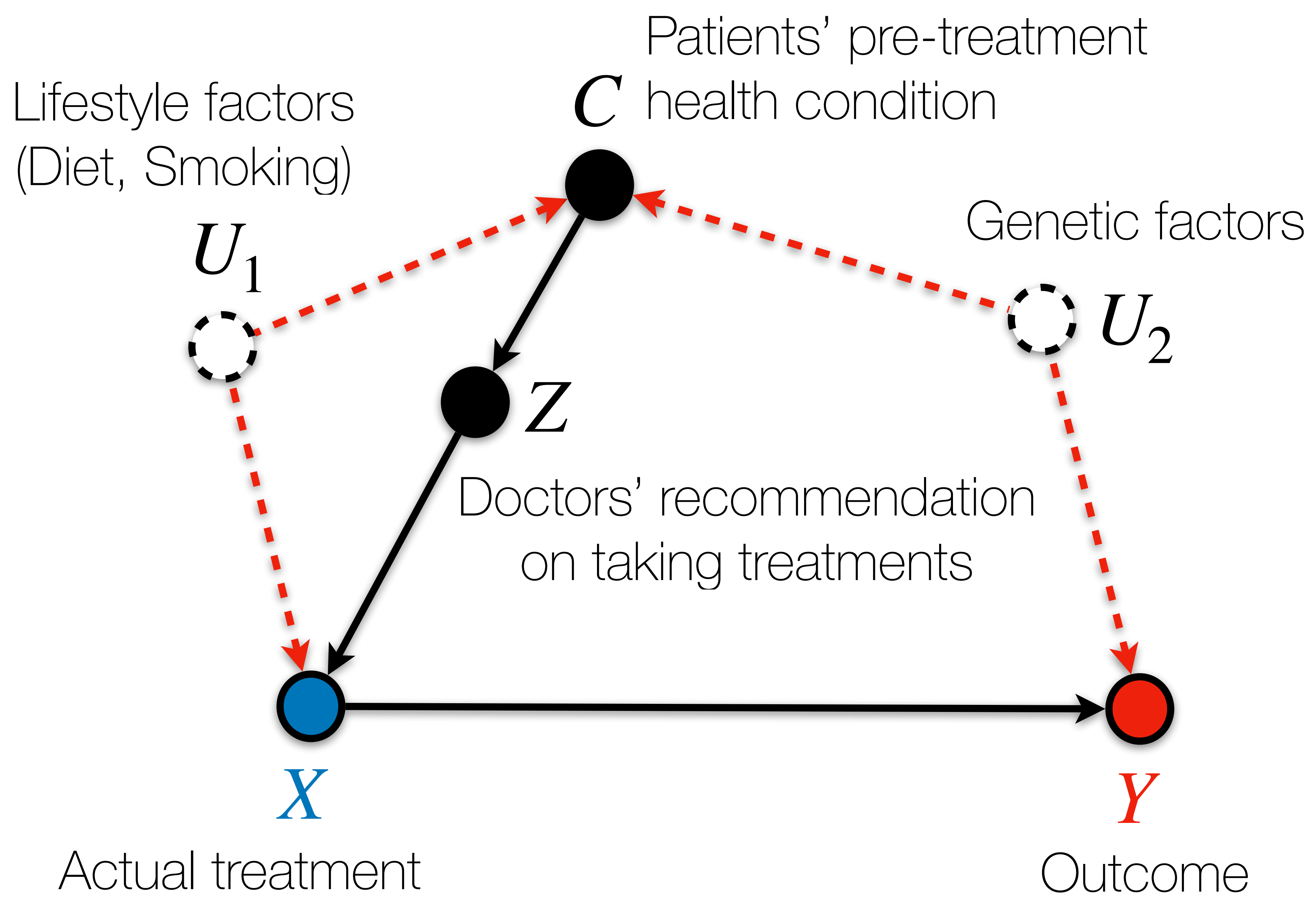
(i.e., $\text{avg}(\ \hat{\pi}(XC)(Y - \hat{\mu}(XC)) + \hat{\mu}(xC)$) with sample-splitting)

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

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

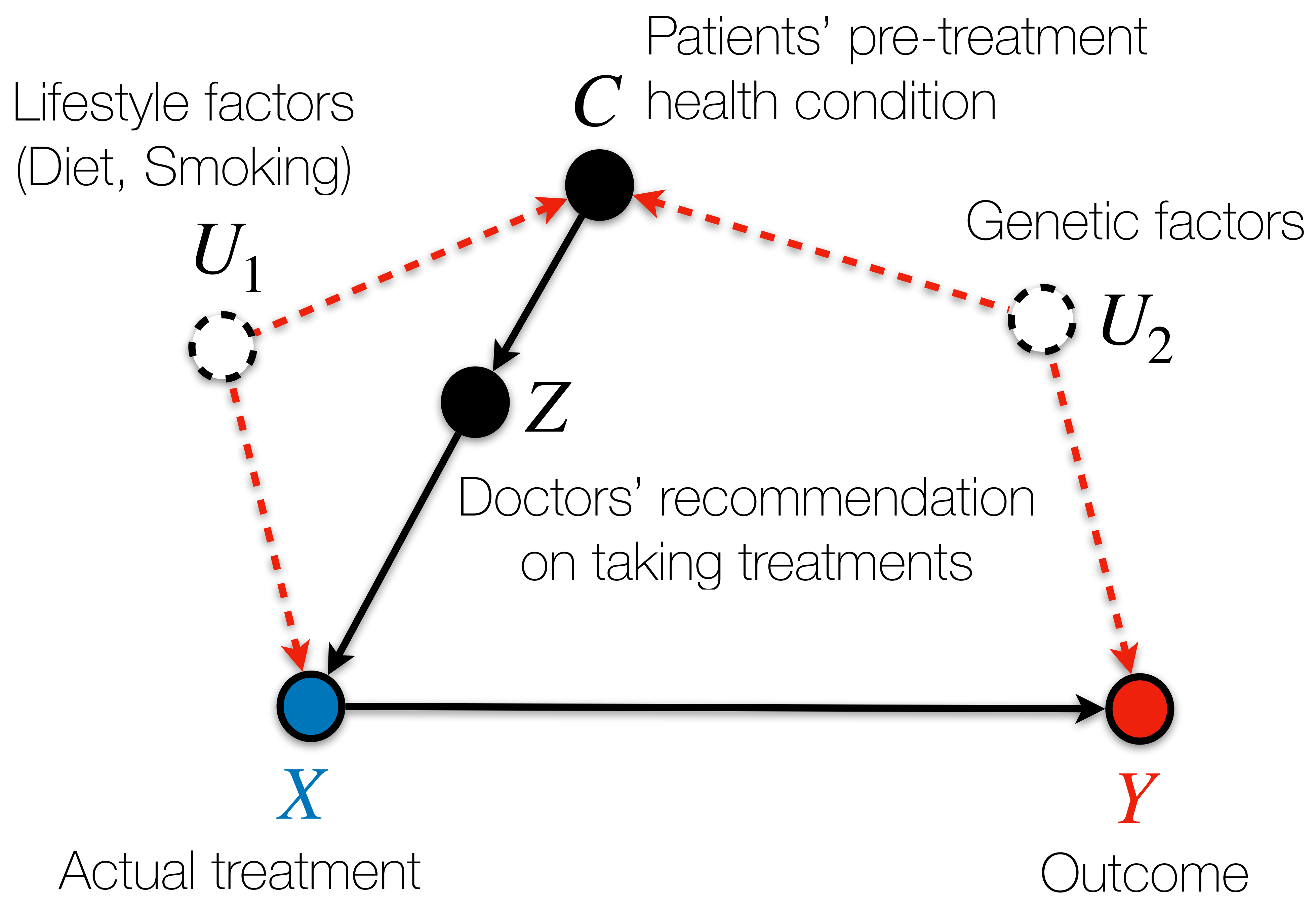
Non-BD Example: “Napkin Graph”

Non-BD Example: “Napkin Graph”



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

Non-BD Example: "Napkin Graph"

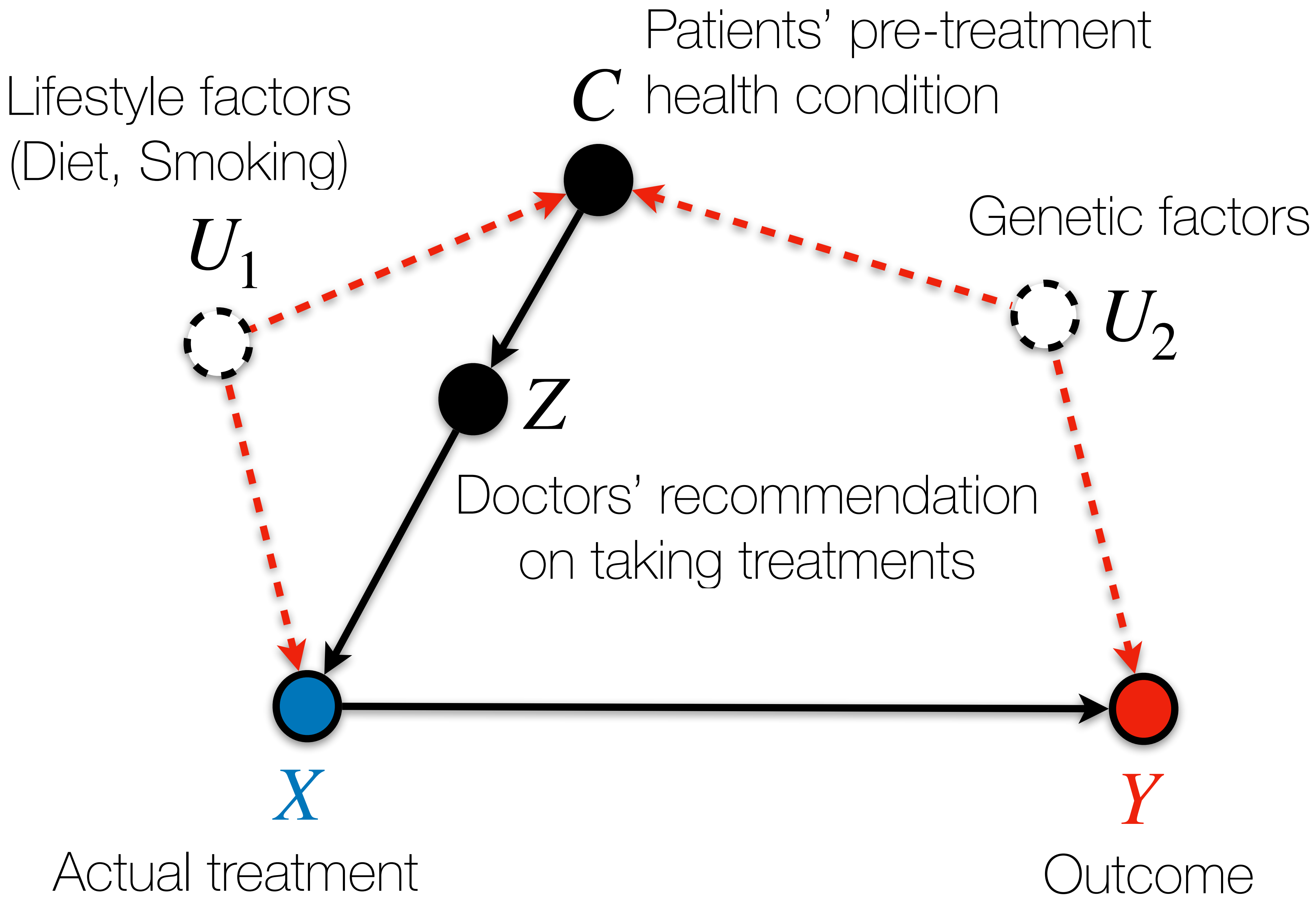


$$\mathbb{E}[Y \mid \mathrm{do}(x)]$$

Identification

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

Non-BD Example: “Napkin Graph”



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

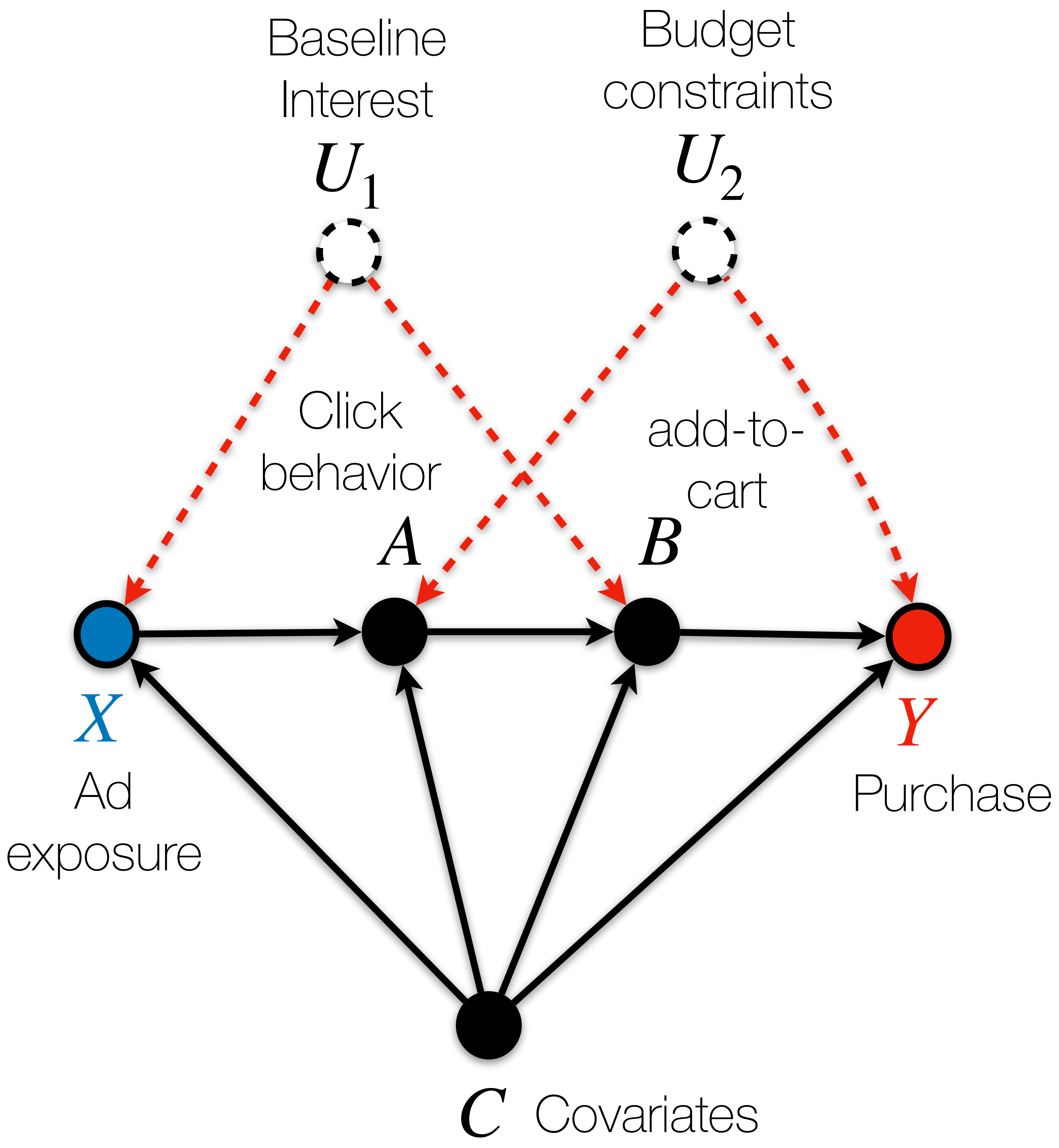
Identification

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

Estimation

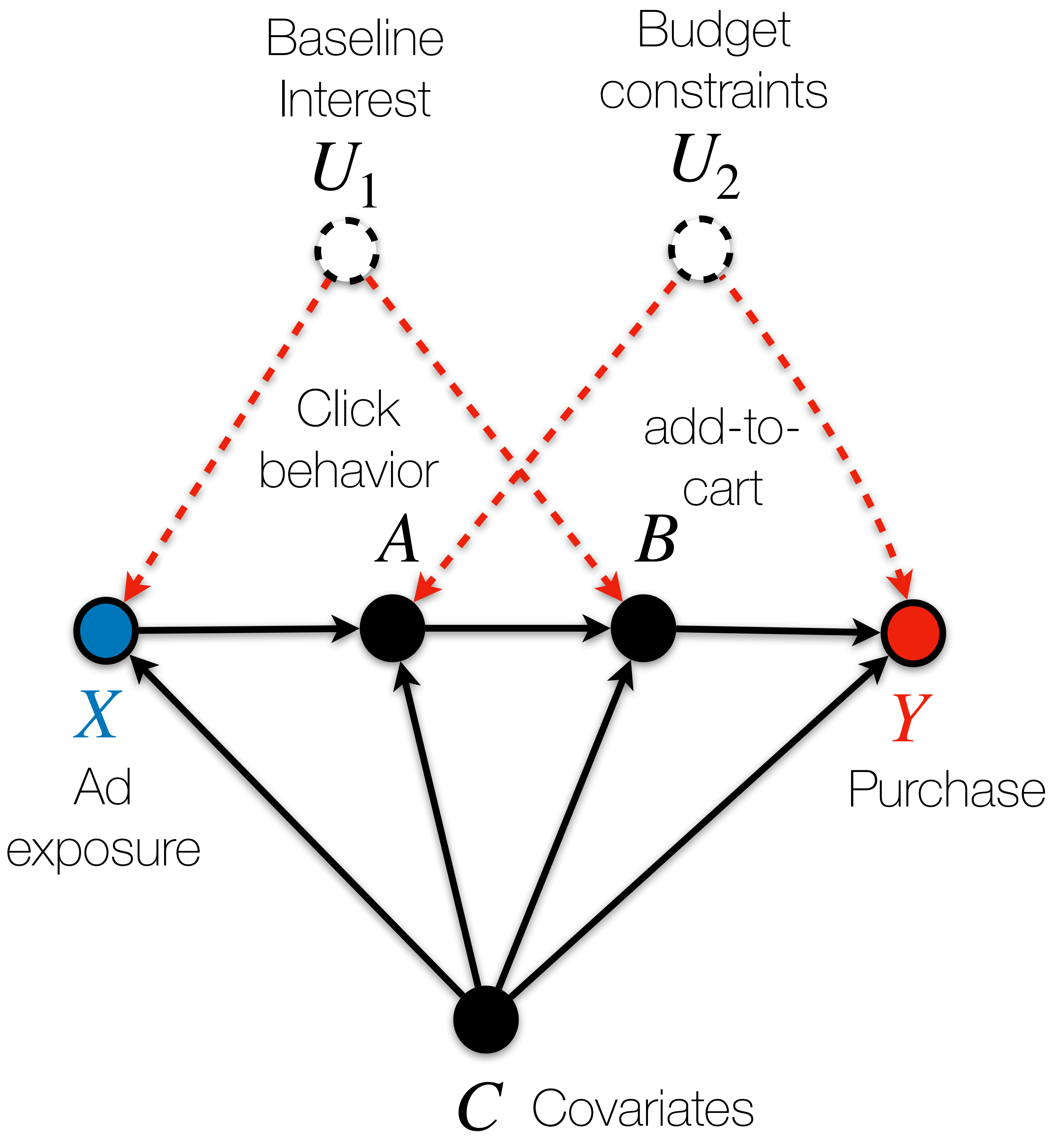
$$\widehat{\mathbb{E}[Y \mid \text{do}(x)]} = \textbf{?}$$

Non-BD Example: “Verma Graph”



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

Non-BD Example: “Verma Graph”



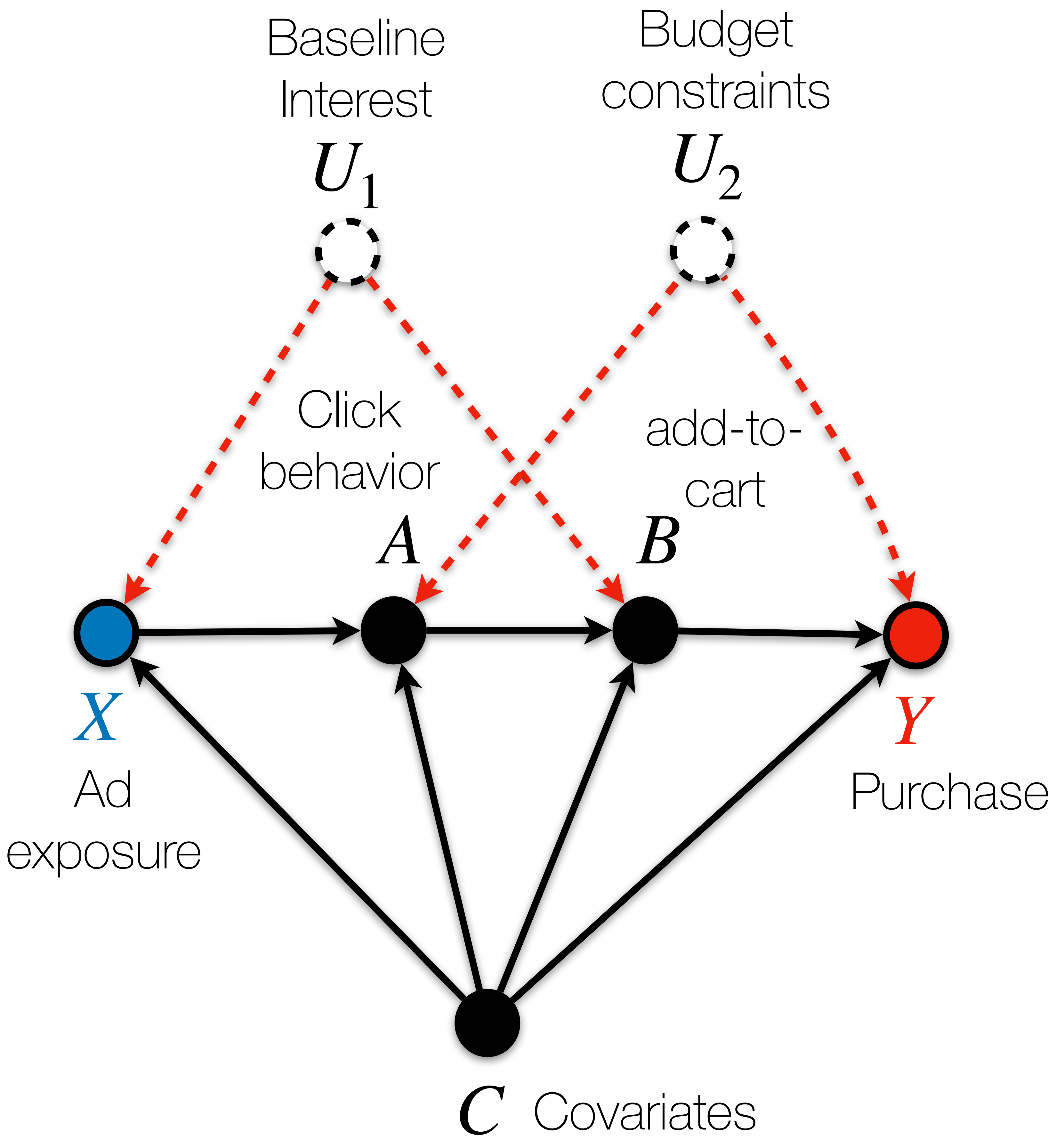
$$\mathbb{E}[Y \mid \text{do}(x)]$$

Identification

$$\mathbb{E}[Y \mid \text{do}(x)] = \sum_{bax'c} \mathbb{E}[Y \mid baxc] P(b|ax'c) P(a|xc) P(xc)$$

X is fixed to x and marginalized out (x') at the same time

Non-BD Example: “Verma Graph”



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

Identification

$$\mathbb{E}[Y \mid \mathrm{do}(x)] = \sum_{bax'c} \mathbb{E}[Y \mid baxc] P(b|ax'c) P(a|xc) P(xc)$$

X is fixed to x and marginalized out (x') at the same time

Estimation

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

Gap bw Identification & Estimation

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

Gap bw Identification & Estimation

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

Gap bw Identification & Estimation

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

Gap bw Identification & Estimation

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

Background: Causal Effect Identification

Background: Causal Effect Identification

Causal Effect 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

Causal Effect Identification

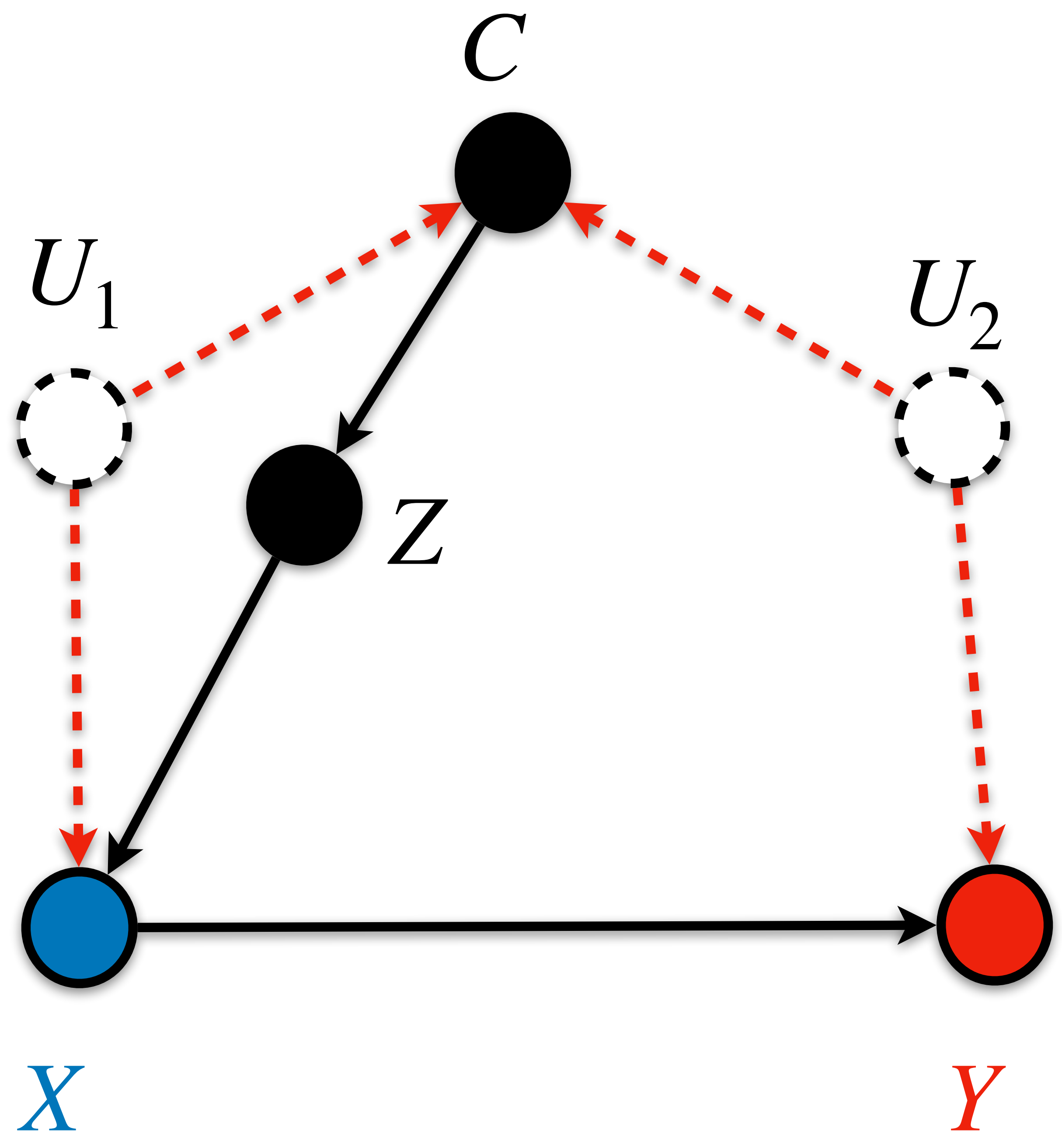
- spanning a *tree* from $P(\mathbf{V})$
- to reach to causal distribution $P(Y \mid \mathrm{do}(X))$
- through factorization & marginalization of distributions

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

Background: Causal Effect Identification

Causal Effect Identification

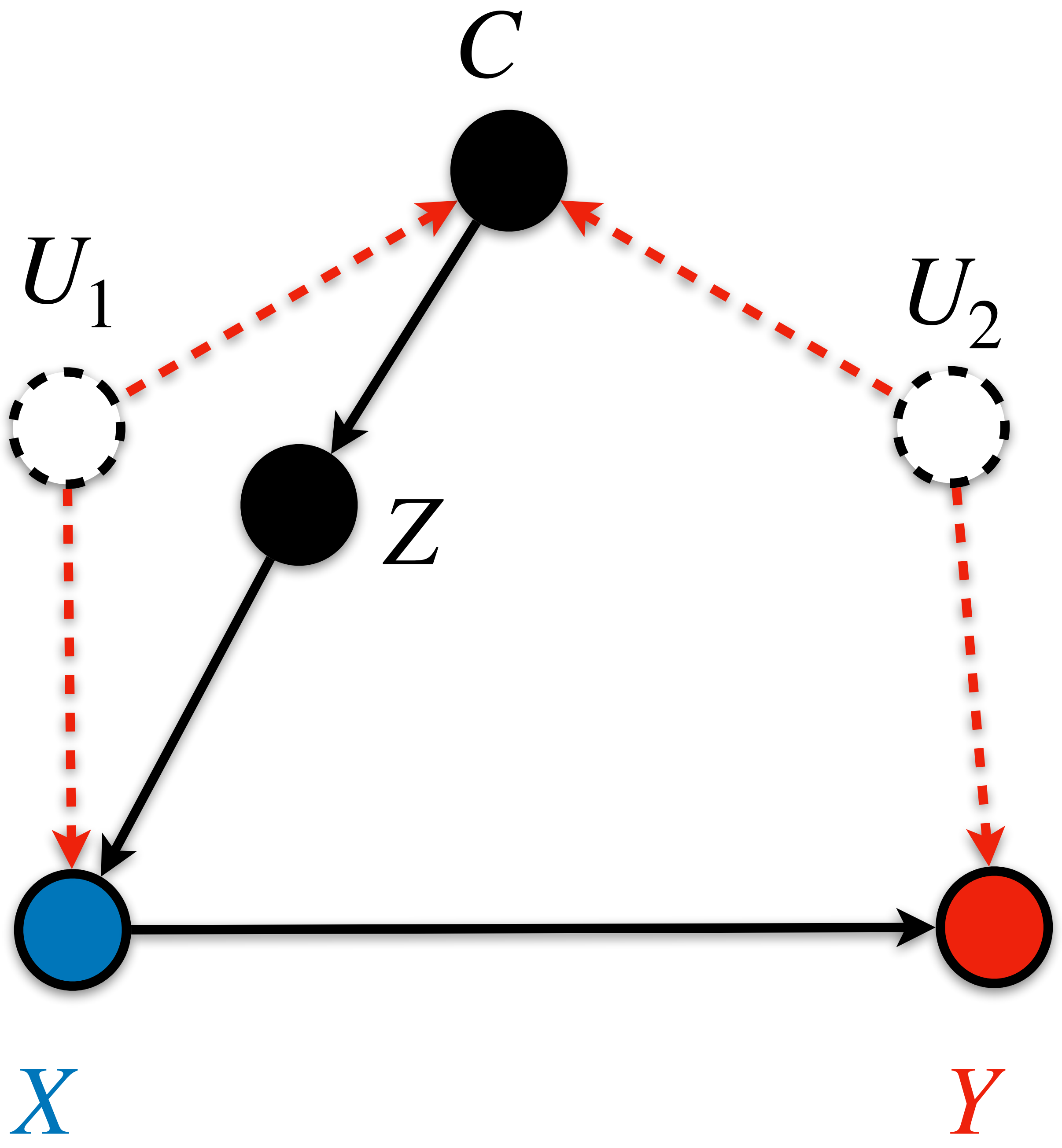
- spanning a *tree* from $P(\mathbf{V})$
- to reach to causal distribution $P(Y \mid \mathrm{do}(X))$
- through factorization & marginalization of distributions



Background: Causal Effect Identification

Causal Effect Identification

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

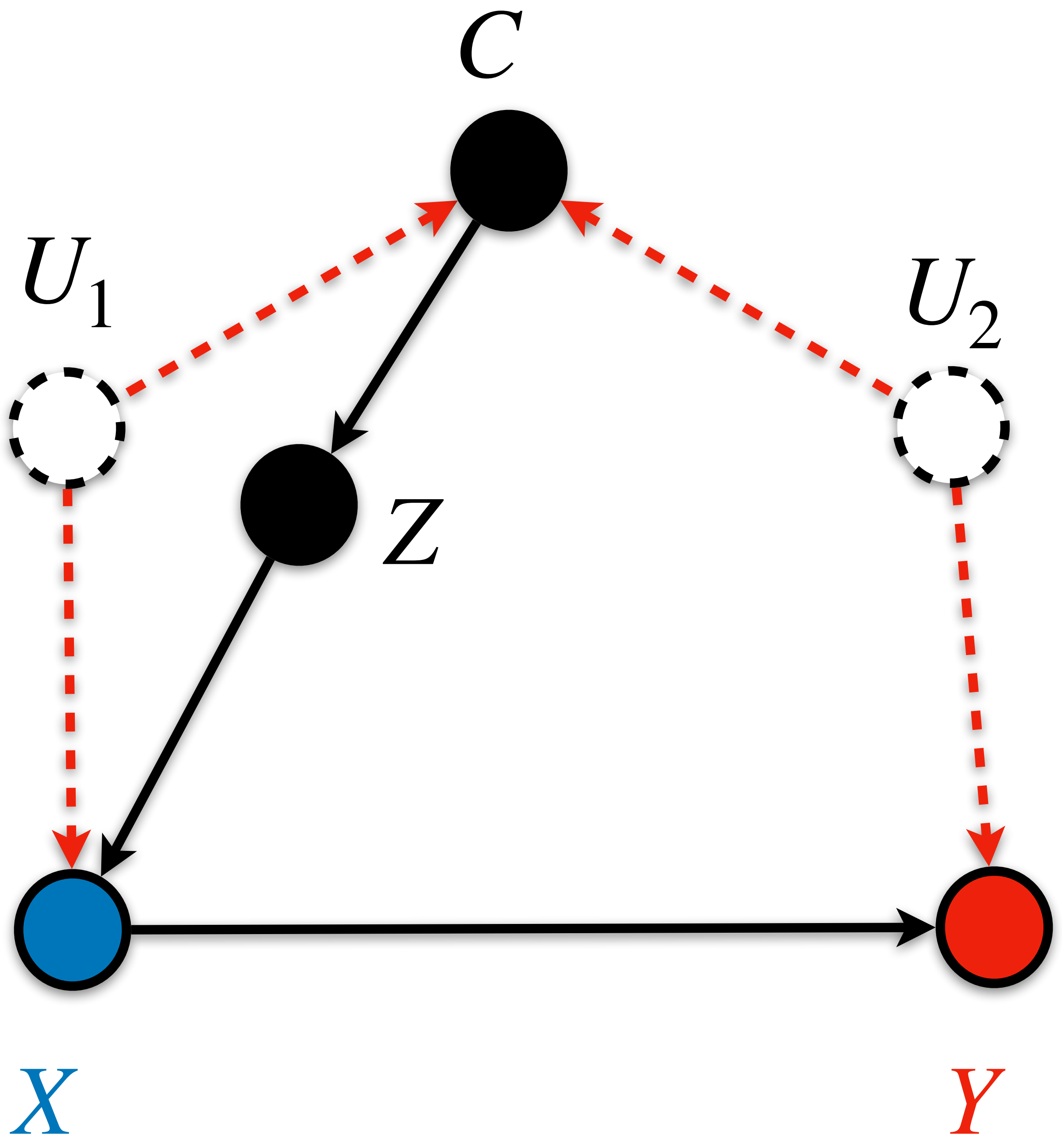


$P(CZXY)$

Background: Causal Effect Identification

Causal Effect Identification

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



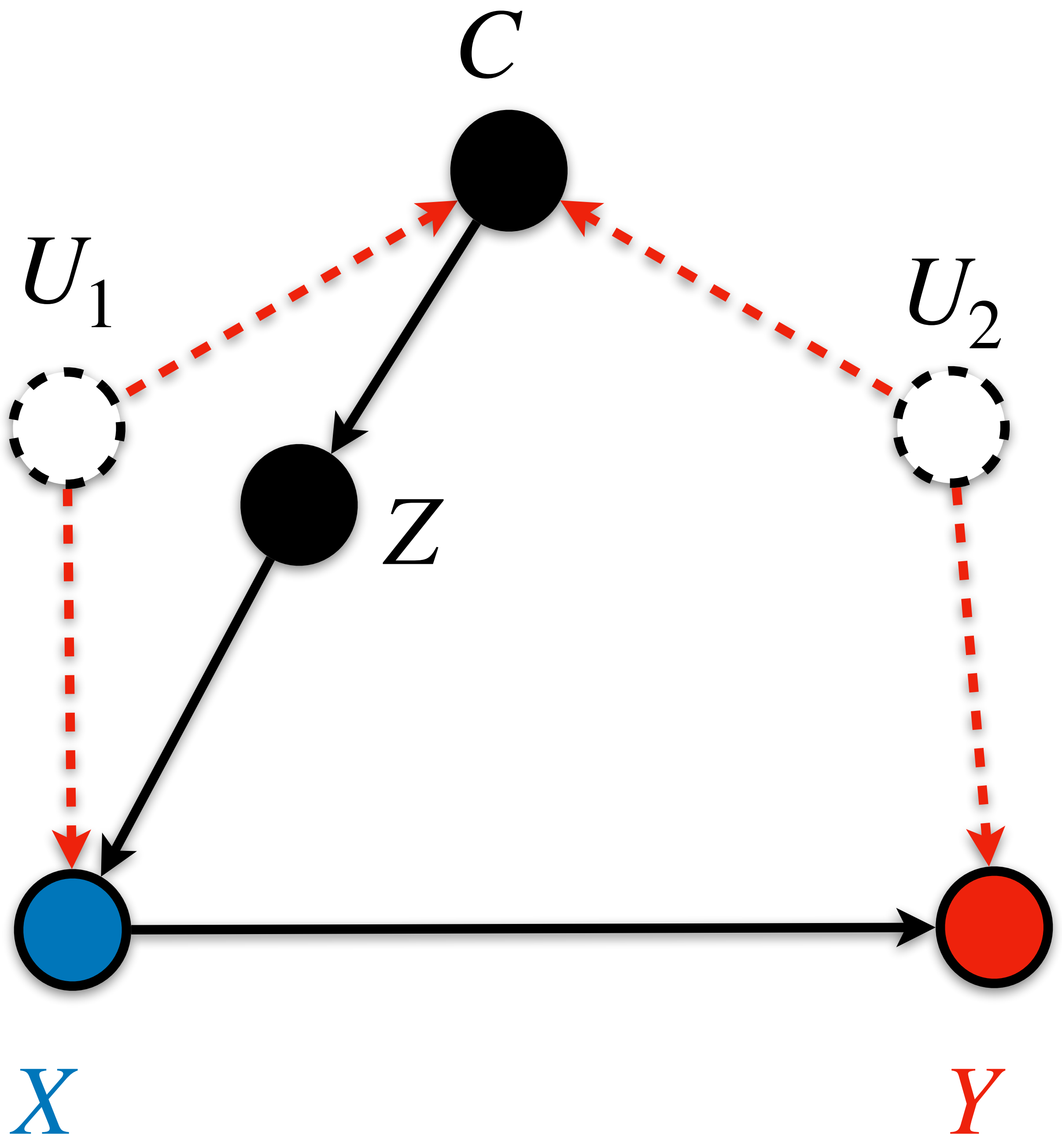
$$P(CZXY) \xrightarrow{\text{Factorization}} P_{\mathrm{do}(Z)}(CXY)$$

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

Background: Causal Effect Identification

Causal Effect Identification

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



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

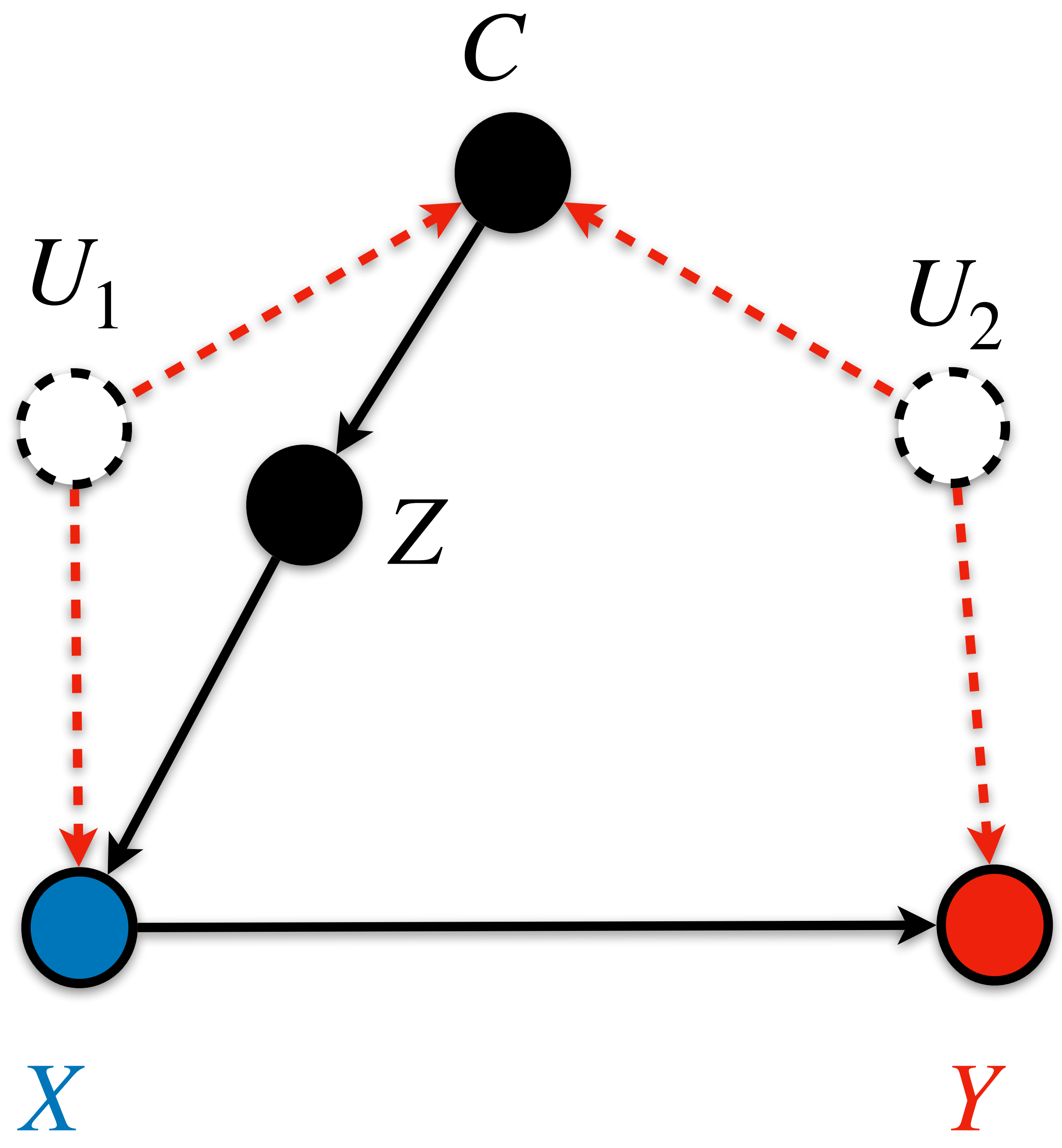
$P(C)P(XY \mid ZC)$

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

Background: Causal Effect Identification

Causal Effect 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,

- *BDs (or mSBDs) can be estimated sample-efficiently using robust estimators*
- The computation tree for the effect identification is composed of *interventional distributions as intermediate nodes.*

My Approach: 3-Step

To connect BD & Identification,

My Approach: 3-Step

To connect BD & Identification,

❶ **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,

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

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

❸ **Construct** robust estimators by using robust BD estimators

Complete Criterion for mSBD Adjustment

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

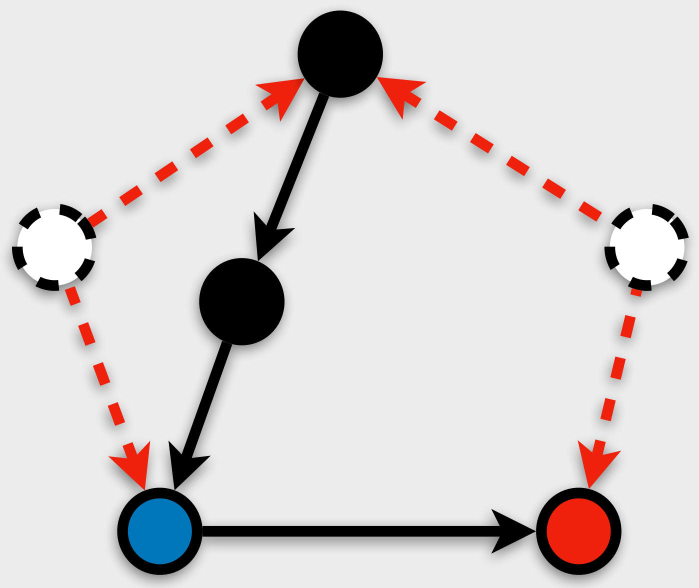
Complete Criterion for mSBD Adjustment

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

Complete Criterion for mSBD Adjustment

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

Graph

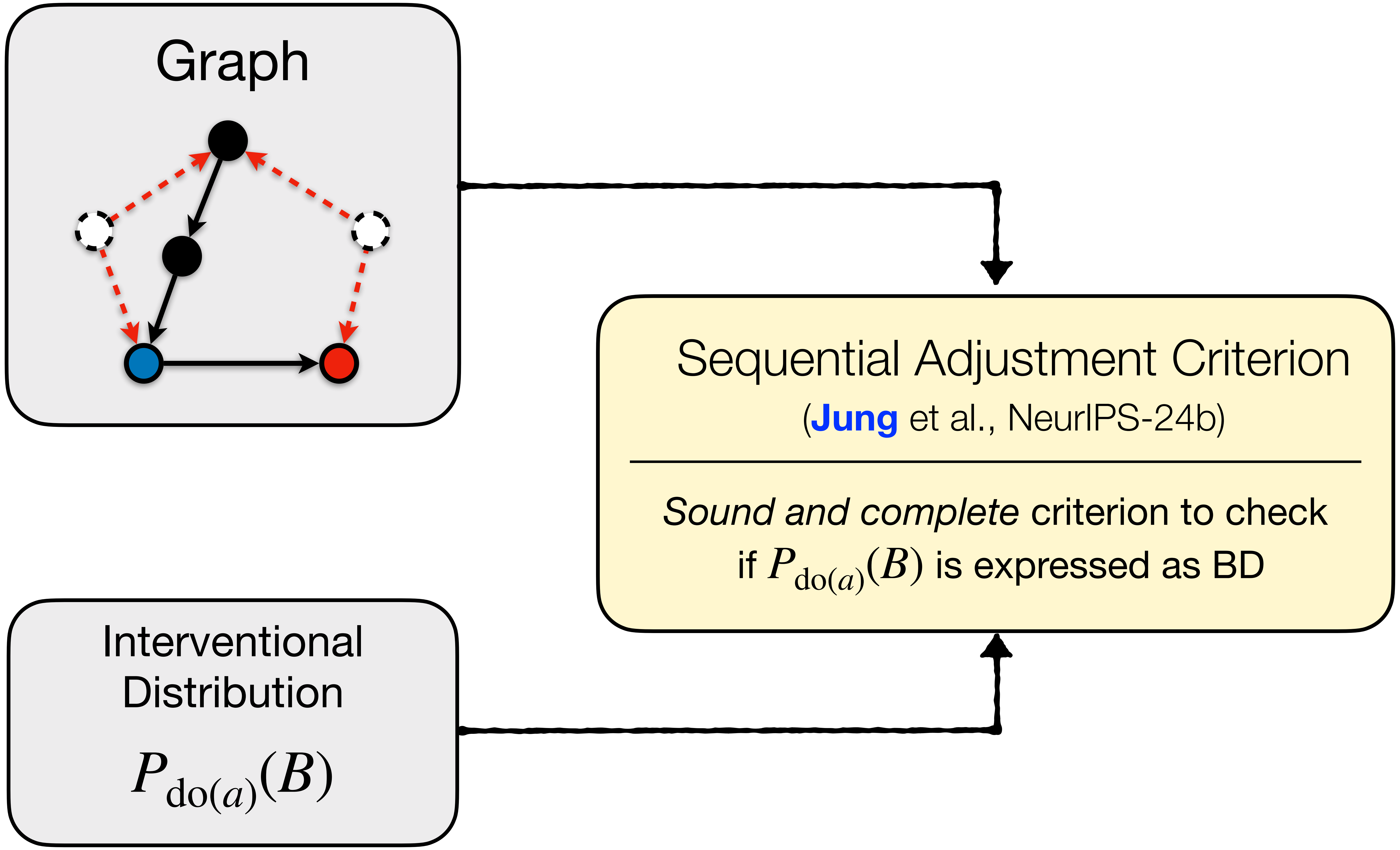


Interventional Distribution

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

Complete Criterion for mSBD Adjustment

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



Graph

Interventional Distribution

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

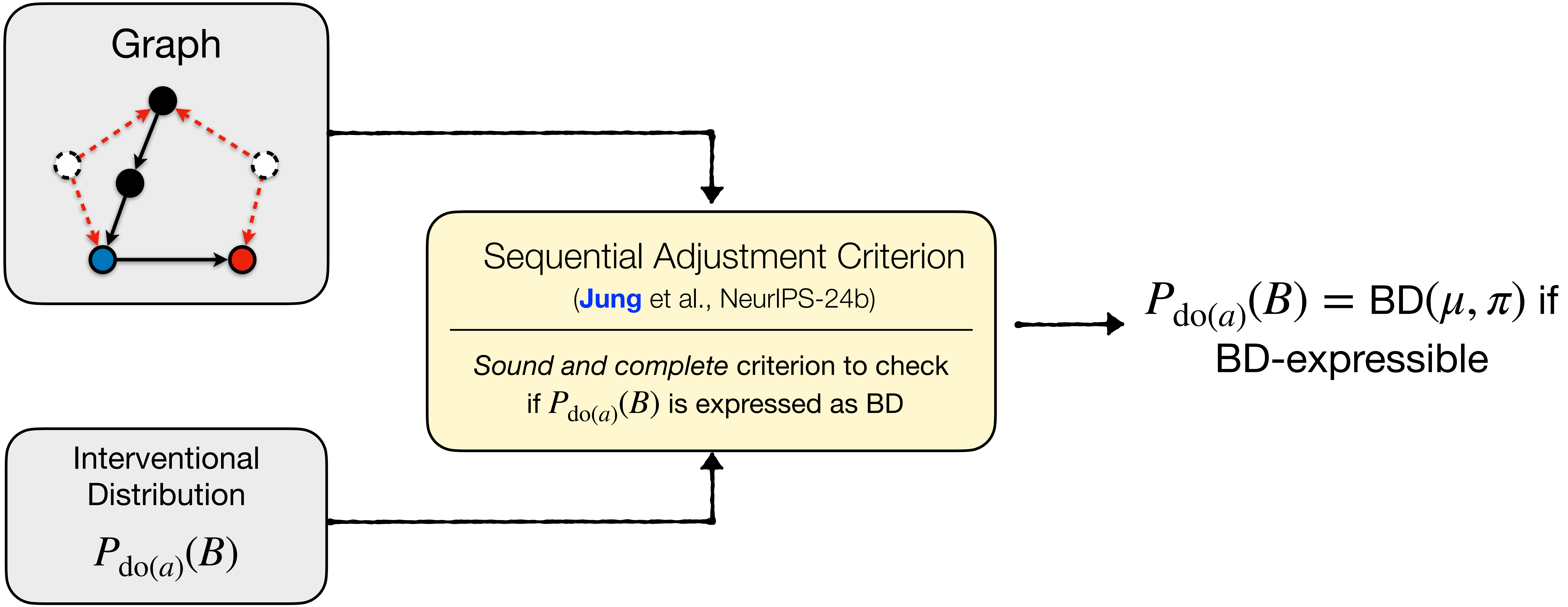
Sequential Adjustment Criterion

(**Jung** et al., NeurIPS-24b)

Sound and complete criterion to check if $P_{\mathrm{do}(a)}(B)$ is expressed as BD

Complete Criterion for mSBD Adjustment

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



Graph

Interventional Distribution

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

Sequential Adjustment Criterion

(**Jung** et al., NeurIPS-24b)

Sound and complete criterion to check if $P_{\text{do}(a)}(B)$ is expressed as BD

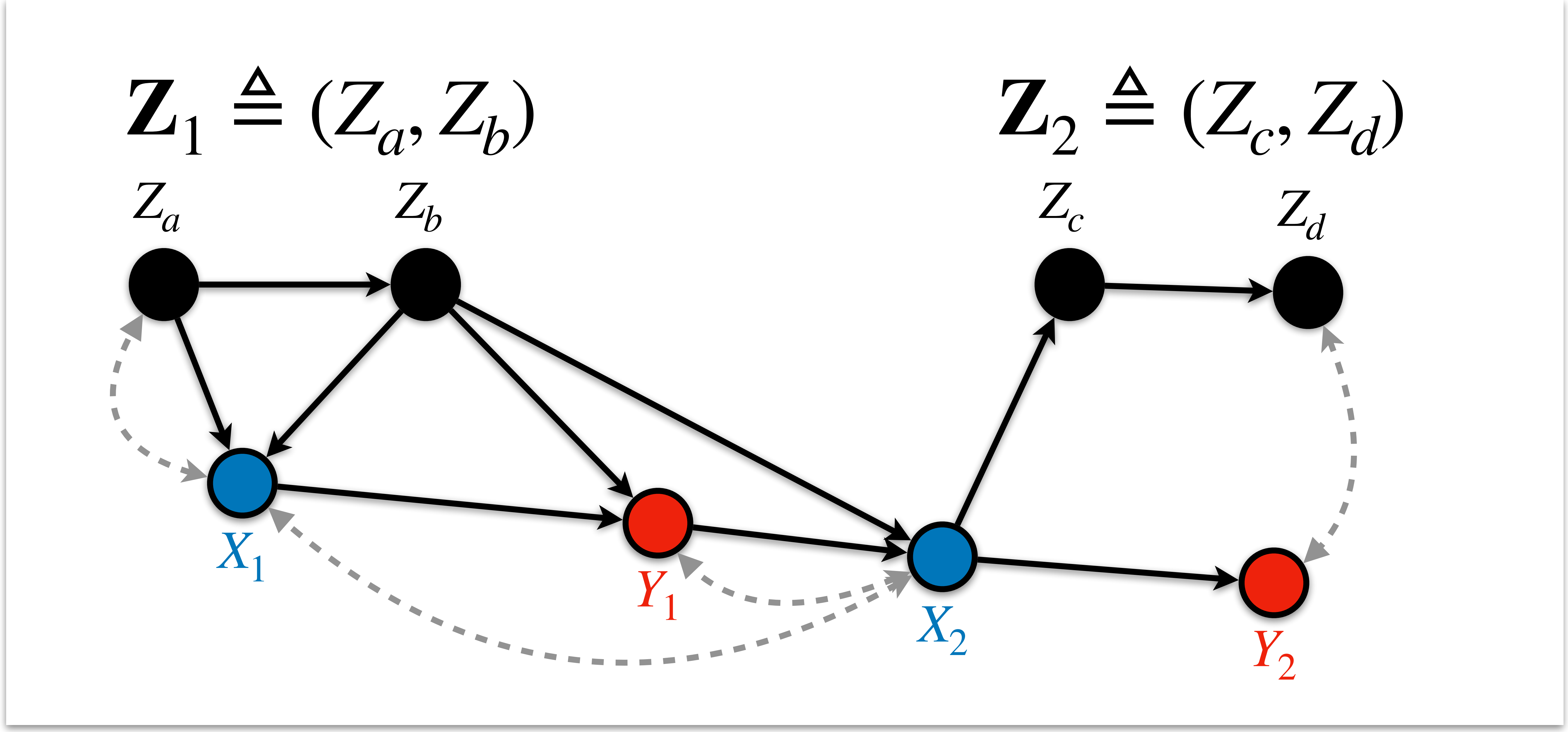
$P_{\text{do}(a)}(B) = \text{BD}(\mu, \pi)$ if BD-expressible

Motivation: Incompleteness of BD/mSBD

∃ examples s.t. $P(\mathbf{y} \mid \mathrm{do}(\mathbf{x}))$ is BD adjustment even if BD criterion fails.

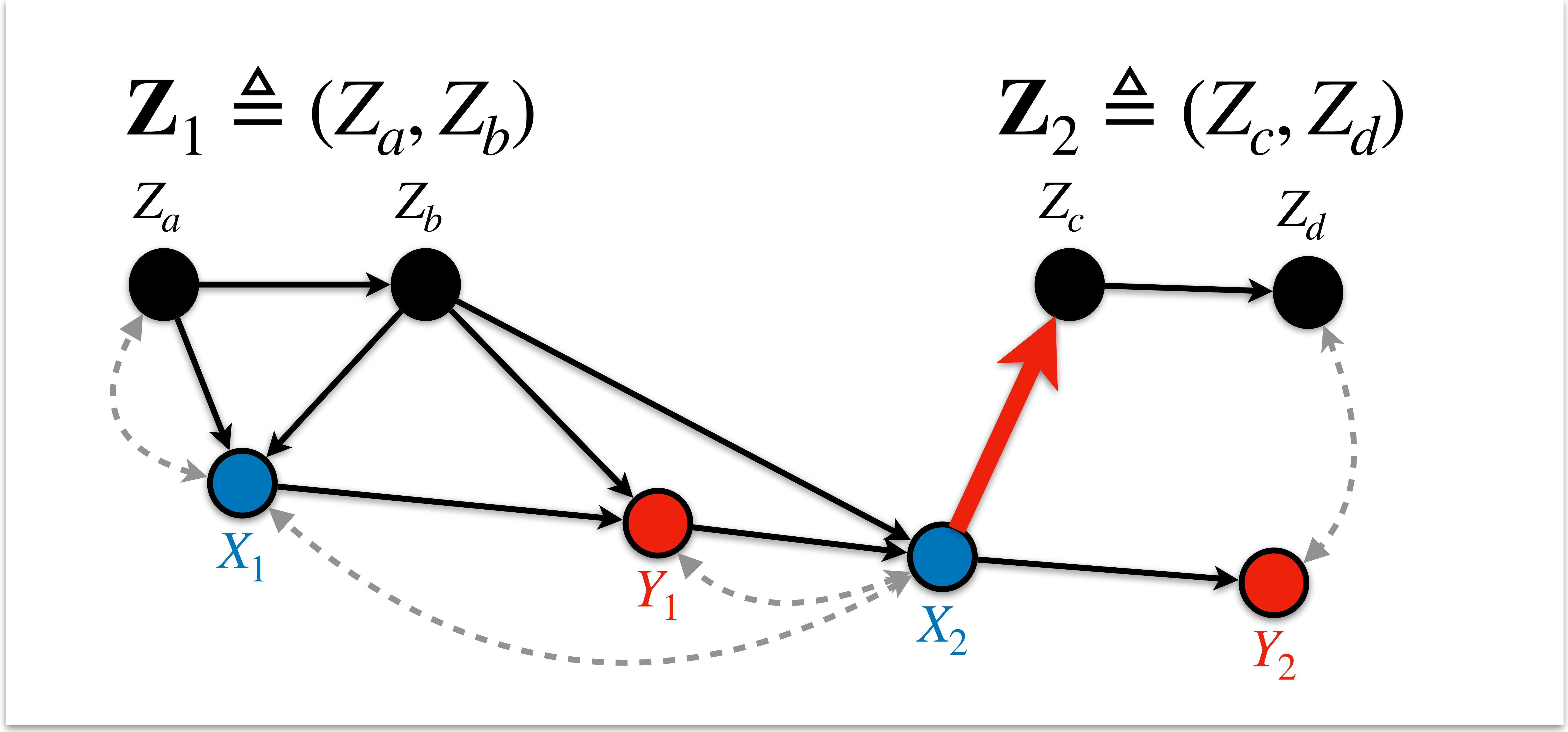
Motivation: Incompleteness of BD/mSBD

∃ examples s.t. $P(\mathbf{y} \mid \text{do}(\mathbf{x}))$ is BD adjustment even if BD criterion fails.



Motivation: Incompleteness of BD/mSBD

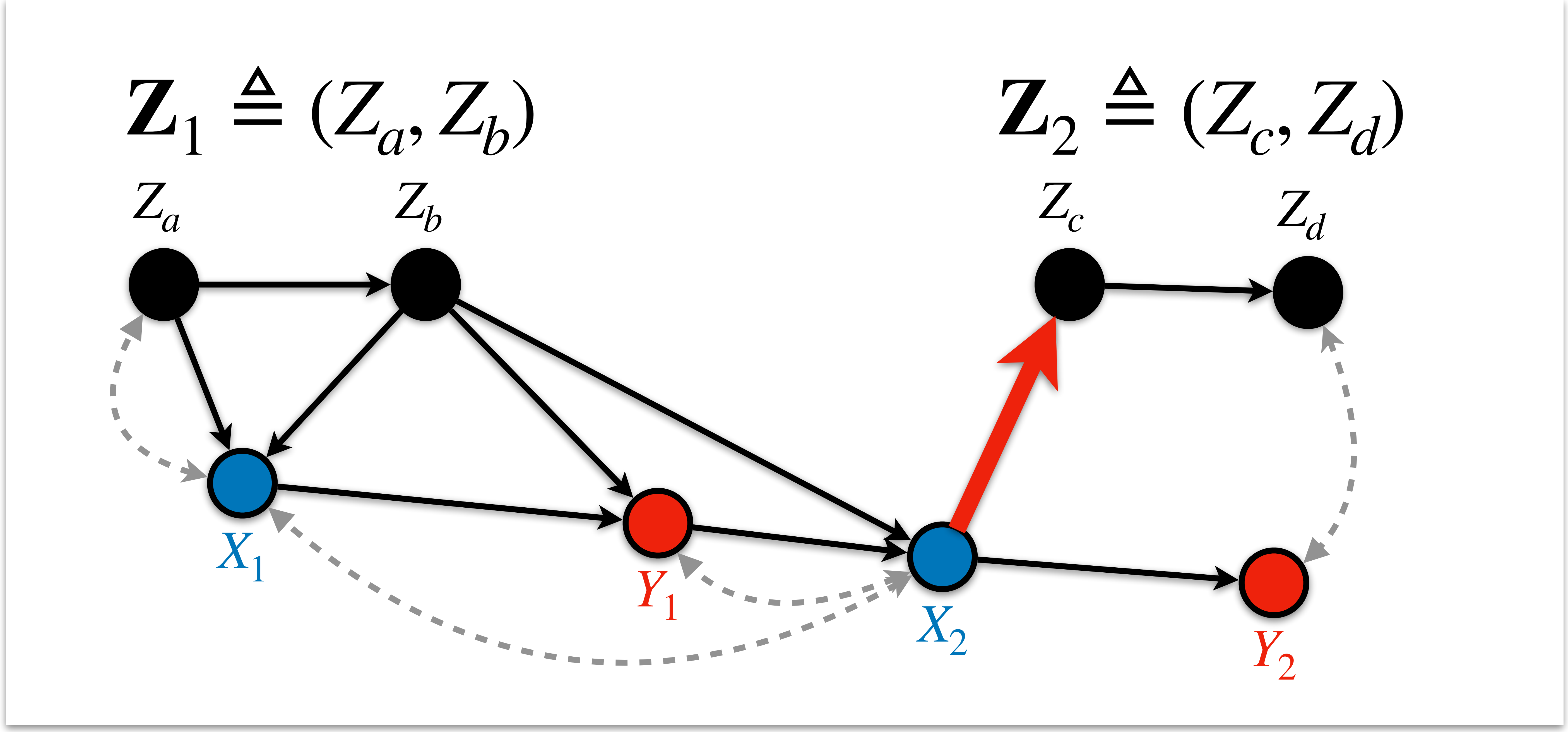
∃ examples s.t. $P(\mathbf{y} \mid do(\mathbf{x}))$ is BD adjustment even if BD criterion fails.



① $\mathbf{Z}$ doesn't satisfy the mSBD criterion

Motivation: Incompleteness of BD/mSBD

∃ examples s.t. $P(\mathbf{y} \mid \text{do}(\mathbf{x}))$ is BD adjustment even if BD criterion fails.



❶ $\mathbf{Z}$ doesn't satisfy the mSBD criterion

❷ $$P(y_1y_2 \mid \text{do}(x_1x_2)) = \sum_{\mathbf{z}_1\mathbf{z}_2} \overbrace{P(y_2 \mid \text{prev}_1, \mathbf{z}_2x_2)P(y_1\mathbf{z}_2 \mid \mathbf{z}_1x_1)P(\mathbf{z}_1)}^{\text{“mSBD adjustment”}}$$

Complete Seq. Adjustment Criterion

Complete Seq. Adjustment Criterion

Sequential Adjustment Criterion (SAC)

A seq. $\mathbf{Z} = (\mathbf{Z}_1, \cdots, \mathbf{Z}_m)$ satisfies the SAC if, for $i = 1, \cdots, m$, $\mathbf{Z}_i \cup \mathbf{prev}_{i-1}$ blocks confounding path between $(\mathbf{X}_i, \mathbf{Y}^{\geq i})$

Complete Seq. Adjustment Criterion

Sequential Adjustment Criterion (SAC)

A seq. $\mathbf{Z} = (\mathbf{Z}_1, \cdots, \mathbf{Z}_m)$ satisfies the SAC if, for $i = 1, \cdots, m$, $\mathbf{Z}_i \cup \mathbf{prev}_{i-1}$ blocks confounding path between $(\mathbf{X}_i, \mathbf{Y}^{\geq i})$

$\Leftrightarrow$

Completeness

$P(\mathbf{y} \mid \mathrm{do}(\mathbf{x}))$ is given as mSBD.

Complete Seq. Adjustment Criterion

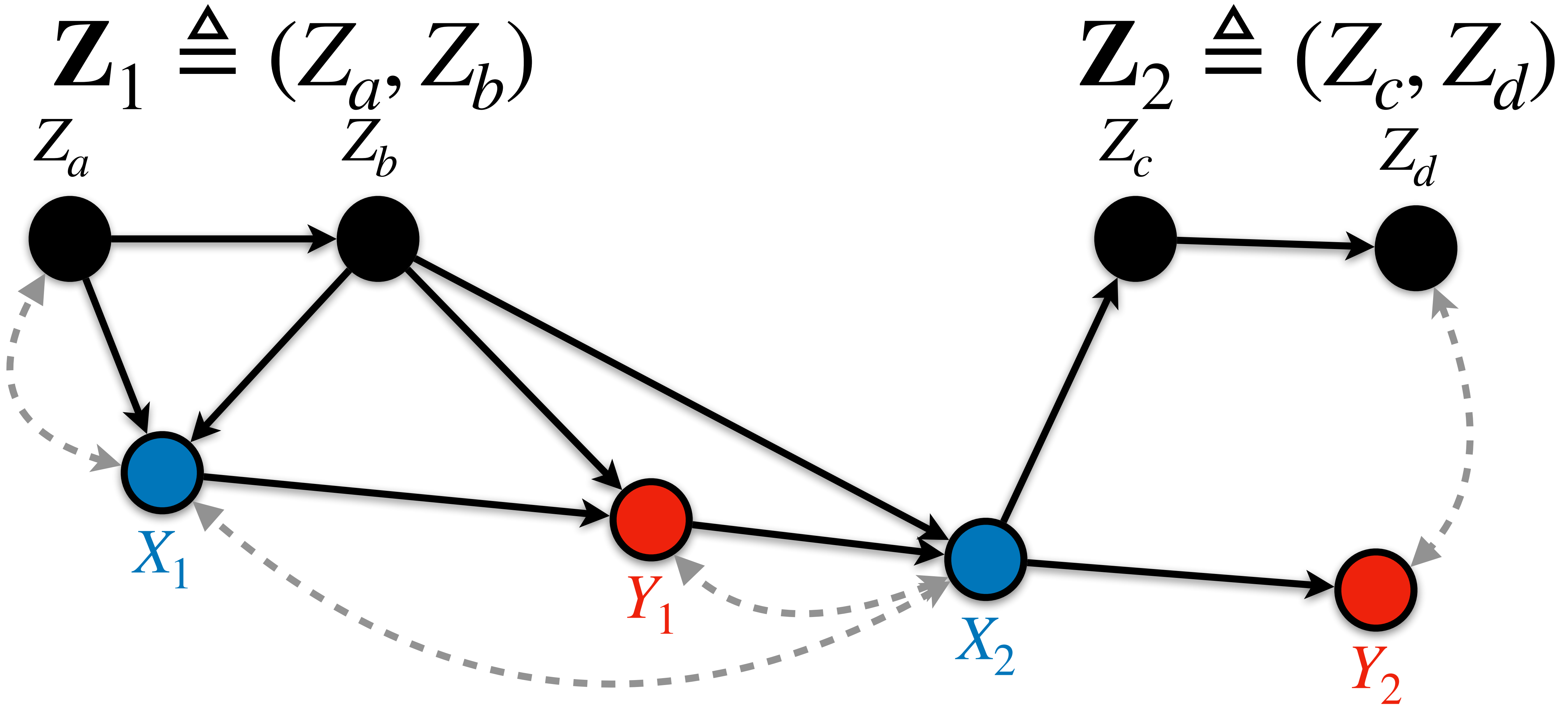
Sequential Adjustment Criterion (SAC)

A seq. $\mathbf{Z} = (\mathbf{Z}_1, \cdots, \mathbf{Z}_m)$ satisfies the SAC if, for $i = 1, \cdots, m$, $\mathbf{Z}_i \cup \mathbf{prev}_{i-1}$ blocks confounding path between $(\mathbf{X}_i, \mathbf{Y}^{\geq i})$

$\Leftrightarrow$

Completeness

$P(\mathbf{y} \mid do(\mathbf{x}))$ is given as mSBD.



✗ mSBD fails

Complete Seq. Adjustment Criterion

Sequential Adjustment Criterion (SAC)

A seq. $\mathbf{Z} = (\mathbf{Z}_1, \cdots, \mathbf{Z}_m)$ satisfies the SAC if, for $i = 1, \cdots, m$, $\mathbf{Z}_i \cup \mathbf{prev}_{i-1}$ blocks confounding path between $(\mathbf{X}_i, \mathbf{Y}^{\geq i})$

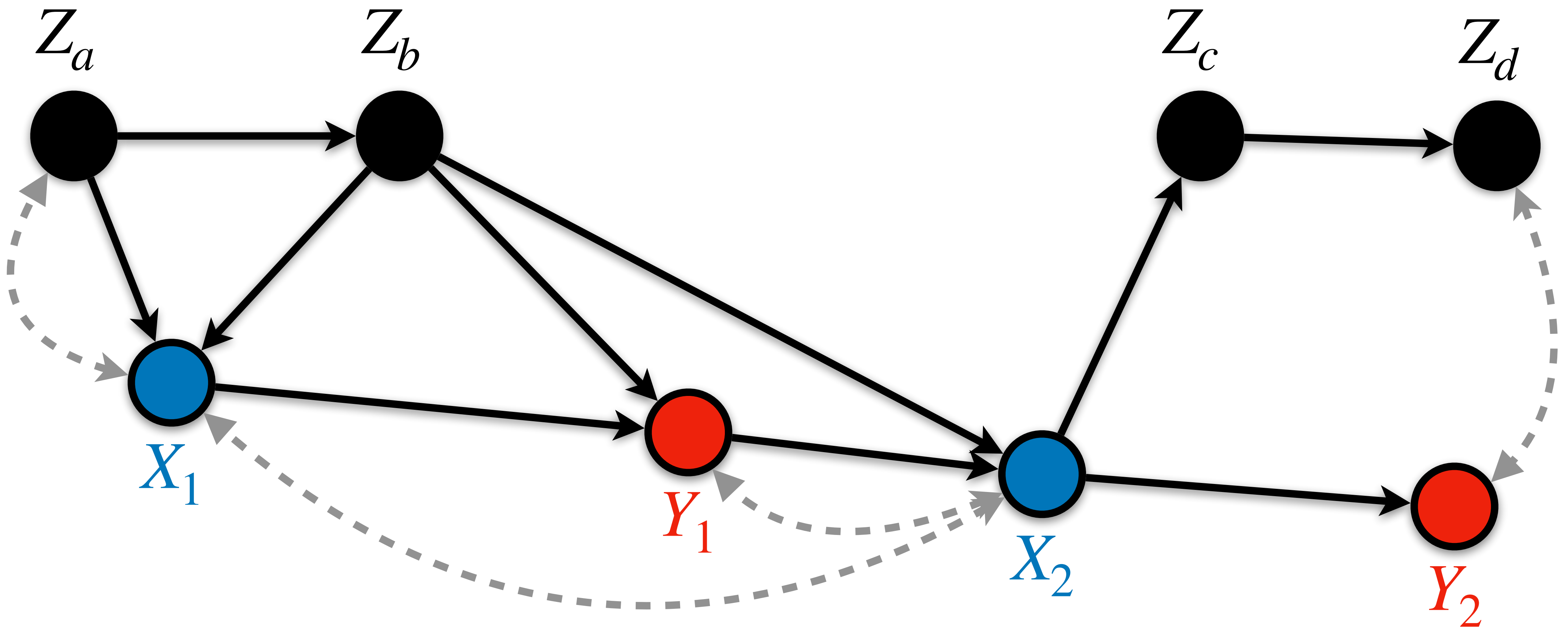
$\Leftrightarrow$

Completeness

$P(\mathbf{y} \mid \mathrm{do}(\mathbf{x}))$ is given as mSBD.

$\mathbf{Z}_1 \triangleq (Z_a, Z_b)$

$\mathbf{Z}_2 \triangleq (Z_c, Z_d)$



✗ mSBD fails

✓ SAC holds

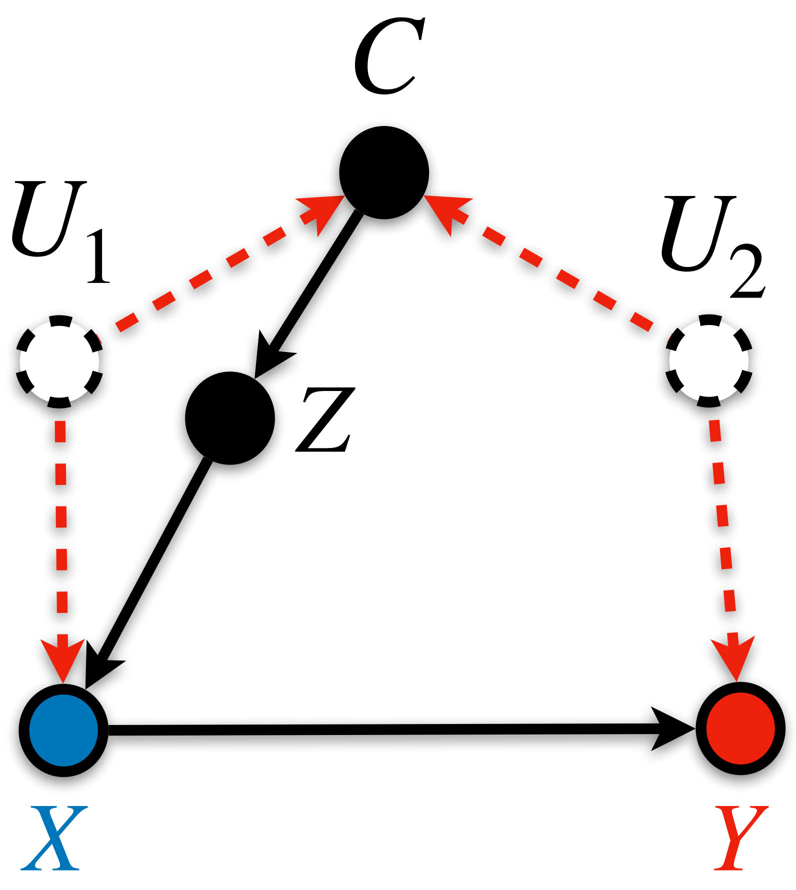
Estimating Causal Effects in 3-Steps

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



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

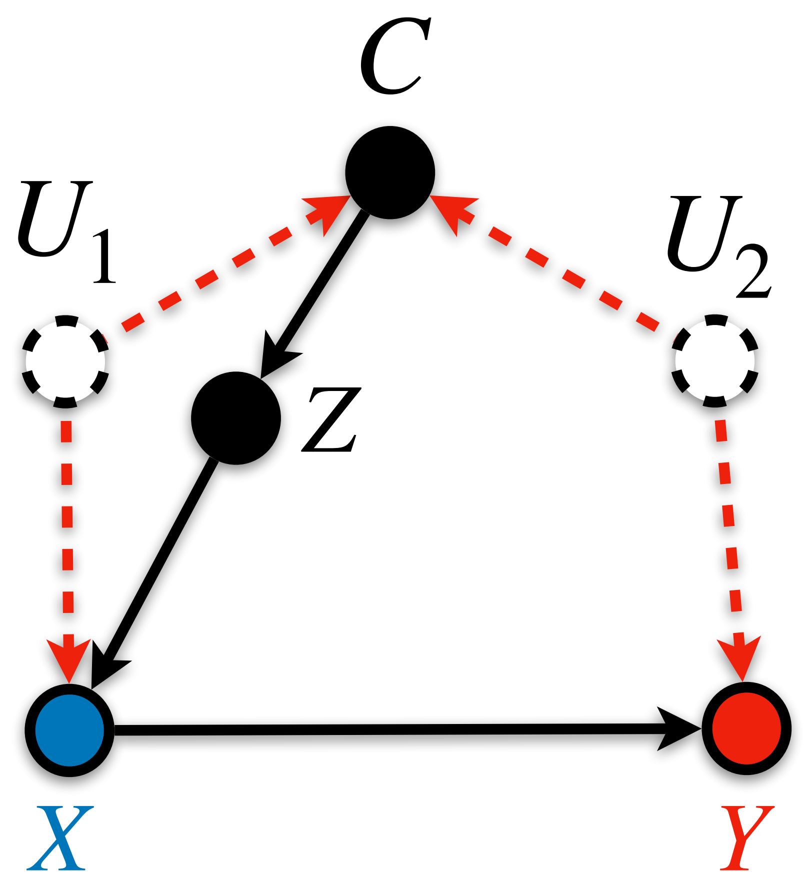
$$\sum_c P(c)P(XY \mid Zc)$$

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

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

Estimating Causal Effects in 3-Steps

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



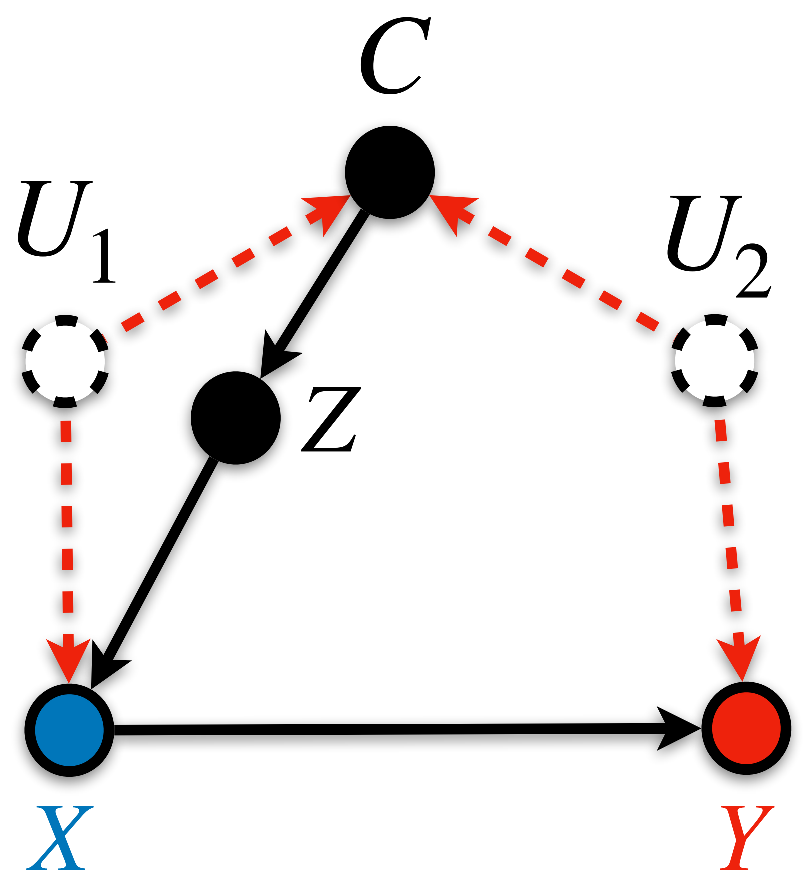
$$P(CZXY) \xrightarrow{\text{Factorization}} P_{\text{do}(Z)}(CXY) \xrightarrow{\text{Marginalization}} P_{\text{do}(Z)}(XY) \xrightarrow{\text{Factorization}} P(Y \mid \text{do}(X))$$

🔍

SAC

Estimating Causal Effects in 3-Steps

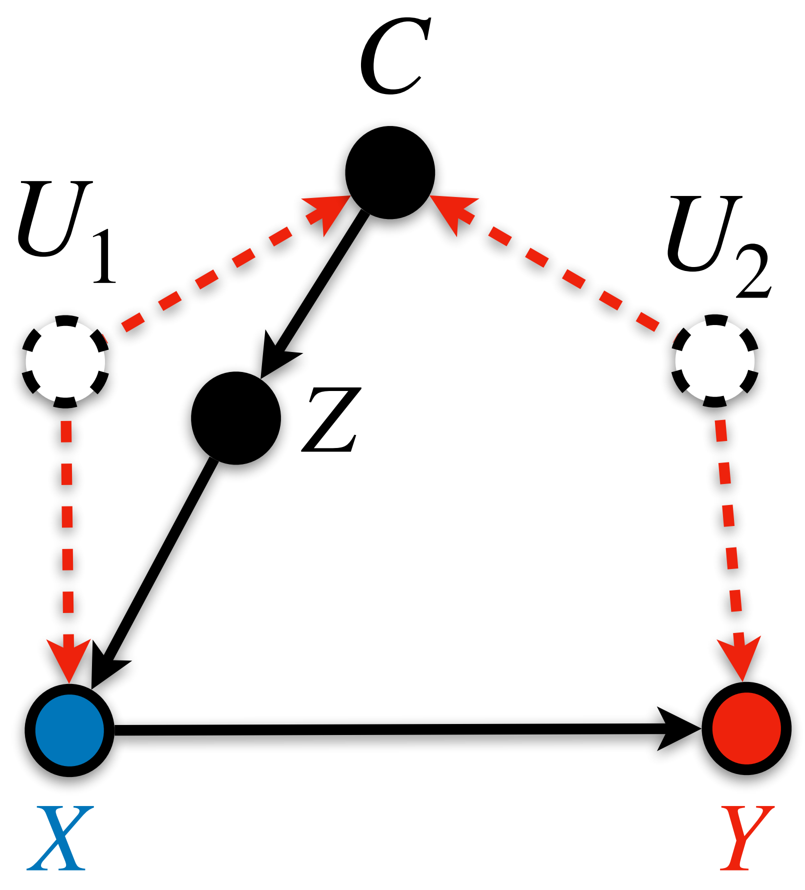
② **Express** causal effects as a function of BD



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

Estimating Causal Effects in 3-Steps

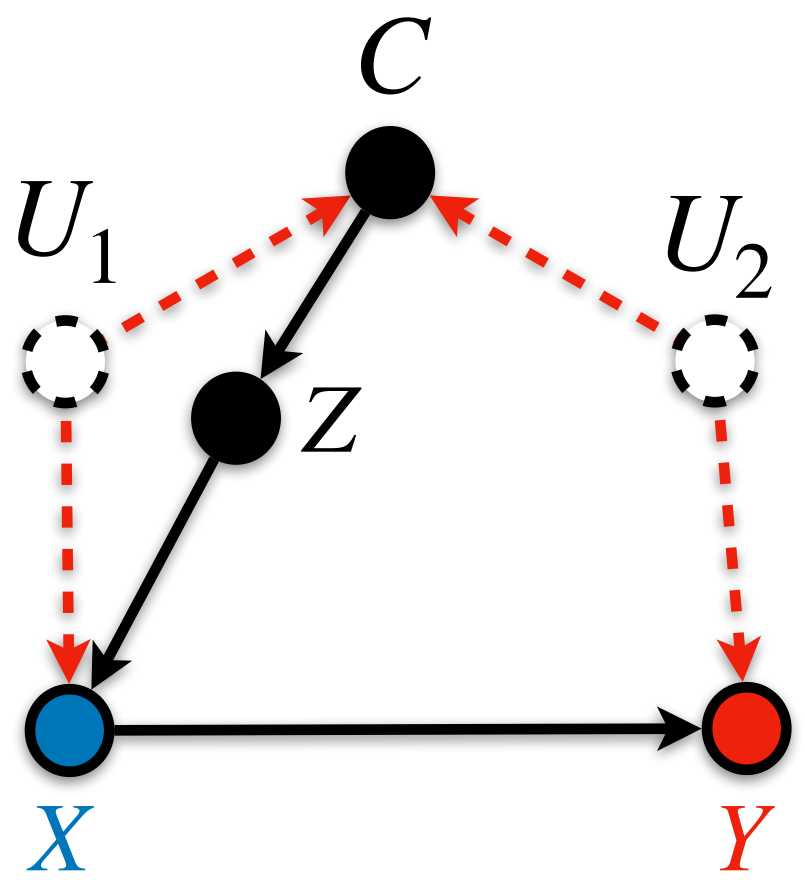
② **Express** causal effects as a function of BD



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

Estimating Causal Effects in 3-Steps

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



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

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

Estimating Causal Effects in 3-Steps

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

C

Theorem (Jung et al., 2021, AAAI)

The followings are equivalent:

1. $P(\mathbf{y} \mid \mathrm{do}(\mathbf{x}))$ is identifiable from $(\mathcal{G}, P)$
2. $P(\mathbf{y} \mid \mathrm{do}(\mathbf{x}))$ is expressible as a ***function of BDs*** through AdmissibleID

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

$\mathrm{do}(X))$

DML-ID: Estimator for Identifiable Causal Effects

DML-ID: Estimator for Identifiable Causal Effects

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

DML-ID: Estimator for Identifiable Causal Effects

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

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

DML-ID: Estimator for Identifiable Causal Effects

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

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

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

"DML-ID"

DML-ID: Estimator for Identifiable Causal Effects

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

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

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

"DML-ID"

DML-ID: Estimator for Identifiable Causal Effects

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

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

DML-BD DML-BD ⋯ DML-BD

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

"DML-ID"

Robustness of DML-ID

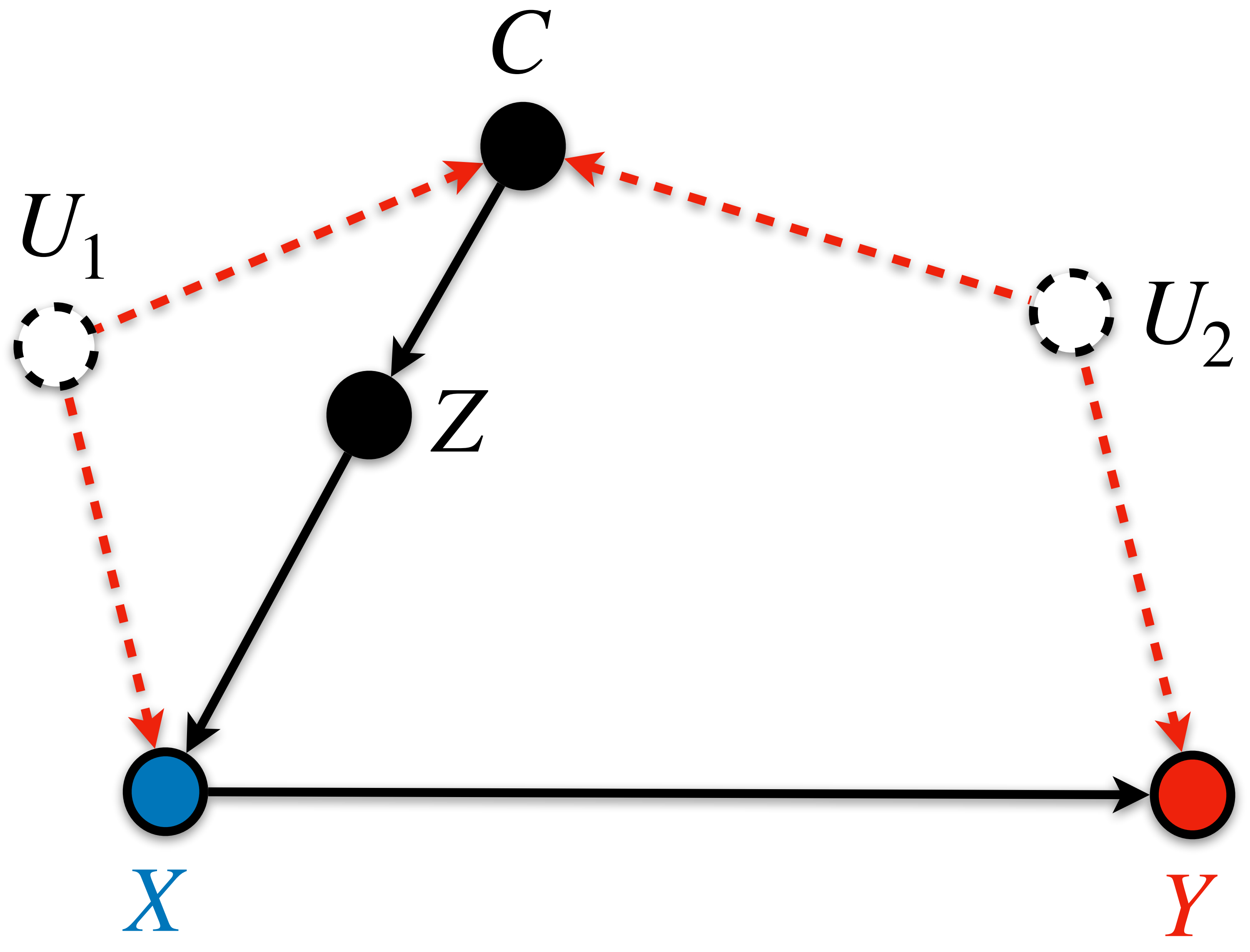
Theorem

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

- **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

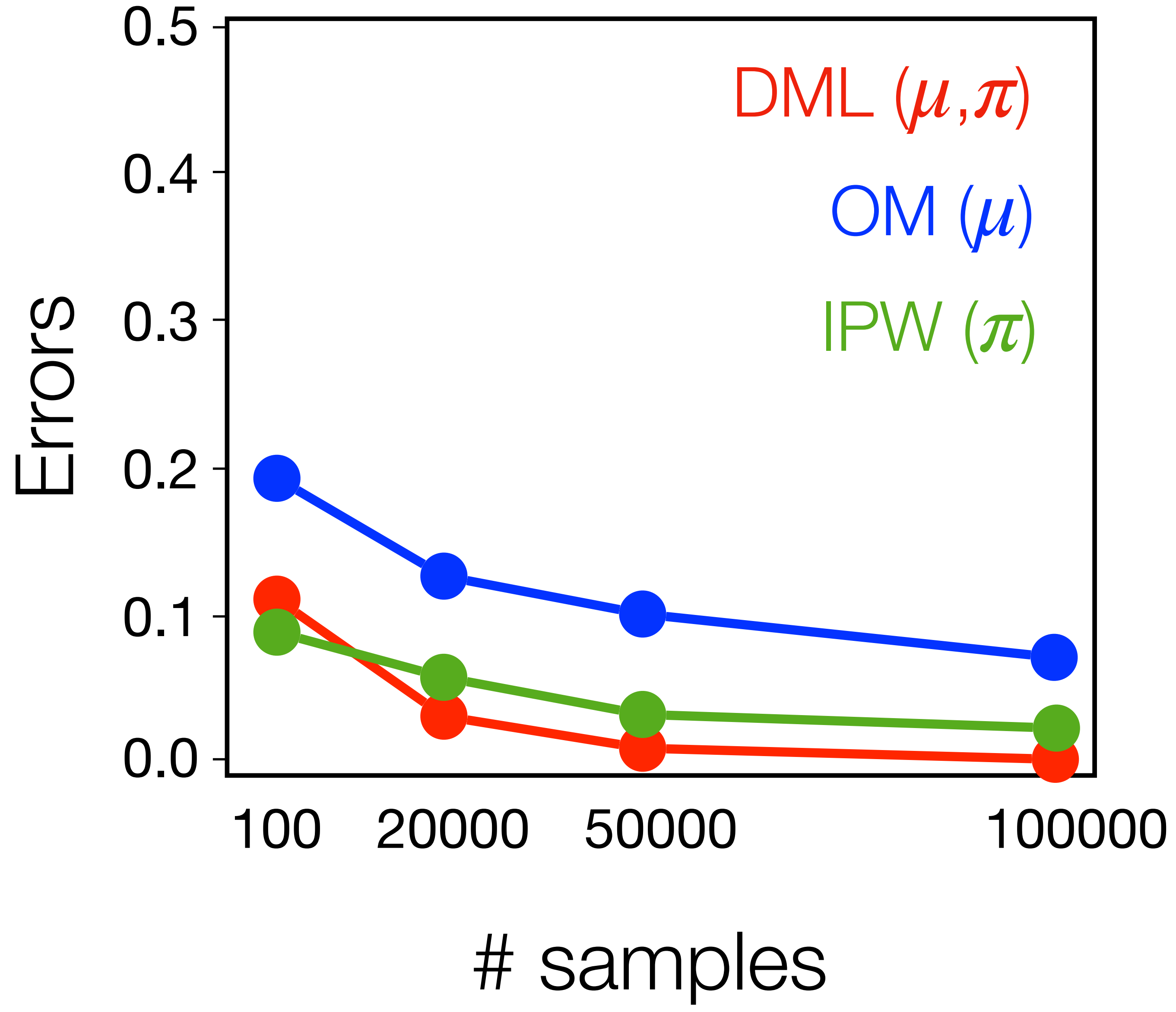
DML-ID - Simulation



DML-ID - Simulation

Fast Convergence

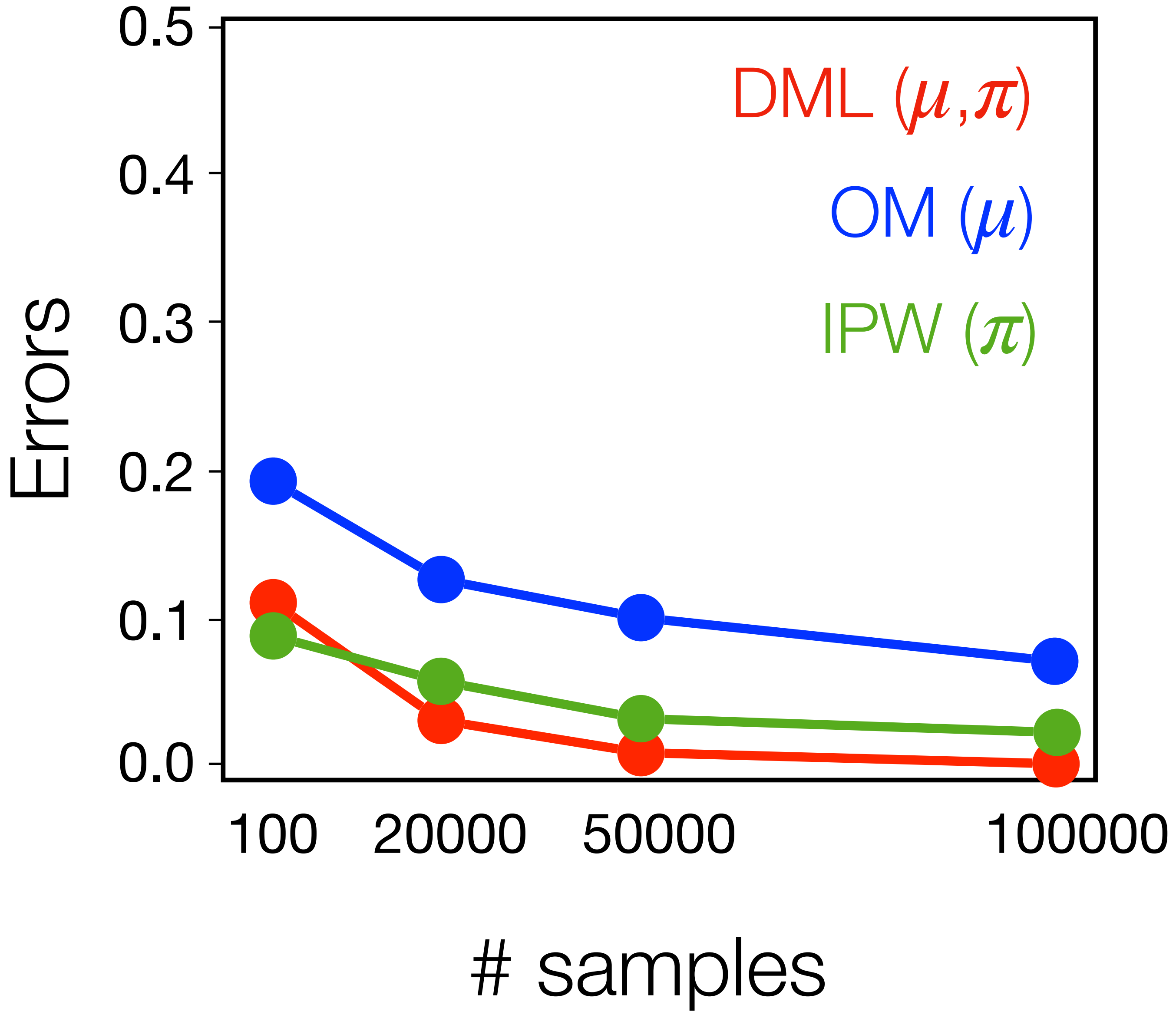
$(\hat{\boldsymbol{\mu}}, \hat{\boldsymbol{\pi}}) \rightarrow (\boldsymbol{\mu}_0, \boldsymbol{\pi}_0)$ slowly



DML-ID - Simulation

Fast Convergence

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

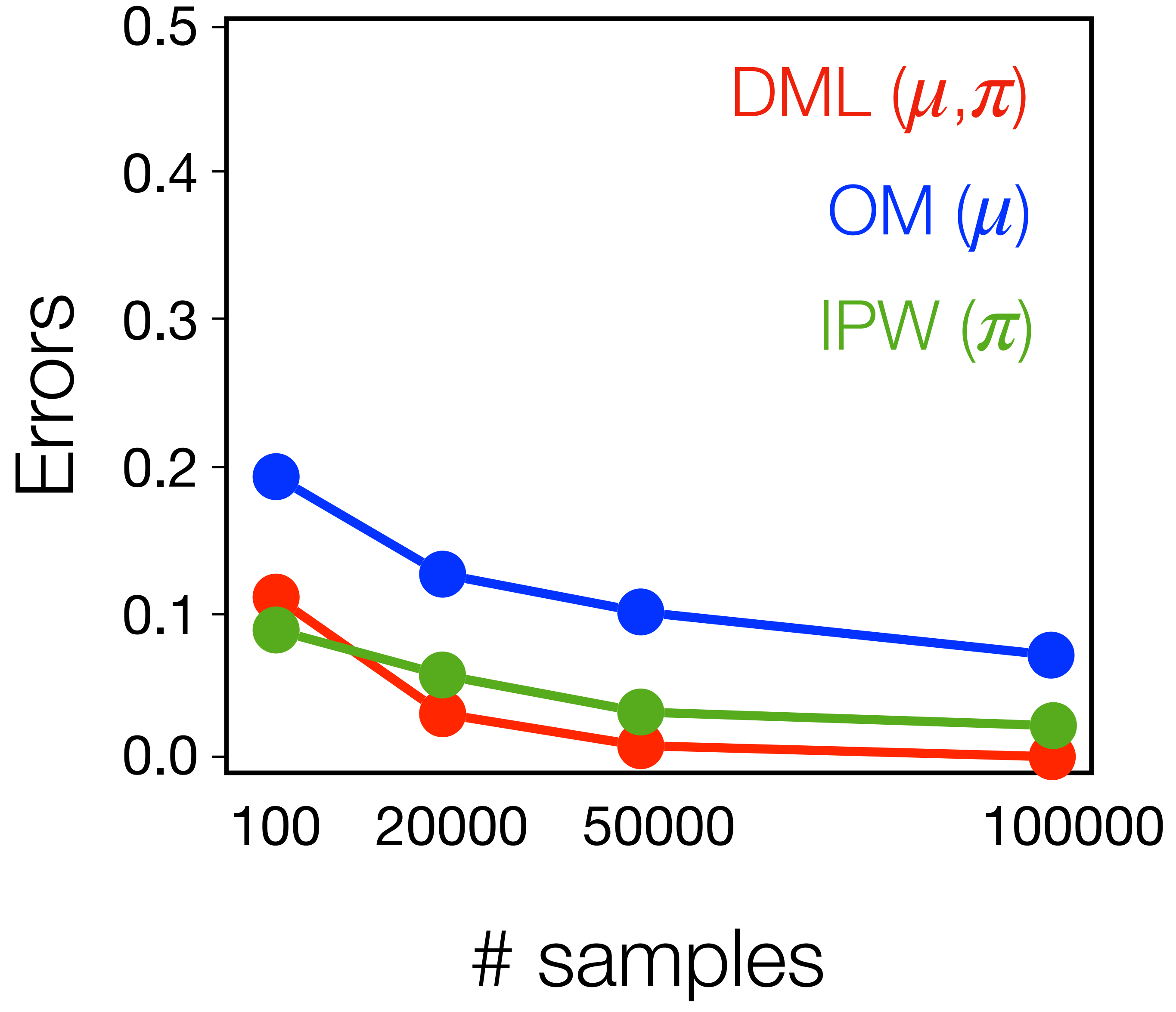


DML-ID converges fast, even when $(\hat{\boldsymbol{\mu}},\hat{\boldsymbol{\pi}})$ converge slowly

DML-ID - Simulation

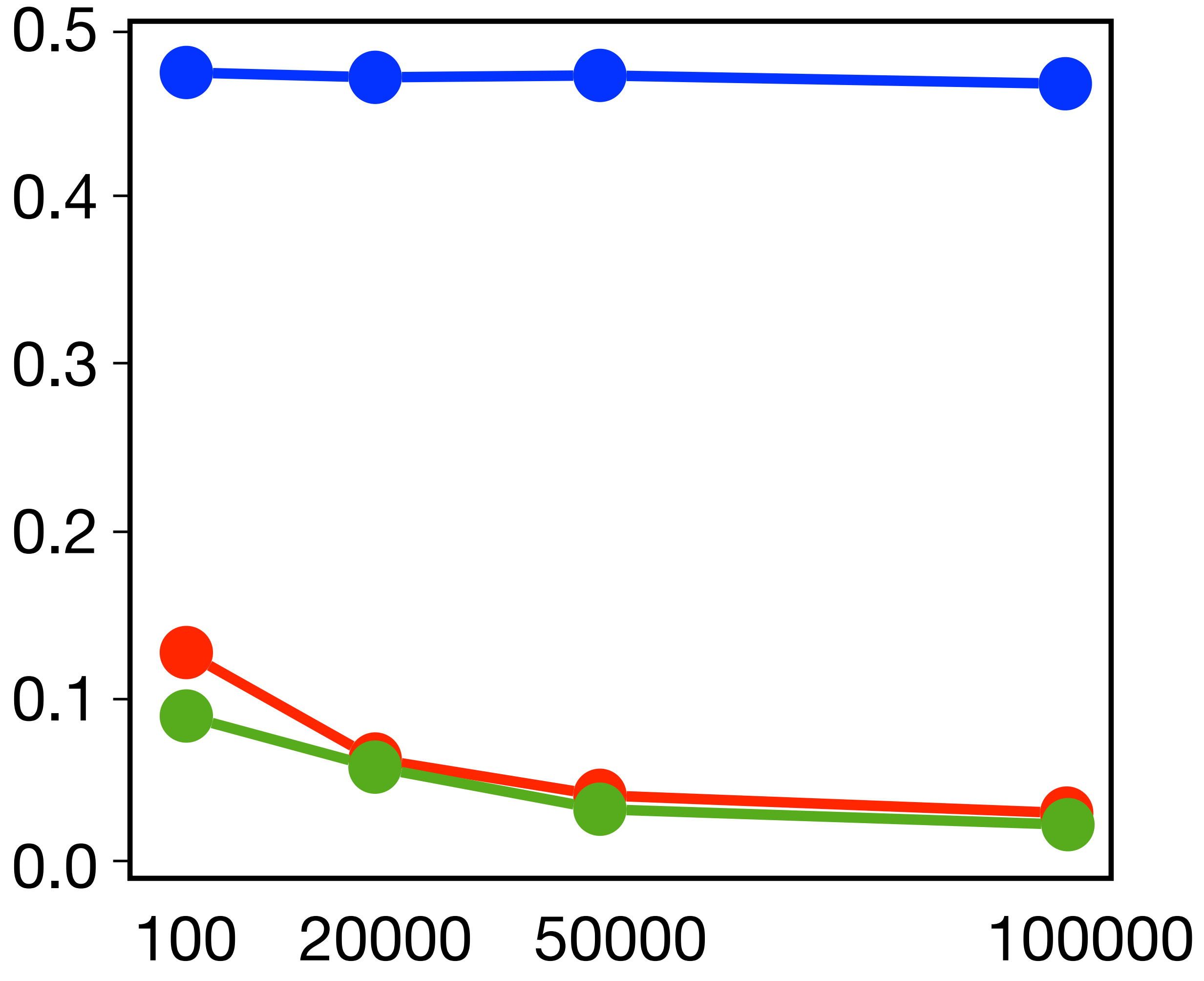
Fast Convergence

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

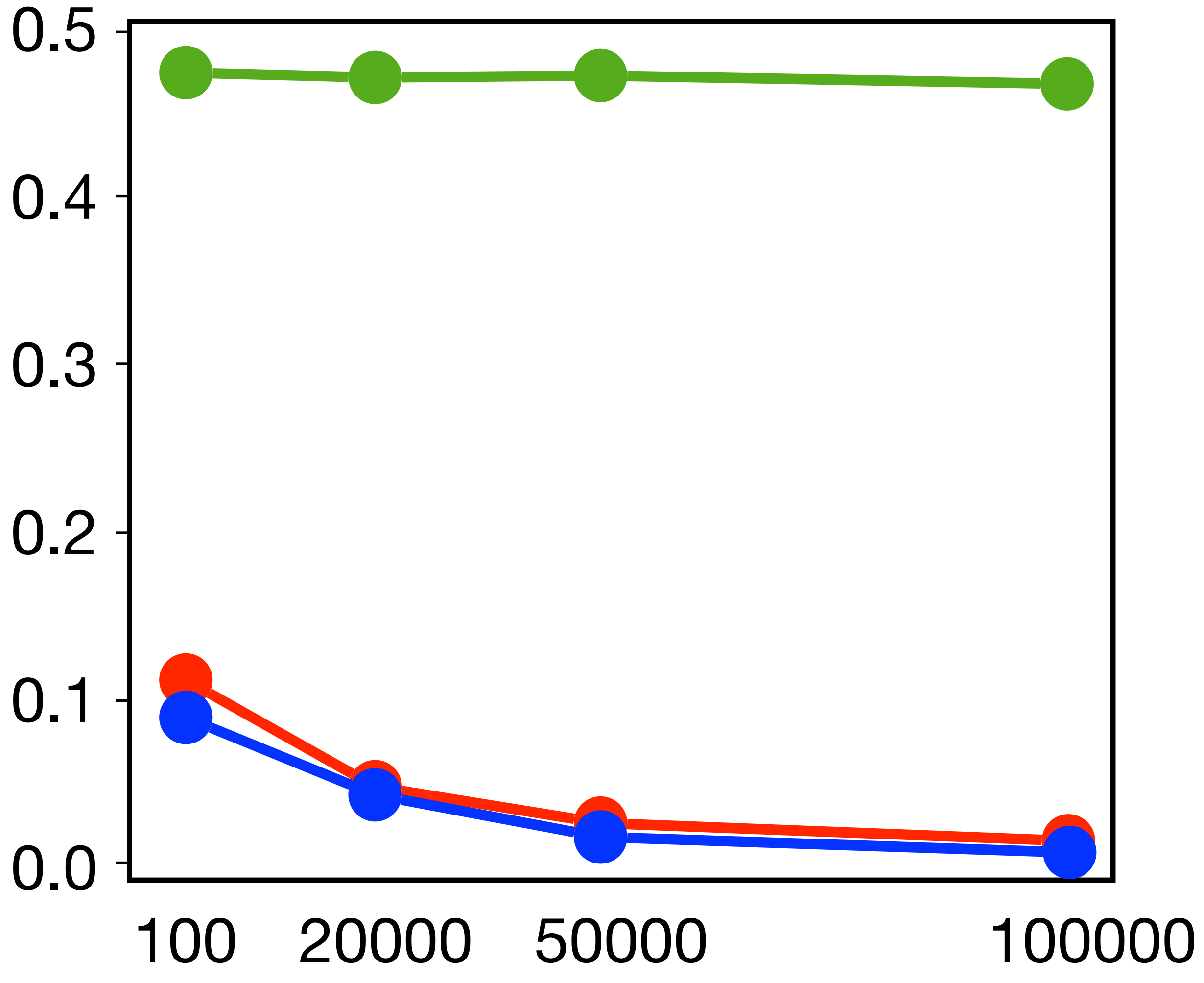


Double Robustness

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



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

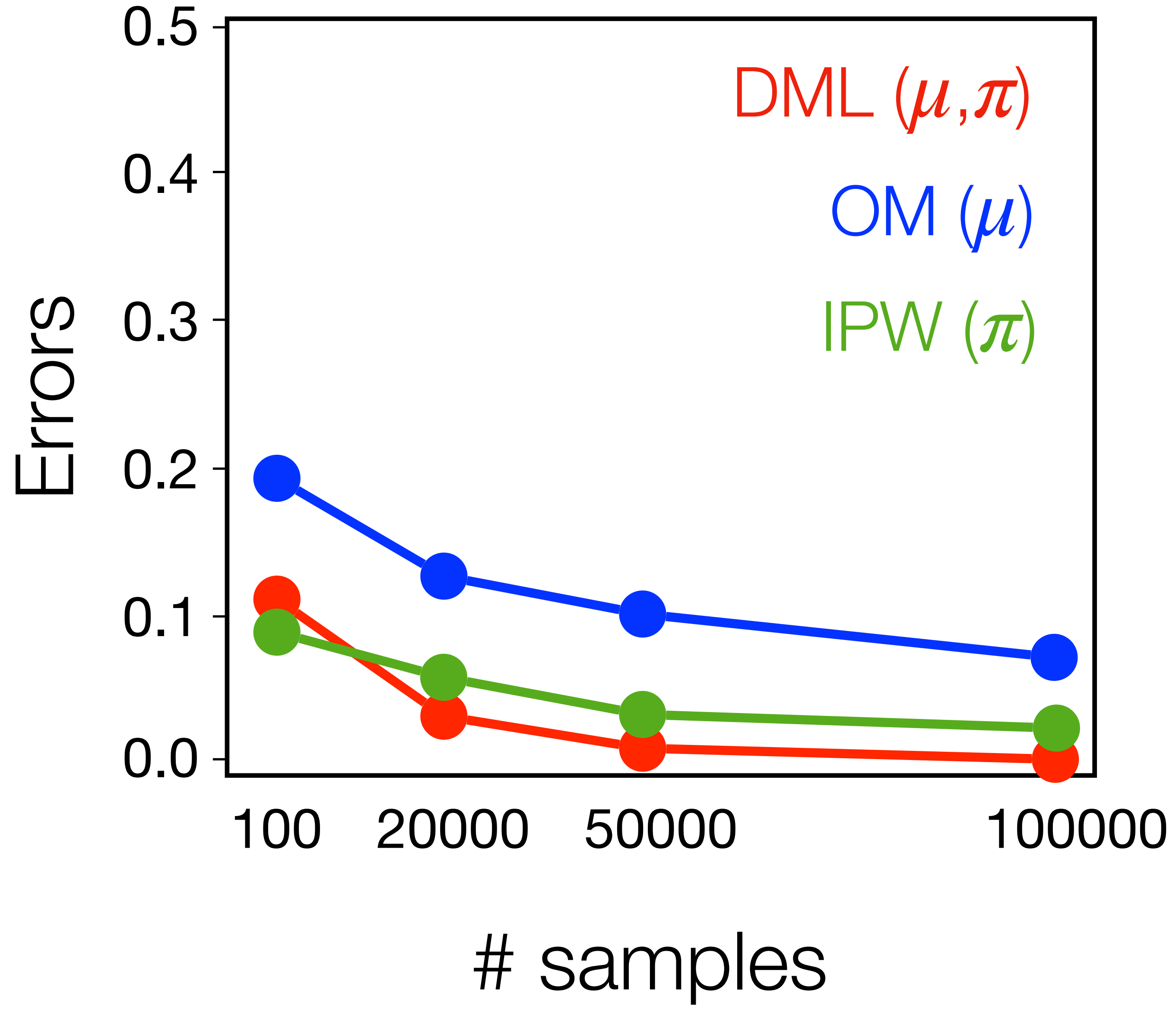


DML-ID converges fast, even when $(\hat{\boldsymbol{\mu}},\hat{\boldsymbol{\pi}})$ converge slowly

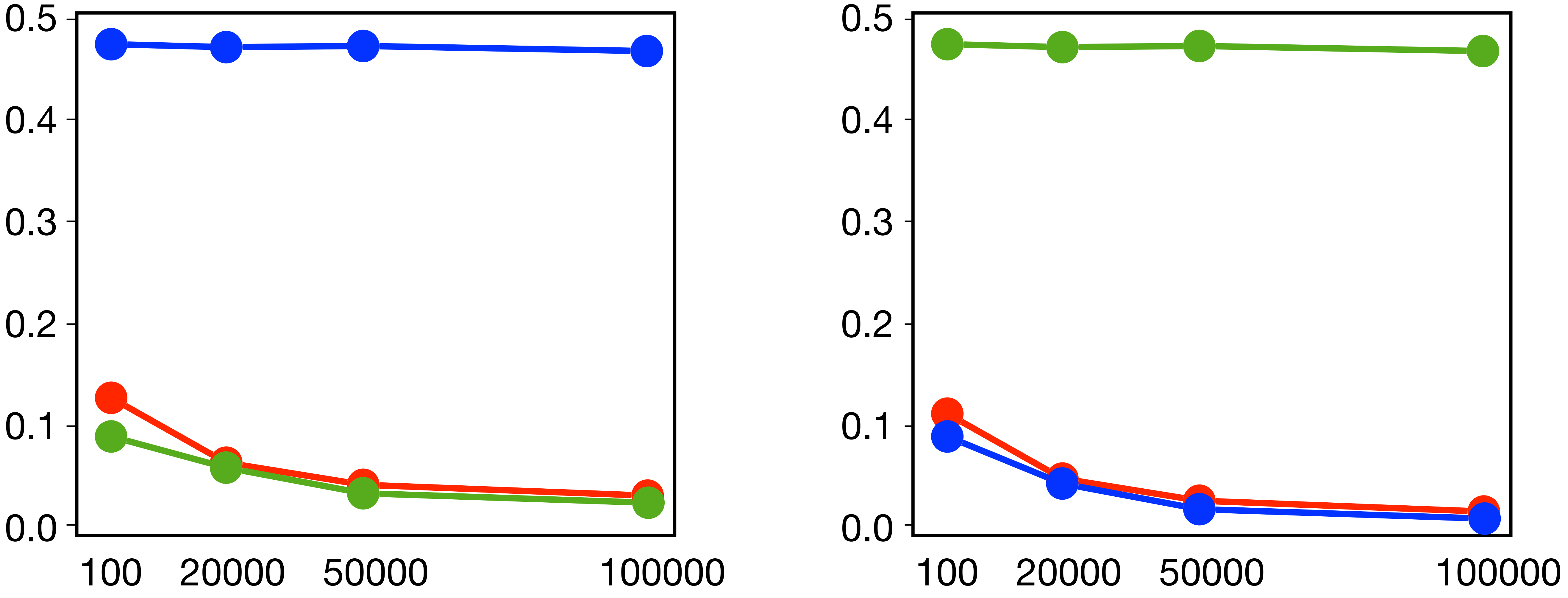
DML-ID - Simulation

Fast Convergence

$(\hat{\boldsymbol{\mu}},\hat{\boldsymbol{\pi}}) \rightarrow (\boldsymbol{\mu}_0, \boldsymbol{\pi}_0)$ 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 fast, even when $(\hat{\boldsymbol{\mu}},\hat{\boldsymbol{\pi}})$ converge slowly

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

Talk Outline

➡ ❶ Estimating causal effects from observations

❷ Estimating causal effects from data fusion

❸ Unified and scalable estimation method

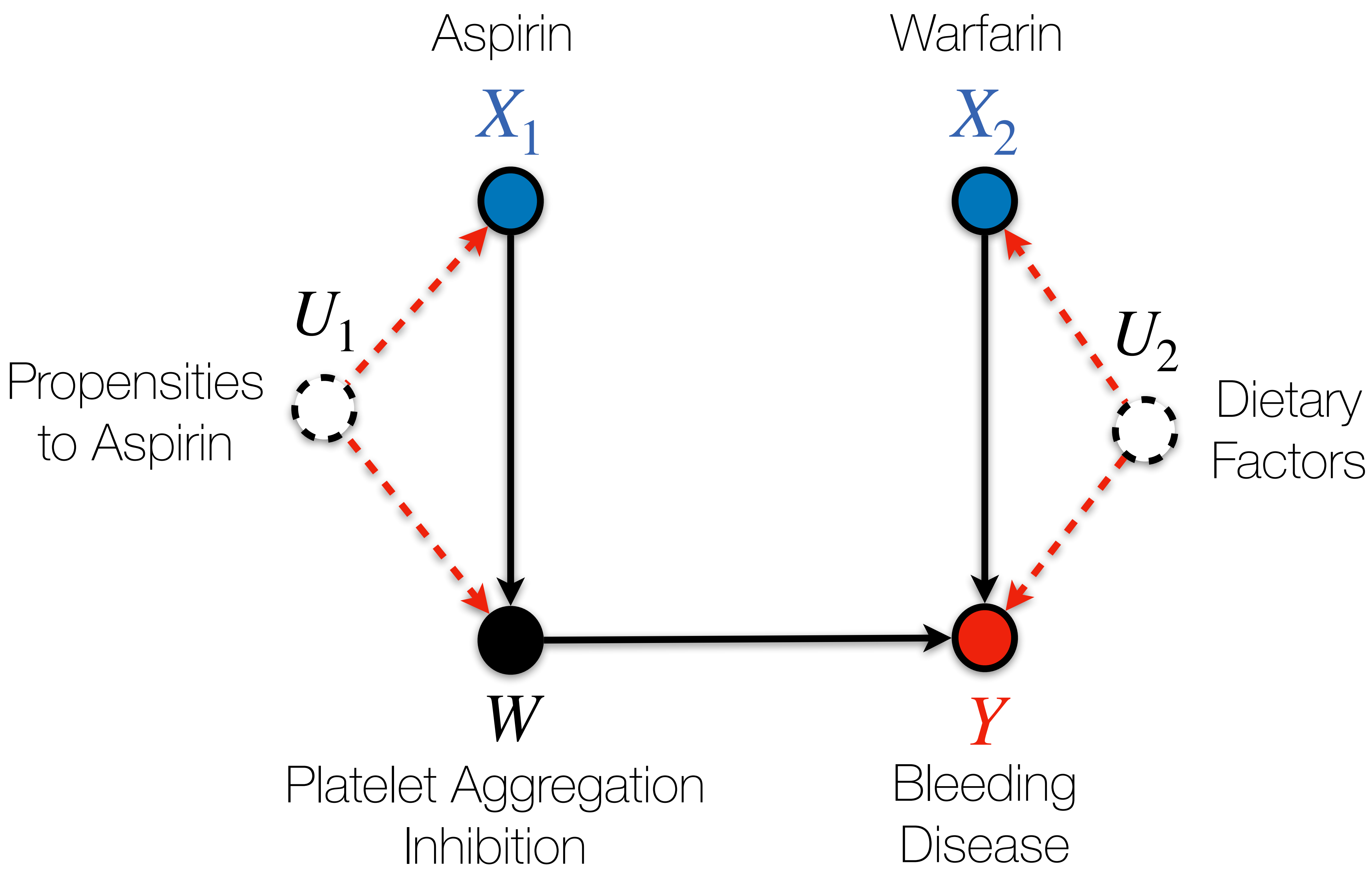
❹ Conclusion

Talk Outline

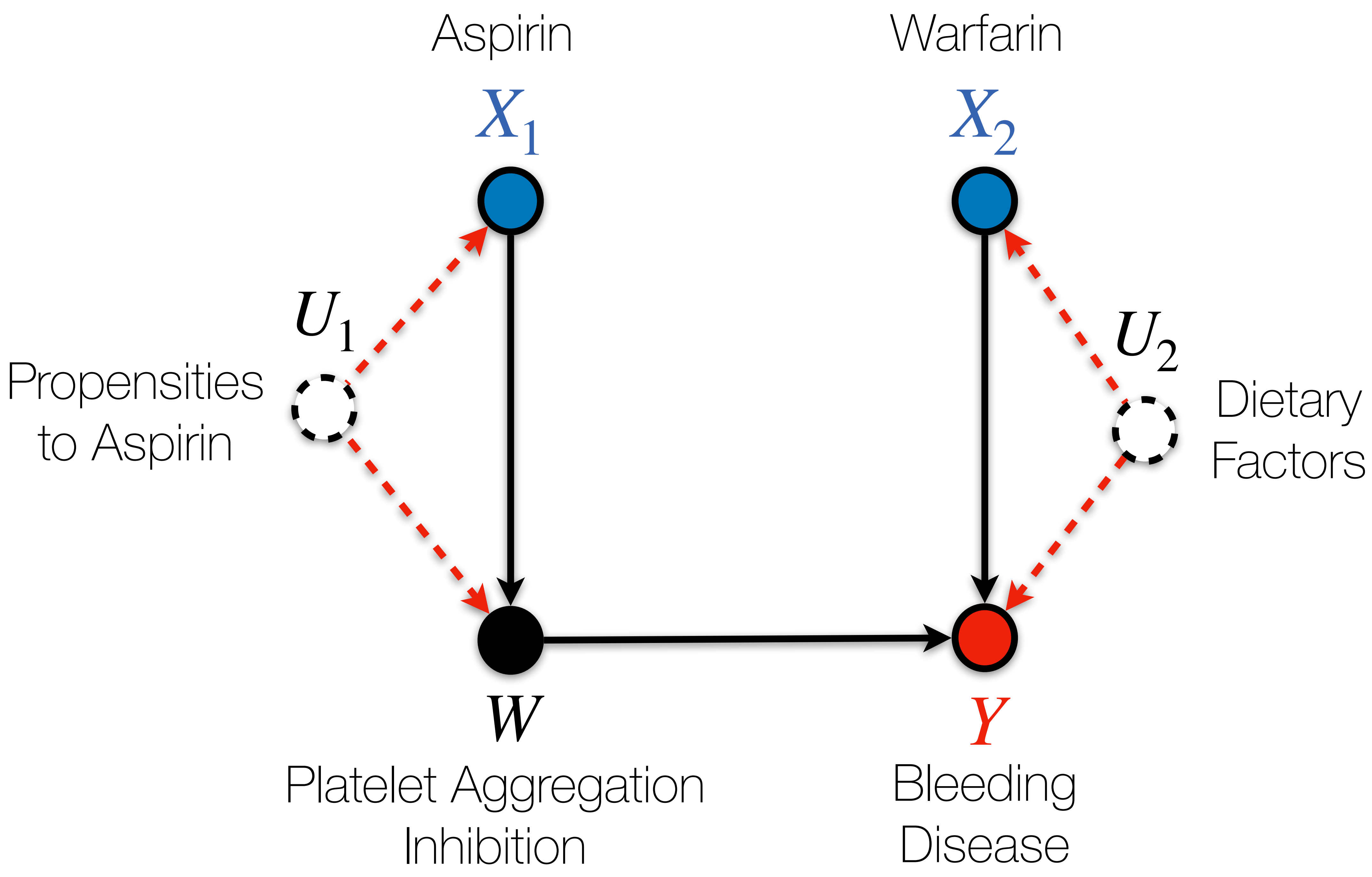
✓

➡ ❷ Estimating causal effects from data fusion

Motivation: Joint Treatment Effect Estimation

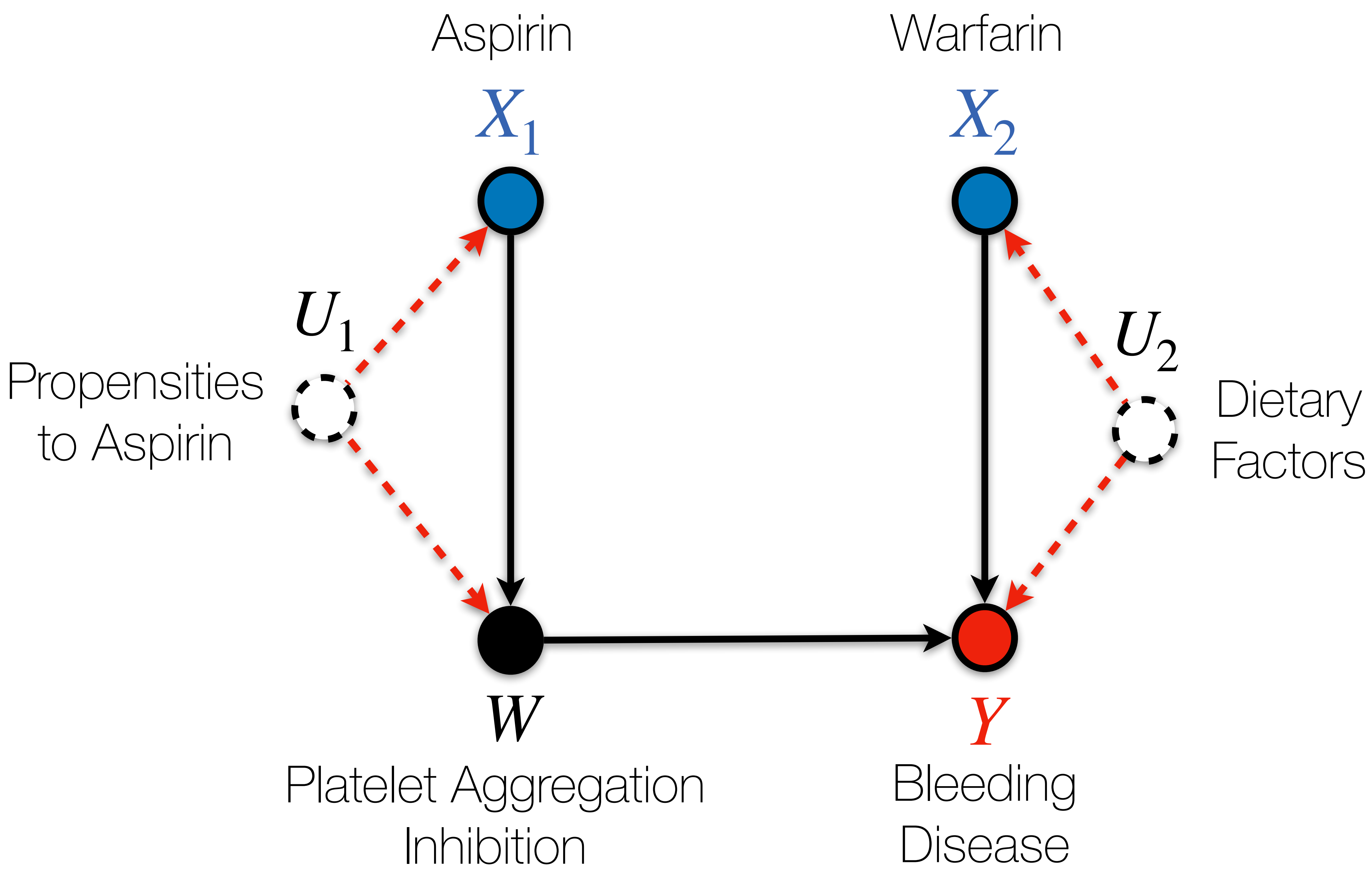


Motivation: Joint Treatment Effect Estimation



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

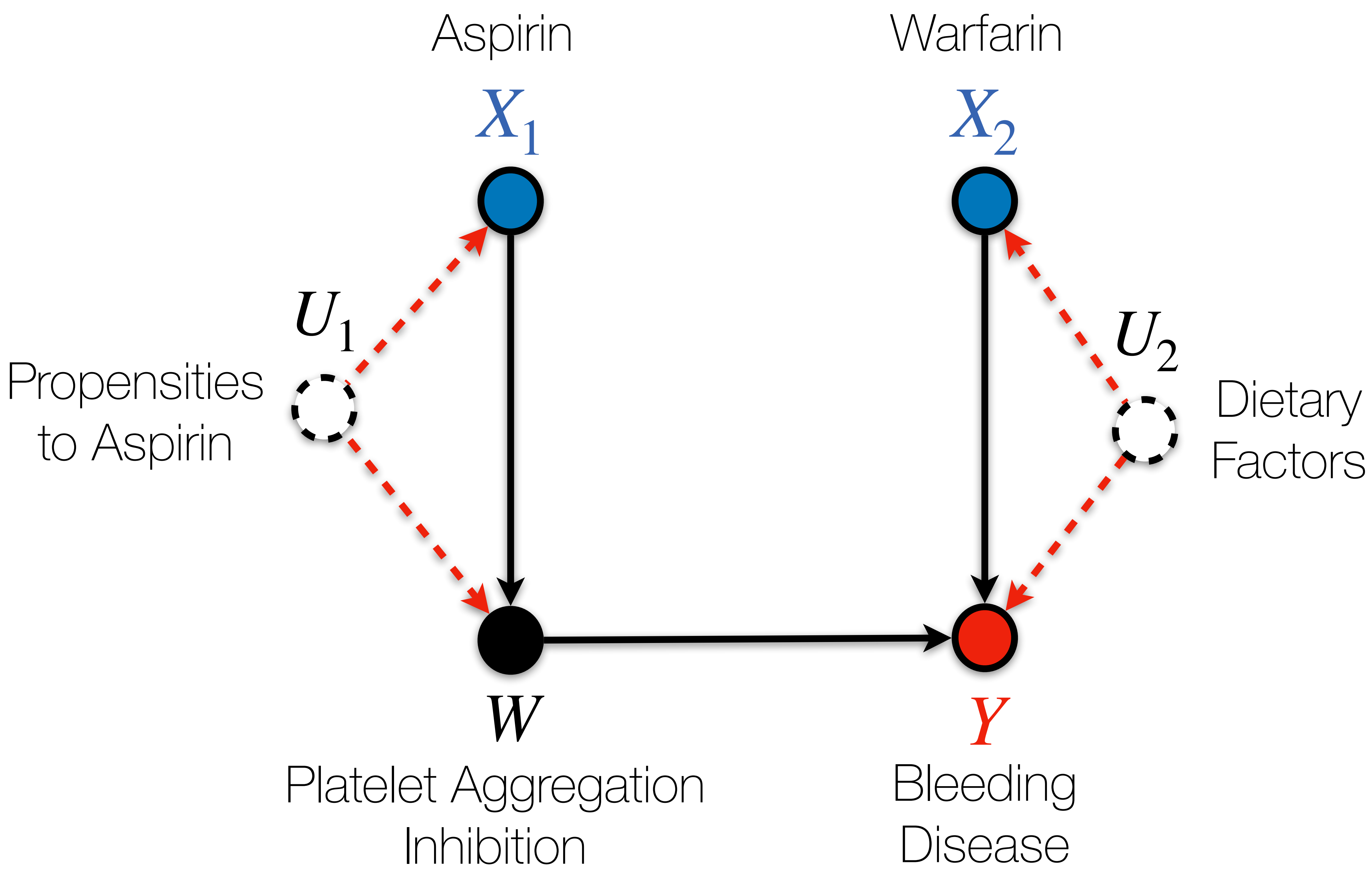
Motivation: Joint Treatment Effect Estimation



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

- BD is not applicable

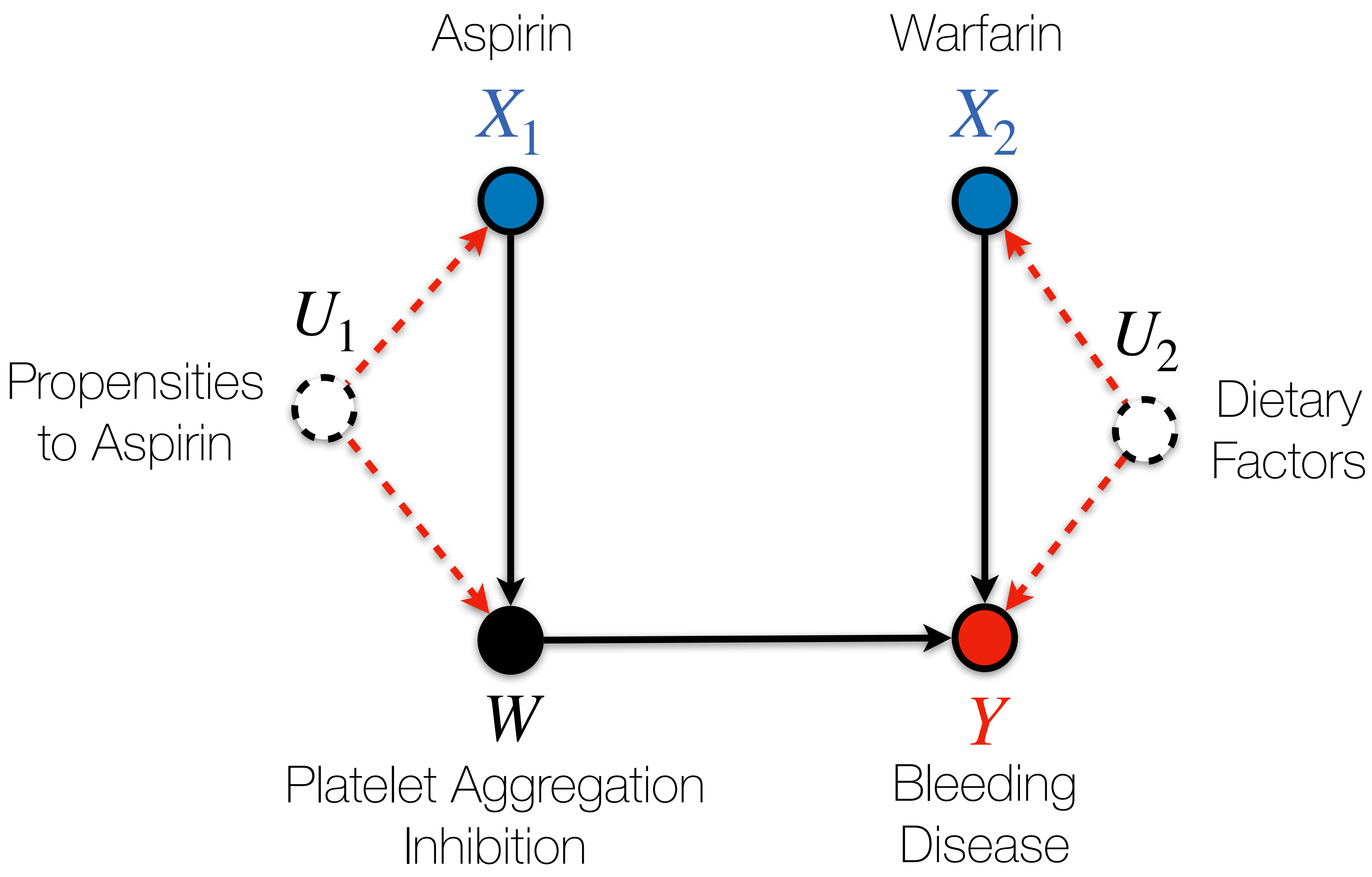
Motivation: Joint Treatment Effect Estimation



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

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

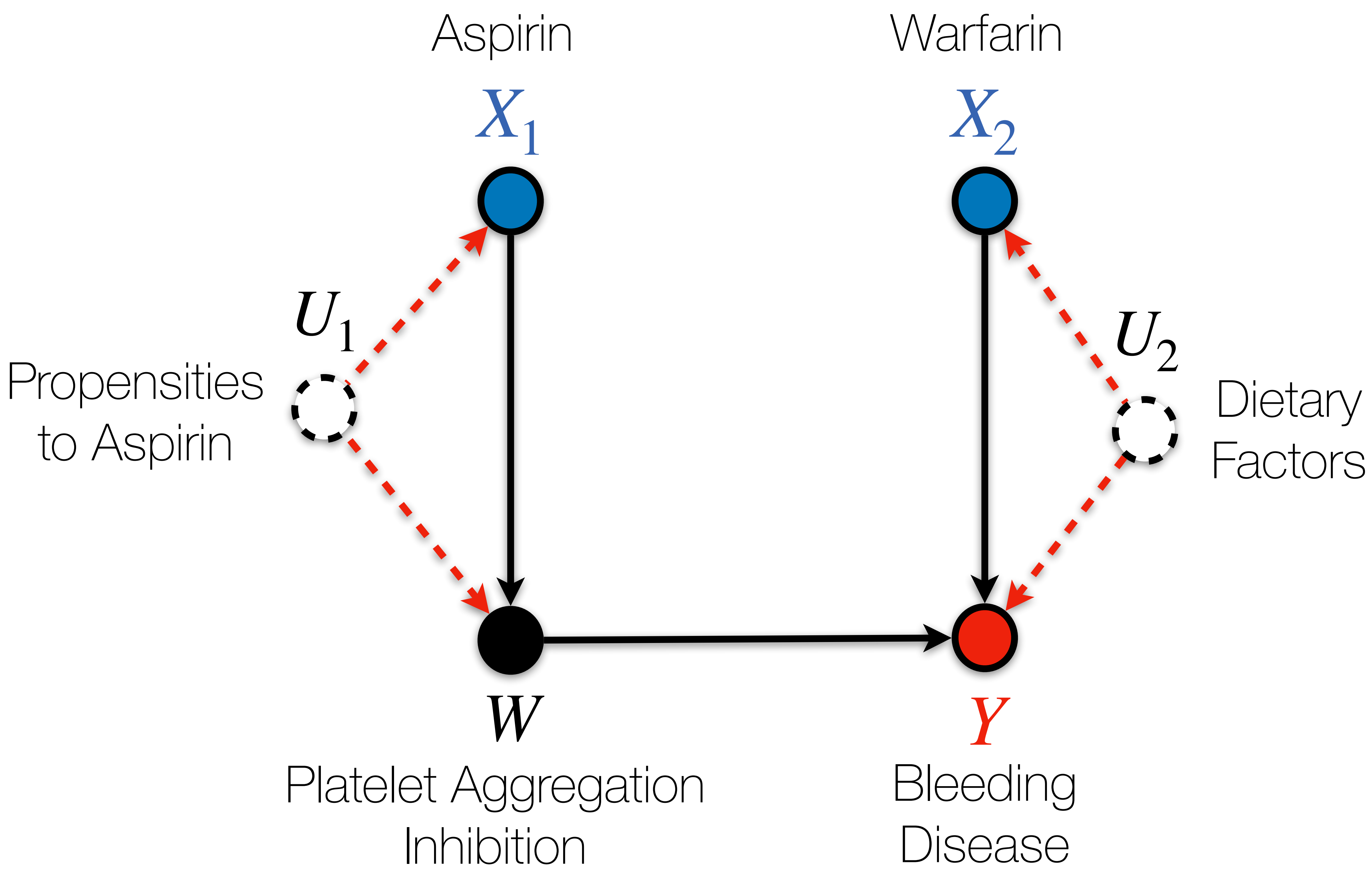
Motivation: Joint Treatment Effect Estimation



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 $P_{\mathrm{do}(x_2)}(\mathbf{V})$?

Joint Treatment Effect Identification

BD Criterion for Joint Treatment Effect (BD$^+$)

(**Jung** et al., ICML 2023)

A set $\mathbf{Z}$ satisfies the *BD criterion from marginal experiments* $P_{\mathrm{do}(\mathbf{x}_1)}$ *and* $P_{\mathrm{do}(\mathbf{x}_2)}$ relative to the outcome $\mathbf{Y}$ for the *joint treatment effect* $(\mathbf{X}_1, \mathbf{X}_2)$ in $\mathscr{G}$ if

1. $\mathbf{Z}$ is not a descendent of $\mathbf{X}_2$ in $\mathscr{G}$ (instead of non-descendant of (X1,X2)); and
2. $\mathbf{Z}$ blocks every spurious path between $\mathbf{X}_1$ and $\mathbf{Y}$ in the experiment $\mathrm{do}(\mathbf{X}_2)$

Joint Treatment Effect Identification

BD Criterion for Joint Treatment Effect (BD$^+$)

(**Jung** et al., ICML 2023)

A set $\mathbf{Z}$ satisfies the *BD criterion from marginal experiments* $P_{\mathrm{do}(\mathbf{x}_1)}$ *and* $P_{\mathrm{do}(\mathbf{x}_2)}$ relative to the outcome $\mathbf{Y}$ for the *joint treatment effect* $(\mathbf{X}_1, \mathbf{X}_2)$ in $\mathcal{G}$ if

1. $\mathbf{Z}$ is not a descendent of $\mathbf{X}_2$ in $\mathcal{G}$ (instead of non-descendant of (X1,X2)); and
2. $\mathbf{Z}$ blocks every spurious path between $\mathbf{X}_1$ and $\mathbf{Y}$ in the experiment $\mathrm{do}(\mathbf{X}_2)$

$$\mathbb{E}[Y \mid \mathrm{do}(\mathbf{x}_1, \mathbf{x}_2)] = \sum_{\mathbf{z}} \mathbb{E}_{\mathrm{do}(\mathbf{x}_2)}[Y \mid \mathbf{x}_1, \mathbf{z}] P_{\mathrm{do}(\mathbf{x}_1)}(\mathbf{z})$$

Joint Treatment Effect Identification

BD Criterion for Joint Treatment Effect (BD$^+$)

(**Jung** et al., ICML 2023)

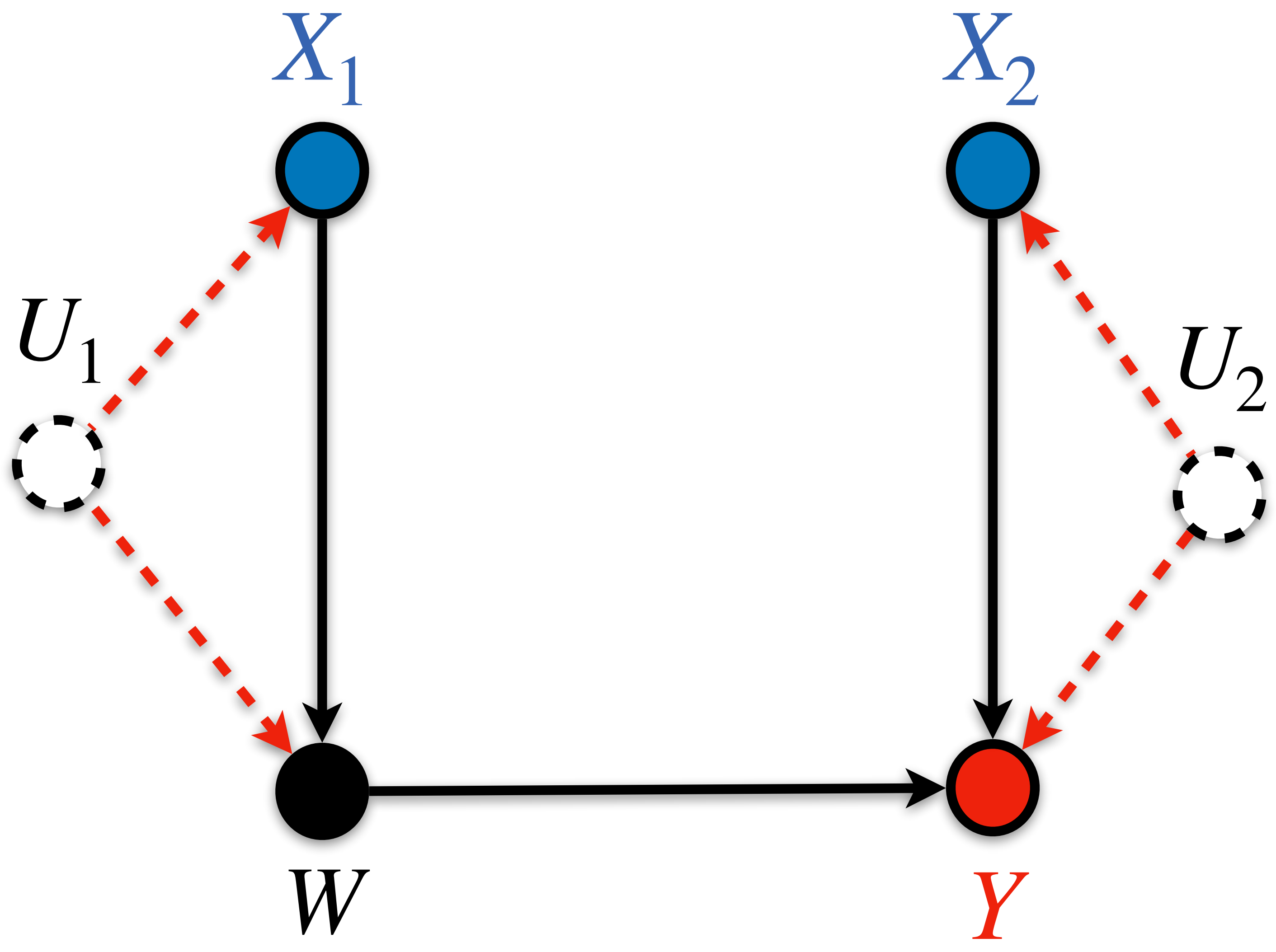
A set $\mathbf{Z}$ satisfies the *BD criterion from marginal experiments* $P_{\mathrm{do}(\mathbf{x}_1)}$ *and* $P_{\mathrm{do}(\mathbf{x}_2)}$ relative to the outcome $\mathbf{Y}$ for the *joint treatment effect* $(\mathbf{X}_1, \mathbf{X}_2)$ in $\mathcal{G}$ if

1. $\mathbf{Z}$ is not a descendent of $\mathbf{X}_2$ in $\mathcal{G}$ (instead of non-descendant of (X1,X2)); and
2. $\mathbf{Z}$ blocks every spurious path between $\mathbf{X}_1$ and $\mathbf{Y}$ in the experiment $\mathrm{do}(\mathbf{X}_2)$

$$\mathbb{E}[Y \mid \mathrm{do}(\mathbf{x}_1, \mathbf{x}_2)] = \sum_{\mathbf{z}} \underbrace{\mathbb{E}_{\mathrm{do}(\mathbf{x}_2)}[Y \mid \mathbf{x}_1, \mathbf{z}]}_{\text{Trial on } X_2} \underbrace{P_{\mathrm{do}(\mathbf{x}_1)}(\mathbf{z})}_{\text{Trial on } X_1}$$

Example of BD+

1. $\mathbf{Z} = \{W\}$ is not a descendent of $\mathbf{X}_2$ in $\mathcal{G}$; and
2. $\mathbf{Z} = \{W\}$ blocks every spurious path between $\mathbf{X}_1$ and $\mathbf{Y}$ in the experiment $\mathrm{do}(\mathbf{X}_2)$



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

Parametrization of BD+

$$\mathbb{E}[Y \mid \mathrm{do}(\mathbf{x}_1, \mathbf{x}_2)] = \sum_{\mathbf{z}} \mathbb{E}_{\mathrm{do}(\mathbf{x}_2)}[Y \mid \mathbf{x}_1, \mathbf{z}] P_{\mathrm{do}(x_1)}(\mathbf{z})$$

Parametrization of BD+

$$\mathbb{E}[Y \mid do(\mathbf{x}_1, \mathbf{x}_2)] = \sum_{\mathbf{z}} \mathbb{E}_{do(\mathbf{x}_2)}[Y \mid \mathbf{x}_1, \mathbf{z}] P_{do(x_1)}(\mathbf{z})$$

$$\mu(\mathbf{X}_1, \mathbf{Z}) \triangleq \mathbb{E}_{do(\mathbf{x}_2)}[Y \mid \mathbf{X}_1, \mathbf{Z}]$$

Parametrization of BD+

$$\mathbb{E}[Y \mid \text{do}(\mathbf{x}_1, \mathbf{x}_2)] = \sum_{\mathbf{z}} \mathbb{E}_{\text{do}(\mathbf{x}_2)}[Y \mid \mathbf{x}_1, \mathbf{z}] P_{\text{do}(x_1)}(\mathbf{z})$$

$$\mu(\mathbf{X}_1, \mathbf{Z}) \triangleq \mathbb{E}_{\text{do}(\mathbf{x}_2)}[Y \mid \mathbf{X}_1, \mathbf{Z}]$$

$$\mathbb{E}_{\text{do}(\mathbf{x}_1)}[\mu(\mathbf{x}_1, \mathbf{Z})]$$

$$= \sum_{\mathbf{z}} \mu(\mathbf{x}_1, \mathbf{z}) P_{\text{do}(\mathbf{x}_1)}(\mathbf{z})$$

$$= \mathbb{E}[Y \mid \text{do}(\mathbf{x}_1, \mathbf{x}_2)]$$

Parametrization of BD+

$$\mathbb{E}[Y \mid \mathrm{do}(\mathbf{x}_1, \mathbf{x}_2)] = \sum_{\mathbf{z}} \mathbb{E}_{\mathrm{do}(\mathbf{x}_2)}[Y \mid \mathbf{x}_1, \mathbf{z}] P_{\mathrm{do}(x_1)}(\mathbf{z})$$

$$\mu(\mathbf{X}_1,\mathbf{Z}) \triangleq \mathbb{E}_{\mathrm{do}(\mathbf{x}_2)}[Y \mid \mathbf{X}_1, \mathbf{Z}]$$

$$\mathbb{E}_{\mathrm{do}(\mathbf{x}_1)}[\mu(\mathbf{x}_1,\mathbf{Z})]$$

$$= \sum_{\mathbf{z}} \mu(\mathbf{x}_1,\mathbf{z}) P_{\mathrm{do}(\mathbf{x}_1)}(\mathbf{z})$$

$$= \mathbb{E}[Y \mid \mathrm{do}(\mathbf{x}_1, \mathbf{x}_2)]$$

$\pi(\mathbf{X}_1,\mathbf{Z})$: Solution of

$$\mathbb{E}_{\mathrm{do}(\mathbf{x}_2)}[\pi(\mathbf{X}_1\mathbf{Z}) \times \mu(\mathbf{X}_1\mathbf{Z})] = \mathbb{E}_{\mathrm{do}(\mathbf{x}_1)}[\mu(\mathbf{x}_1,\mathbf{Z})]$$

Parametrization of BD+

$$\mathbb{E}[Y \mid do(\mathbf{x}_1, \mathbf{x}_2)] = \sum_{\mathbf{z}} \mathbb{E}_{do(\mathbf{x}_2)}[Y \mid \mathbf{x}_1, \mathbf{z}] P_{do(x_1)}(\mathbf{z})$$

$$\mu(\mathbf{X}_1, \mathbf{Z}) \triangleq \mathbb{E}_{do(\mathbf{x}_2)}[Y \mid \mathbf{X}_1, \mathbf{Z}]$$

$$\mathbb{E}_{do(\mathbf{x}_1)}[\mu(\mathbf{x}_1, \mathbf{Z})]$$

$$= \sum_{\mathbf{z}} \mu(\mathbf{x}_1, \mathbf{z}) P_{do(\mathbf{x}_1)}(\mathbf{z})$$

$$= \mathbb{E}[Y \mid do(\mathbf{x}_1, \mathbf{x}_2)]$$

$\pi(\mathbf{X}_1, \mathbf{Z})$: Solution of

$$\mathbb{E}_{do(\mathbf{x}_2)}[\pi(\mathbf{X}_1 \mathbf{Z}) \times \mu(\mathbf{X}_1 \mathbf{Z})] = \mathbb{E}_{do(\mathbf{x}_1)}[\mu(\mathbf{x}_1, \mathbf{Z})]$$

$$\mathbb{E}_{do(\mathbf{x}_2)}[\pi(\mathbf{X}_1 \mathbf{Z}) \times Y]$$

$$= \mathbb{E}_{do(\mathbf{x}_2)}[\pi(\mathbf{X}_1 \mathbf{Z}) \times \mu(\mathbf{X}_1 \mathbf{Z})]$$

$$= \mathbb{E}_{do(\mathbf{x}_1)}[\mu(\mathbf{x}_1, \mathbf{Z})]$$

$$= \mathbb{E}[Y \mid do(\mathbf{x}_1, \mathbf{x}_2)]$$

Doubly Robust Estimator for BD+

$$\mathbb{E}[Y \mid do(\mathbf{x}_1, \mathbf{x}_2)] = \text{BD}^{+}(\mu, \pi) \triangleq \mathbb{E}_{do(x_2)}[\mu \times \pi]$$

Doubly Robust Estimator for BD+

$$\mathbb{E}[Y \mid \mathrm{do}(\mathbf{x}_1, \mathbf{x}_2)] = \mathrm{BD}^{+}(\mu,\pi) \triangleq \mathbb{E}_{\mathrm{do}(\mathrm{x}_2)}[\mu \times \pi]$$

“Double Robustness”

$$\mathbf{?}(\hat{\mu},\hat{\pi}) - \mathbb{E}_{\mathrm{do}(x_2)}[\mu \times \pi] = \mathbb{E}_{\mathrm{do}(x_2)}[\{\hat{\mu}-\mu\}\times\{\pi-\hat{\pi}\}]$$

Doubly Robust Estimator for BD+

$$\mathbb{E}[Y \mid \mathrm{do}(\mathbf{x}_1, \mathbf{x}_2)] = \text{BD}^{+}(\mu, \pi) \triangleq \mathbb{E}_{\mathrm{do}(\mathrm{x}_2)}[\mu \times \pi]$$

$$\mathbf{?}(\hat{\mu}, \hat{\pi}) = \mathbb{E}_{\mathrm{do}(x_2)}[\{\hat{\mu} - \mu\} \times \{\pi - \hat{\pi}\}] + \mathbb{E}_{\mathrm{do}(x_2)}[\mu \times \pi]$$

Doubly Robust Estimator for BD$^+$

$$\mathbb{E}[Y \mid \mathrm{do}(\mathbf{x}_1, \mathbf{x}_2)] = \mathrm{BD}^+(\mu,\pi) \triangleq \mathbb{E}_{\mathrm{do}(\mathrm{x}_2)}[\mu\times\pi]$$

$$\textbf{?}(\hat{\mu},\hat{\pi}) = \mathbb{E}_{\mathrm{do}(x_2)}[\{\hat{\mu}-\mu\}\times\{\pi-\hat{\pi}\}] + \mathbb{E}_{\mathrm{do}(x_2)}[\mu\times\pi]$$

$$= \mathbb{E}_{\mathrm{do}(x_2)}[\hat{\pi}\{\mu-\hat{\mu}\}+\pi\hat{\mu}]$$

Doubly Robust Estimator for BD+

$$\mathbb{E}[Y \mid \mathrm{do}(\mathbf{x}_1, \mathbf{x}_2)] = \mathrm{BD}^{+}(\mu,\pi) \triangleq \mathbb{E}_{\mathrm{do}(\mathrm{x}_2)}[\mu\times\pi]$$

$$\begin{aligned}
\textbf{?}(\hat{\mu},\hat{\pi}) &= \mathbb{E}_{\mathrm{do}(x_2)}[\{\hat{\mu}-\mu\}\times\{\pi-\hat{\pi}\}] + \mathbb{E}_{\mathrm{do}(x_2)}[\mu\times\pi] \\
&= \mathbb{E}_{\mathrm{do}(x_2)}[\hat{\pi}\{\mu-\hat{\mu}\}+\pi\hat{\mu}] \\
&= \mathbb{E}_{\mathrm{do}(x_2)}[\hat{\pi}\{Y-\hat{\mu}\}] + \mathbb{E}_{\mathrm{do}(x_1)}[\hat{\mu}(x, C)]
\end{aligned}$$

Doubly Robust Estimator for BD+

$$\mathbb{E}[Y \mid \mathrm{do}(\mathbf{x}_1, \mathbf{x}_2)] = \text{BD}^{+}(\mu, \pi) \triangleq \mathbb{E}_{\mathrm{do}(\mathrm{x}_2)}[\mu \times \pi]$$

$$\begin{aligned}
\mathbf{?}(\hat{\mu}, \hat{\pi}) &= \mathbb{E}_{\mathrm{do}(x_2)}[\{\hat{\mu} - \mu\} \times \{\pi - \hat{\pi}\}] + \mathbb{E}_{\mathrm{do}(x_2)}[\mu \times \pi] \\
&= \mathbb{E}_{\mathrm{do}(x_2)}[\hat{\pi}\{\mu - \hat{\mu}\} + \pi\hat{\mu}] \\
&= \mathbb{E}_{\mathrm{do}(x_2)}[\hat{\pi}\{Y - \hat{\mu}\}] + \mathbb{E}_{\mathrm{do}(x_1)}[\hat{\mu}(x, C)]
\end{aligned}$$

DML-BD+

$$\widehat{\text{BD}^{+}}(\hat{\mu}, \hat{\pi}) \triangleq \mathbb{E}_{\mathrm{do}(x_2)}[\hat{\pi}\{Y - \hat{\mu}\}] + \mathbb{E}_{\mathrm{do}(x_1)}[\hat{\mu}(x, C)]$$

Robustness of DML-BD$^+$

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

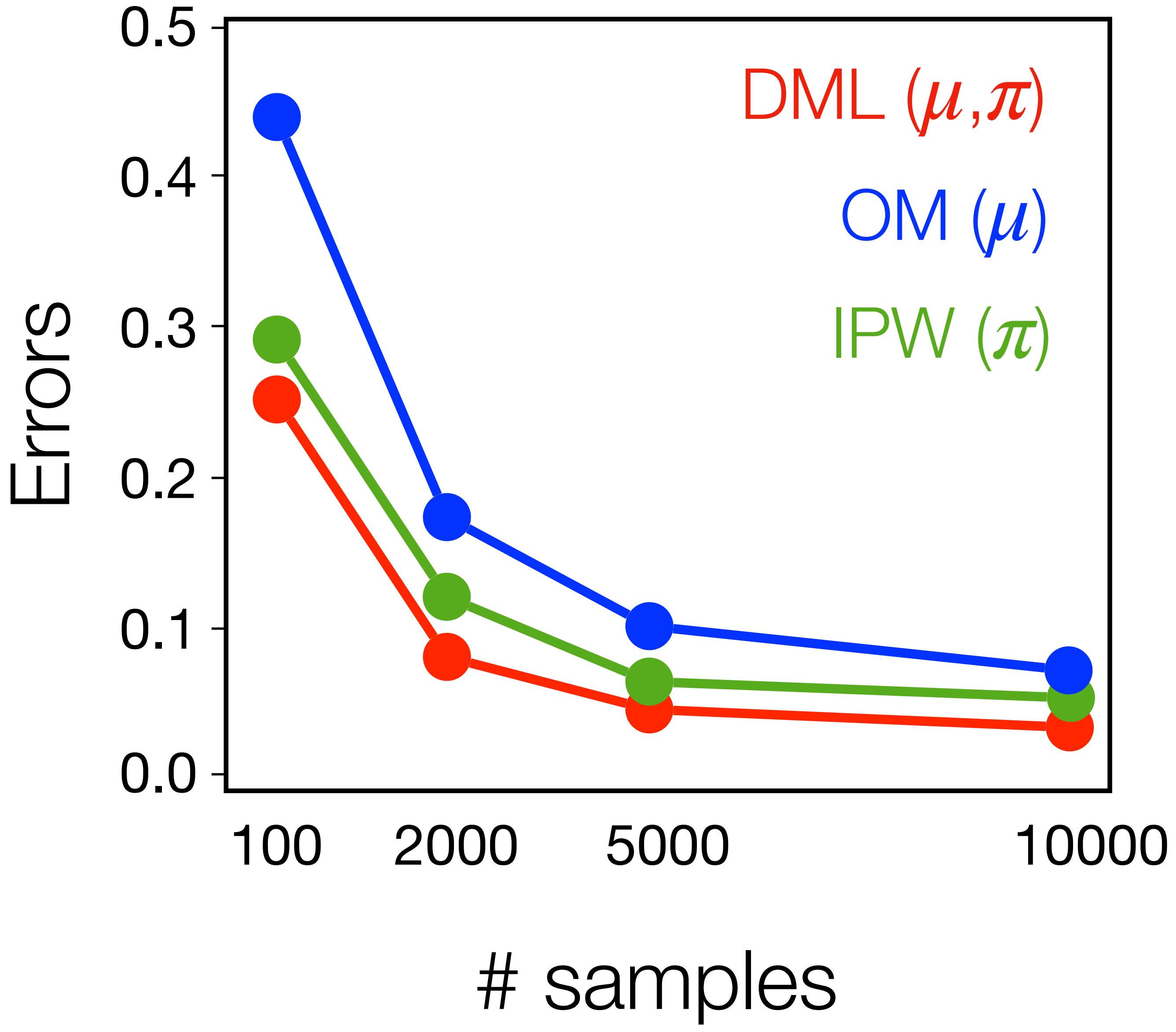
- **Double Robustness:** Error $= 0$ if either $\hat{\mu} = \mu$ or $\hat{\pi} = \pi$
- **Fast Convergence:** Error $\to 0$ *fast* even when $\hat{\mu} \to \mu$ and $\hat{\pi} \to \pi$ *slowly*.

Simulation: DML-BD+

Simulation: DML-BD+

Fast Convergence

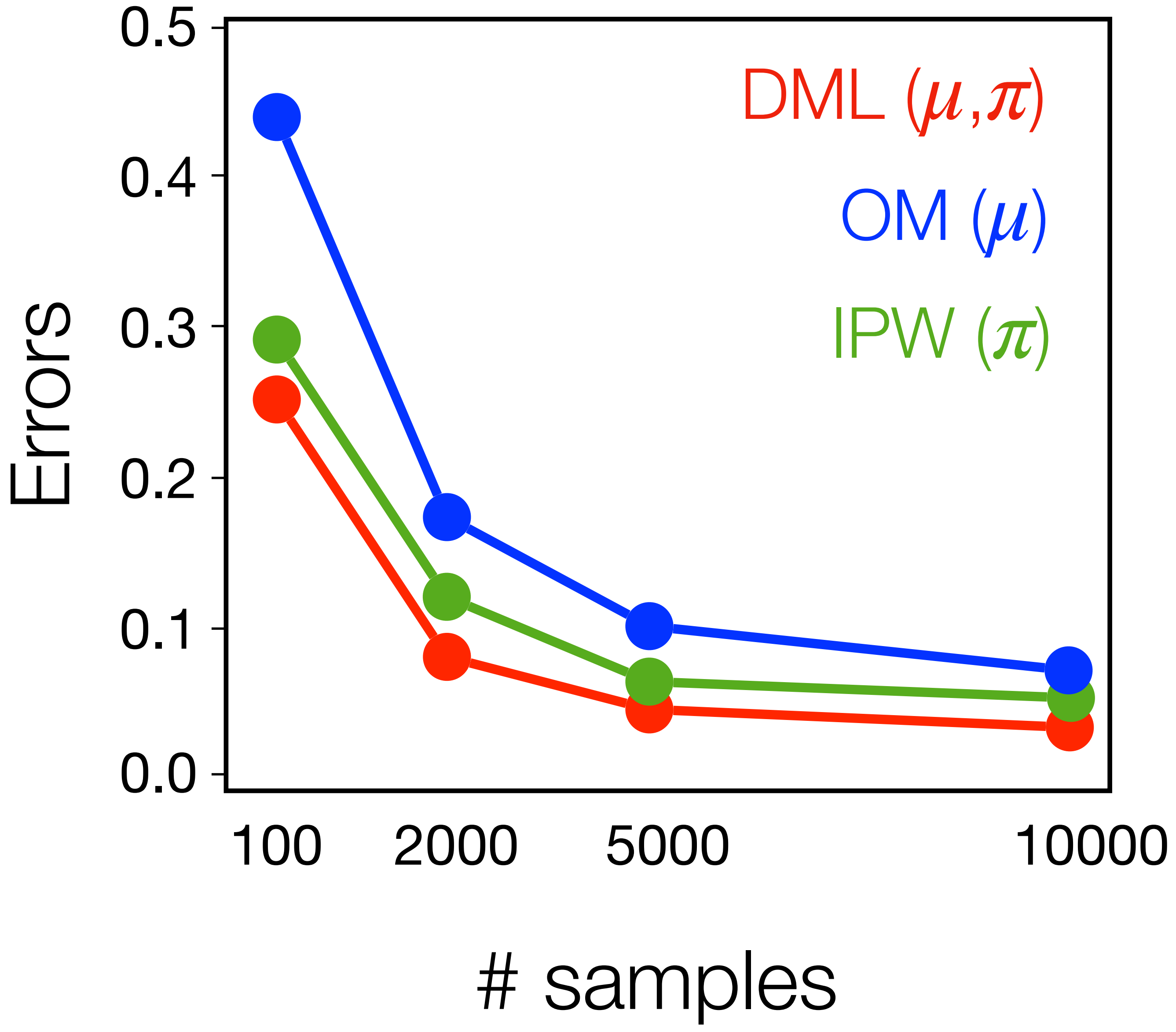
$(\hat{\boldsymbol{\mu}}, \hat{\boldsymbol{\pi}}) \rightarrow (\boldsymbol{\mu}_0, \boldsymbol{\pi}_0)$ slowly



Simulation: DML-BD$^+$

Fast Convergence

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

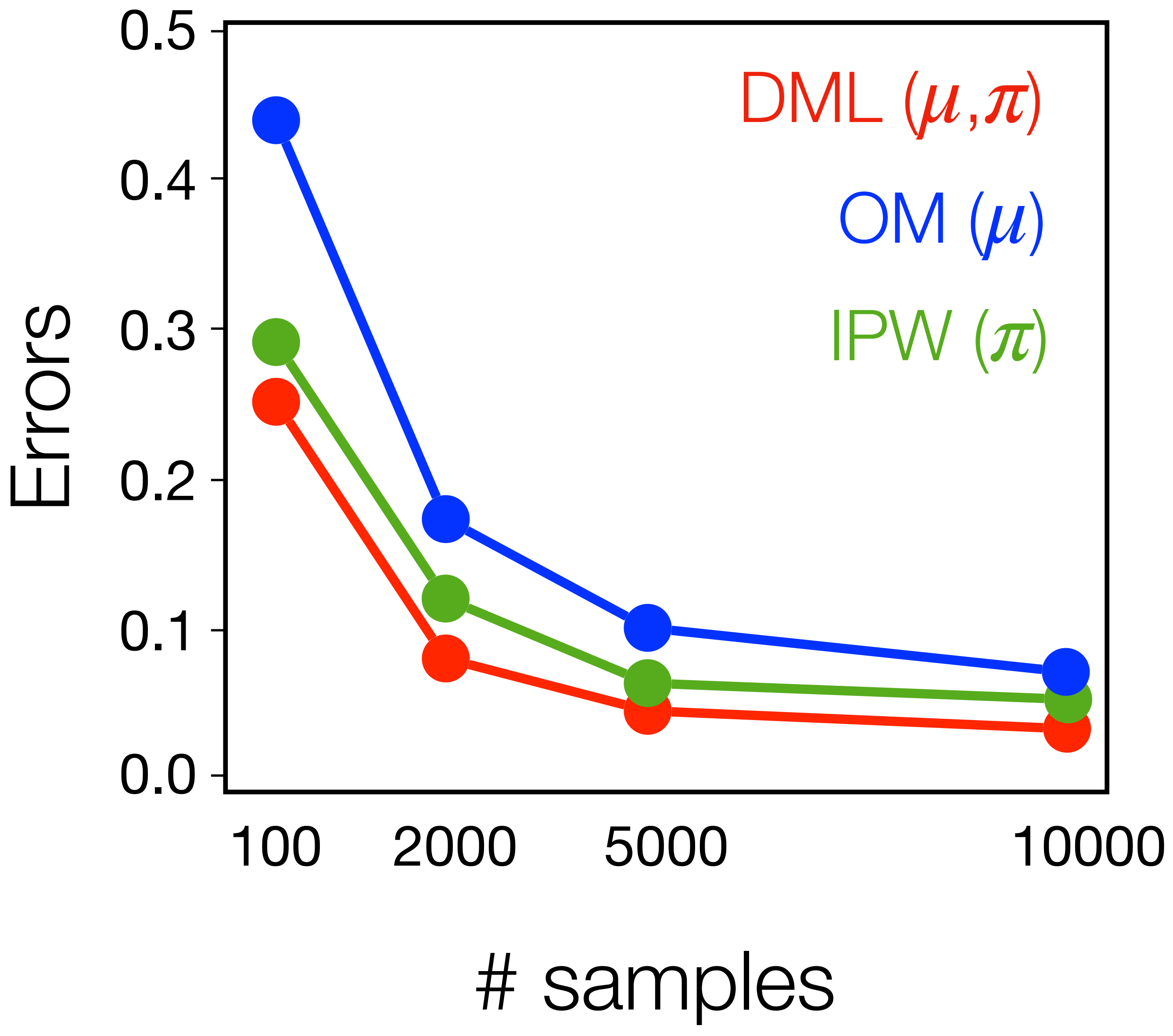


DML-BD$^+$ converges fast, even when $(\hat{\boldsymbol{\mu}}, \hat{\boldsymbol{\pi}})$ converge slowly

Simulation: DML-BD$^+$

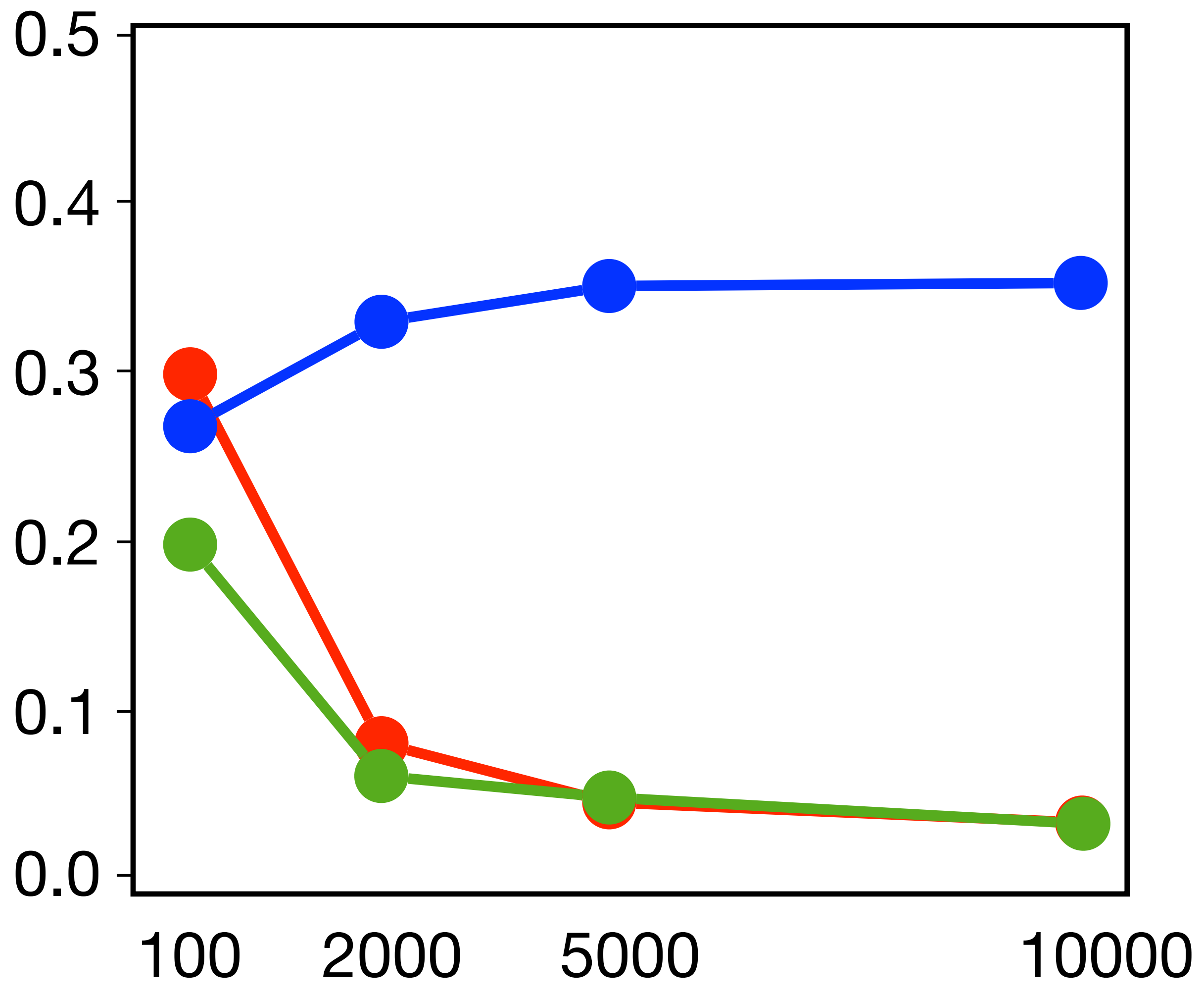
Fast Convergence

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

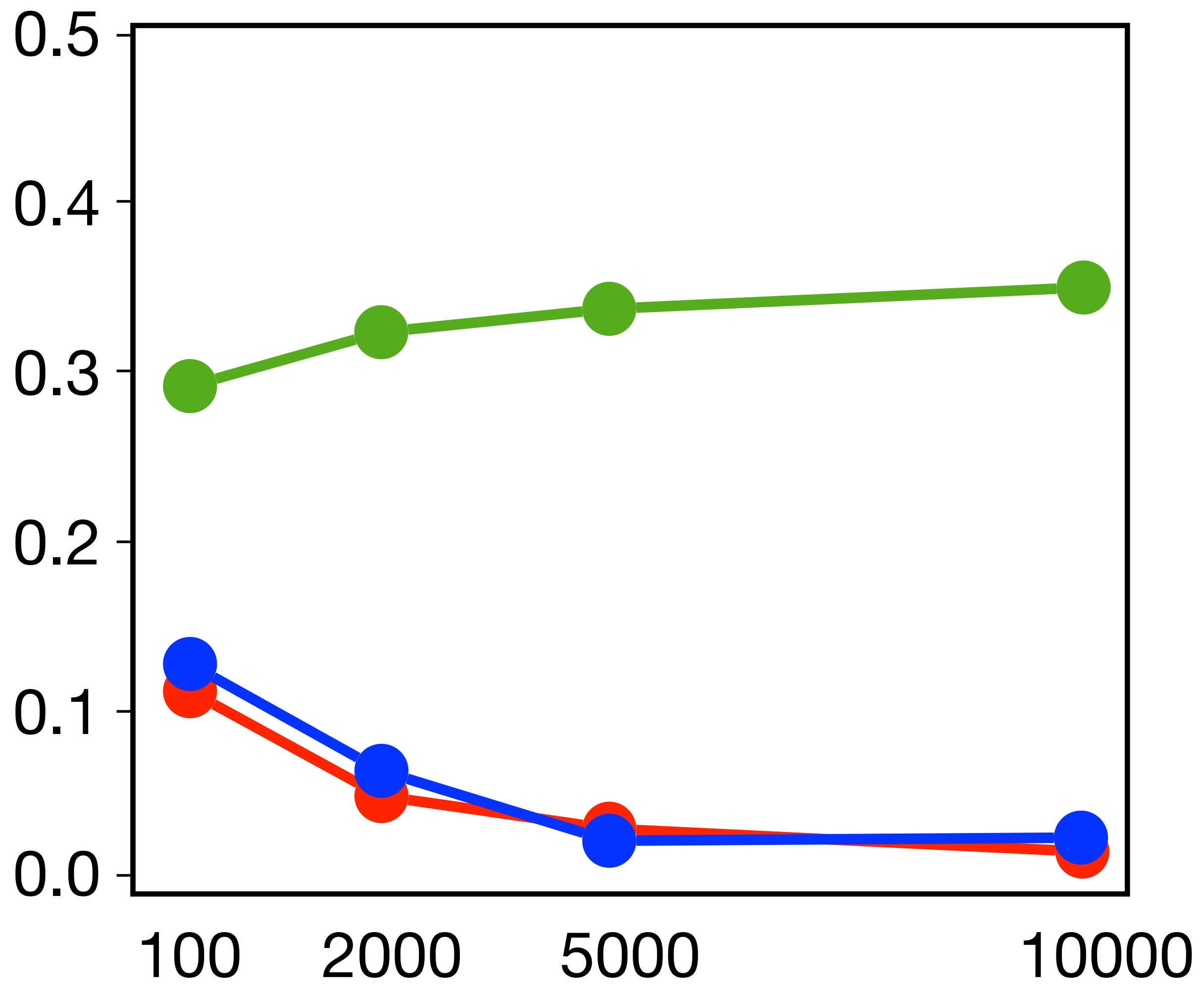


Double Robustness

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



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

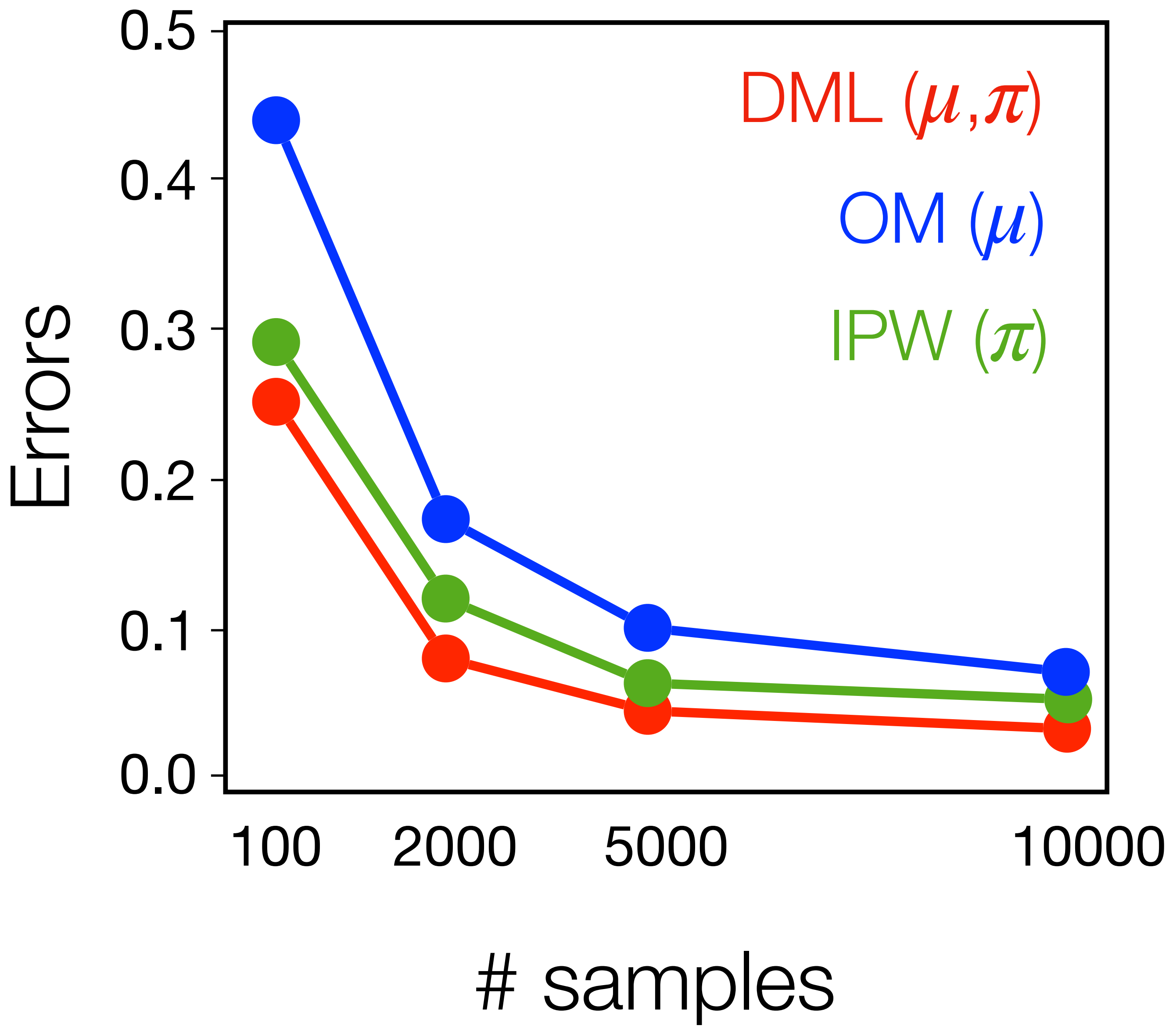


DML-BD$^+$ converges fast, even when $(\hat{\boldsymbol{\mu}},\hat{\boldsymbol{\pi}})$ converge slowly

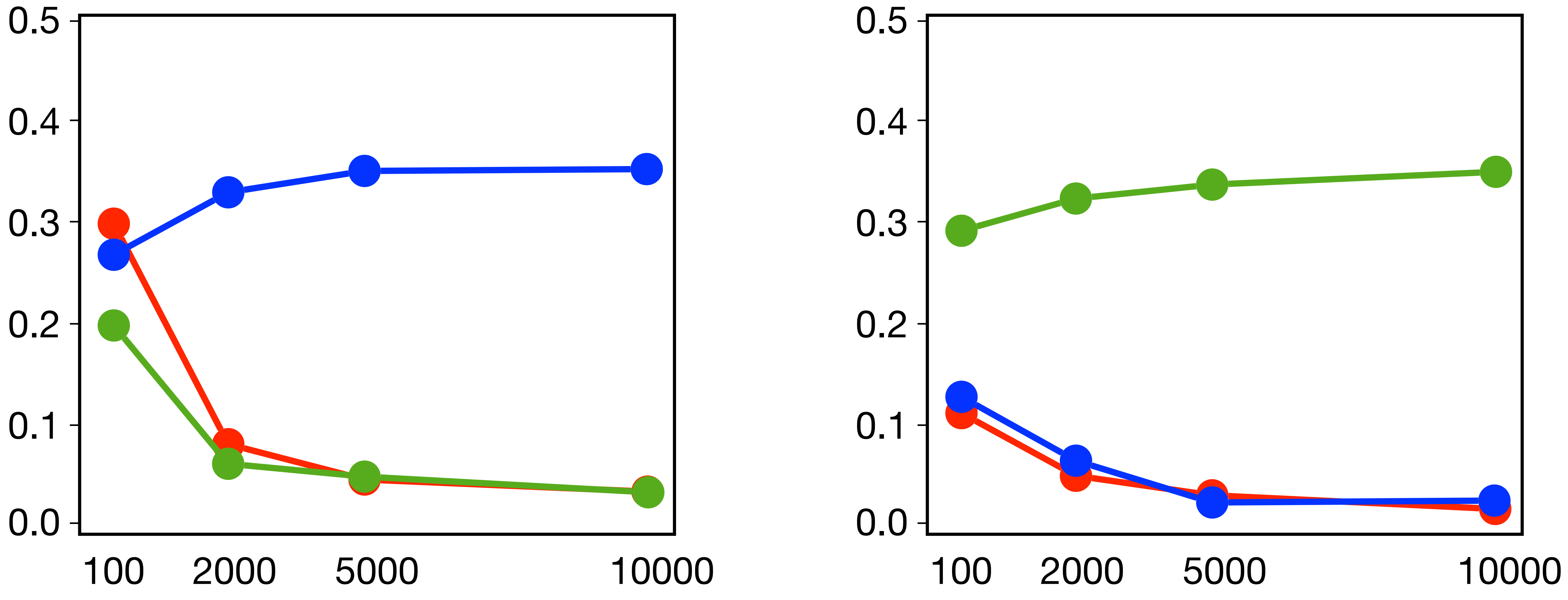
Simulation: DML-BD$^+$

Fast Convergence

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



Double Robustness



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

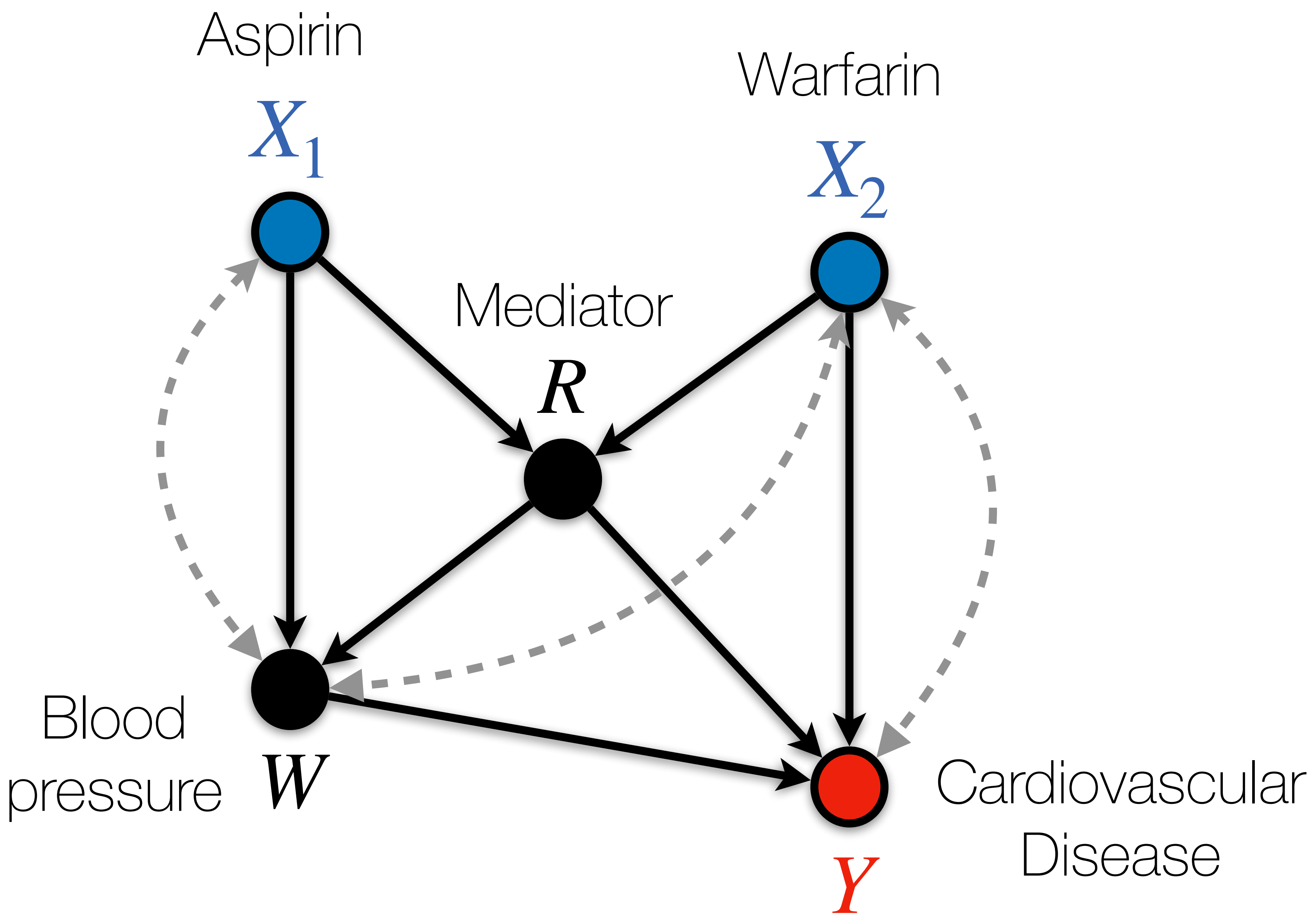
$\hat{\boldsymbol{\pi}}$ misspecified ($\hat{\boldsymbol{\pi}} \neq \boldsymbol{\pi}$)

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

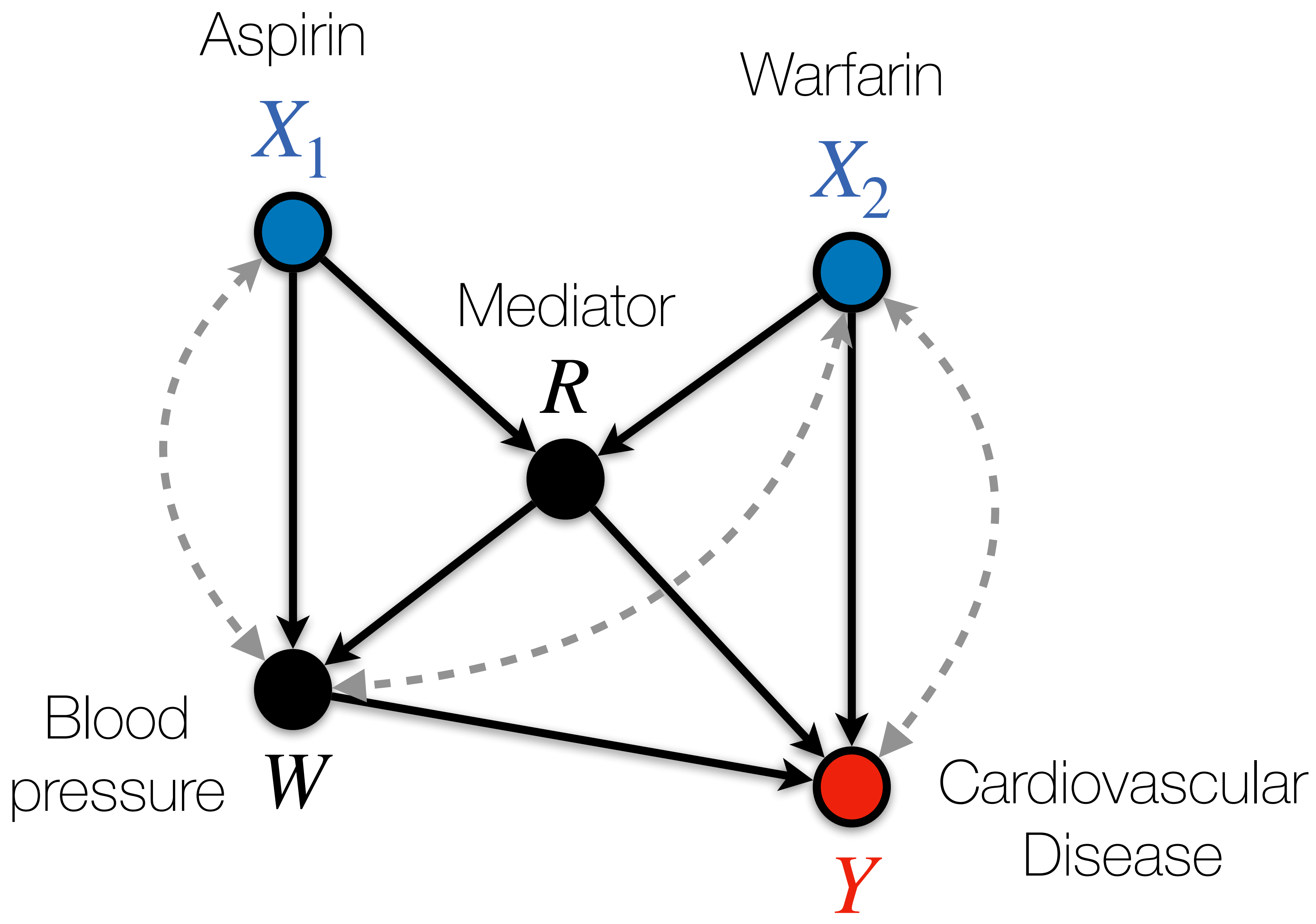
DML-BD$^+$ converges to the true causal effect even when $\hat{\boldsymbol{\mu}}$ or $\hat{\boldsymbol{\pi}}$ are misspecified.

Example where BD^+ Fails

Example where BD+ Fails

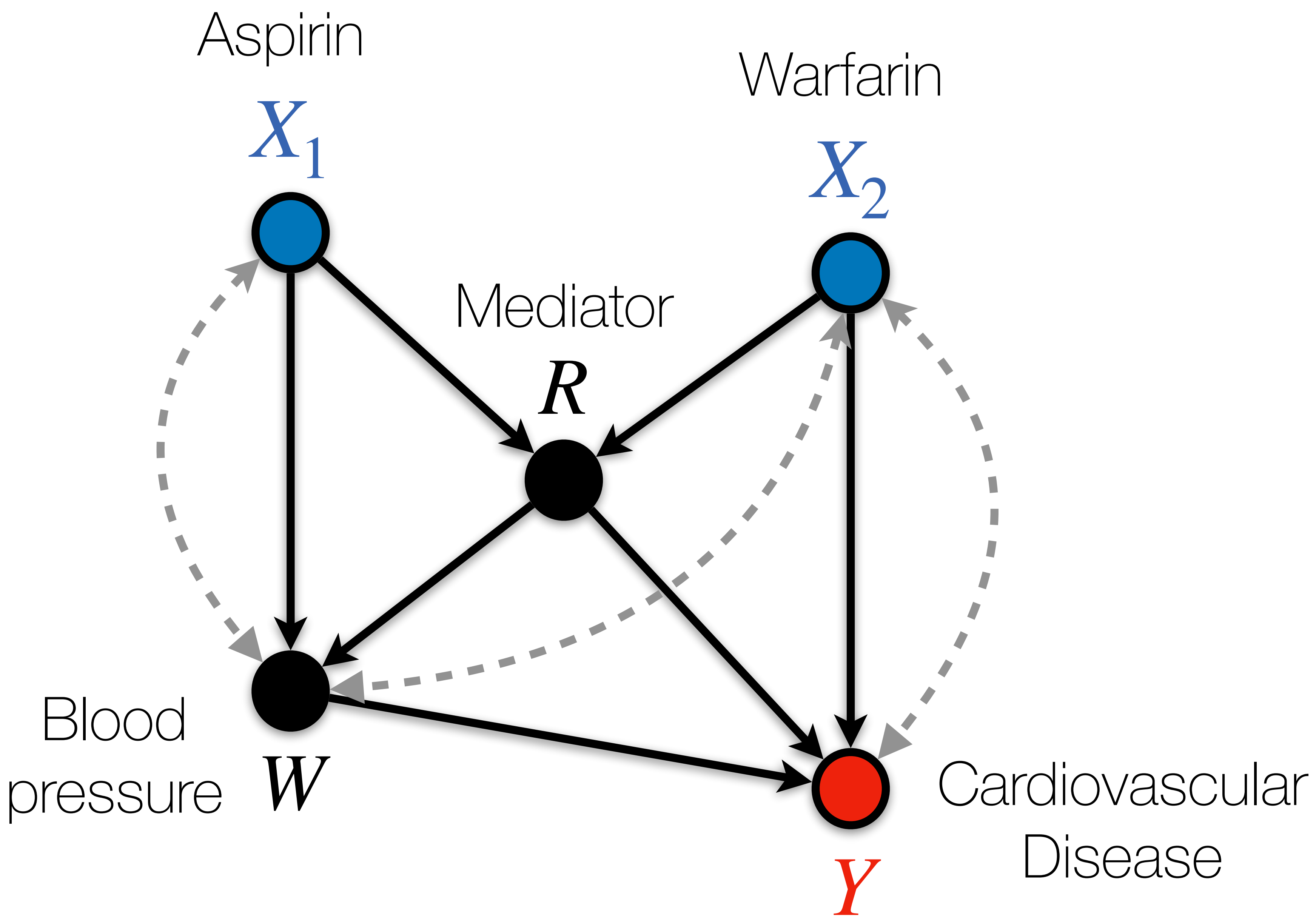


Example where BD+ Fails



$$\sum_{rw} P_{\text{do}(x_1)}(r \mid x_2) P_{\text{do}(x_2)}(y \mid rwx_1) \sum_{x_2'} P_{\text{do}(x_1)}(w \mid r, x_2') P_{\text{do}(x_1)}(x_2')$$

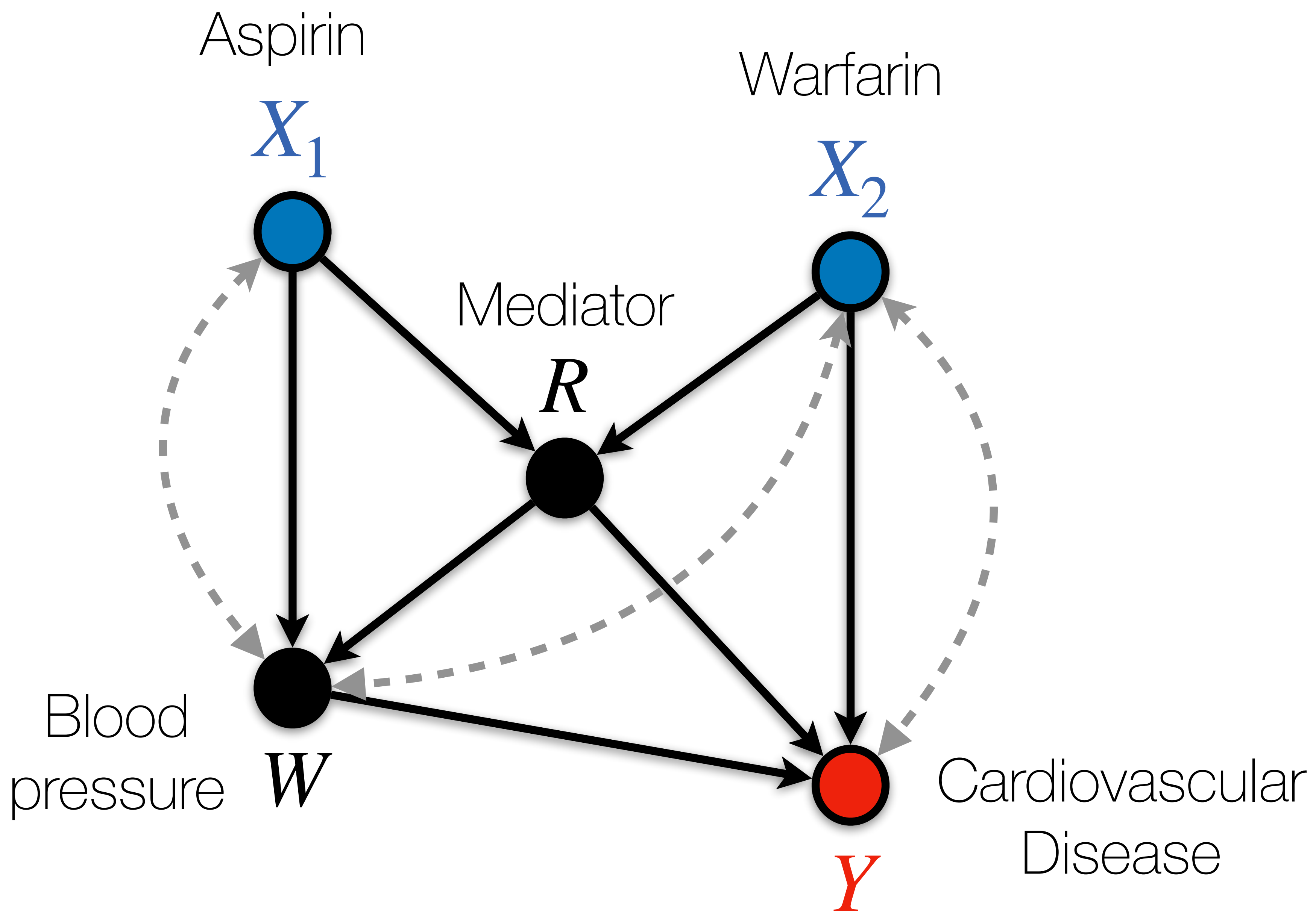
Example where BD+ Fails



✗ BD+ fails

$$\sum_{rw} P_{\mathrm{do}(x_1)}(r \mid x_2) P_{\mathrm{do}(x_2)}(y \mid rwx_1) \sum_{x_2'} P_{\mathrm{do}(x_1)}(w \mid r, x_2') P_{\mathrm{do}(x_1)}(x_2')$$

Example where BD+ Fails



✗ BD+ fails

$$\sum_{rw} P_{\text{do}(x_1)}(r \mid x_2) P_{\text{do}(x_2)}(y \mid rwx_1) \sum_{x_2'} P_{\text{do}(x_1)}(w \mid r, x_2') P_{\text{do}(x_1)}(x_2')$$

Can $\mathbb{E}[Y \mid \text{do}(x_1, x_2)]$ be sample-efficiently estimated?

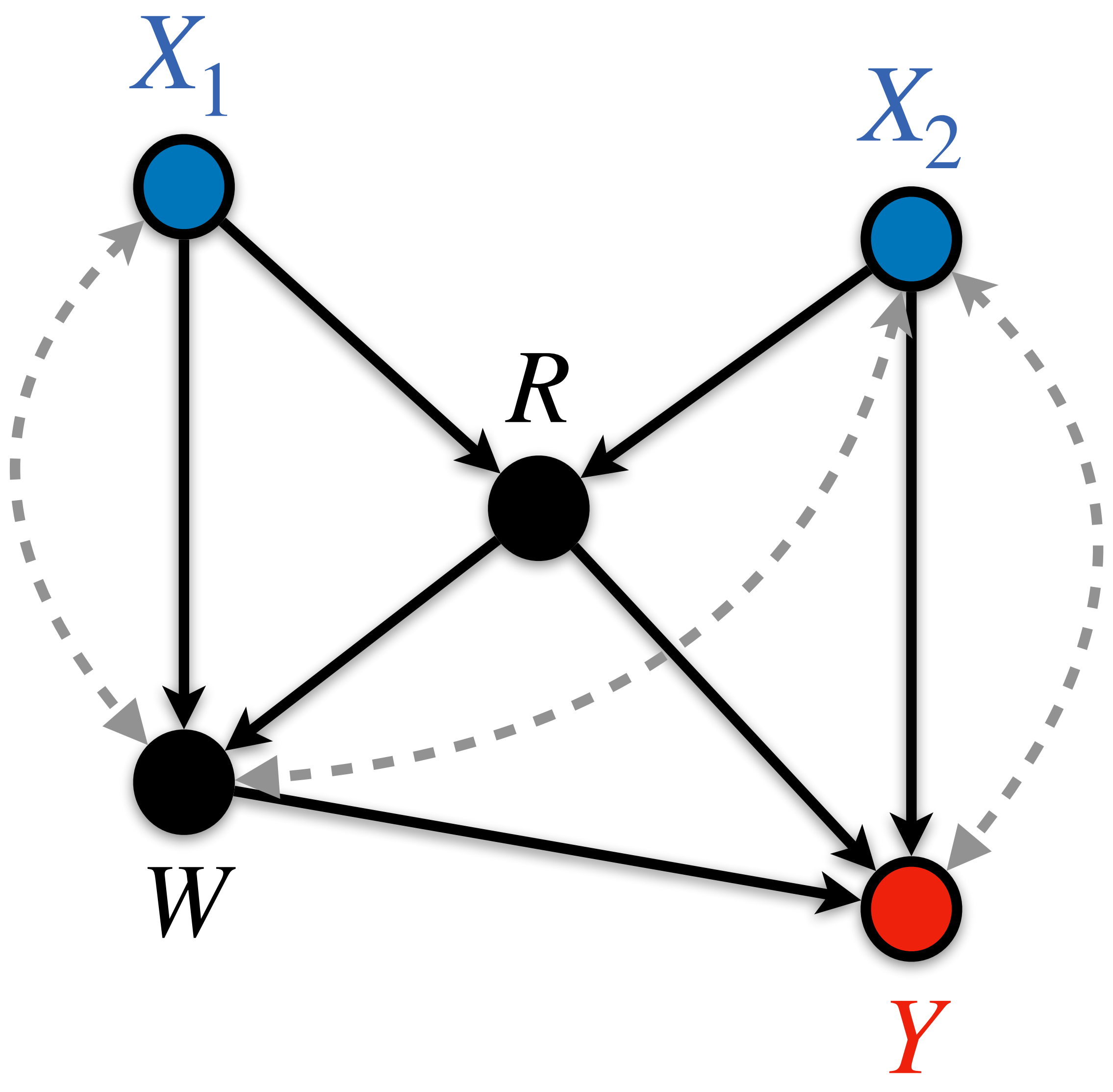
Background: General Identification from Data Fusion

General Identification (gID)

Bareinboim and Pearl, 2012; Lee et al. 2019

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

Background: General Identification from Data Fusion



Available distributions

$P_{\mathrm{do}(x_1)}(RWX_2Y)$

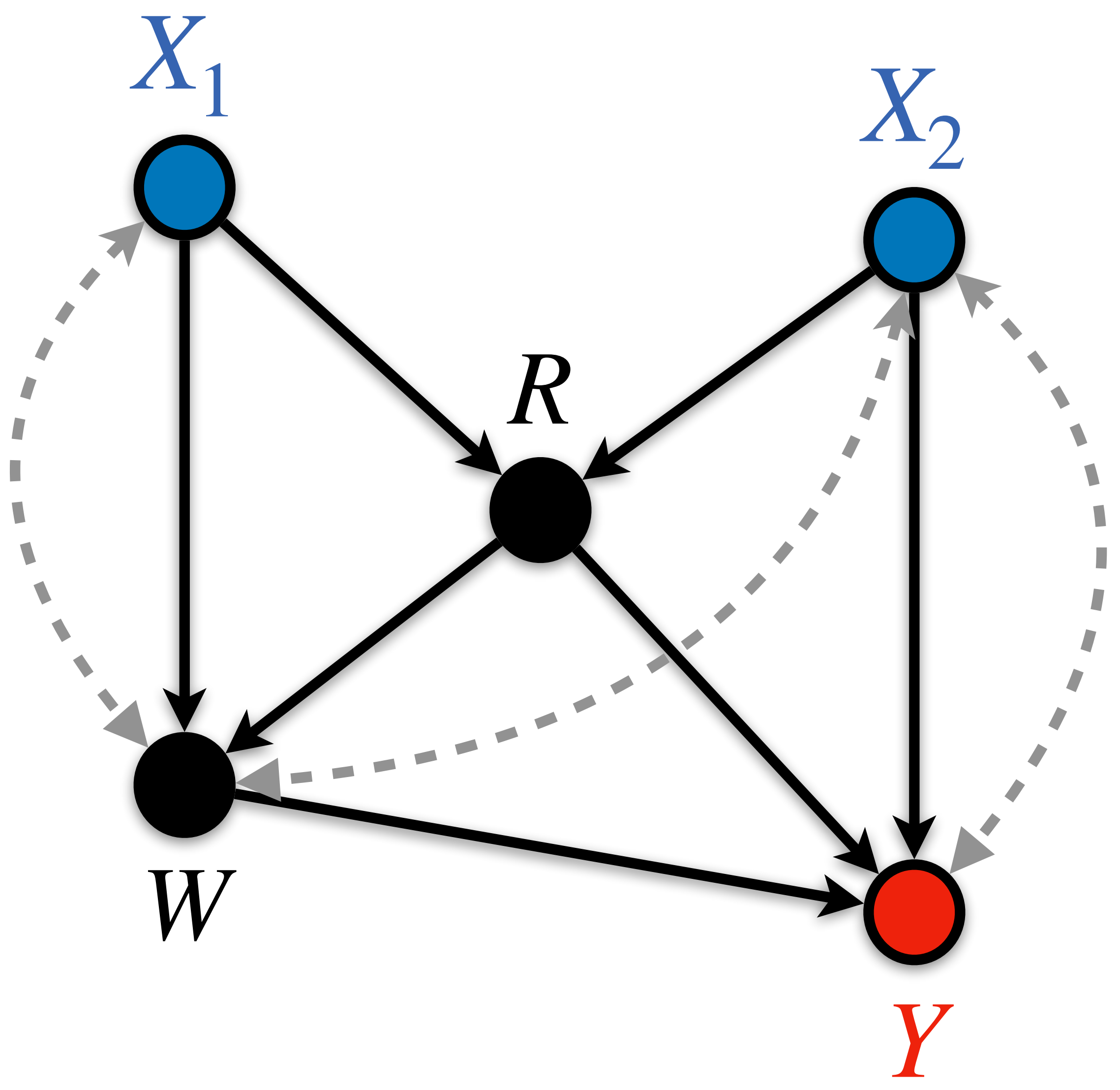
$P_{\mathrm{do}(x_2)}(RWX_1Y)$

General Identification (gID)

Bareinboim and Pearl, 2012; Lee et al. 2019

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

Background: General Identification from Data Fusion



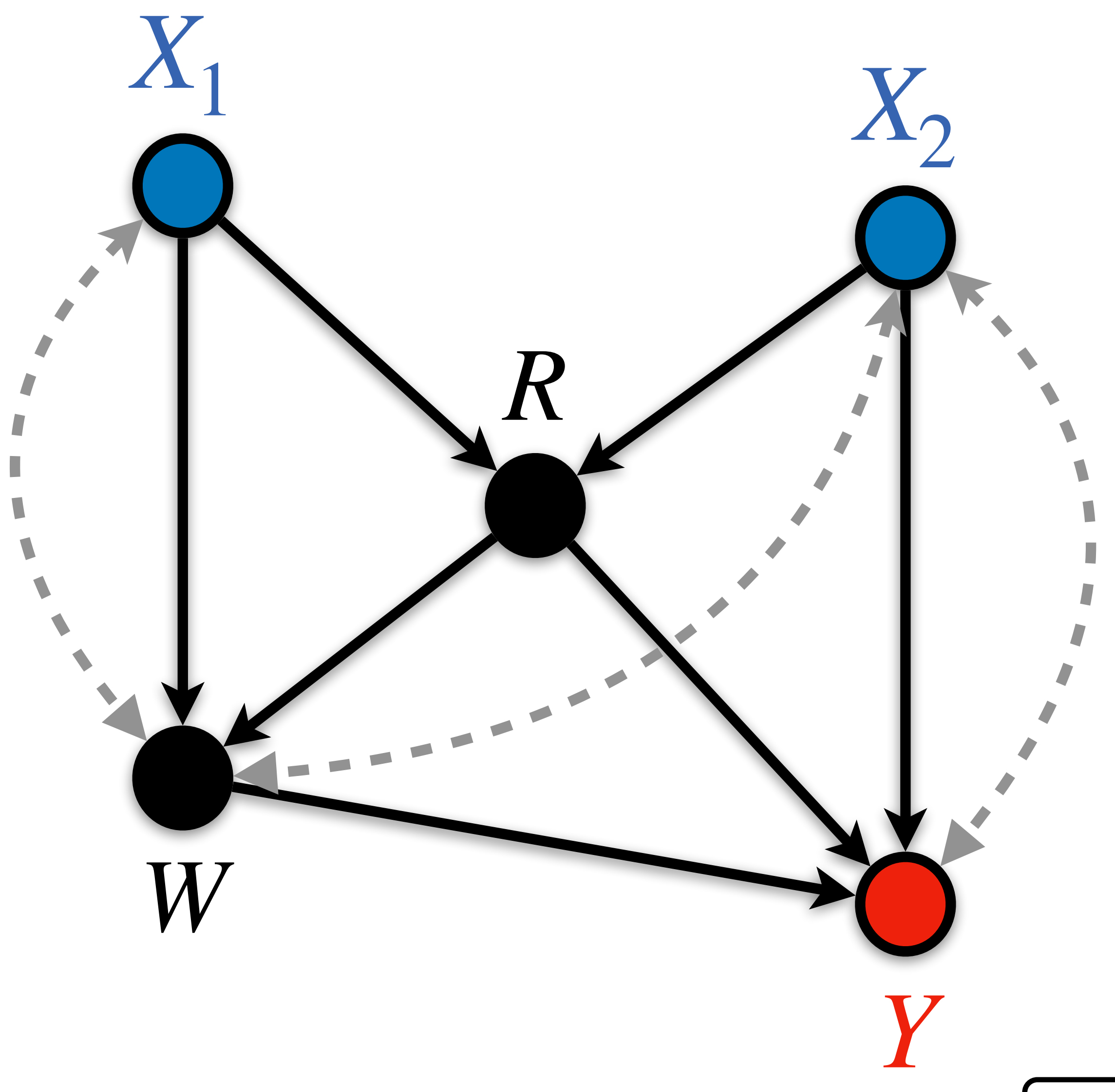
General Identification (gID)

Bareinboim and Pearl, 2012; Lee et al. 2019

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

$$\begin{tikzcd}
 & P_{\mathrm{do}(x_1,r)}(WX_2Y) \arrow[r, "\text{Marginalization}"] & P_{\mathrm{do}(x_1,r)}(W) \\
P_{\mathrm{do}(x_1)}(RWX_2Y) \arrow[ur, "\text{Factorization}"] \arrow[rr, "\text{Factorization}"] & & P_{\mathrm{do}(x_1,x_2)}(R) \\
P_{\mathrm{do}(x_2)}(RWX_1Y) \arrow[rr, "\text{Factorization}"'] & & P_{\mathrm{do}(w,r,x_2)}(Y)
\end{tikzcd}$$

Background: General Identification from Data Fusion



General Identification (gID)

Bareinboim and Pearl, 2012; Lee et al. 2019

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

$P_{\mathrm{do}(x_1)}(y \mid wrx_2)P_{\mathrm{do}(x_1)}(w \mid rx_2)P_{\mathrm{do}(x_1)}(x_2)$

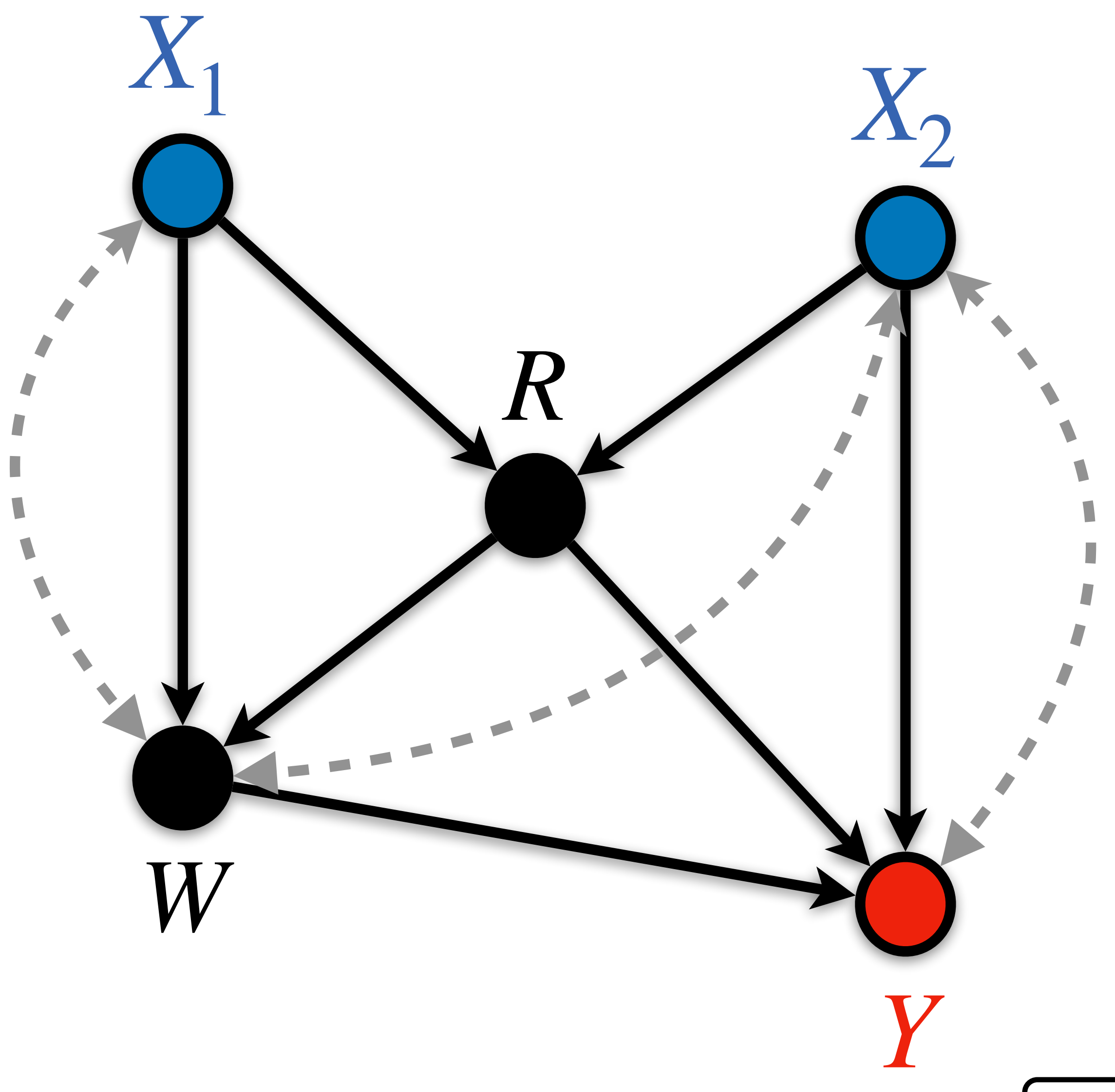
$\sum_{x_2'} P_{\mathrm{do}(x_1)}(w \mid r, x_2')P_{\mathrm{do}(x_1)}(x_2')$

$$\begin{tikzcd}
 & P_{\mathrm{do}(x_1,r)}(WX_2Y) \arrow[r, "\text{Marginalization}"] & P_{\mathrm{do}(x_1,r)}(W) \\
P_{\mathrm{do}(x_1)}(RWX_2Y) \arrow[ur, "\text{Factorization}"] \arrow[rr, "\text{Factorization}"] & & P_{\mathrm{do}(x_1,x_2)}(R) \\
P_{\mathrm{do}(x_2)}(RWX_1Y) \arrow[rr, "\text{Factorization}"'] & & P_{\mathrm{do}(w,r,x_2)}(Y)
\end{tikzcd}$$

$P_{\mathrm{do}(x_1)}(r \mid x_2)$

$P_{\mathrm{do}(x_2)}(y \mid rwx_1)$

Background: General Identification from Data Fusion



General Identification (gID)

Bareinboim and Pearl, 2012; Lee et al. 2019

- spanning a *tree* from available distributions $\{P_{\text{do}(\mathbf{r}_i)}(\mathbf{V})\}_{\mathbf{R}_i \subseteq \mathbf{V}}$
- to reach to causal distribution $P(\mathbf{Y} \mid \text{do}(\mathbf{X}))$
- through factorization & marginalization of distributions

$P_{\text{do}(x_1)}(y \mid wrx_2)P_{\text{do}(x_1)}(w \mid rx_2)P_{\text{do}(x_1)}(x_2)$

$\sum_{x_2'} P_{\text{do}(x_1)}(w \mid r, x_2')P_{\text{do}(x_1)}(x_2')$

$P_{\text{do}(x_1,r)}(WX_2Y)$ Marginalization $P_{\text{do}(x_1,r)}(W)$

Factorization

$P_{\text{do}(x_1)}(r \mid x_2)P_{\text{do}(x_2)}(y \mid rwx_1) \sum_{x_2'} P_{\text{do}(x_1)}(w \mid r, x_2')P_{\text{do}(x_1)}(x_2')$

Factorization

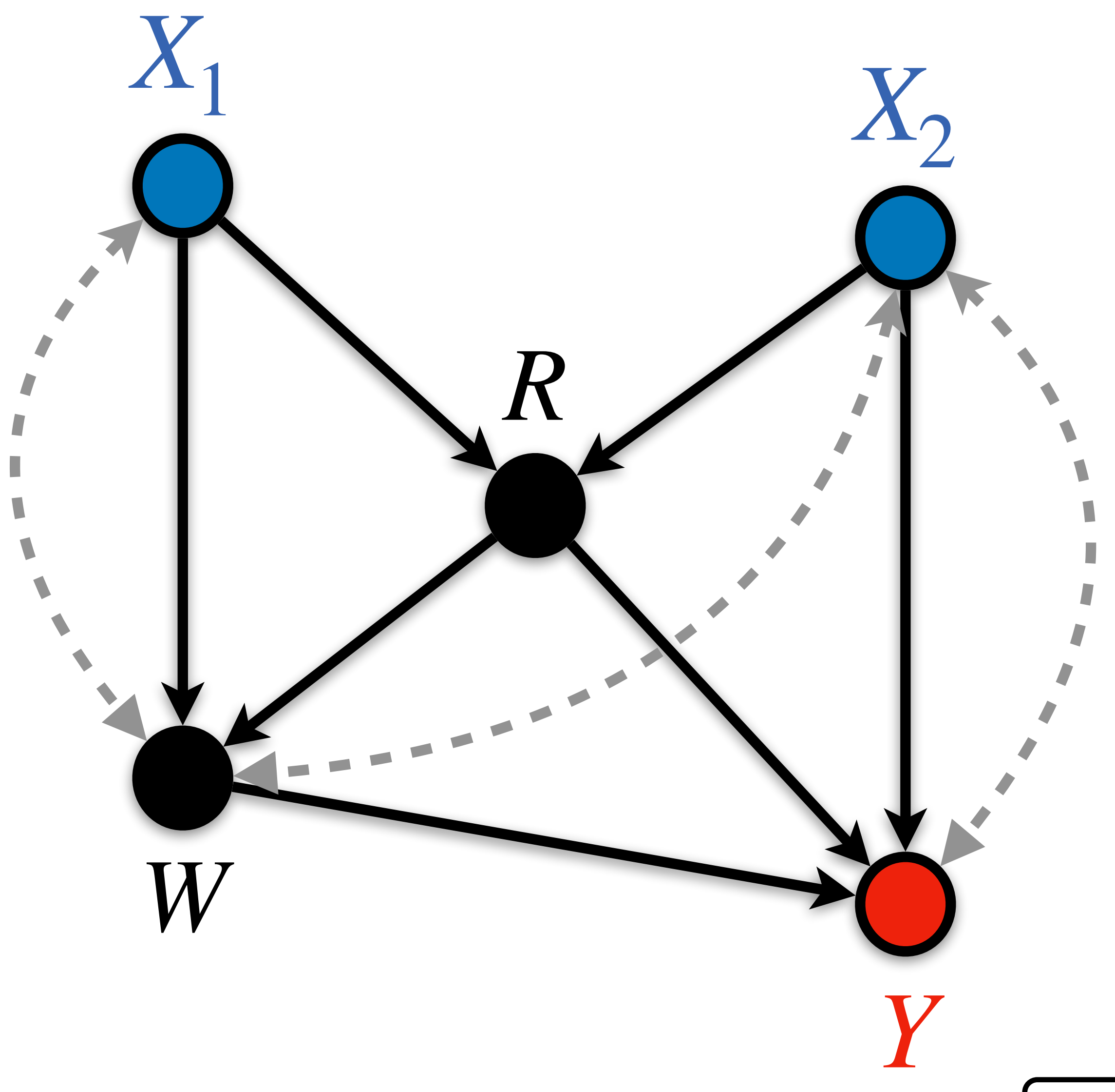
$P_{\text{do}(x_1)}(r \mid x_2)$

$P_{\text{do}(x_1)}(RWX_2Y)$ Factorization $P_{\text{do}(x_1,x_2)}(R)$ $P_{\text{do}(x_1,x_2)}(RWY)$

$P_{\text{do}(x_2)}(y \mid rwx_1)$

$P_{\text{do}(x_2)}(RWX_1Y)$ Factorization $P_{\text{do}(w,r,x_2)}(Y)$

Background: General Identification from Data Fusion



General Identification (gID)

Bareinboim and Pearl, 2012; Lee et al. 2019

- spanning a *tree* from available distributions $\{P_{\text{do}(\mathbf{r}_i)}(\mathbf{V})\}_{\mathbf{R}_i \subseteq \mathbf{V}}$
- to reach to causal distribution $P(\mathbf{Y} \mid \text{do}(\mathbf{X}))$
- through factorization & marginalization of distributions

$P_{\text{do}(x_1)}(y \mid wrx_2) P_{\text{do}(x_1)}(w \mid rx_2) P_{\text{do}(x_1)}(x_2)$

$\sum_{x_2'} P_{\text{do}(x_1)}(w \mid r, x_2') P_{\text{do}(x_1)}(x_2')$

Factorization

$P_{\text{do}(x_1,r)}(WX_2Y) \xrightarrow{\text{Marginalization}} P_{\text{do}(x_1,r)}(W)$

Factorization

$P_{\text{do}(x_1)}(r \mid x_2) P_{\text{do}(x_2)}(y \mid rwx_1) \sum_{x_2'} P_{\text{do}(x_1)}(w \mid r, x_2') P_{\text{do}(x_1)}(x_2')$

$P_{\text{do}(x_1)}(r \mid x_2)$

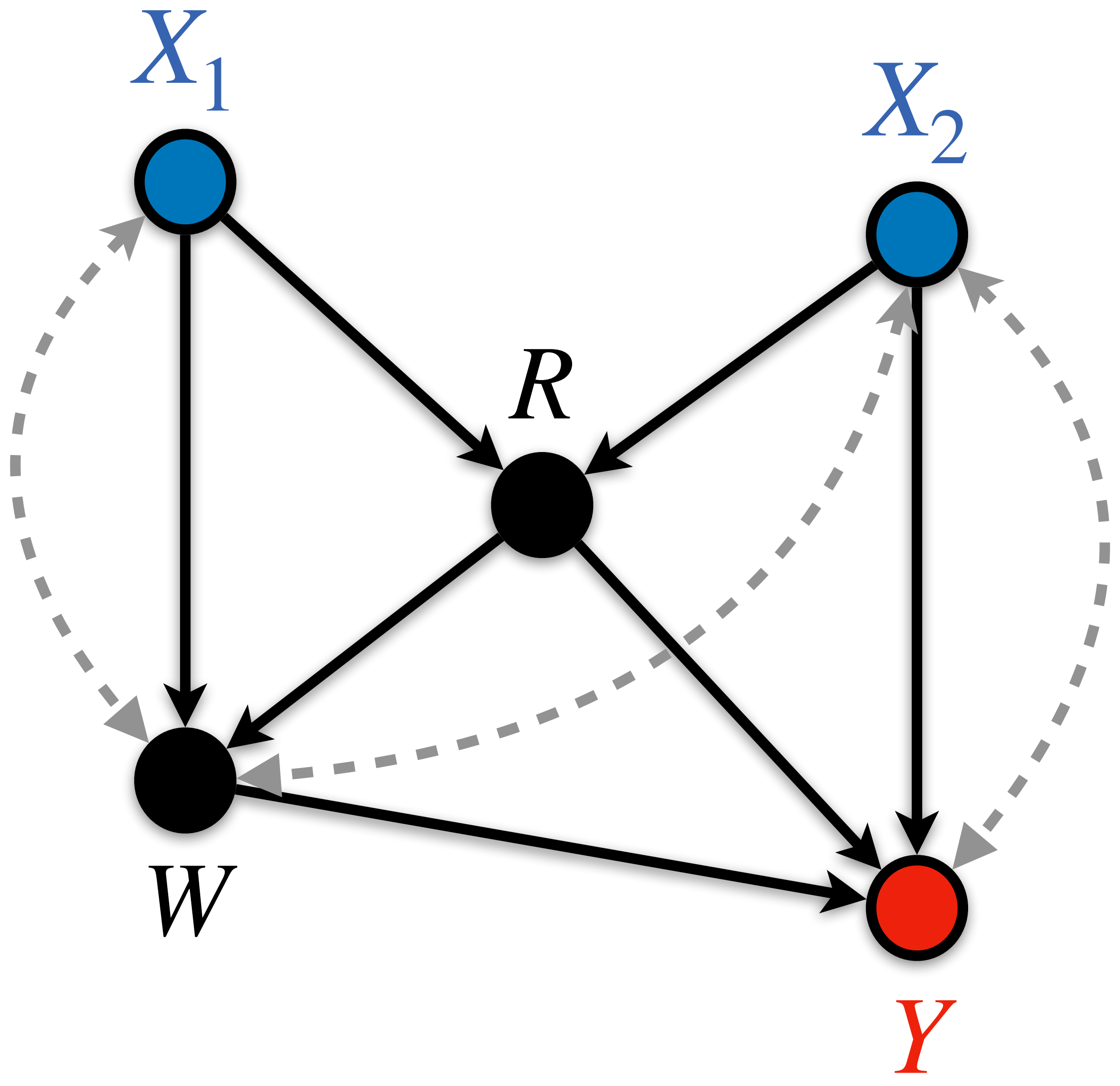
$P_{\text{do}(x_1)}(RWX_2Y) \xrightarrow{\text{Factorization}} P_{\text{do}(x_1,x_2)}(R) \rightarrow P_{\text{do}(x_1,x_2)}(RWY) \xrightarrow{\text{Marginalization}} P_{\text{do}(x_1,x_2)}(Y)$

$P_{\text{do}(x_2)}(y \mid rwx_1)$

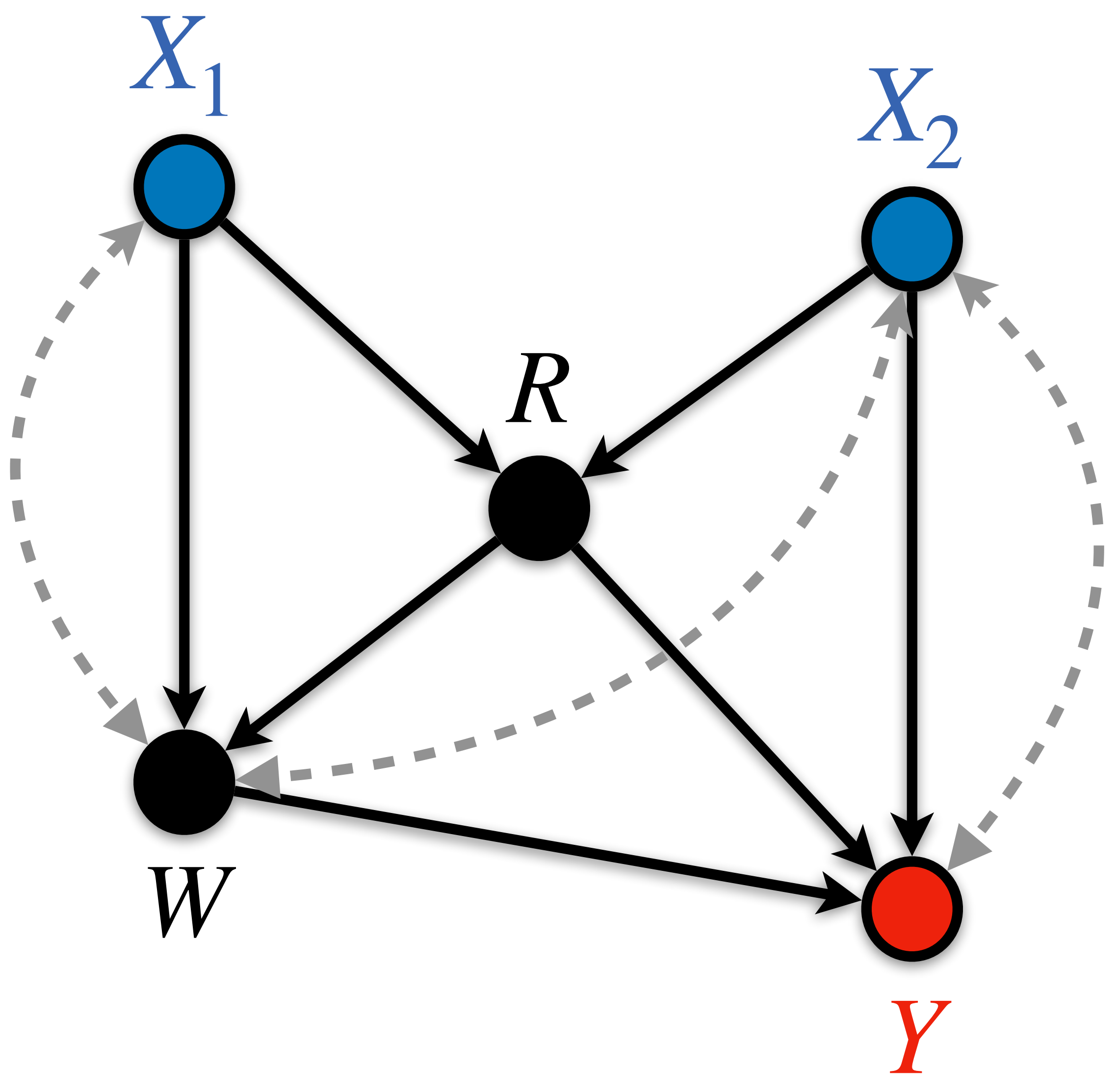
$P_{\text{do}(x_2)}(RWX_1Y) \xrightarrow[\text{Factorization}]{} P_{\text{do}(w,r,x_2)}(Y)$

$\sum_{rw} P_{\text{do}(x_1)}(r \mid x_2) P_{\text{do}(x_2)}(y \mid rwx_1) \sum_{x_2'} P_{\text{do}(x_1)}(w \mid r, x_2') P_{\text{do}(x_1)}(x_2')$

Causal effects as a function of BD$^+$

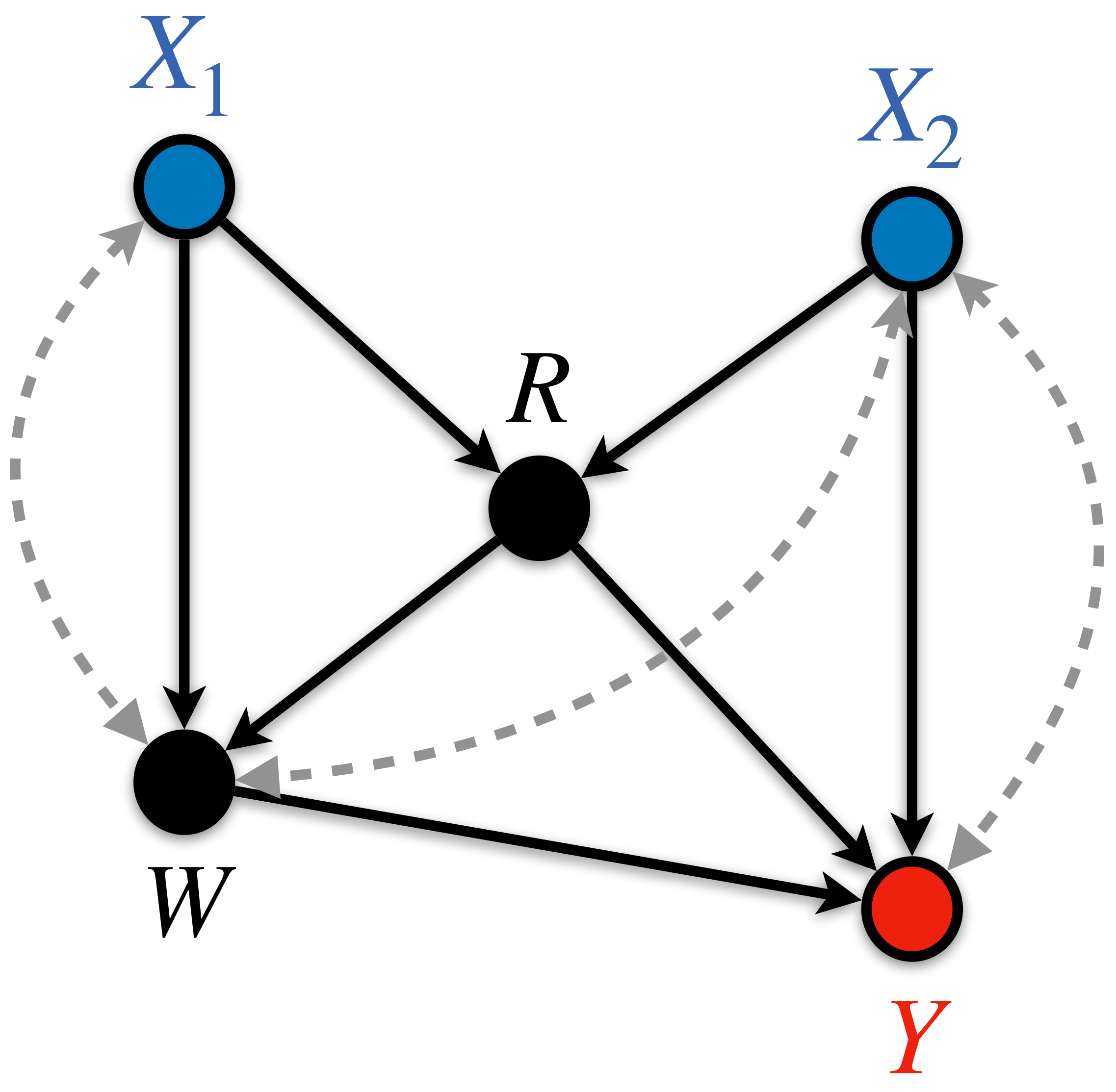


Causal effects as a function of BD+



$$\begin{tikzcd}
 & P_{\mathrm{do}(x_1,r)}(WX_2Y) \arrow[r] & P_{\mathrm{do}(x_1,r)}(W) \arrow[dr, "\text{Factorization}"] & & \\
P_{\mathrm{do}(x_1)}(RWX_2Y) \arrow[ur] \arrow[rr] & & P_{\mathrm{do}(x_1,x_2)}(R) \arrow[r] & P_{\mathrm{do}(x_1,x_2)}(RWY) \arrow[r, "\text{Marginalization}"] & P_{\mathrm{do}(x_1,x_2)}(Y) \\
P_{\mathrm{do}(x_2)}(RWX_1Y) \arrow[rr] & & P_{\mathrm{do}(w,r,x_2)}(Y) \arrow[ur] & &
\end{tikzcd}$$

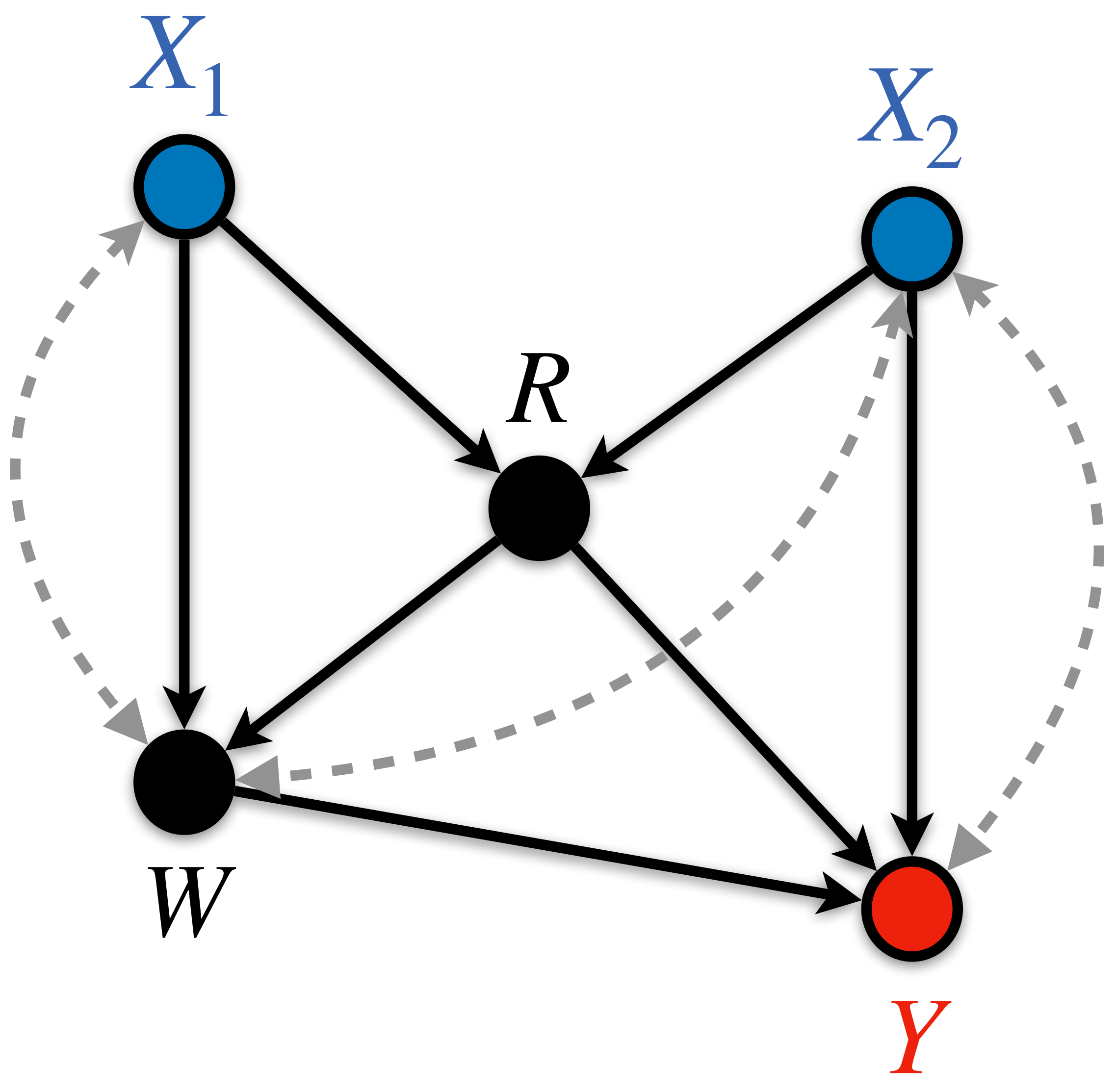
Causal effects as a function of BD+



$P_{\mathrm{do}(x_1 x_2)}(WR)$ is BD+-expressible from $\{P_{\mathrm{do}(x_i)}\}_{i=1,2}$

$$\begin{tikzcd}
 & P_{\mathrm{do}(x_1,r)}(WX_2Y) \arrow[r] & P_{\mathrm{do}(x_1,r)}(W) \arrow[dr, "\text{Factorization}"] & & \\
P_{\mathrm{do}(x_1)}(RWX_2Y) \arrow[ur] \arrow[rr] & & P_{\mathrm{do}(x_1,x_2)}(R) \arrow[r] & P_{\mathrm{do}(x_1,x_2)}(RWY) \arrow[r, "\text{Marginalization}"] & P_{\mathrm{do}(x_1,x_2)}(Y) \\
P_{\mathrm{do}(x_2)}(RWX_1Y) \arrow[rr] & & P_{\mathrm{do}(w,r,x_2)}(Y) \arrow[ur] & &
\end{tikzcd}$$

Causal effects as a function of BD+

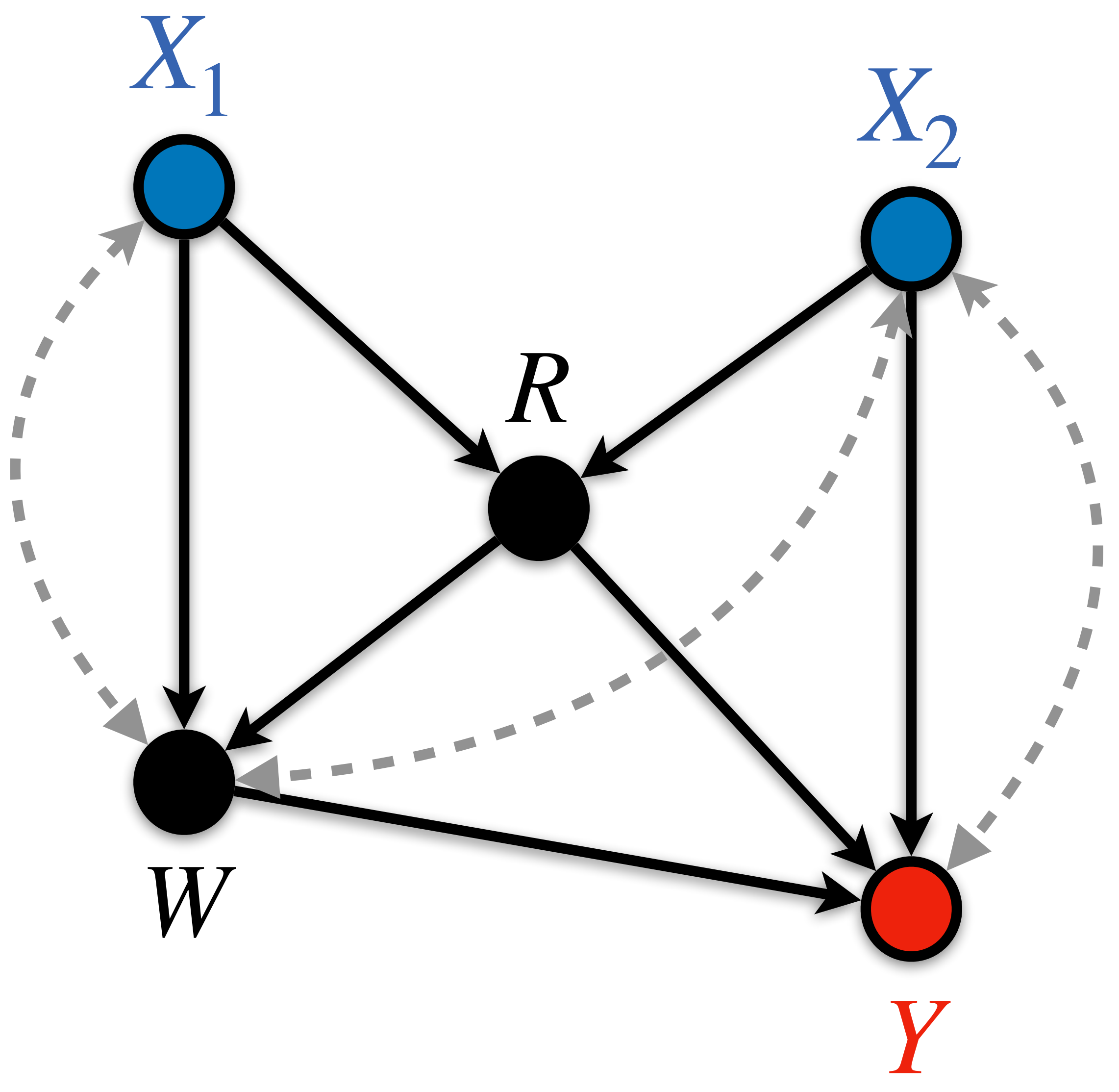


$P_{\mathrm{do}(x_1x_2)}(WR)$ is BD+-expressible from $\{P_{\mathrm{do}(x_i)}\}_{i=1,2}$

$$\begin{tikzcd}
 & P_{\mathrm{do}(x_1,r)}(WX_2Y) \arrow[r] & P_{\mathrm{do}(x_1,r)}(W) \arrow[dr, "\text{Factorization}"] & & \\
P_{\mathrm{do}(x_1)}(RWX_2Y) \arrow[ur] \arrow[rr] & & P_{\mathrm{do}(x_1,x_2)}(R) \arrow[r] & P_{\mathrm{do}(x_1,x_2)}(RWY) \arrow[r, "\text{Marginalization}"] & P_{\mathrm{do}(x_1,x_2)}(Y) \\
P_{\mathrm{do}(x_2)}(RWX_1Y) \arrow[rr] & & P_{\mathrm{do}(w,r,x_2)}(Y) \arrow[ur] & &
\end{tikzcd}$$

BD+-expressible from $P_{\mathrm{do}(x_2)}$

Causal effects as a function of BD+



$P_{\mathrm{do}(x_1 x_2)}(WR)$ is BD+-expressible from $\{P_{\mathrm{do}(x_i)}\}_{i=1,2}$

$P_{\mathrm{do}(x_1 x_2)}(Y)$ is a function of BD+

$$\begin{tikzcd}
 & P_{\mathrm{do}(x_1,r)}(WX_2Y) \arrow[r] & \boxed{P_{\mathrm{do}(x_1,r)}(W)} \arrow[dr, "\text{Factorization}"] & & \\
P_{\mathrm{do}(x_1)}(RWX_2Y) \arrow[ur] \arrow[rr] & & \boxed{P_{\mathrm{do}(x_1,x_2)}(R)} \arrow[r] & P_{\mathrm{do}(x_1,x_2)}(RWY) \arrow[r, "\text{Marginalization}"] & \boxed{P_{\mathrm{do}(x_1,x_2)}(Y)} \\
P_{\mathrm{do}(x_2)}(RWX_1Y) \arrow[rr] & & \boxed{P_{\mathrm{do}(w,r,x_2)}(Y)} \arrow[ur] & &
\end{tikzcd}$$

BD+-expressible from $P_{\mathrm{do}(x_2)}$

Causal effects as a function of BD+

Theorem

The followings are equivalent:

1. $P(\mathbf{y} \mid \text{do}(\mathbf{x}))$ is identifiable from $(\mathcal{G}, \{P_{\text{do}(\mathbf{r}_i)}\})$
2. $P(\mathbf{y} \mid \text{do}(\mathbf{x}))$ is expressible as a ***function of BD+s*** through AdmissibleGID (**Jung** et al., NeurIPS23)

DML-gID: Estimator for Causal Effects from Fusion

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

DML-gID: Estimator for Causal Effects from Fusion

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

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

"DML-gID"

DML-gID: Estimator for Causal Effects from Fusion

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

DML-BD$^+$ DML-BD$^+$ ⋯ DML-BD$^+$

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

“DML-gID”

Robustness of DML-gID

Theorem

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

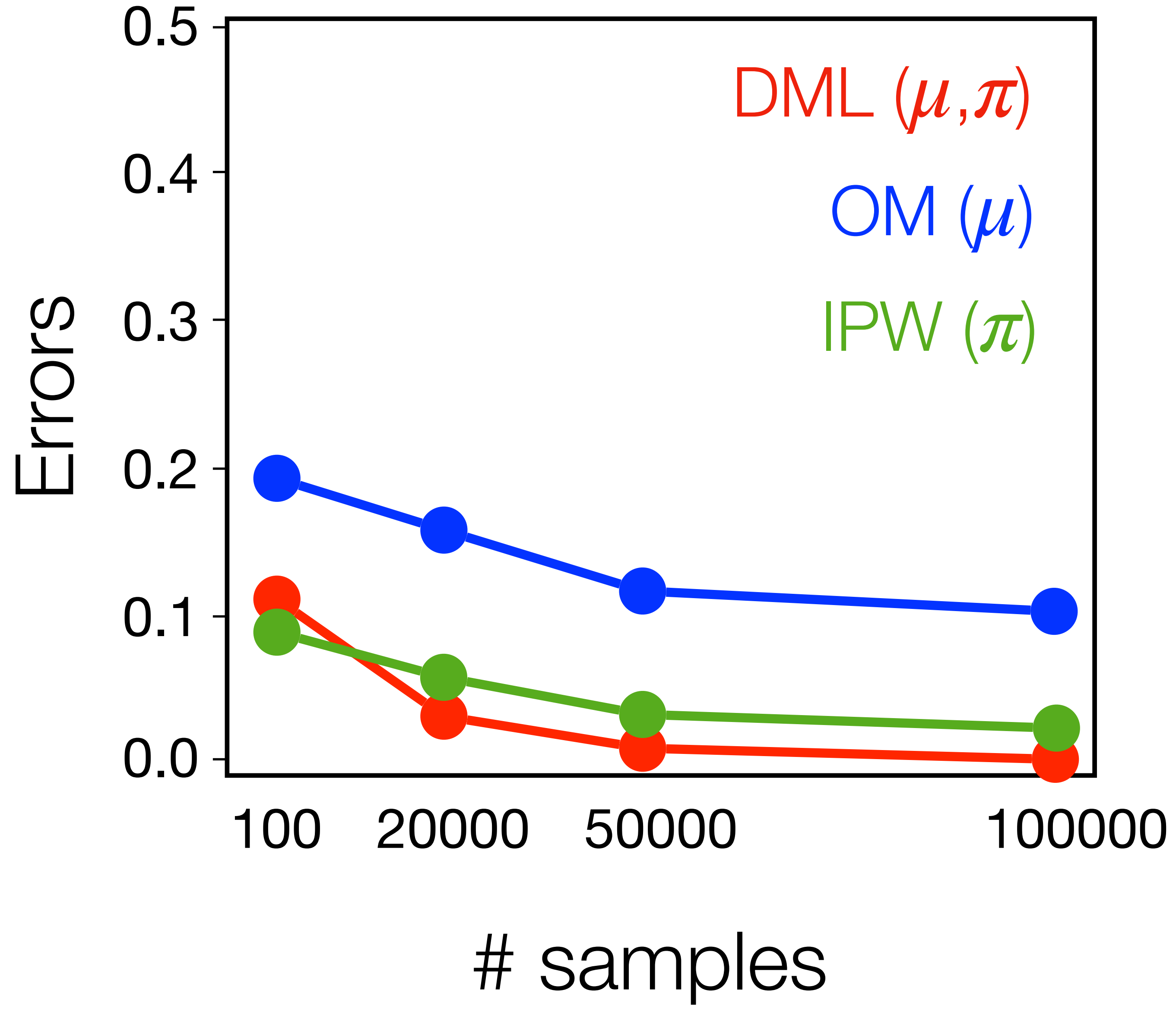
- **Double Robustness:** Error $= 0$ if either $\hat{\mu}_i = \mu_i$ or $\hat{\pi}_i = \pi_i$ for all $i = 1, \cdots, m$.
- **Fast Convergence:** Error $\rightarrow 0$ *fast* even when $\hat{\mu}_i \rightarrow \mu_i$ and $\hat{\pi}_i \rightarrow \pi_i$ *slow*.

DML-gID - Simulation

DML-gID - Simulation

Fast Convergence

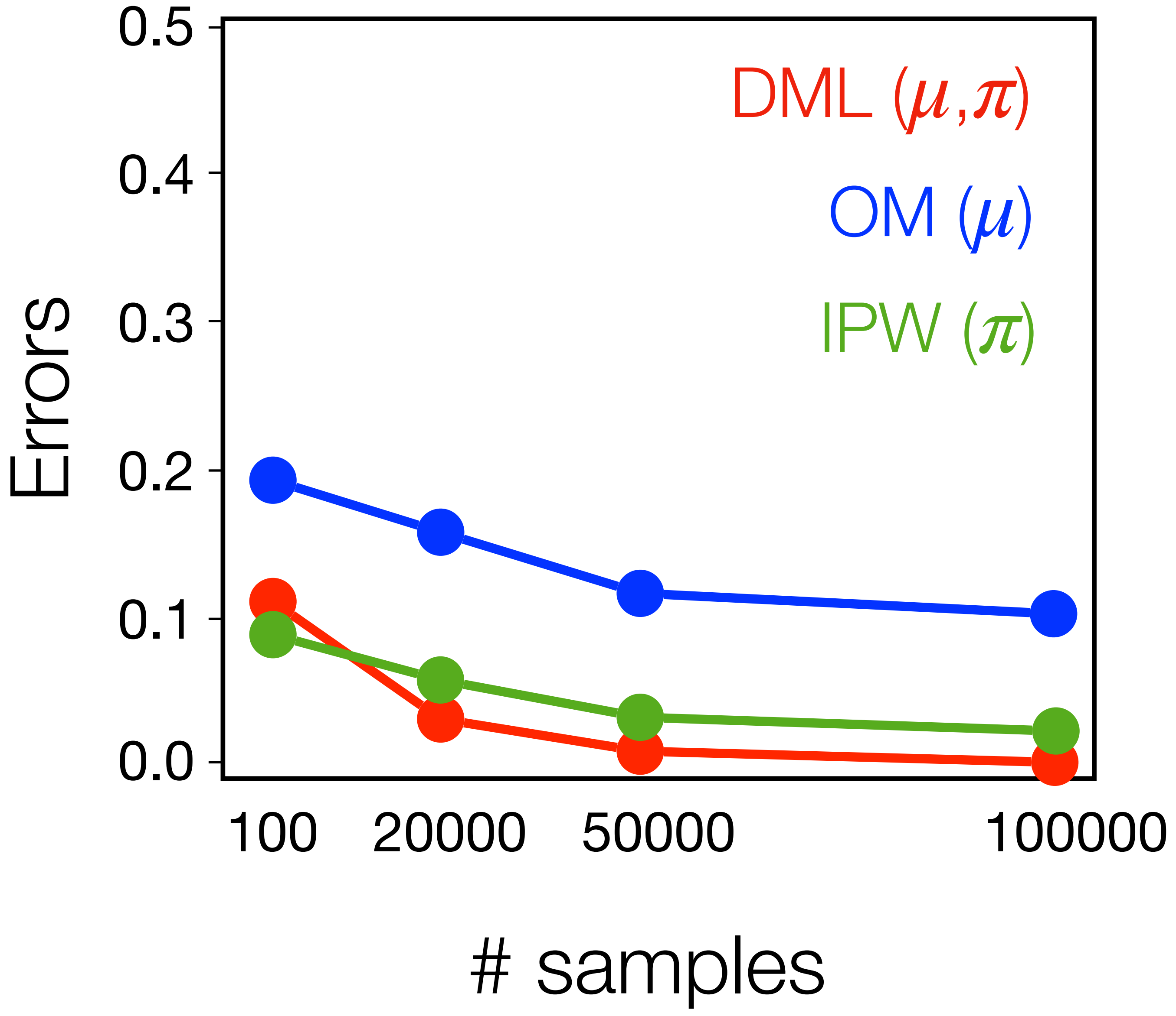
$(\hat{\boldsymbol{\mu}}, \hat{\boldsymbol{\pi}}) \rightarrow (\boldsymbol{\mu}_0, \boldsymbol{\pi}_0)$ slowly



DML-gID - Simulation

Fast Convergence

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

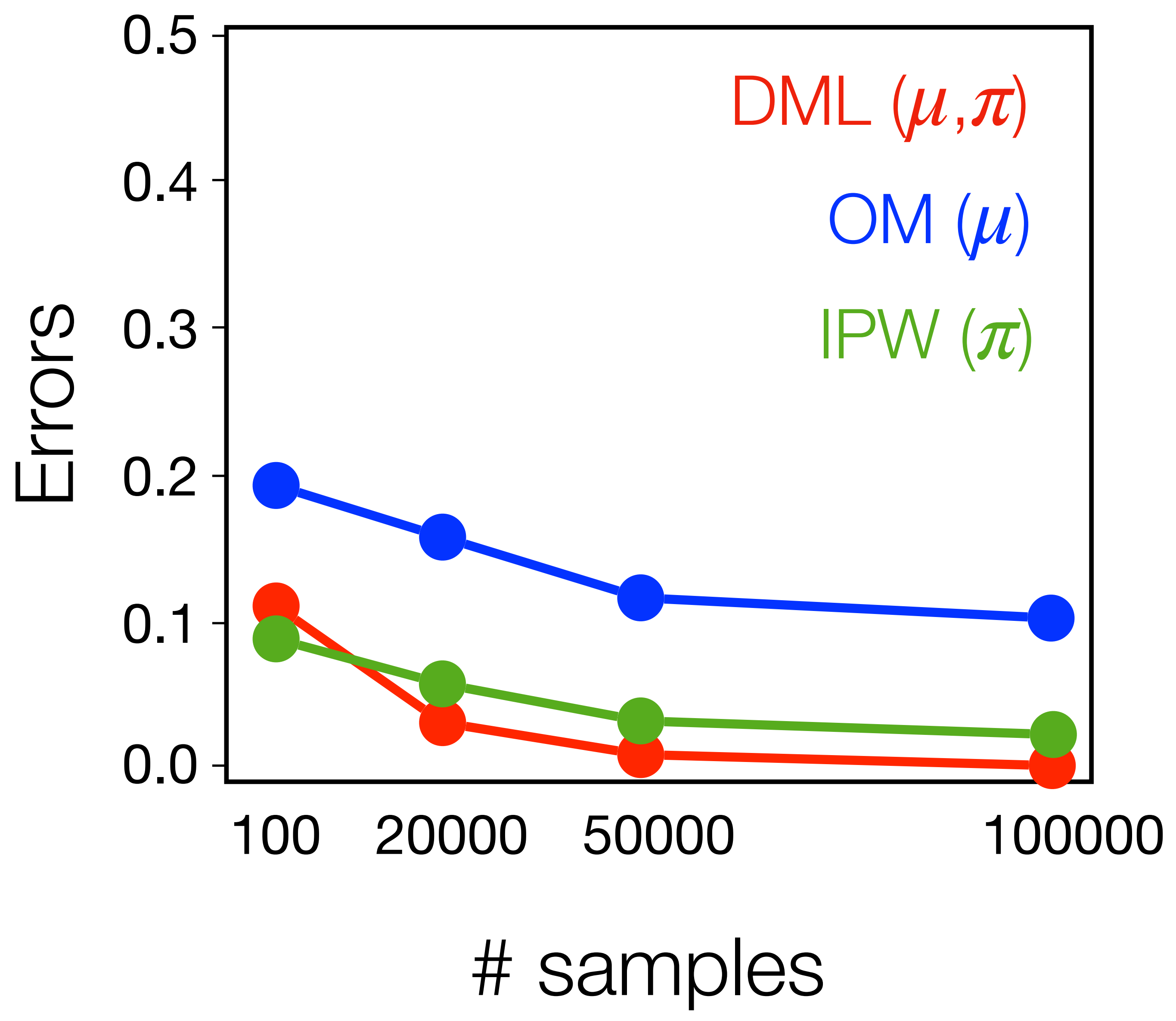


DML-gID converges fast, even when $(\hat{\boldsymbol{\mu}}, \hat{\boldsymbol{\pi}})$ converge slowly

DML-gID - Simulation

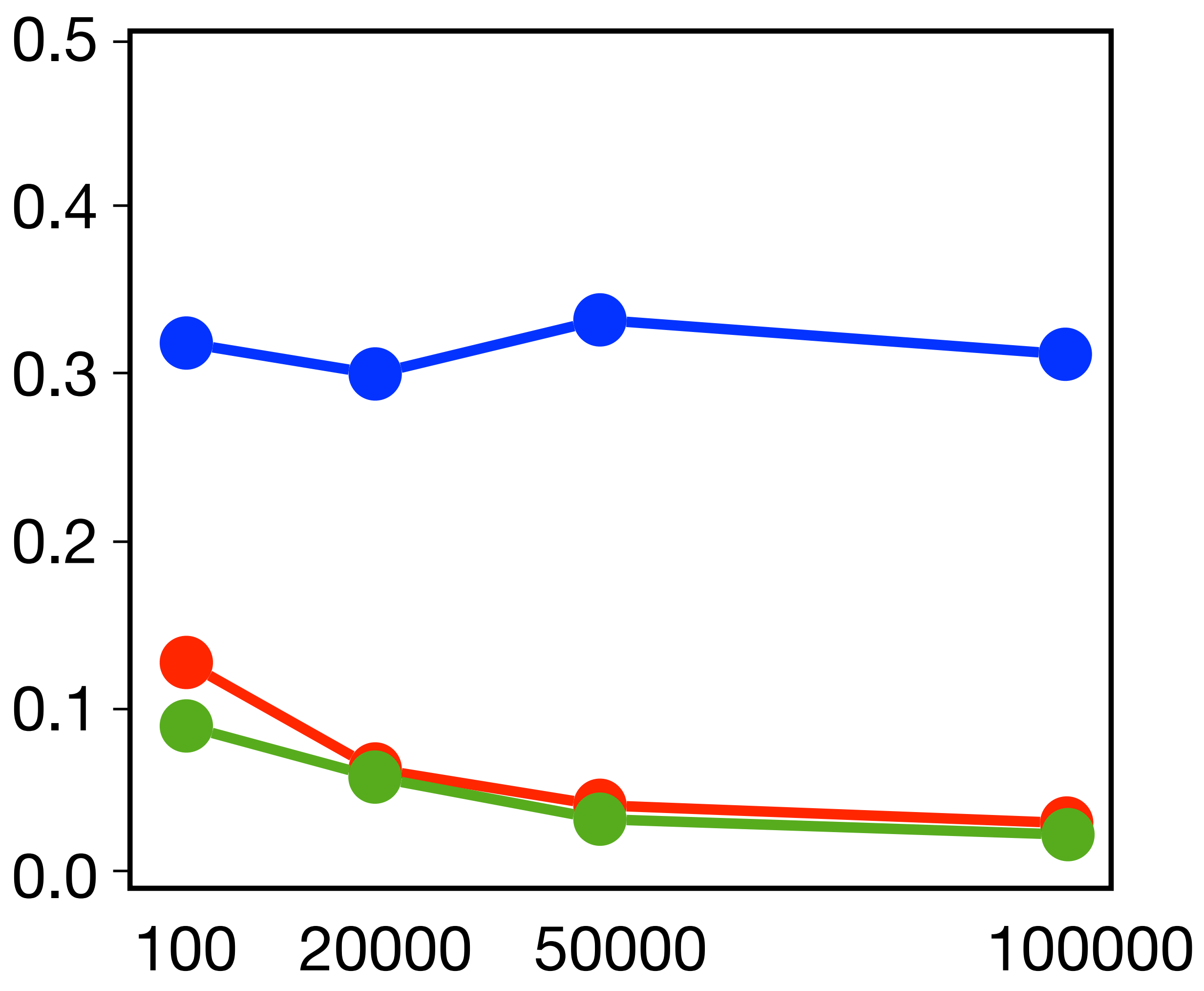
Fast Convergence

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

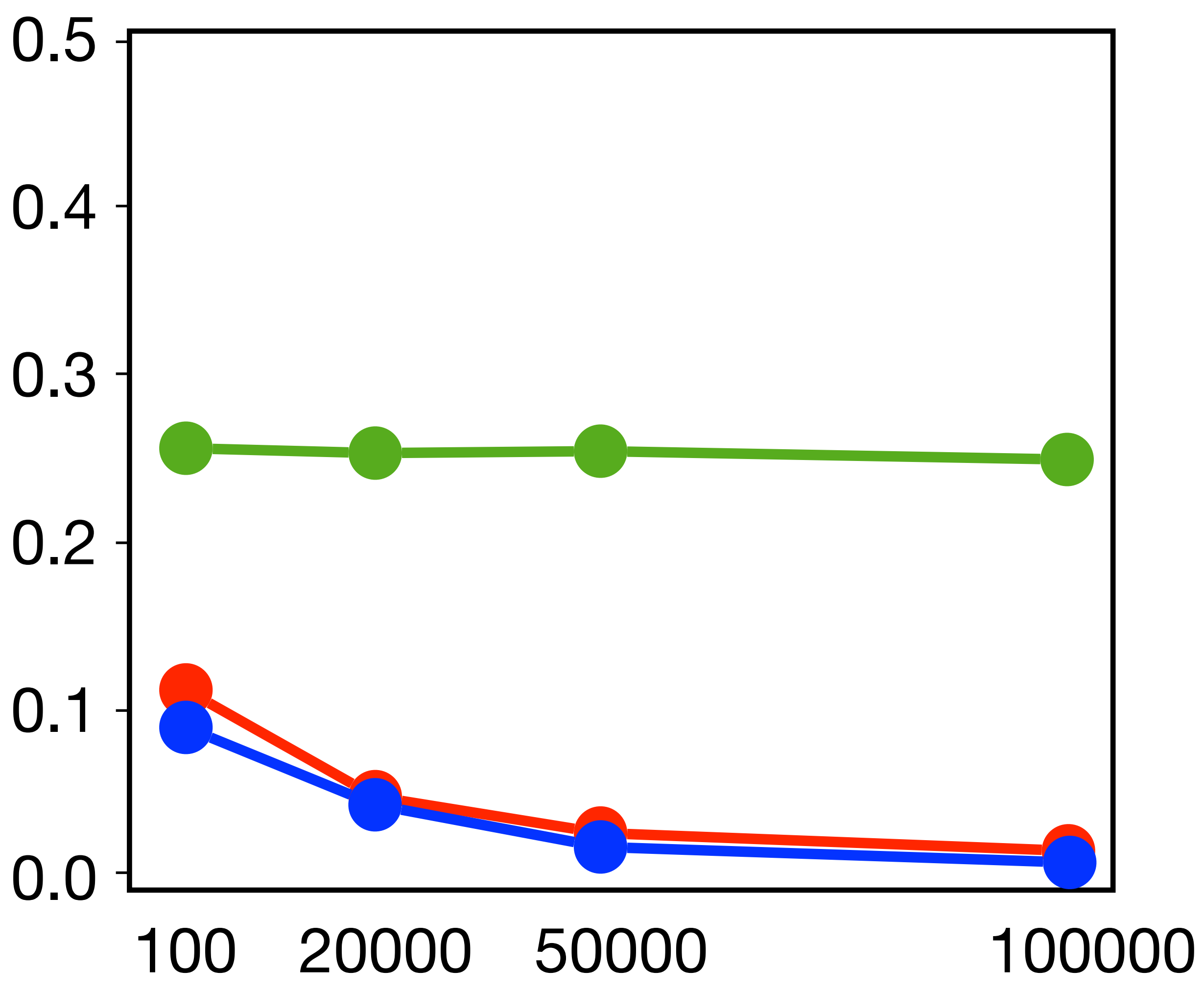


Double Robustness

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



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

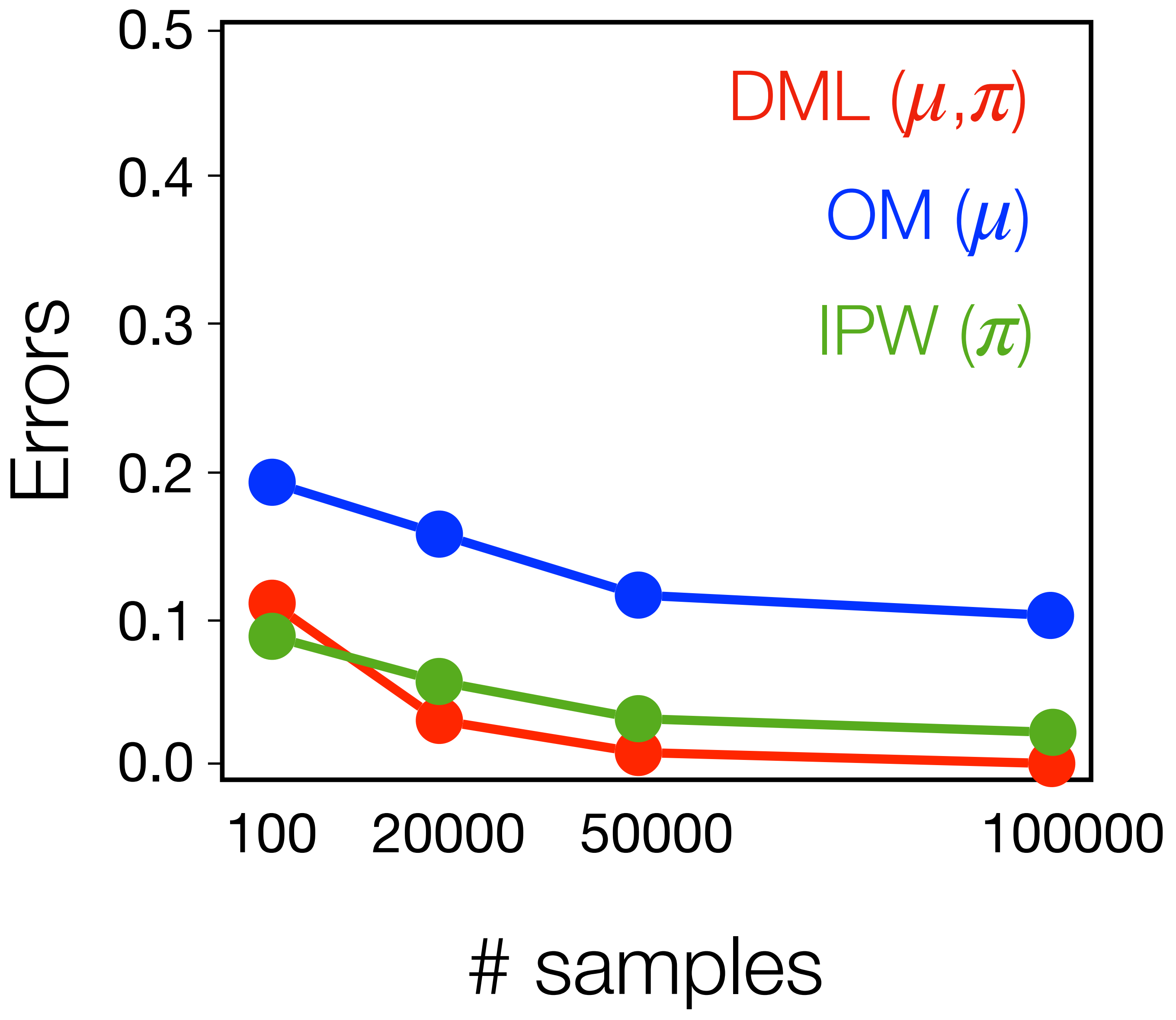


DML-gID converges fast, even when $(\hat{\boldsymbol{\mu}},\hat{\boldsymbol{\pi}})$ converge slowly

DML-gID - Simulation

Fast Convergence

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

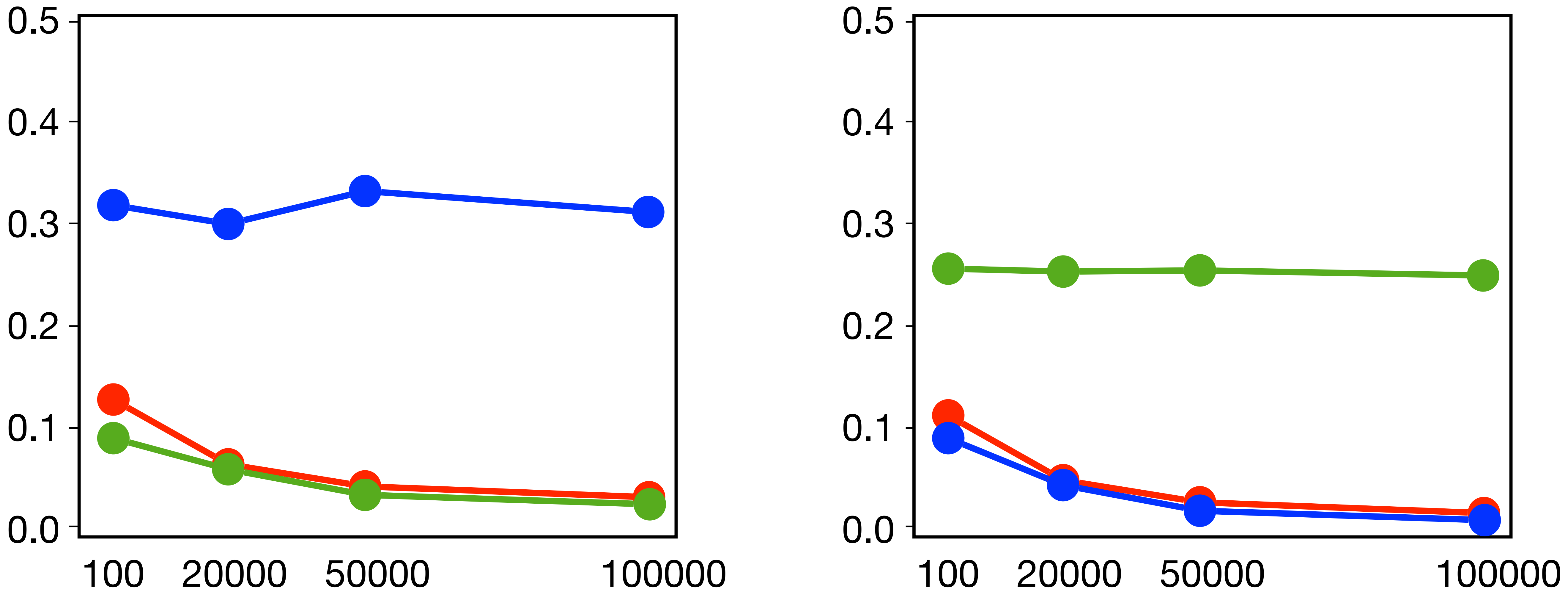


DML-gID converges fast, even when $(\hat{\boldsymbol{\mu}},\hat{\boldsymbol{\pi}})$ converge slowly

Double Robustness

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

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



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

Talk Outline

❶ Estimating causal effects from observations

❷ Estimating causal effects from data fusion

❸ Unified and scalable estimation method

❹ Conclusion

Talk Outline

✓

➡ ❸ Unified and scalable estimation method

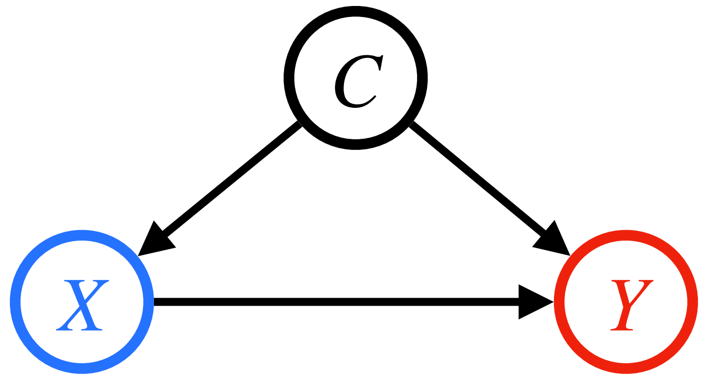
Motivation: Multilinear Causal Estimands

A causal effect $\mathbb{E}[Y \mid \mathrm{do}(\mathbf{x})]$ is often identified as a multilinear functional.

Motivation: Multilinear Causal Estimands

A causal effect $\mathbb{E}[Y \mid \mathrm{do}(\mathbf{x})]$ is often identified as a multilinear functional.

Back-Door (BD)

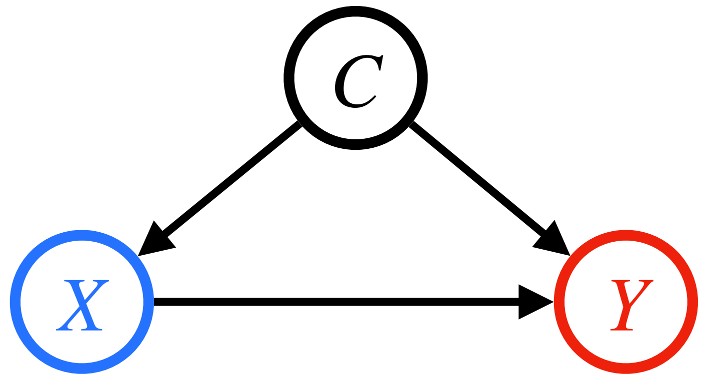


$$\sum_c \mathbb{E}[Y \mid x, c]P(c)$$

Motivation: Multilinear Causal Estimands

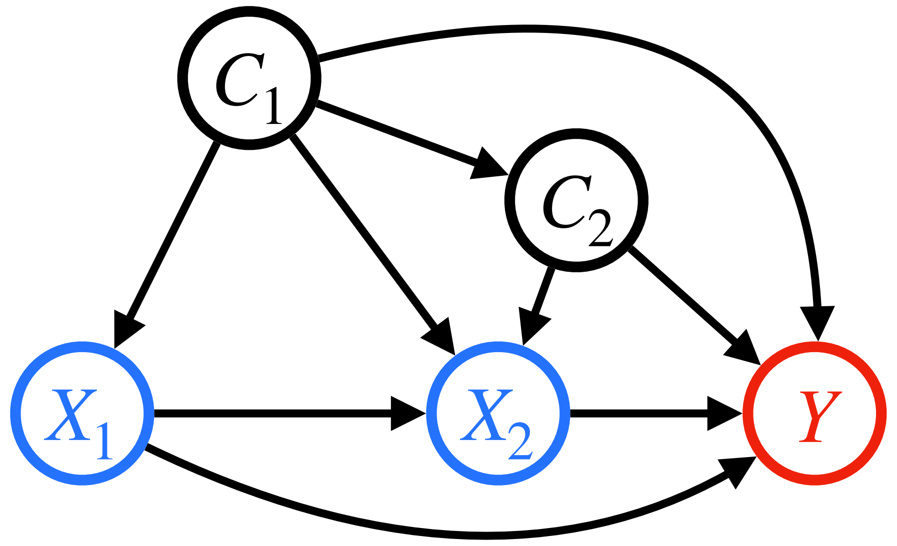
A causal effect $\mathbb{E}[Y \mid \mathrm{do}(\mathbf{x})]$ is often identified as a multilinear functional.

Back-Door (BD)



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

Sequential BD (SBD)

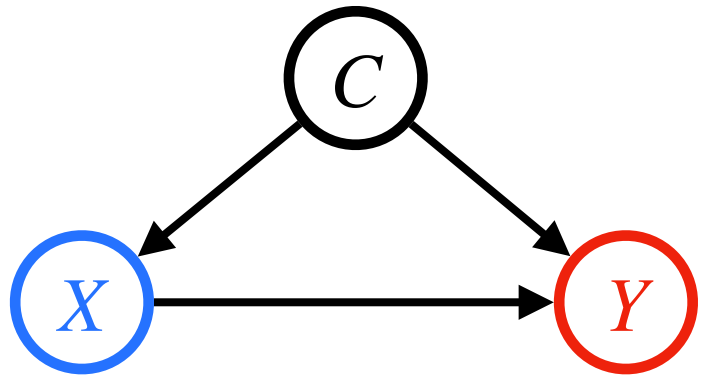


$$\sum_{z_1,z_2} \mathbb{E}[Y \mid x_2, z_2, x_1, z_1] P(z_2 \mid x_1, z_1) P(z_1)$$

Motivation: Multilinear Causal Estimands

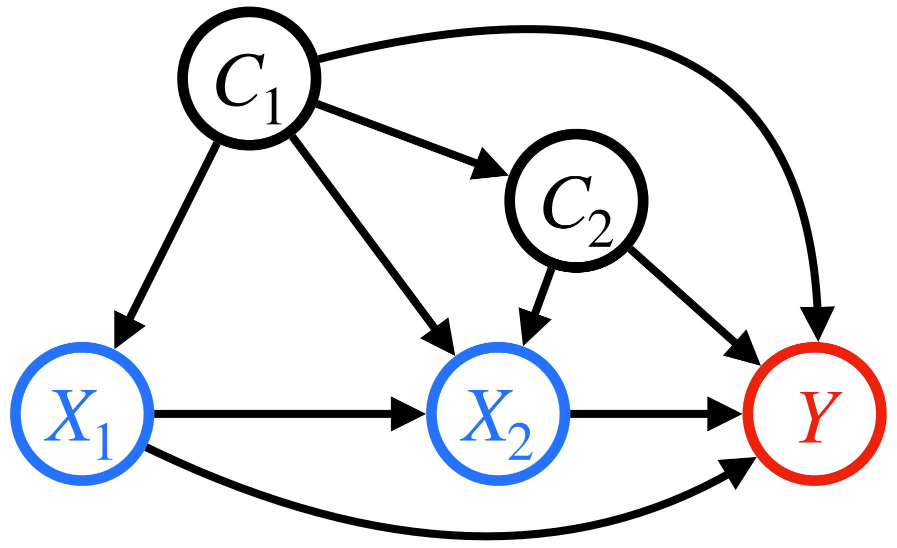
A causal effect $\mathbb{E}[Y \mid \mathrm{do}(\mathbf{x})]$ is often identified as a multilinear functional.

Back-Door (BD)



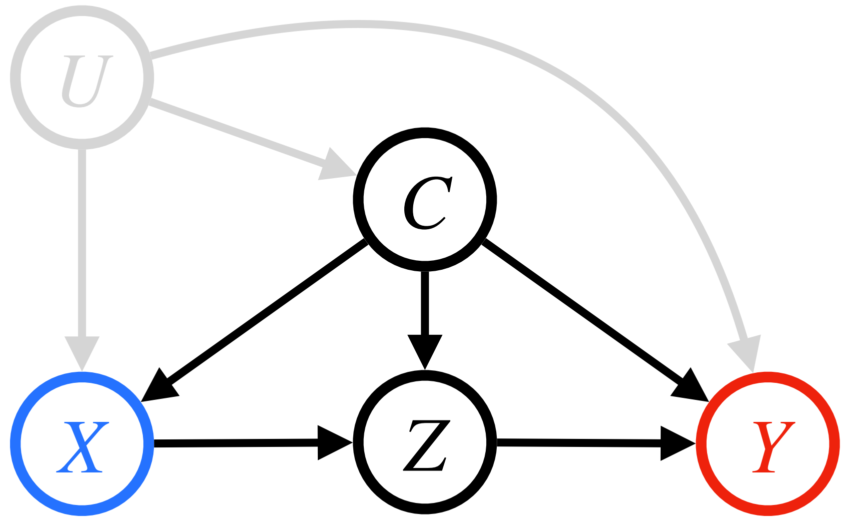
$$\sum_c \mathbb{E}[Y \mid x, c] P(c)$$

Sequential BD (SBD)



$$\sum_{z_1, z_2} \mathbb{E}[Y \mid x_2, z_2, x_1, z_1] P(z_2 \mid x_1, z_1) P(z_1)$$

Front-Door (FD)

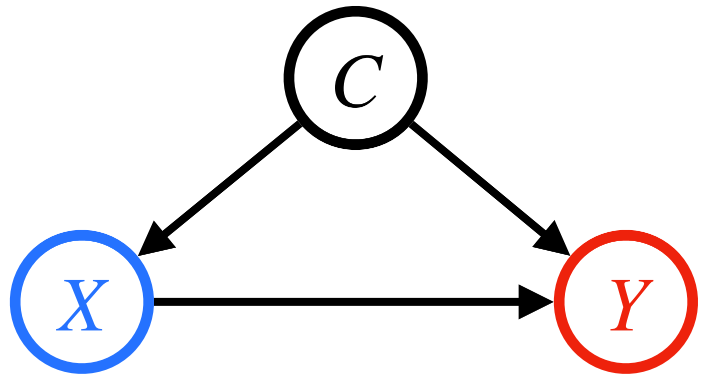


$$\sum_{z, x', c} \mathbb{E}[Y \mid z, x', c] P(z \mid x', c) P(x' c)$$

Motivation: Multilinear Causal Estimands

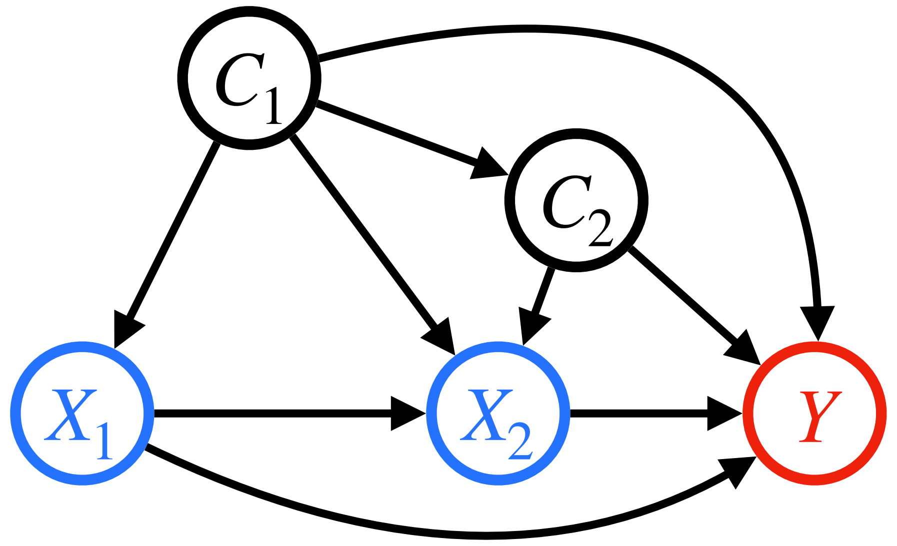
A causal effect $\mathbb{E}[Y \mid \mathrm{do}(\mathbf{x})]$ is often identified as a multilinear functional.

Back-Door (BD)



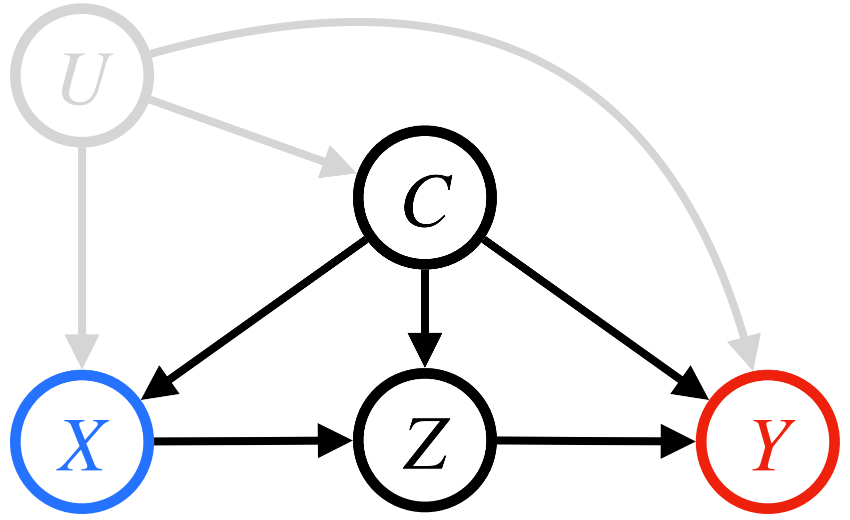
$$\sum_c \mathbb{E}[Y \mid x, c]P(c)$$

Sequential BD (SBD)



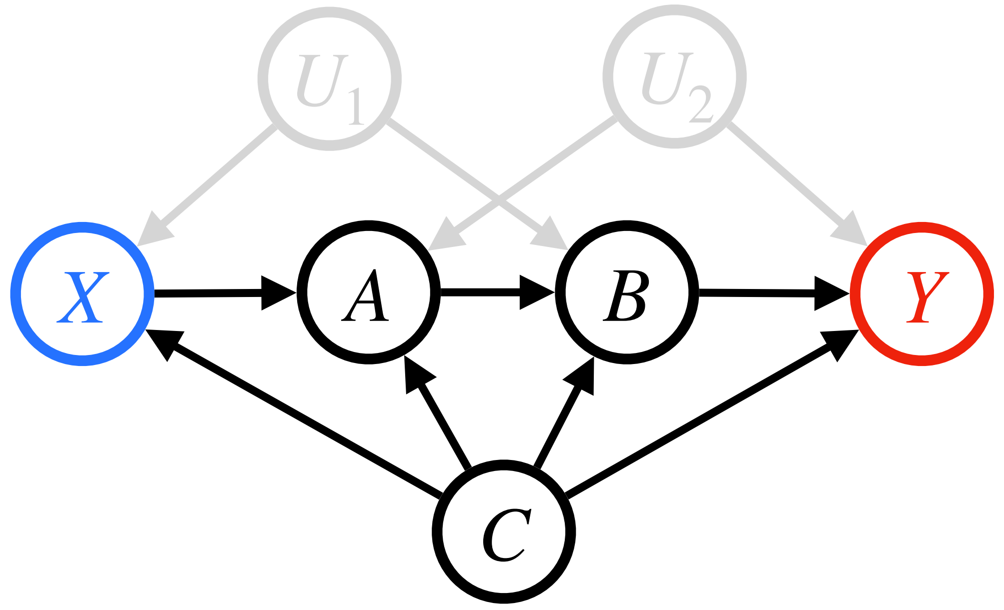
$$\sum_{z_1,z_2} \mathbb{E}[Y \mid x_2, z_2, x_1, z_1]P(z_2 \mid x_1, z_1)P(z_1)$$

Front-Door (FD)



$$\sum_{z,x',c} \mathbb{E}[Y \mid z, x', c]P(z \mid x', c)P(x'c)$$

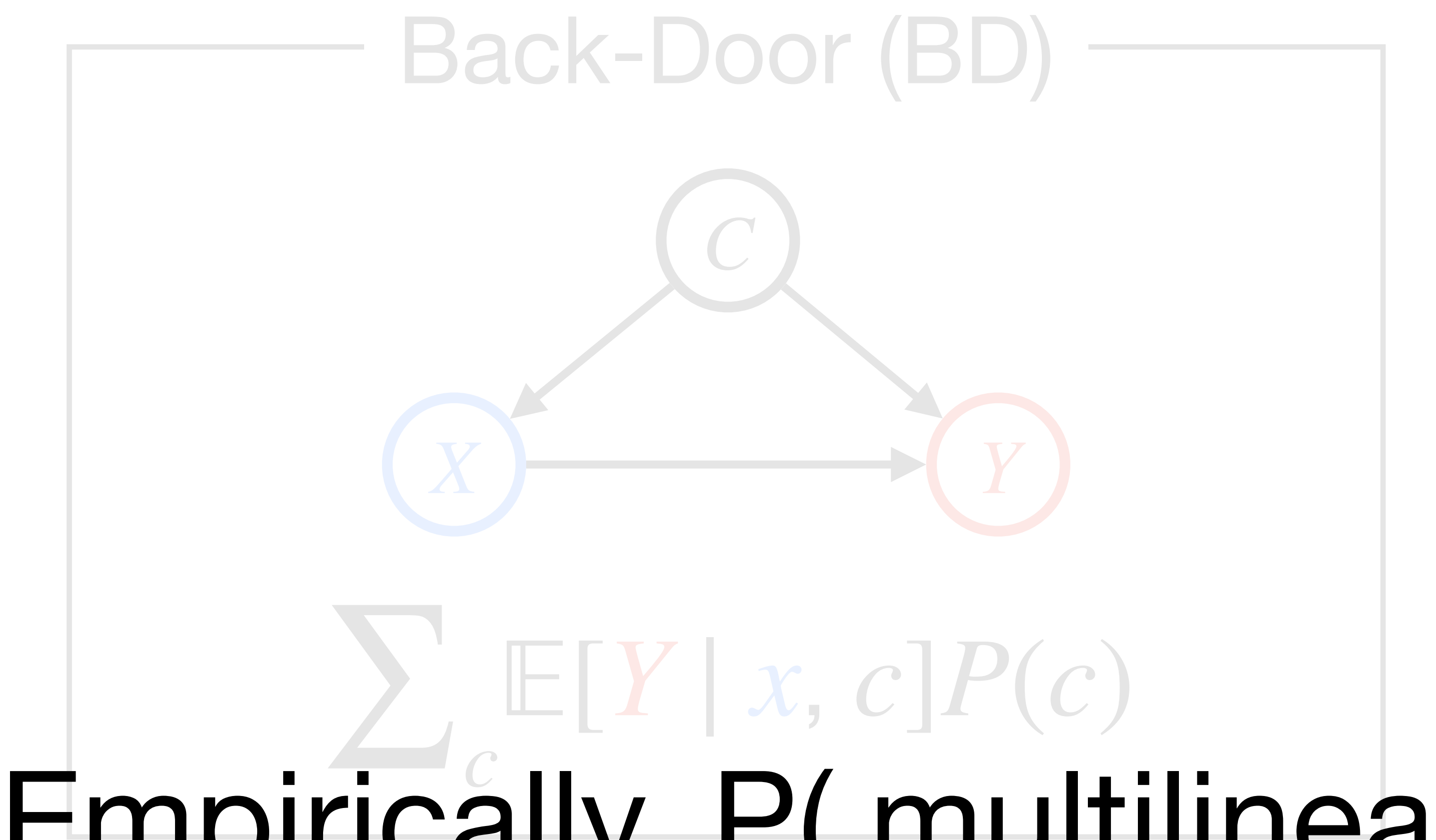
Verma

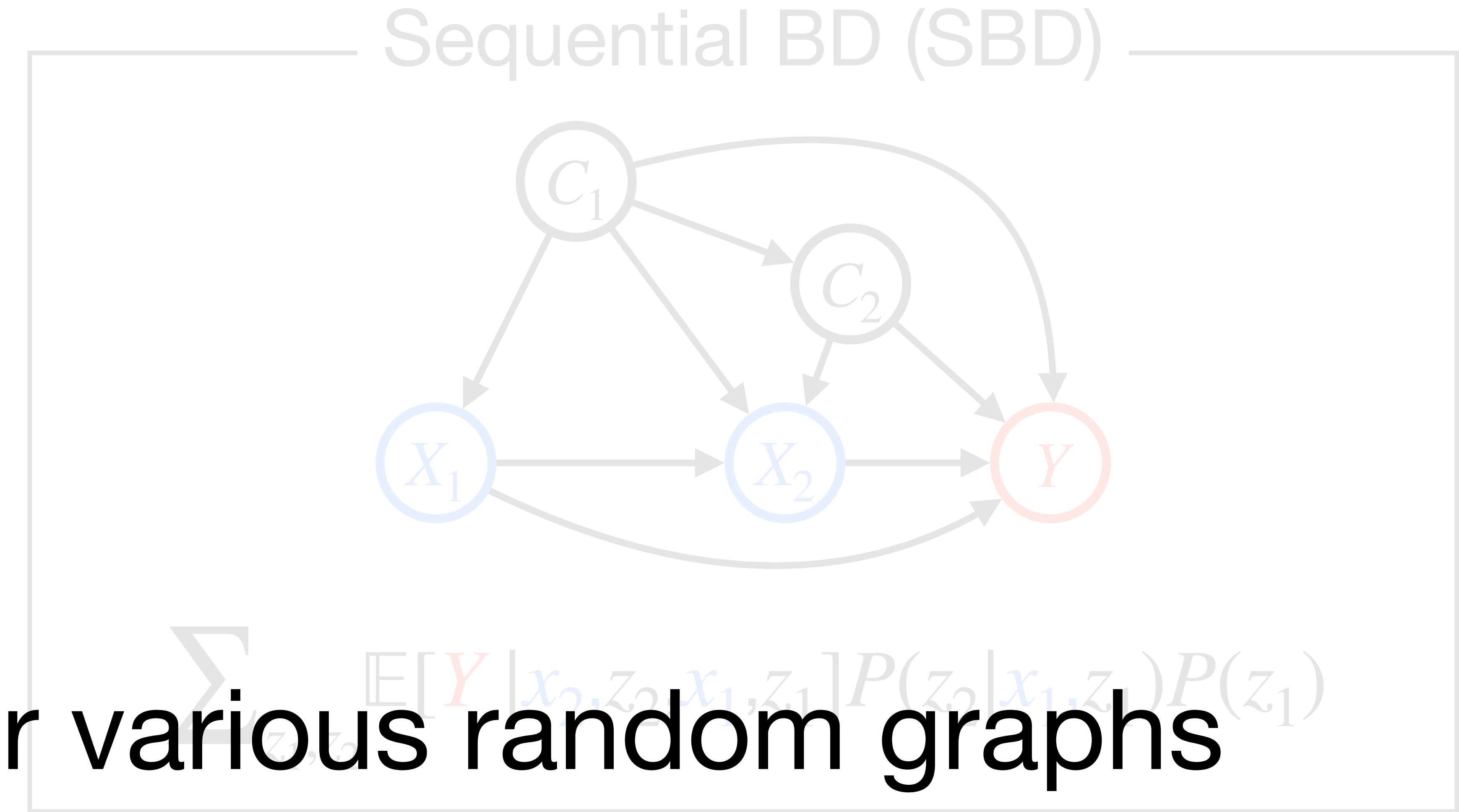


$$\sum_{bax'c} \mathbb{E}[Y \mid baxc]P(r \mid bax)P(b \mid \overline{a}x)P(ax)$$

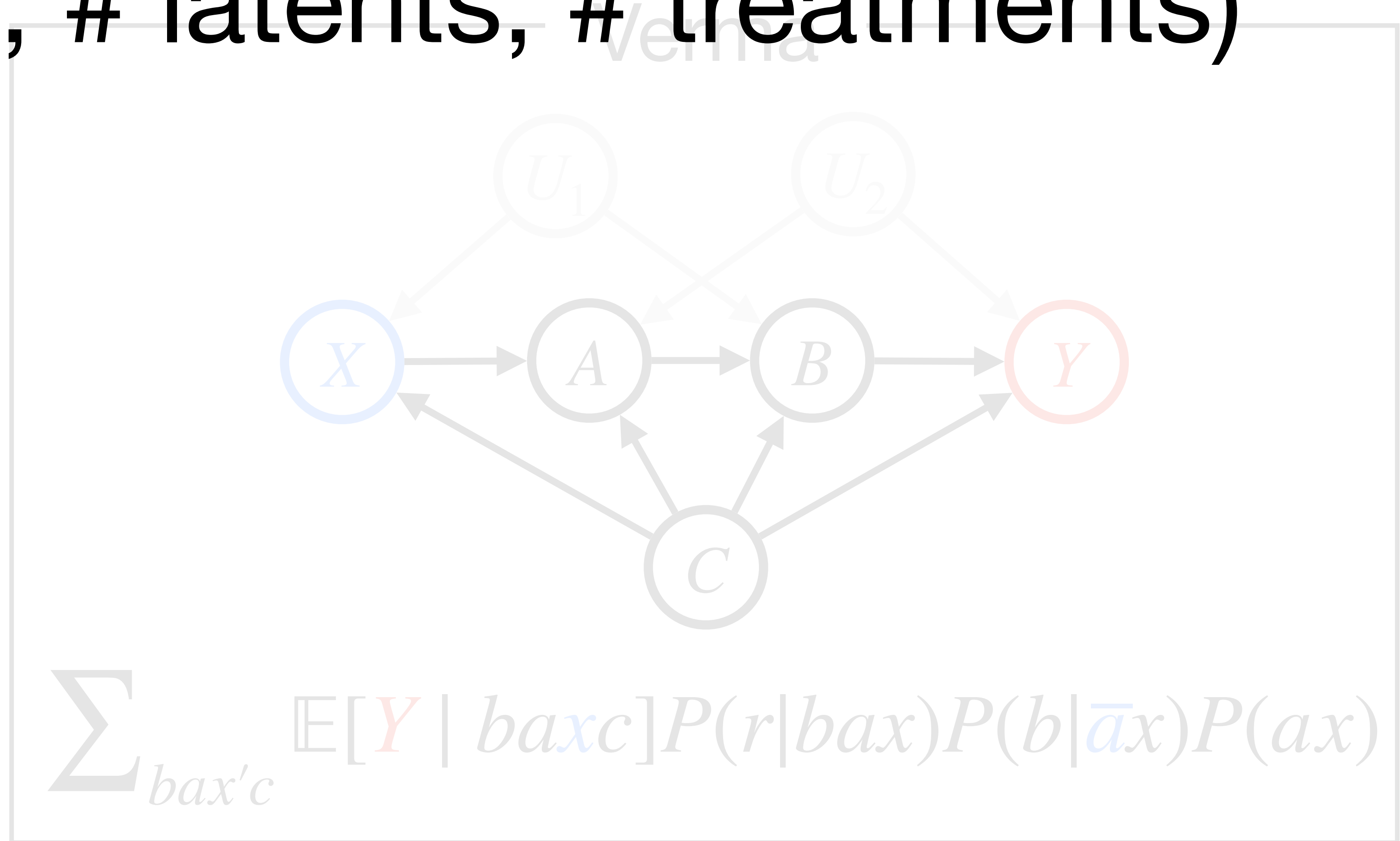
Motivation: Multilinear Causal Estimands

A causal effect $\mathbb{E}[Y \mid \mathrm{do}(\mathbf{x})]$ is often identified as a multilinear functional.





Empirically, P(multilinear | ID) > 99% for various random graphs
(randomness imposed to # observables, # latents, # treatments)



Tasks

Tasks

❶ **Identification**: Sound and complete graphical criterion for identifying causal effects as a multilinear causal estimand

Tasks

❶ **Identification**: Sound and complete graphical criterion for identifying causal effects as a multilinear causal estimand

❷ **Computationally efficiency**: A new formulation for multilinear causal estimands with computational efficiency.

Tasks

❶ **Identification**: Sound and complete graphical criterion for identifying causal effects as a multilinear causal estimand

❷ **Computationally efficiency**: A new formulation for multilinear causal estimands with computational efficiency.

❸ **Sample efficiency:** A doubly robust and sample efficient estimation framework

Multilniear Estimands Criterion (MEC)

Multilniear Estimands Criterion (MEC)

Assumption & Setup

- $\mathbb{E}[Y \mid \mathrm{do}(x)]$ is identifiable
- $\mathbf{D}$: $\mathrm{an}_{\mathcal{G}(\mathbf{V} \setminus \mathbf{X})}(Y)$.
- $\mathbf{D}_{\mathbf{X}} \subseteq \mathbf{D}$: Variables containing $\mathbf{X}$ & sharing the same hidden confounders

Multilniear Estimands Criterion (MEC)

Assumption & Setup

- $\mathbb{E}[Y \mid \mathrm{do}(x)]$ is identifiable
- $\mathbf{D}$: $\mathrm{an}_{\mathcal{G}(\mathbf{V} \setminus \mathbf{X})}(Y)$.
- $\mathbf{D}_{\mathbf{X}} \subseteq \mathbf{D}$: Variables containing $\mathbf{X}$ & sharing the same hidden confounders

Soundness and Completeness of MEC

$\mathbb{E}[Y \mid \mathrm{do}(x)]$ is *multilinear* estimand iff

- $\mathbf{D}_{\mathbf{X}} = \emptyset$;
- $P(\mathbf{D}_{\mathbf{X}} \mid \mathrm{do}(\mathrm{parent}(\mathbf{D}_{\mathbf{X}}))$ is SBD; or given as a product of two SBDs.

Multilniear Estimands Criterion (MEC)

Assumption & Setup

- $\mathbb{E}[Y \mid \mathrm{do}(x)]$ is identifiable
- $\mathbf{D}$: $\mathrm{an}_{\mathscr{G}(\mathbf{V} \setminus \mathbf{X})}(Y)$.
- $\mathbf{D}_\mathbf{X} \subseteq \mathbf{D}$: Variables containing $\mathbf{X}$ & sharing the same hidden confounders

Soundness and Completeness of MEC

$\mathbb{E}[Y \mid \mathrm{do}(x)]$ is *multilinear* estimand iff

- $\mathbf{D}_\mathbf{X} = \emptyset$;
- $P(\mathbf{D}_\mathbf{X} \mid \mathrm{do}(\mathrm{parent}(\mathbf{D}_\mathbf{X}))$ is SBD; or given as a product of two SBDs.
- **MEC for** $|\mathbf{X}| = 1$: X and its children are not connected by bidirected paths.

Multilniear Estimands Criterion (MEC)

Assumption & Setup

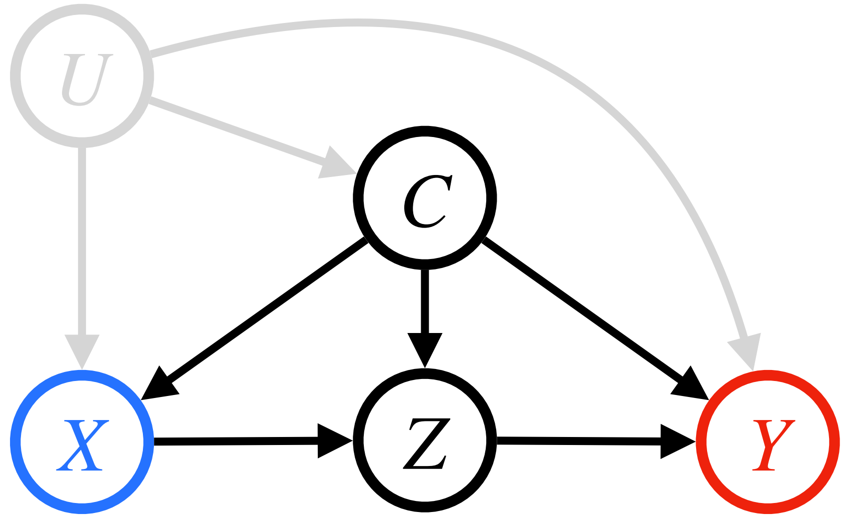
- $\mathbb{E}[Y \mid \mathrm{do}(x)]$ is identifiable
- $\mathbf{D}$: $\mathrm{an}_{\mathcal{G}(\mathbf{V}\setminus\mathbf{X})}(Y)$.
- $\mathbf{D}_{\mathbf{X}} \subseteq \mathbf{D}$: Variables containing $\mathbf{X}$ & sharing the same hidden confounders

Soundness and Completeness of MEC

$\mathbb{E}[Y \mid \mathrm{do}(x)]$ is *multilinear* estimand iff

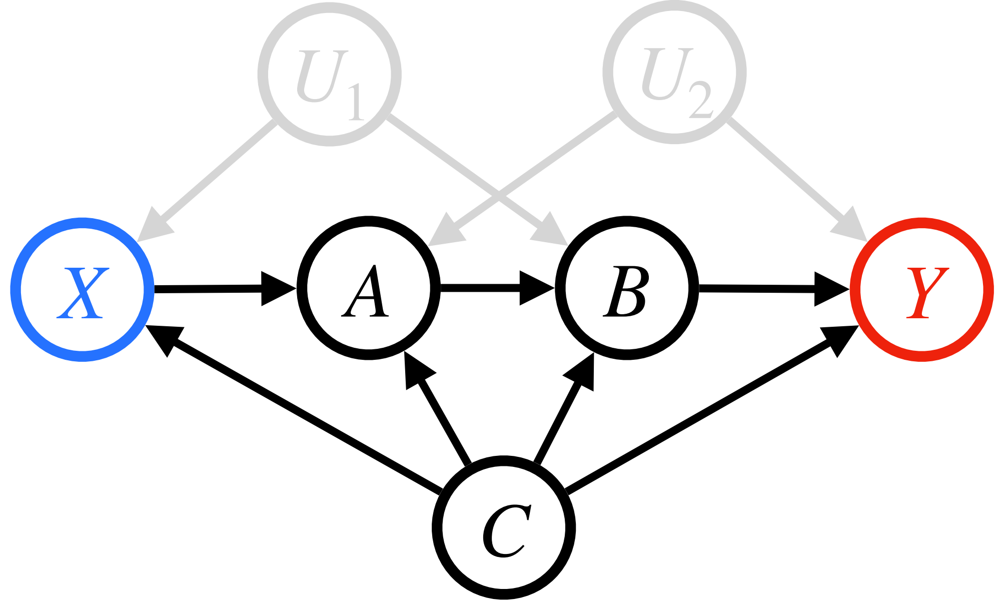
- $\mathbf{D}_{\mathbf{X}} = \varnothing$;
- $P(\mathbf{D}_{\mathbf{X}} \mid \mathrm{do}(\mathrm{parent}(\mathbf{D}_{\mathbf{X}})))$ is SBD; or given as a product of two SBDs.
- **MEC for $|\mathbf{X}| = 1$**: X and its children are not connected by bidirected paths.

Front-Door (FD)



$$\sum_{z,x',c} \mathbb{E}[Y \mid z,x',c]P(z \mid x,c)P(x'c)$$

Verma



$$\sum_{bax'c} \mathbb{E}[Y \mid baxc]P(b|ax'c)P(a|xc)P(x'c)$$

Example: g-computation

Example: g-computation

$$\mathbb{E}[Y \mid \mathrm{do}(x_1,x_2)] \triangleq \sum_{c_1',c_2'} \mathbb{E}[Y \mid x_2,c_2,x_1,c_1]P(c_2|x_1,c_1)P(c_1)$$

Example: g-computation

$$\mathbb{E}[Y \mid \mathrm{do}(x_1, x_2)] \triangleq \boxed{\sum_{c_1', c_2'}} \mathbb{E}[Y \mid x_2, c_2, x_1, c_1] P(c_2 | x_1, c_1) P(c_1)$$

Evaluating $\sum_{c_1', c_2'}$ is computationally expensive, but can be circumvented by nested expectation.

Example: g-computation

$$\mathbb{E}[Y \mid \mathrm{do}(x_1,x_2)] \triangleq \boxed{\sum_{c_1',c_2'}} \mathbb{E}[Y \mid x_2,c_2,x_1,c_1]P(c_2|x_1,c_1)P(c_1)$$

❶ $\mu_2(X_2, C_2, X_1,C_1) \triangleq \mathbb{E}[Y \mid X_2, C_2, X_1,C_1]$

Example: g-computation

$$\mathbb{E}[Y \mid \mathrm{do}(x_1, x_2)] \triangleq \boxed{\sum_{c_1', c_2'}} \mathbb{E}[Y \mid x_2, c_2, x_1, c_1] P(c_2 \mid x_1, c_1) P(c_1)$$

❶ $\mu_2(X_2, C_2, X_1, C_1) \triangleq \mathbb{E}[Y \mid X_2, C_2, X_1, C_1]$

❷ $\mu_1(X_1, C_1) \triangleq \mathbb{E}[\mu_2(x_2, C_2, X_1, C_1) \mid X_1, C_1]$

Example: g-computation

$$\mathbb{E}[Y \mid \mathrm{do}(x_1,x_2)] \triangleq \boxed{\sum_{c_1',c_2'}} \mathbb{E}[Y \mid x_2, c_2, x_1, c_1] P(c_2 \mid x_1, c_1) P(c_1)$$

❶ $\mu_2(X_2, C_2, X_1, C_1) \triangleq \mathbb{E}[Y \mid X_2, C_2, X_1, C_1]$

❷ $\mu_1(X_1, C_1) \triangleq \mathbb{E}[\mu_2(x_2, C_2, X_1, C_1) \mid X_1, C_1]$

$$\boxed{\sum_{c_2} \mathbb{E}[Y \mid x_2, c_2, X_1, C_1] P(c_2 \mid X_1, C_1)}$$

Example: g-computation

$$\mathbb{E}[Y \mid \text{do}(x_1, x_2)] \triangleq \boxed{\sum_{c_1', c_2'}} \mathbb{E}[Y \mid x_2, c_2, x_1, c_1] P(c_2 | x_1, c_1) P(c_1)$$

❶ $\mu_2(X_2, C_2, X_1, C_1) \triangleq \mathbb{E}[Y \mid X_2, C_2, X_1, C_1]$

❷ $\mu_1(X_1, C_1) \triangleq \mathbb{E}[\mu_2(x_2, C_2, X_1, C_1) \mid X_1, C_1]$

$$\sum_{c_2} \mathbb{E}[Y \mid x_2, c_2, X_1, C_1] P(c_2 \mid X_1, C_1)$$

❸ $\mathbb{E}[Y \mid \text{do}(x_1, x_2)] = \mathbb{E}[\mu_1(x_1, C_1)]$

Example: g-computation

$$\mathbb{E}[Y \mid \mathrm{do}(x_1,x_2)] \triangleq \boxed{\sum_{c_1',c_2'}} \mathbb{E}[Y \mid x_2,c_2,x_1,c_1]P(c_2 \mid x_1,c_1)P(c_1)$$

❶ $\mu_2(X_2, C_2, X_1,C_1) \triangleq \mathbb{E}[Y \mid X_2, C_2, X_1,C_1]$

❷ $\mu_1(X_1, C_1) \triangleq \mathbb{E}[\mu_2(x_2, C_2, X_1,C_1) \mid X_1,C_1]$

$$\sum_{c_2} \mathbb{E}[Y \mid x_2,c_2,X_1,C_1]P(c_2 \mid X_1,C_1)$$

❸ $\mathbb{E}[Y \mid \mathrm{do}(x_1,x_2)] = \mathbb{E}[\mu_1(x_1, C_1)]$

$$\sum_{c_1,c_2} \mathbb{E}[Y \mid x_2,c_2,x_1,c_1]P(c_2 \mid x_1,c_1)P(c_1)$$

Example: g-computation

$$\mathbb{E}[Y \mid \mathrm{do}(x_1,x_2)] \triangleq \boxed{\sum_{c_1',c_2'}} \mathbb{E}[Y \mid x_2,c_2,x_1,c_1]P(c_2 \mid x_1,c_1)P(c_1)$$

❶ $\mu_2(X_2, C_2, X_1, C_1) \triangleq \mathbb{E}[Y \mid X_2, C_2, X_1, C_1]$

❷ $\mu_1(X_1, C_1) \triangleq \mathbb{E}[\mu_2(x_2, C_2, X_1, C_1) \mid X_1, C_1]$

$\sum_{c_2} \mathbb{E}[Y \mid x_2, c_2, X_1, C_1]P(c_2 \mid X_1, C_1)$

❸ $\mathbb{E}[Y \mid \mathrm{do}(x_1,x_2)] = \mathbb{E}[\mu_1(x_1, C_1)]$

$\sum_{c_1,c_2} \mathbb{E}[Y \mid x_2, c_2, x_1, c_1]P(c_2 \mid x_1, c_1)P(c_1)$

The SBD estimand can be estimated in a computationally efficient manner using nested conditional expectations.

Limitation of Nested Expectation: FD

Front-Door (FD): $\mathbb{E}[Y \mid \mathrm{do}(x_1)] \triangleq \sum_{z,x',c} \mathbb{E}[Y \mid z,x',c]P(z \mid x,c)P(x'c)$

Limitation of Nested Expectation: FD

Front-Door (FD): $\mathbb{E}[Y \mid \mathrm{do}(x_1)] \triangleq \sum_{z,x',c} \mathbb{E}[Y \mid z,x',c]P(z \mid x,c)P(x'c)$

❶ $\mu_2(Z, X, C) \triangleq \mathbb{E}[Y \mid Z, X, C]$

Limitation of Nested Expectation: FD

Front-Door (FD): $\mathbb{E}[Y \mid \mathrm{do}(x_1)] \triangleq \sum_{z,x',c} \mathbb{E}[Y \mid z,x',c] P(z \mid x,c) P(x'c)$

❶ $\mu_2(Z, X, C) \triangleq \mathbb{E}[Y \mid Z, X, C]$

❷ $\mu_1(X,C) \triangleq \mathbb{E}[\mu_2(Z, X, C) \mid X,C]$

$\sum_z \mathbb{E}[Y \mid z,X,C] P(z \mid X,C)$

Limitation of Nested Expectation: FD

Front-Door (FD): $\mathbb{E}[Y \mid \mathrm{do}(x_1)] \triangleq \sum_{z,x',c} \mathbb{E}[Y \mid z,x',c]P(z \mid x,c)P(x'c)$

❶ $\mu_2(Z, X, C) \triangleq \mathbb{E}[Y \mid Z, X, C]$

❷ $\mu_1(X,C) \triangleq \mathbb{E}[\mu_2(Z, X, C) \mid X,C]$

$\sum_z \mathbb{E}[Y \mid z,X,C]P(z \mid X,C)$

❸ $\mathbb{E}[\mu_1(x, C)] \neq \mathbb{E}[Y \mid \mathrm{do}(x)]$

$\sum_{z,x',c} \mathbb{E}[Y \mid z,x,c]P(z \mid x,c)P(c)$

Limitation of Nested Expectation: FD

Front-Door (FD): $\mathbb{E}[Y \mid \mathrm{do}(x_1)] \triangleq \sum_{z,x',c} \mathbb{E}[Y \mid z,x',c]P(z \mid x,c)P(x'c)$

❶ $\mu_2(Z, X, C) \triangleq \mathbb{E}[Y \mid Z, X, C]$

❷ $\mu_1(X,C) \triangleq \mathbb{E}[\mu_2(Z, X, C) \mid X,C]$

$\sum_z \mathbb{E}[Y \mid z,X,C]P(z \mid X,C)$

❸ $\mathbb{E}[\mu_1(x, C)] \neq \mathbb{E}[Y \mid \mathrm{do}(x)]$

$\sum_{z,x',c} \mathbb{E}[Y \mid z,x,c]P(z \mid x,c)P(c)$

The standard nested expectation cannot represent multilinear estimands when treatments are both marginalized and fixed simultaneously.

Limitation of Nested Expectation: Multilinear Estimand

Front-Door (FD) $\sum_{z,x',c} \mathbb{E}[Y \mid z,x',c]P(z \mid x,c)P(x'c)$

Verma $\sum_{bax'c} \mathbb{E}[Y \mid baxc]P(b|ax'c)P(a|xc)P(x'c)$

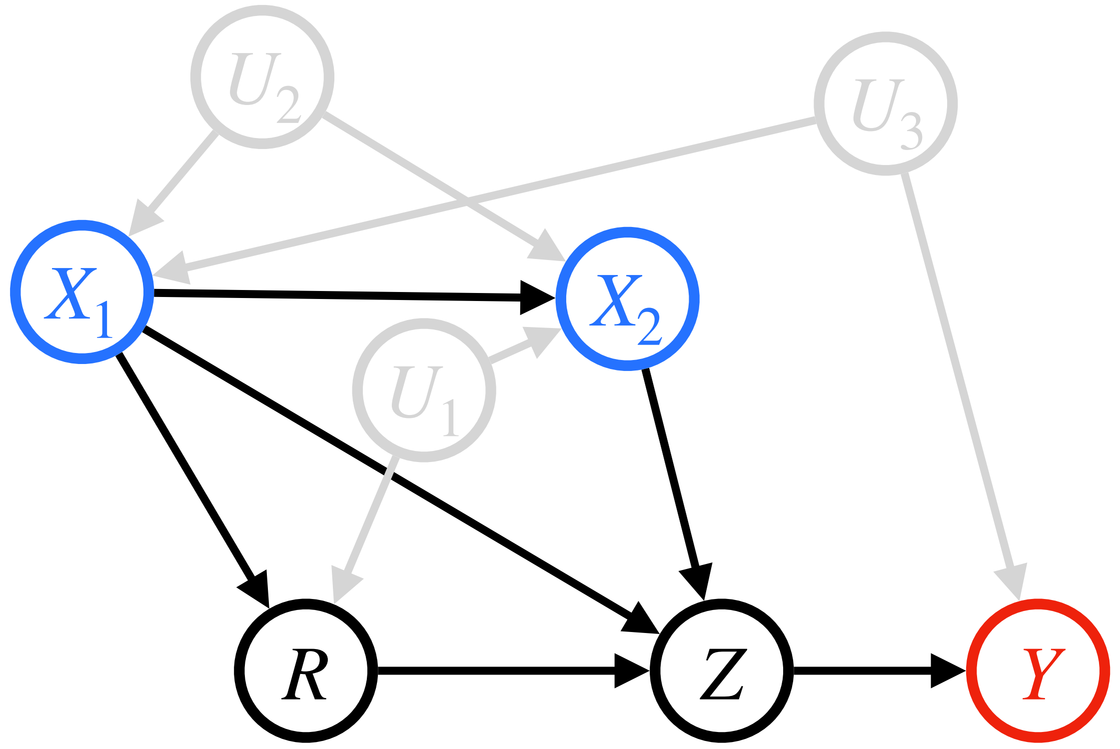
Treatments $\mathbf{X}$ are fixed to $\mathbf{x}$ and marginalized $\mathbf{x}'$ simultaneously.

Limitation of Nested Expectation: Multilinear Estimand

$$\left.\begin{array}{ll} \textbf{Front-Door (FD)} & \sum_{z,x',c} \mathbb{E}[Y\,|z,x',c]P(z\,|\,x,c)P(x'c) \\ \textbf{Verma} & \sum_{bax'c} \mathbb{E}[Y\,|\,baxc]P(b|ax'c)P(a|xc)P(x'c) \end{array}\right\}$$

Treatments $\mathbf{X}$ are fixed to $\mathbf{x}$ and marginalized $\mathbf{x}'$ simultaneously.

Example



$$\sum_{r,z,x_1',x_2',r'} \mathbb{E}[Y\,|z,\,r,\,x_2',\,x_1']P(z\,|\,r,x_1,x_2)P(r\,|\,x_1)P(r',x_1',x_2')$$

∃variable that is marginalized multiple times.

Kernel Policy Product: Representation of MCE

Kernel Policy Product: Representation of MCE

Front-Door (FD) $\sum_{z,x',c} \mathbb{E}[Y \mid z,x',c]P(z \mid x,c)P(x'c)$

Kernel Policy Product: Representation of MCE

Front-Door (FD) $\sum_{z,x',c} \mathbb{E}[Y\,|z,x',c]P(z \mid x,c)P(x'c)$

$=$ Expectation of Y over $P(Y \mid Z,X,C)P(Z \mid \dot{X},C)\mathbb{I}(\dot{X}=x)P(X,C)$

Kernel Policy Product: Representation of MCE

Front-Door (FD) $\sum_{z,x',c} \mathbb{E}[Y\,|z,x',c]P(z\,|\,x,c)P(x'c)$

$=$ Expectation of Y over $P(Y\,|\,Z,X,C)P(Z\,|\,\dot{X},C)\mathbb{I}(\dot{X}=x)P(X,C)$

Copied Proxy: $\dot{X}$ is an *independent* copy of X s.t.
$P(Z|X,C) = P(Z|\dot{X},C)$.

Kernel Policy Product: Representation of MCE

Front-Door (FD) $\sum_{z,x',c} \mathbb{E}[Y\,|z,x',c]P(z\,|\,x,c)P(x'c)$

$=$ Expectation of Y over $P(Y\,|\,Z,X,C)P(Z\,|\,\dot{X},C)\mathbb{I}(\dot{X}=x)P(X,C)$

Copied Proxy: $\dot{X}$ is an *independent* copy of X s.t.
$P(Z|X,C) = P(Z|\dot{X},C)$.

Verma $\sum_{bax'c} \mathbb{E}[Y\,|\,baxc]P(b|ax'c)P(a|xc)P(x'c)$

Kernel Policy Product: Representation of MCE

Front-Door (FD) $\sum_{z,x',c} \mathbb{E}[Y\,|z,x',c]P(z\,|\,x,c)P(x'c)$

$= \text{Expectation of } Y \text{ over } P(Y\,|\,Z,X,C)P(Z\,|\,\dot{X},C)\mathbb{I}(\dot{X}{=}x)P(X,C)$

Copied Proxy: $\dot{X}$ is an *independent* copy of X s.t.
$P(Z|X,C) = P(Z|\dot{X},C)$.

Verma $\sum_{bax'c} \mathbb{E}[Y\,|\,baxc]P(b|ax'c)P(a|xc)P(x'c)$

$= \text{Expectation of } Y \text{ over } P(Y\,|\,B,A,X,C)\mathbb{I}(X{=}x)P(B\,|\,A,\dot{X},C)P(A\,|X,C)P(\dot{X},C)$

Kernel Policy Product: Representation of MCE

Front-Door (FD) $\sum_{z,x',c} \mathbb{E}[Y \,|z,x',c]P(z \,|\, x,c)P(x'c)$

$=$ Expectation of Y over $P(Y \,|\, Z,X,C)P(Z \,|\, \dot{X},C)\mathbb{I}(\dot{X}=x)P(X,C)$

Copied Proxy: $\dot{X}$ is an *independent* copy of X s.t.
$P(Z|X,C) = P(Z|\dot{X},C)$.

Kernel Policy Product: A product of conditional probability kernels and policies over variables & their copied proxies

Verma $\sum_{bax'c} \mathbb{E}[Y \,|\, baxc]P(b|ax'c)P(a|xc)P(x'c)$

$=$ Expectation of Y over $P(Y \,|\, B,A,X,C)\mathbb{I}(X=x)P(B \,|\, A,\dot{X},C)P(A \,|X,C)P(\dot{X},C)$

Computational Efficiency Gain

Front-Door (FD): $\mathbb{E}[Y \mid \mathrm{do}(x)] \triangleq \sum_{z,x',c} \mathbb{E}[Y \mid z,x',c] P(z \mid x,c) P(x'c)$

$=$ Expectation of Y over $P(Y \mid Z,X,C) P(Z \mid \dot{X},C) \mathbb{I}(\dot{X}=x) P(X,C)$

Computational Efficiency Gain

Front-Door (FD): $\mathbb{E}[Y \mid \mathrm{do}(x)] \triangleq \sum_{z,x',c} \mathbb{E}[Y \mid z,x',c]P(z \mid x,c)P(x'c)$

$=$ Expectation of Y over $P(Y \mid Z,X,C)P(Z \mid \dot{X},C)\mathbb{I}(\dot{X}=x)P(X,C)$

❶ Learn $\mu_2(Z, X, C) \triangleq \mathbb{E}[Y \mid Z, X, C]$

Computational Efficiency Gain

Front-Door (FD): $\mathbb{E}[Y \mid \mathrm{do}(x)] \triangleq \sum_{z,x',c} \mathbb{E}[Y \,|z,x',c]P(z \mid x,c)P(x'c)$

$=$ Expectation of Y over $P(Y \mid Z,X,C)P(Z \mid \dot{X},C)\mathbb{I}(\dot{X}=x)P(X,C)$

❶ Learn $\mu_2(Z, X, C) \triangleq \mathbb{E}[Y \mid Z, X, C]$

❷ Evaluate μ_2 on $(Z, \dot{X}, C)$ ($\dot{X}$ is a copied proxy of X). $\mu_2(Z, \dot{X}, C)$:= $\mathbb{E}[Y \,|Z,X \leftarrow \dot{X},C]$

Computational Efficiency Gain

Front-Door (FD): $\mathbb{E}[Y \mid \mathrm{do}(x)] \triangleq \sum_{z,x',c} \mathbb{E}[Y \mid z,x',c]P(z \mid x,c)P(x'c)$

$= \text{Expectation of } Y \text{ over } P(Y \mid Z,X,C)P(Z \mid \dot{X},C)\mathbb{I}(\dot{X}=x)P(X,C)$

❶ Learn $\mu_2(Z, X, C) \triangleq \mathbb{E}[Y \mid Z, X, C]$

❷ Evaluate μ_2 on $(Z, \dot{X}, C)$ ($\dot{X}$ is a copied proxy of X). $\mu_2(Z, \dot{X}, C) := \mathbb{E}[Y \mid Z, X \leftarrow \dot{X}, C]$

❸ Learn $\mu_1(X,\dot{X},C)=\mathbb{E}[\mu_2(Z,\dot{X},C) \mid X,\dot{X},C]$

Computational Efficiency Gain

Front-Door (FD): $\mathbb{E}[Y \mid \mathrm{do}(x)] \triangleq \sum_{z,x',c} \mathbb{E}[Y \mid z,x',c]P(z \mid x,c)P(x'c)$

$=$ Expectation of Y over $P(Y \mid Z,X,C)P(Z \mid \dot{X},C)\mathbb{I}(\dot{X}=x)P(X,C)$

❶ Learn $\mu_2(Z, X, C) \triangleq \mathbb{E}[Y \mid Z, X, C]$

❷ Evaluate μ_2 on $(Z, \dot{X}, C)$ ($\dot{X}$ is a copied proxy of X). $\mu_2(Z, \dot{X}, C) := \mathbb{E}[Y \mid Z, X \leftarrow \dot{X}, C]$

❸ Learn $\mu_1(X,\dot{X},C)=\mathbb{E}[\mu_2(Z,\dot{X},C) \mid X,\dot{X},C]$ $\boxed{= \sum_z \mu_2(z,\dot{X},C)P(z \mid X,\dot{X},C) = \sum_z \mathbb{E}[Y \mid Z,X \leftarrow \dot{X},C]P(z \mid X,C)}$

Computational Efficiency Gain

Front-Door (FD): $\mathbb{E}[Y \mid \mathrm{do}(x)] \triangleq \sum_{z,x',c} \mathbb{E}[Y \,|z,x',c]P(z \mid x,c)P(x'c)$

$=$ Expectation of Y over $P(Y \mid Z,X,C)P(Z \mid \dot{X},C)\mathbb{I}(\dot{X}{=}x)P(X,C)$

❶ Learn $\mu_2(Z, X, C) \triangleq \mathbb{E}[Y \mid Z, X, C]$

❷ Evaluate μ_2 on $(Z, \dot{X}, C)$ ($\dot{X}$ is a copied proxy of X). $\mu_2(Z, \dot{X}, C)$:= $\mathbb{E}[Y \,|Z,X \leftarrow \dot{X},C]$

❸ Learn $\mu_1(X,\dot{X},C){=}\mathbb{E}[\mu_2(Z,\dot{X},C) \mid X,\dot{X},C]$ $\boxed{= \sum_z \mu_2(z,\dot{X},C)P(z \mid X,\dot{X},C) = \sum_z \mathbb{E}[Y \,|Z,X \leftarrow \dot{X},C]P(z \mid X,C)}$

❹ Evaluate μ_1 on (x, X, C) to have $\mu_1(x, X, C)$

Computational Efficiency Gain

Front-Door (FD): $\mathbb{E}[Y \mid \mathrm{do}(x)] \triangleq \sum_{z,x',c} \mathbb{E}[Y \mid z,x',c]P(z \mid x,c)P(x'c)$

$=$ Expectation of Y over $P(Y \mid Z,X,C)P(Z \mid \dot{X},C)\mathbb{I}(\dot{X}=x)P(X,C)$

❶ Learn $\mu_2(Z, X, C) \triangleq \mathbb{E}[Y \mid Z, X, C]$

❷ Evaluate μ_2 on $(Z, \dot{X}, C)$ ($\dot{X}$ is a copied proxy of X). $\mu_2(Z, \dot{X}, C) := \mathbb{E}[Y \mid Z, X \leftarrow \dot{X}, C]$

❸ Learn $\mu_1(X,\dot{X},C)=\mathbb{E}[\mu_2(Z,\dot{X},C) \mid X,\dot{X},C]$ $\boxed{= \sum_z \mu_2(z,\dot{X},C)P(z \mid X,\dot{X},C) = \sum_z \mathbb{E}[Y \mid Z,X \leftarrow \dot{X},C]P(z \mid X,C)}$

❹ Evaluate μ_1 on (x, X, C) to have $\mu_1(x, X, C)$ $\boxed{= \sum_z \mathbb{E}[Y \mid z,X,C]P(z \mid x,C)}$

Computational Efficiency Gain

Front-Door (FD): $\mathbb{E}[Y \mid \mathrm{do}(x)] \triangleq \sum_{z,x',c} \mathbb{E}[Y \mid z,x',c]P(z \mid x,c)P(x'c)$

$= \text{Expectation of } Y \text{ over } P(Y \mid Z,X,C)P(Z \mid \dot{X},C)\mathbb{I}(\dot{X}=x)P(X,C)$

❶ Learn $\mu_2(Z, X, C) \triangleq \mathbb{E}[Y \mid Z, X, C]$

❷ Evaluate μ_2 on $(Z, \dot{X}, C)$ ($\dot{X}$ is a copied proxy of X). $\mu_2(Z, \dot{X}, C) := \mathbb{E}[Y \mid Z, X \leftarrow \dot{X}, C]$

❸ Learn $\mu_1(X,\dot{X},C) = \mathbb{E}[\mu_2(Z,\dot{X},C) \mid X,\dot{X},C]$ $\boxed{= \sum_z \mu_2(z,\dot{X},C)P(z \mid X,\dot{X},C) = \sum_z \mathbb{E}[Y \mid Z,X \leftarrow \dot{X},C]P(z \mid X,C)}$

❹ Evaluate μ_1 on (x, X, C) to have $\mu_1(x, X, C)$ $\boxed{= \sum_z \mathbb{E}[Y \mid z,X,C]P(z \mid x,C)}$

❺ $\mathbb{E}[Y \mid \mathrm{do}(x)] = \mathbb{E}[\mu_1(x, X, C)]$

Computational Efficiency Gain

Front-Door (FD): $\mathbb{E}[Y \mid \mathrm{do}(x)] \triangleq \sum_{z,x',c} \mathbb{E}[Y \mid z,x',c]P(z \mid x,c)P(x'c)$

$= \text{Expectation of } Y \text{ over } P(Y \mid Z,X,C)P(Z \mid \dot{X},C)\mathbb{I}(\dot{X}=x)P(X,C)$

❶ Learn $\mu_2(Z, X, C) \triangleq \mathbb{E}[Y \mid Z, X, C]$

❷ Evaluate μ_2 on $(Z, \dot{X}, C)$ ($\dot{X}$ is a copied proxy of X). $\mu_2(Z, \dot{X}, C) := \mathbb{E}[Y \mid Z, X \leftarrow \dot{X}, C]$

❸ Learn $\mu_1(X,\dot{X},C) = \mathbb{E}[\mu_2(Z,\dot{X},C) \mid X,\dot{X},C]$ $\boxed{= \sum_z \mu_2(z,\dot{X},C)P(z \mid X,\dot{X},C) = \sum_z \mathbb{E}[Y \mid Z, X \leftarrow \dot{X}, C]P(z \mid X,C)}$

❹ Evaluate μ_1 on (x, X, C) to have $\mu_1(x, X, C)$ $\boxed{= \sum_z \mathbb{E}[Y \mid z,X,C]P(z \mid x,C)}$

❺ $\mathbb{E}[Y \mid \mathrm{do}(x)] = \mathbb{E}[\mu_1(x, X, C)]$ $\boxed{= \sum_{z,x',c} \mathbb{E}[Y \mid z,x',c]P(z \mid x,c)P(x'c)}$

Computational Efficiency Gain

Front-Door (FD): $\mathbb{E}[Y \mid \text{do}(x)] \triangleq \sum_{z,x',c} \mathbb{E}[Y \mid z,x',c] P(z \mid x,c) P(x'c)$

$=$ Expectation of Y over $P(Y \mid Z,X,C) P(Z \mid \dot{X},C) \mathbb{I}(\dot{X}=x) P(X,C)$

Computational efficiency gain via replacing $\sum_{z,x',c}$ through KPP expression with copied proxies and nested conditional expectations

① Learn $\mu_2(Z, X, C) \triangleq \mathbb{E}[Y \mid Z, X, C]$

② Evaluate μ_2 on $(Z, \dot{X}, C)$ ($\dot{X}$ is a copied proxy of X): $\mu_2(Z, \dot{X}, C) := \mathbb{E}[Y \mid Z, X \leftarrow \dot{X}, C]$

③ L[illegible] $(X, \dot{X}, C) \triangleq \mathbb{E}[$[illegible]$(Z, \dot{X}, C) \mid X, \dot{X}, C]$ $= \sum_z$ [illegible] $\sum_z$ [illegible] $(z \mid X,C)$

④ Evaluate μ_1 on (x, X, C) to have $\mu_1(x, X, C)$ $= \sum_z \mathbb{E}[Y \mid z,X,C] P(z \mid x,C)$

⑤ $\mathbb{E}[Y \mid \text{do}(x)] = \mathbb{E}[\mu_1(x, X, C)]$ $= \sum_{z,x',c} \mathbb{E}[Y \mid z,x',c] P(z \mid x,c) P(x'c)$

Sample Efficient Estimation

Sample Efficient Estimation

❶ MCE($\boldsymbol{\mu}$,$\boldsymbol{\pi}$), where $\boldsymbol{\pi}$ is a set of the 1st-order error correction parameters.

Sample Efficient Estimation

❶ MCE($\boldsymbol{\mu}$,$\boldsymbol{\pi}$), where $\boldsymbol{\pi}$ is a set of the 1st-order error correction parameters.

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

Sample Efficient Estimation

❶ MCE($\boldsymbol{\mu}$,$\boldsymbol{\pi}$), where $\boldsymbol{\pi}$ is a set of the 1st-order error correction parameters.

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

$$\mu_2(Z, X, C) \triangleq \mathbb{E}[Y \mid Z, X, C]$$

Sample Efficient Estimation

❶ MCE($\boldsymbol{\mu}$,$\boldsymbol{\pi}$), where $\boldsymbol{\pi}$ is a set of the 1st-order error correction parameters.

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

$\mu_2(Z, X, C) \triangleq \mathbb{E}[Y \mid Z, X, C]$

$\pi_2 > 0$ s.t. $\mathbb{E}[\mu_2 \pi_2] = \mathbb{E}[\mu_2(Z, \dot{A}, X)]$

Sample Efficient Estimation

① MCE($\boldsymbol{\mu}$,$\boldsymbol{\pi}$), where $\boldsymbol{\pi}$ is a set of the 1st-order error correction parameters.

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

$\mu_2(Z, X, C) \triangleq \mathbb{E}[Y \mid Z, X, C]$

$\pi_2 > 0$ s.t. $\mathbb{E}[\mu_2 \pi_2] = \mathbb{E}[\mu_2(Z, \dot{A}, X)]$

$\mu_1(A,\dot{A},X)=\mathbb{E}[\mu_2(Z,\dot{X},C) \mid A,\dot{A},X]$

Sample Efficient Estimation

① MCE($\boldsymbol{\mu}$,$\boldsymbol{\pi}$), where $\boldsymbol{\pi}$ is a set of the 1st-order error correction parameters.

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

$\mu_2(Z, X, C) \triangleq \mathbb{E}[Y \mid Z, X, C]$ $\pi_2 > 0$ s.t. $\mathbb{E}[\mu_2\pi_2] = \mathbb{E}[\mu_2(Z, \dot{A}, X)]$

$\mu_1(A,\dot{A},X)=\mathbb{E}[\mu_2(Z,\dot{X},C) \mid A,\dot{A},X]$ $\pi_1 > 0$ s.t. $\mathbb{E}[\mu_1\pi_1] = \mathbb{E}[\mu_1(\bar{a}, A, X)]$

Sample Efficient Estimation

❶ MCE($\boldsymbol{\mu}$,$\boldsymbol{\pi}$), where $\boldsymbol{\pi}$ is a set of the 1st-order error correction parameters.

$$\mathbb{E}[Y \mid \mathrm{do}(x)] \triangleq \sum\nolimits_{z,x',c} \mathbb{E}[Y \mid z,x',c]P(z \mid x,c)P(x'c)$$

$\mu_2(Z, X, C) \triangleq \mathbb{E}[Y \mid Z, X, C]$ $\quad$ $\pi_2 > 0$ s.t. $\mathbb{E}[\mu_2\pi_2] = \mathbb{E}[\mu_2(Z, \dot{A}, X)]$

$\mu_1(A,\dot{A},X)$=$\mathbb{E}[\mu_2(Z,\dot{X},C) \mid A,\dot{A},X]$ $\quad$ $\pi_1 > 0$ s.t. $\mathbb{E}[\mu_1\pi_1] = \mathbb{E}[\mu_1(\bar{a}, A, X)]$

❷ DML-MCE($\hat{\mu}$,$\hat{\pi}$) is an AIPW-style estimator tailored to KPP.

Sample Efficient Estimation

❶ MCE($\boldsymbol{\mu}$,$\boldsymbol{\pi}$), where π is a set of the 1st-order error correction parameters.

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

$$\mu_2(Z, X, C) \triangleq \mathbb{E}[Y \mid Z, X, C] \qquad \pi_2 > 0 \text{ s.t. } \mathbb{E}[\mu_2\pi_2] = \mathbb{E}[\mu_2(Z, \dot{A}, X)]$$

$$\mu_1(A,\dot{A},X)=\mathbb{E}[\mu_2(Z,\dot{X},C) \mid A,\dot{A},X] \qquad \pi_1 > 0 \text{ s.t. } \mathbb{E}[\mu_1\pi_1] = \mathbb{E}[\mu_1(\bar{a}, A, X)]$$

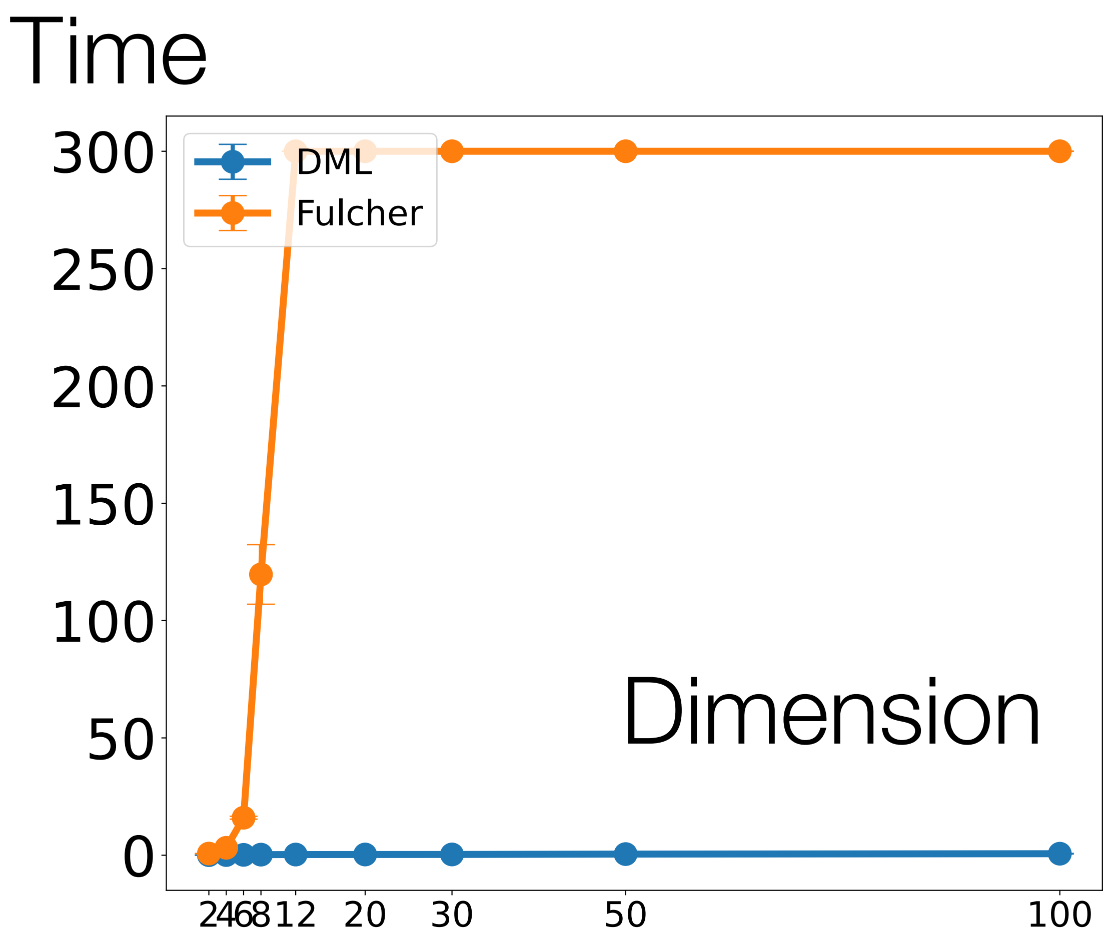
❷ DML-MCE($\hat{\mu}$,$\hat{\pi}$) is an AIPW-style estimator tailored to KPP.

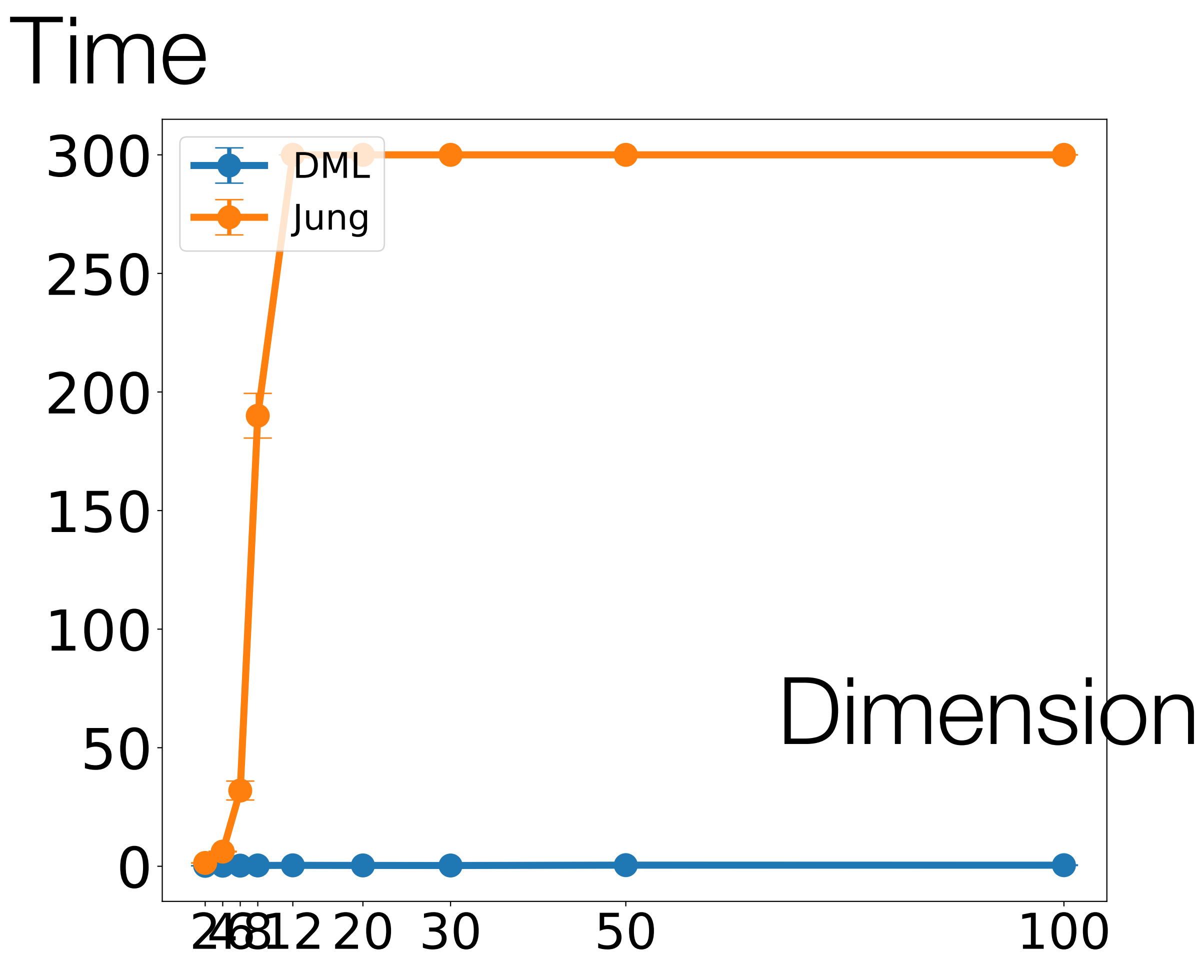
$$\text{Error}(\text{DML-MCE}(\hat{\boldsymbol{\mu}},\hat{\boldsymbol{\pi}}), \text{DML}(\boldsymbol{\mu},\boldsymbol{\pi})) = \sum_i \text{Error}(\hat{\mu}_i,\mu_i) \times \text{Error}(\hat{\pi}_i,\pi_i)$$

Simulation Results

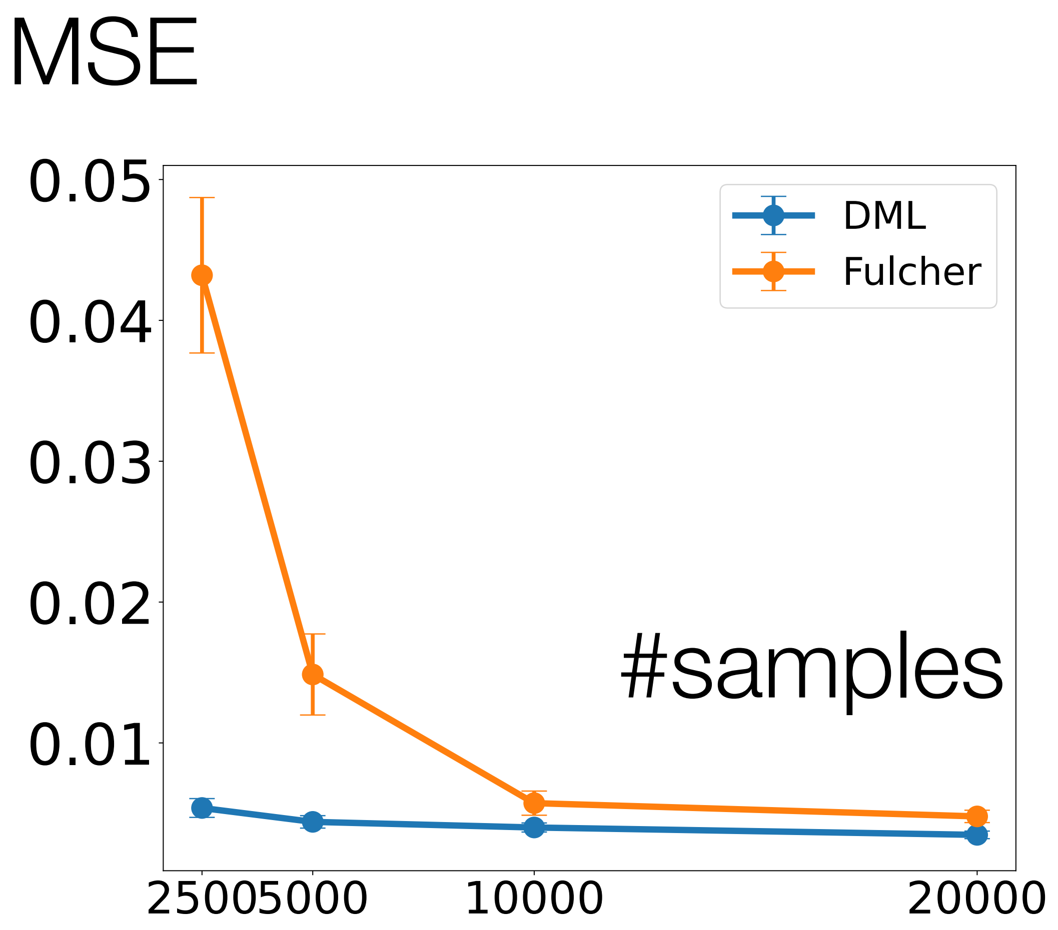
FD

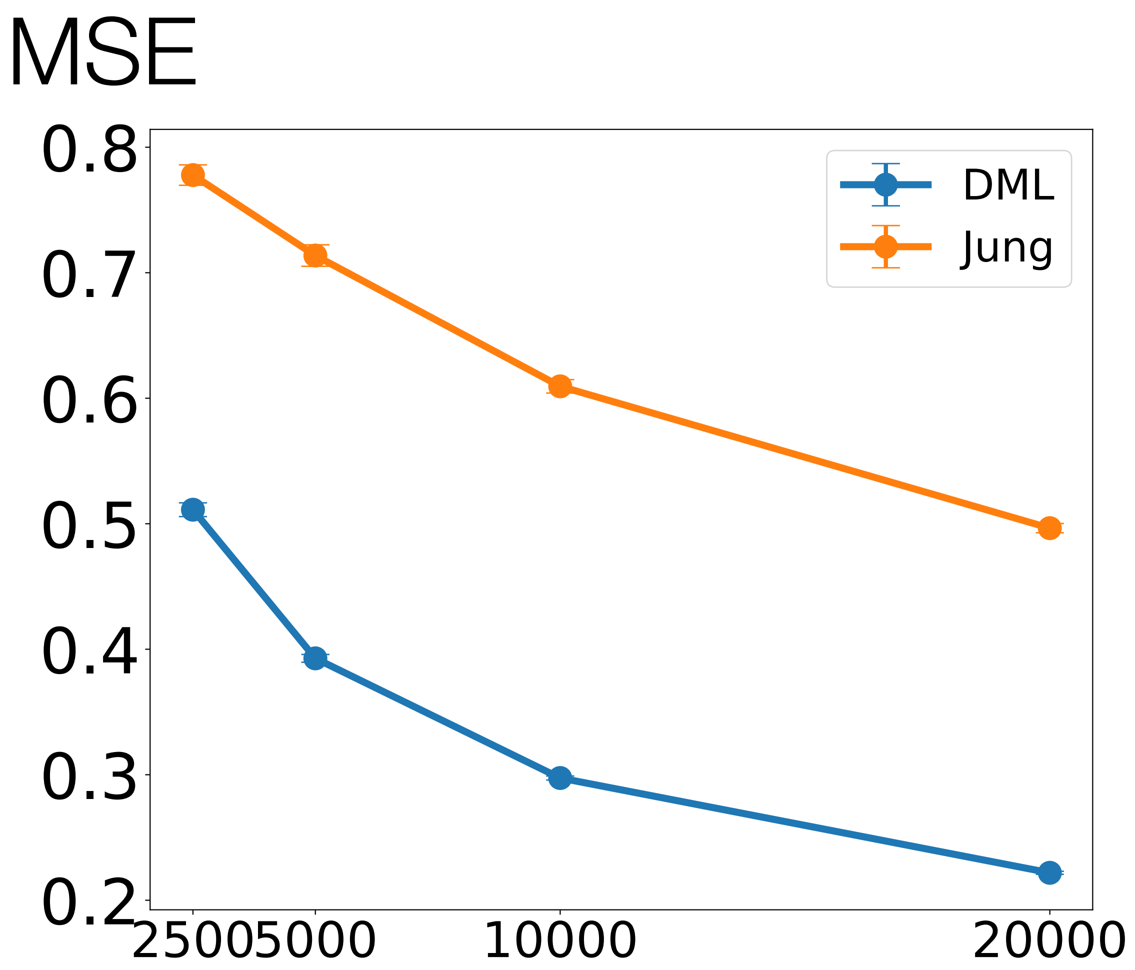
Verma





- Existing estimators' evaluation time increase as dimensions increases
- DML estimator exhibits computational efficiency gains.





- DML estimator exhibits sample efficiency.

Talk Outline

❶ Estimating causal effects from observations

❷ Estimating causal effects from data fusion

➡ ❸ Unified causal effect estimation method

❹ Conclusion

Talk Outline

✓

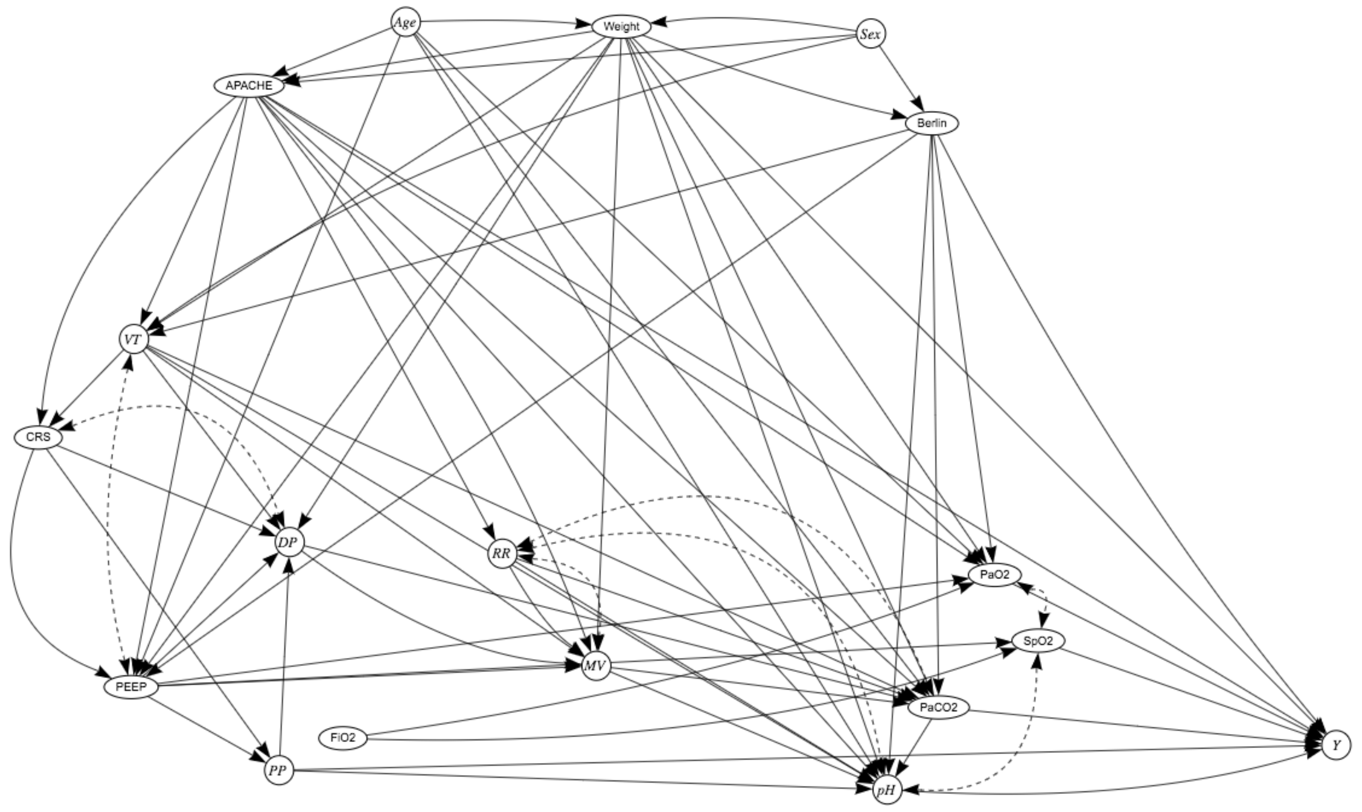
➡ ④ Conclusion

This Talk: Estimating Causal Effects

This Talk: Estimating Causal Effects

Tasks

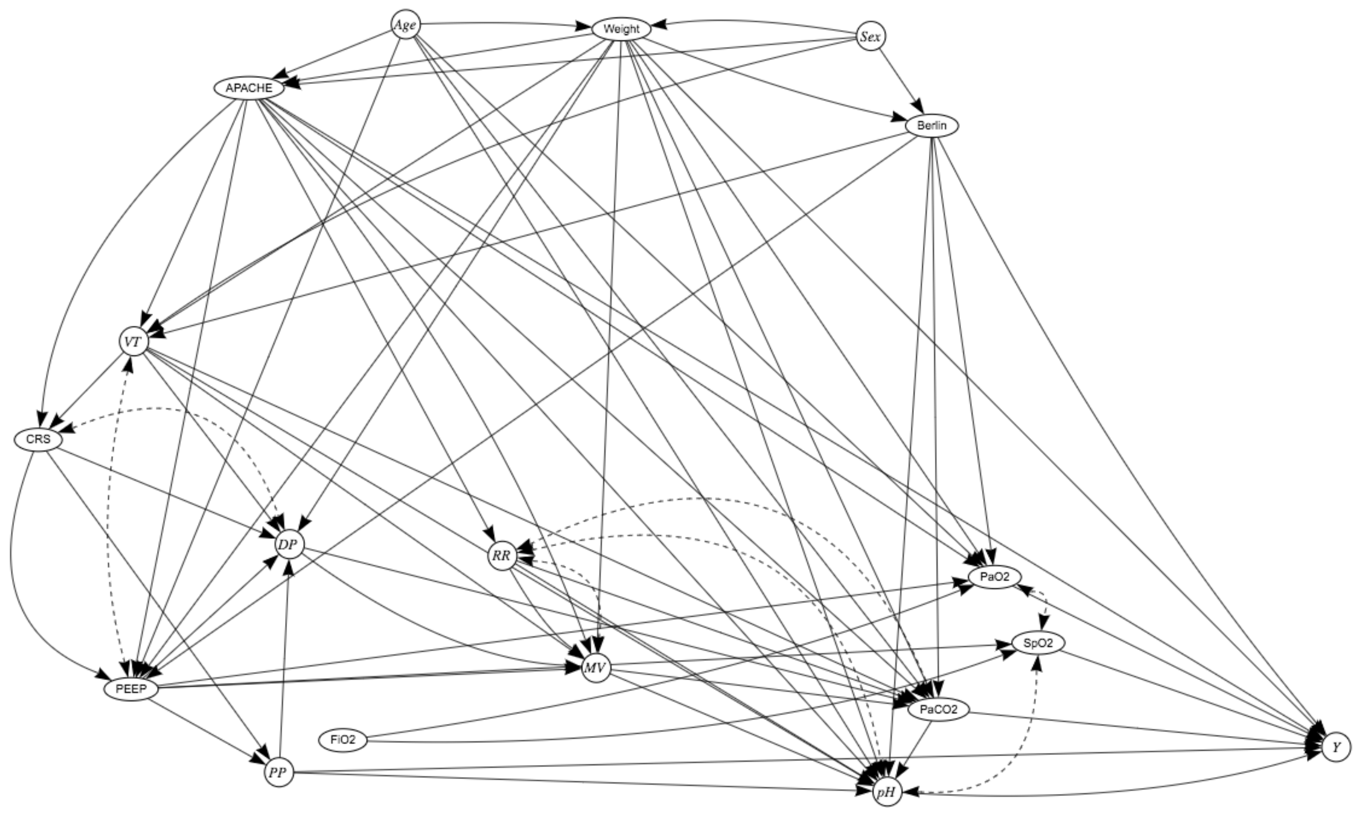
1. From Observation



This Talk: Estimating Causal Effects

Tasks

1. From Observation



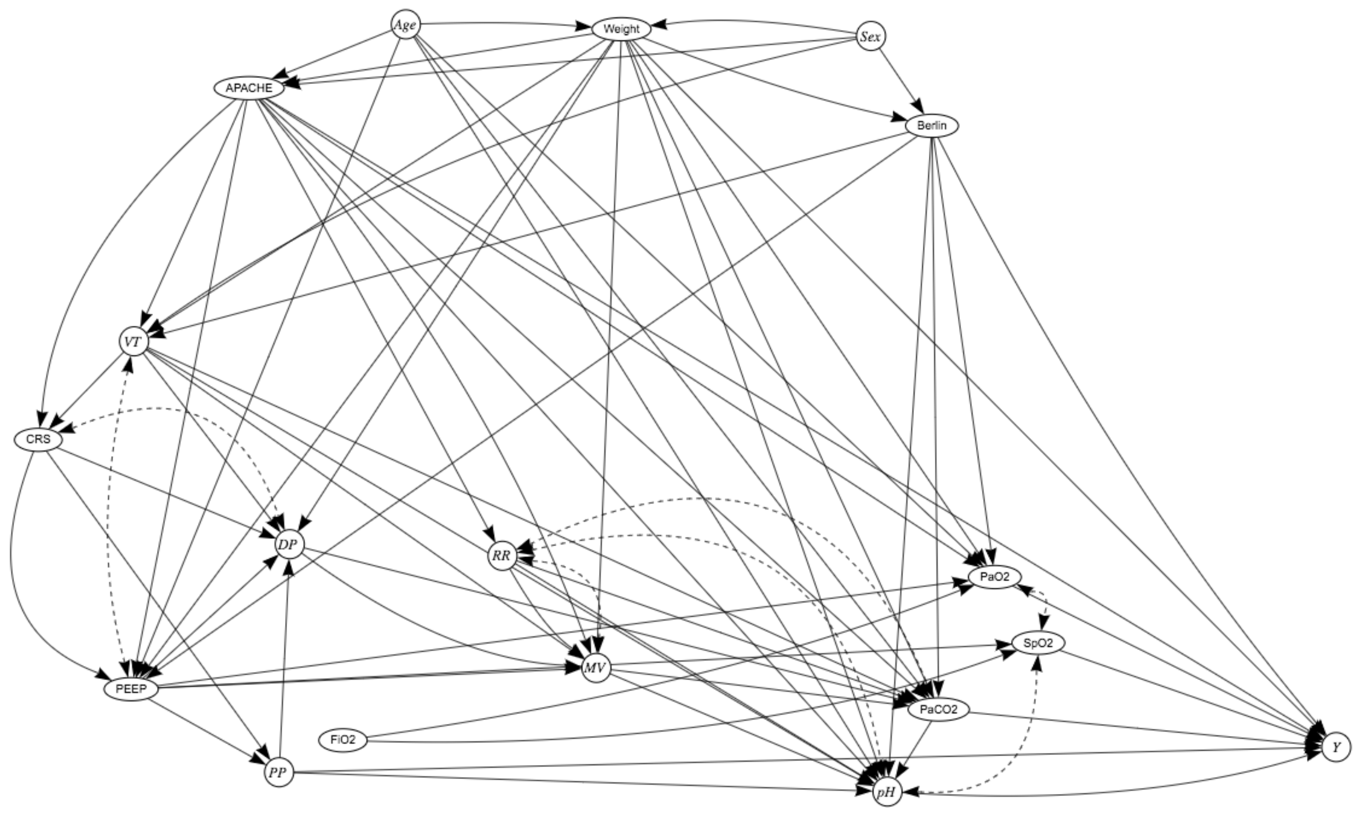
Solution

DML-ID

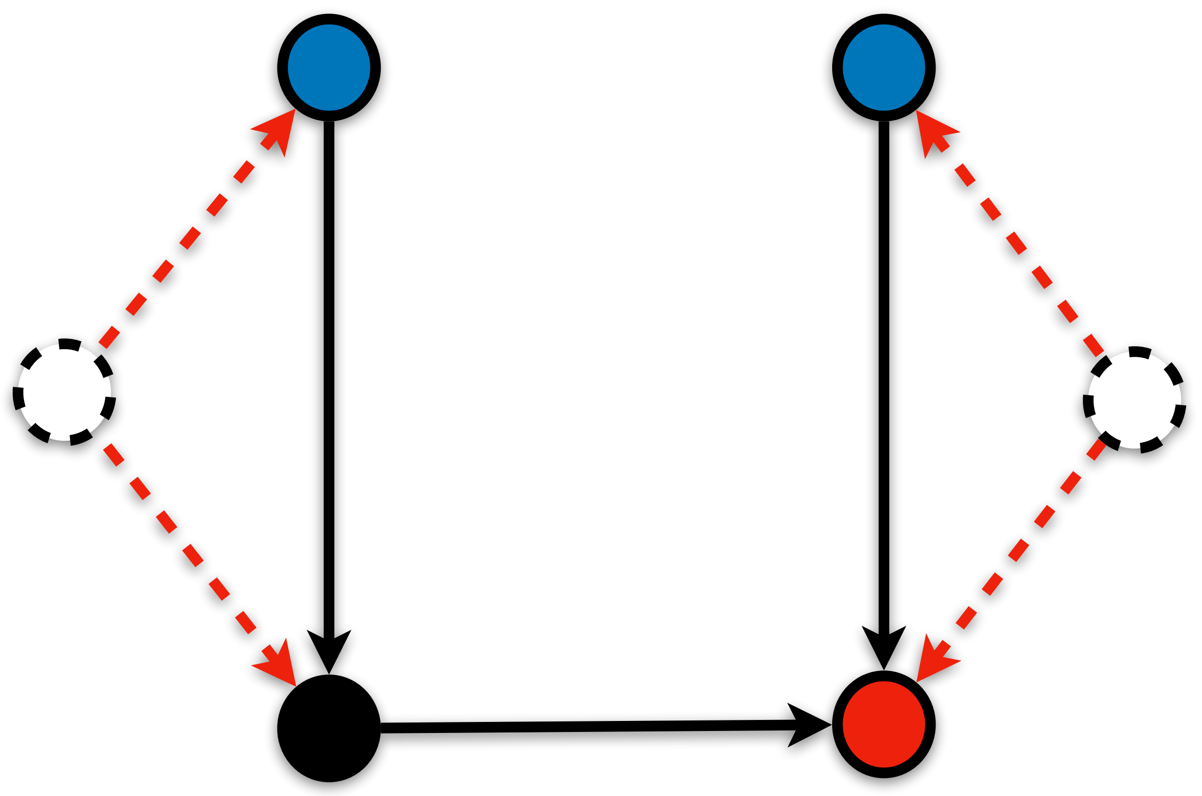
This Talk: Estimating Causal Effects

Tasks

1. From Observation



2. From Data Fusion



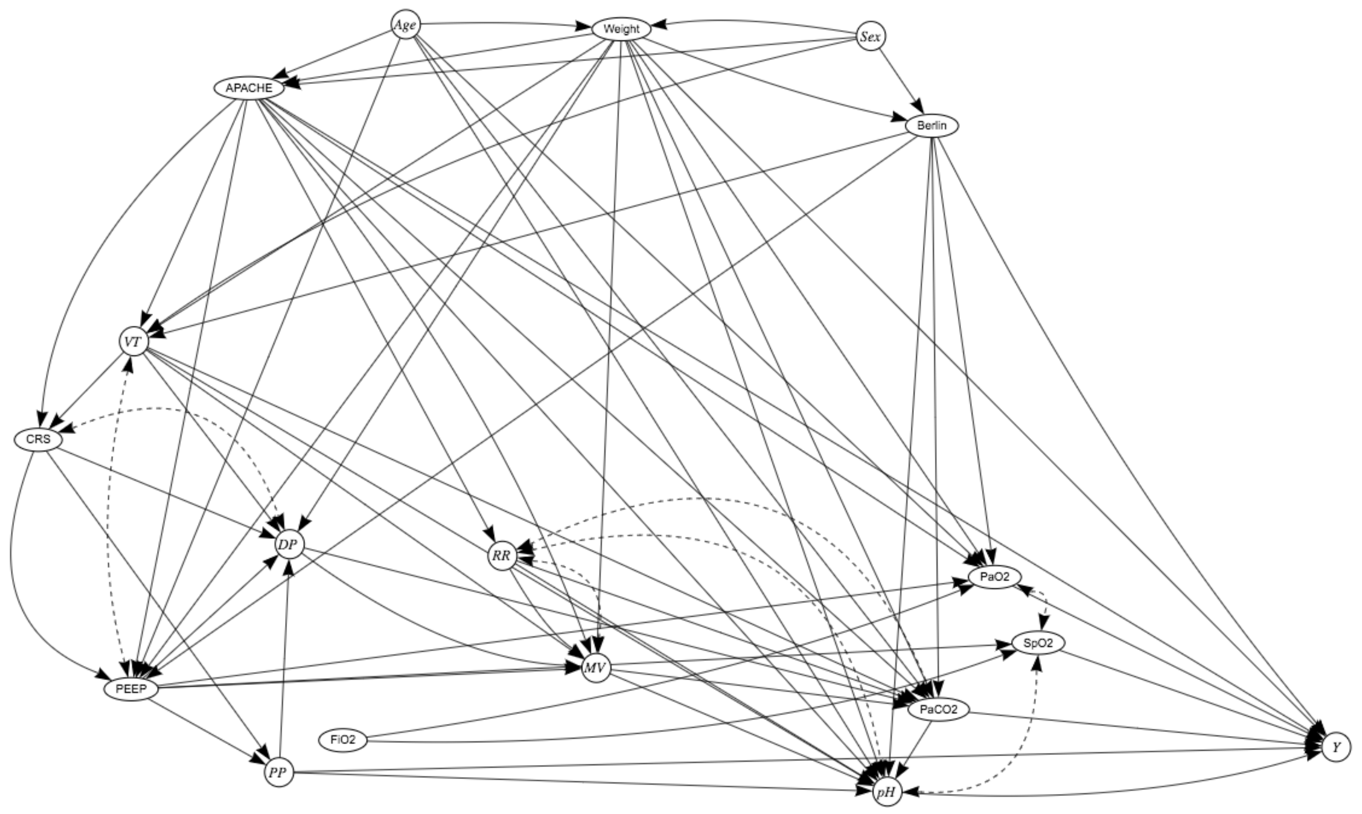
Solution

DML-ID

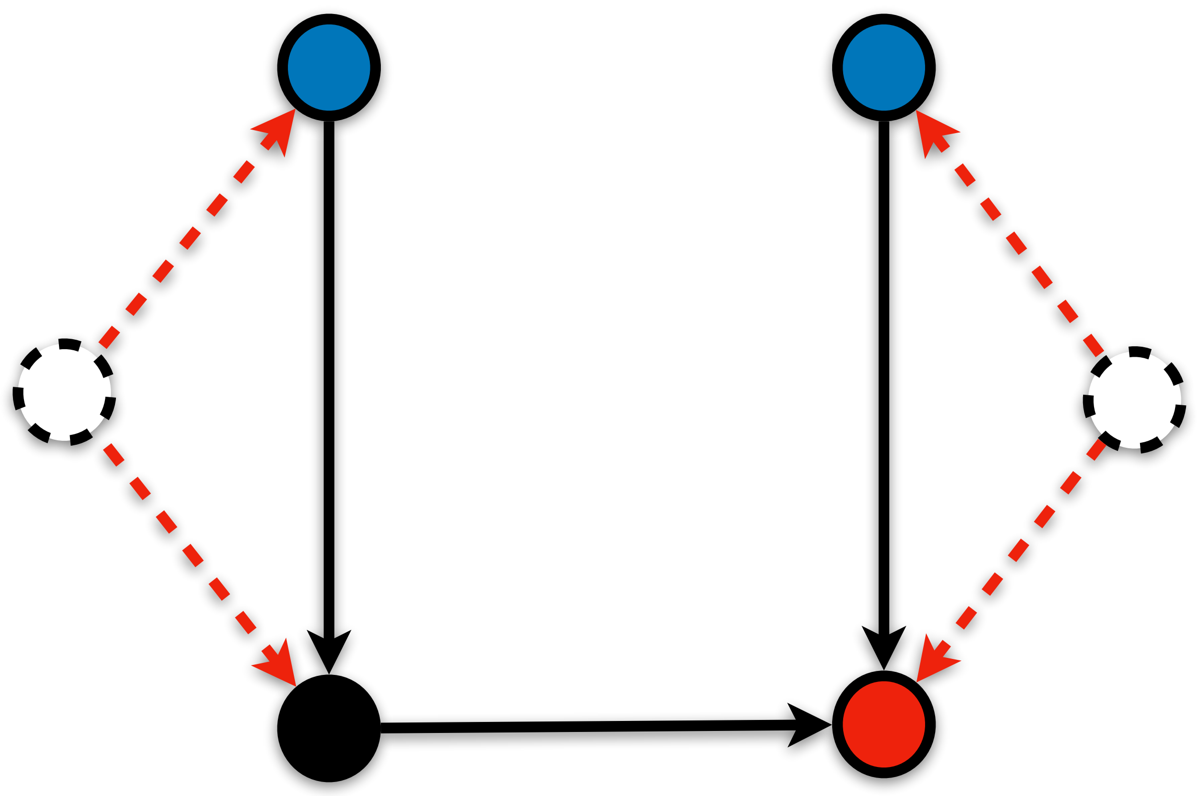
This Talk: Estimating Causal Effects

Tasks

1. From Observation



2. From Data Fusion



Solution

DML-ID

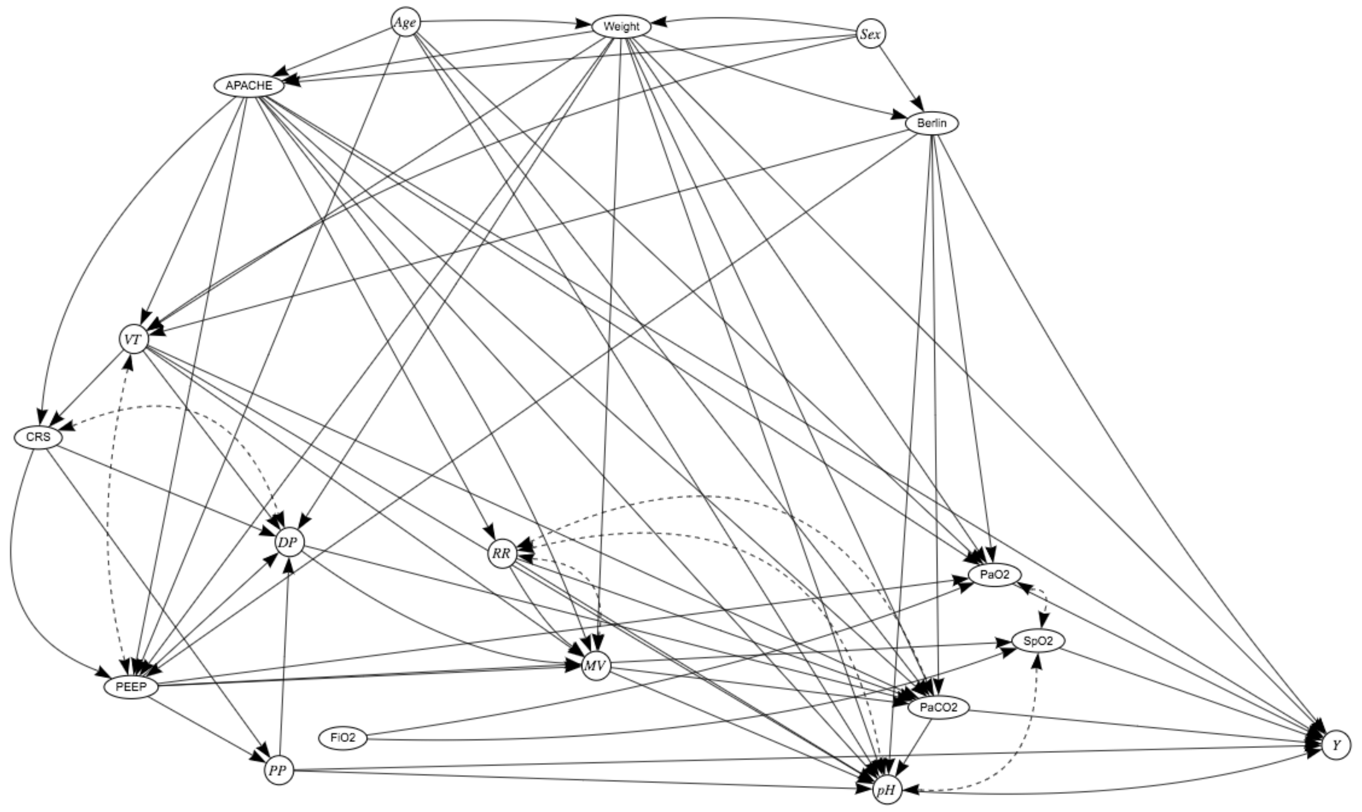
- DML-BD$^+$
- DML-gID

This Talk: Estimating Causal Effects

Tasks

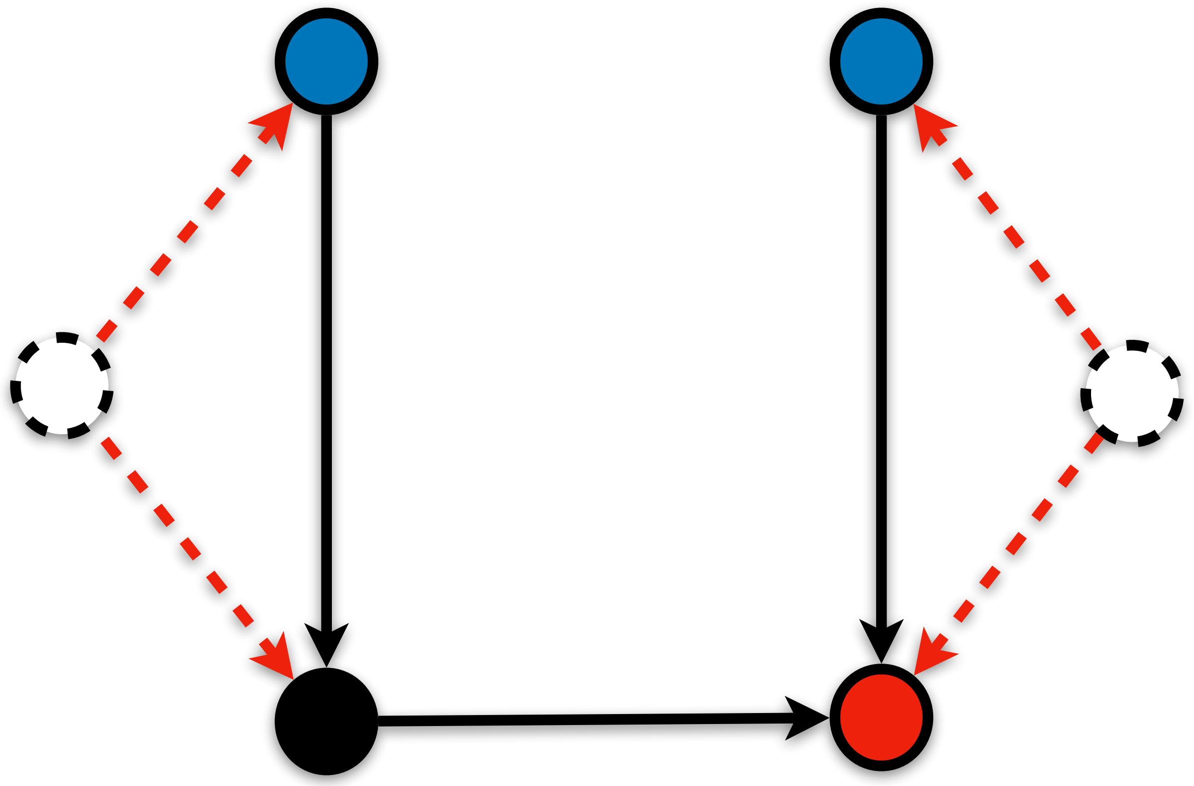
Solution

1. From Observation



DML-ID

2. From Data Fusion



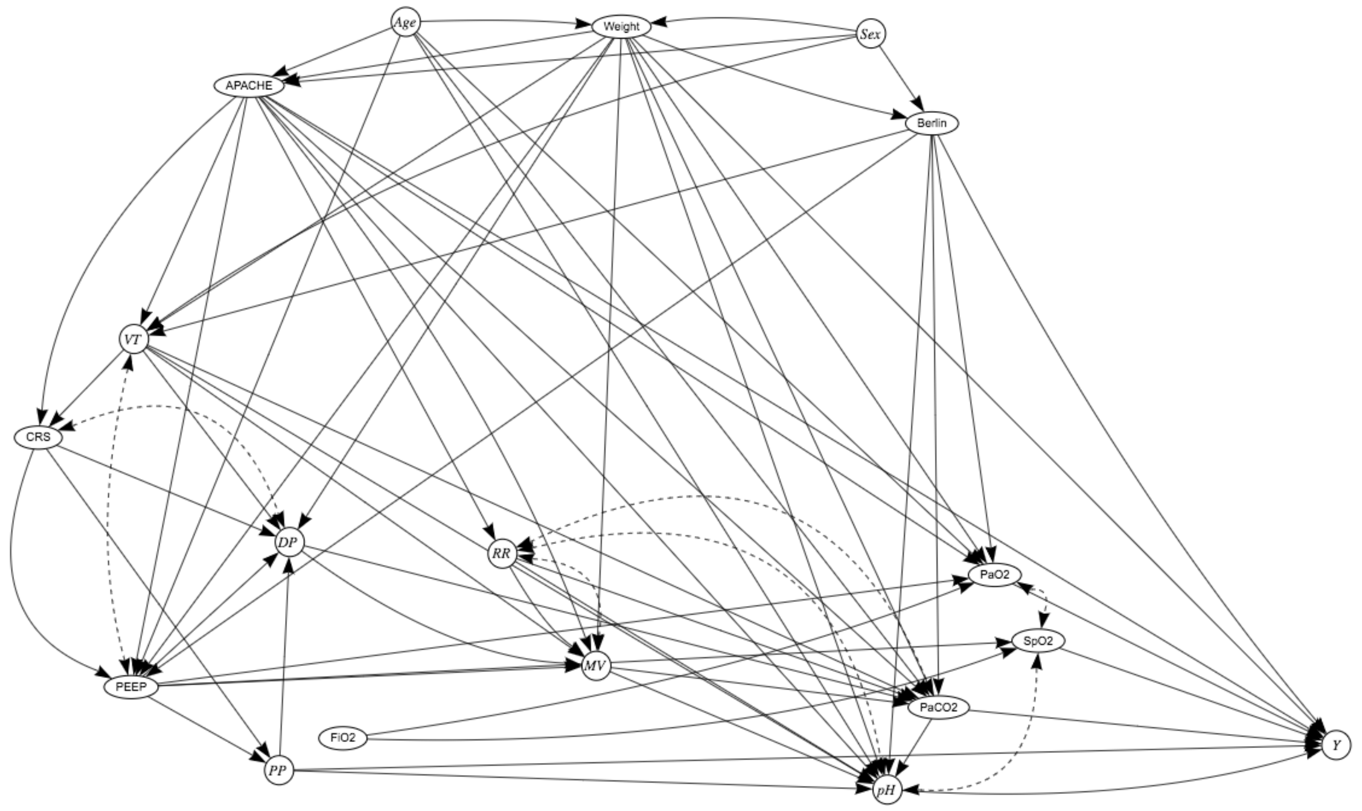
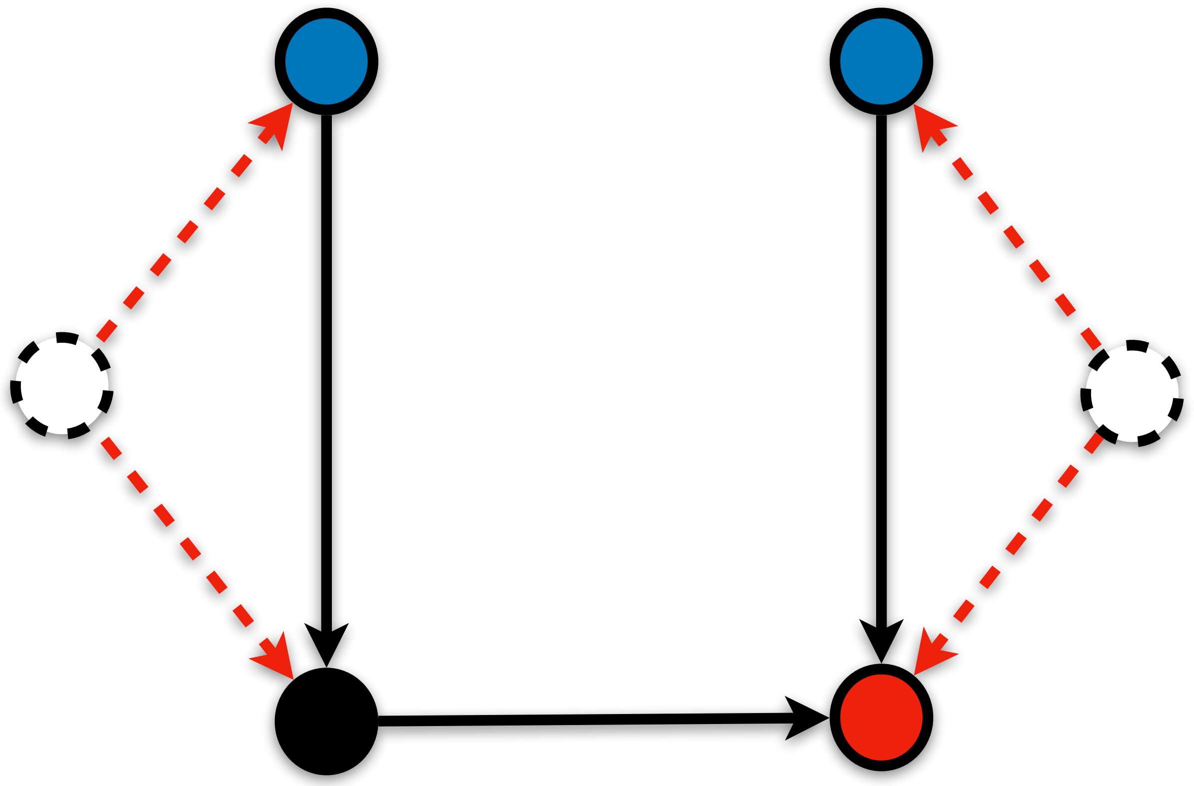
- DML-BD$^+$
- DML-gID

3. Scalable Estimation

FD

$$\sum_{z,x',c}\mathbb{E}[Y \mid z,x',c]P(z \mid x,c)P(x'c)$$

This Talk: Estimating Causal Effects

	1. From Observation	2. From Data Fusion	3. Scalable Estimation
Tasks			**FD** $\sum_{z,x',c} \mathbb{E}[Y \mid z,x',c]P(z \mid x,c)P(x'c)$
Solution	DML-ID	- DML-BD$^+$ - DML-gID	DML-MCE

Future Directions

Future Directions

1. **Causal Inference under Real-World Imperfections**

Future Directions

1. **Causal Inference under Real-World Imperfections**
    - unmeasured confounding, limited overlap, or high-dimensionality (e.g., images)

Future Directions

1. **Causal Inference under Real-World Imperfections**
    - unmeasured confounding, limited overlap, or high-dimensionality (e.g., images)

2. **Robust and Scalable Estimation**

Future Directions

1. **Causal Inference under Real-World Imperfections**
    - unmeasured confounding, limited overlap, or high-dimensionality (e.g., images)

2. **Robust and Scalable Estimation**
    - Inference using large-size pre-trained models.
    - computationally efficient method that scale to high-dimensional and large-scale datasets.

Future Directions

1. **Causal Inference under Real-World Imperfections**
    - unmeasured confounding, limited overlap, or high-dimensionality (e.g., images)

2. **Robust and Scalable Estimation**
    - Inference using large-size pre-trained models.
    - computationally efficient method that scale to high-dimensional and large-scale datasets.

3. **Causal AI for Diverse Modalities**

Future Directions

1. **Causal Inference under Real-World Imperfections**
    - unmeasured confounding, limited overlap, or high-dimensionality (e.g., images)

2. **Robust and Scalable Estimation**
    - Inference using large-size pre-trained models.
    - computationally efficient method that scale to high-dimensional and large-scale datasets.

3. **Causal AI for Diverse Modalities**
    - Inference for multi-modal covariates, treatments, and outcomes.

Thank you

www.yonghanjung.me/

Appendix

Logistics

Logistics

- **Professor Neville**
  - Please initiate and sign “Form 11: Report of the Final Examination”
  - Please approve the “Form 9: Electronic Thesis Acceptance Form (ETAF)” after reviewing the thesis.

Logistics

- **Professor Neville**
  - Please initiate and sign "Form 11: Report of the Final Examination"
  - Please approve the "Form 9: Electronic Thesis Acceptance Form (ETAF)" after reviewing the thesis.
- **Other Professors**
  - Please sign "Form 11: Report of the Final Examination"
  - Please approve the "Form 9: Electronic Thesis Acceptance Form (ETAF)" after reviewing the thesis.

Logistics

- **Professor Neville**
  - Please initiate and sign “Form 11: Report of the Final Examination”
  - Please approve the “Form 9: Electronic Thesis Acceptance Form (ETAF)” after reviewing the thesis.
- **Other Professors**
  - Please sign “Form 11: Report of the Final Examination”
  - Please approve the “Form 9: Electronic Thesis Acceptance Form (ETAF)” after reviewing the thesis.

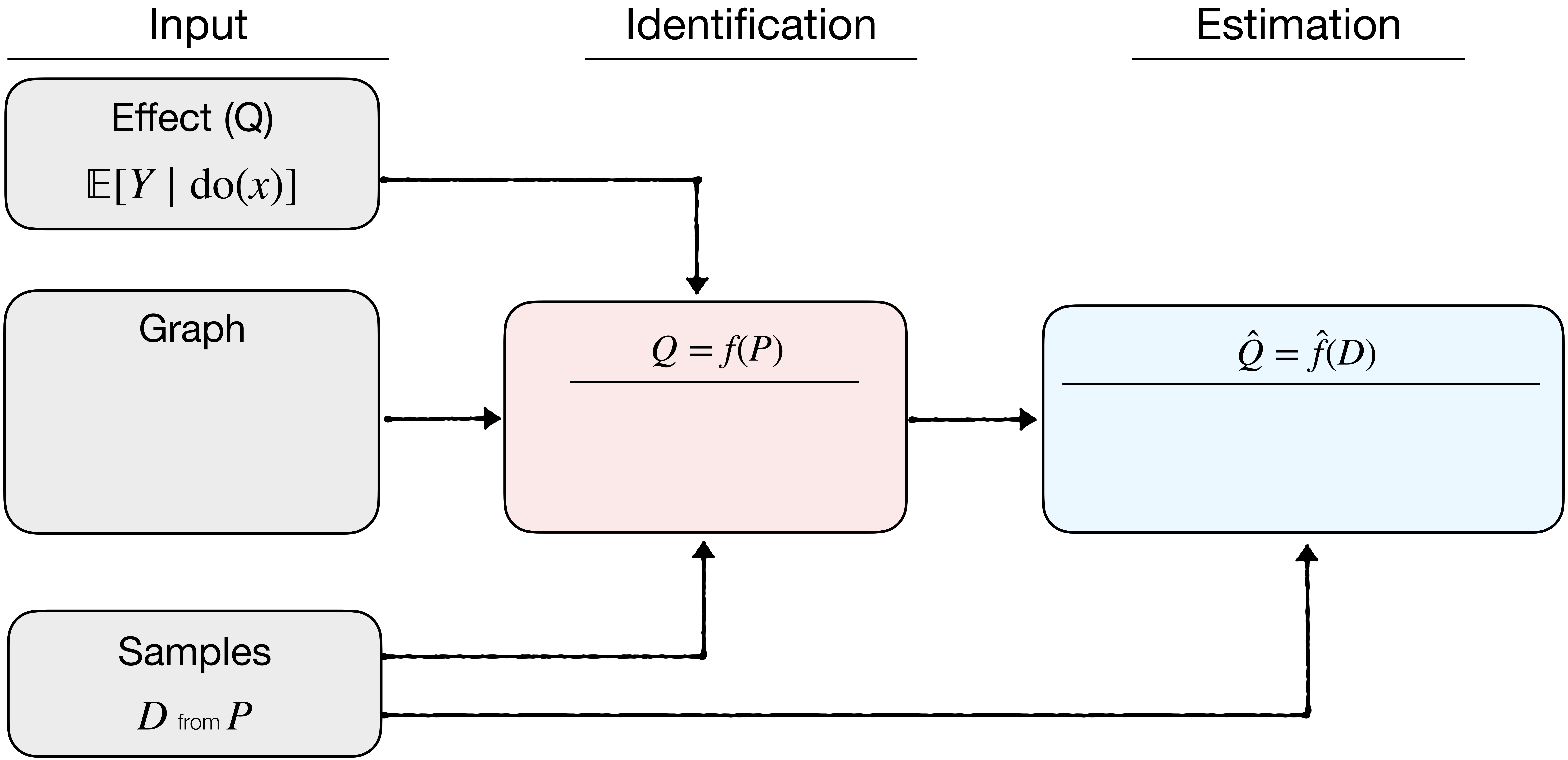
I kindly ask that you complete these by **June 12** to meet the PhD completion deadline for my next job appointment — Assistant Professor at UIUC's School of Information Sciences.

Logistics

- Please initiate and sign “Form 11: Report of the Final Examination”
- Please approve the “Form 9: Electronic Thesis Acceptance Form (ETAF)” after reviewing the thesis.

Omitted Works

Other Work 1: Causal inference Without Graphs



Input

Identification

Estimation

Effect (Q)

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

Graph

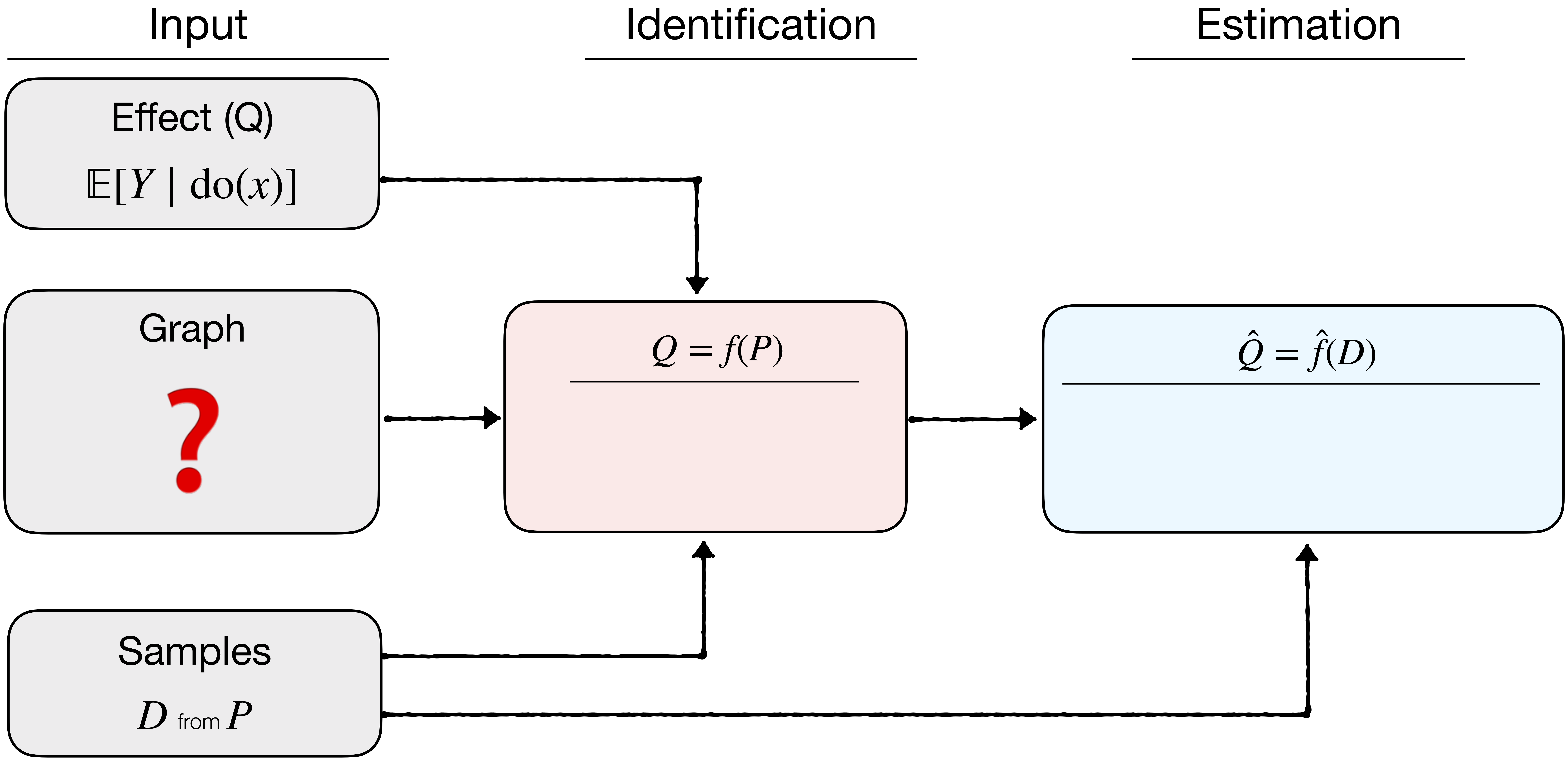
$Q = f(P)$

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

Samples

D from P

Other Work 1: Causal inference Without Graphs



Input

Identification

Estimation

Effect (Q)

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

Graph

?

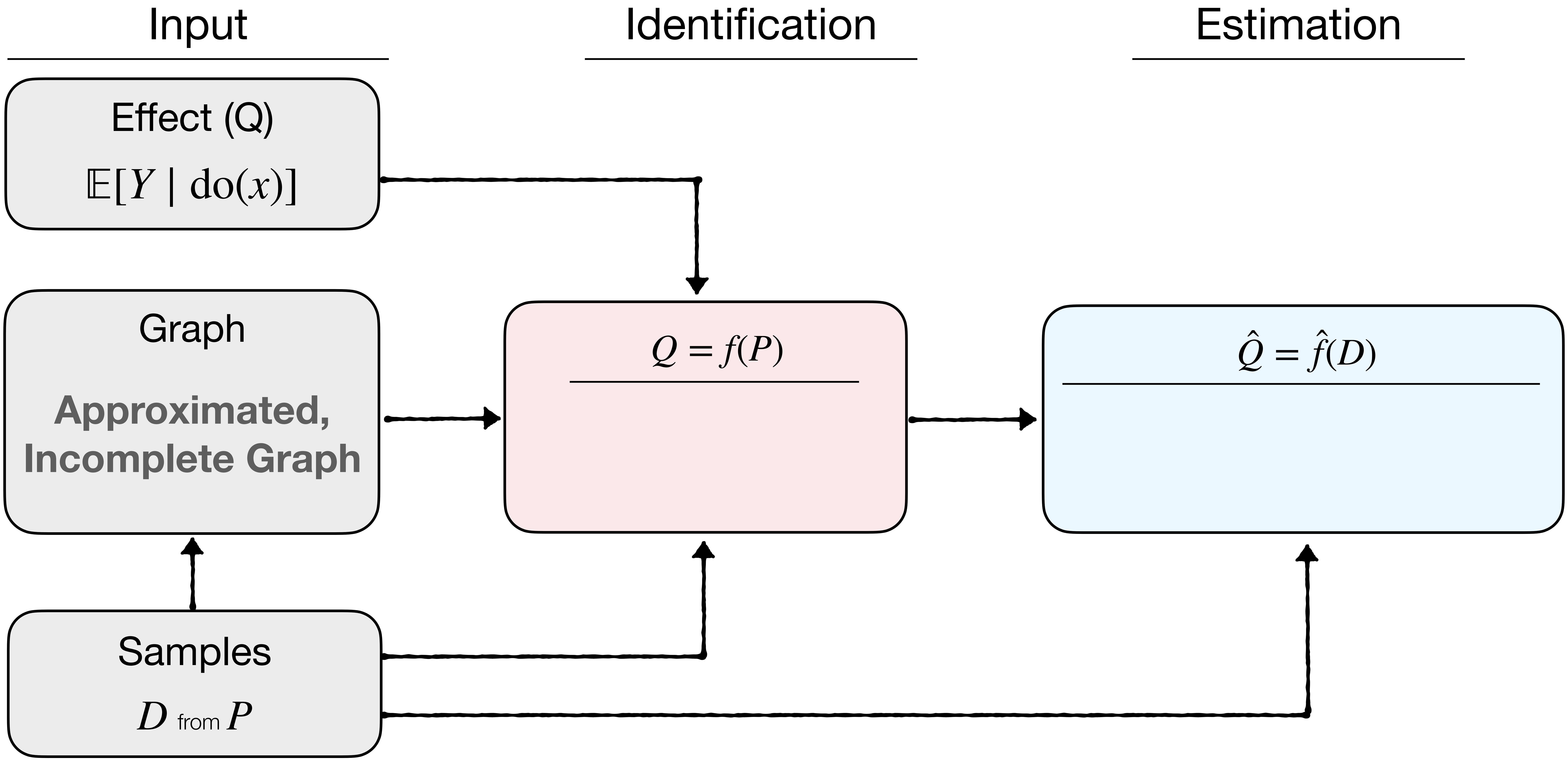
Samples

D from P

$Q = f(P)$

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

Other Work 1: Causal inference Without Graphs



Input

Identification

Estimation

Effect (Q)

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

Graph

Approximated, Incomplete Graph

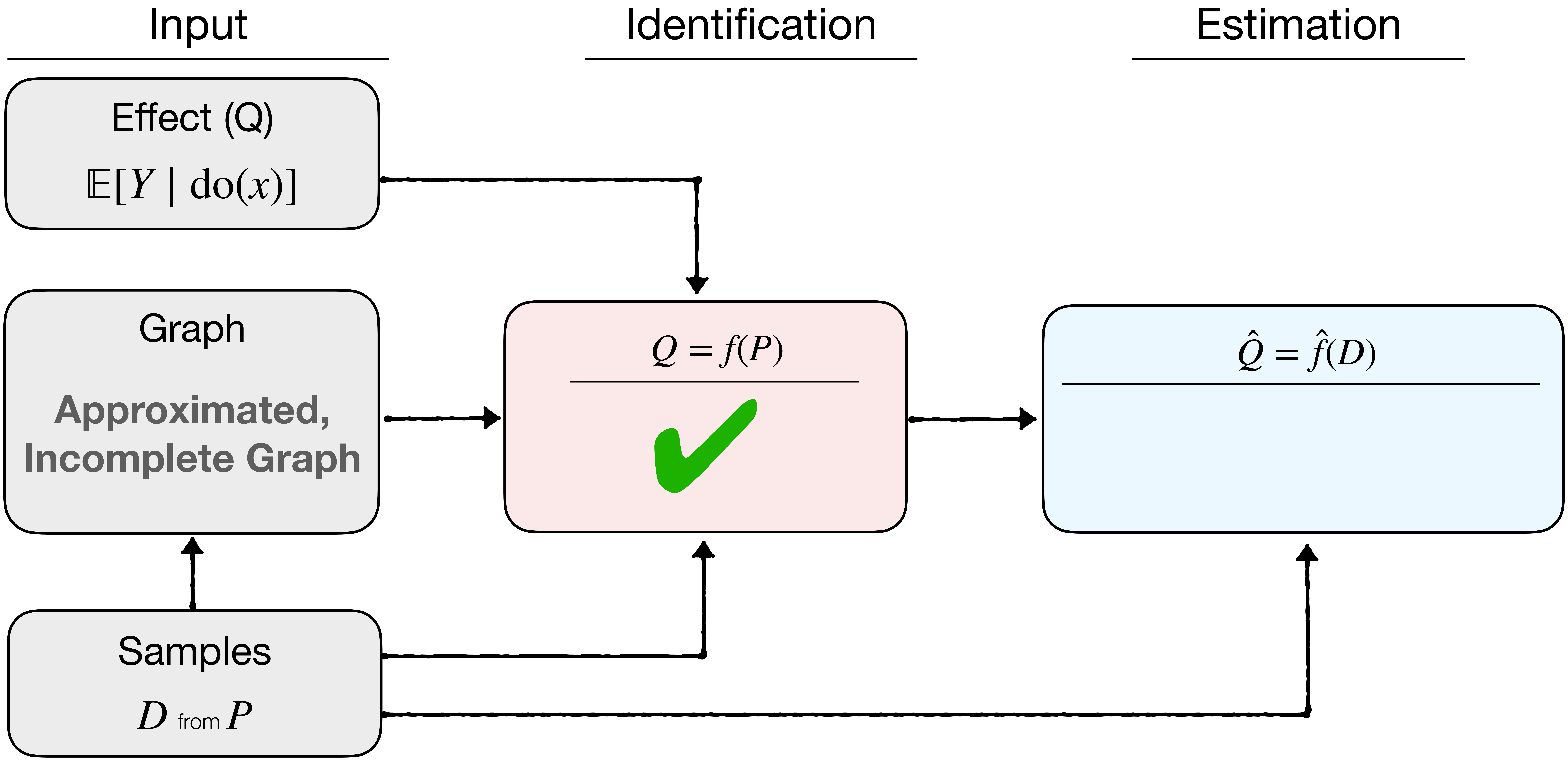
Samples

D from P

$Q = f(P)$

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

Other Work 1: Causal inference Without Graphs



Input

Identification

Estimation

Effect (Q)

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

Graph

Approximated, Incomplete Graph

Samples

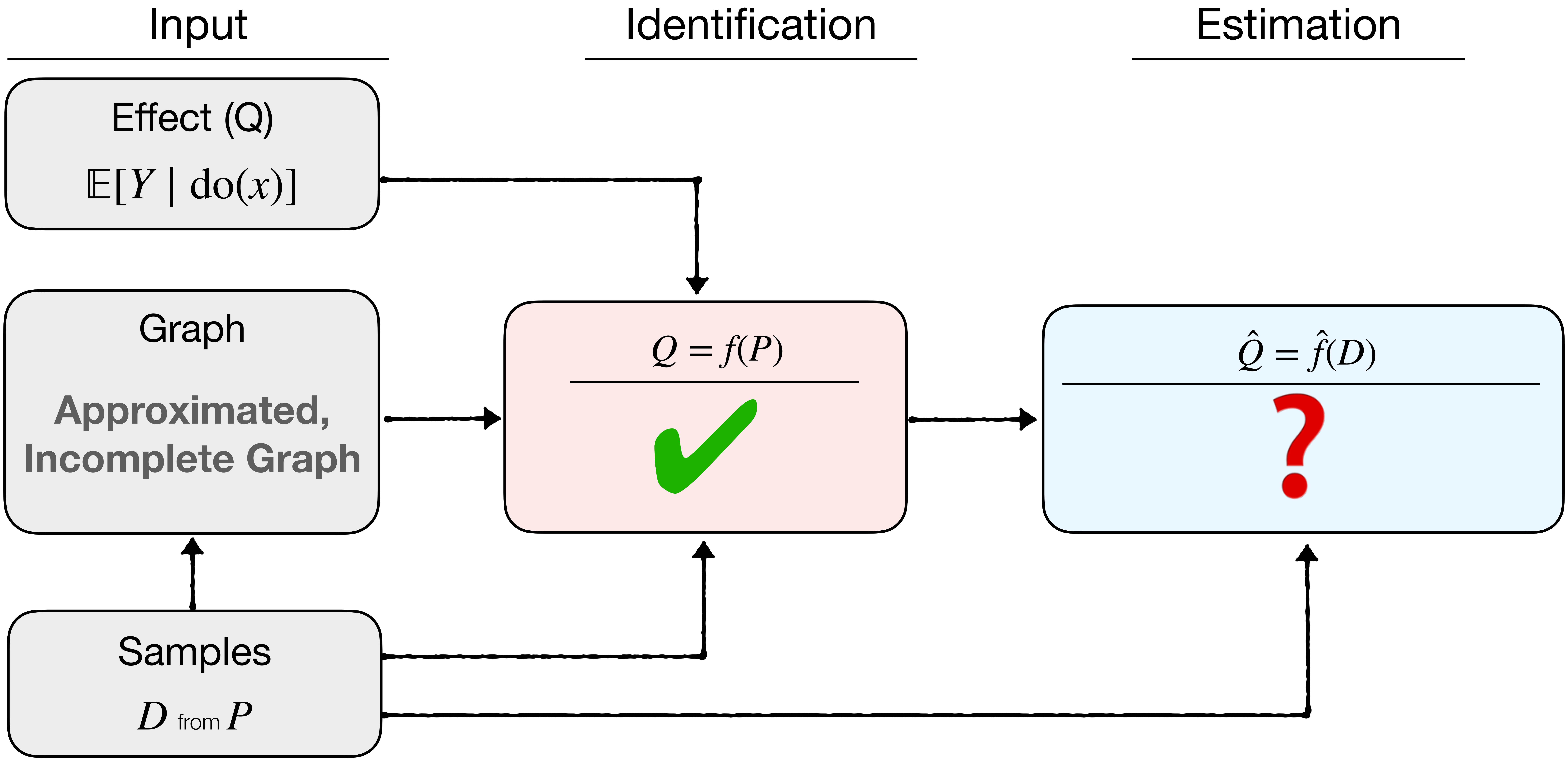
D from P

$Q = f(P)$

✓

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

Other Work 1: Causal inference Without Graphs



Input

Identification

Estimation

Effect (Q)

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

Graph

Approximated, Incomplete Graph

Samples

D from P

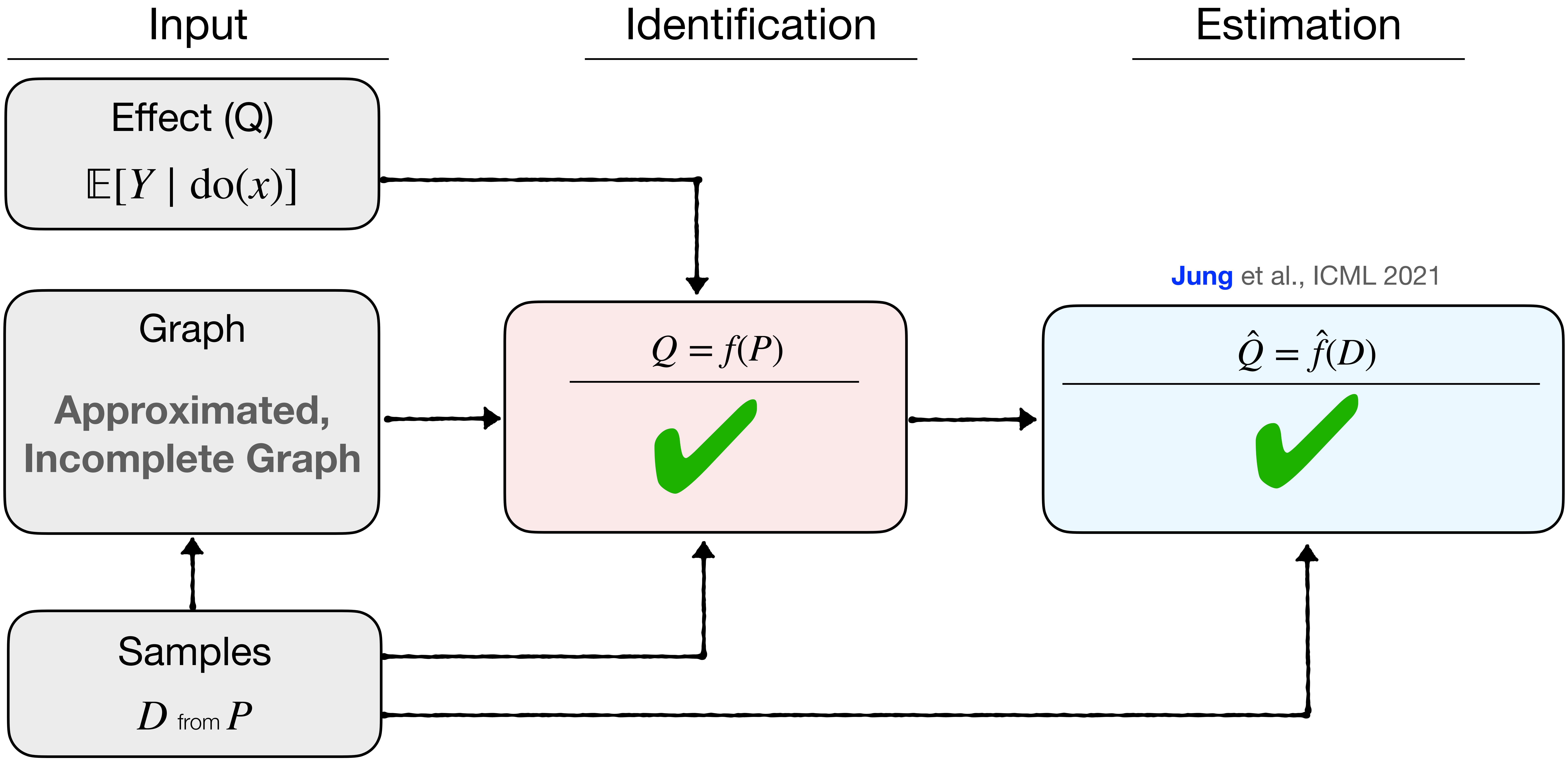
$Q = f(P)$

✓

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

?

Other Work 1: Causal inference Without Graphs



Input

Effect (Q)

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

Graph

Approximated, Incomplete Graph

Samples

D from P

Identification

$Q = f(P)$

✓

Estimation

Jung et al., ICML 2021

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

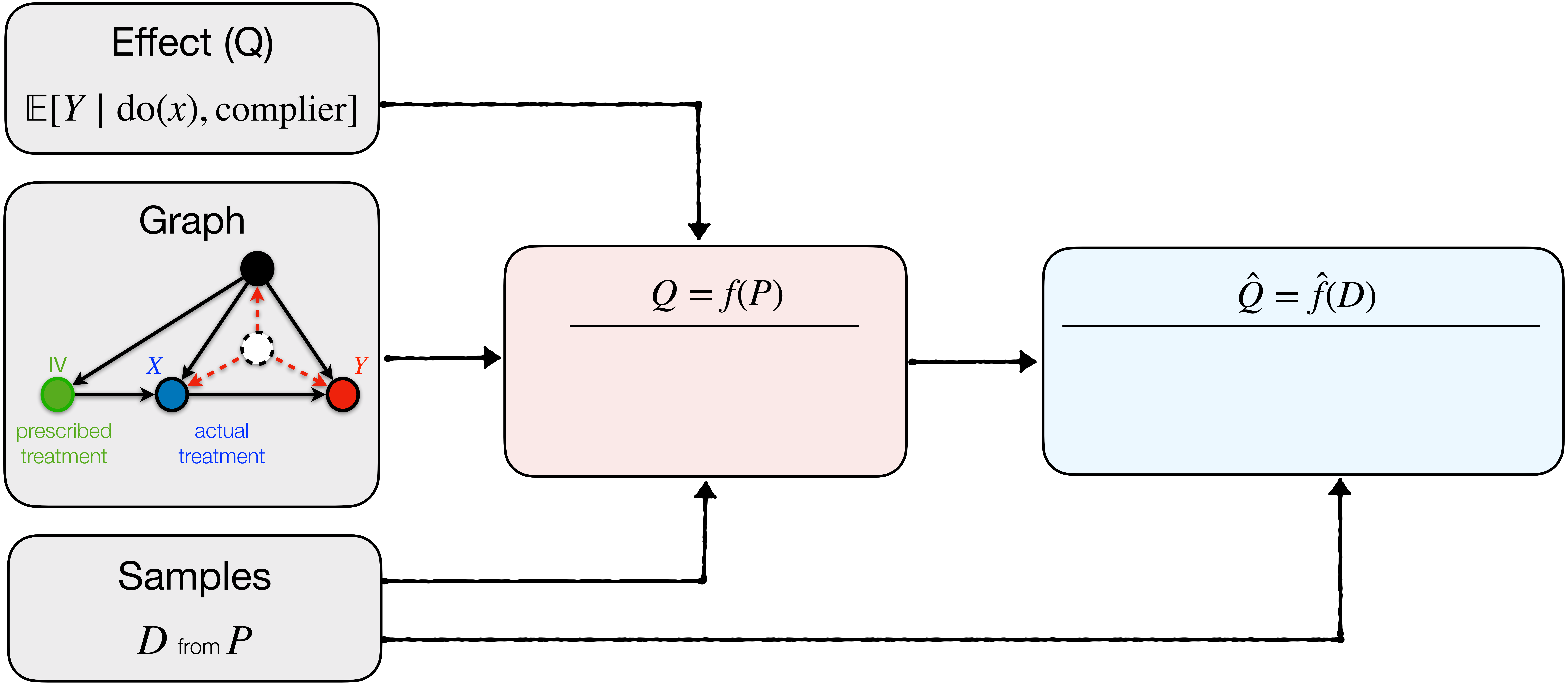
✓

Other Work 2: Instrumental Variable (IV) Analysis

Input

Identification

Estimation



Effect (Q)

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

Graph

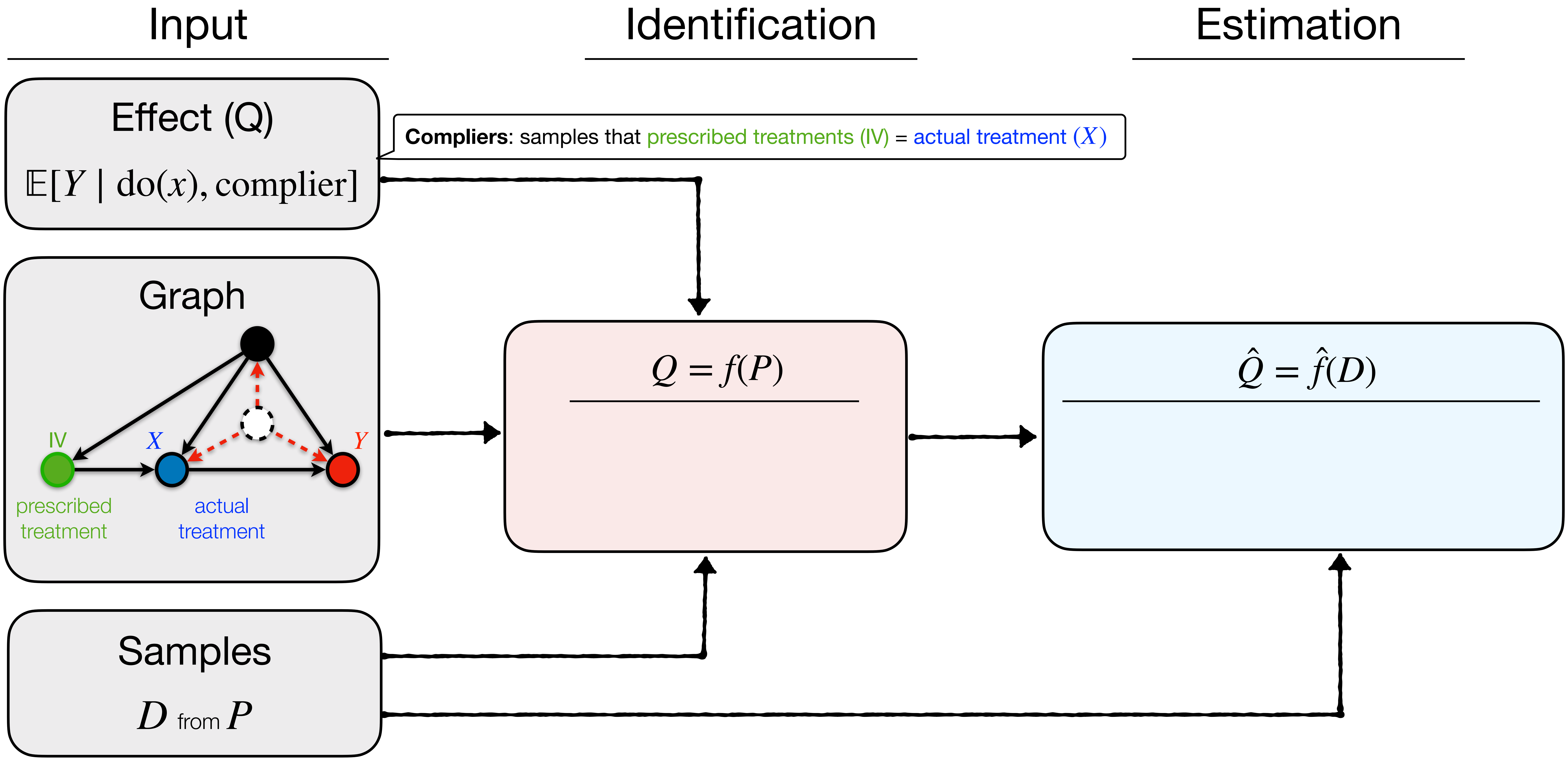
Samples

D from P

$Q = f(P)$

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

Other Work 2: Instrumental Variable (IV) Analysis



Input

Identification

Estimation

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

$Q = f(P)$

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

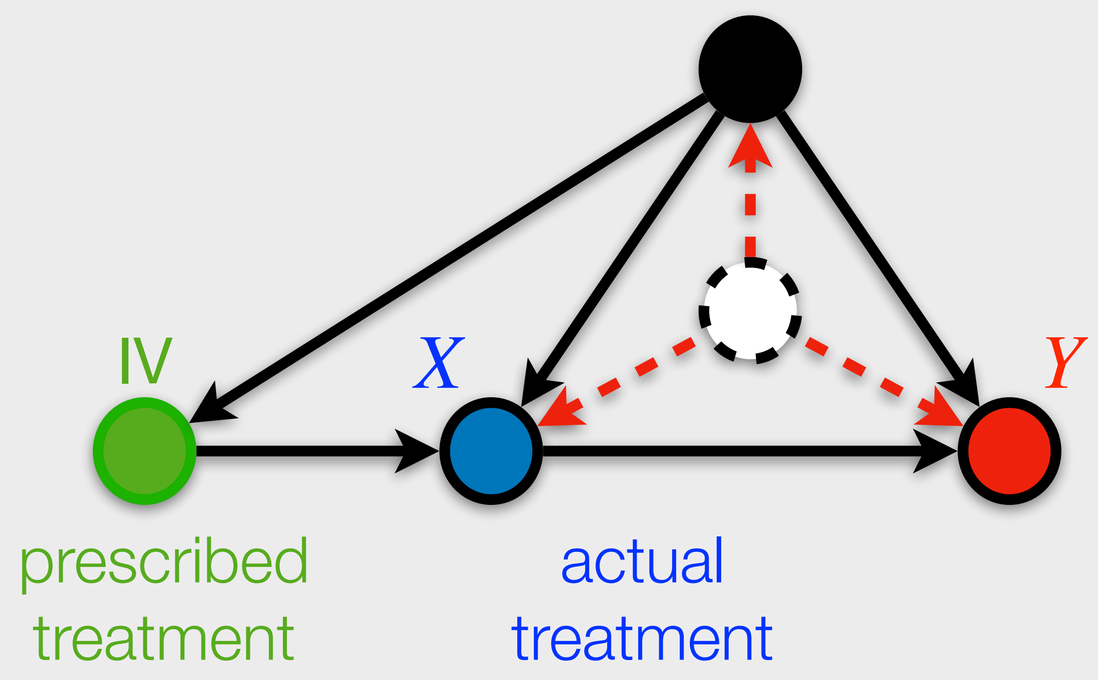
Other Work 2: Instrumental Variable (IV) Analysis

Input

Effect (Q)

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

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

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

✓

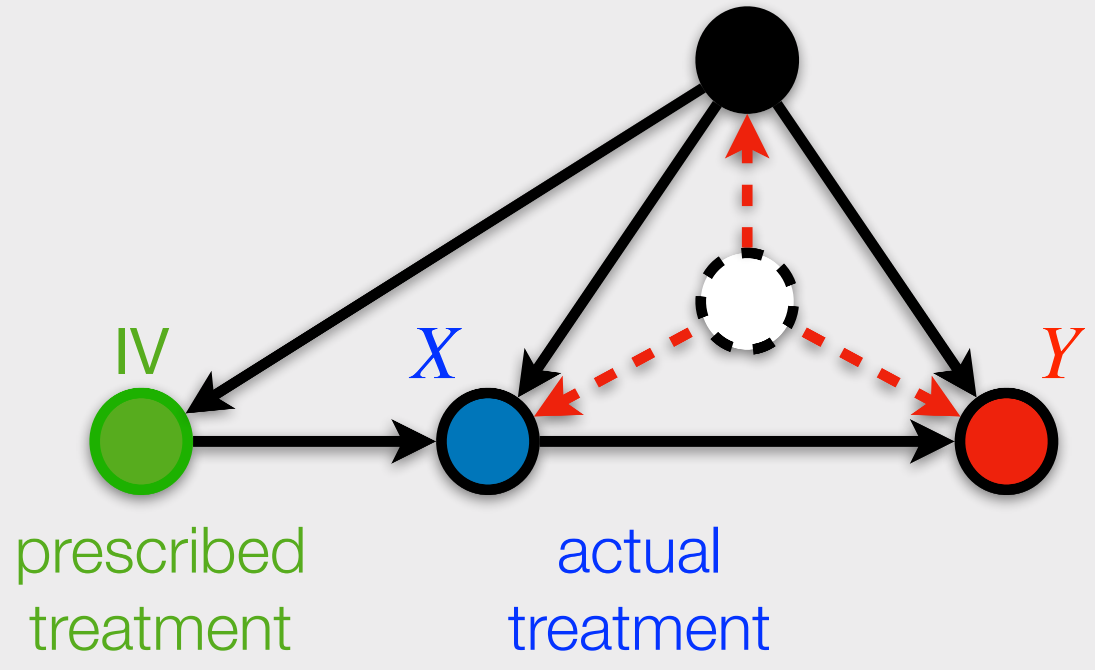
Other Work 2: Instrumental Variable (IV) Analysis

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

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

?

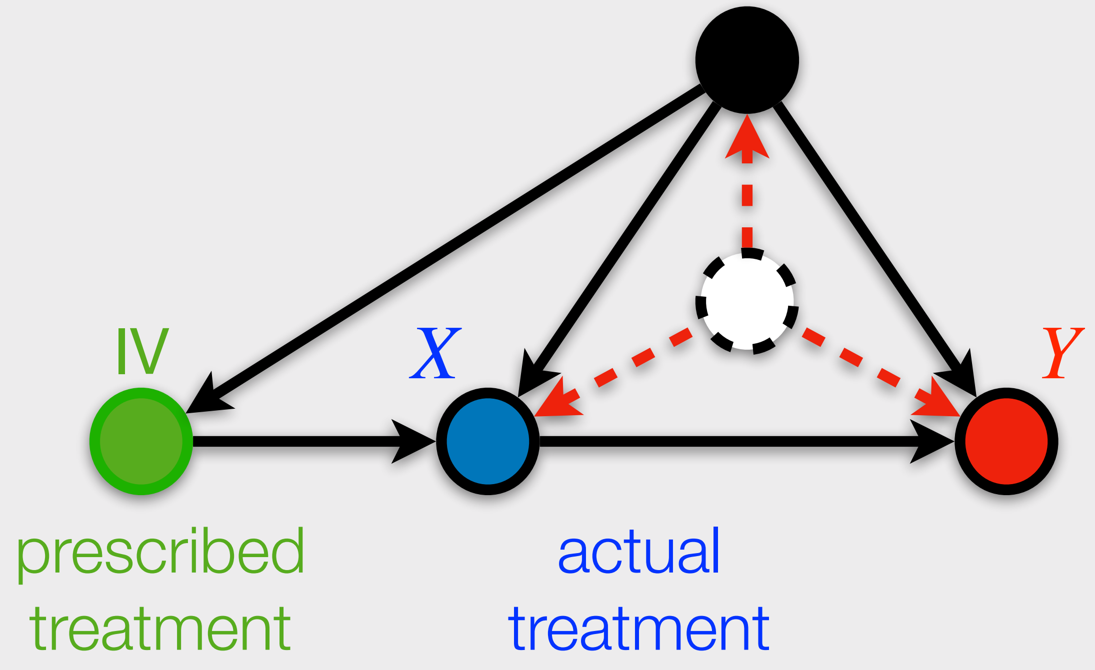
Other Work 2: Instrumental Variable (IV) Analysis

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

Jung et al., NeurIPS 2021 (Spotlight)

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

✓

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)

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

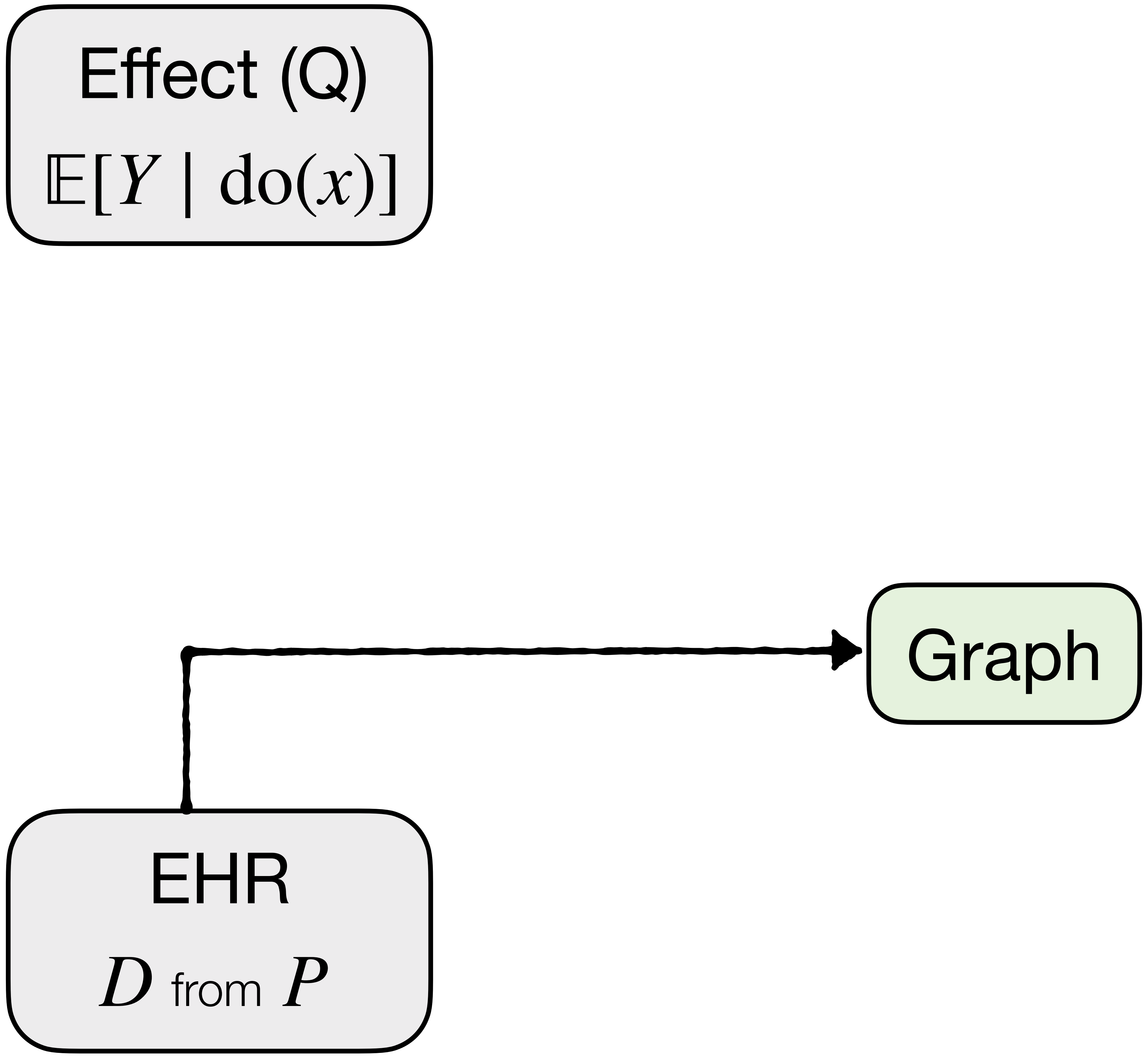
EHR

D from P

Application 1. Emulating RCT from EHR

Input

Graph Discovery



Effect (Q)

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

EHR

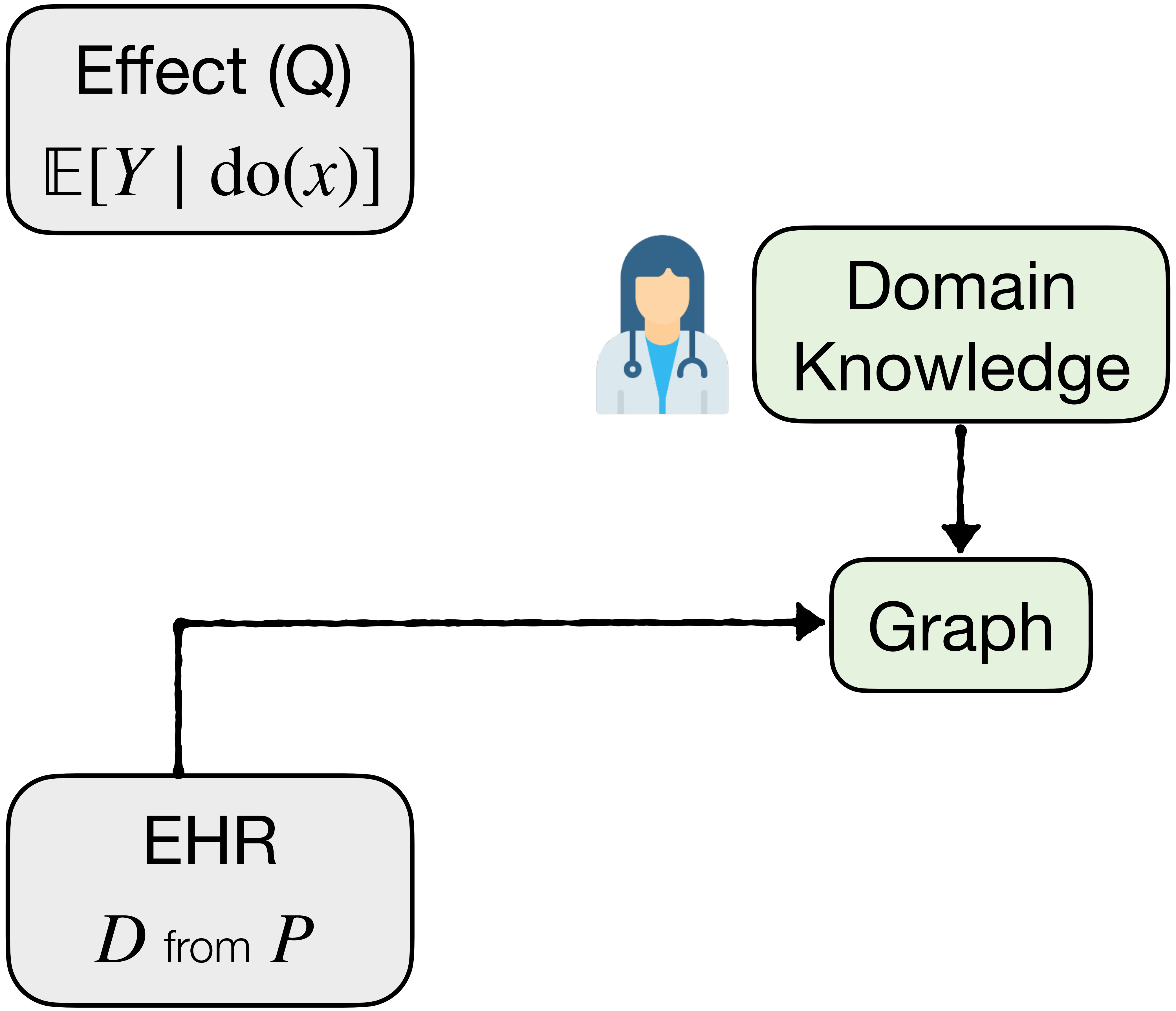
D from P

Graph

Application 1. Emulating RCT from EHR

Input

Graph Discovery



Effect (Q)

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

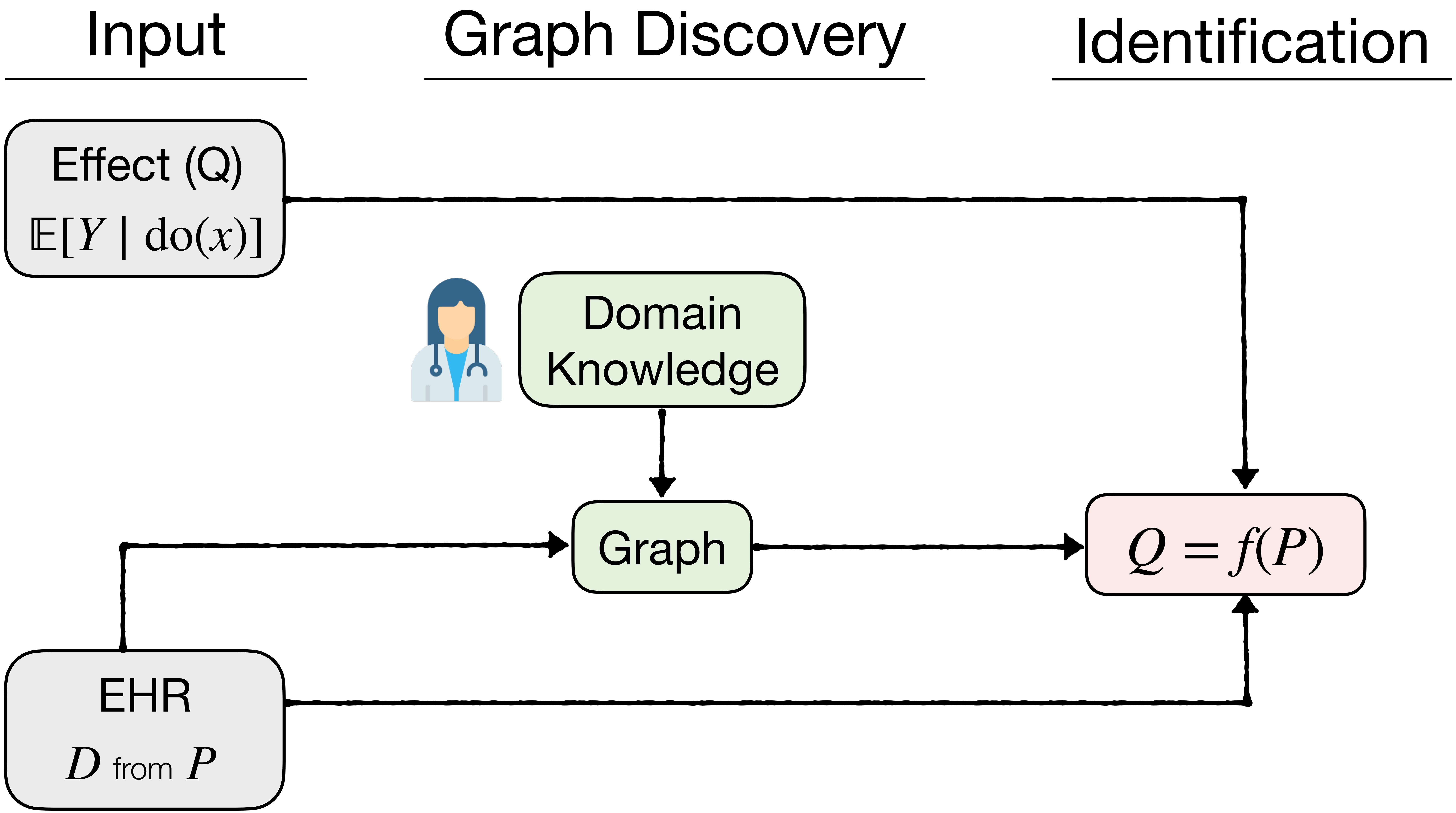
Domain Knowledge

Graph

EHR

D from P

Application 1. Emulating RCT from EHR



Input

Graph Discovery

Identification

Effect (Q)

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

Domain Knowledge

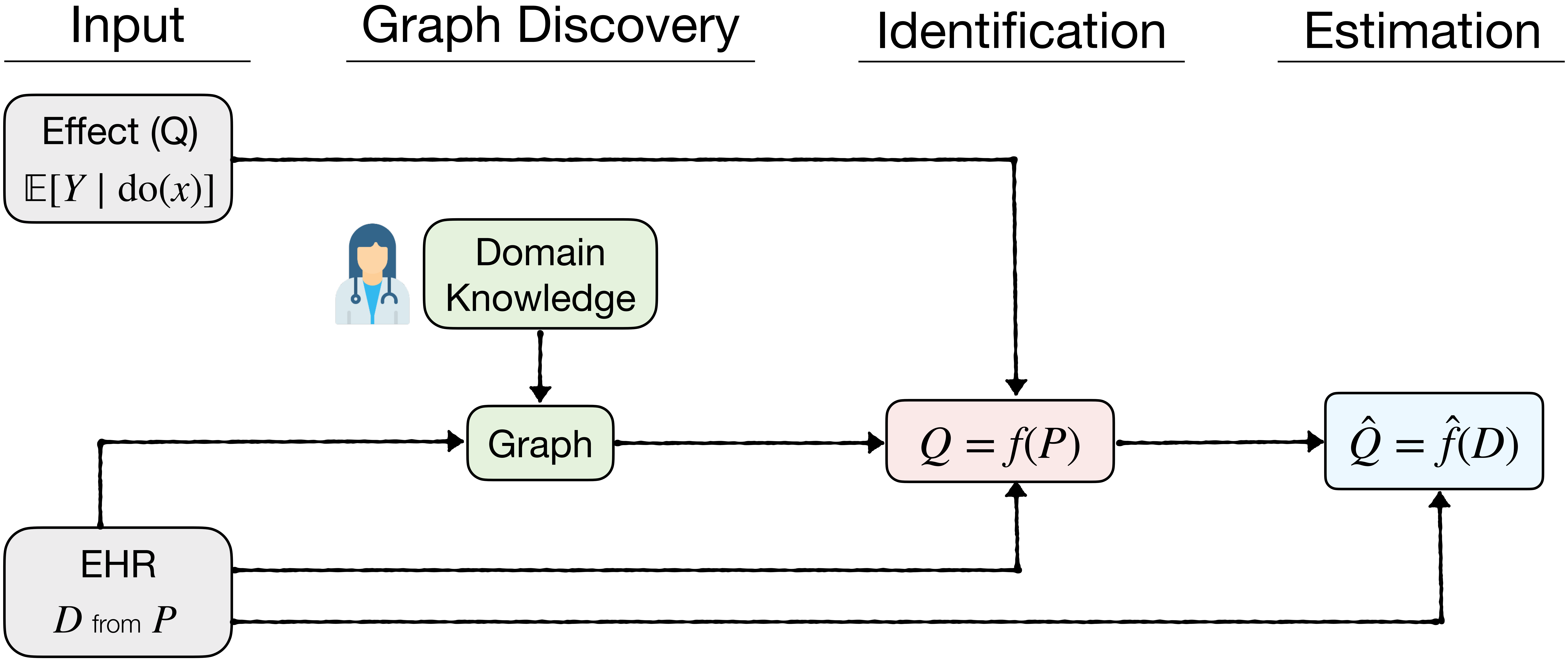
Graph

$Q = f(P)$

EHR

D from P

Application 1. Emulating RCT from EHR



Input

Graph Discovery

Identification

Estimation

Effect (Q)

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

Domain Knowledge

Graph

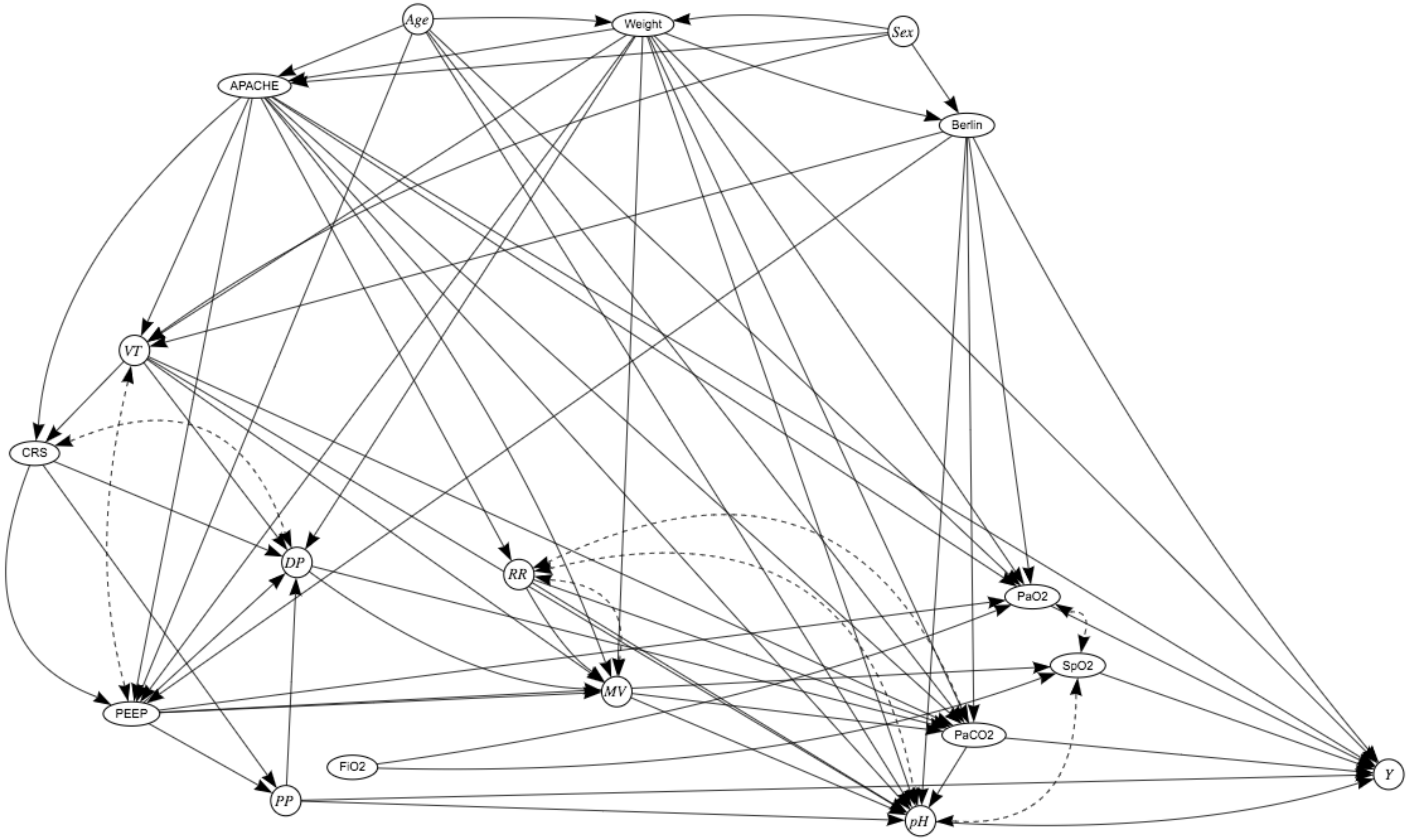
$Q = f(P)$

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

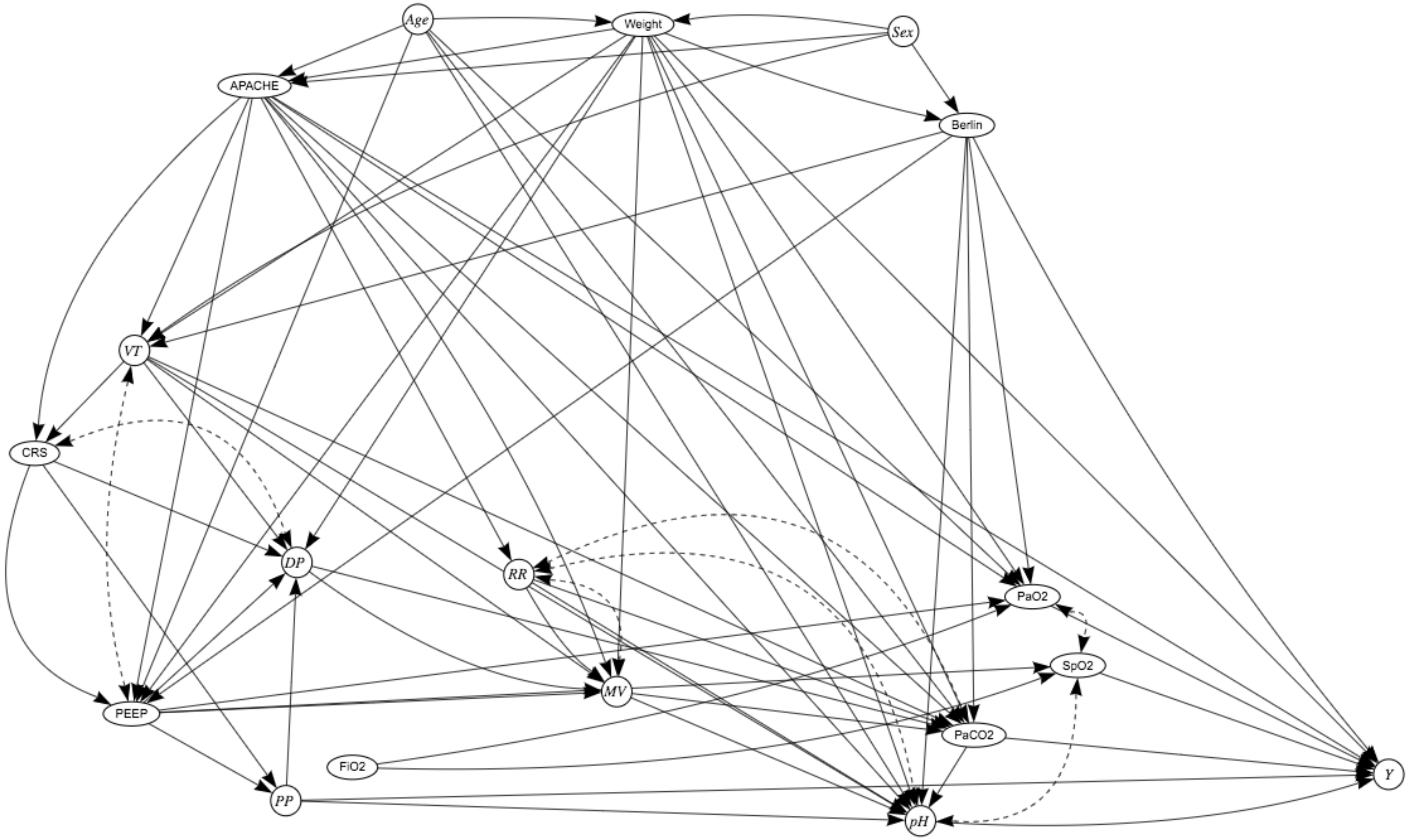
EHR

D from P

Application 1. Emulating RCT from EHR



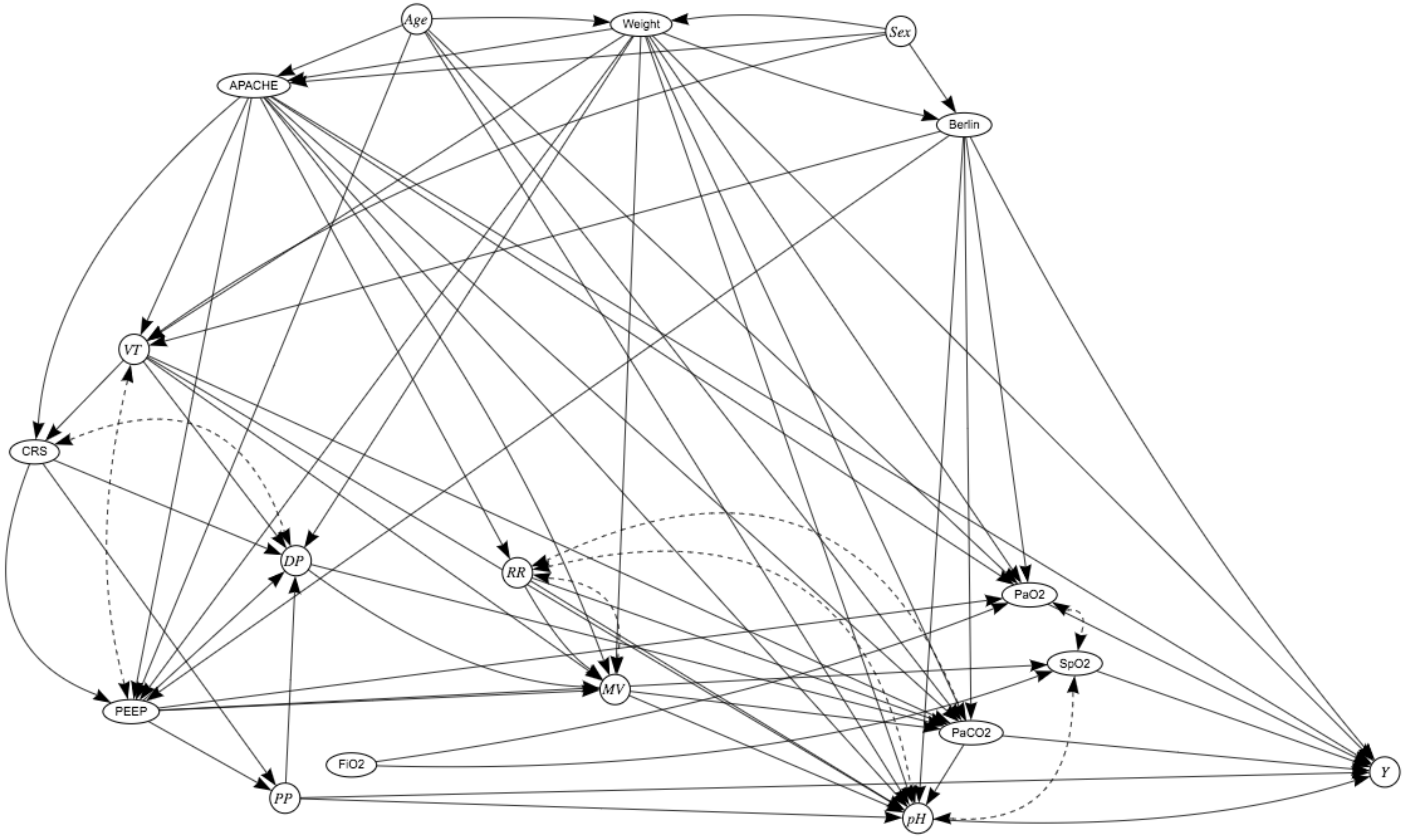
Application 1. Emulating RCT from EHR



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

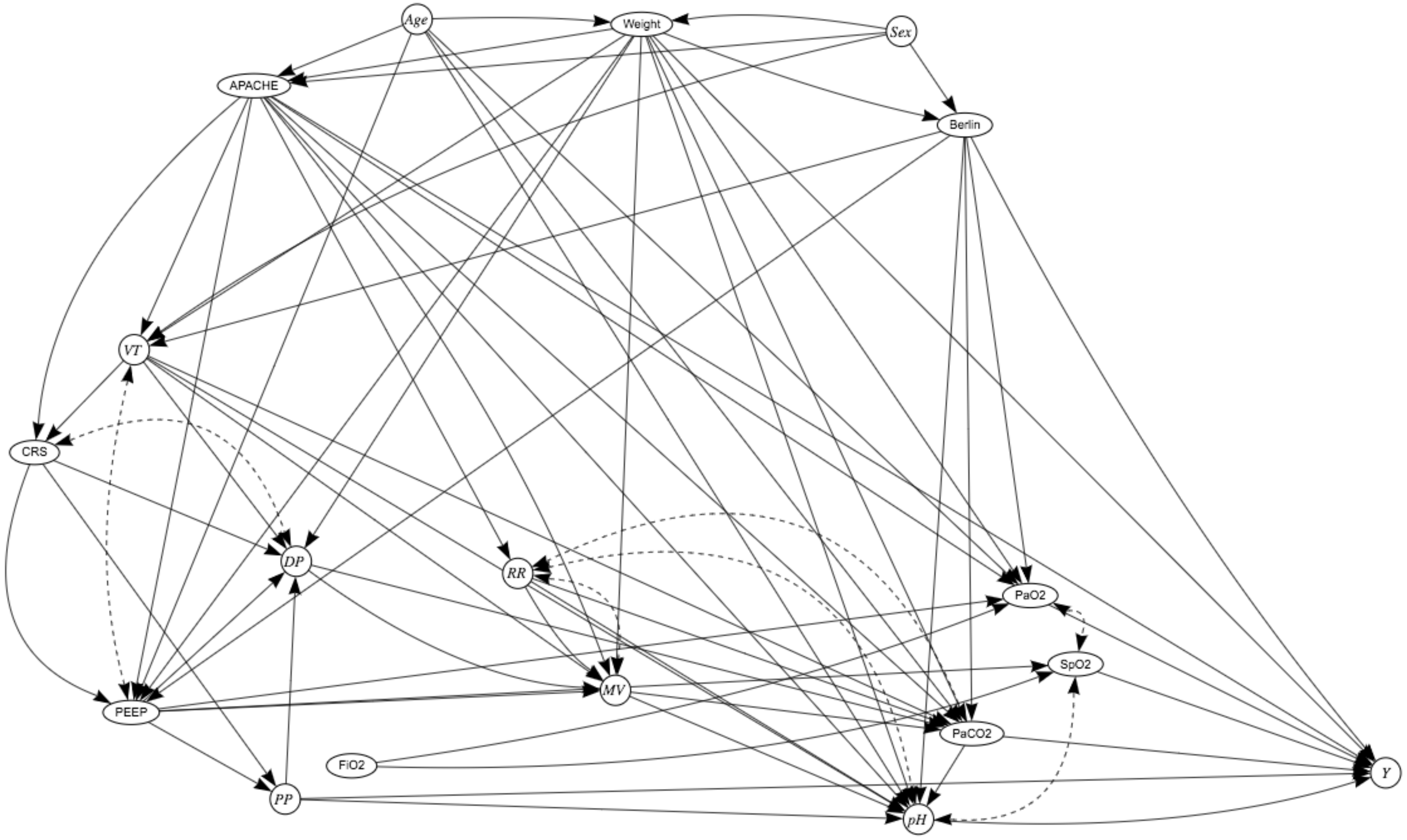
For seminal RCTs,

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

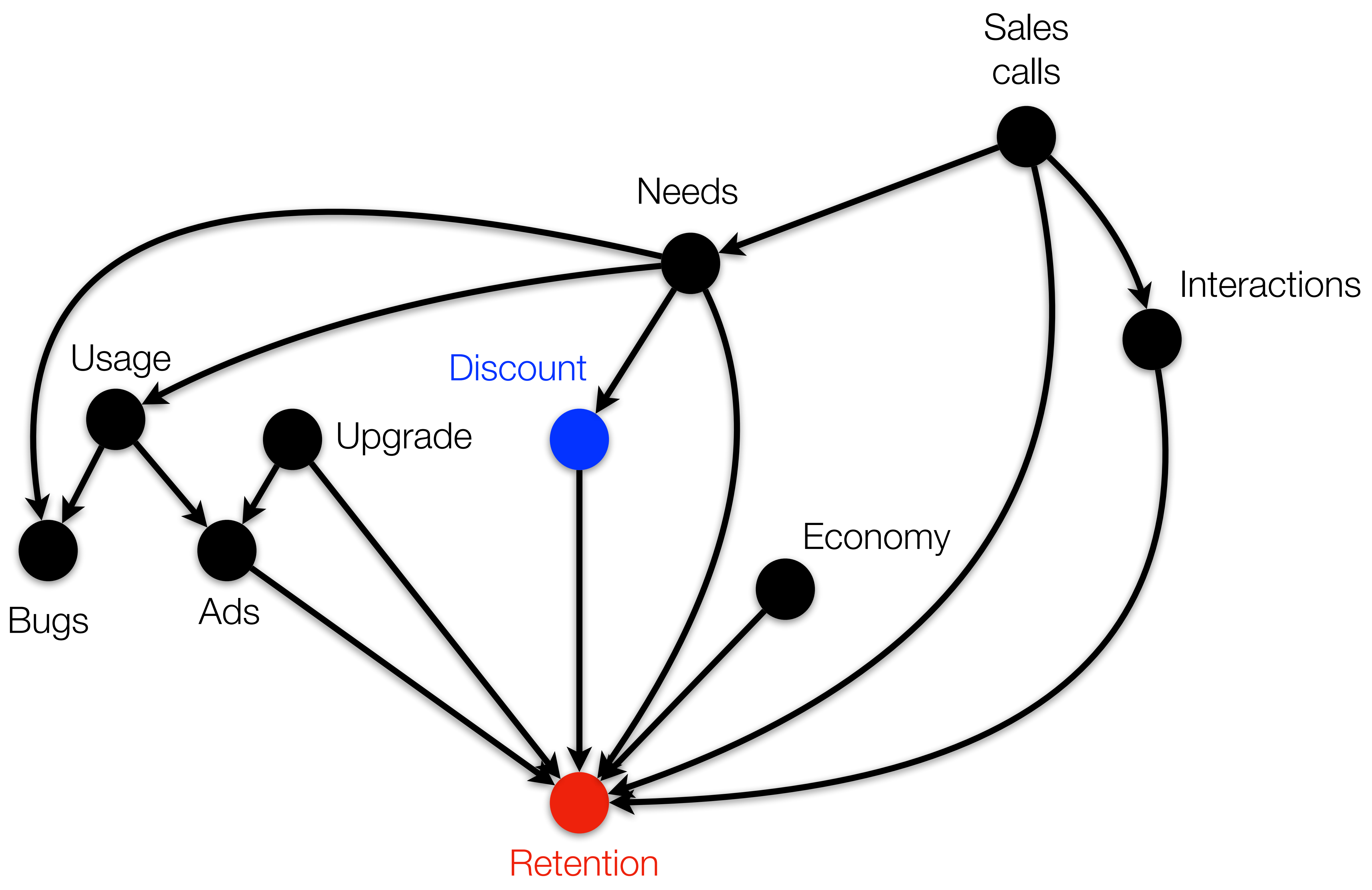
= Trials' treatment recommendation

Impact

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

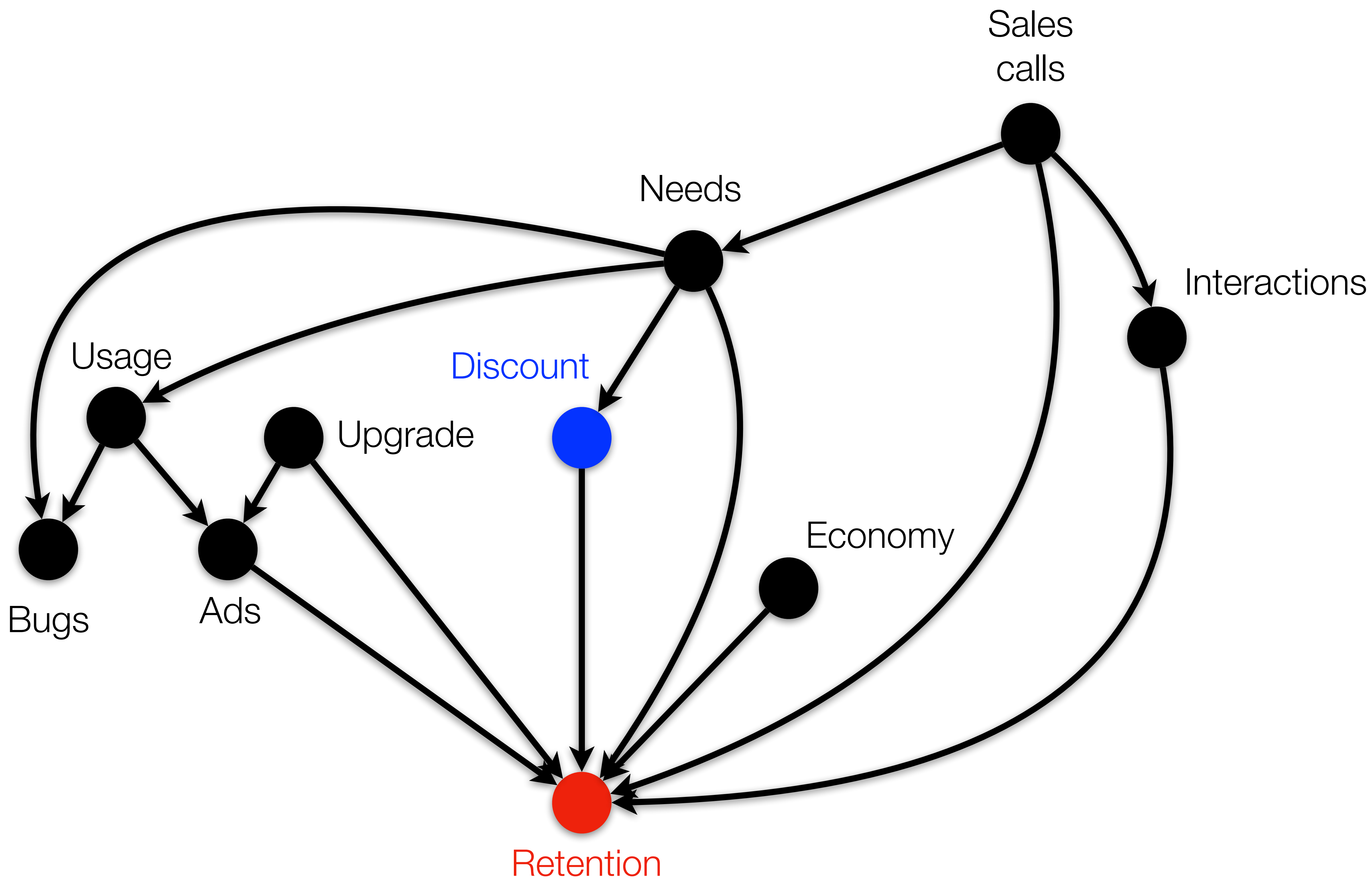
Application 2. Explainable AI

Application 2. Explainable AI



Contribution of Discount to the Retention?

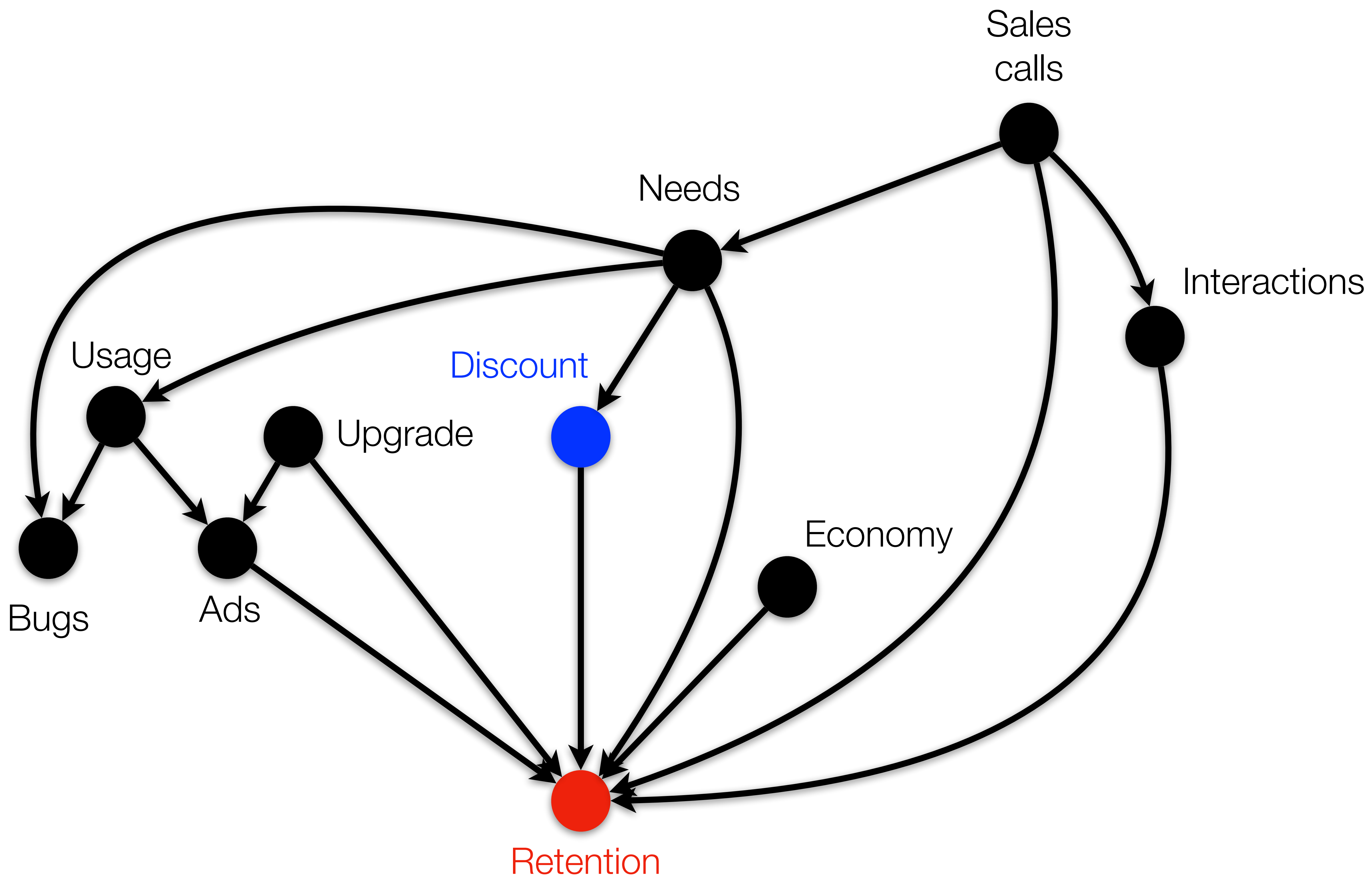
Application 2. Explainable AI



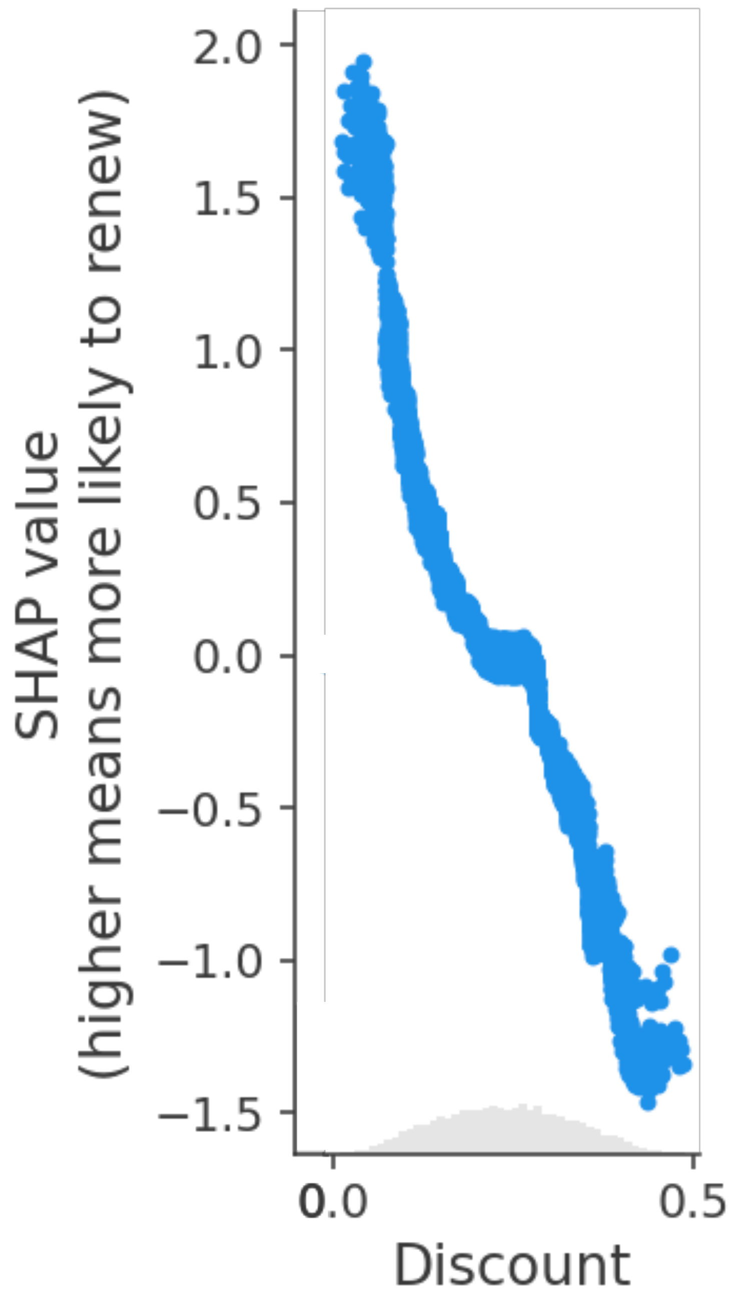
Contribution of Discount to the Retention?

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

Application 2. Explainable AI

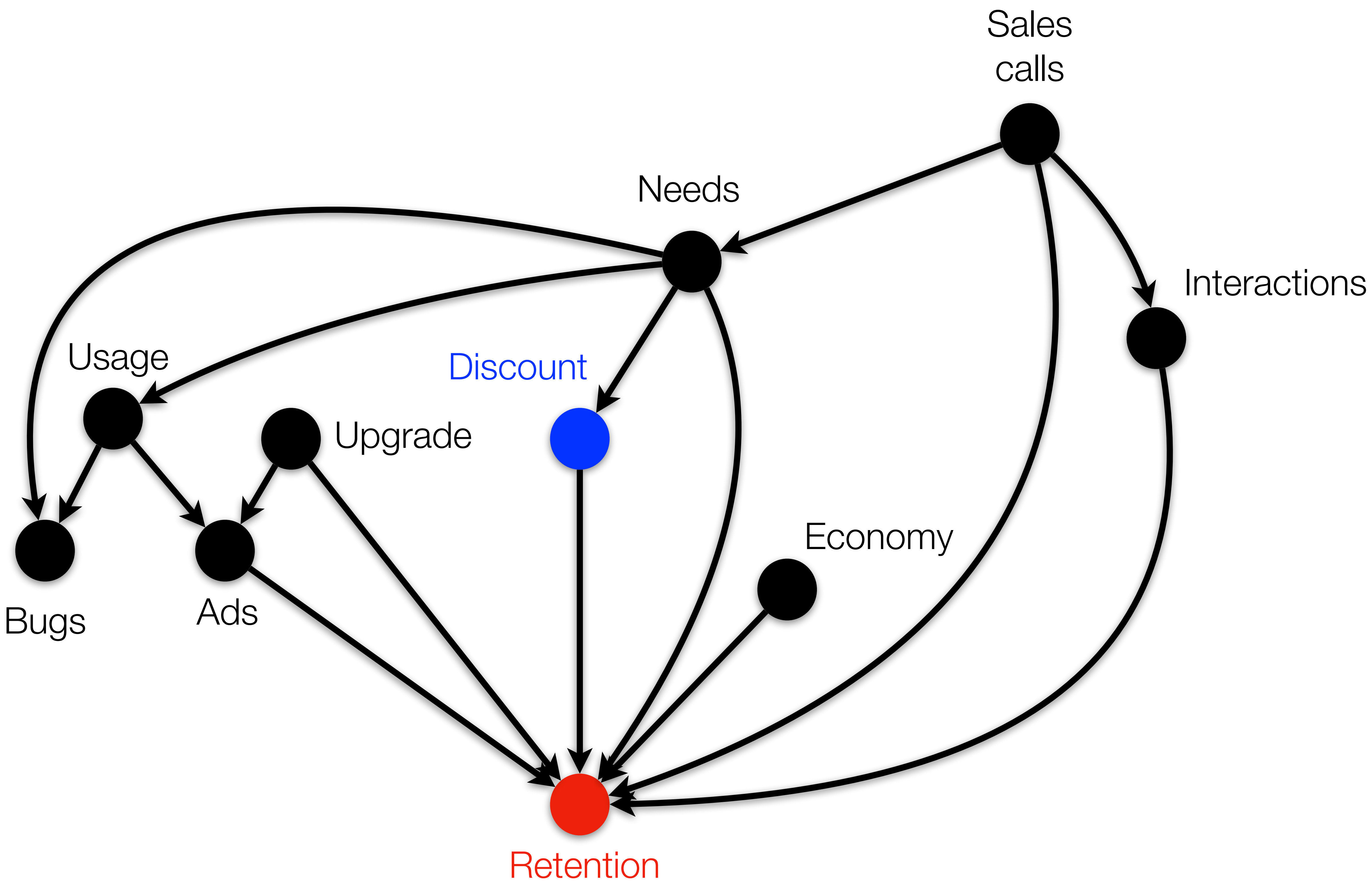


Contribution of Discount to the Retention?

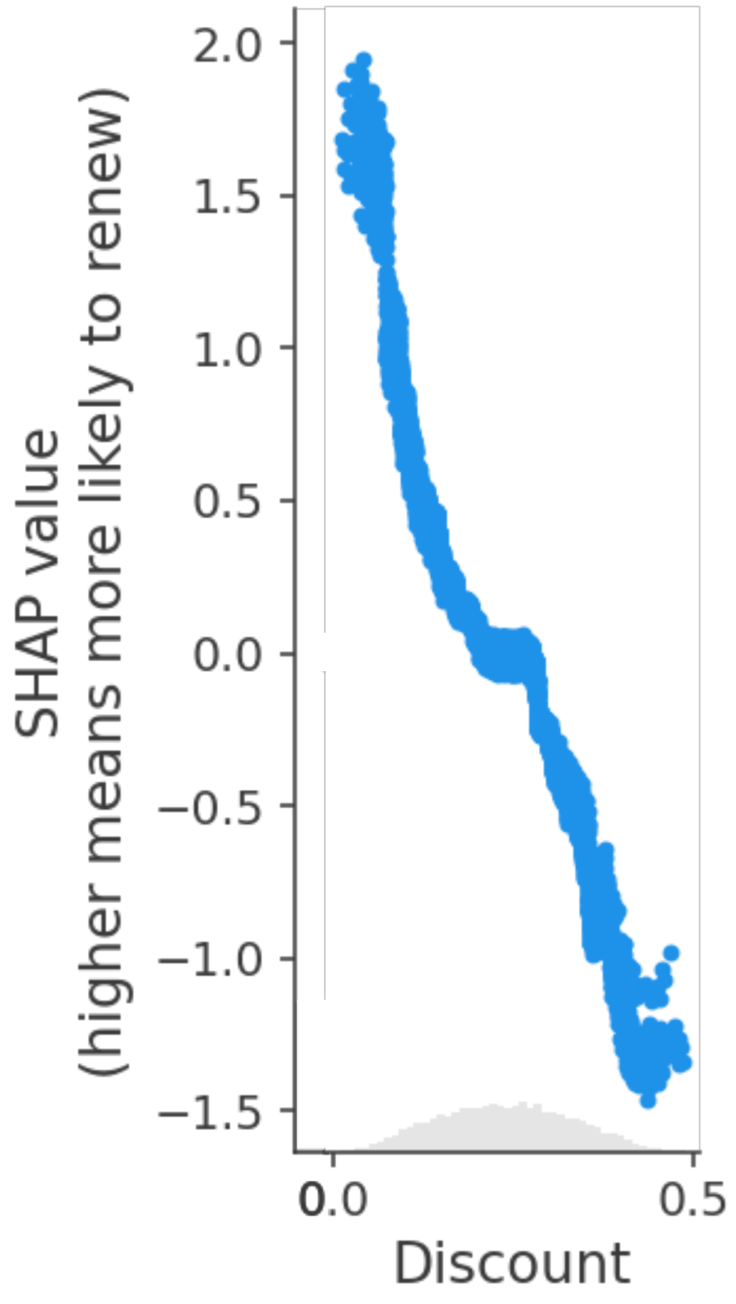


- SHAP value: one of the most cited measure for the feature importance

Application 2. Explainable AI

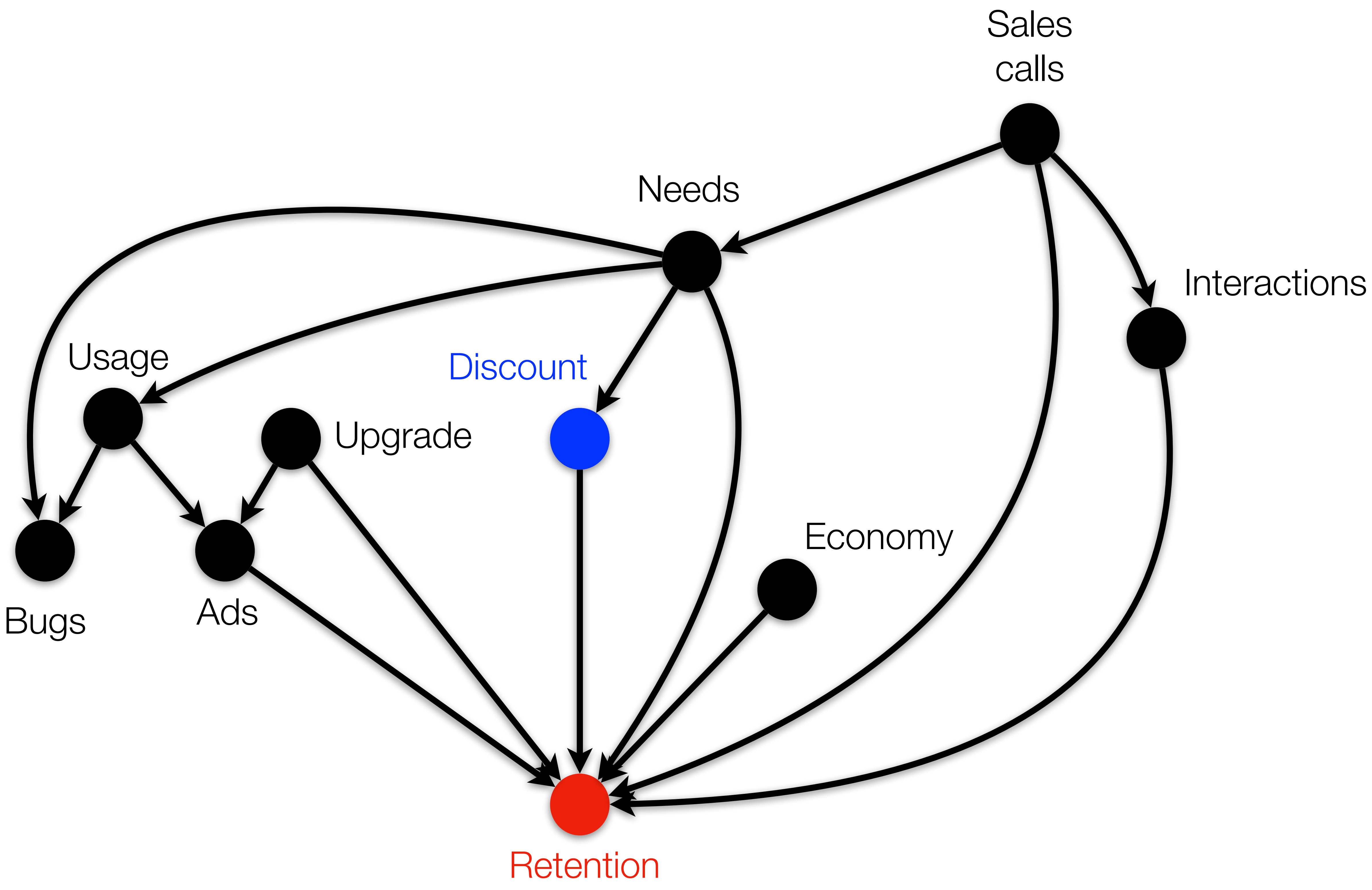


Contribution of Discount to the Retention?

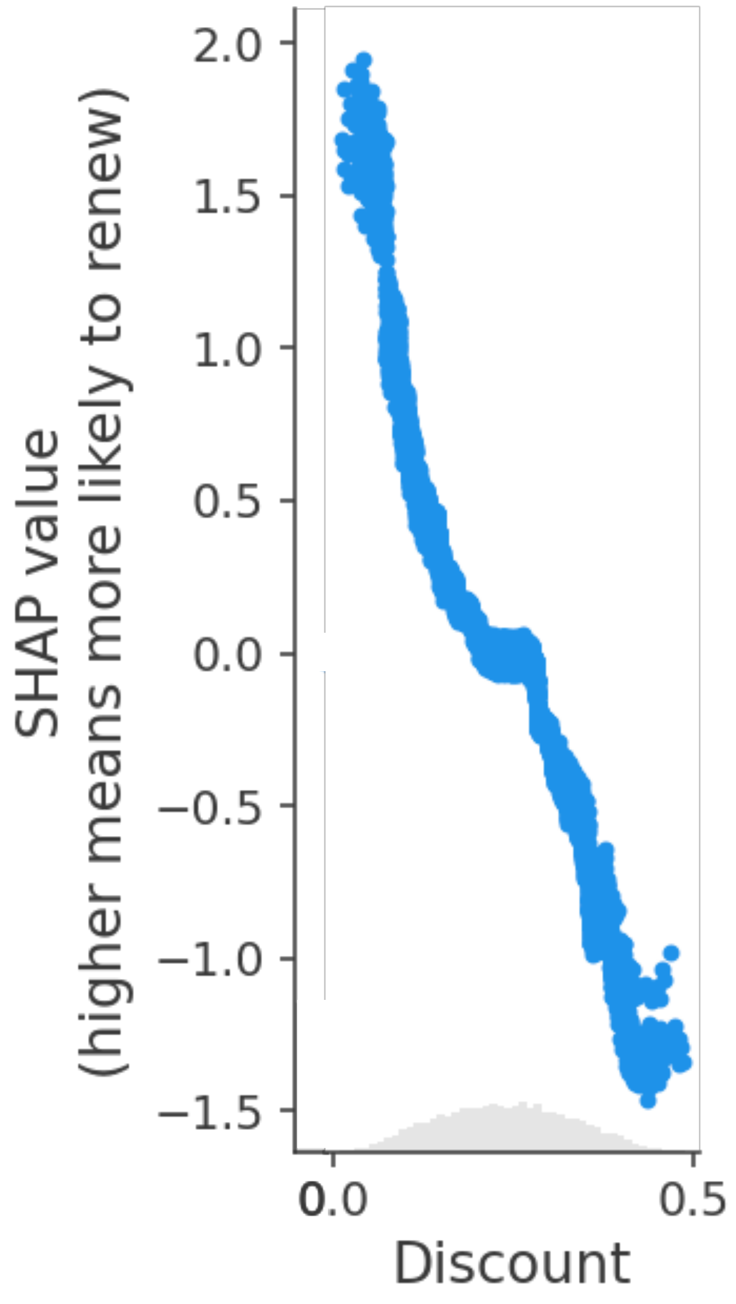


- SHAP value: one of the most cited measure for the feature importance
- *Larger* discounts contribute *less* to retention?

Application 2. Explainable AI

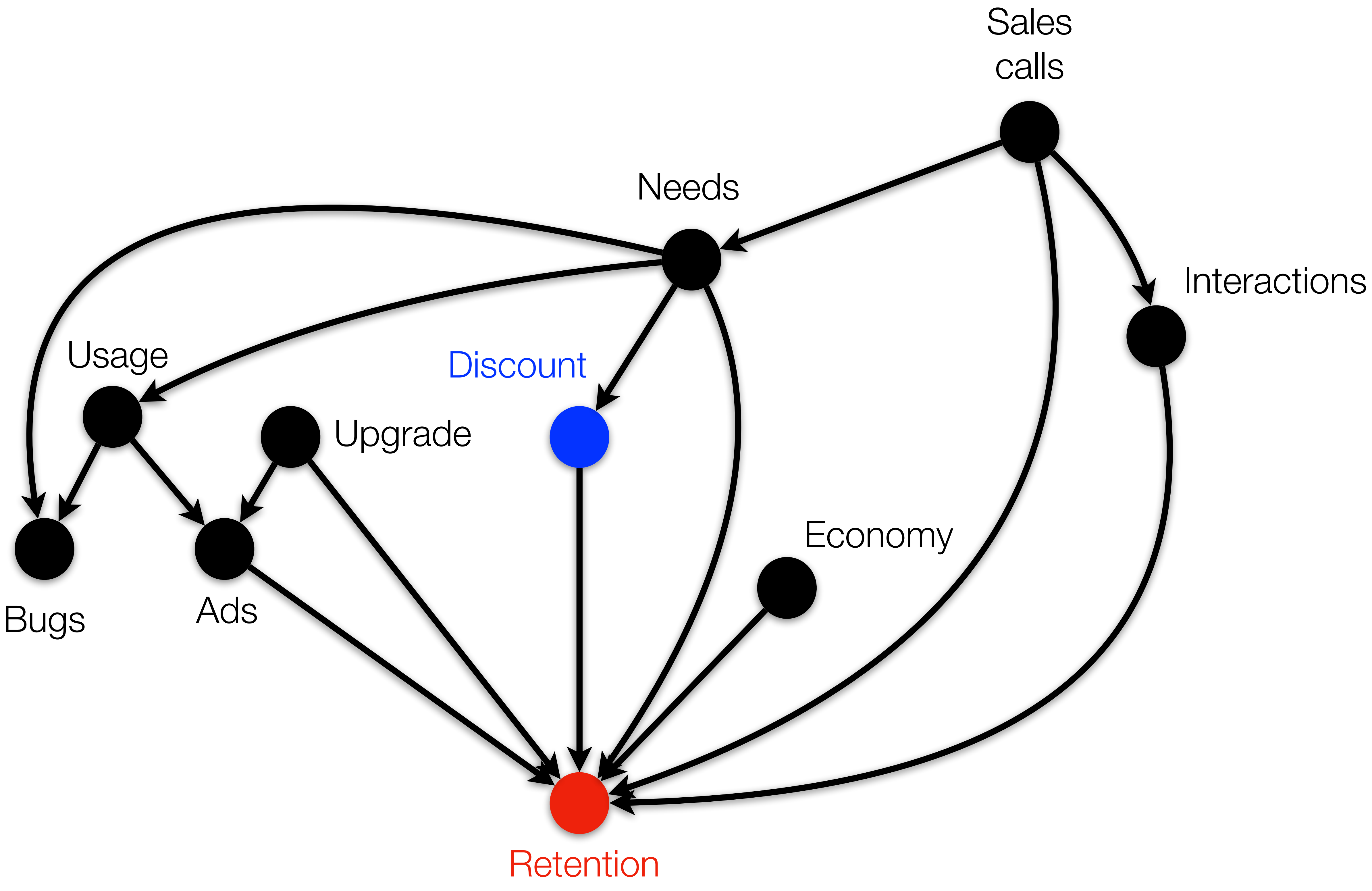


Contribution of Discount to the Retention?

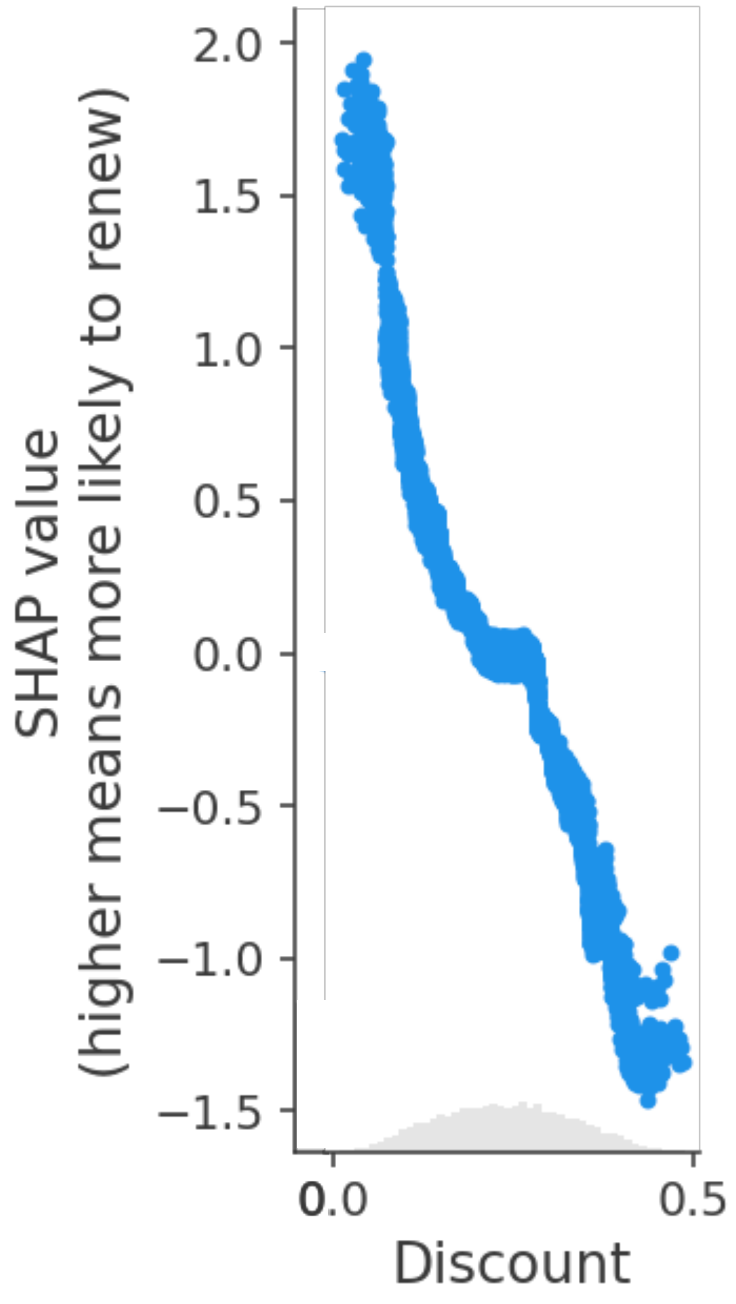


- SHAP value: one of the most cited measure for the feature importance
- *Larger* discounts contribute *less* to retention?
- Mismatch with human intuition is due to computing the importance based on correlation (e.g. $\mathbb{E}[\text{retention}|\text{discount}]$)

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

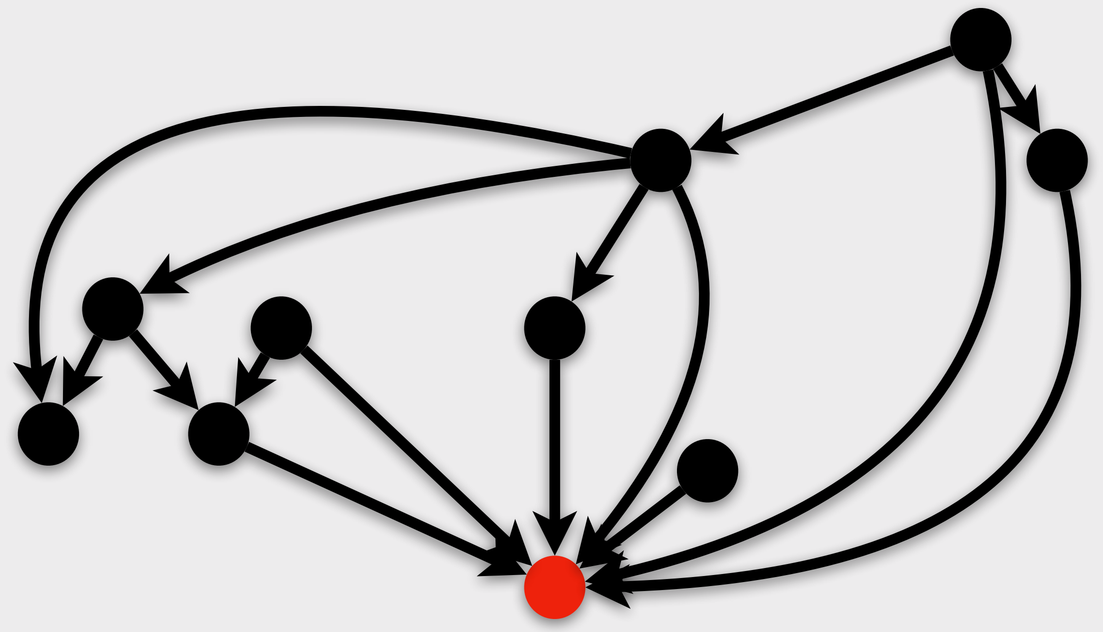
Jung et al.,ICML 2022

do-Shapley: Causality-based Feature Attribution

Jung et al.,ICML 2022

Input

Graph



Data

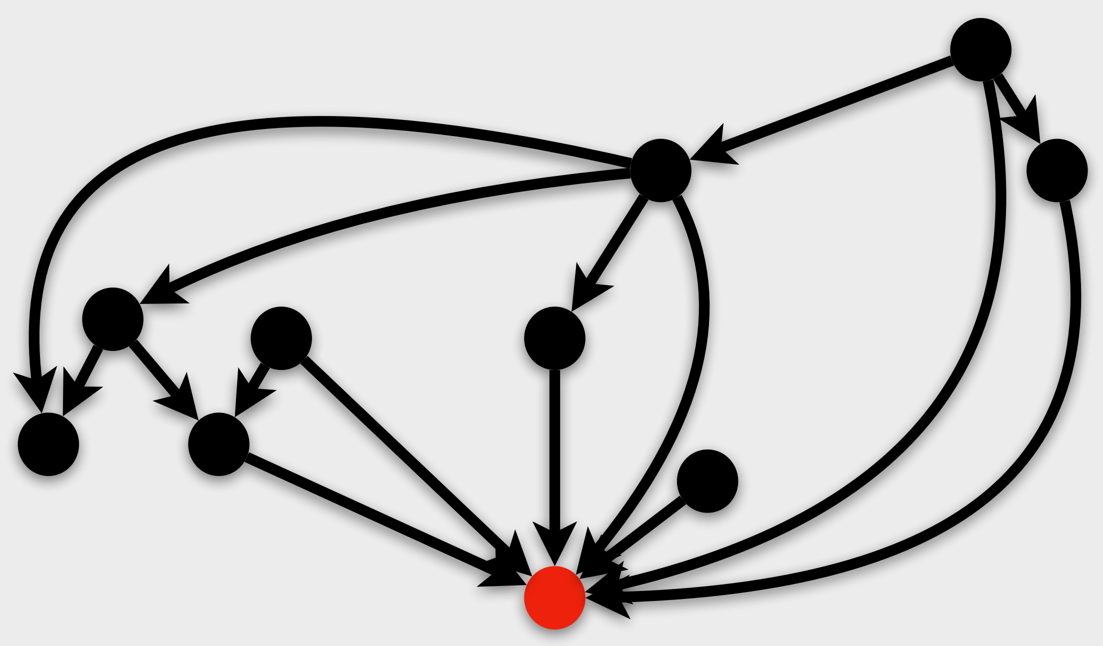
- Input: $(X_1, \cdots, X_m)$
- output: $f(X_1, \cdots, X_m)$

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

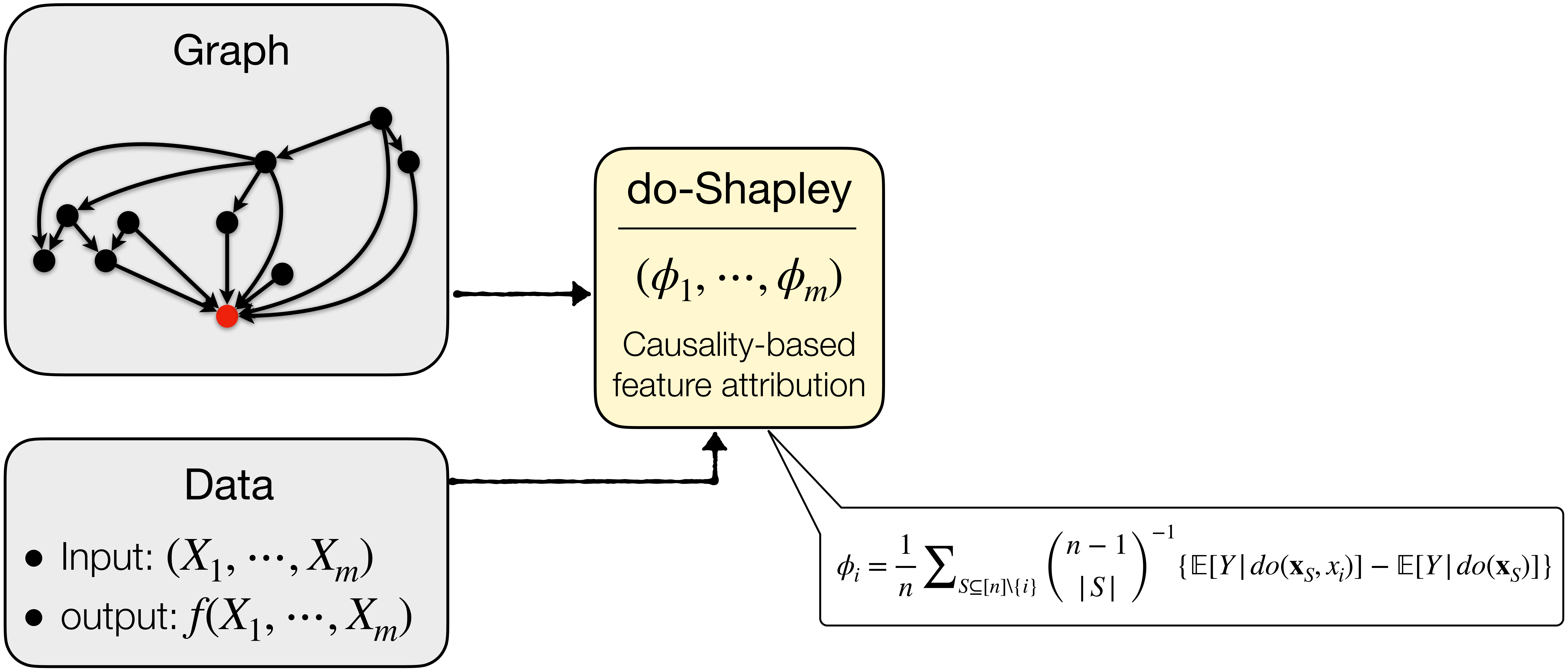
do-Shapley

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

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

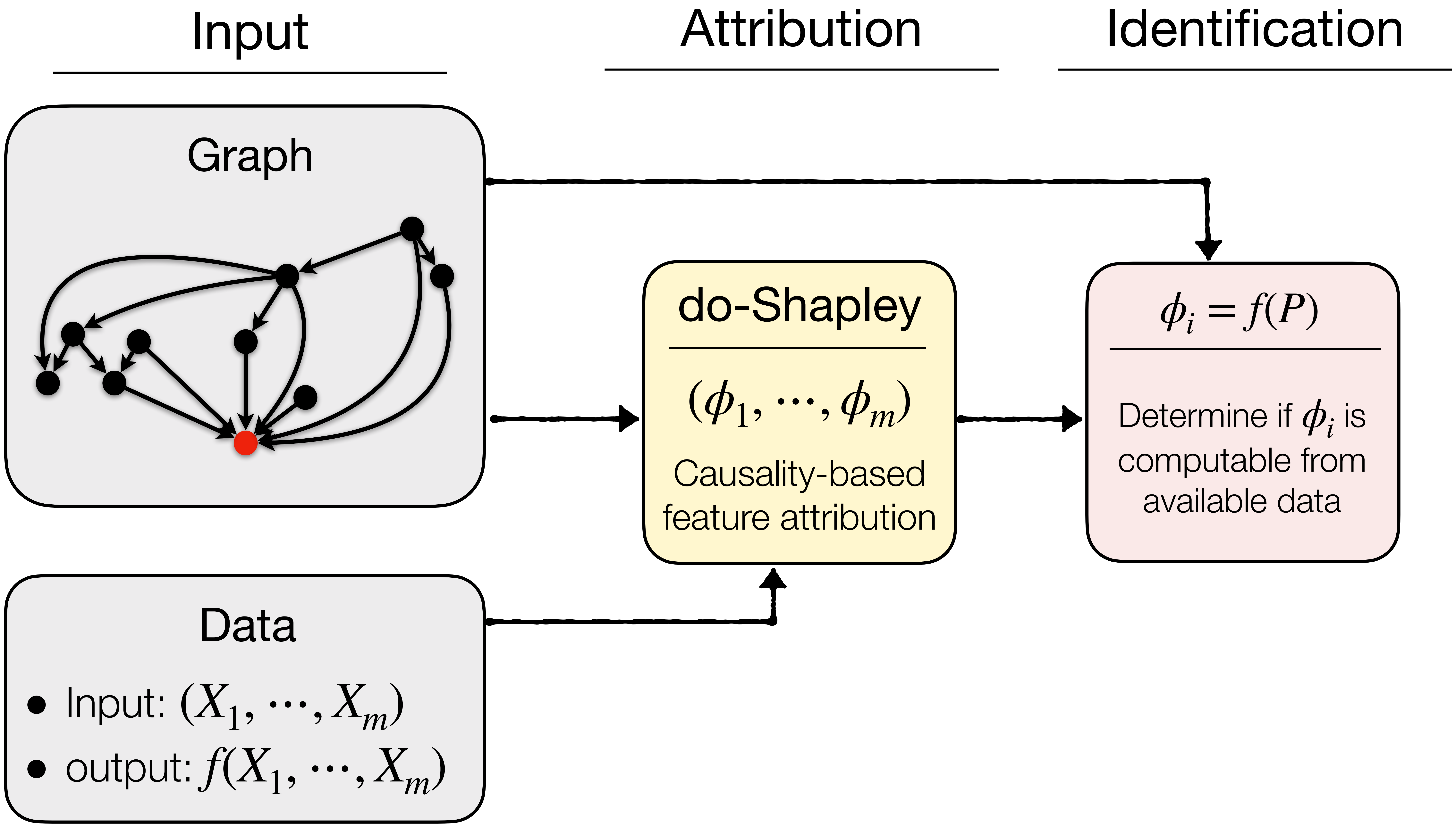
$(\phi_1, \cdots, \phi_m)$

Causality-based feature attribution

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

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

$\phi_i = f(P)$

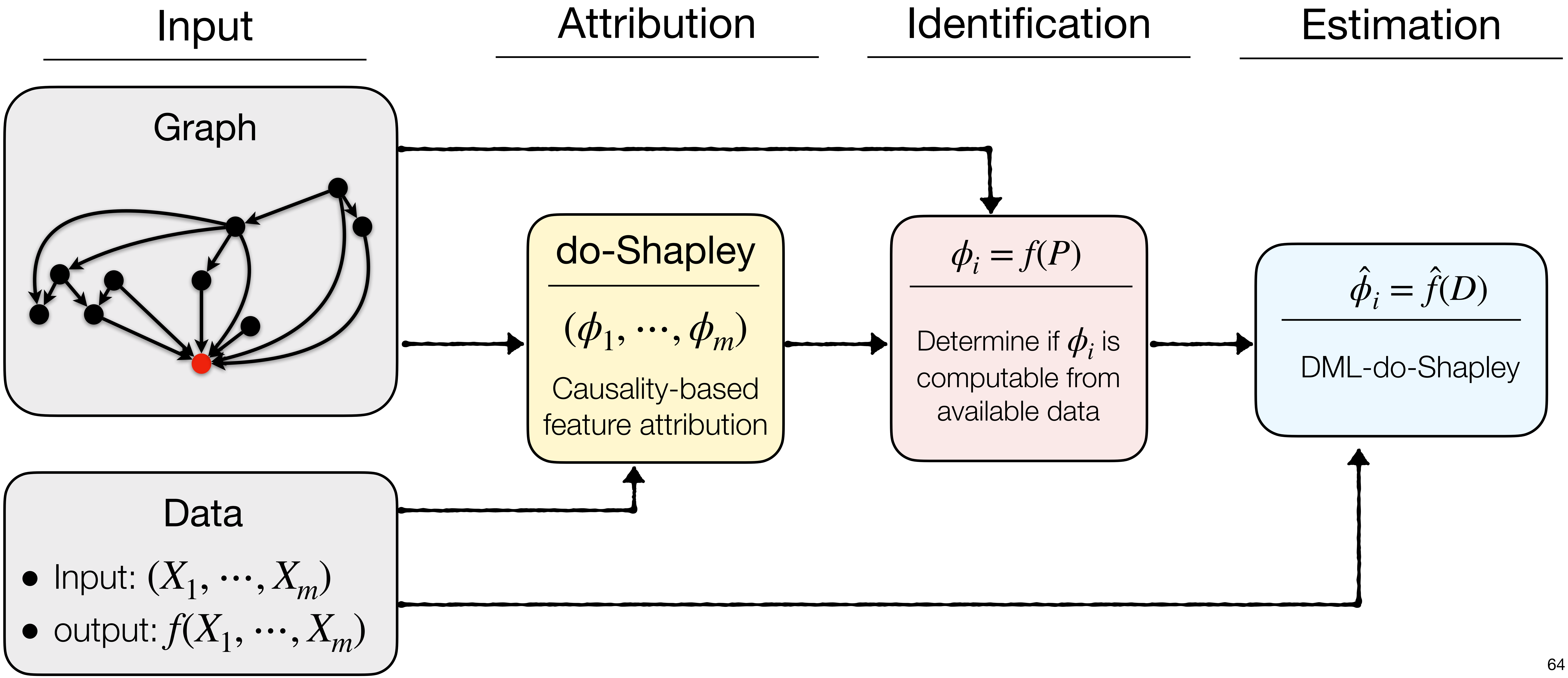
Determine if ϕ_i is computable from available data

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



Input

Attribution

Identification

Estimation

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

$\phi_i = f(P)$

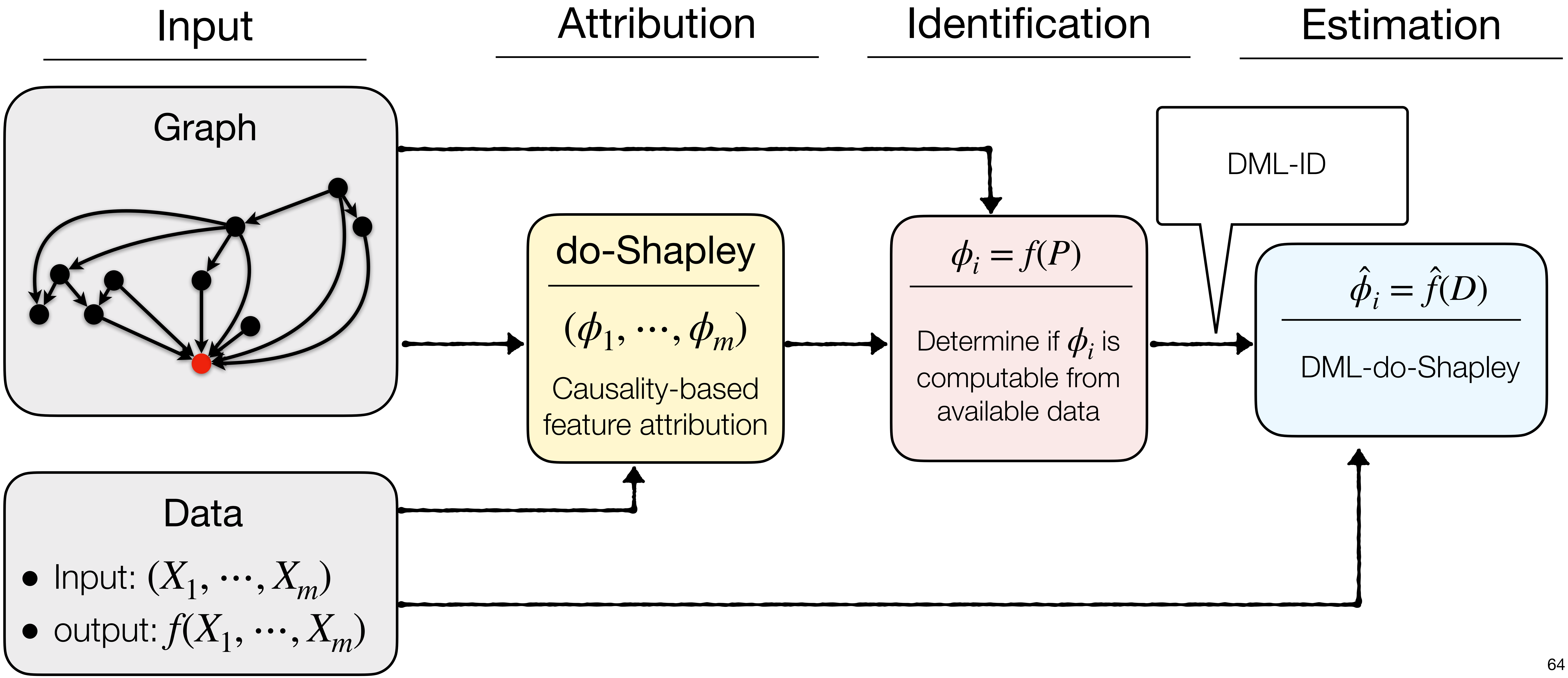
Determine if ϕ_i is computable from available data

$\hat{\phi}_i = \hat{f}(D)$

DML-do-Shapley

do-Shapley: Causality-based Feature Attribution

Jung et al.,ICML 2022



Input

Attribution

Identification

Estimation

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

$\phi_i = f(P)$

Determine if ϕ_i is computable from available data

DML-ID

$\hat{\phi}_i = \hat{f}(D)$

DML-do-Shapley

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

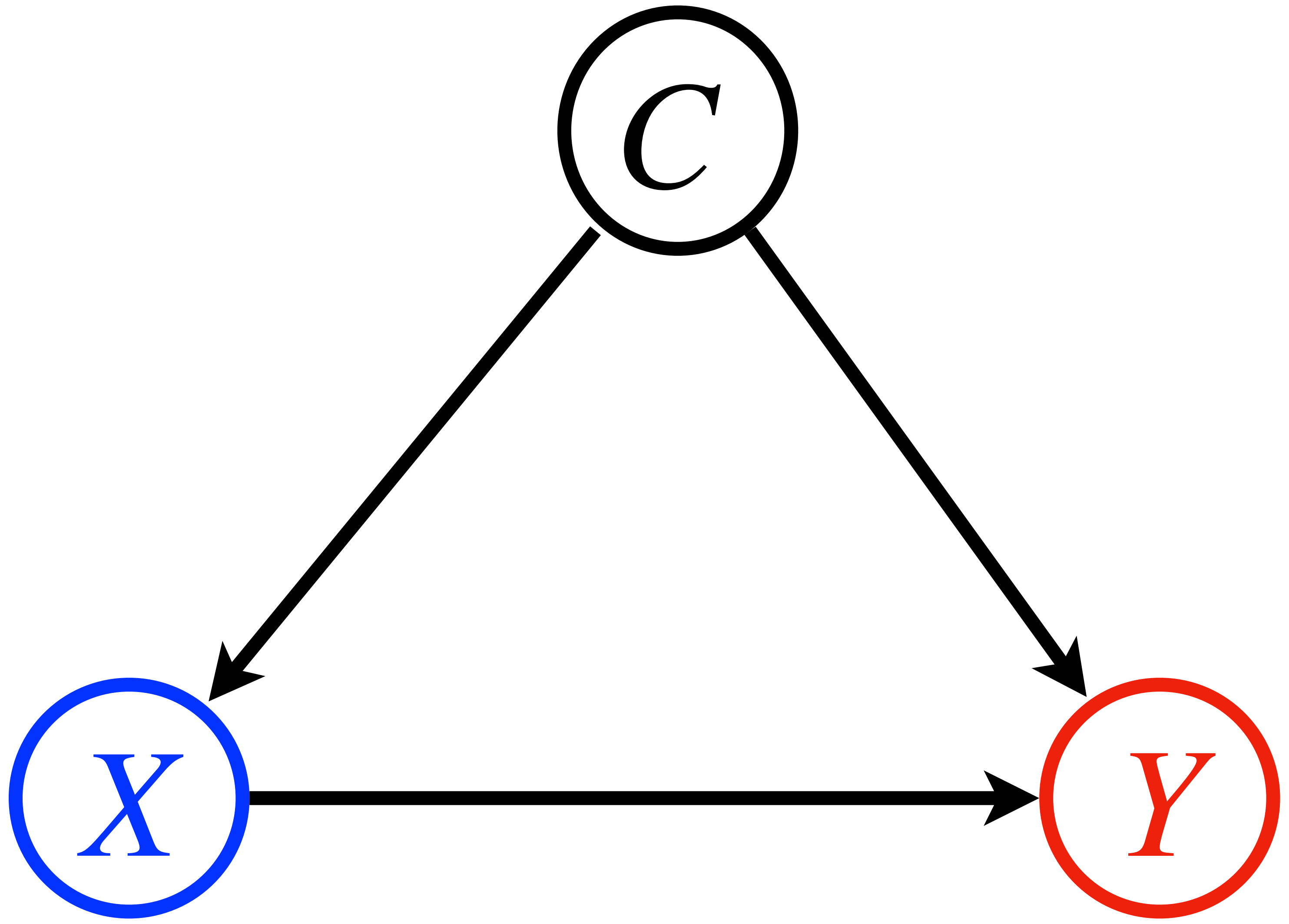
Impact on Explainable AI

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

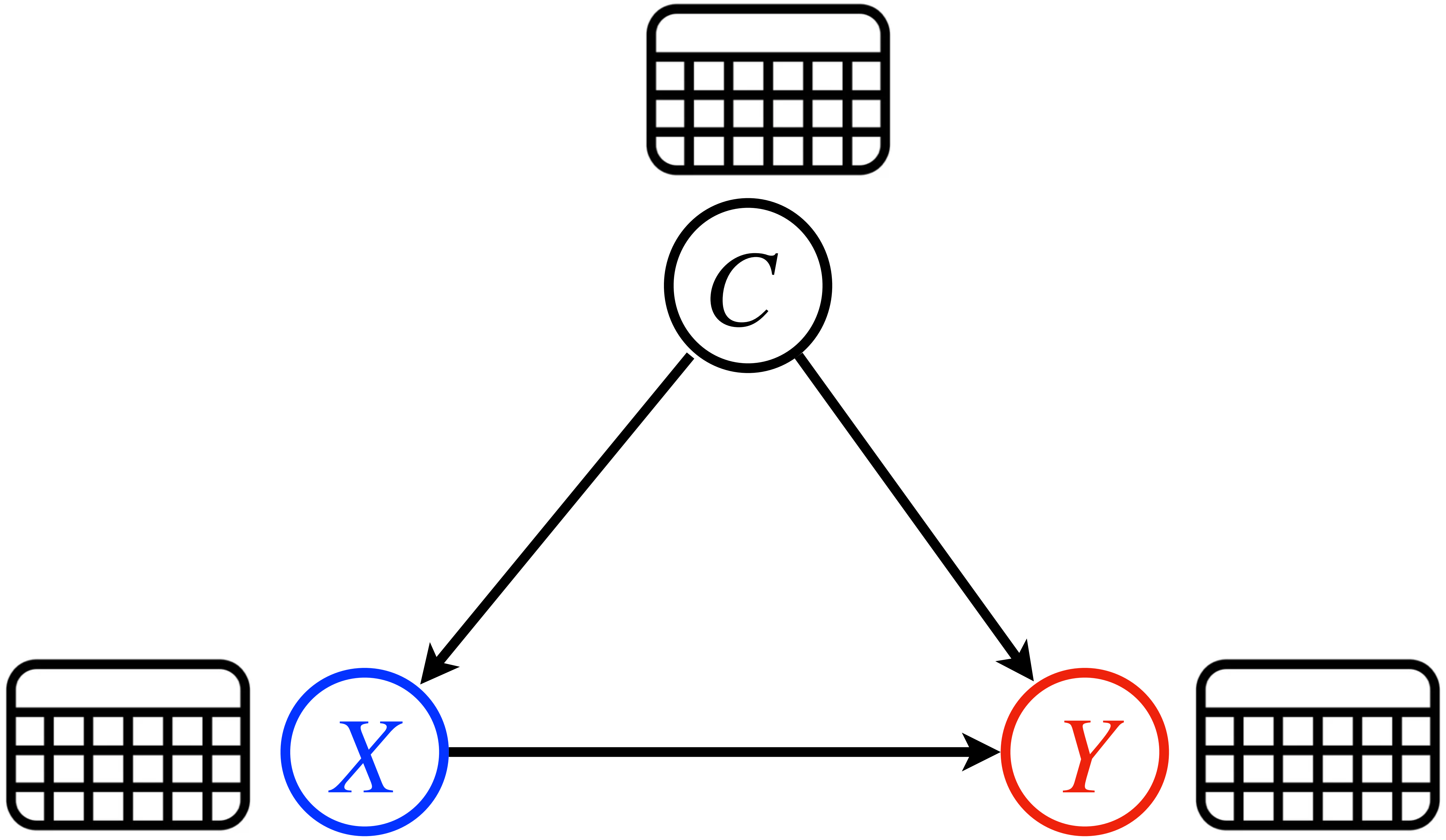
- Two features receive equal contributions whenever their causal effects are the same.
- Feature's contribution = 0 if it has no causal effect
- Feature contributions closely approximate their causal effects on the outcome
- The sum of feature contributions = The outcome $f(X_1, \cdots, X_m)$

Future Direction

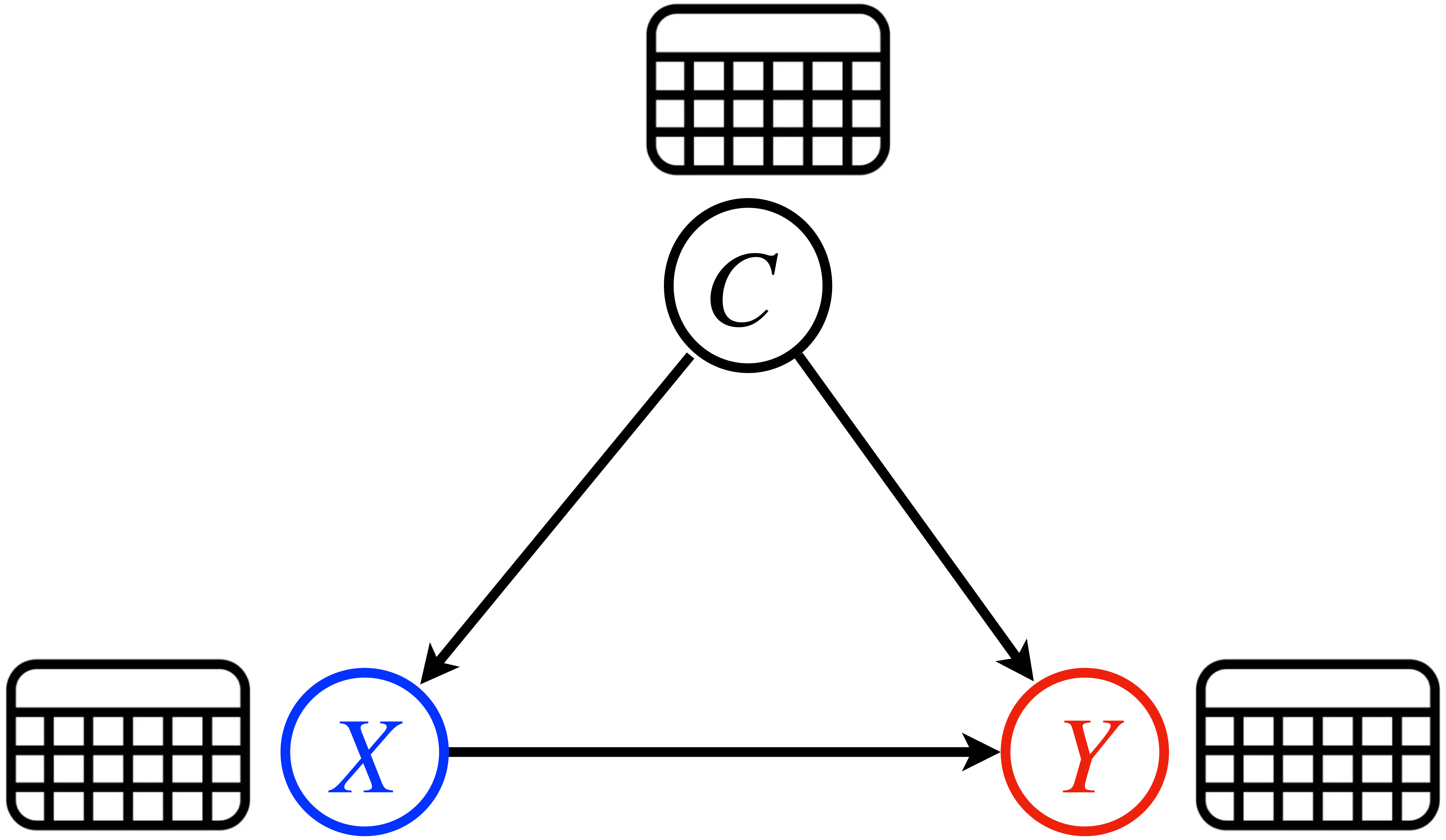
Future 1: Inference with Multi-modal Data



Future 1: Inference with Multi-modal Data

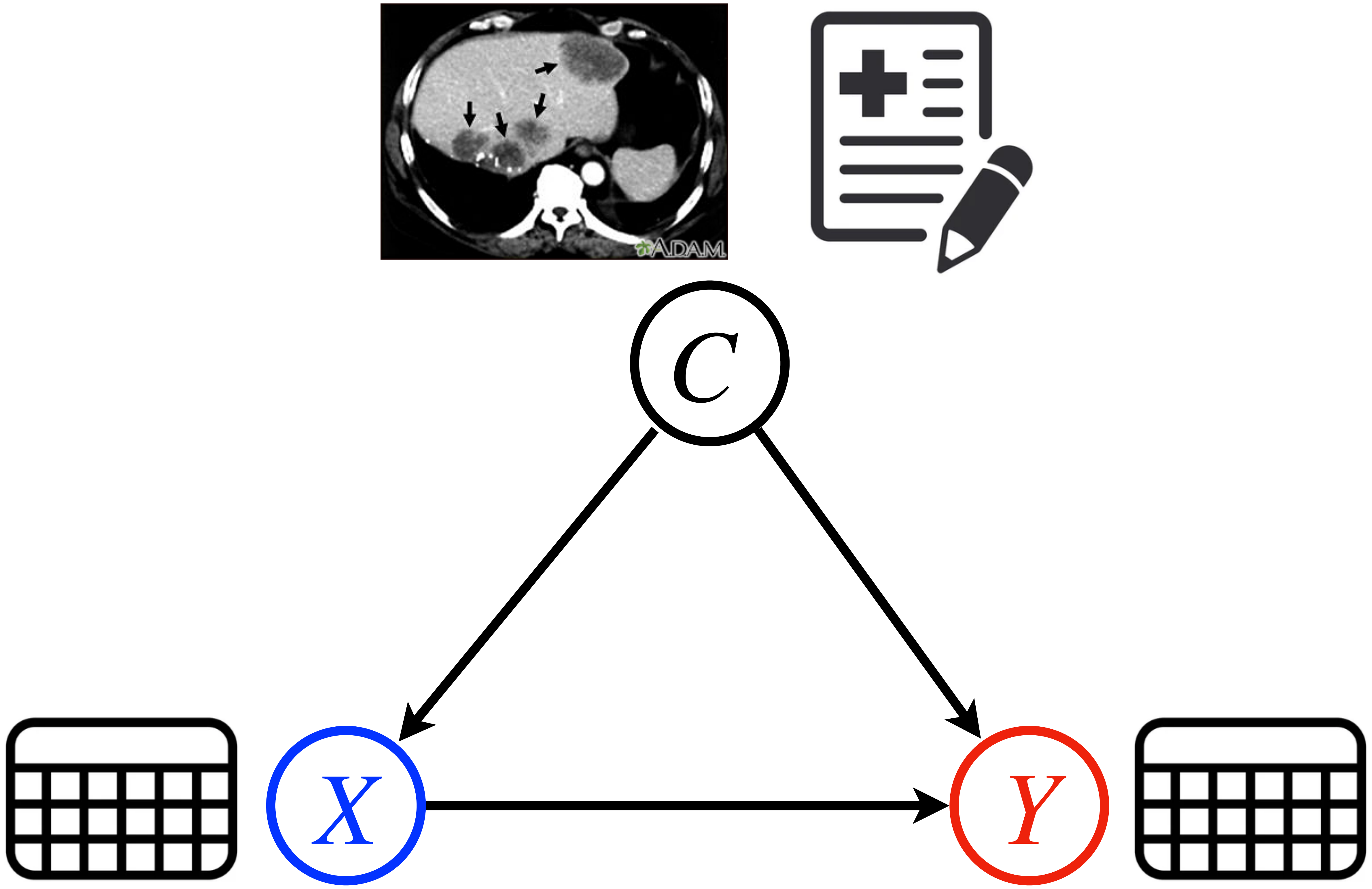


Future 1: Inference with Multi-modal Data



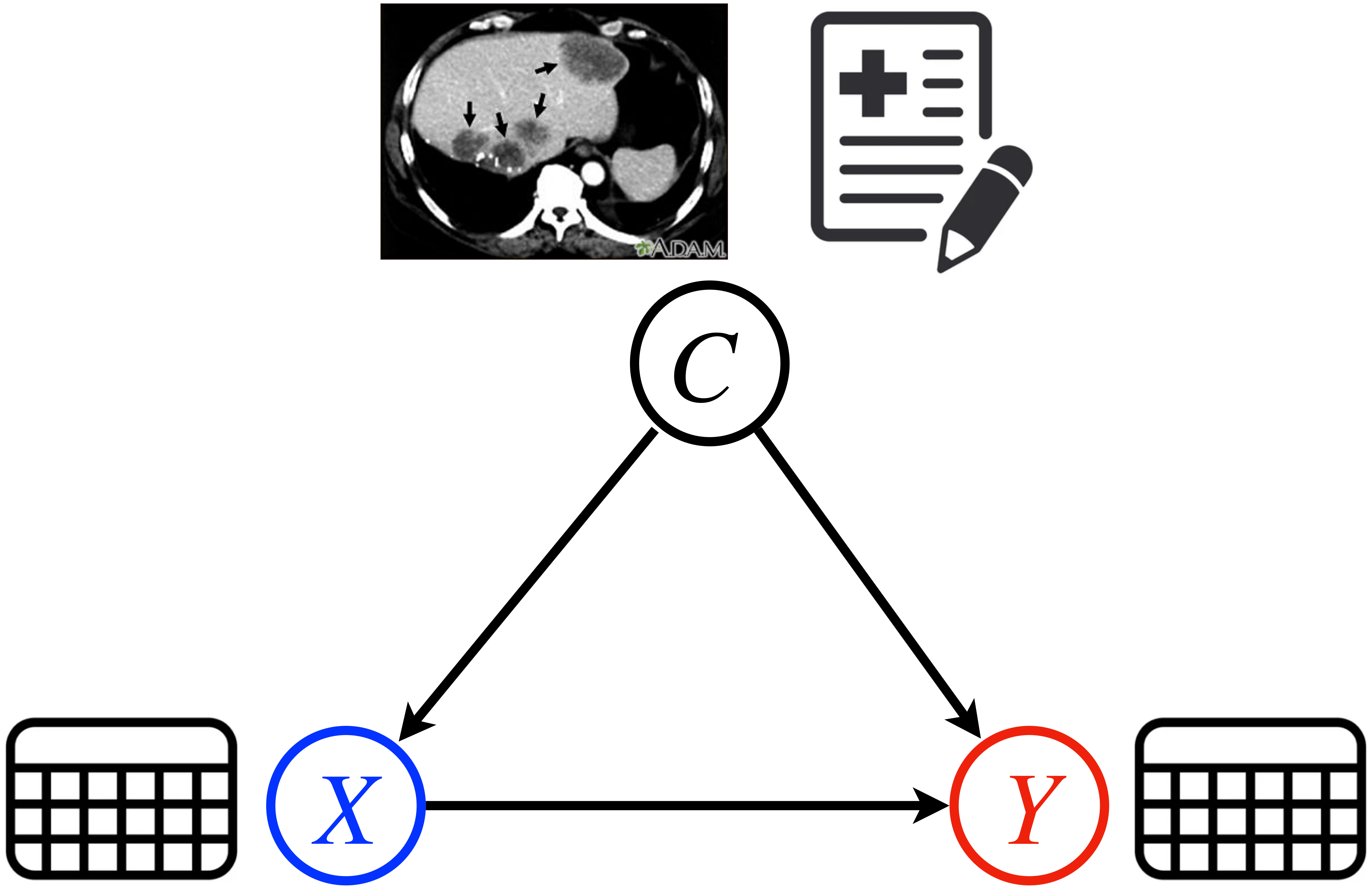
$$\mathbb{E}[Y \mid \mathrm{do}(x)] = \sum_c \mathbb{E}[Y \mid x, c]\, P(c)$$

Future 1: Inference with Multi-modal Data



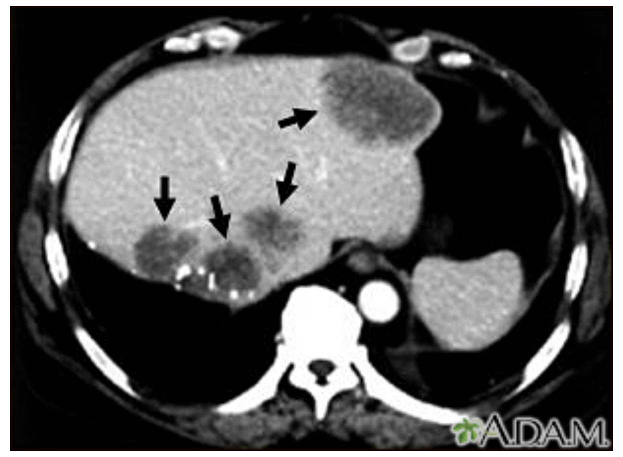
$$\mathbb{E}[Y \mid \mathrm{do}(x)] = \sum_c \mathbb{E}[Y \mid x, c]\, P(c)$$

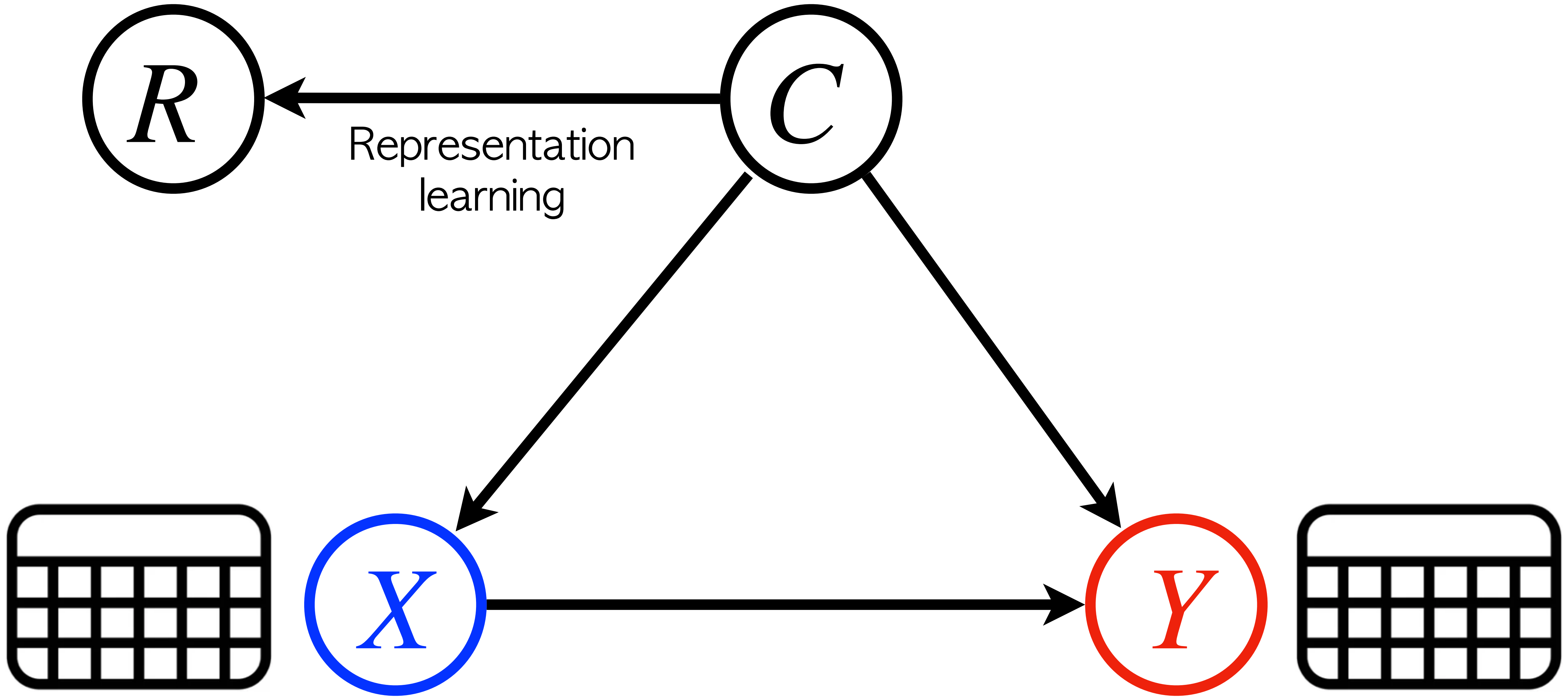
Future 1: Inference with Multi-modal Data



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

Future 1: Inference with Multi-modal Data

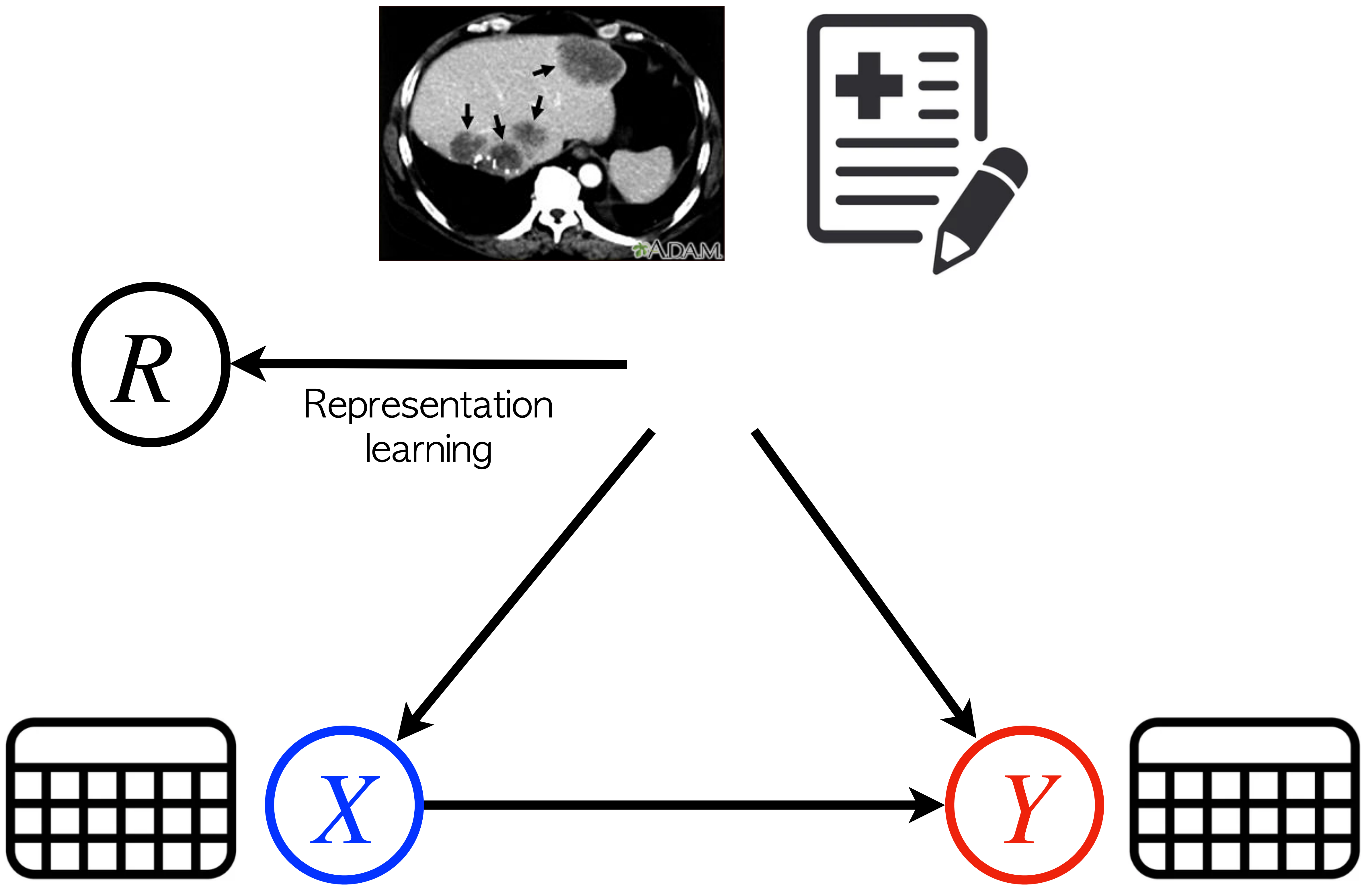




Representation learning

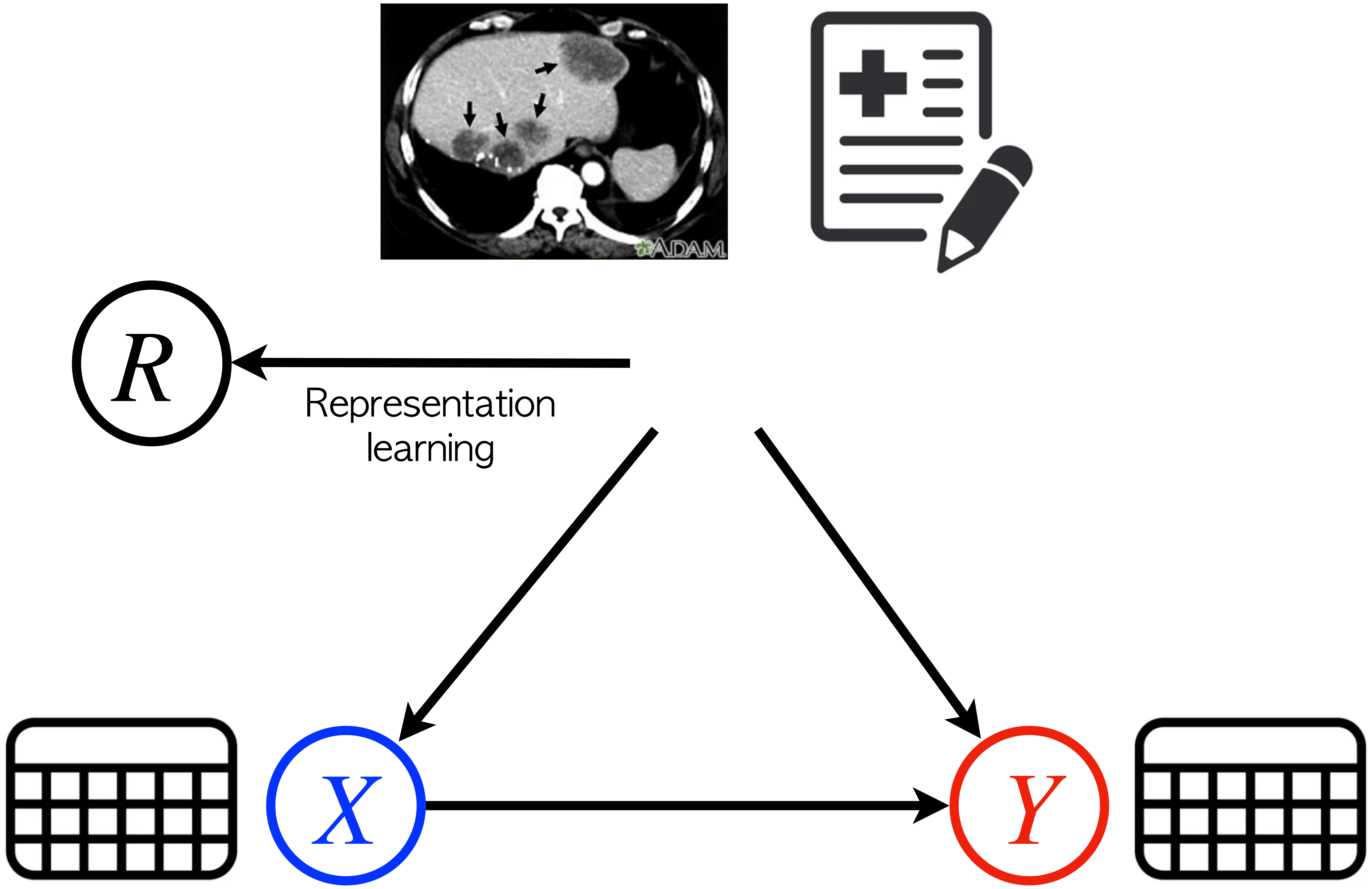
$$\mathbb{E}[Y \mid \mathrm{do}(x)] = \quad ?$$

Future 1: Inference with Multi-modal Data



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

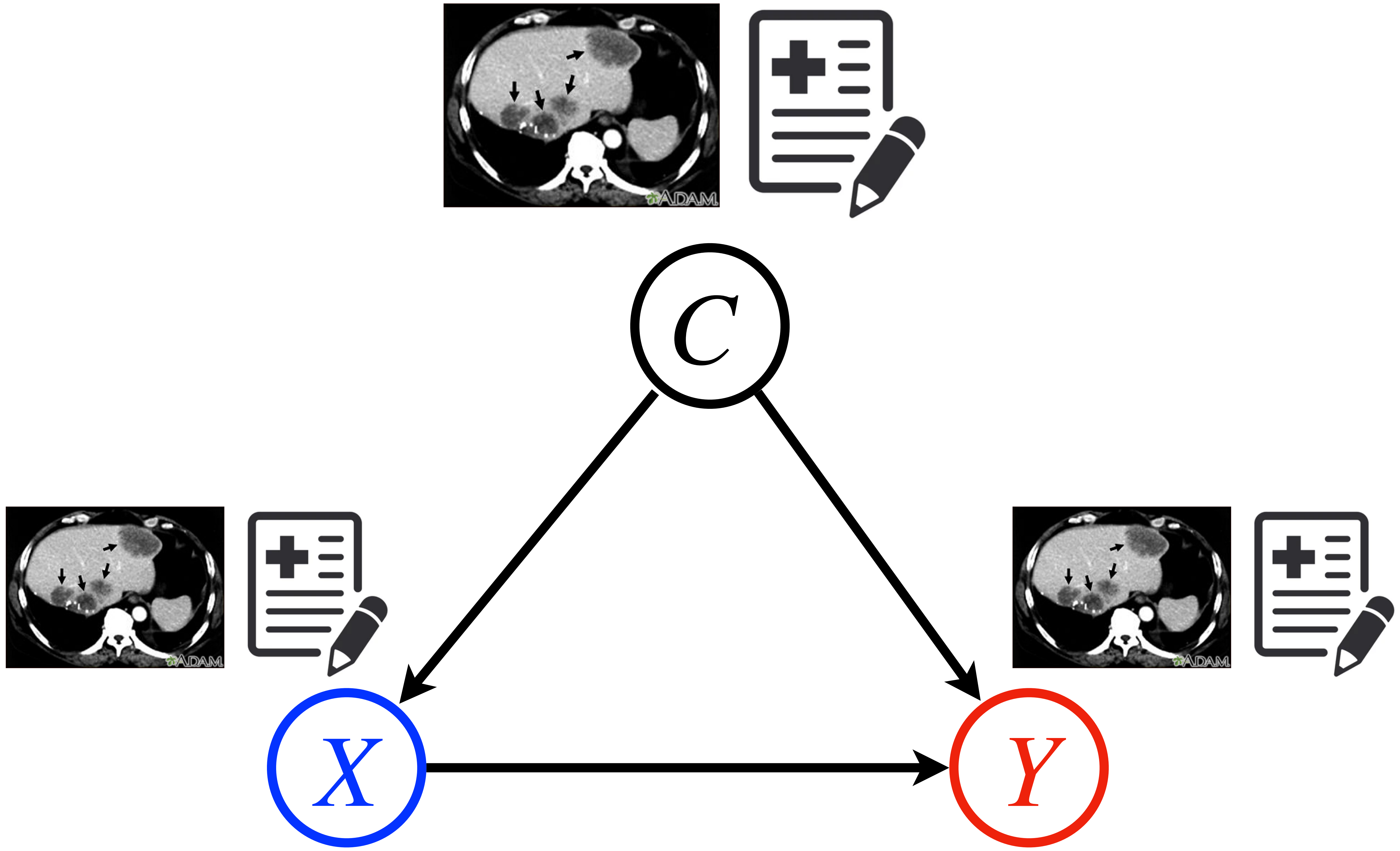
Future 1: Inference with Multi-modal Data



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

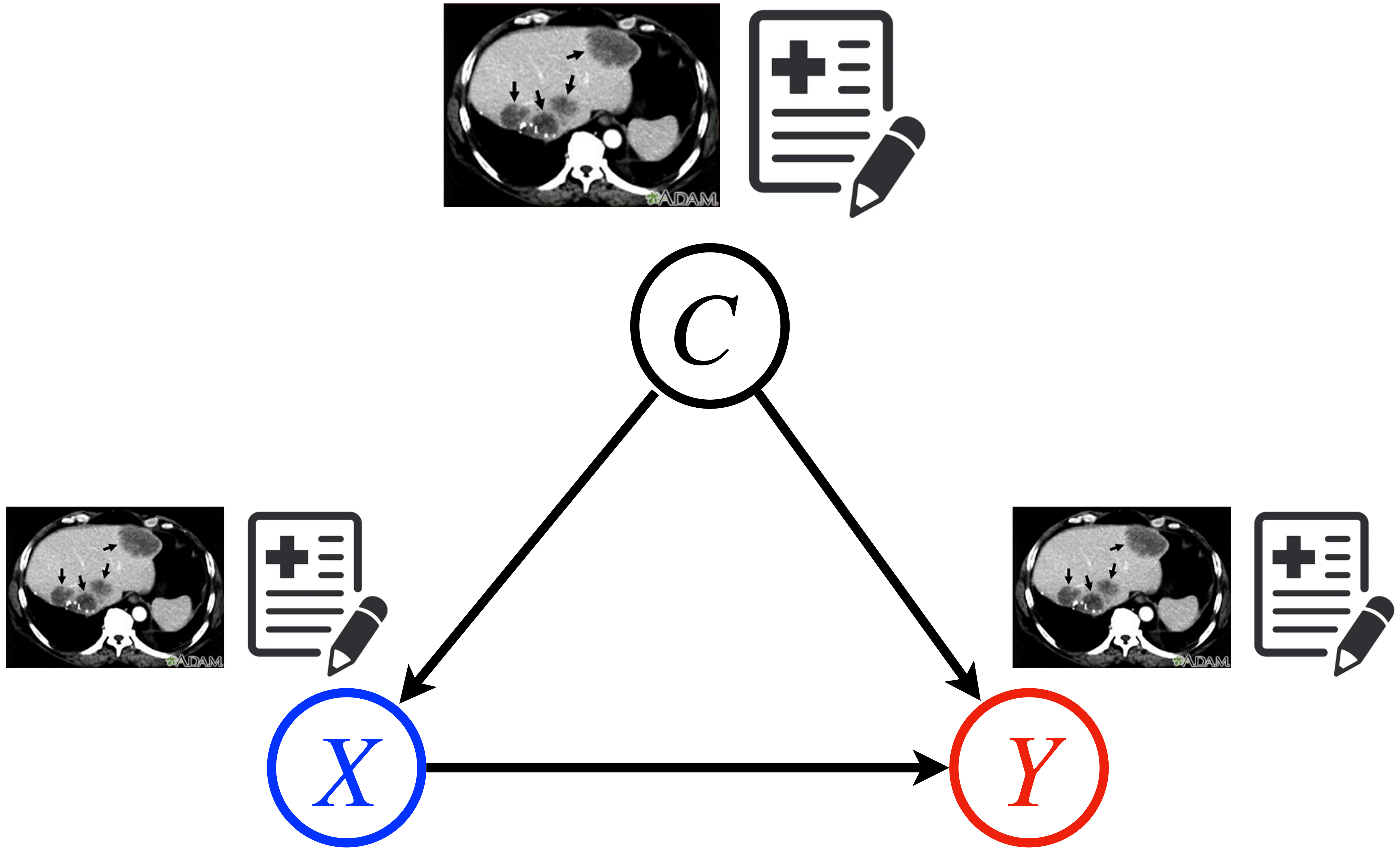
R doesn't satisfy the BD criterion

Future 1: Inference with Multi-modal Data



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

Future 1: Inference with Multi-modal Data



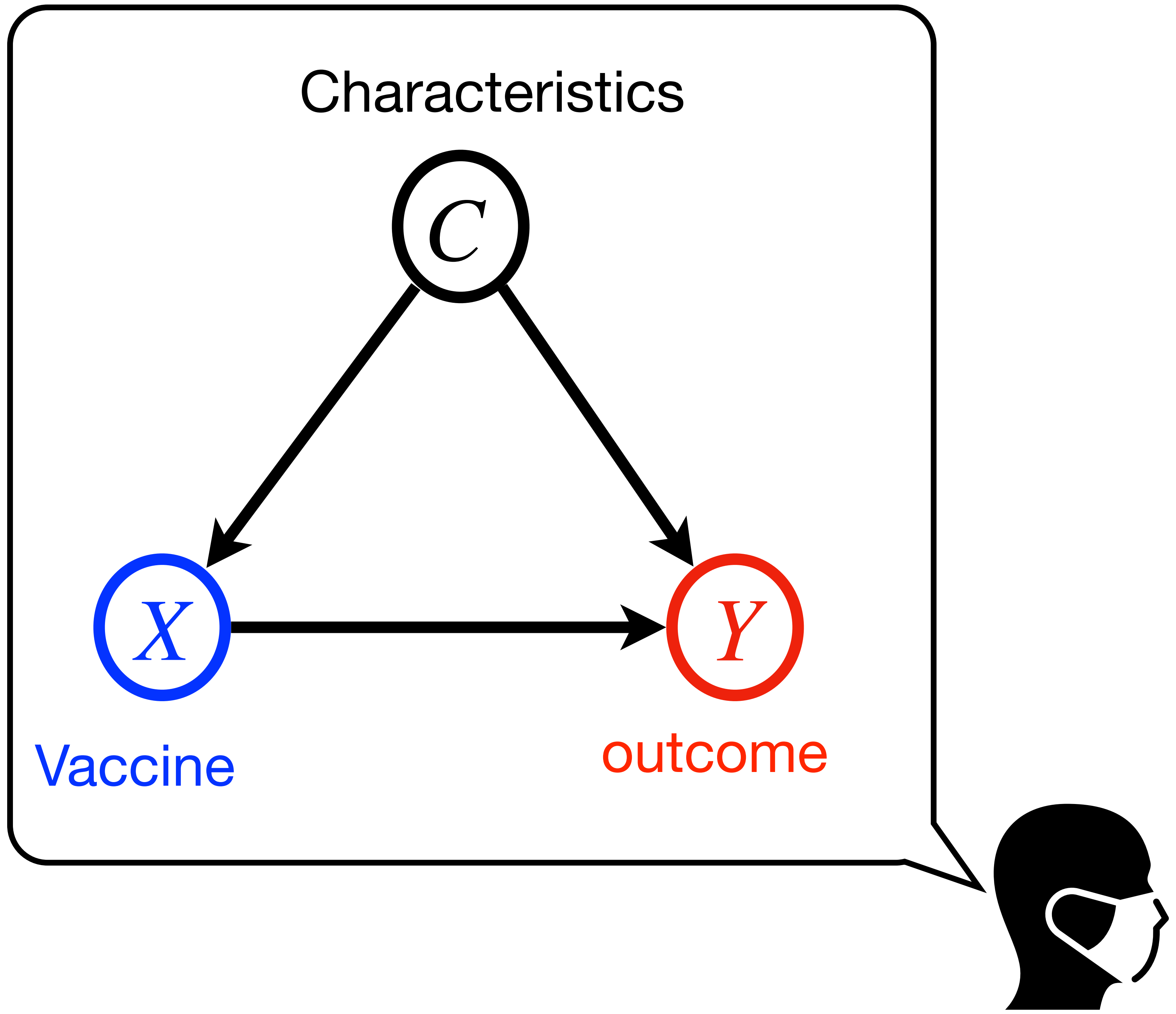
$$\mathbb{E}[Y \mid \mathrm{do}(x)] = \ ?$$

Approach

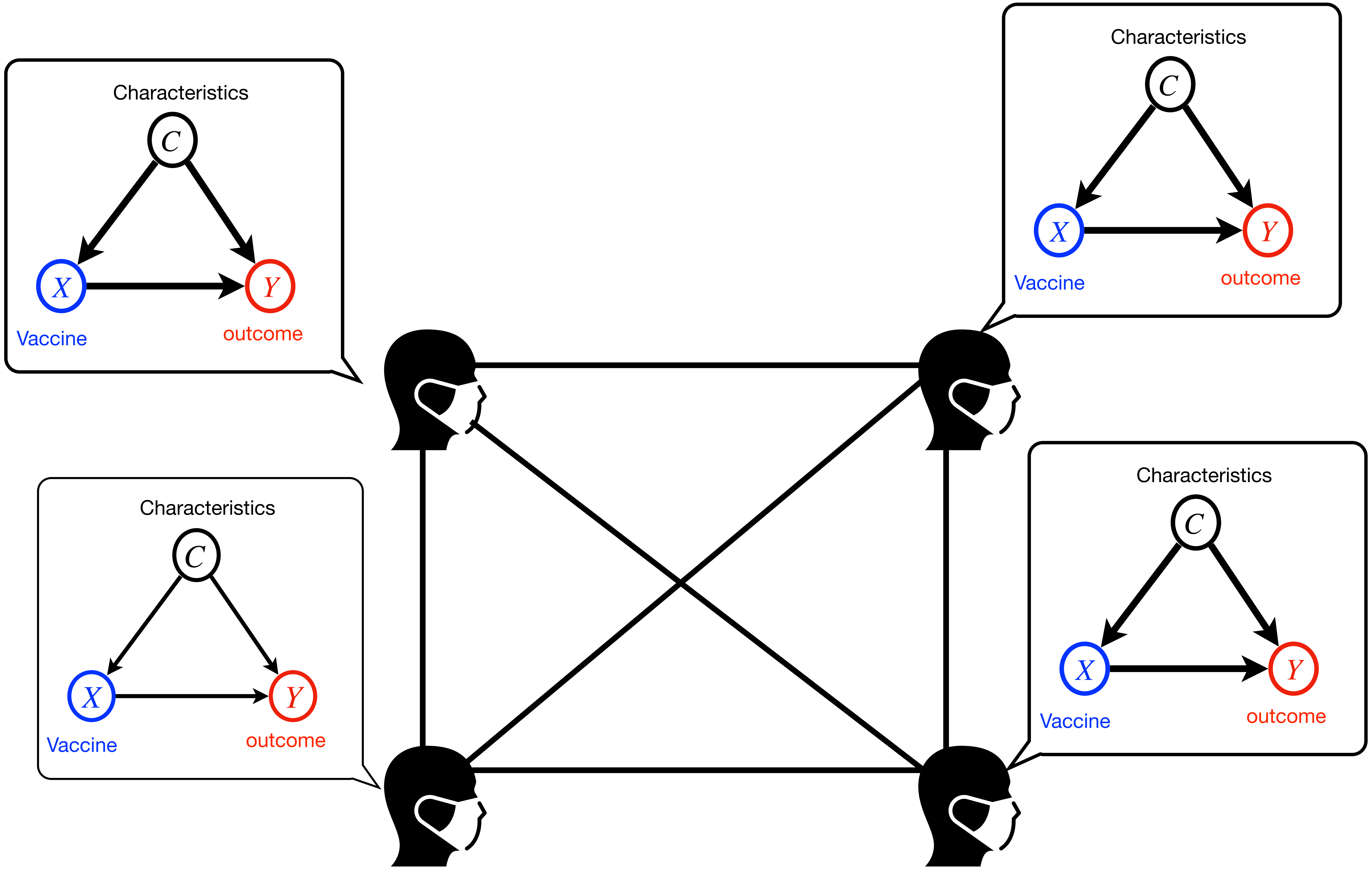
- Representation learning taking account of causal dependencies
- New causal inference methods that allows us to use existing representation learning models

Future 2: Causal Inference with Spatiotemporal Data

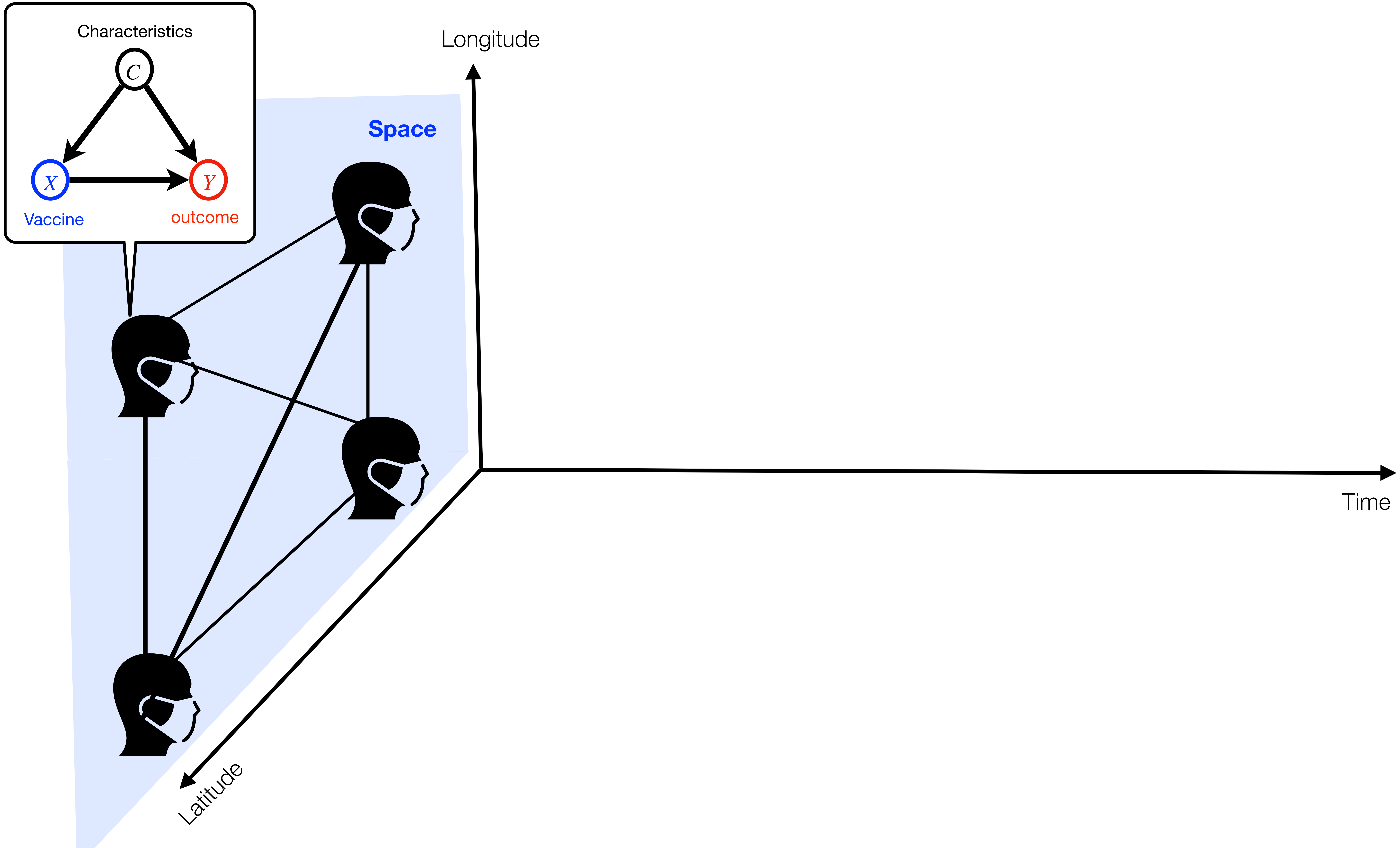
Future 2: Causal Inference with Spatiotemporal Data



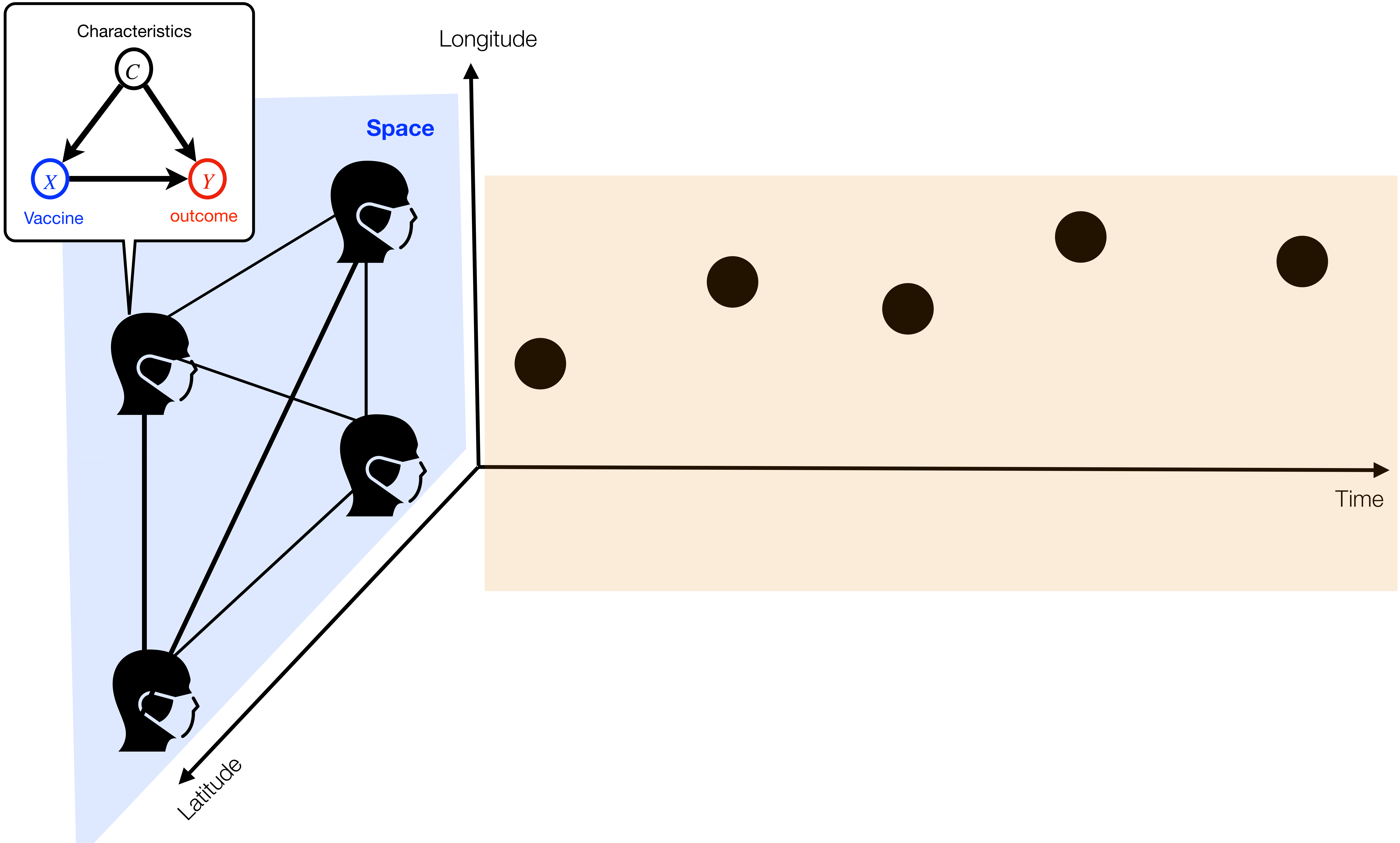
Future 2: Causal Inference with Spatiotemporal Data



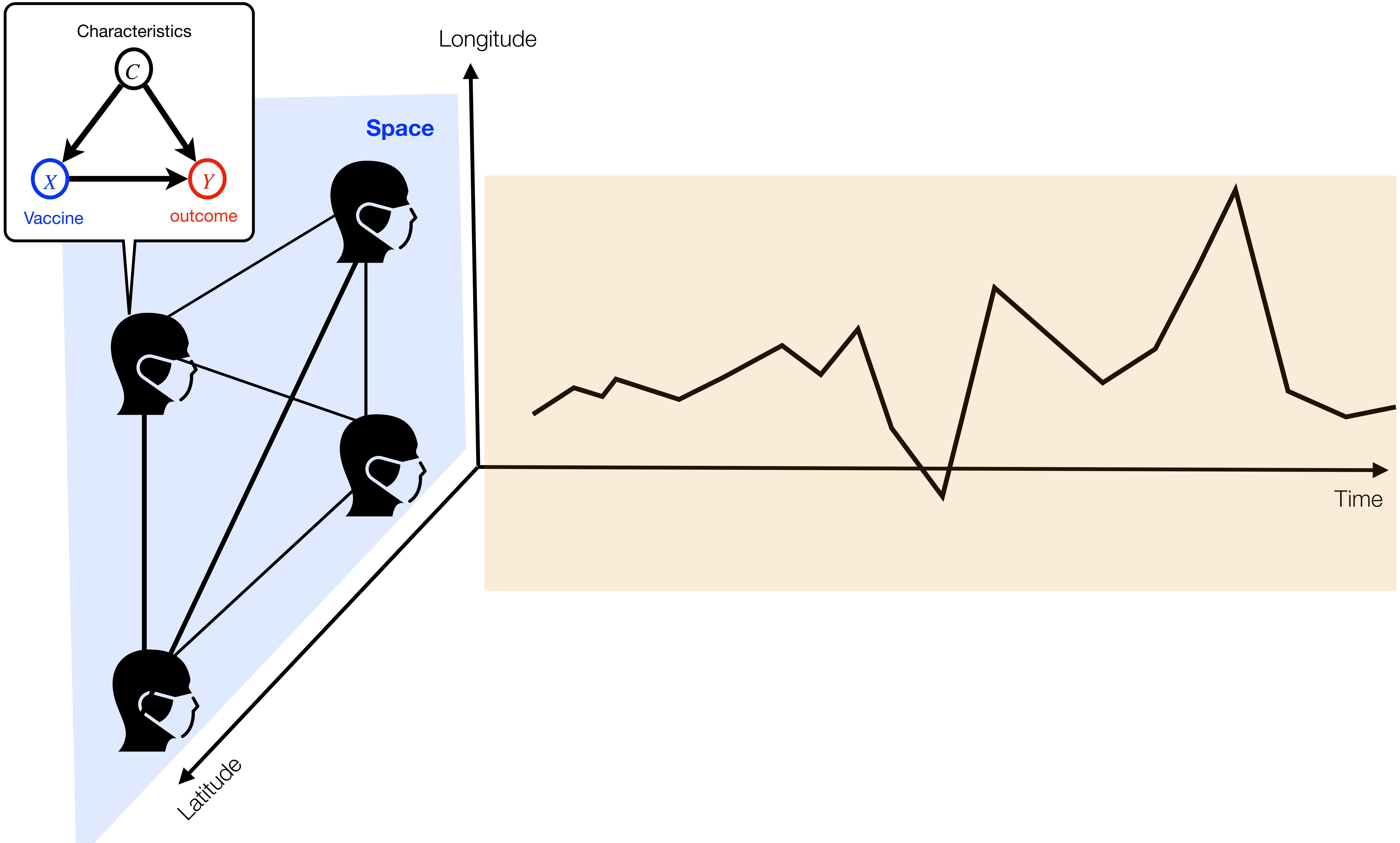
Future 2: Causal Inference with Spatiotemporal Data



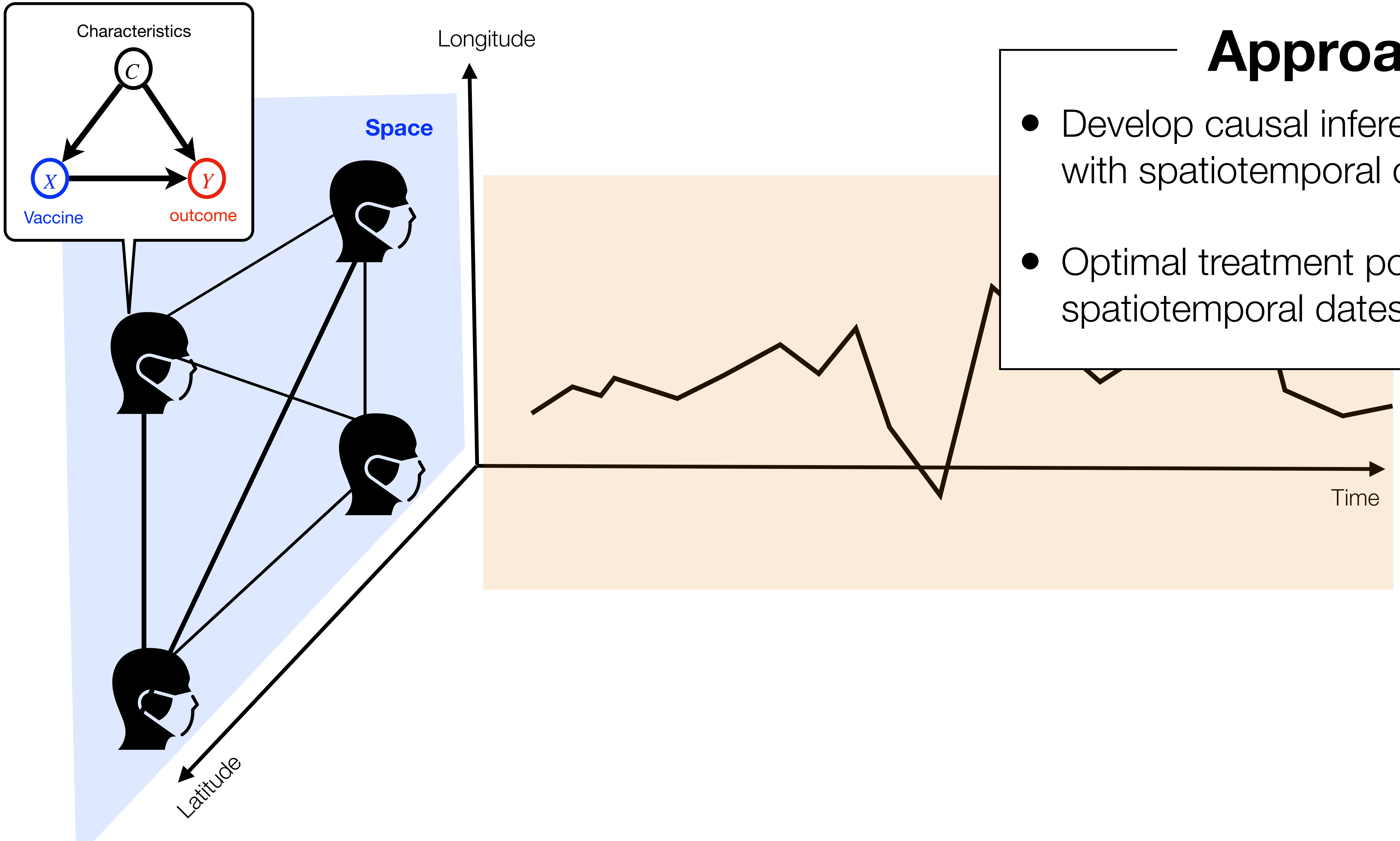
Future 2: Causal Inference with Spatiotemporal Data



Future 2: Causal Inference with Spatiotemporal Data



Future 2: Causal Inference with Spatiotemporal Data



Approach

- Develop causal inference methods with spatiotemporal dataset
- Optimal treatment policy with spatiotemporal dates

Future 3: Causal Inference Loop with Uncertainty

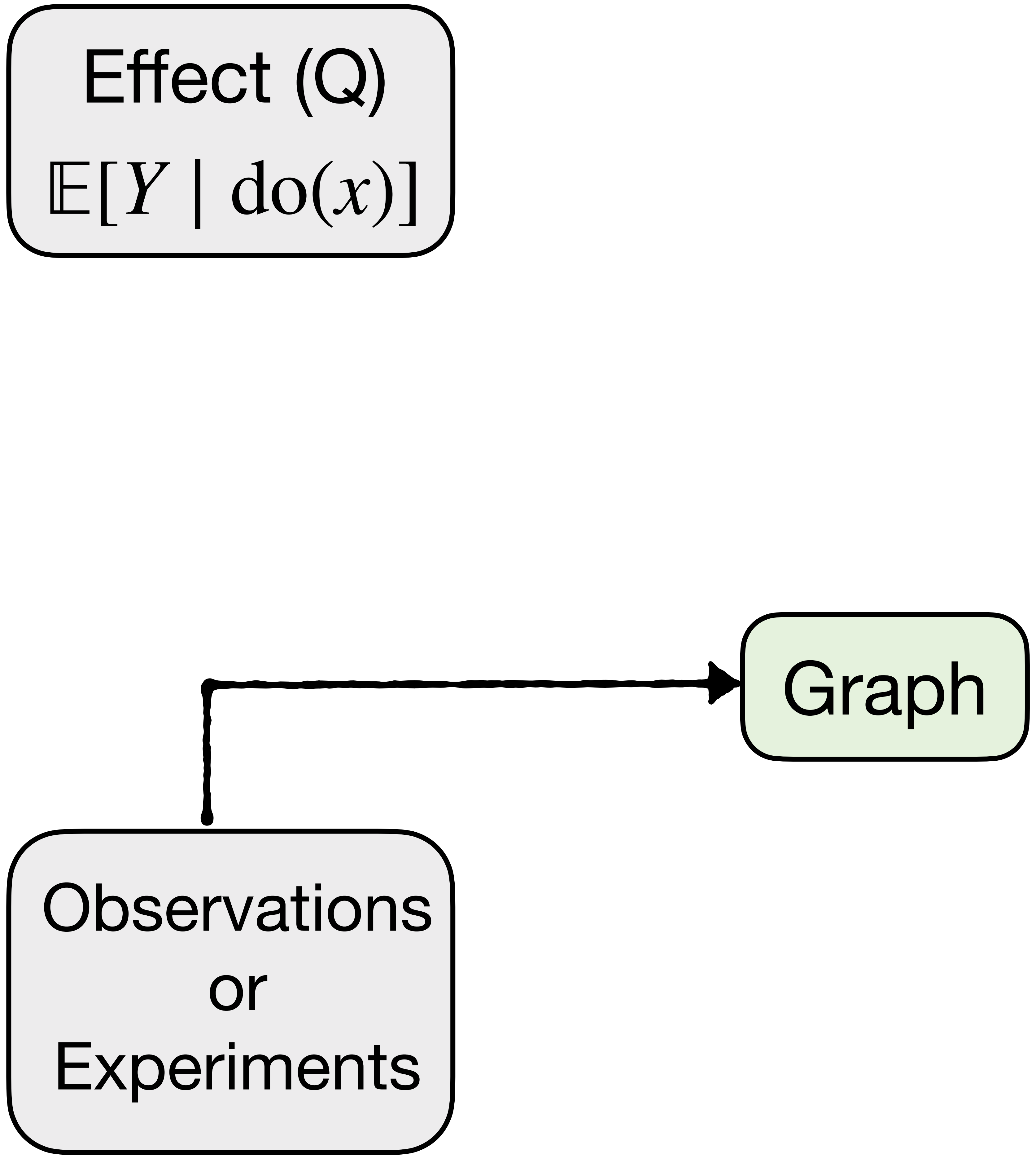
Future 3: Causal Inference Loop with Uncertainty

Effect (Q)

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

Observations
or
Experiments

Future 3: Causal Inference Loop with Uncertainty



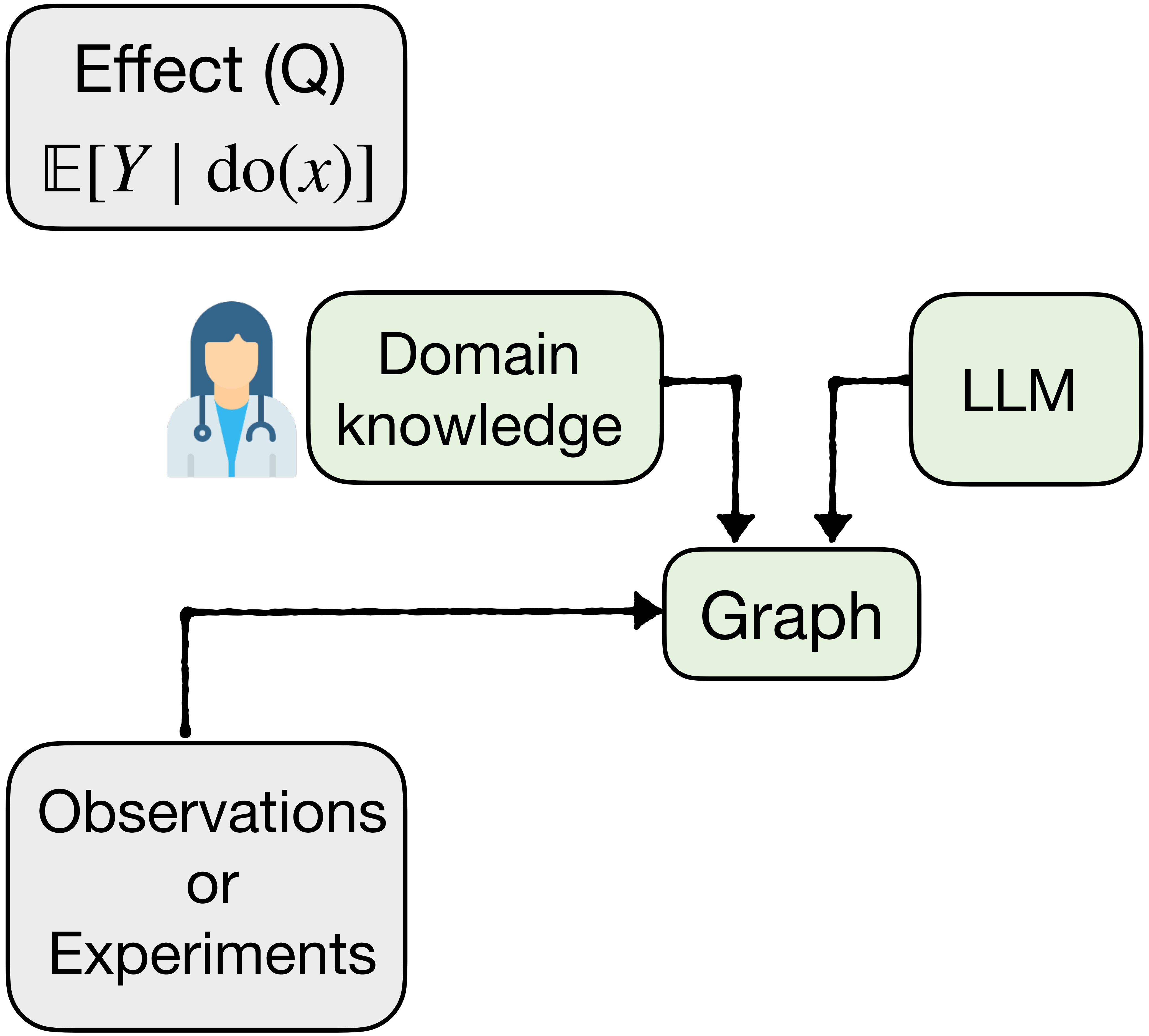
Effect (Q)

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

Graph

Observations
or
Experiments

Future 3: Causal Inference Loop with Uncertainty



Effect (Q)

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

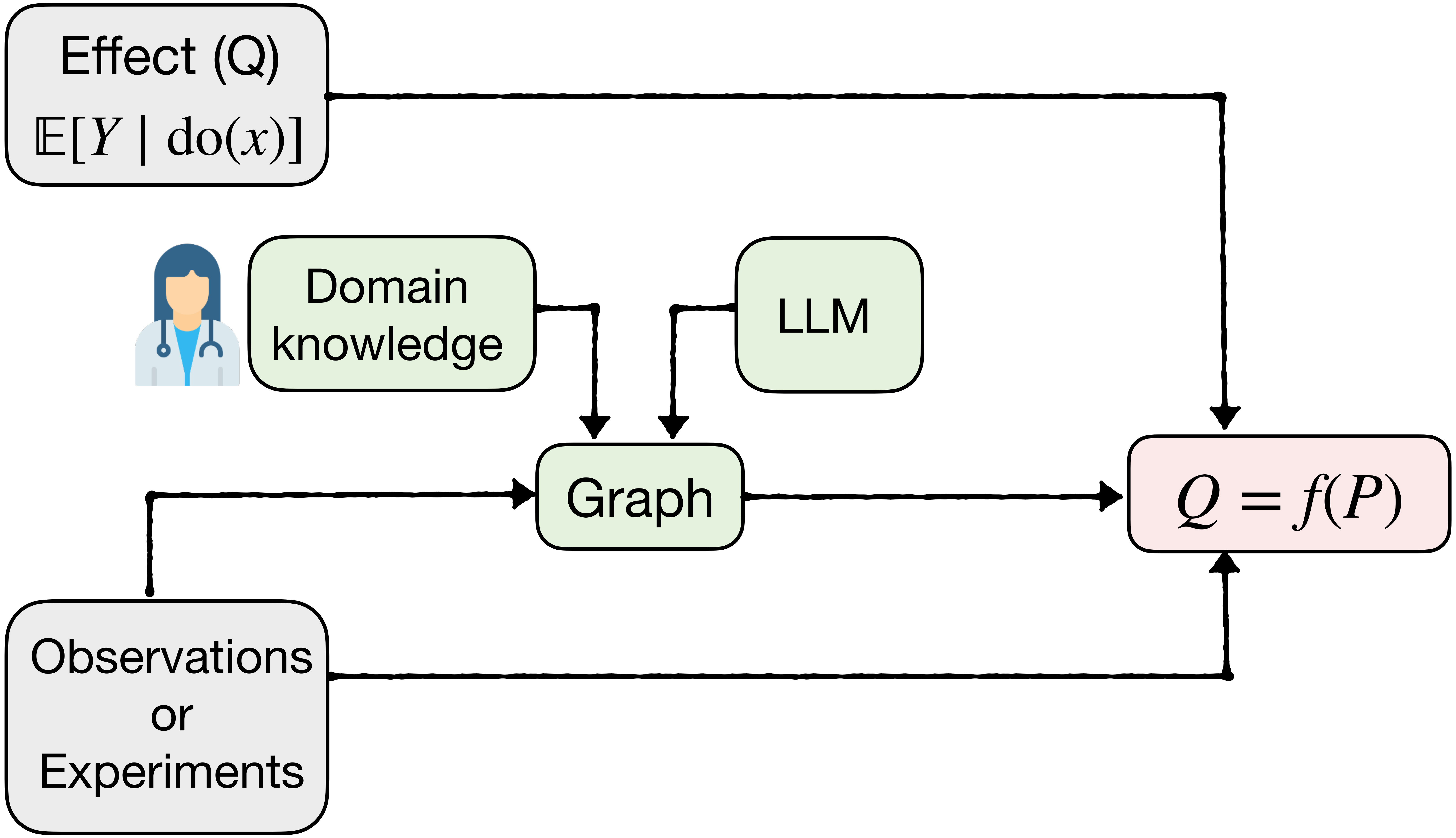
Domain knowledge

LLM

Graph

Observations or Experiments

Future 3: Causal Inference Loop with Uncertainty



Effect (Q)

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

Domain knowledge

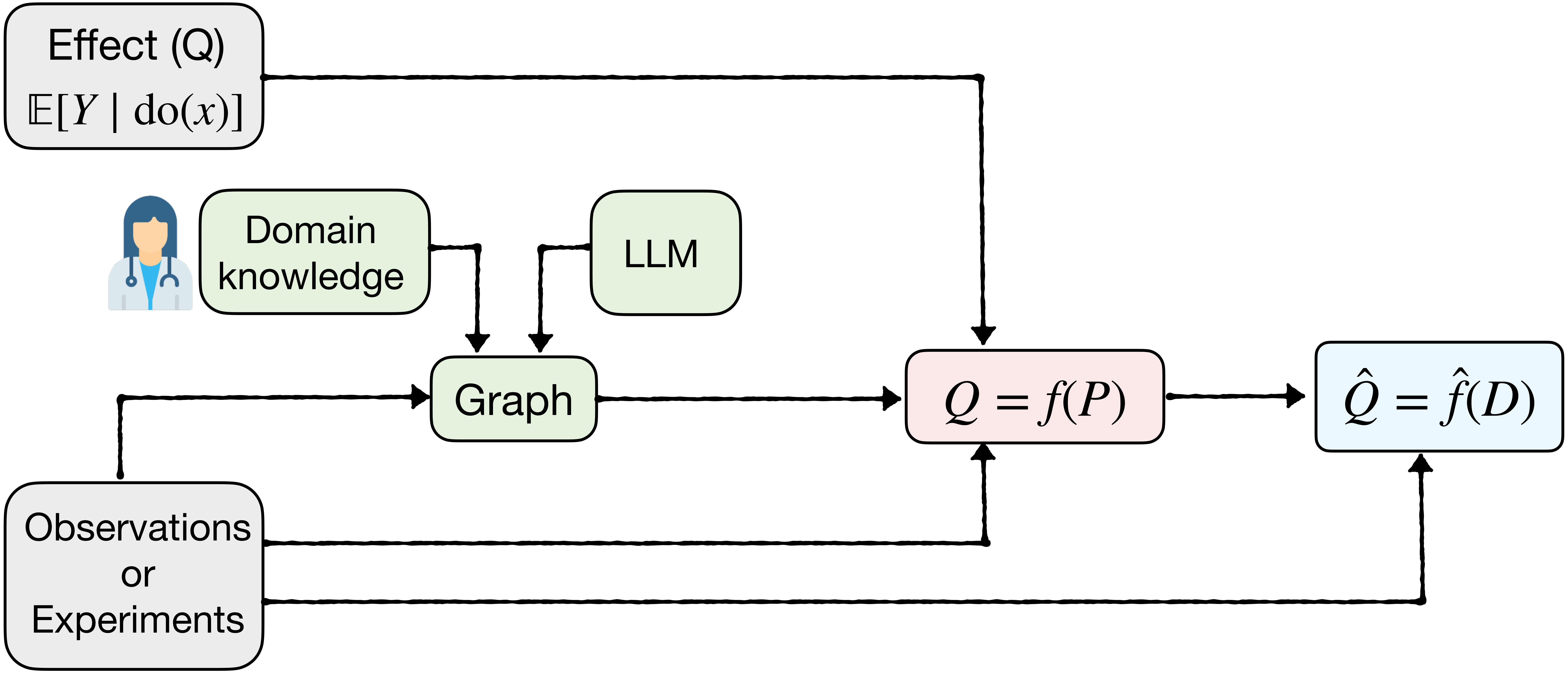
LLM

Graph

$Q = f(P)$

Observations or Experiments

Future 3: Causal Inference Loop with Uncertainty



Effect (Q)

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

Domain knowledge

LLM

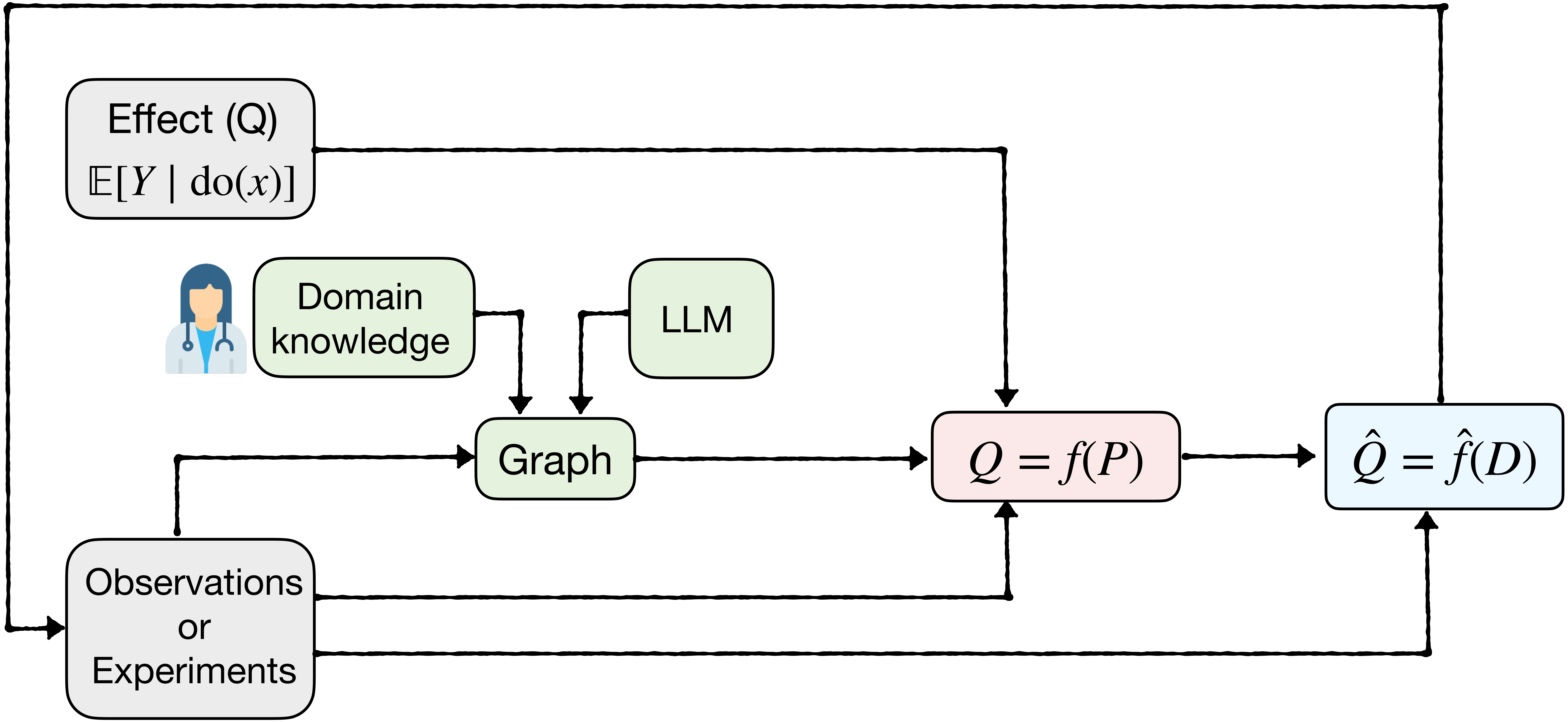
Graph

$Q = f(P)$

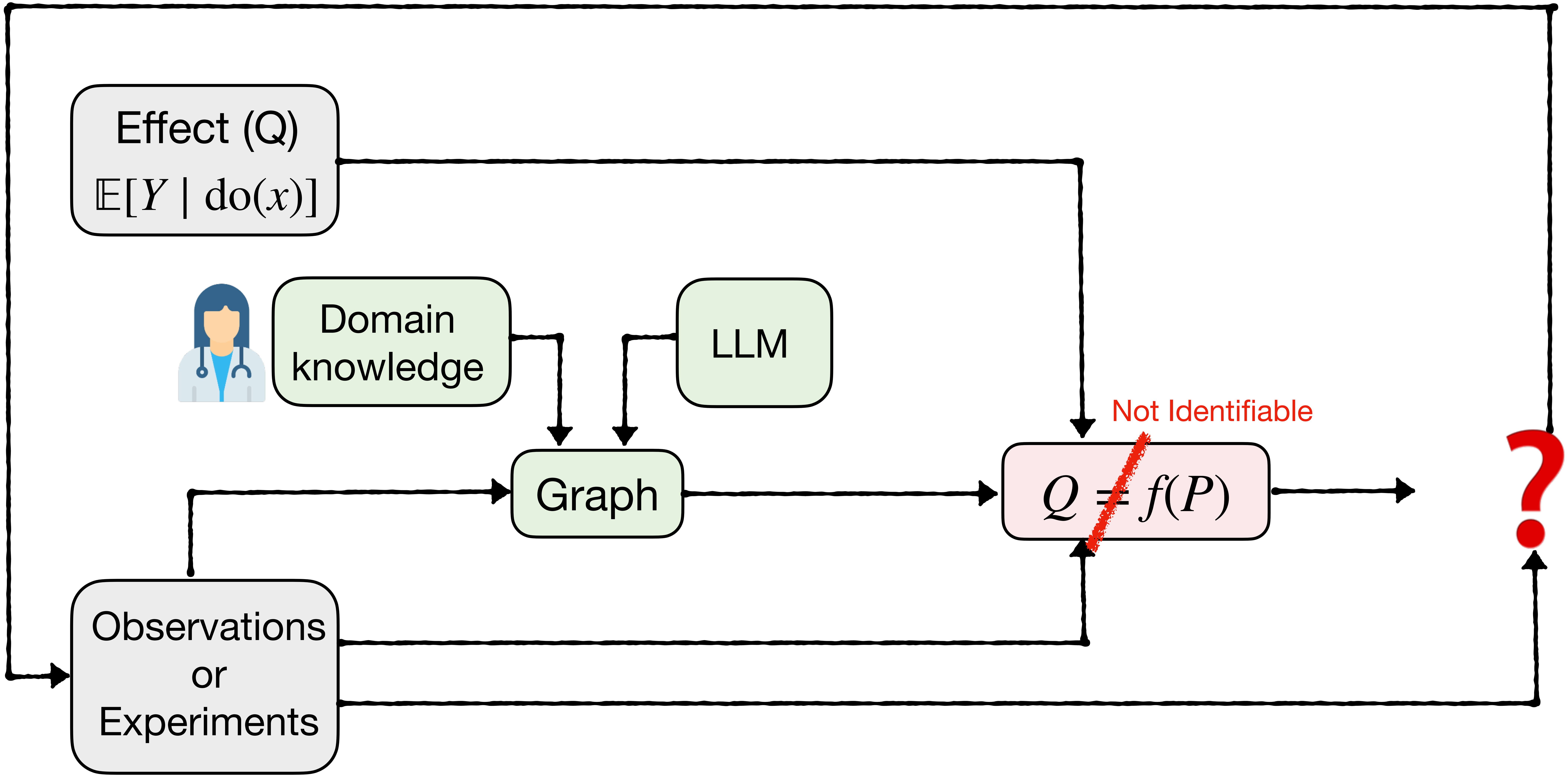
$\hat{Q} = \hat{f}(D)$

Observations or Experiments

Future 3: Causal Inference Loop with Uncertainty



Future 3: Causal Inference Loop with Uncertainty



Effect (Q)

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

Domain knowledge

LLM

Graph

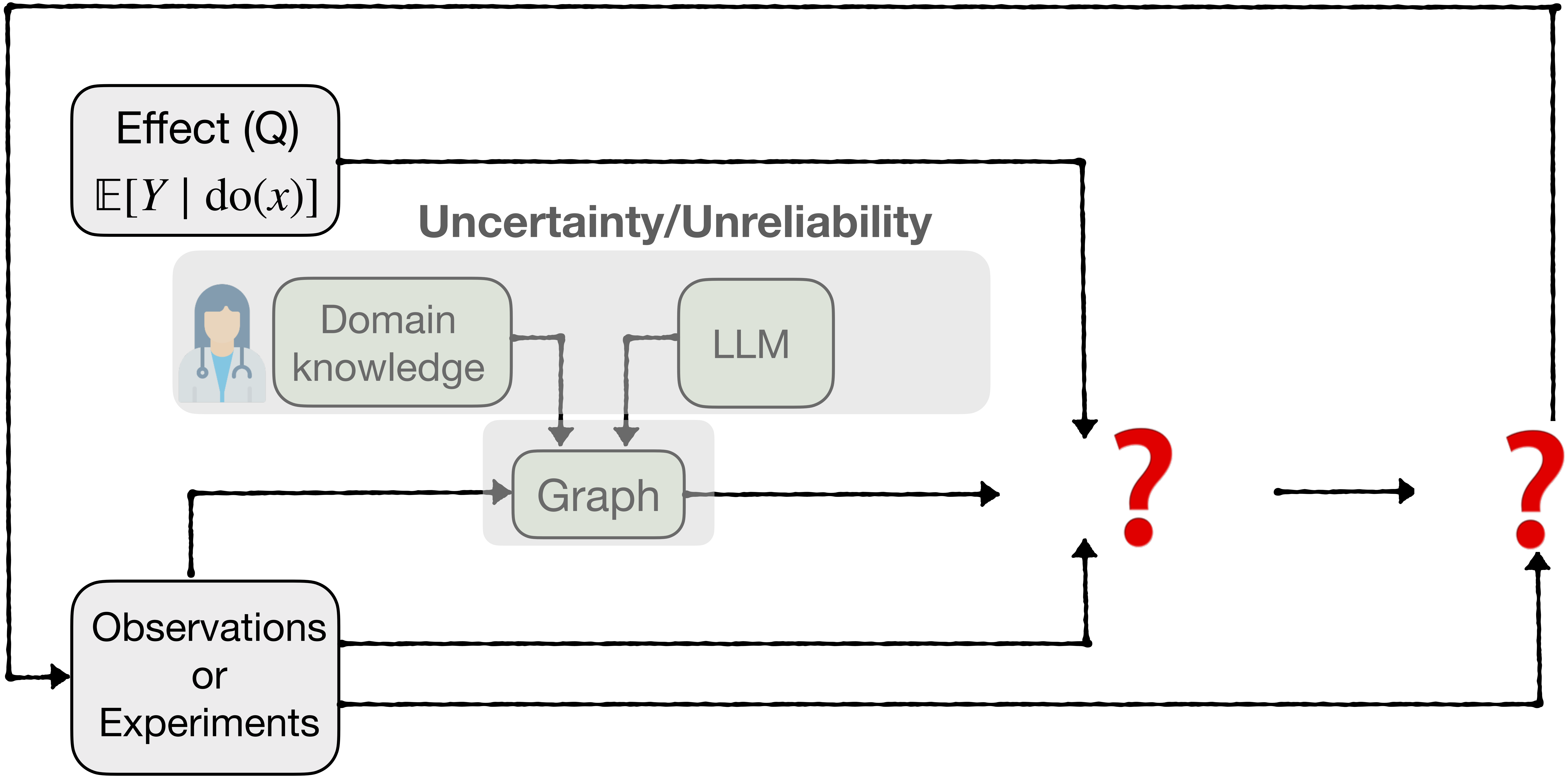
Not Identifiable

$Q \cancel{=} f(P)$

?

Observations or Experiments

Future 3: Causal Inference Loop with Uncertainty



Effect (Q)

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

Uncertainty/Unreliability

Domain knowledge

LLM

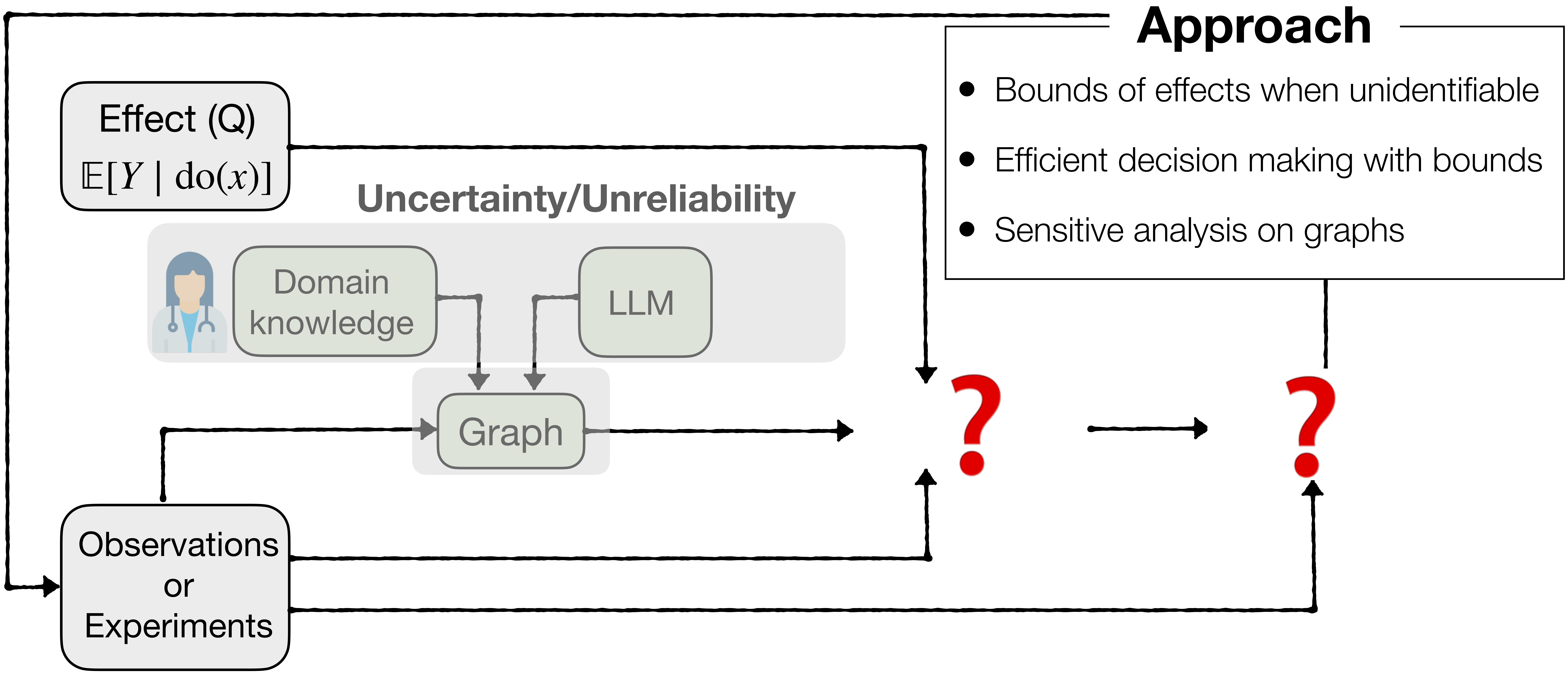
Graph

Observations or Experiments

?

?

Future 3: Causal Inference Loop with Uncertainty



Effect (Q)

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

Uncertainty/Unreliability

Domain knowledge

LLM

Graph

Observations or Experiments

?

?

Approach

- Bounds of effects when unidentifiable
- Efficient decision making with bounds
- Sensitive analysis on graphs