Causal Data Science:

Estimating Identifiable Causal Effects

Yonghan Jung

Department of Computer Science

Purdue University

www.yonghanjung.me

PhD Defense, Purdue Computer Science | June 06, 2025

References

1. **Jung, Y.**, Tian, J. and Bareinboim, E. (AAAI-2021)
   Estimating Identifiable Causal Effects through Double Machine Learning.

2. **Jung, Y.**, Tian, J. and Bareinboim, E. (ICML-2023)
   Estimating Joint Treatment Effects by Combining Multiple Experiments.

3. **Jung, Y.**, Tian, J., Díaz, I. and Bareinboim, E. (NeurIPS-2023)
   Estimating Causal Effects Identifiable from a Combination of Observations and Experiments.

4. **Jung, Y.**, Tian, J. and Bareinboim, E. (NeurIPS-2024)
   Unified Covariate Adjustment for Causal Inference

5. **Jung, Y.**, Park, W., and Lee, S. (NeurIPS-2024)
   Complete Graphical Criterion for Sequential Covariate Adjustment in Causal Inference

References (II)

6. **Jung, Y.**, Tian, J. and Bareinboim, E.. (AAAI-2020)
   Estimating Causal Effects using Weighting-based Estimators.
7. **Jung, Y.**, Tian, J. and Bareinboim, E. (NeurIPS-2020)
   Learning Causal Effects via Weighted Empirical Risk Minimization.
8. **Jung, Y.**, Tian, J. and Bareinboim, E. (ICML-2021)
   Estimating Identifiable Causal Effects on Markov Equivalence Class through Double Machine Learning
9. **Jung, Y.**, Tian, J. and Bareinboim, E. (NeurIPS-2021)
   Double Machine Learning Density Estimation for Local Treatment Effects with Instruments.
10. **Jung, Y.**, Kasiviswanathan, S., Tian, J., Janzing D., Blöbaum, P., Bareinboim, E. (ICML-2022)
    On Measuring Causal Contributions via do-Interventions.
11. **Jung, Y*.** and Bellot, A*. (NeurIPS-2024)
    Efficient Policy Evaluation Across Multiple Different Experimental Datasets

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

Clinical Infectious Diseases, 2019

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

Clinical Infectious Diseases, 2019

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

CNN, 2020

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

Clinical Infectious Diseases, 2019

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

CNN, 2020

“ *WHO recommends against use of Remdesivir for COVID patients*

CNN, 2020

What’s going on?

Story Behind the Data

Observational Study (FDA)

	Mortality Rate
Remdesivir	11%
Non Remdesivir	20%

Positive Correlation with Lower Mortality

vs.

Randomized Trial (WHO)

	Mortality Rate
Remdesivir	15%
Non Remdesivir	15%

No Causal Effect to Lower Mortality

Story Behind the Data

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

Observational Study (FDA)

	Mortality Rate
Remdesivir	11%
Non Remdesivir	20%

Positive Correlation with Lower Mortality

vs.

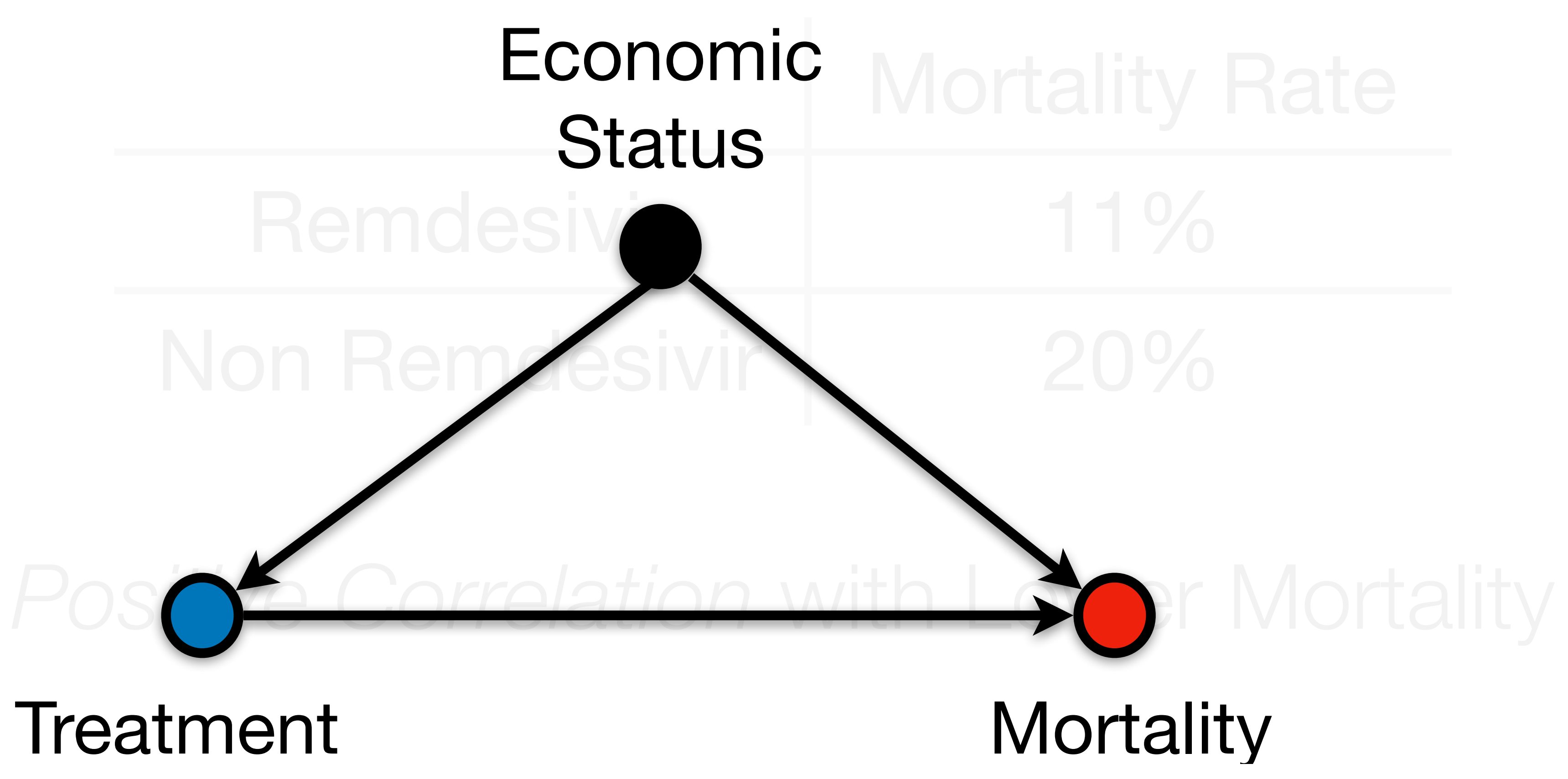
Randomized Trial (WHO)

	Mortality Rate
Remdesivir	15%
Non Remdesivir	15%

No Causal Effect to Lower Mortality

Story Behind the Data

Observational Study (FDA)



vs.

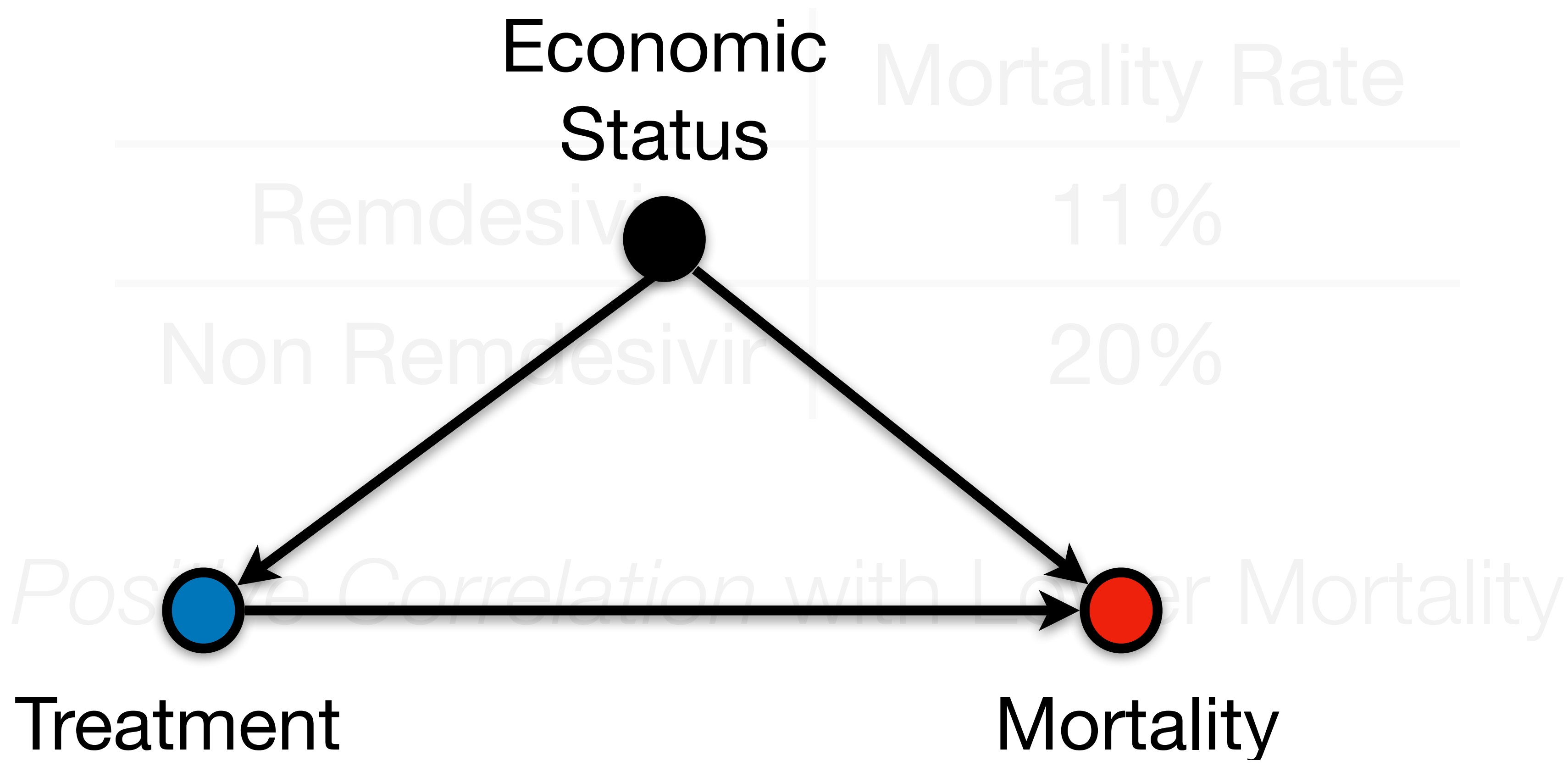
Randomized Trial (WHO)

	Mortality Rate
Remdesivir	15%
Non Remdesivir	15%

No Causal Effect to Lower Mortality

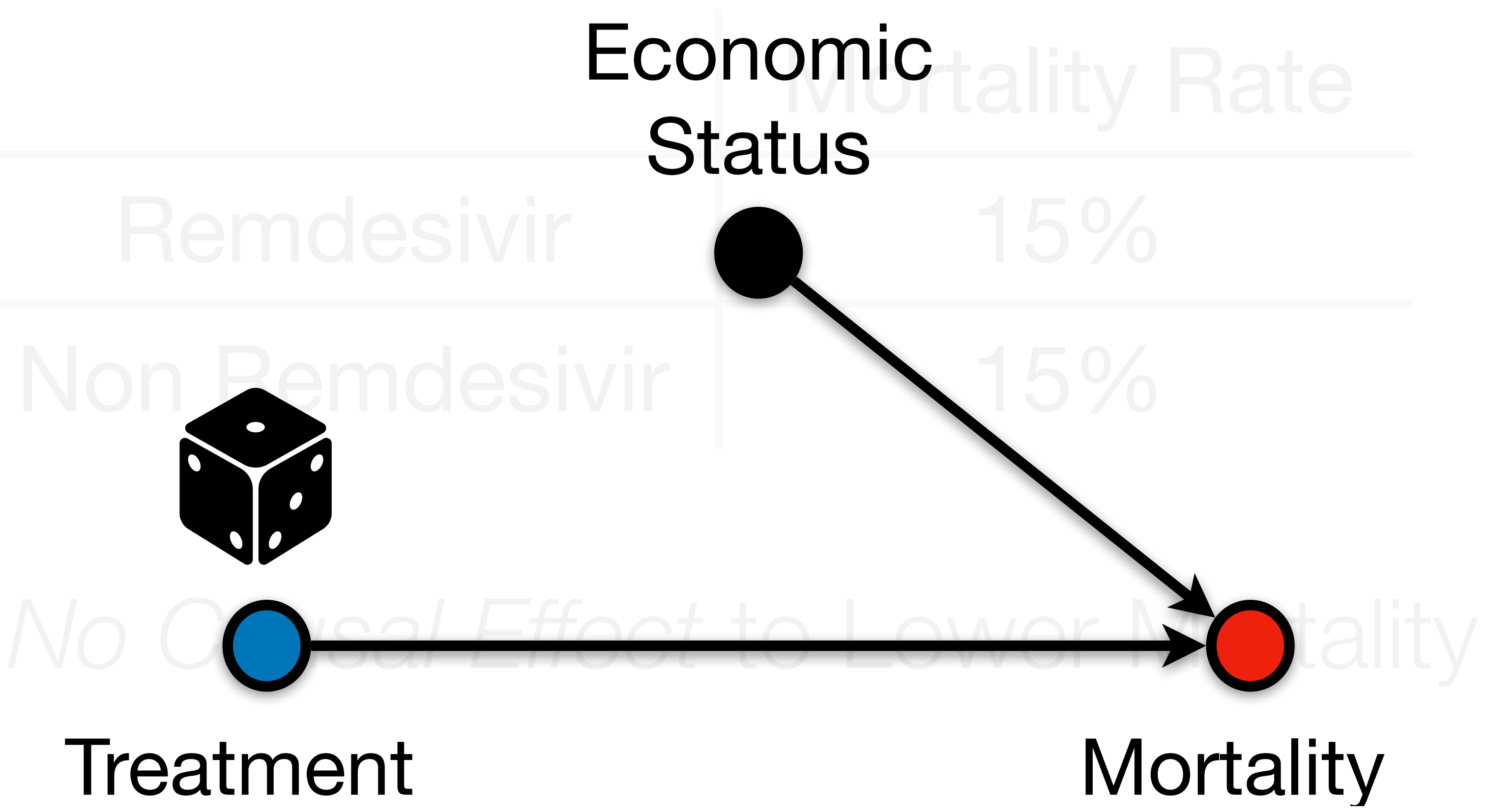
Story Behind the Data

Observational Study (FDA)



vs.

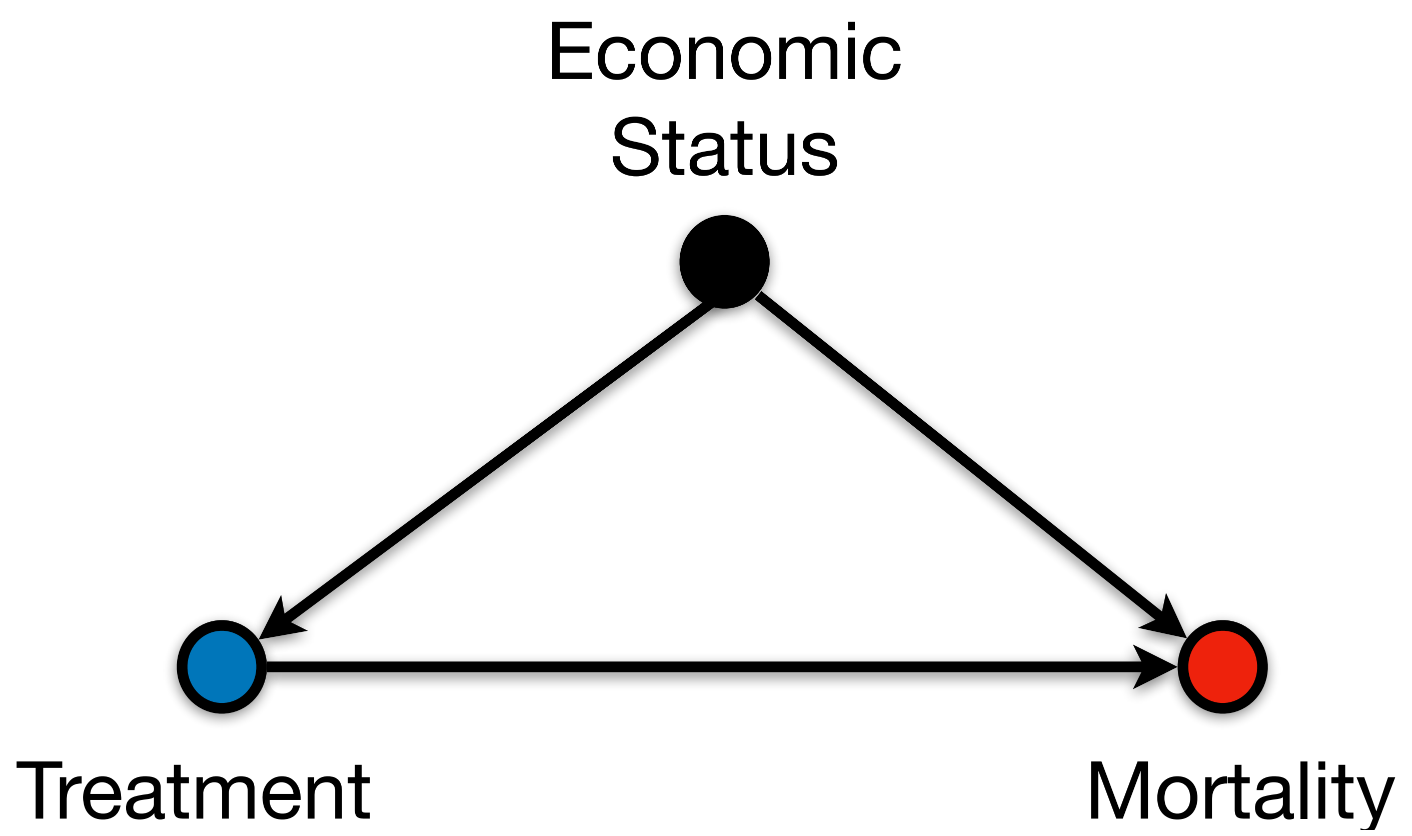
Randomized Trial (WHO)



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%

Structural Causal Models & Causal Graph

Structural Causal Models & Causal Graph

[Def 1] Structural Causal Model

(Pearl 95)

A structural causal model (SCM) is a 4-tuple $\langle \mathbf{V}, \mathbf{U}, \mathbf{F}, P(\mathbf{U}) \rangle$ where

- $\mathbf{V} = \{V_1, \ldots, V_n\}$ are endogenous variables;
- $\mathbf{U} = \{U_1, \ldots, U_m\}$ are exogenous variables;
- $\mathbf{F} = \{f_1, \ldots, f_n\}$are functions determining $\mathbf{V}$ ($V_i \leftarrow f_i(\mathbf{PA}_i, \mathbf{U}_i)$ for $\mathbf{PA}_i \subseteq \mathbf{V}$, $\mathbf{U}_i \subseteq \mathbf{U}$)
- $P(\mathbf{U})$ is a distribution over $\mathbf{U}$.

Structural Causal Models & Causal Graph

[Def 1] Structural Causal Model

(Pearl 95)

A structural causal model (SCM) is a 4-tuple $\langle \mathbf{V}, \mathbf{U}, \mathbf{F}, P(\mathbf{U}) \rangle$ where

- $\mathbf{V} = \{V_1, \ldots, V_n\}$ are endogenous variables;
- $\mathbf{U} = \{U_1, \ldots, U_m\}$ are exogenous variables;
- $\mathbf{F} = \{f_1, \ldots, f_n\}$ are functions determining $\mathbf{V}$ ($V_i \leftarrow f_i(\mathbf{PA}_i, \mathbf{U}_i)$ for $\mathbf{PA}_i \subseteq \mathbf{V}$, $\mathbf{U}_i \subseteq \mathbf{U}$)
- $P(\mathbf{U})$ is a distribution over $\mathbf{U}$.

$$\mathbf{U} = \{U_1, U_2, U_3\}, \mathbf{V} = \{X, Y, Z\},$$

$$\mathbf{F} = \begin{cases} Z & \leftarrow U_1 \\ X & \leftarrow U_1 \oplus U_2 \oplus Z \\ Y & \leftarrow X \oplus Z \oplus U_3 \end{cases}$$

and $\mathbf{U} = \{U_1, U_2, U_3\}$ are independent.

Structural Causal Models & Causal Graph

[Def 1] Structural Causal Model

(Pearl 95)

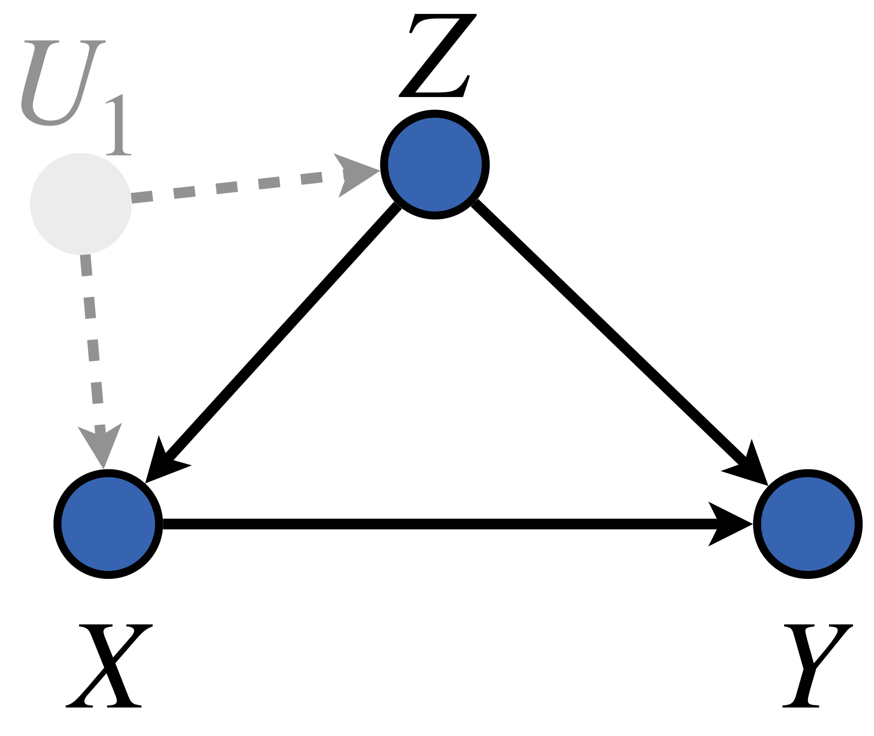
A structural causal model (SCM) is a 4-tuple $\langle \mathbf{V}, \mathbf{U}, \mathbf{F}, P(\mathbf{U}) \rangle$ where

- $\mathbf{V} = \{V_1, \ldots, V_n\}$ are endogenous variables;
- $\mathbf{U} = \{U_1, \ldots, U_m\}$ are exogenous variables;
- $\mathbf{F} = \{f_1, \ldots, f_n\}$ are functions determining $\mathbf{V}$ ($V_i \leftarrow f_i(\mathbf{PA}_i, \mathbf{U}_i)$ for $\mathbf{PA}_i \subseteq \mathbf{V}$, $\mathbf{U}_i \subseteq \mathbf{U}$)
- $P(\mathbf{U})$ is a distribution over $\mathbf{U}$.

$$\mathbf{U} = \{U_1, U_2, U_3\}, \mathbf{V} = \{X, Y, Z\},$$

$$\mathbf{F} = \begin{cases} Z & \leftarrow U_1 \\ X & \leftarrow U_1 \oplus U_2 \oplus Z \\ Y & \leftarrow X \oplus Z \oplus U_3 \end{cases}$$

and $\mathbf{U} = \{U_1, U_2, U_3\}$ are independent.



Causal Diagram $\mathcal{G}$

Do-Interventions & Counterfactual

Do-Interventions & Counterfactual

[Def 4] Intervention on $\mathbf{X} = \mathbf{x}$

(Pearl 95)

For an SCM $\mathscr{M} \triangleq \langle \mathbf{V}, \mathbf{U}, \mathbf{F}, P(\mathbf{U}) \rangle$, an *intervention* is to replace $\mathbf{F}$ to
$\mathbf{F}_{\mathbf{x}} \triangleq \{f_i : V_i \notin \mathbf{X}\} \cup \{\mathbf{X} \leftarrow \mathbf{x}\}$ (“do($\mathbf{x}$)”),

which induces an *interventional SCM*
$\mathscr{M}_{\mathbf{x}} \triangleq \langle \mathbf{V}, \mathbf{U}, \mathbf{F}_{\mathbf{x}}, P(\mathbf{U}) \rangle$

Do-Interventions & Counterfactual

[Def 4] Intervention on $\mathbf{X} = \mathbf{x}$

(Pearl 95)

For an SCM $\mathscr{M} \triangleq \langle \mathbf{V}, \mathbf{U}, \mathbf{F}, P(\mathbf{U}) \rangle$, an *intervention* is to replace $\mathbf{F}$ to $\mathbf{F}_{\mathbf{x}} \triangleq \{f_i : V_i \notin \mathbf{X}\} \cup \{\mathbf{X} \leftarrow \mathbf{x}\}$ ("do($\mathbf{x}$)"),

which induces an *interventional SCM* $\mathscr{M}_{\mathbf{x}} \triangleq \langle \mathbf{V}, \mathbf{U}, \mathbf{F}_{\mathbf{x}}, P(\mathbf{U}) \rangle$

Potential Response (Pearl 2000)

The potential response of $\mathbf{Y} \subseteq \mathbf{V}$ to an intervention $\text{do}(\mathbf{x})$ is $\mathbf{Y}_{\mathbf{x}} \triangleq \mathbf{Y}_{\mathscr{M}_{\mathbf{x}}}$, induced by the interventional SCM

Do-Interventions & Counterfactual

[Def 4] Intervention on $\mathbf{X} = \mathbf{x}$

(Pearl 95)

For an SCM $\mathcal{M} \triangleq \langle \mathbf{V}, \mathbf{U}, \mathbf{F}, P(\mathbf{U}) \rangle$, an *intervention* is to replace $\mathbf{F}$ to

$\mathbf{F}_{\mathbf{x}} \triangleq \{f_i : V_i \notin \mathbf{X}\} \cup \{\mathbf{X} \leftarrow \mathbf{x}\}$ ("do($\mathbf{x}$)"),

which induces an *interventional SCM*

$\mathcal{M}_{\mathbf{x}} \triangleq \langle \mathbf{V}, \mathbf{U}, \mathbf{F}_{\mathbf{x}}, P(\mathbf{U}) \rangle$

Potential Response (Pearl 2000)

The potential response of $\mathbf{Y} \subseteq \mathbf{V}$ to an intervention $\mathrm{do}(\mathbf{x})$ is $\mathbf{Y}_{\mathbf{x}} \triangleq \mathbf{Y}_{\mathcal{M}_{\mathbf{x}}}$, induced by the interventional SCM

$$\mathbf{U} = \{U_1, U_2, U_3\}, \mathbf{V} = \{X, Y, Z\},$$

$$\mathbf{F}_x = \begin{cases} Z & \leftarrow U_1 \\ X & \leftarrow x \\ Y_x & \leftarrow x \oplus Z \oplus U_3 \end{cases}$$

and (U_1, U_2, U_3) are independent

Do-Interventions & Counterfactual

[Def 4] Intervention on $\mathbf{X} = \mathbf{x}$

(Pearl 95)

For an SCM $\mathcal{M} \triangleq \langle \mathbf{V}, \mathbf{U}, \mathbf{F}, P(\mathbf{U}) \rangle$, an *intervention* is to replace $\mathbf{F}$ to
$\mathbf{F}_{\mathbf{x}} \triangleq \{f_i : V_i \notin \mathbf{X}\} \cup \{\mathbf{X} \leftarrow \mathbf{x}\}$ ("do($\mathbf{x}$)"),

which induces an *interventional SCM*
$\mathcal{M}_{\mathbf{x}} \triangleq \langle \mathbf{V}, \mathbf{U}, \mathbf{F}_{\mathbf{x}}, P(\mathbf{U}) \rangle$

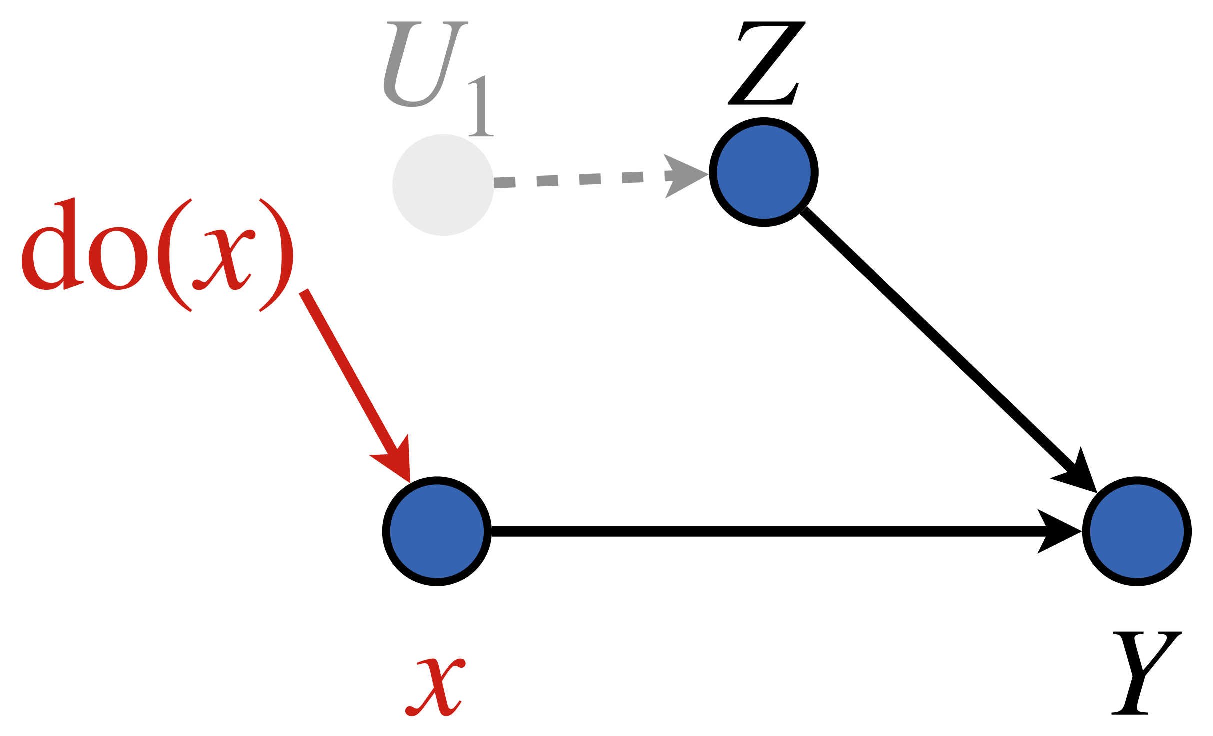
Potential Response (Pearl 2000)

The potential response of $\mathbf{Y} \subseteq \mathbf{V}$ to an intervention $\text{do}(\mathbf{x})$ is $\mathbf{Y}_{\mathbf{x}} \triangleq \mathbf{Y}_{\mathcal{M}_{\mathbf{x}}}$, induced by the interventional SCM

$$\mathbf{U} = \{U_1, U_2, U_3\}, \mathbf{V} = \{X, Y, Z\},$$

$$\mathbf{F}_x = \begin{cases} Z & \leftarrow U_1 \\ X & \leftarrow x \\ Y_x & \leftarrow x \oplus Z \oplus U_3 \end{cases}$$

and (U_1, U_2, U_3) are independent



Causal Diagram $\mathcal{G}_{\overline{X}}$

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	Identification	Estimation
	"When is the causal effect computable from available data?"	"How do we compute the effect from data?"

Effect (Q)

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

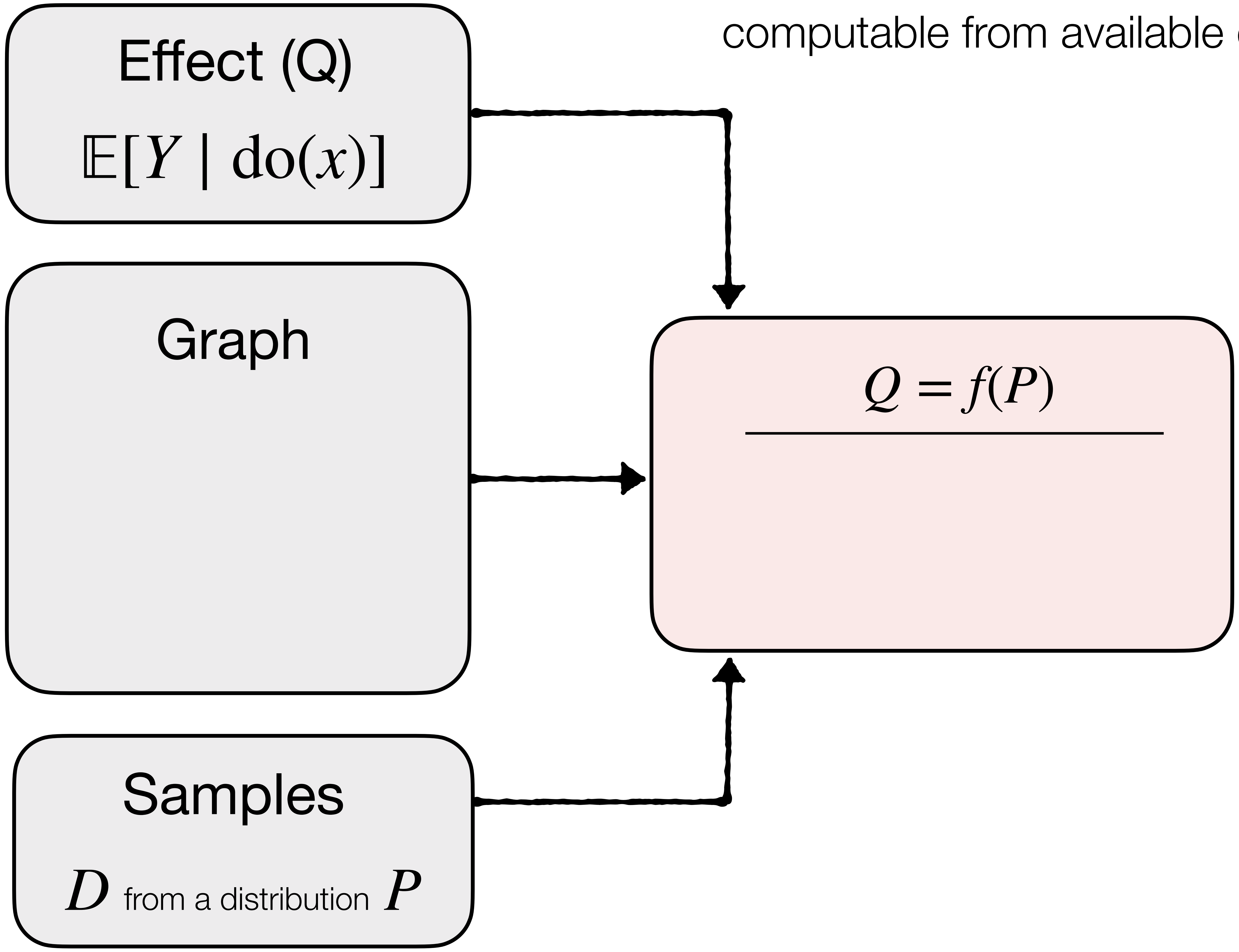
Graph

Samples

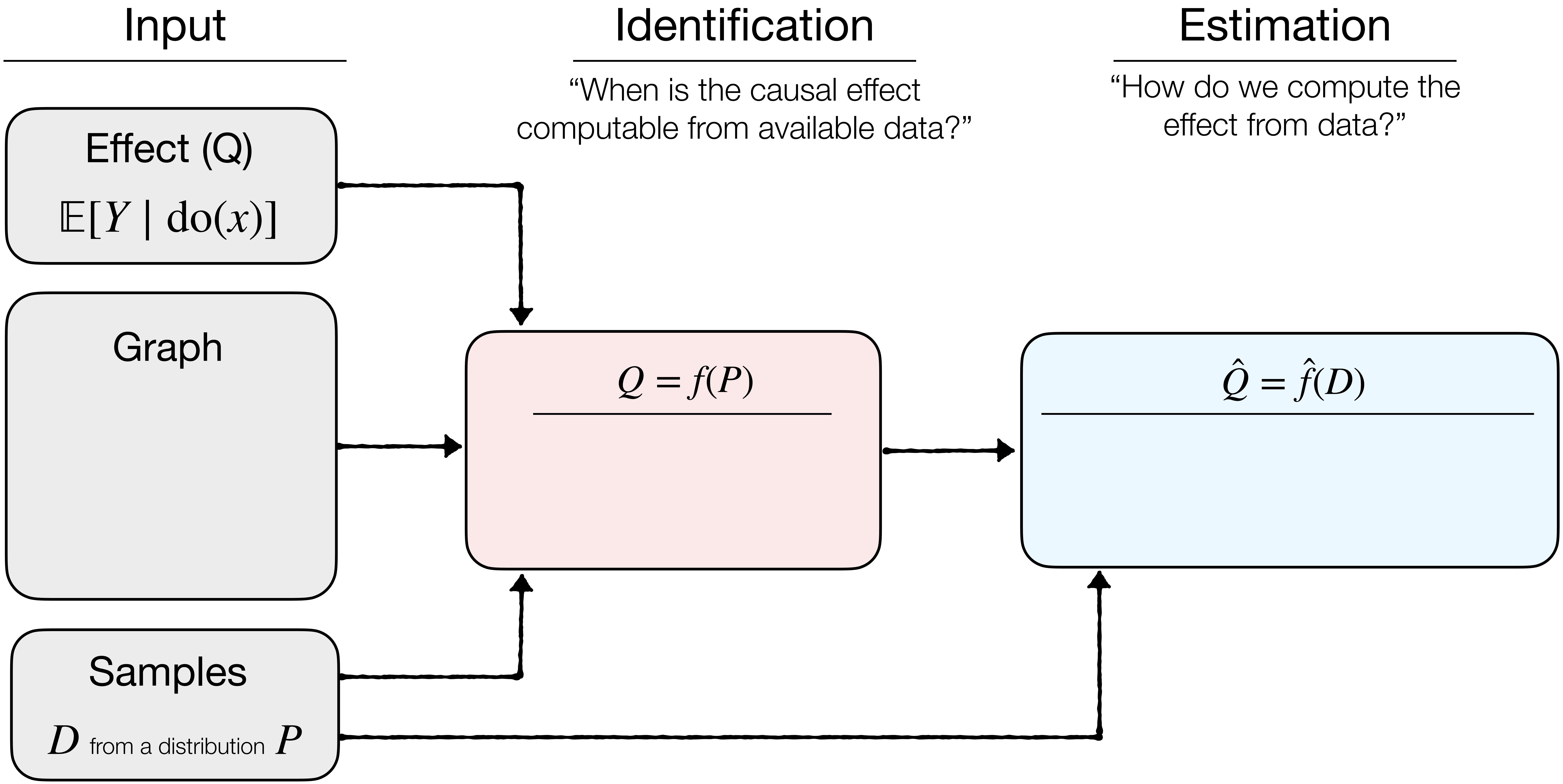
D from a distribution P

Standard Causal Inference Engine

Input	Identification	Estimation
	"When is the causal effect computable from available data?"	"How do we compute the effect from data?"



Standard Causal Inference Engine



Standard Causal Inference Engine

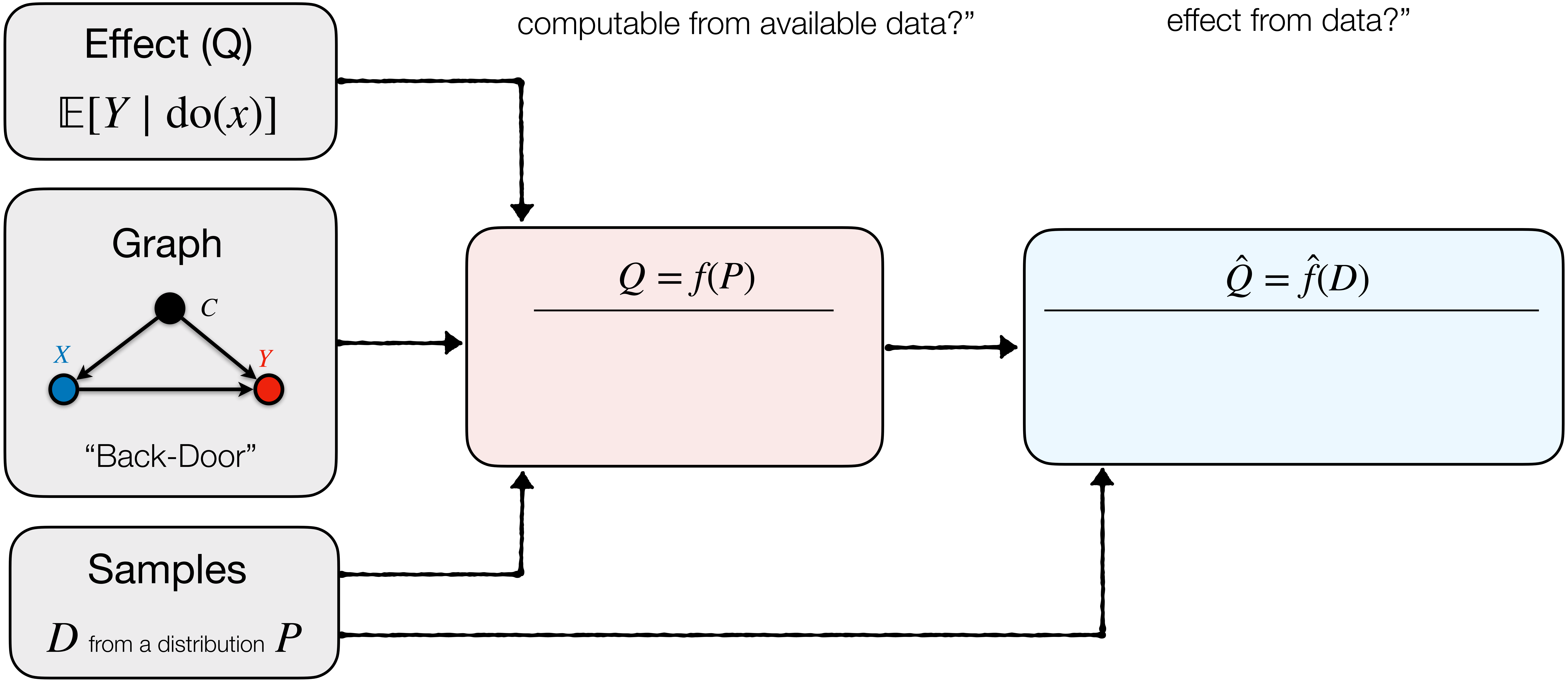
Input

Identification

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

Estimation

"How do we compute the effect from data?"



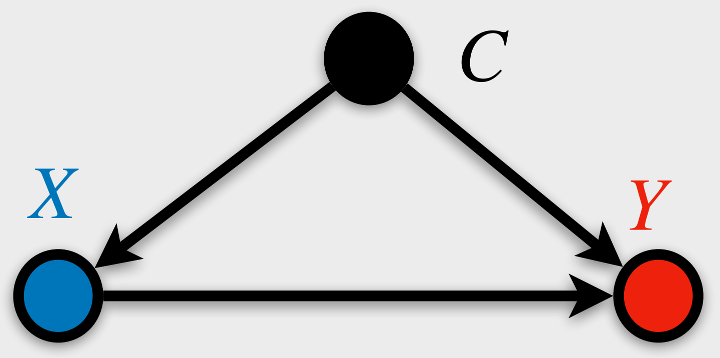
Standard Causal Inference Engine

Input	Identification	Estimation
	"When is the causal effect computable from available data?"	"How do we compute the effect from data?"

Effect (Q)

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

Graph



"Back-Door"

Samples

D from a distribution P

$$Q = f(P)$$

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

"Back-Door Adjustment"

Rubin, 74; Robins, 86; Pearl, 95

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

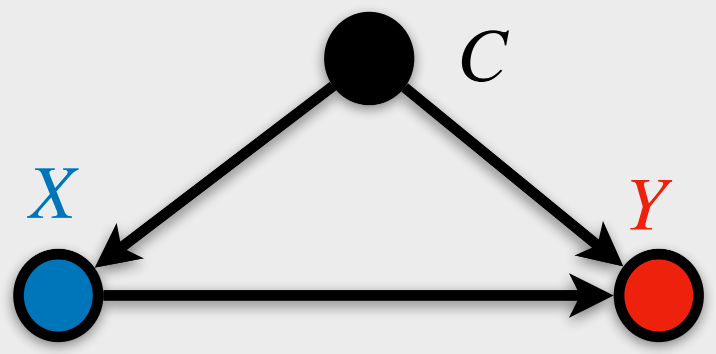
Standard Causal Inference Engine

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

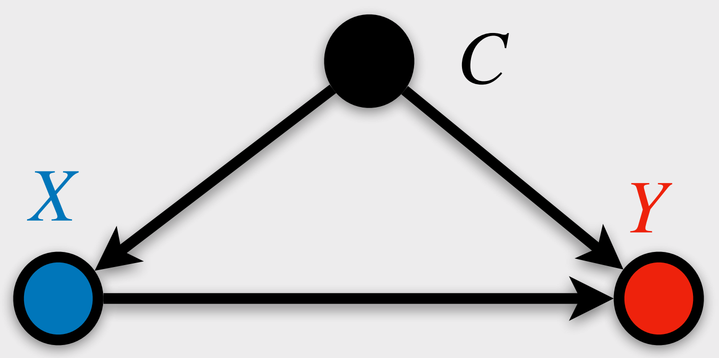
Bickel et al., 93; van der Laan and Rubin 06, van der Laan and Gruber12; Chernozhukov et al, 2018

Challenges in Standard Setting

Effect (Q)

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

Graph



"Back-Door"

Samples

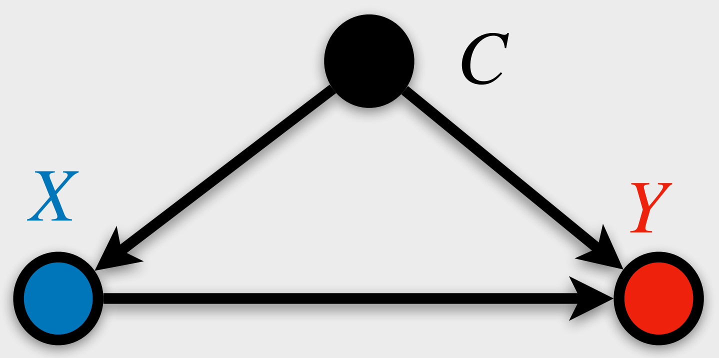
D from P

Challenges in Standard Setting

Effect (Q)

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

Graph



"Back-Door"

Samples

D from P

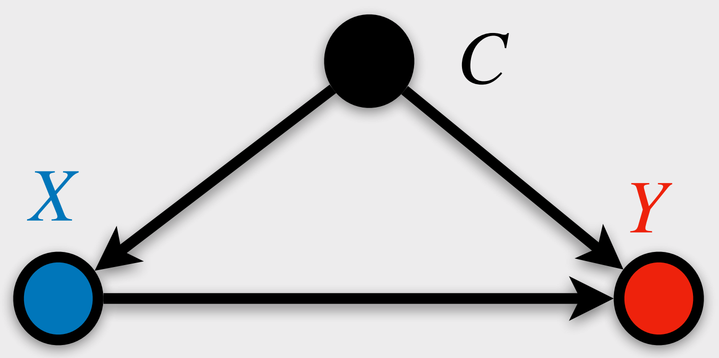
1 Complex dependences

Challenges in Standard Setting

Effect (Q)

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

Graph

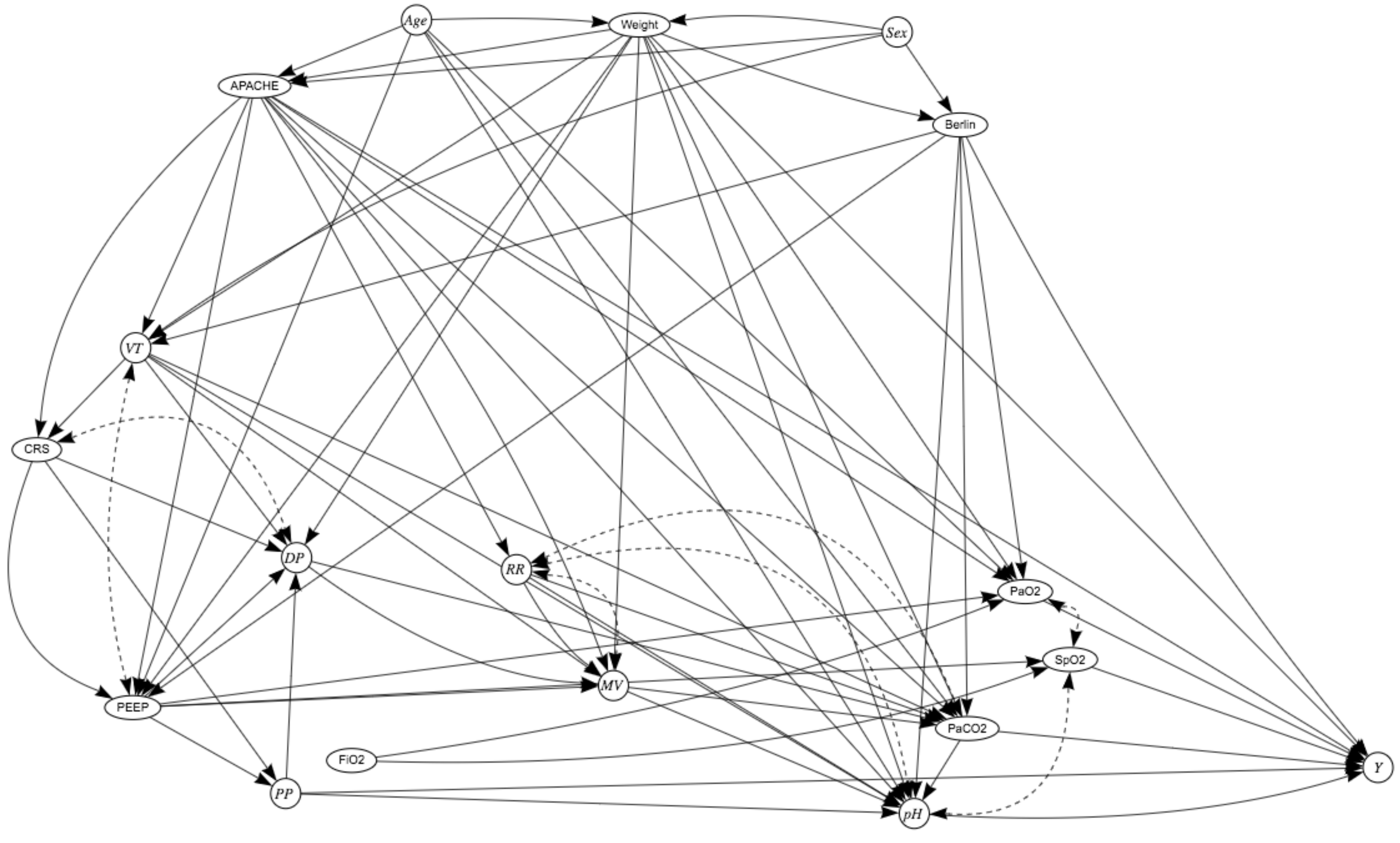


"Back-Door"

Samples

D from P

1 Complex dependences



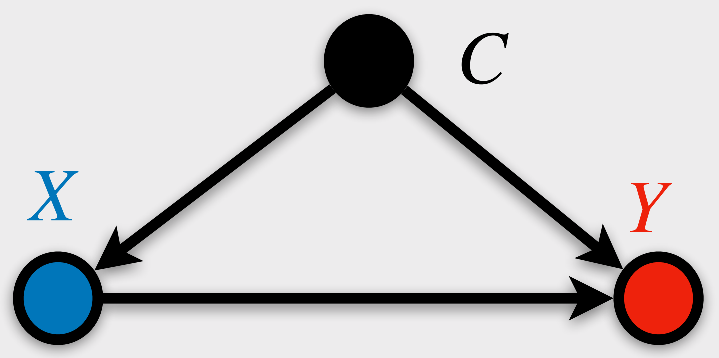
Causal graph on acute respiratory distress syndrome (ARDS)

Challenges in Standard Setting

Effect (Q)

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

Graph



"Back-Door"

Samples

D from P

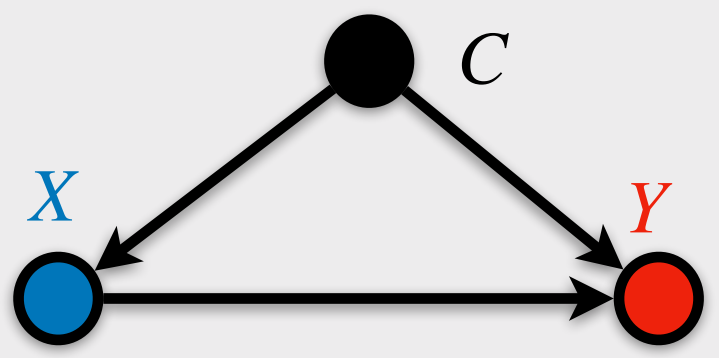
1. Complex dependences
2. Data fusion (observations & experiments)

Challenges in Standard Setting

Effect (Q)

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

Graph

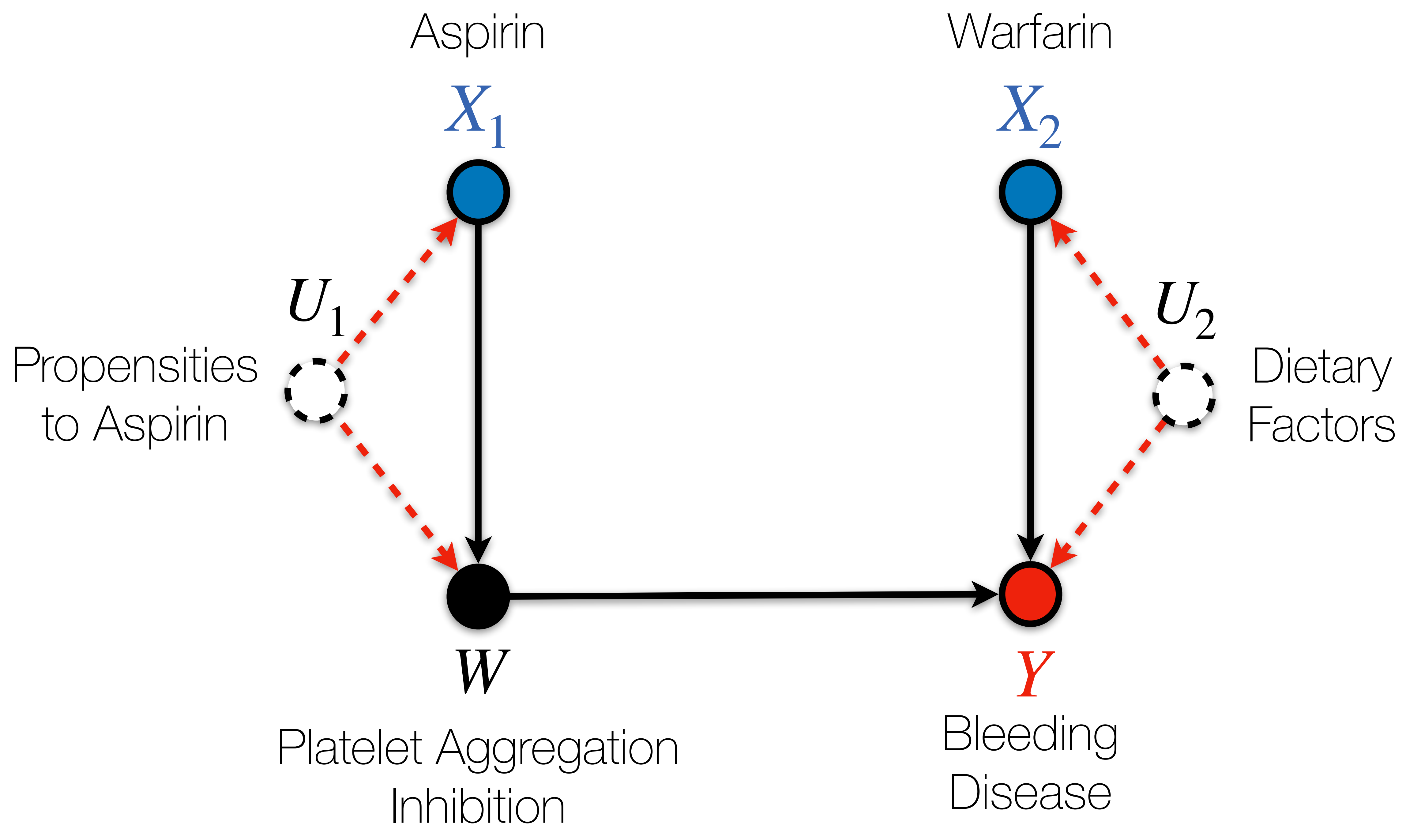


"Back-Door"

Samples

D from P

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



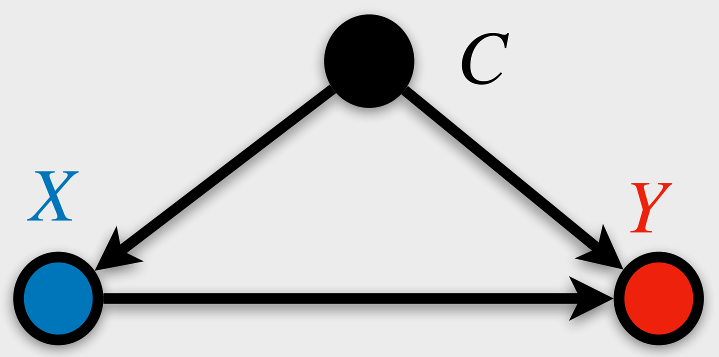
- Goal: Estimate $\mathbb{E}[Y \mid do(x_1, x_2)]$ from single interventions $do(x_1)$ and $do(x_2)$.

Challenges in Standard Setting

Effect (Q)

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

Graph

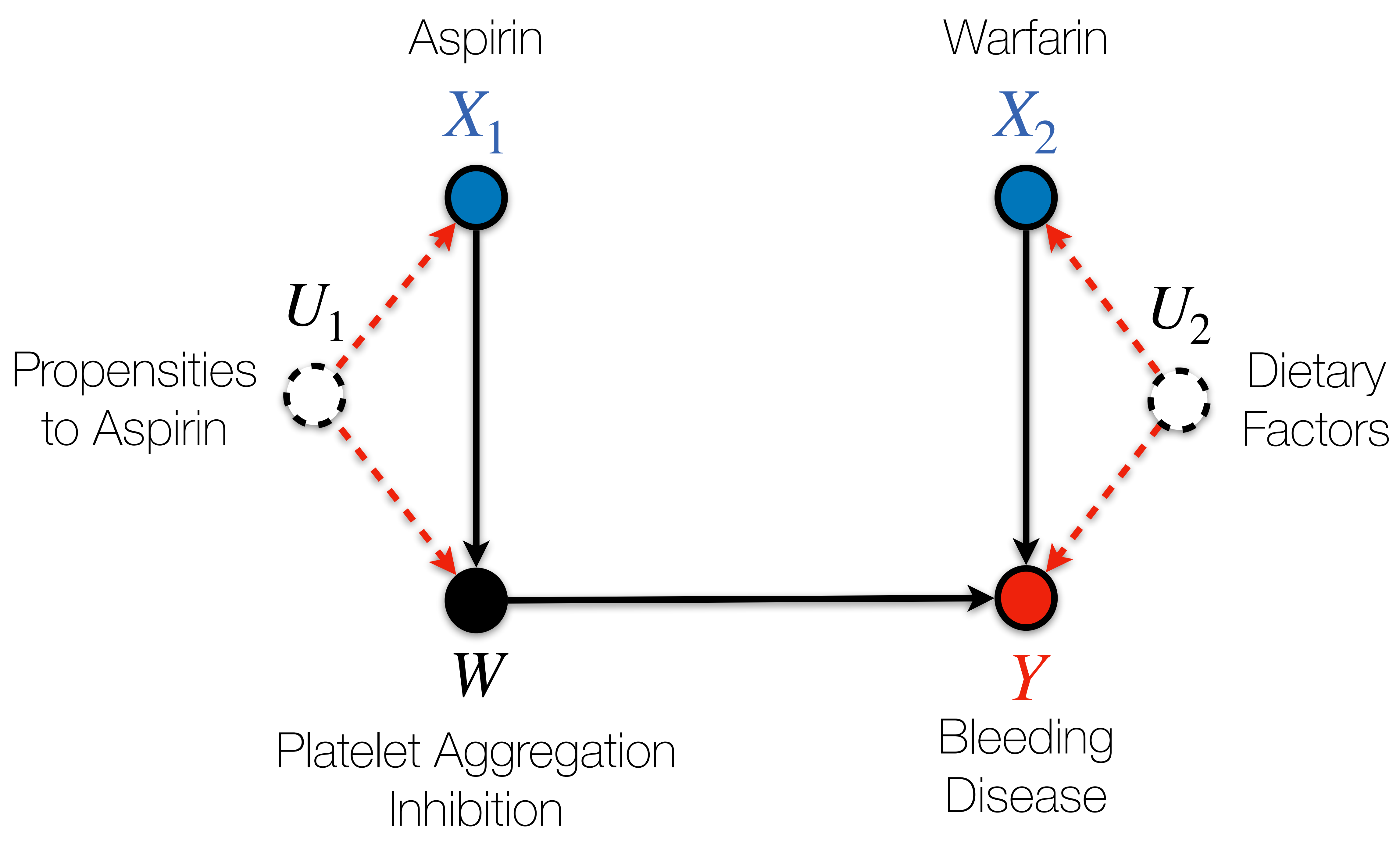


"Back-Door"

Samples

D from P

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



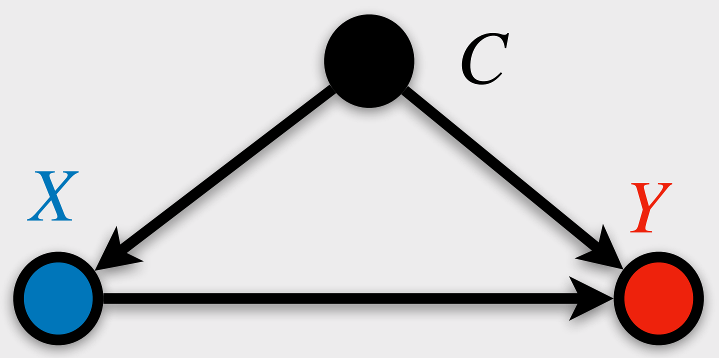
- Goal: Estimate $\mathbb{E}[Y \mid do(x_1, x_2)]$ from single interventions $do(x_1)$ and $do(x_2)$.
- Drug interactions between X_1 and X_2

Challenges in Standard Setting

Effect (Q)

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

Graph

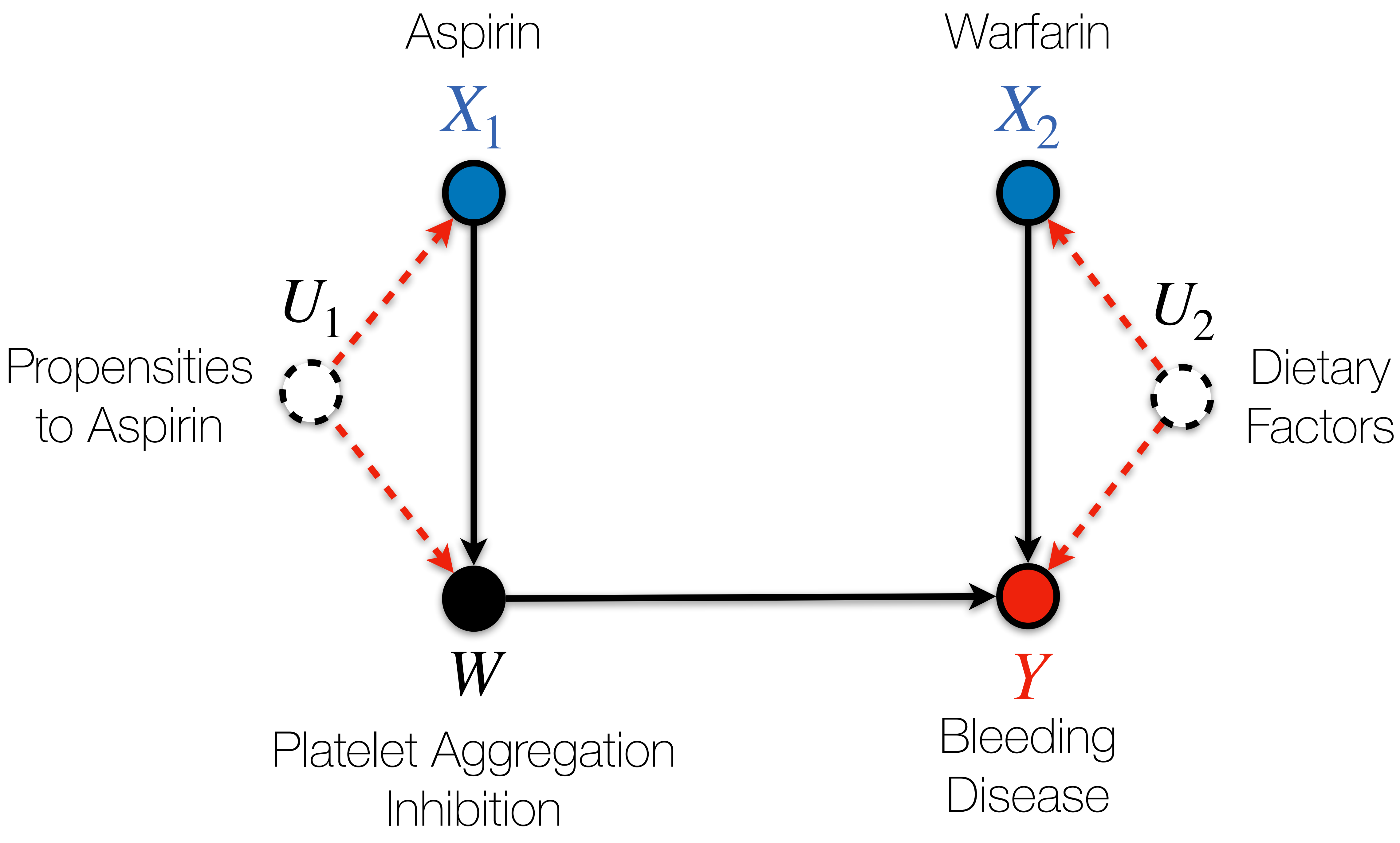


"Back-Door"

Samples

D from P

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



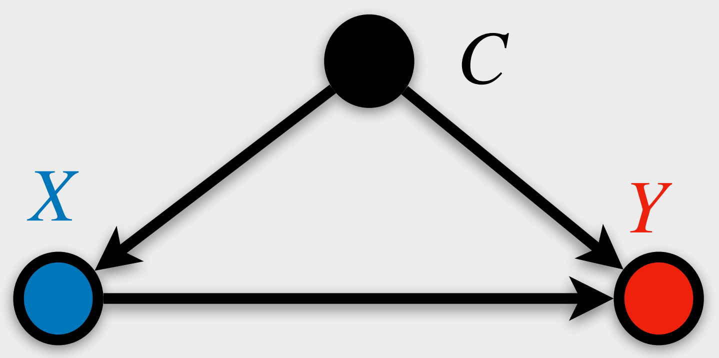
- Goal: Estimate $\mathbb{E}[Y \mid do(x_1, x_2)]$ from single interventions $do(x_1)$ and $do(x_2)$.
- Drug interactions between X_1 and X_2
- Not identifiable from observations

Challenges in Standard Setting

Effect (Q)

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

Graph



"Back-Door"

Samples

D from P

1 Complex dependences

2 Data fusion (observations & experiments)

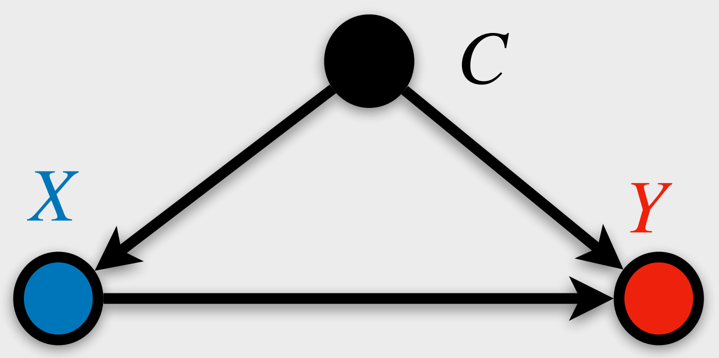
3 More general scenarios

Challenges in Standard Setting

Effect (Q)

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

Graph



"Back-Door"

Samples

D from P

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

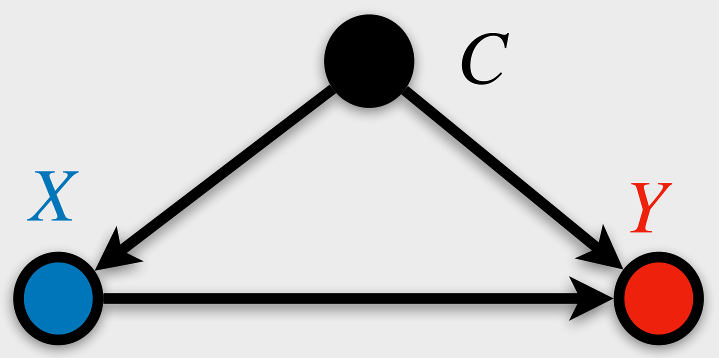
(Fairness) Salary (Y) a man ($X = x$) would earn if he is given the opportunities (M) that other genders ($X \neq x$) had received

Challenges in Standard Setting

Effect (Q)

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

Graph



"Back-Door"

Samples

D from P

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

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

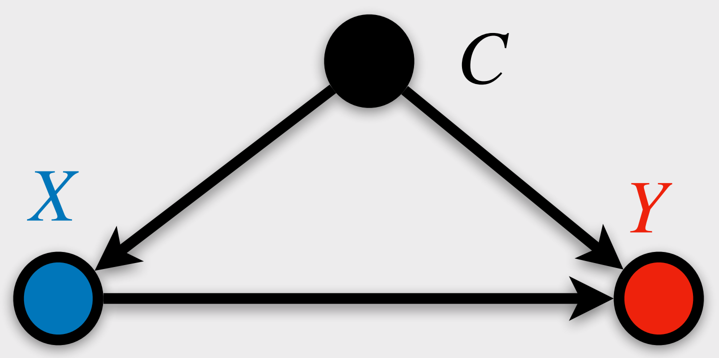
$$\mathbb{E}[\ Y_{x, M_{\neg x}}\ ]$$

Challenges in Standard Setting

Effect (Q)

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

Graph



"Back-Door"

Samples

D from P

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

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

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

expected — salary — a man would earn — given the opportunity — other genders had received

Estimating Identifiable Causal Effects

Tasks	Challenges
	1 Complicated dependences
	2 Data fusion (observations + experiments)
	3 More general scenarios

Estimating Identifiable Causal Effects

Tasks

1 **[Ch. 3]** Estimating causal effects from observations → ✔

Challenges

2 Data fusion (observations + experiments)

3 More general scenarios

Estimating Identifiable Causal Effects

Tasks		Challenges
1 **[Ch. 3]** Estimating causal effects from observations	→	✔
2 **[Ch. 4]** Estimating causal effects from data fusion	→	✔
		3 More general scenarios

Estimating Identifiable Causal Effects

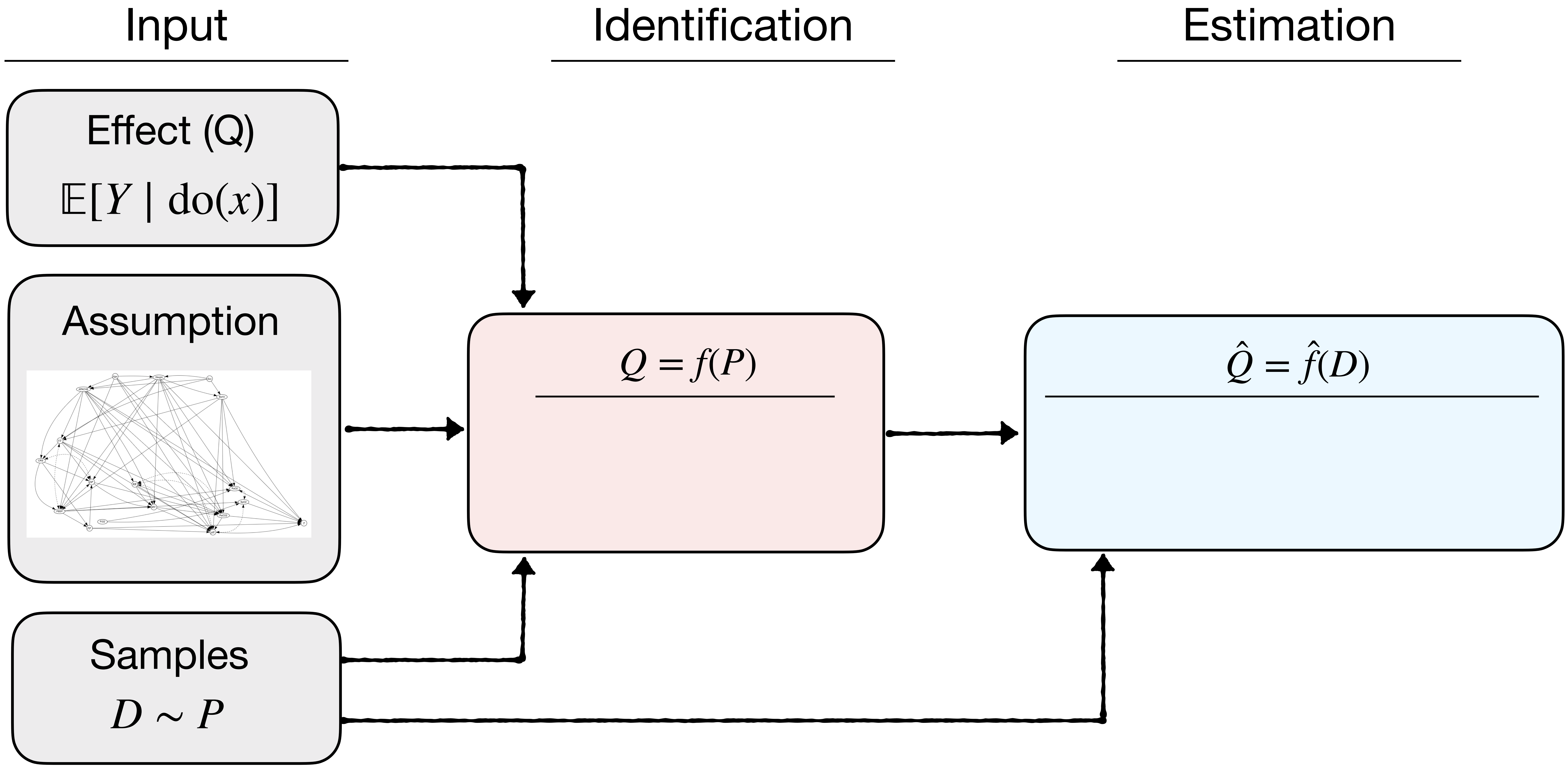
Tasks		Challenges
1 **[Ch. 3]** Estimating causal effects from observations	→	✔
2 **[Ch. 4]** Estimating causal effects from data fusion	→	✔
3 **[Ch. 5]** Unified causal effect estimation method	→	✔

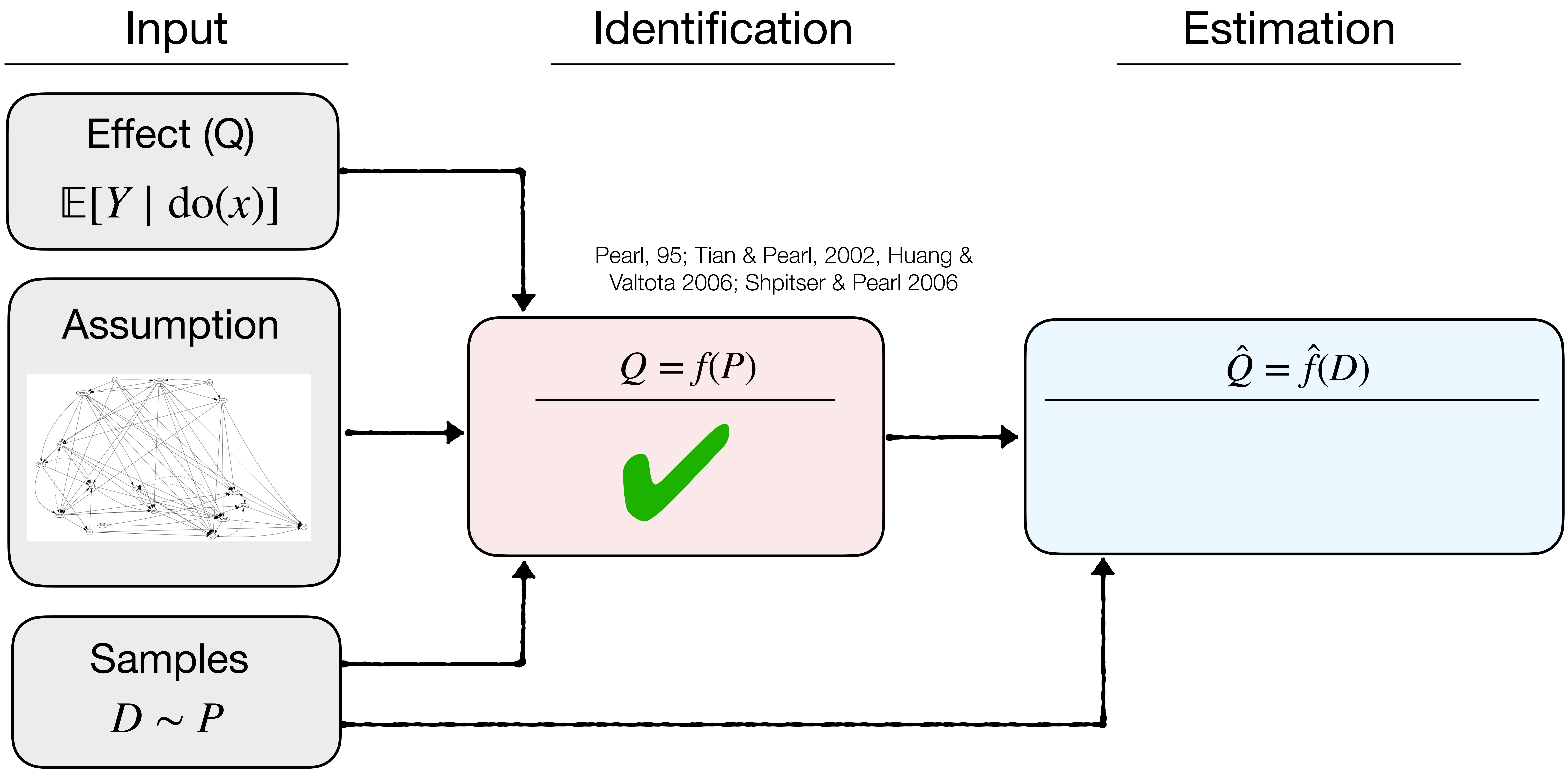
Talk Outline

1. **Ch. 3** Estimating causal effects from observations
2. **Ch. 4** Estimating causal effects from data fusion
3. **Ch. 5** Unified causal effect estimation method
4. Conclusion

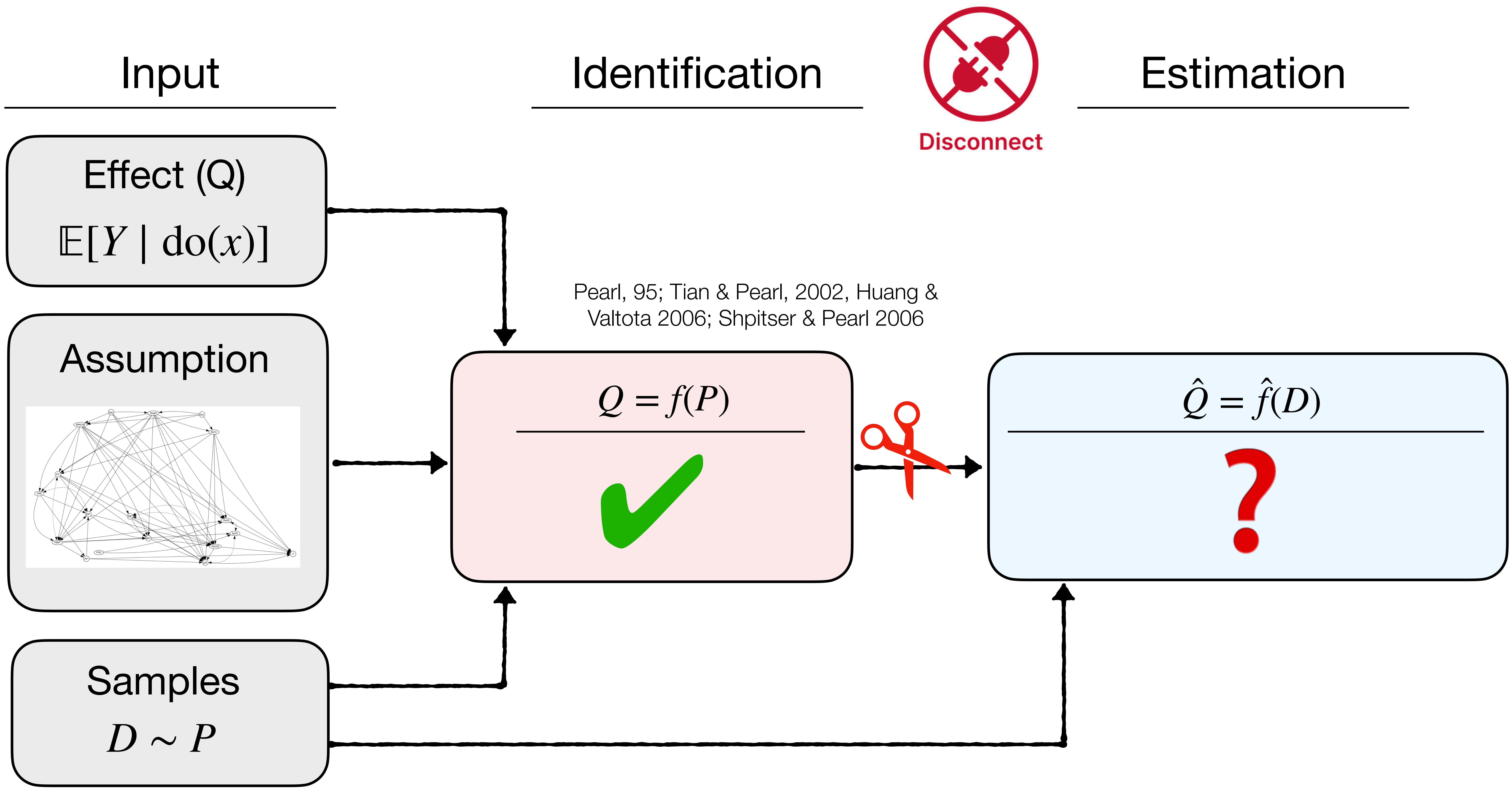
Talk Outline

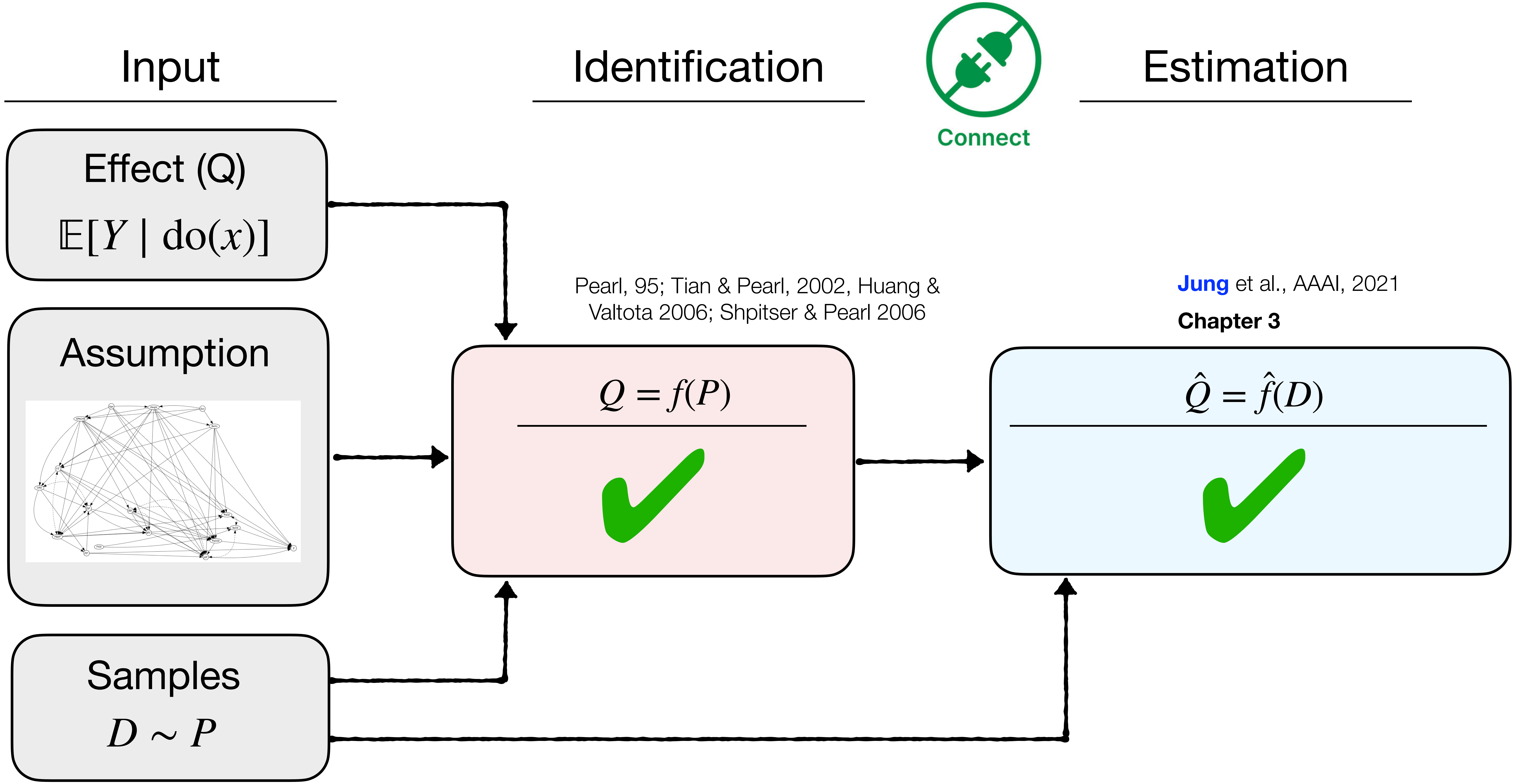
1 **Ch. 3** Estimating causal effects from observations


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



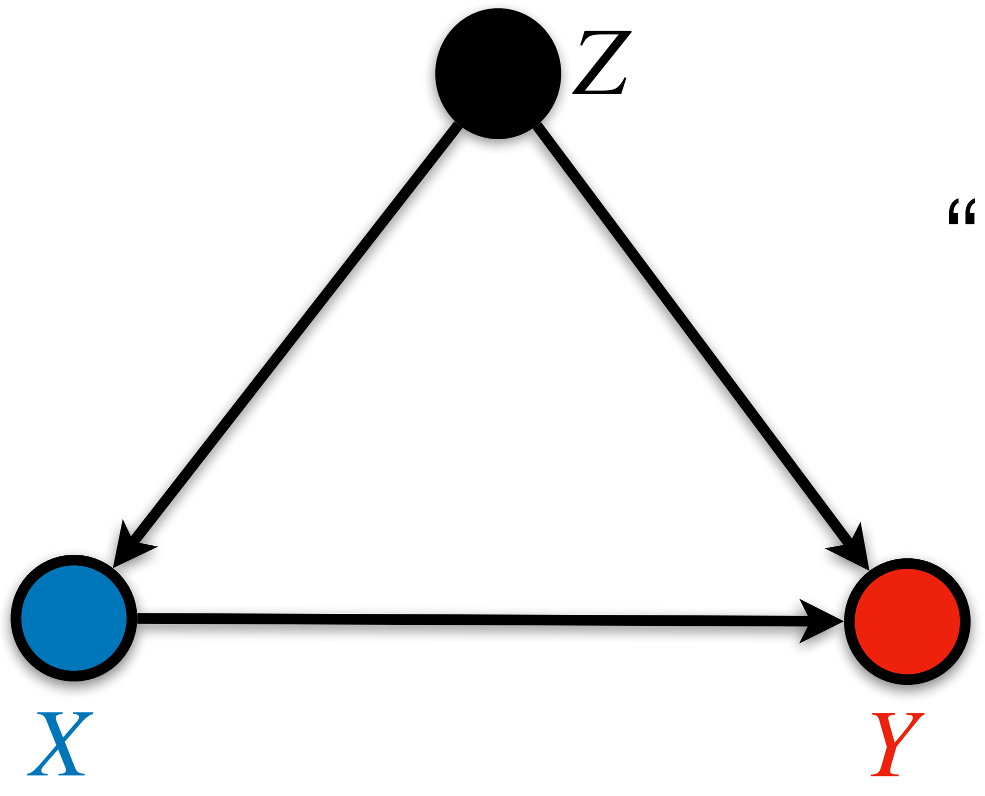
Input
Identification
Estimation
Effect (Q)
$\mathbb{E}[Y \mid \mathrm{do}(x)]$
Assumption
Samples
$D \sim P$
Pearl, 95; Tian & Pearl, 2002, Huang & Valtota 2006; Shpitser & Pearl 2006
$Q = f(P)$
$\hat{Q} = \hat{f}(D)$


Input
Identification
Disconnect
Estimation
Effect (Q)
$\mathbb{E}[Y \mid \mathrm{do}(x)]$
Assumption
Samples
$D \sim P$
Pearl, 95; Tian & Pearl, 2002, Huang & Valtota 2006; Shpitser & Pearl 2006
$Q = f(P)$
$\hat{Q} = \hat{f}(D)$
?


Input
Identification
Connect
Estimation
Effect (Q)
$\mathbb{E}[Y \mid \mathrm{do}(x)]$
Assumption
Samples
$D \sim P$
Pearl, 95; Tian & Pearl, 2002, Huang & Valtota 2006; Shpitser & Pearl 2006
$Q = f(P)$
Jung et al., AAAI, 2021
Chapter 3
$\hat{Q} = \hat{f}(D)$

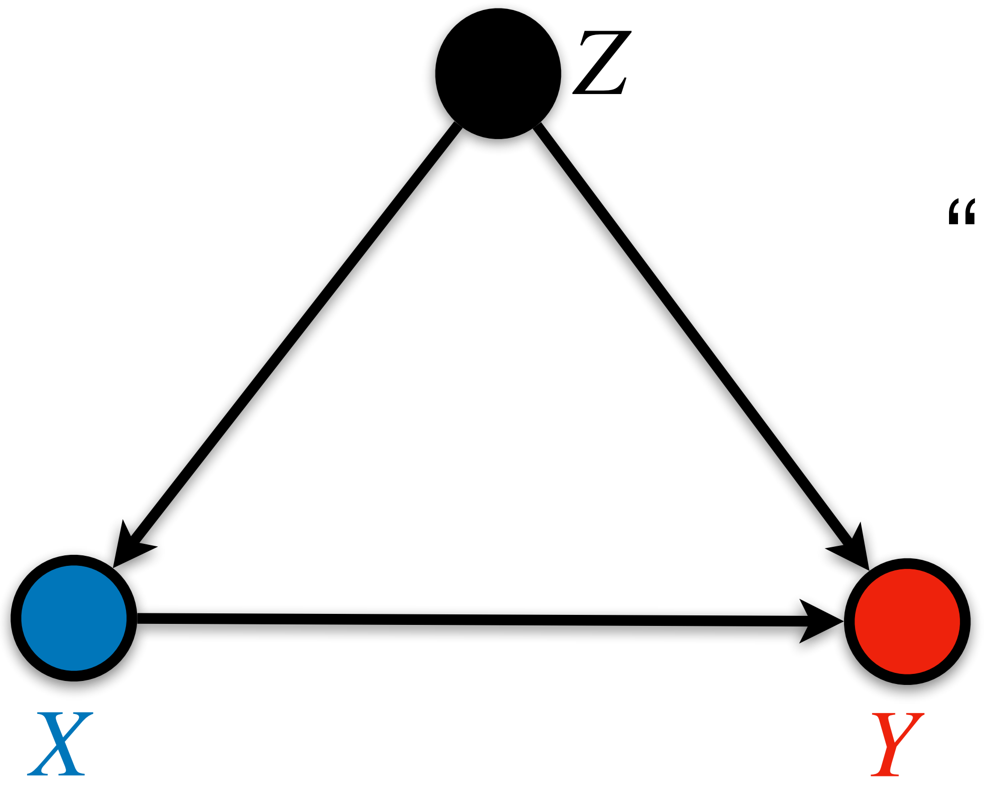
Background: Back-door Adjustment (BD)

Background: Back-door Adjustment (BD)



“Back-door graph”

Background: Back-door Adjustment (BD)



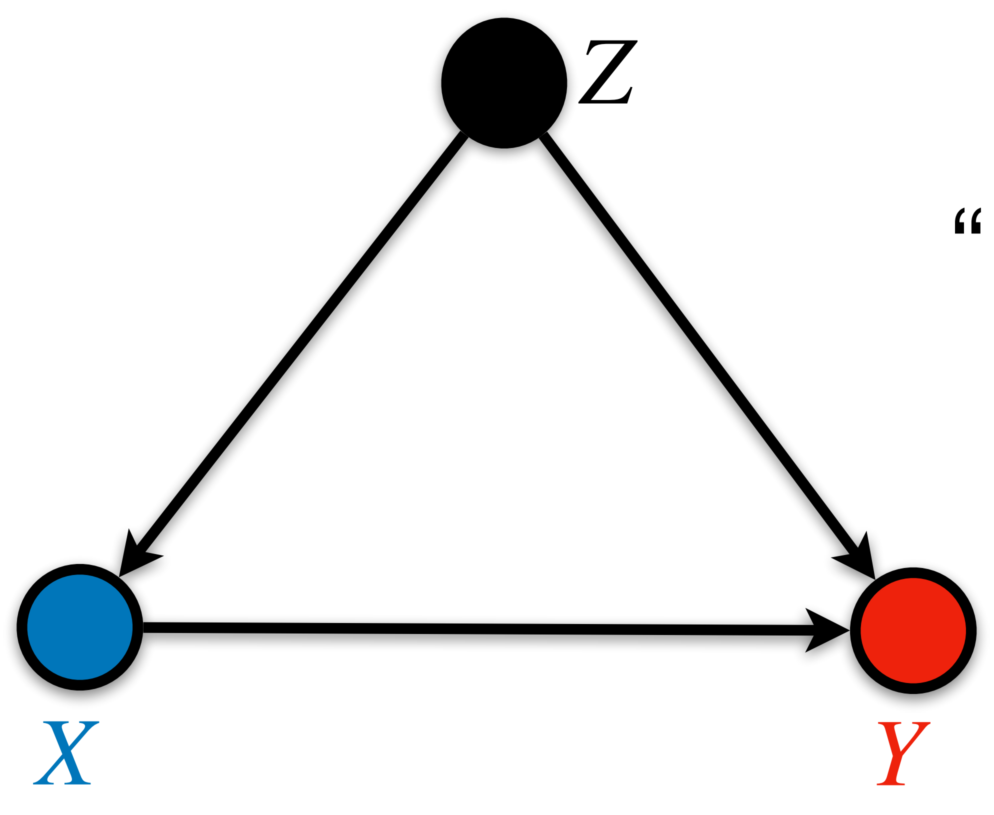
"Back-door graph"

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)

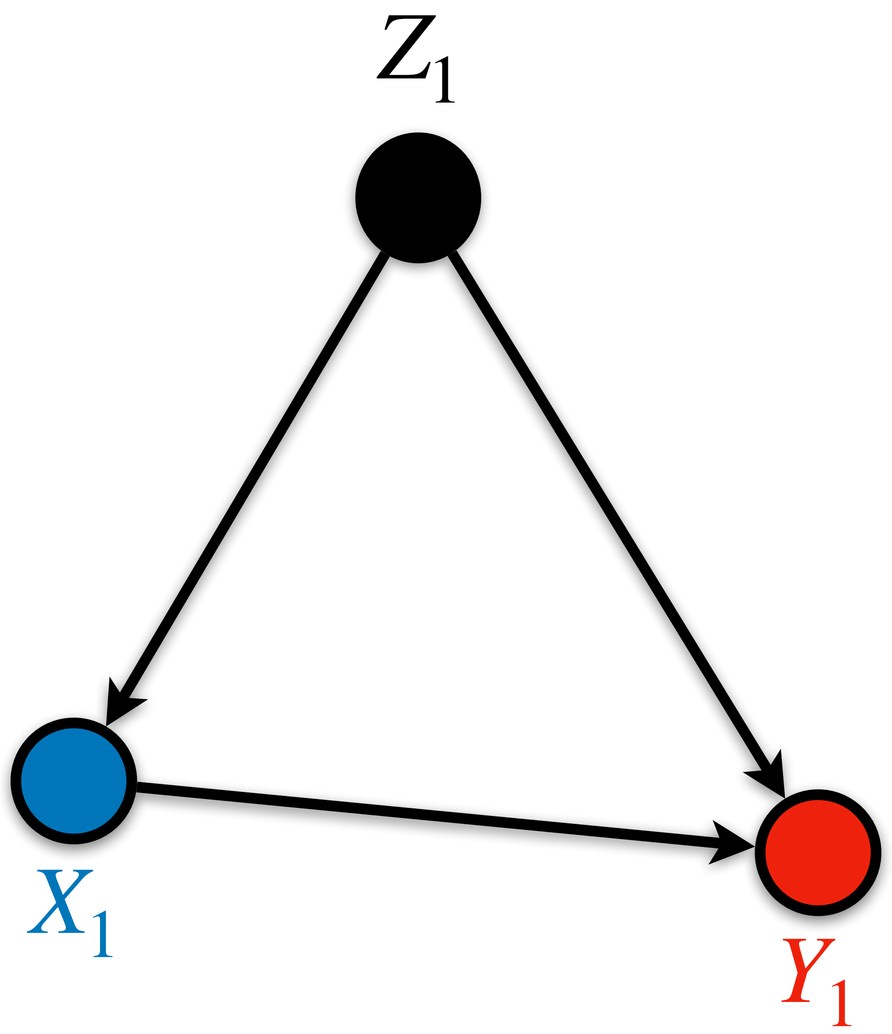
1. $\mathbf{Z}$ is not a descendent of treatment;
2. $\mathbf{Z}$ blocks spurious paths between (treatments, outcome)

“Back-door adjustment (BD)”

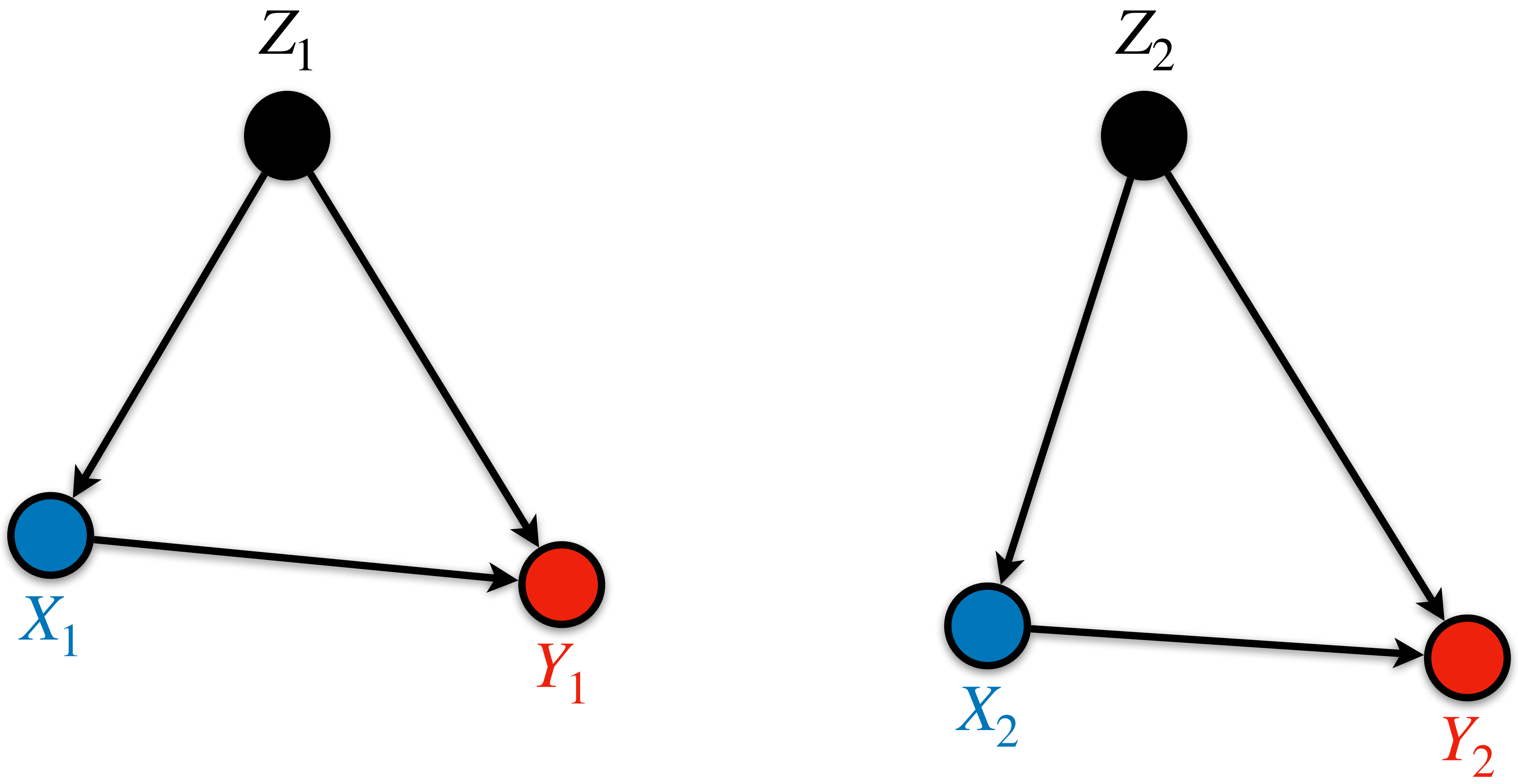
$$P(y \mid do(x)] = \text{BD} \triangleq \sum_z P(y \mid x, z)P(z)$$

Background: Multi-outcome sequential BD

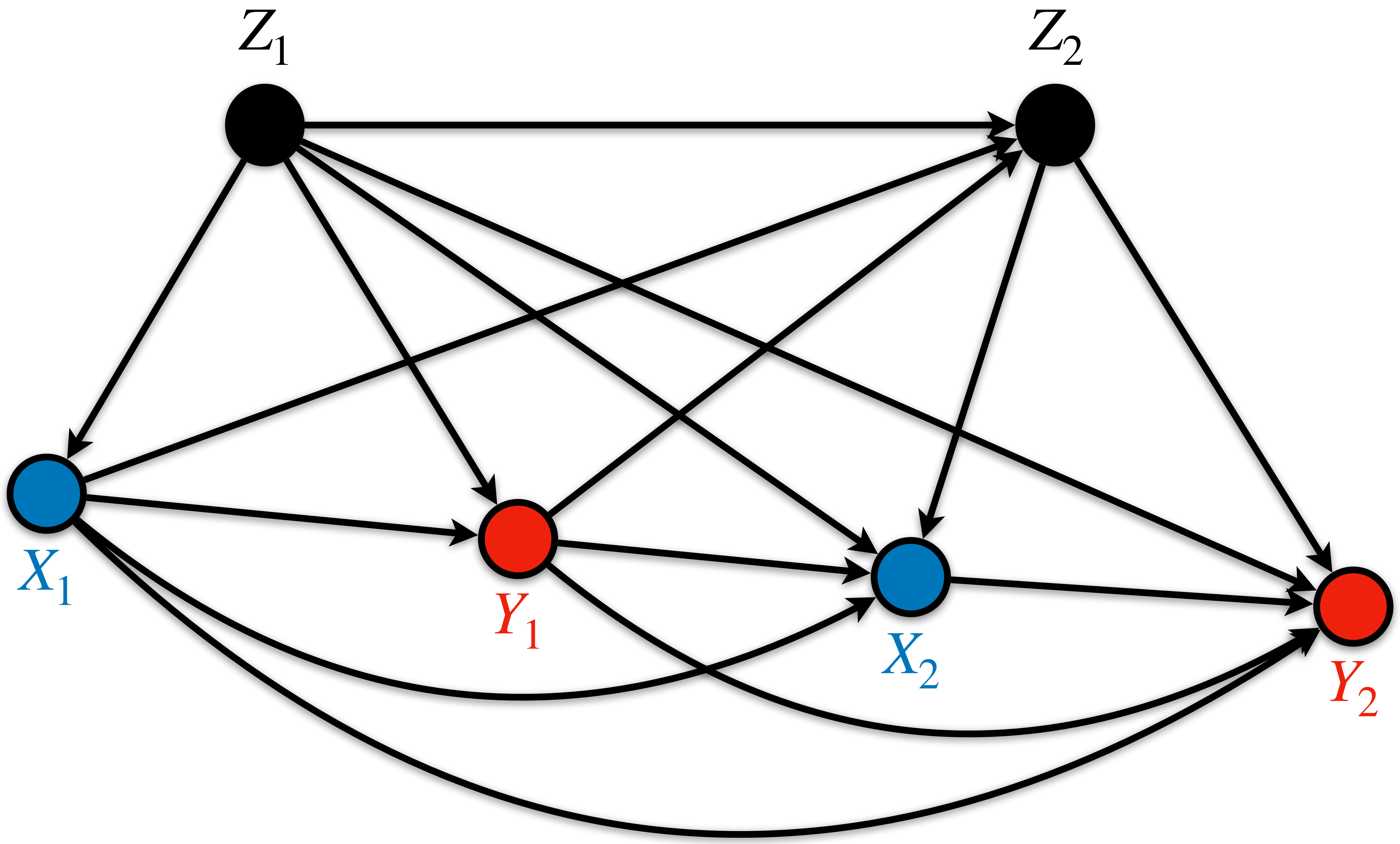
Background: Multi-outcome sequential BD



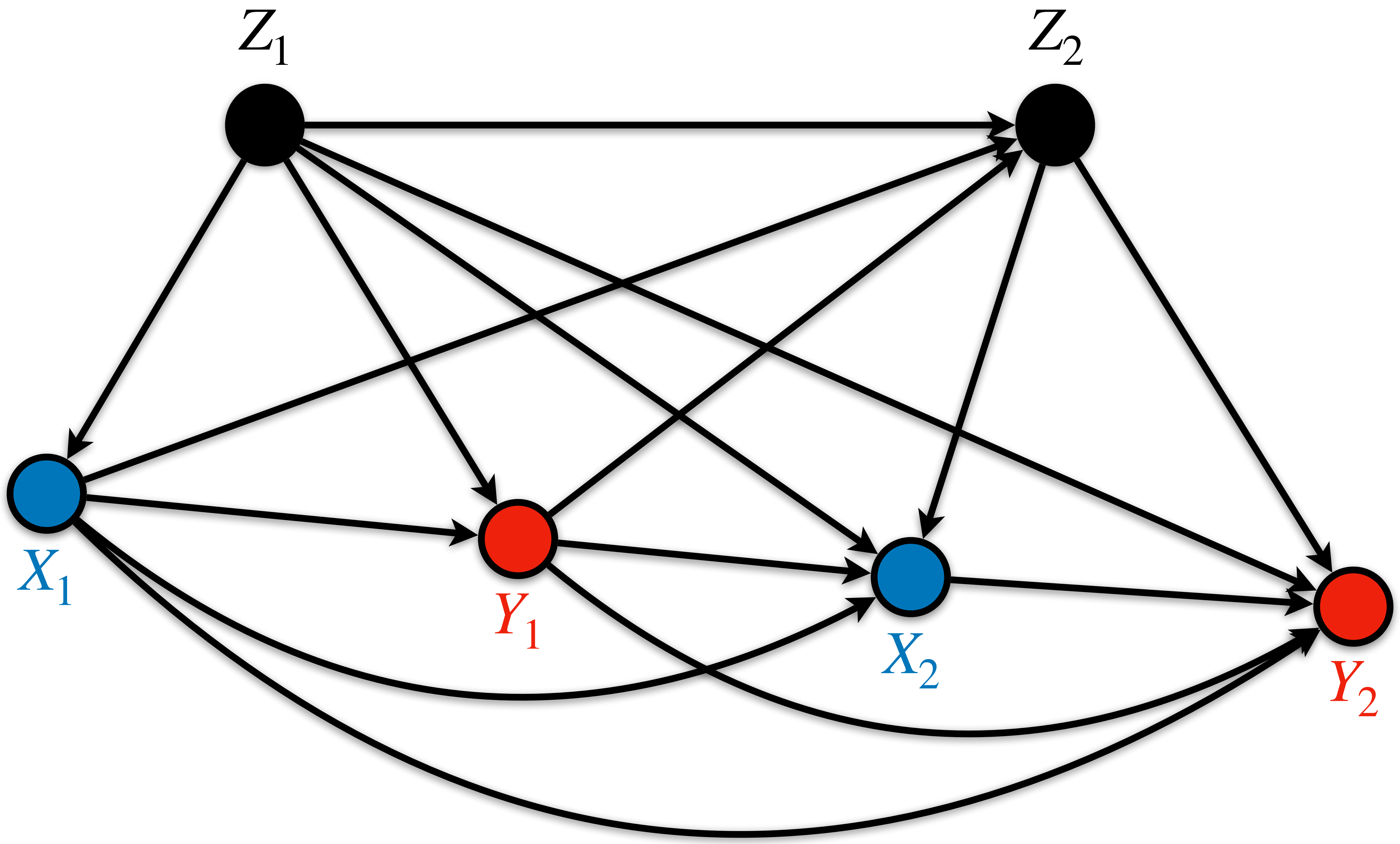
Background: Multi-outcome sequential BD



Background: Multi-outcome sequential BD



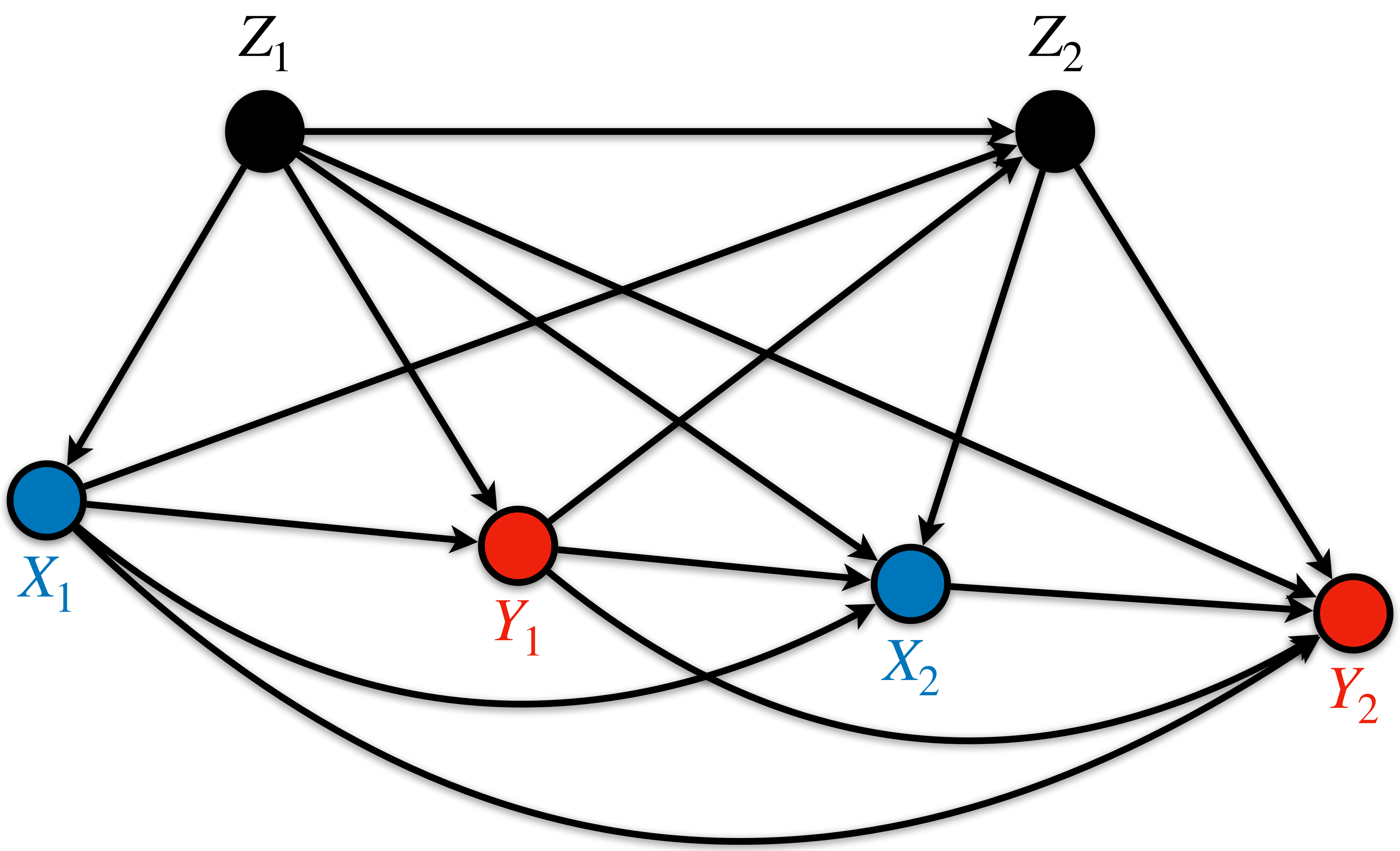
Background: Multi-outcome sequential BD



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



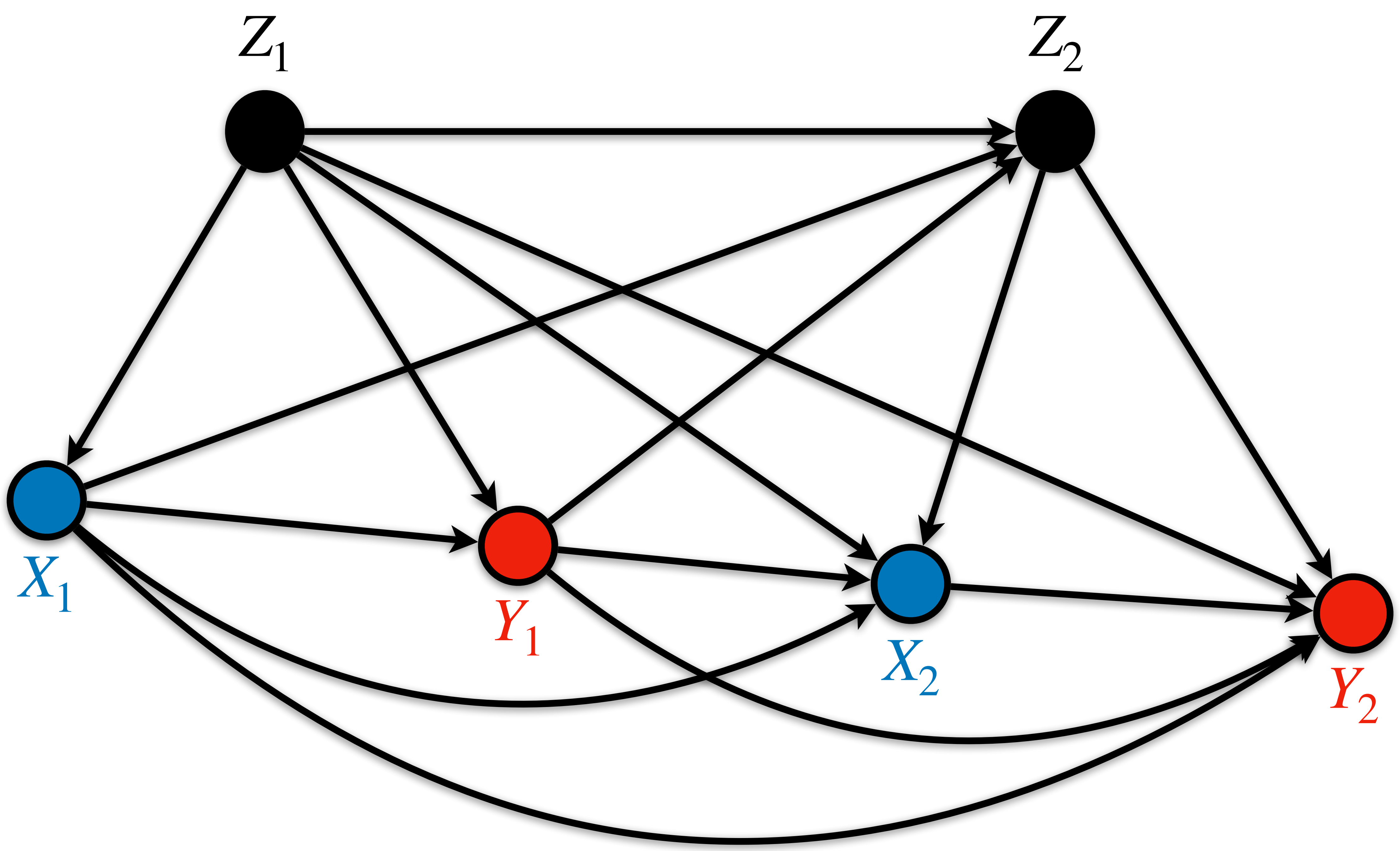
Multi-outcome Sequential BD (mSBD)

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.

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

"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.

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

"mSBD adjustment"

* I'll use "BD" for simplicity, but all results extend to mSBD (as shown in the thesis).

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:

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:

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

Background: Robust Estimator for BD

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

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

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

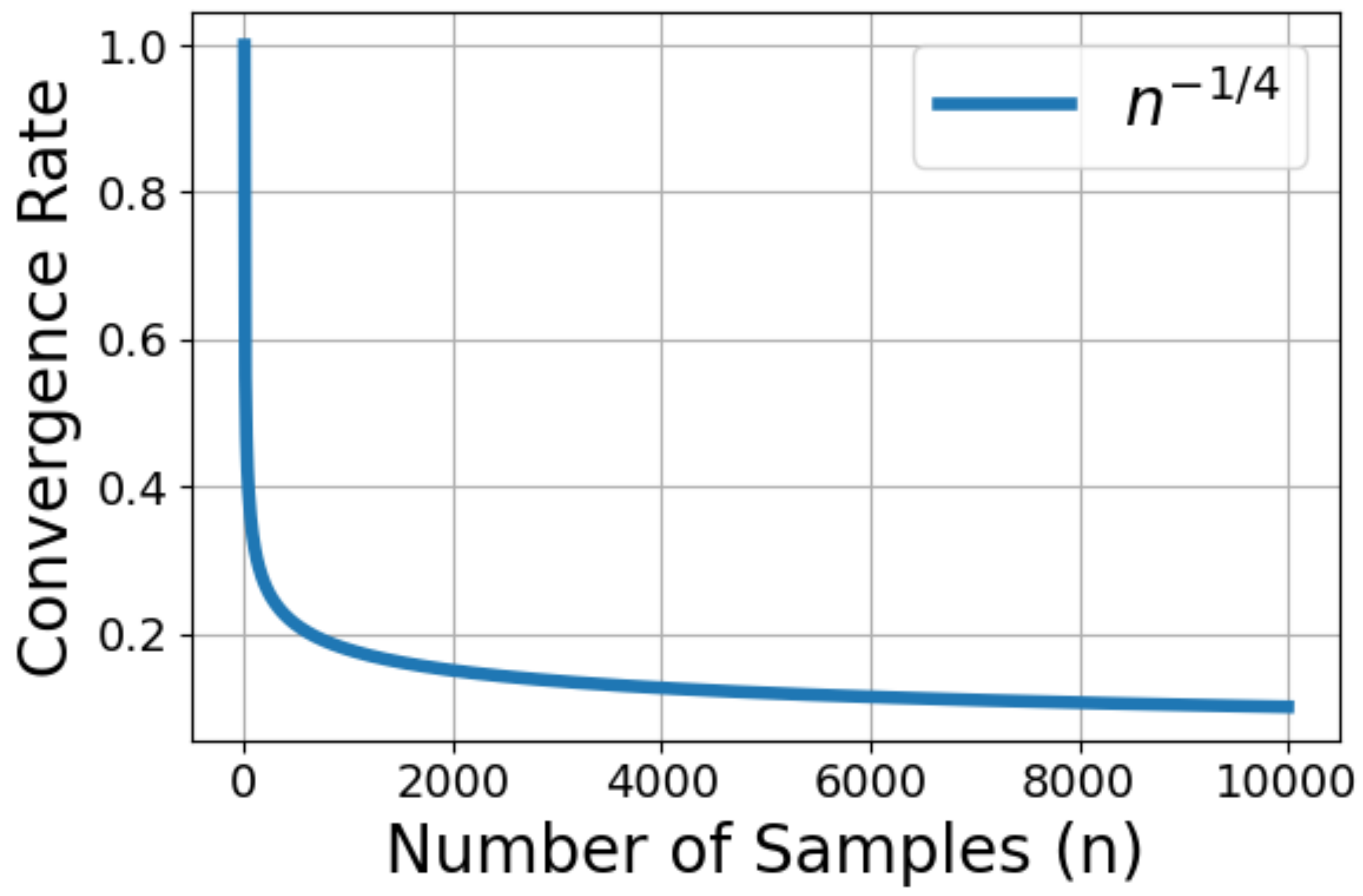
Background: Robust Estimator for BD

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

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

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

Background: Robust Estimator for BD

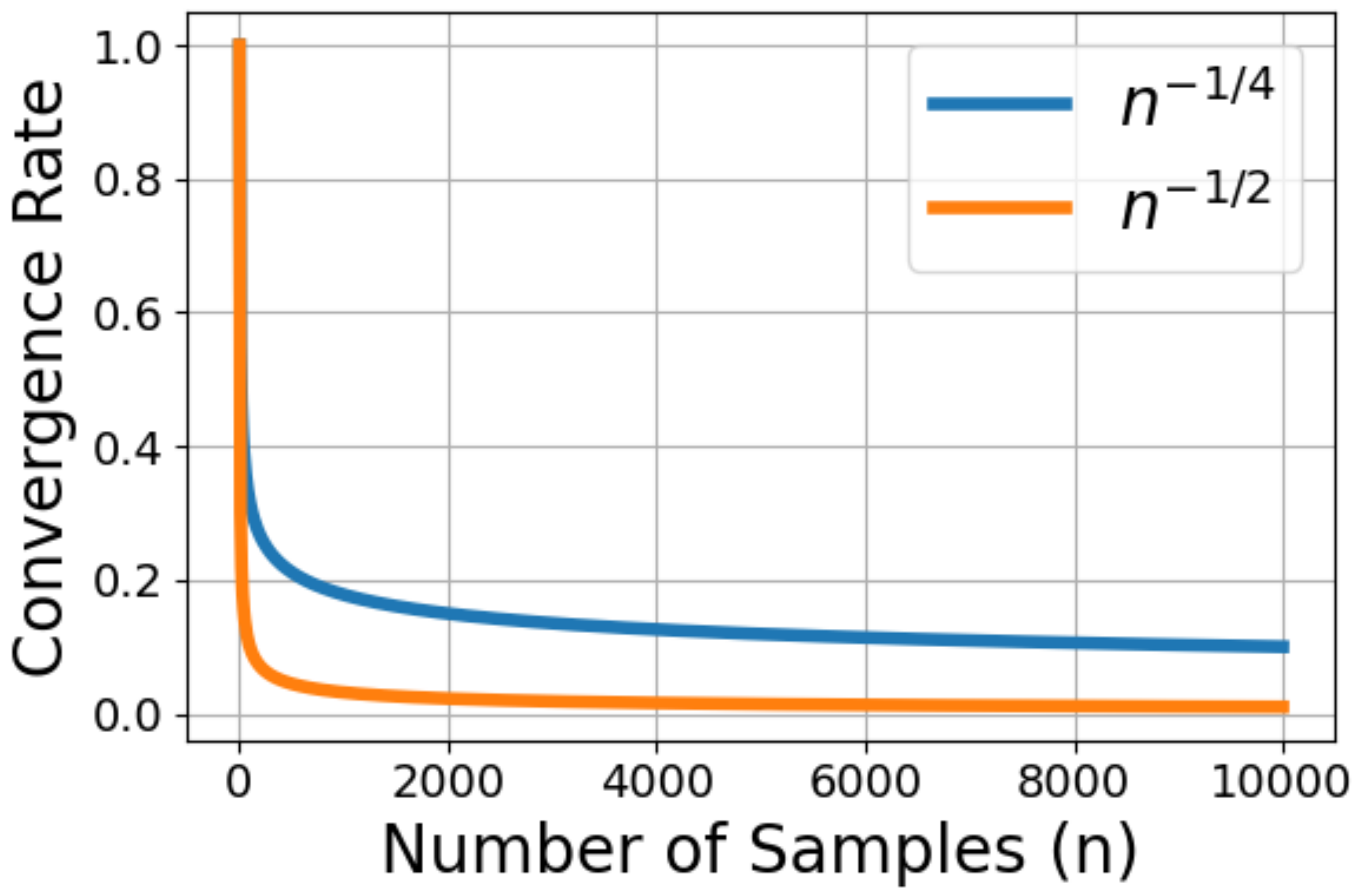


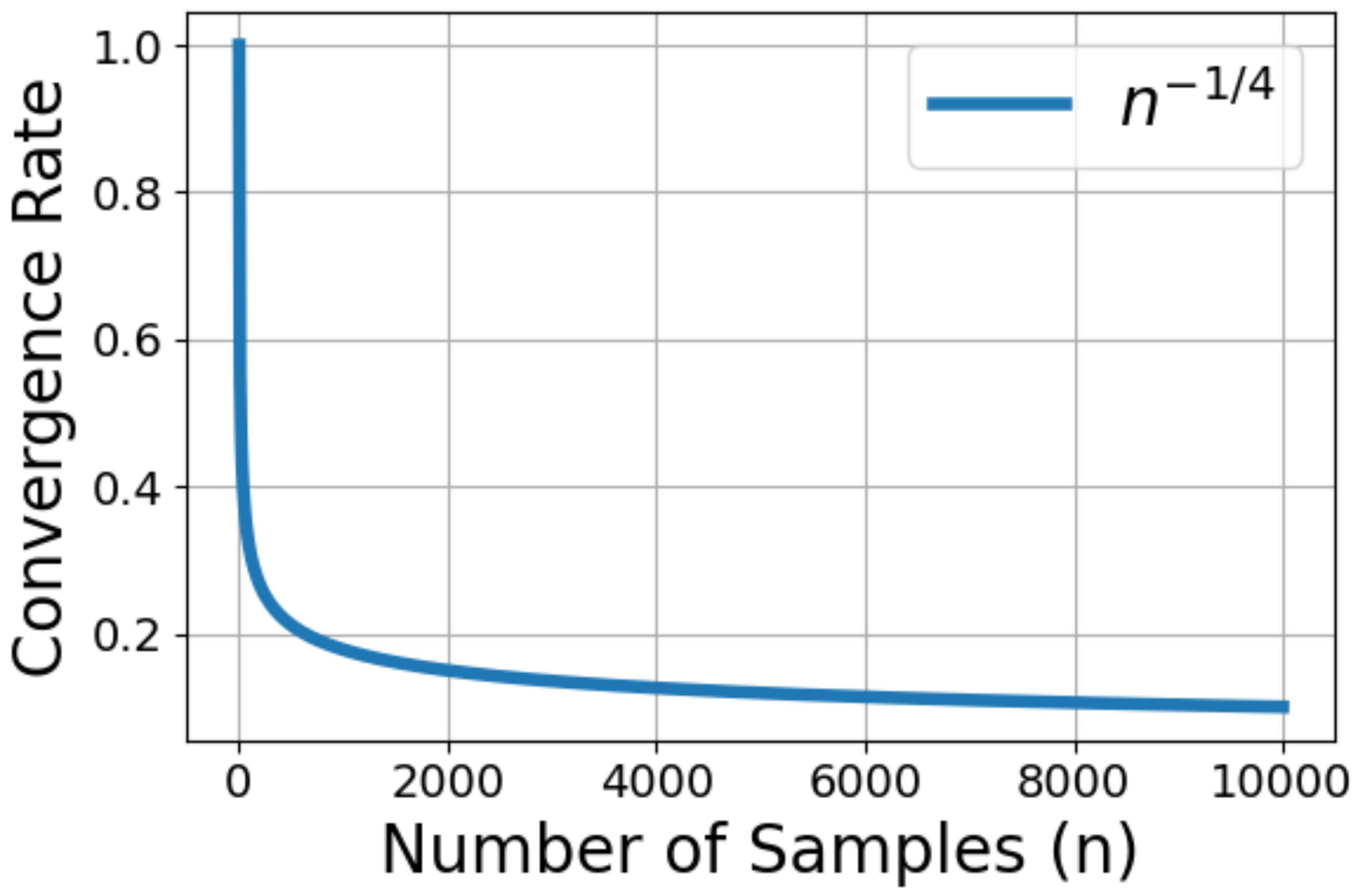
$$\text{Error}(\text{DML-BD}(\hat{\mu},\hat{\pi}), \text{BD}(\mu,\pi)) = \text{Error}(\hat{\mu},\mu) \times \text{Error}(\hat{\pi},\pi)$$

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

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

Background: Robust Estimator for BD





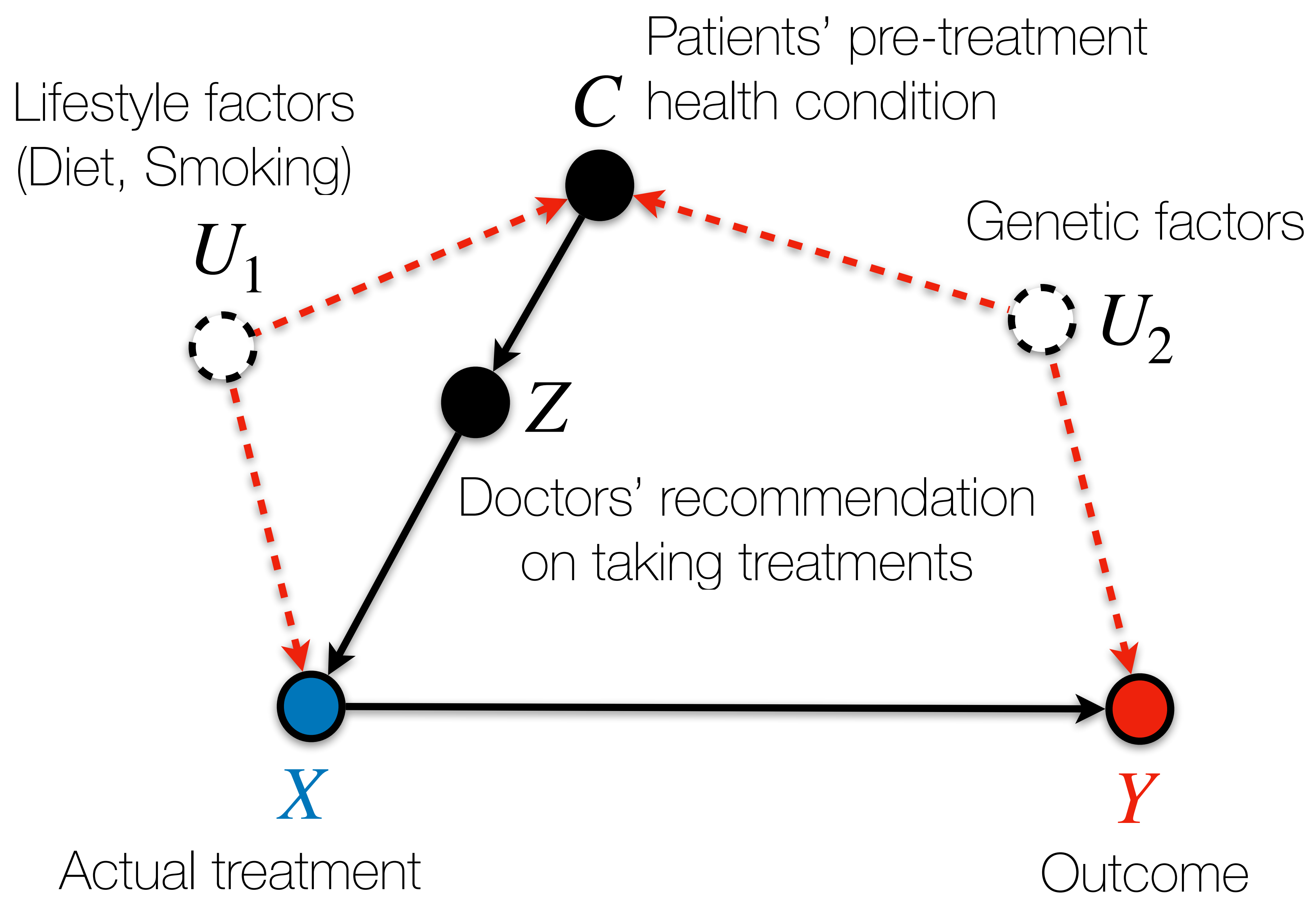
$$\text{Error}(\text{DML-BD}(\hat{\mu},\hat{\pi}),\ \text{BD}(\mu,\pi)) = \text{Error}(\hat{\mu},\mu) \times \text{Error}(\hat{\pi},\pi)$$

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

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

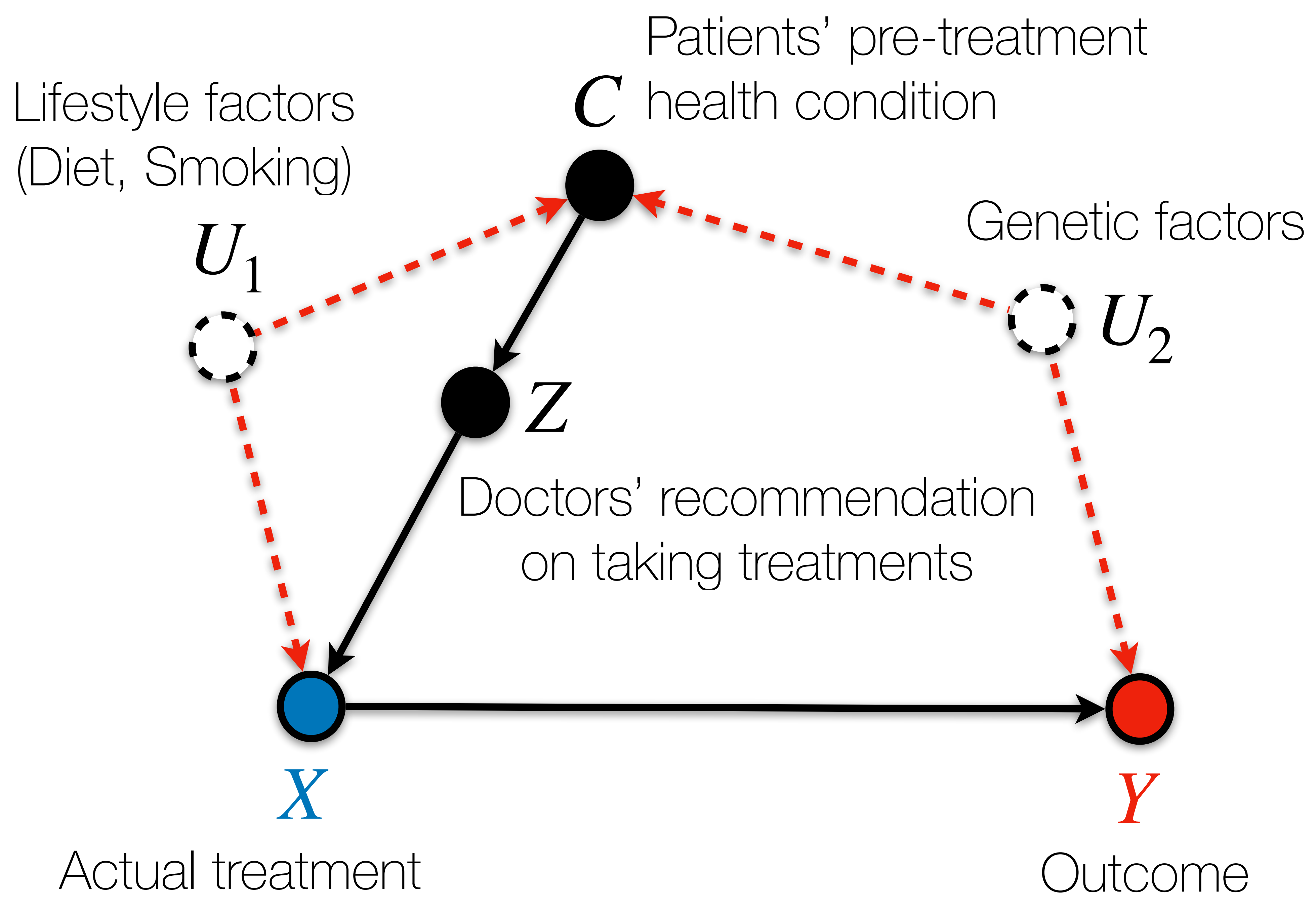
Non-BD Example: “Napkin Graph”

Non-BD Example: "Napkin Graph"



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

Non-BD Example: “Napkin Graph”

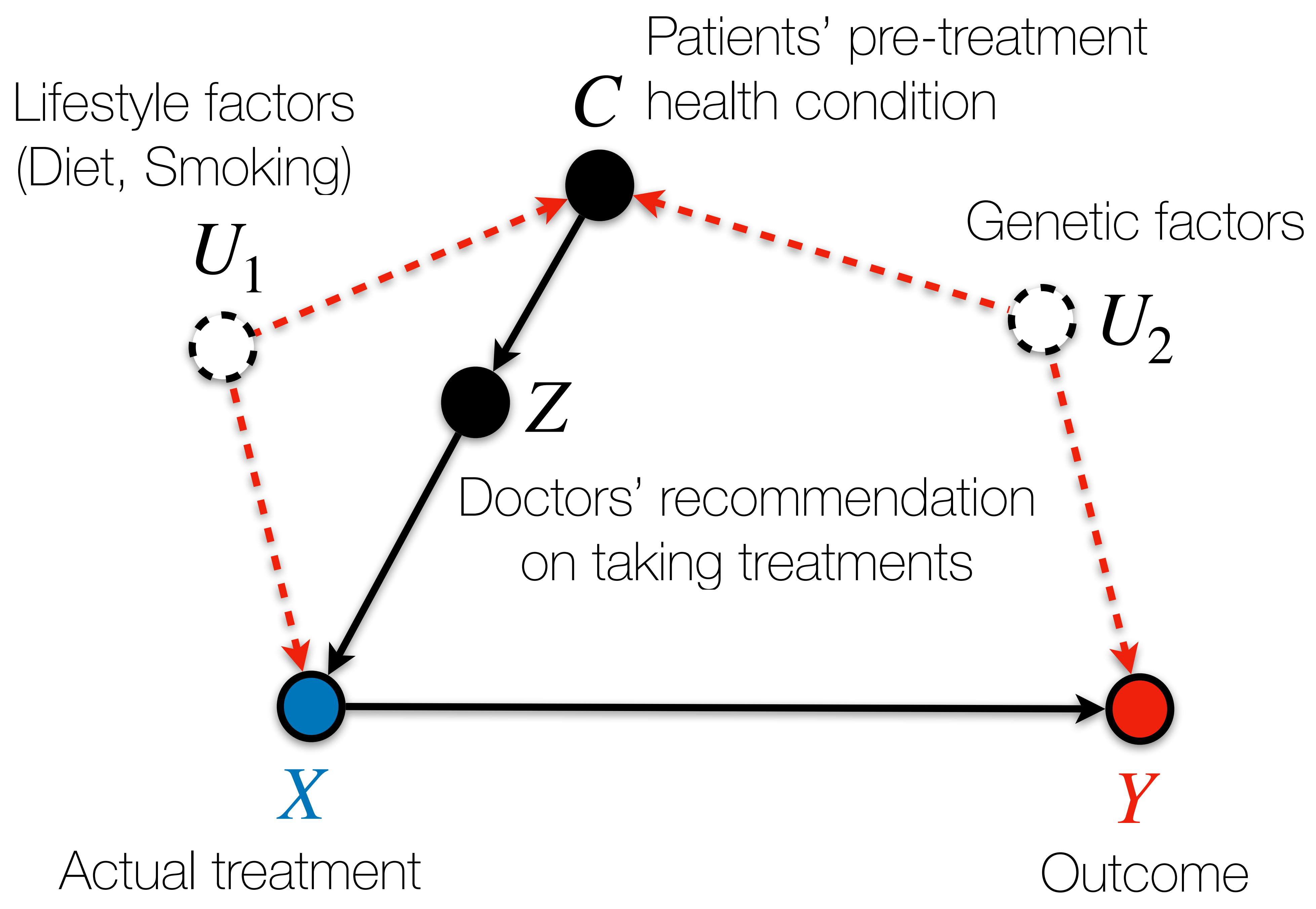


$\mathbb{E}[Y \mid do(x)]$

Identification

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

Non-BD Example: "Napkin Graph"



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

Identification

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

Estimation

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

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	✔	?

Idea for connecting BD and Identification

Idea for connecting BD and Identification

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

Idea for connecting BD and Identification

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

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

Idea for connecting BD and Identification

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

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

by strategically **combining robust BD estimators.**

Background: Causal Effect Identification

Background: Causal Effect Identification

Identification (Algo 1)

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

Background: Causal Effect Identification

Identification (Algo 1)

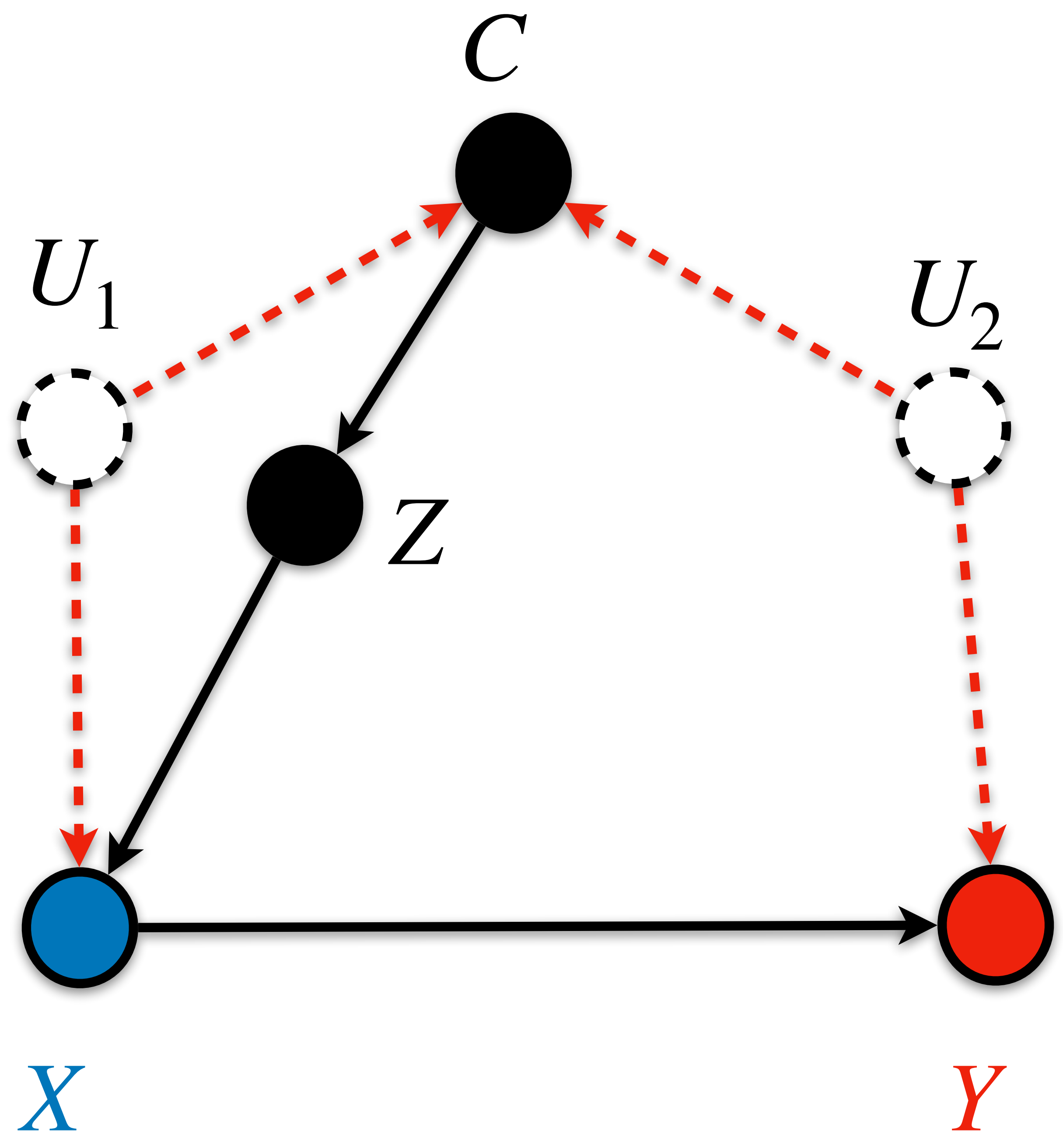
- spanning a *tree* from $P(\mathbf{V})$
- to reach to causal distribution $P(Y \mid \mathrm{do}(X))$
- through factorization & marginalization of distributions

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

Background: Causal Effect Identification

Identification (Algo 1)

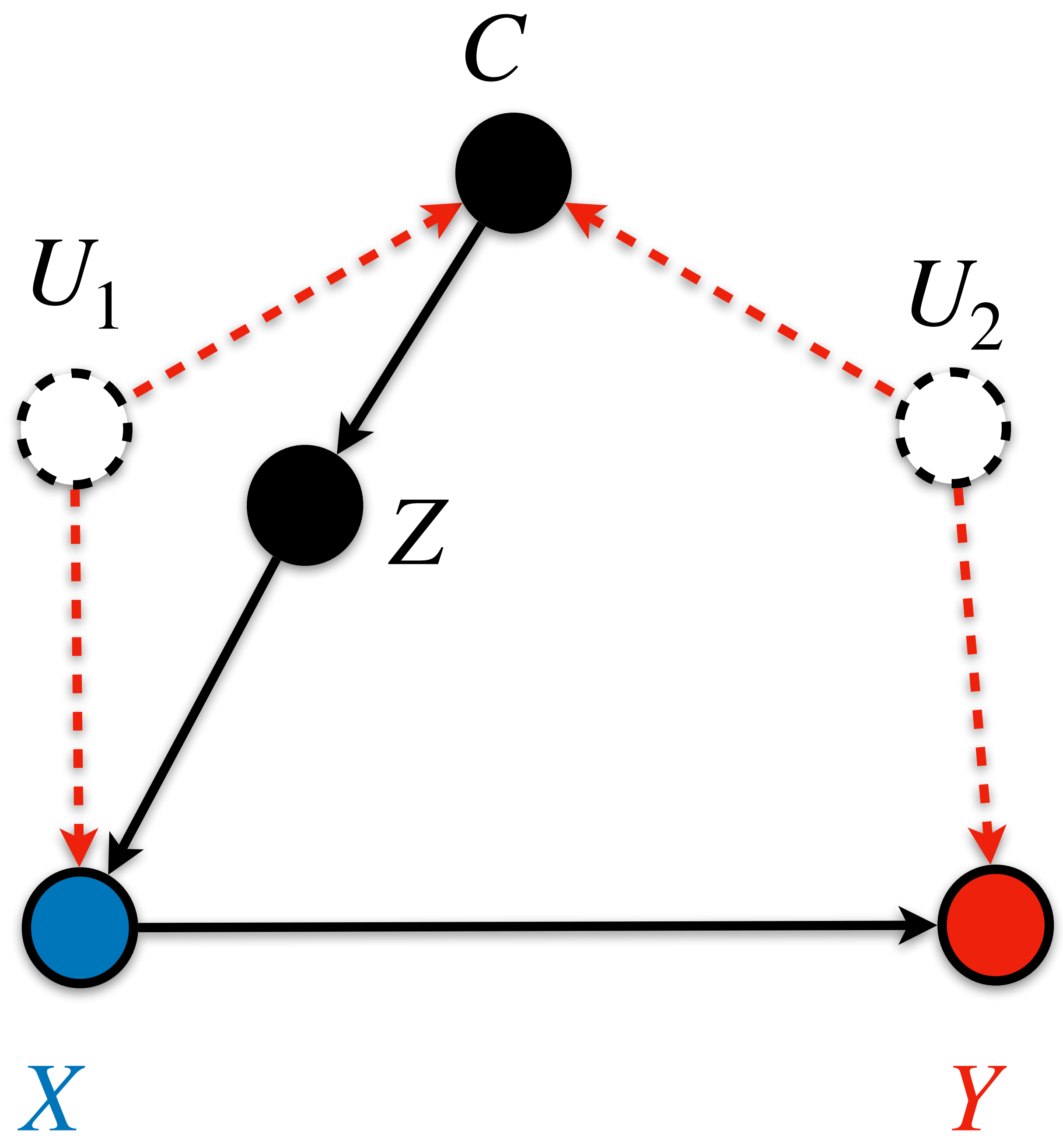
- spanning a *tree* from $P(\mathbf{V})$
- to reach to causal distribution $P(Y \mid \mathrm{do}(X))$
- through factorization & marginalization of distributions



Background: Causal Effect Identification

Identification (Algo 1)

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

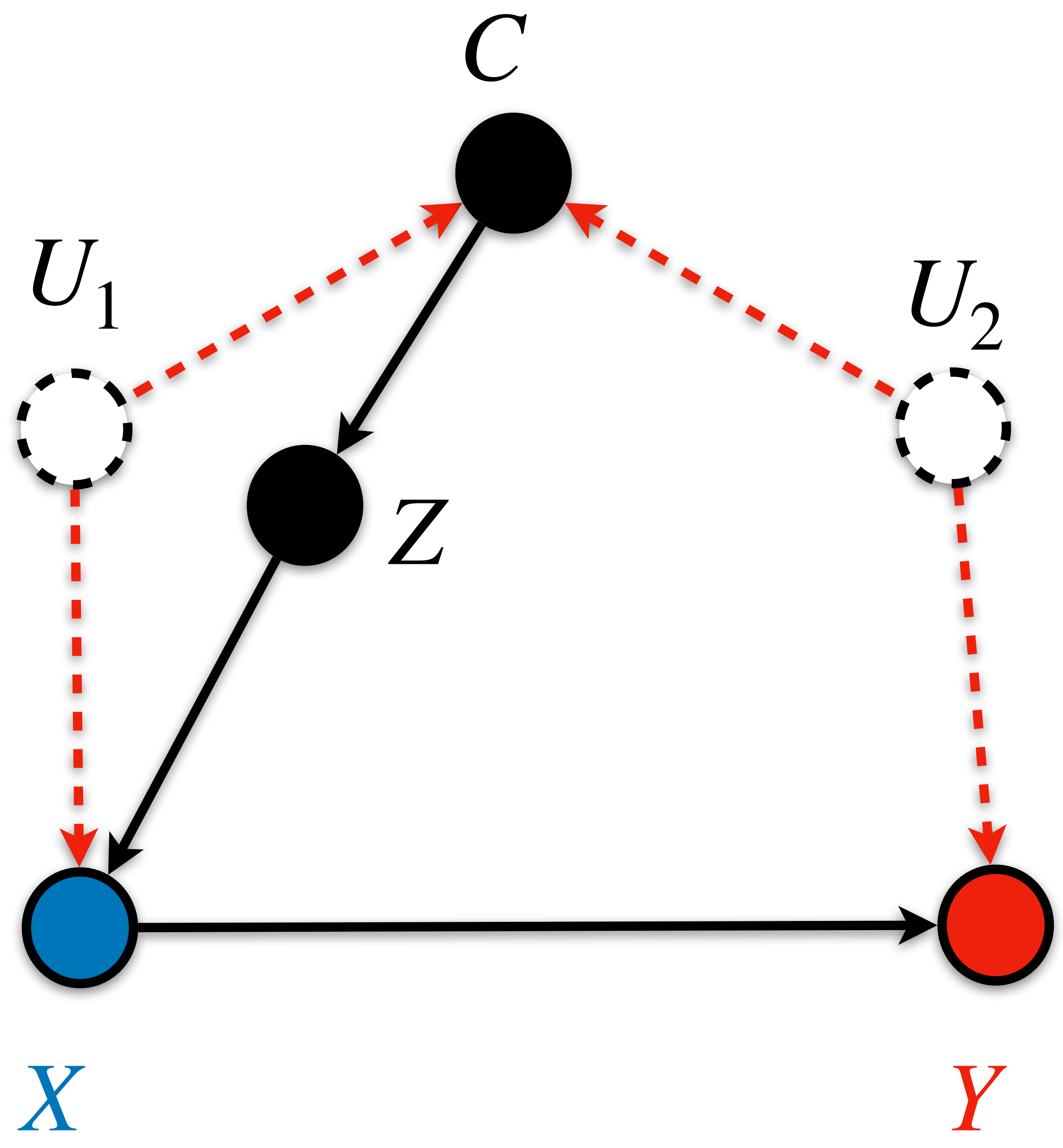


$P(CZXY)$

Background: Causal Effect Identification

Identification (Algo 1)

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



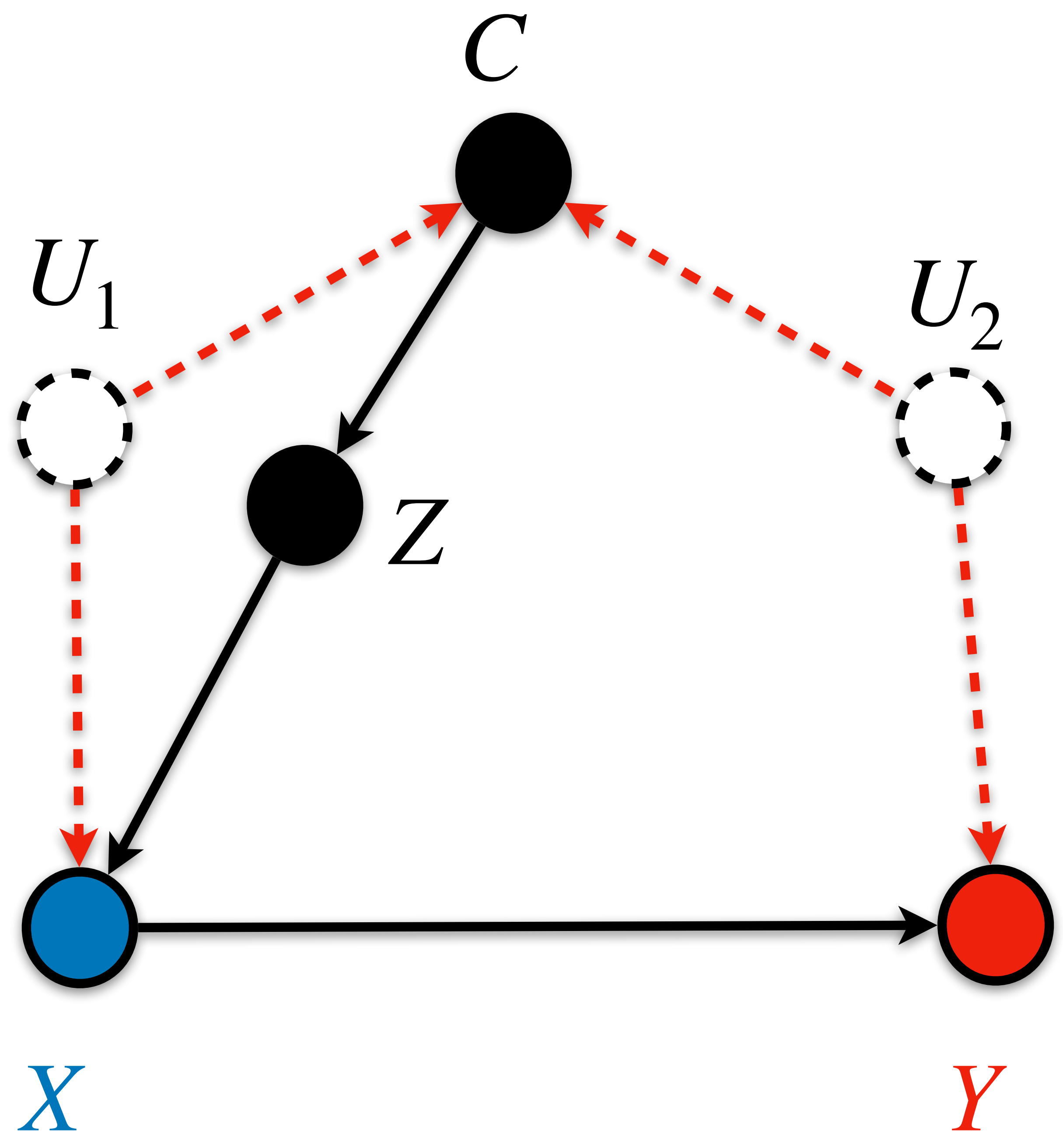
$$P(CZXY) \xrightarrow{\text{Factorization}} P_{\text{do}(Z)}(CXY)$$

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

Background: Causal Effect Identification

Identification (Algo 1)

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



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

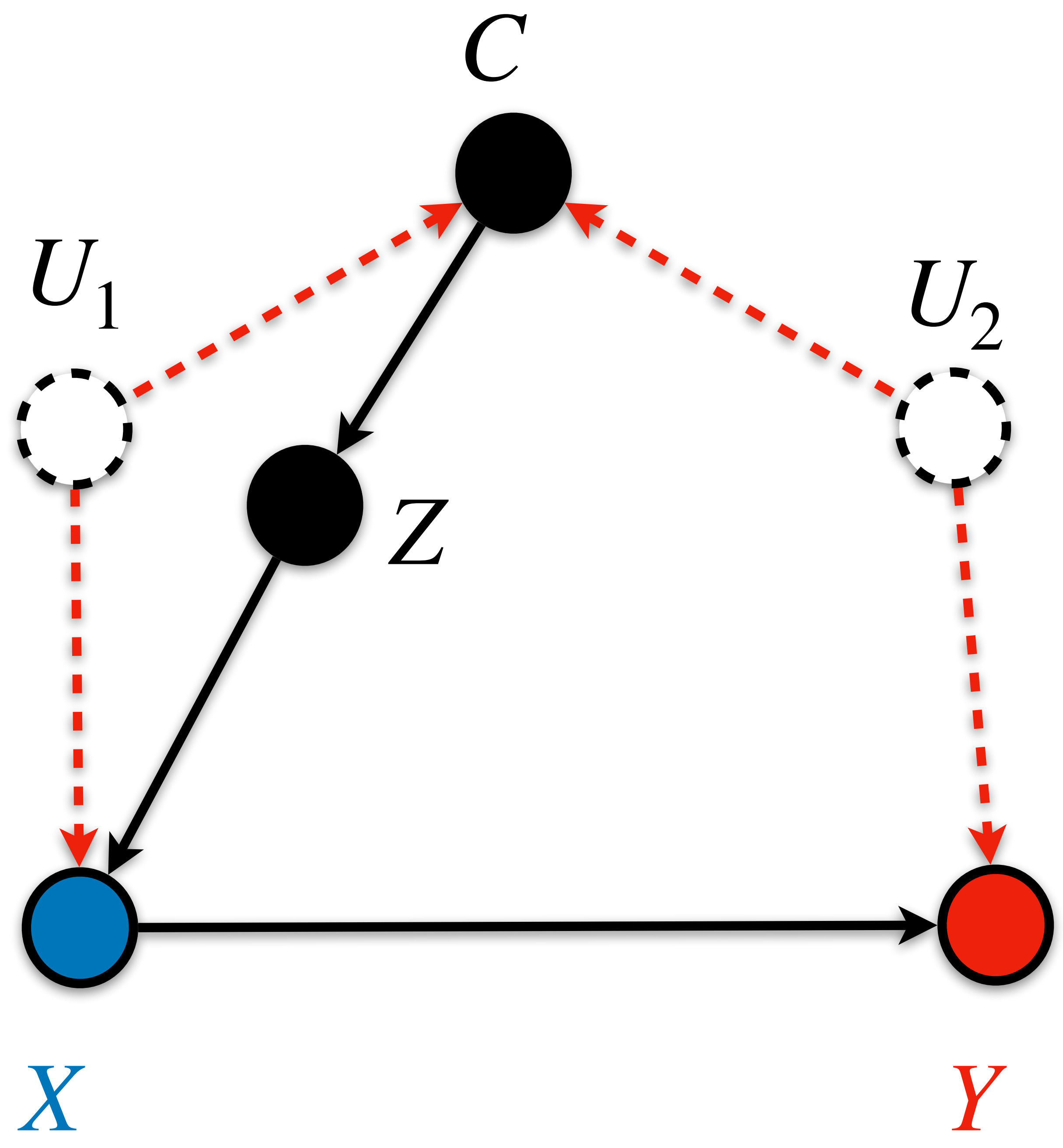
$P_{\mathrm{do}(Z)}(CXY)$: $P(C)P(XY \mid ZC)$

$P_{\mathrm{do}(Z)}(XY)$: $\sum_c P(c)P(XY \mid Zc)$

Background: Causal Effect Identification

Identification (Algo 1)

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



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

$P_{\text{do}(Z)}(CXY)$: $P(C)P(XY \mid ZC)$

$P_{\text{do}(Z)}(XY)$: $\sum_c P(c)P(XY \mid Zc)$

$P(Y \mid \text{do}(X))$: $P_{\text{do}(Z)}(Y \mid X) = \dfrac{\sum_c P(c)P(XY \mid Zc)}{\sum_c P(c)P(X \mid Zc)}$

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,

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

My Approach: 3-Step

To connect BD & Identification,

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

My Approach: 3-Step

To connect BD & Identification,

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

Complete Criterion for mSBD Adjustment

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

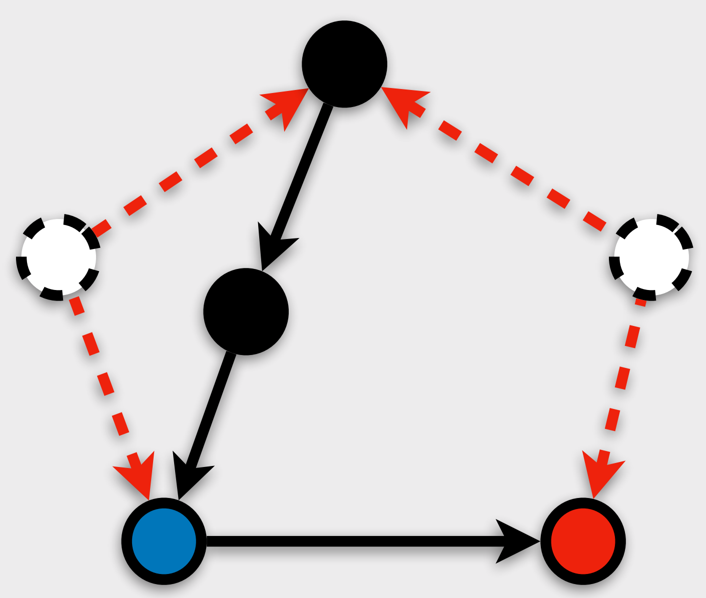
Complete Criterion for mSBD Adjustment

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

Complete Criterion for mSBD Adjustment

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

Graph



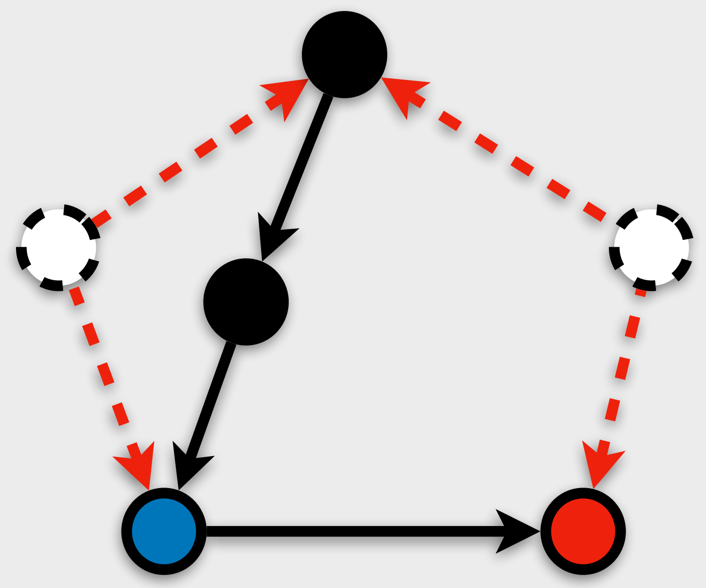
Interventional Distribution

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

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

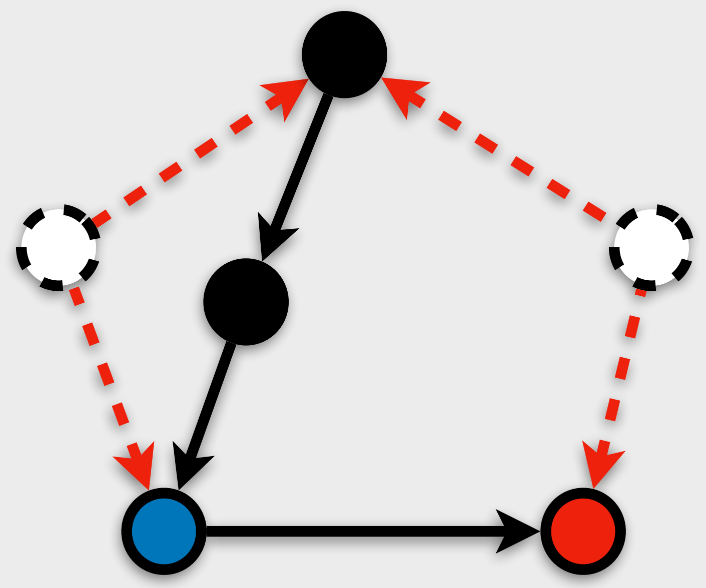
(Sec. 3.2.1)

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

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

Sequential Adjustment Criterion

(Sec. 3.2.1)

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

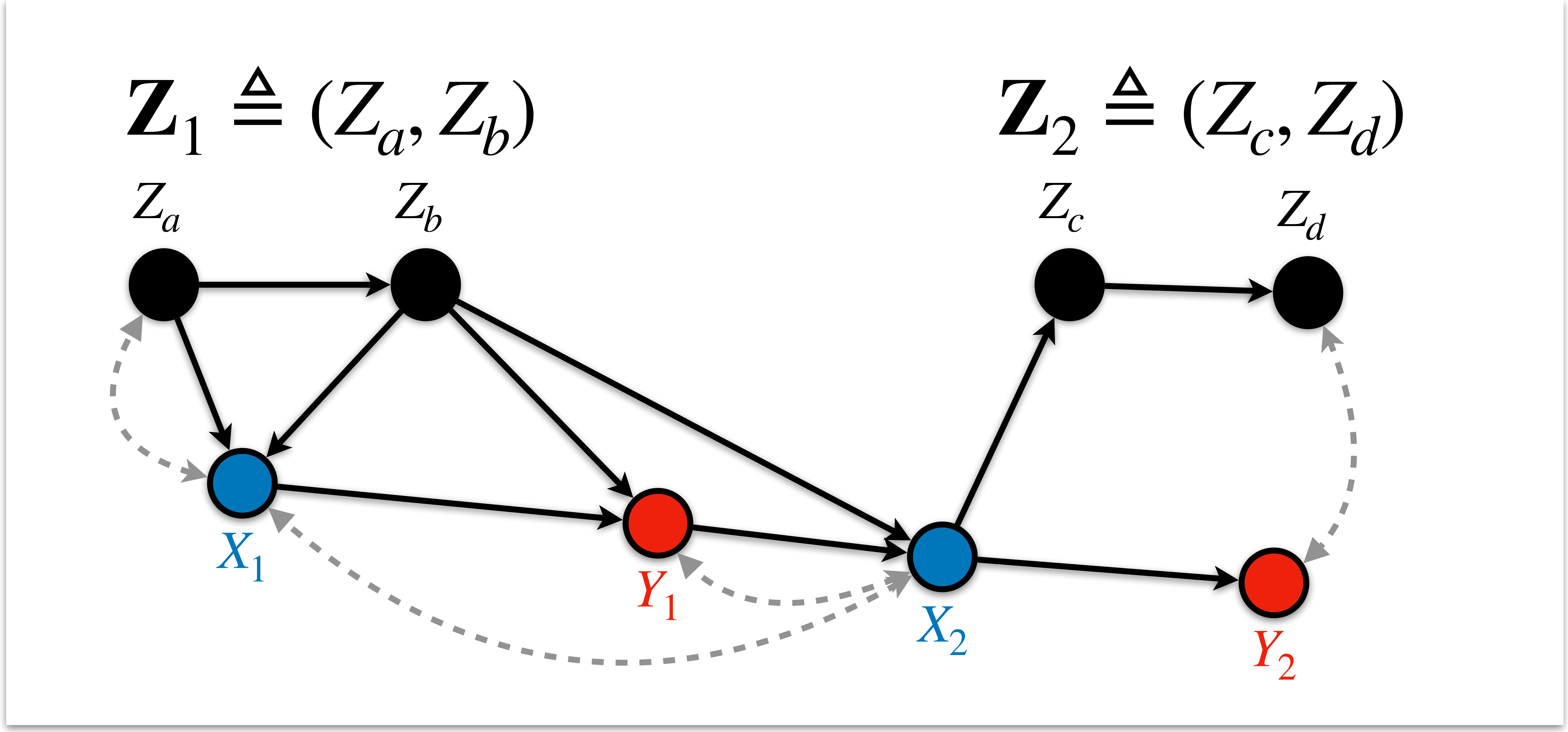
→ $P_{\mathrm{do}(a)}(B) = \mathrm{BD}(\mu, \pi)$ if BD-expressible

Motivation: Incompleteness of BD/mSBD (Sec 3.2)

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

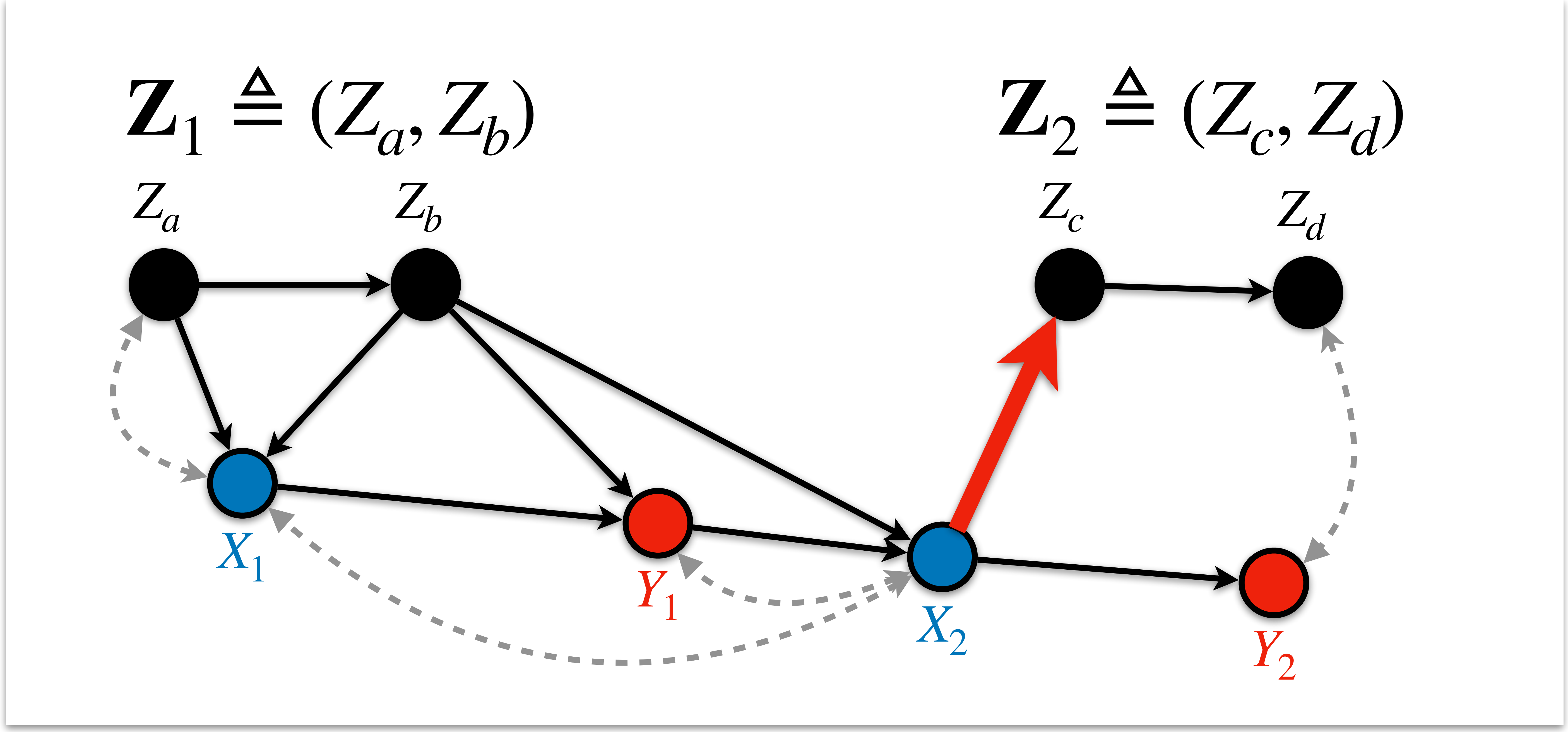
Motivation: Incompleteness of BD/mSBD (Sec 3.2)

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



Motivation: Incompleteness of BD/mSBD (Sec 3.2)

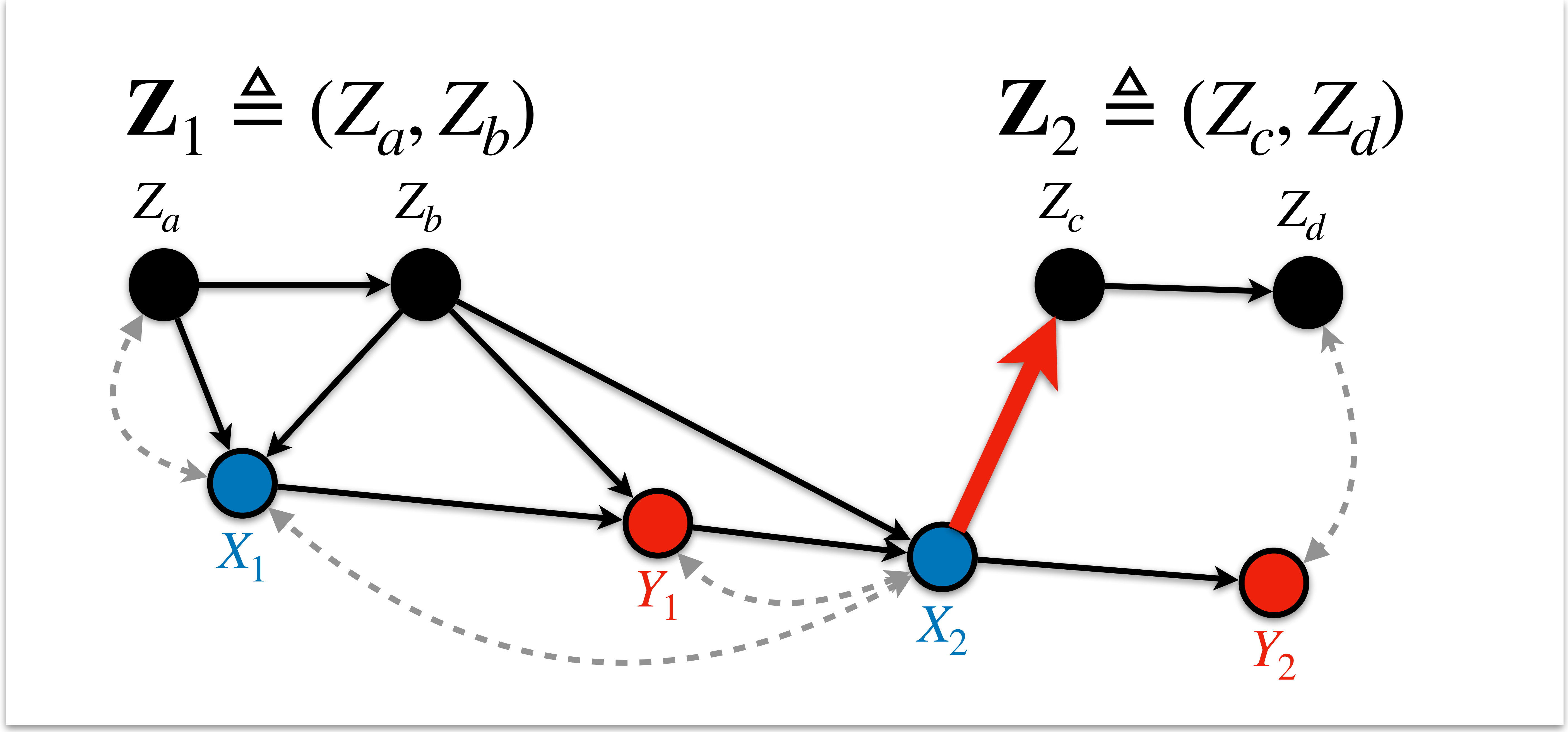
∃ examples s.t. $P(\mathbf{y} \mid do(\mathbf{x}))$ is BD adjustment even if BD criterion fails.



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

Motivation: Incompleteness of BD/mSBD (Sec 3.2)

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



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

"mSBD adjustment"

❷ $P(y_1 y_2 \mid do(x_1 x_2)) = \sum_{\mathbf{z}_1 \mathbf{z}_2} \overbrace{P(y_2 \mid \text{prev}_1, \mathbf{z}_2 x_2) P(y_1 \mathbf{z}_2 \mid \mathbf{z}_1 x_1) P(\mathbf{z}_1)}^{\text{"mSBD adjustment"}}$

Complete Seq. Adjustment Criterion (Sec 3.2)

Complete Seq. Adjustment Criterion (Sec 3.2)

[Def. 29] Sequential Adjustment Criterion (SAC)

A seq. $\mathbf{Z} = (\mathbf{Z}_1, \cdots, \mathbf{Z}_m)$ satisfies the SAC if, for $i = 1, \cdots, m$, $\mathbf{Z}_i \cup \mathbf{prev}_{i-1}$ satisfies the *adjustment criterion* relative to $(\mathbf{X}_i, \mathbf{Y}^{\geq i})$

Complete Seq. Adjustment Criterion (Sec 3.2)

[Def. 29] Sequential Adjustment Criterion (SAC)

A seq. $\mathbf{Z} = (\mathbf{Z}_1, \cdots, \mathbf{Z}_m)$ satisfies the SAC if, for $i = 1,\cdots, m$, $\mathbf{Z}_i \cup \mathbf{prev}_{i-1}$ satisfies the *adjustment criterion* relative to $(\mathbf{X}_i, \mathbf{Y}^{\geq i})$

Complete criterion for the BD adjustment [Shpitser et al., 2010, van der Zander et al., 2014)

Complete Seq. Adjustment Criterion (Sec 3.2)

[Def. 29] Sequential Adjustment Criterion (SAC)

A seq. $\mathbf{Z} = (\mathbf{Z}_1, \cdots, \mathbf{Z}_m)$ satisfies the SAC if, for $i = 1, \cdots, m$, $\mathbf{Z}_i \cup \mathbf{prev}_{i-1}$ satisfies the *adjustment criterion* relative to $(\mathbf{X}_i, \mathbf{Y}^{\geq i})$

$\Leftrightarrow$

[Theorem 10] Completeness

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

Complete Seq. Adjustment Criterion (Sec 3.2)

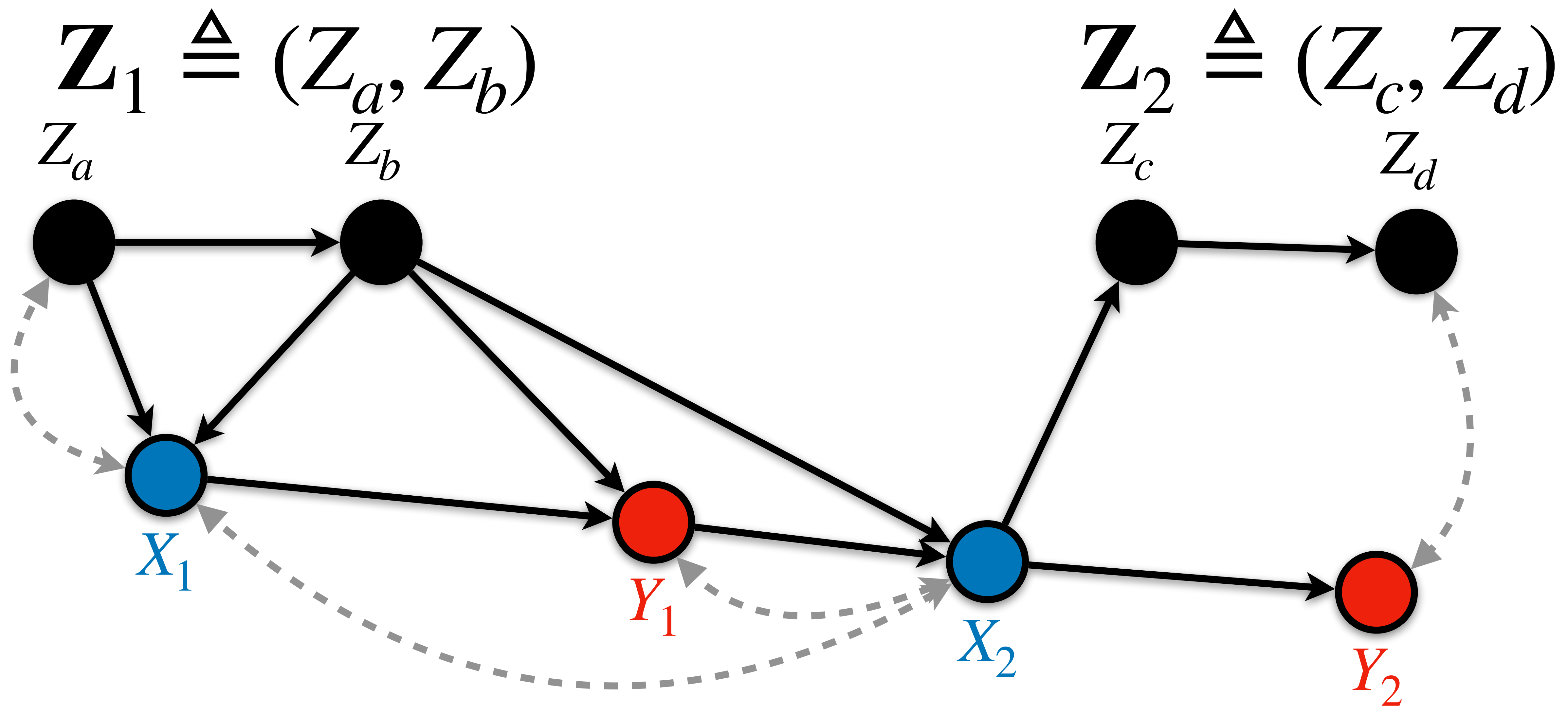
[Def. 29] Sequential Adjustment Criterion (SAC)

A seq. $\mathbf{Z} = (\mathbf{Z}_1, \cdots, \mathbf{Z}_m)$ satisfies the SAC if, for $i = 1, \cdots, m$, $\mathbf{Z}_i \cup \mathbf{prev}_{i-1}$ satisfies the *adjustment criterion* relative to $(\mathbf{X}_i, \mathbf{Y}^{\geq i})$

$\Leftrightarrow$

[Theorem 10] Completeness

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



X mSBD fails

Complete Seq. Adjustment Criterion (Sec 3.2)

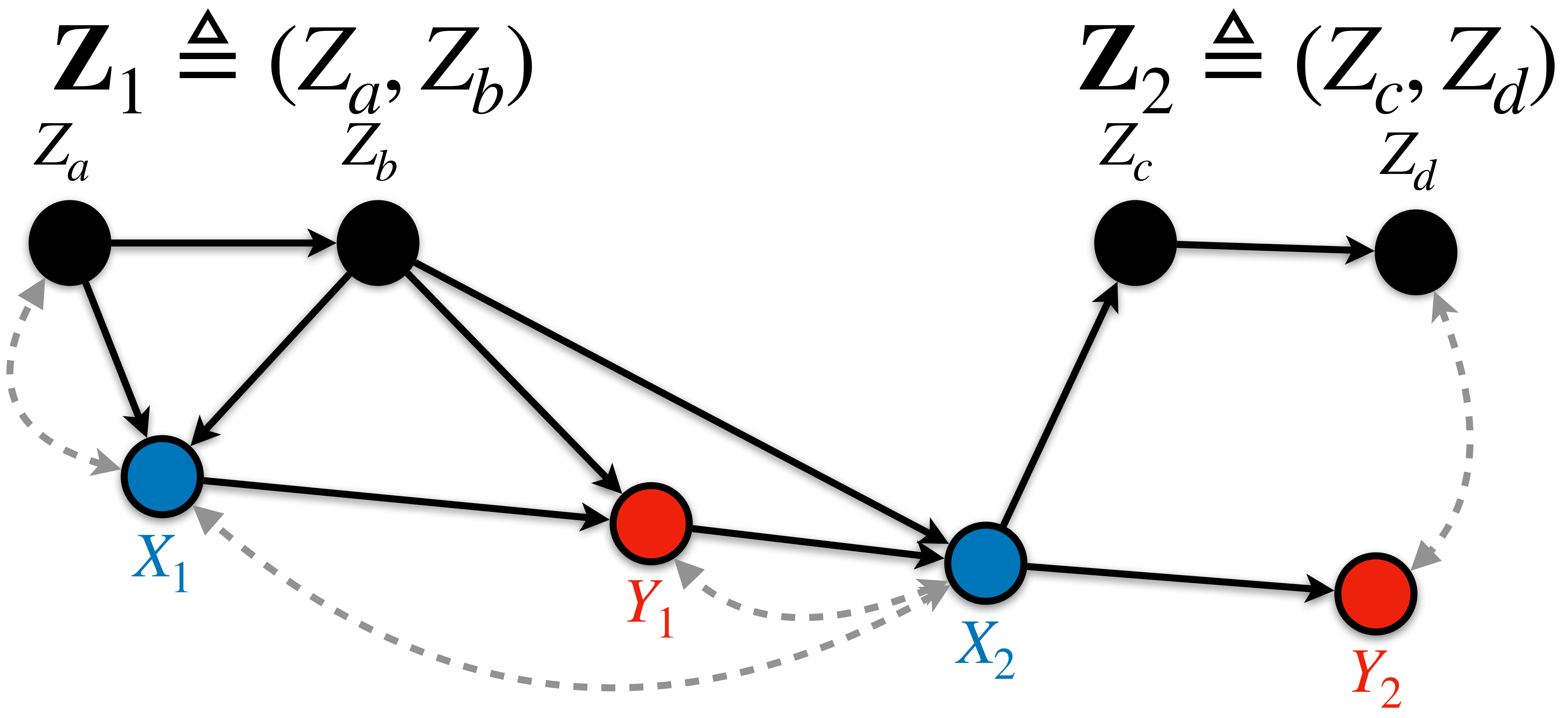
[Def. 29] Sequential Adjustment Criterion (SAC)

A seq. $\mathbf{Z} = (\mathbf{Z}_1, \cdots, \mathbf{Z}_m)$ satisfies the SAC if, for $i = 1, \cdots, m$, $\mathbf{Z}_i \cup \mathbf{prev}_{i-1}$ satisfies the *adjustment criterion* relative to $(\mathbf{X}_i, \mathbf{Y}^{\geq i})$

$\Leftrightarrow$

[Theorem 10] Completeness

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



✗ mSBD fails

✔ SAC holds

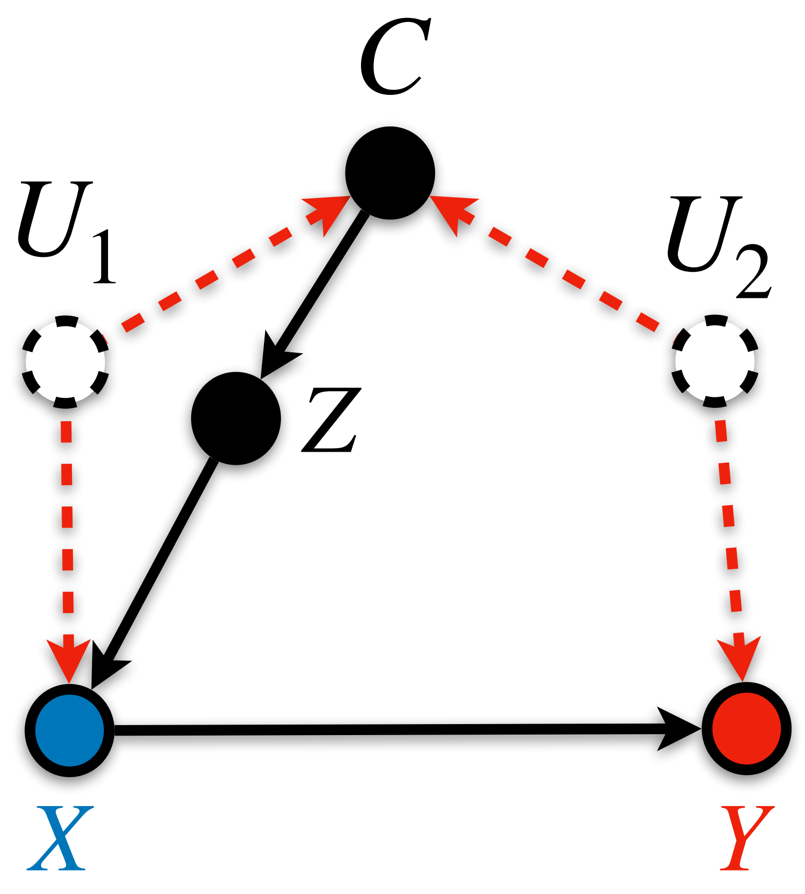
Estimating Causal Effects in 3-Steps

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



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

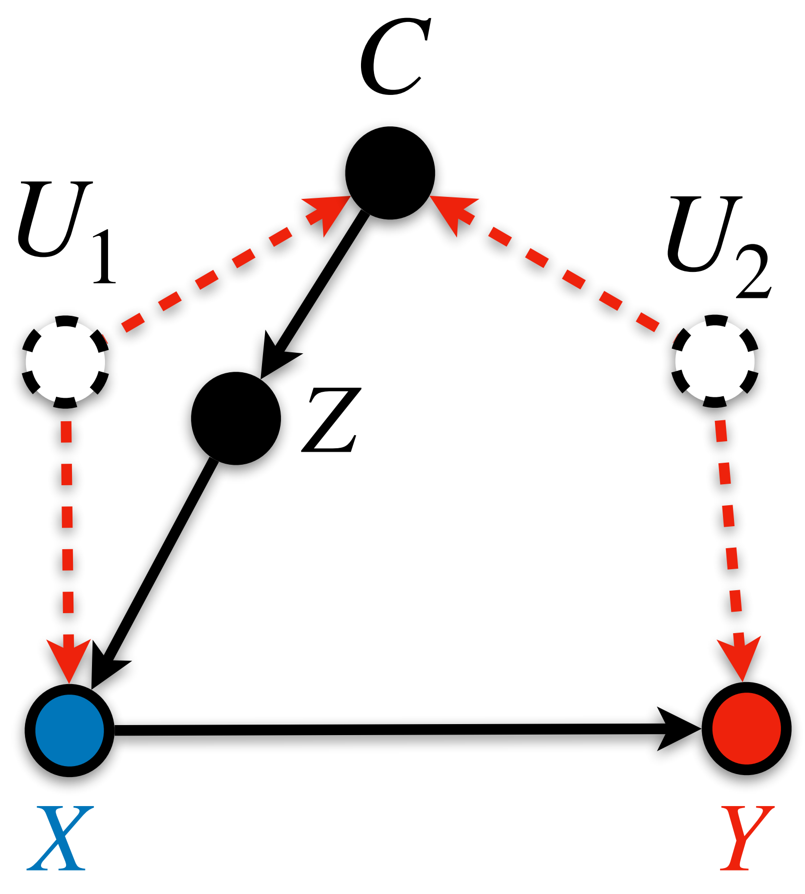
$P_{\text{do}(Z)}(CXY)$: $P(C)P(XY \mid ZC)$

$P_{\text{do}(Z)}(XY)$: $\sum_c P(c)P(XY \mid Zc)$

$P(Y \mid \text{do}(X))$: $\dfrac{\sum_c P(c)P(XY \mid Zc)}{\sum_c P(c)P(X \mid Zc)}$

Estimating Causal Effects in 3-Steps

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

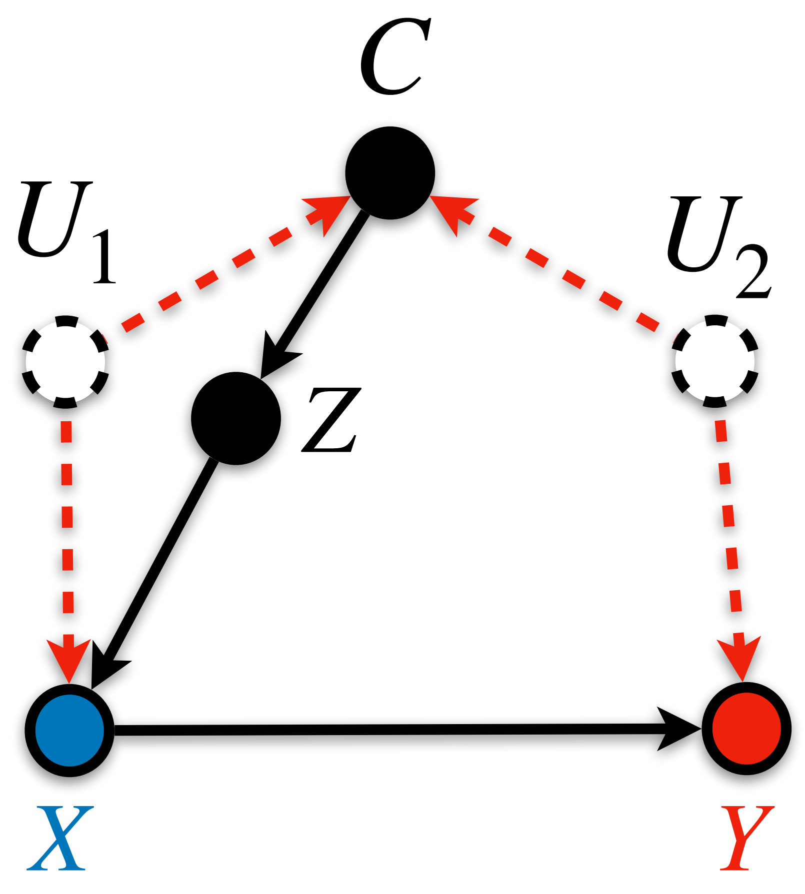


$$P(CZXY) \xrightarrow{\text{Factorization}} P_{\text{do}(Z)}(CXY) \xrightarrow{\text{Marginalization}} P_{\text{do}(Z)}(XY) \xrightarrow{\text{Factorization}} P(Y \mid \text{do}(X))$$

SAC

Estimating Causal Effects in 3-Steps

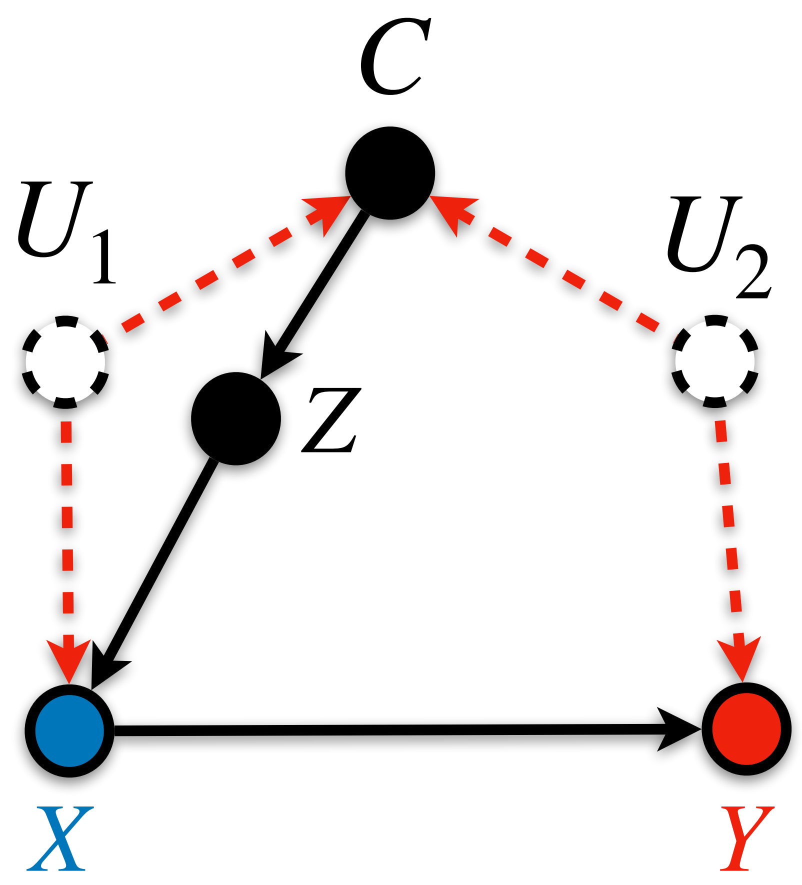
2 **Express** causal effects as a function of BD



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

Estimating Causal Effects in 3-Steps

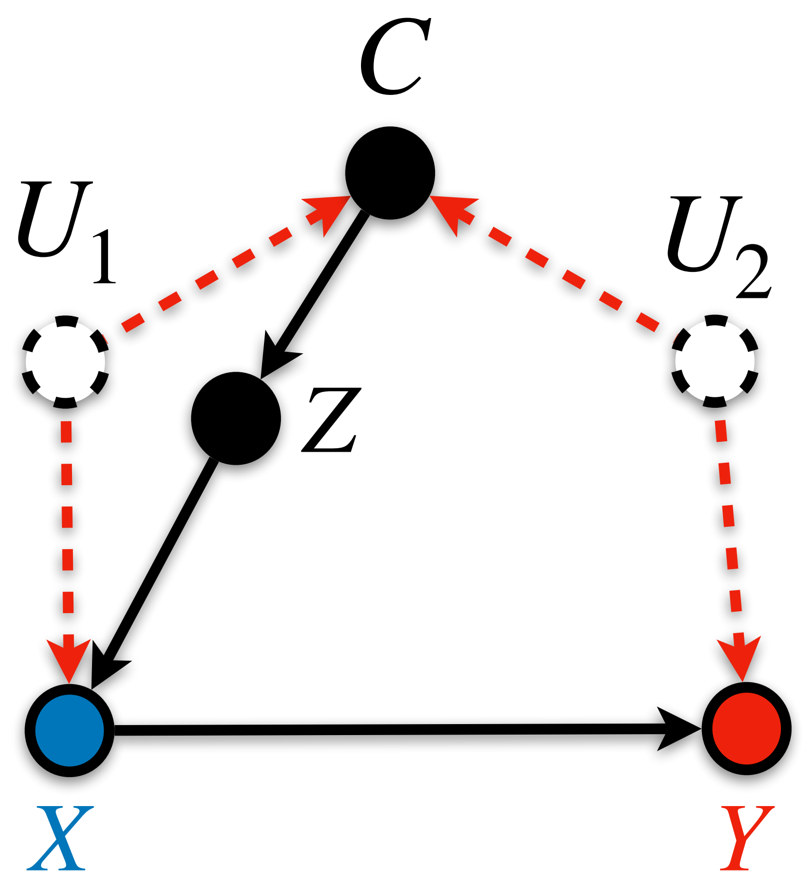
2 **Express** causal effects as a function of BD



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

Estimating Causal Effects in 3-Steps

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



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

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

Estimating Causal Effects in 3-Steps

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

Theorem 14

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（Algo 4）

DML-ID: Estimator for Identifiable Causal Effects

DML-ID: Estimator for Identifiable Causal Effects

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

DML-ID: Estimator for Identifiable Causal Effects

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

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

DML-ID: Estimator for Identifiable Causal Effects

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

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

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

"DML-ID" (Def 36)

DML-ID: Estimator for Identifiable Causal Effects

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

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

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

"DML-ID" (Def 36)

DML-ID: Estimator for Identifiable Causal Effects

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

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

DML-BD ↓ DML-BD ↓ ⋯ DML-BD ↓

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

"DML-ID" (Def 36)

Robustness of DML-ID

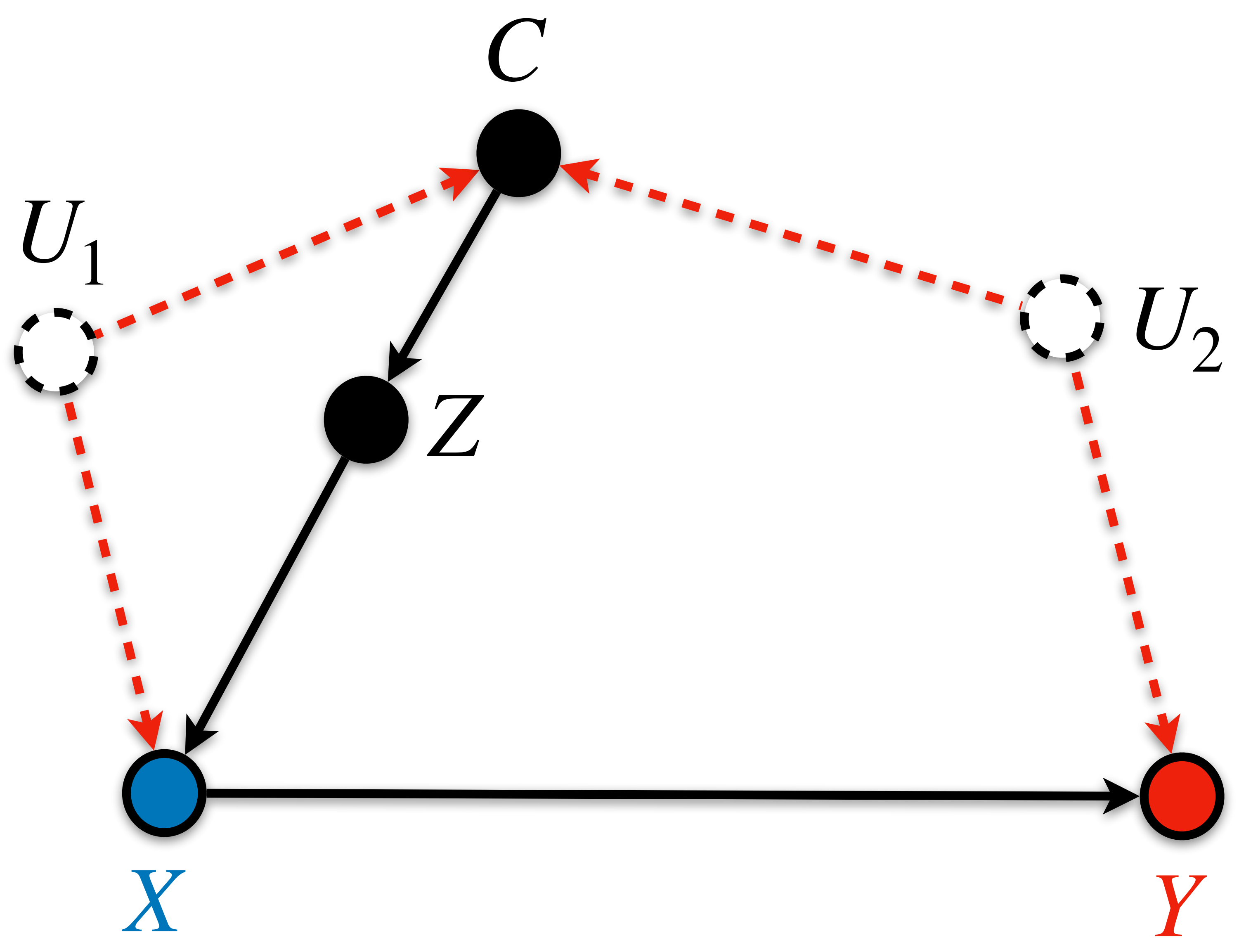
Theorem 16

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

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

DML-ID - Simulation (Sec. 3.5)

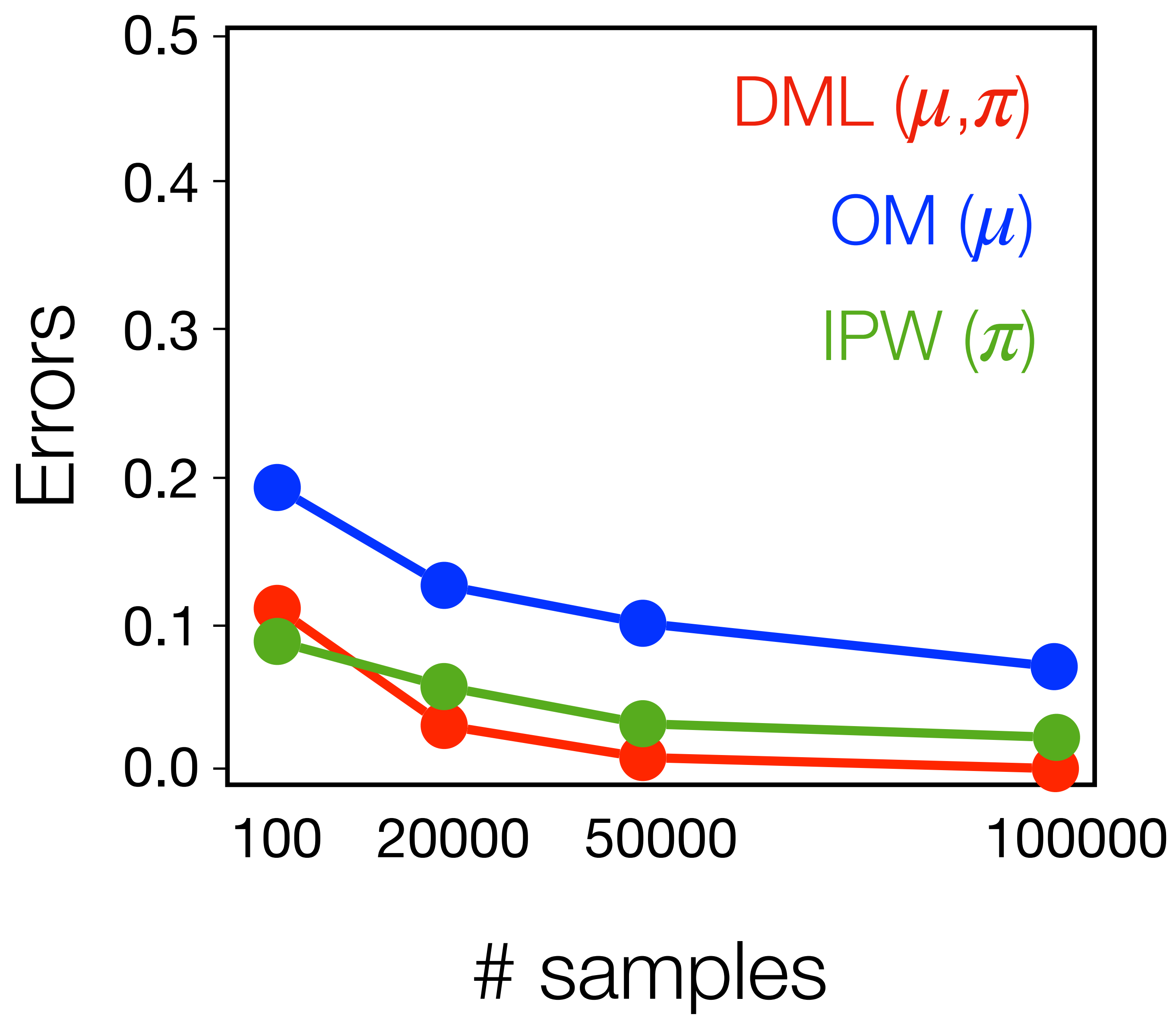
DML-ID - Simulation (Sec. 3.5)



DML-ID - Simulation (Sec. 3.5)

Fast Convergence

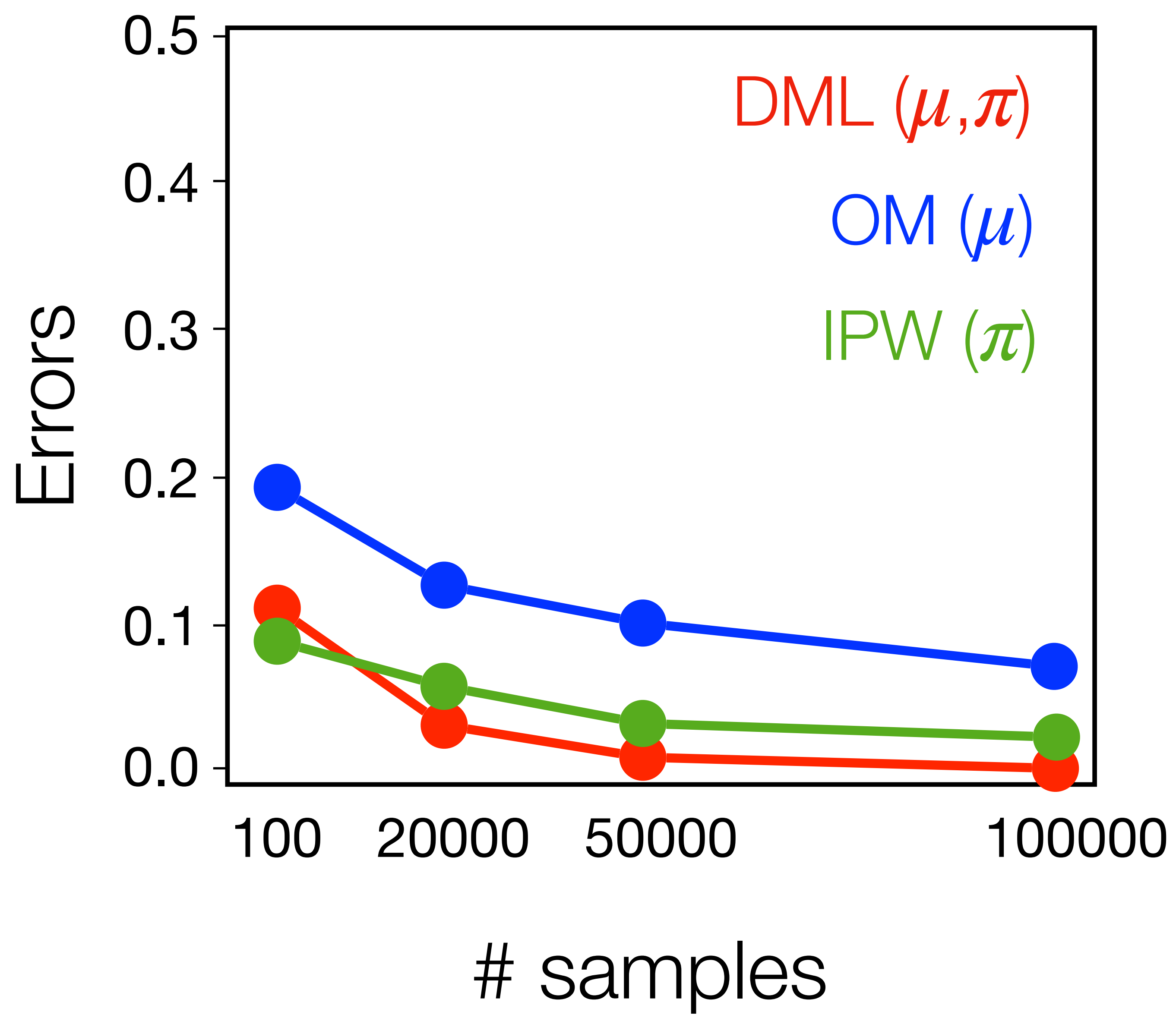
$(\hat{\boldsymbol{\mu}}, \hat{\boldsymbol{\pi}}) \to (\boldsymbol{\mu}_0, \boldsymbol{\pi}_0)$ slowly



DML-ID - Simulation (Sec. 3.5)

Fast Convergence

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

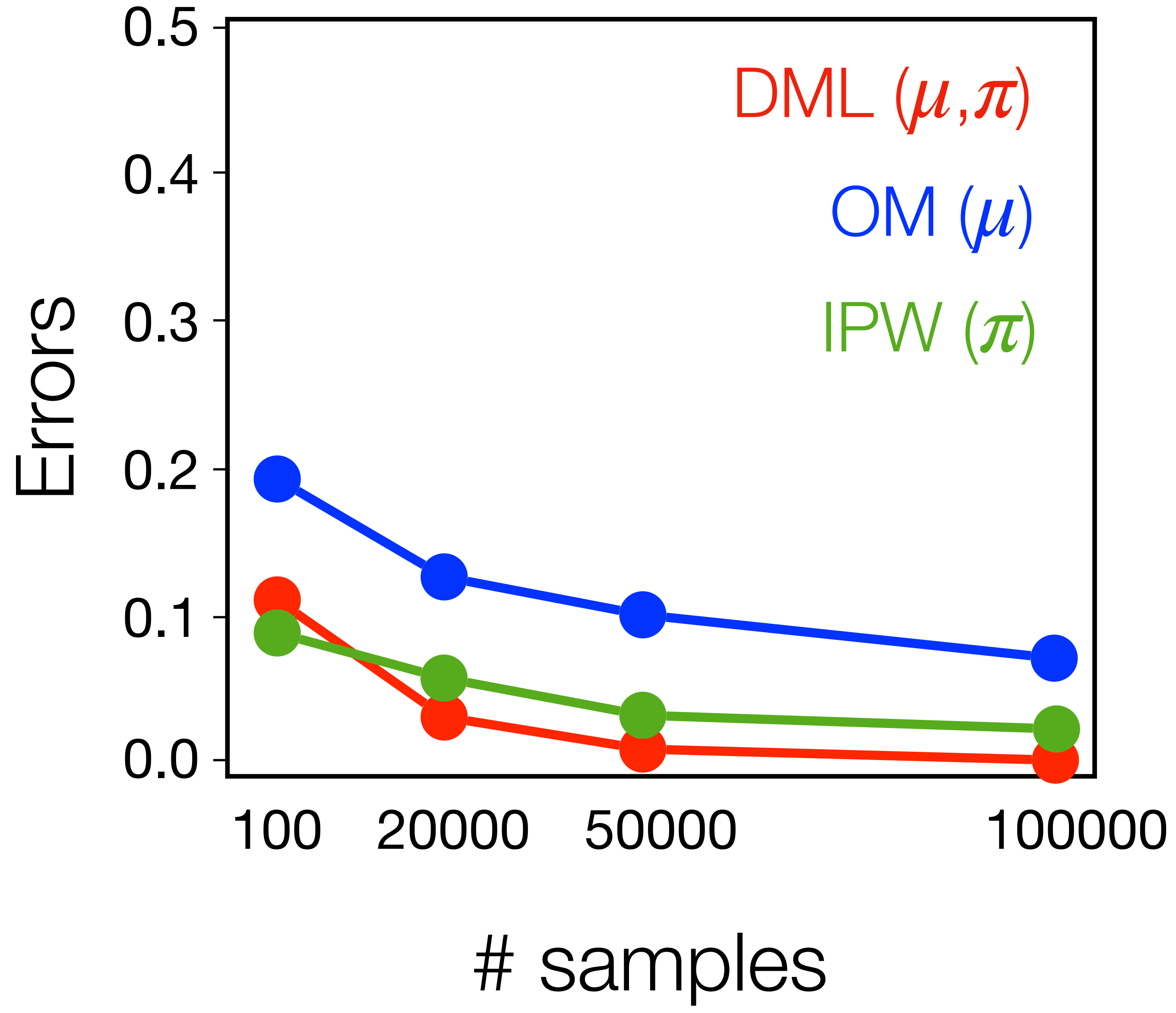


DML-ID converges fast, even when $(\hat{\boldsymbol{\mu}}, \hat{\boldsymbol{\pi}})$ converge slowly

DML-ID - Simulation (Sec. 3.5)

Fast Convergence

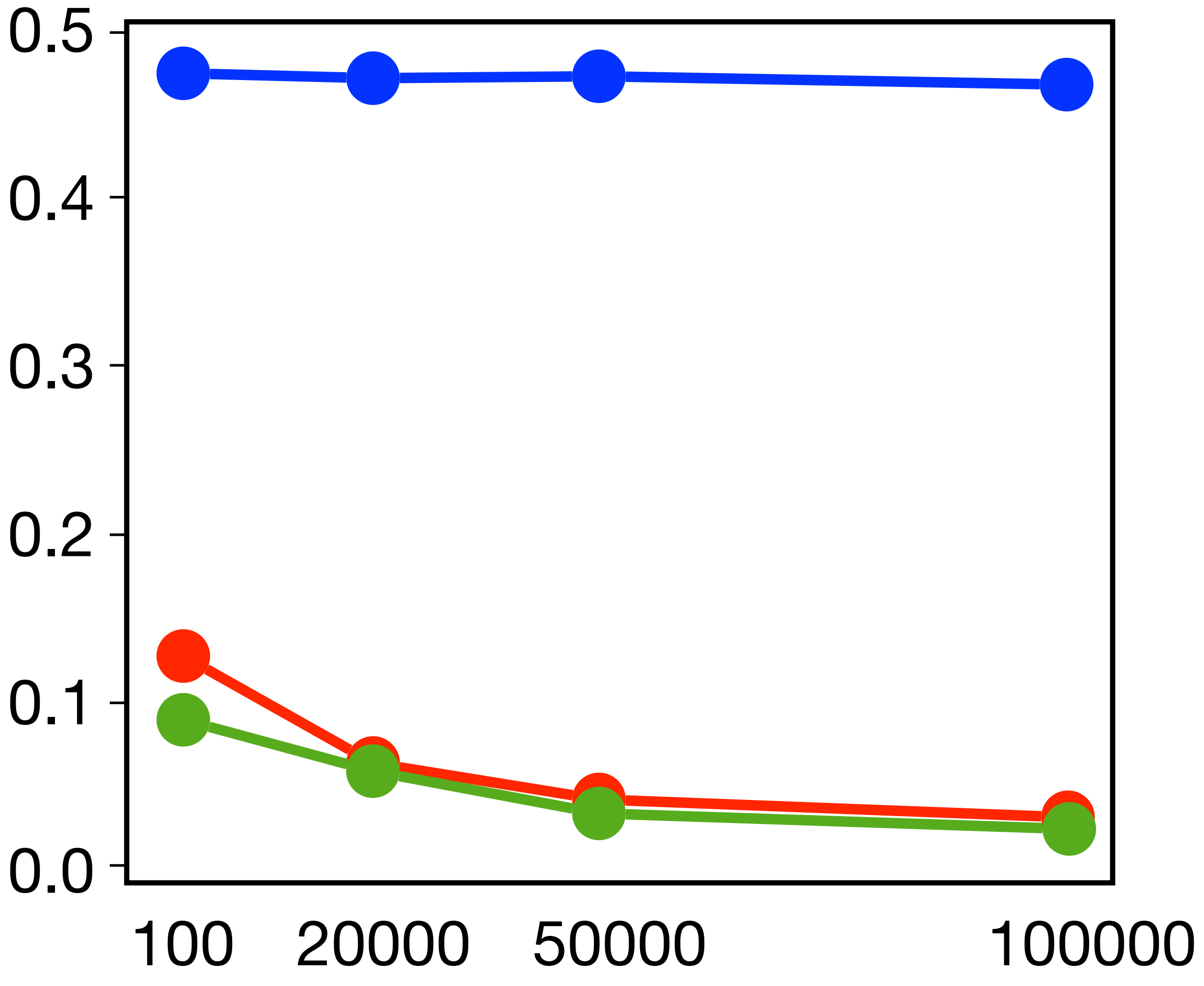
$(\hat{\mu},\hat{\pi}) \rightarrow (\mu_0, \pi_0)$ slowly



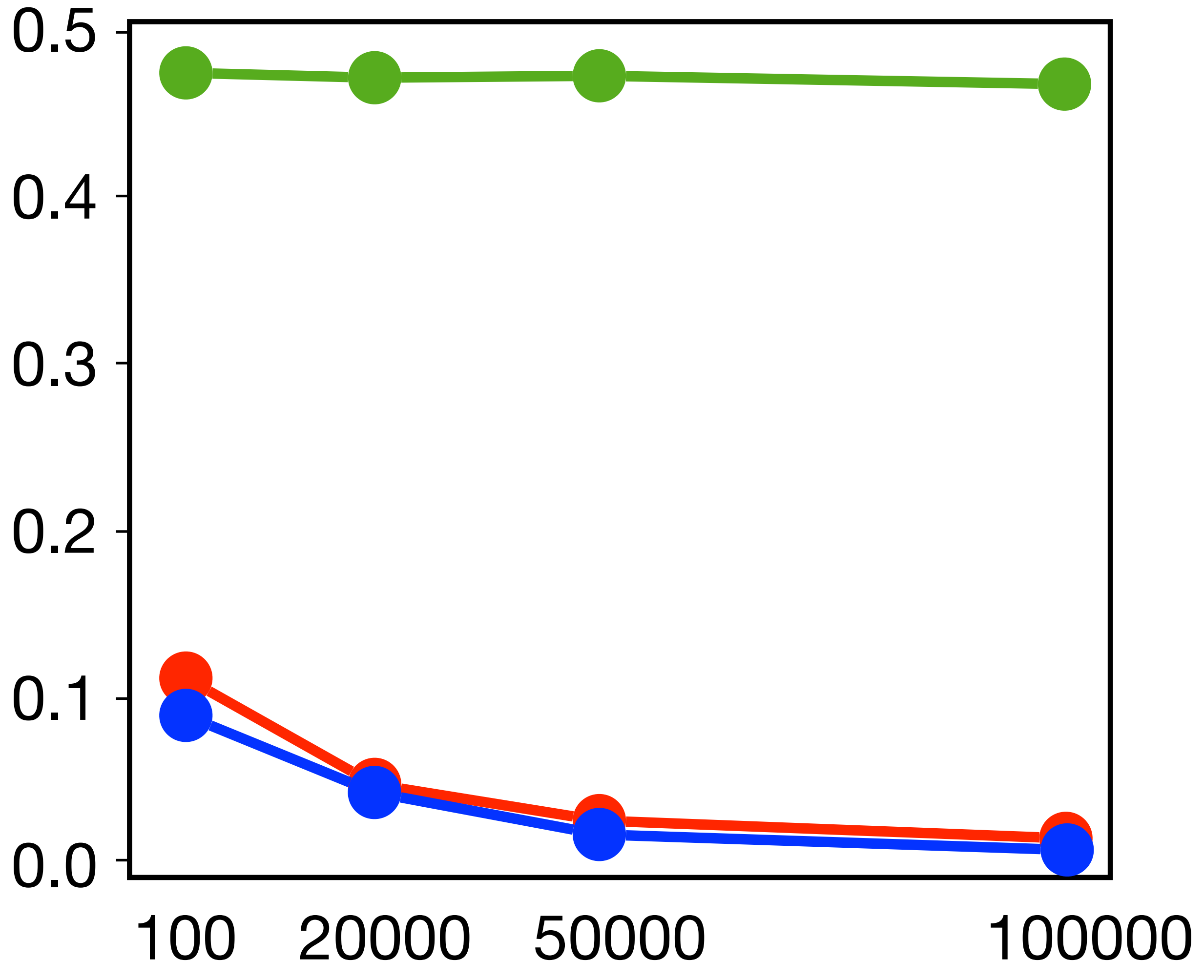
DML-ID converges fast, even when $(\hat{\mu},\hat{\pi})$ converge slowly

Double Robustness

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



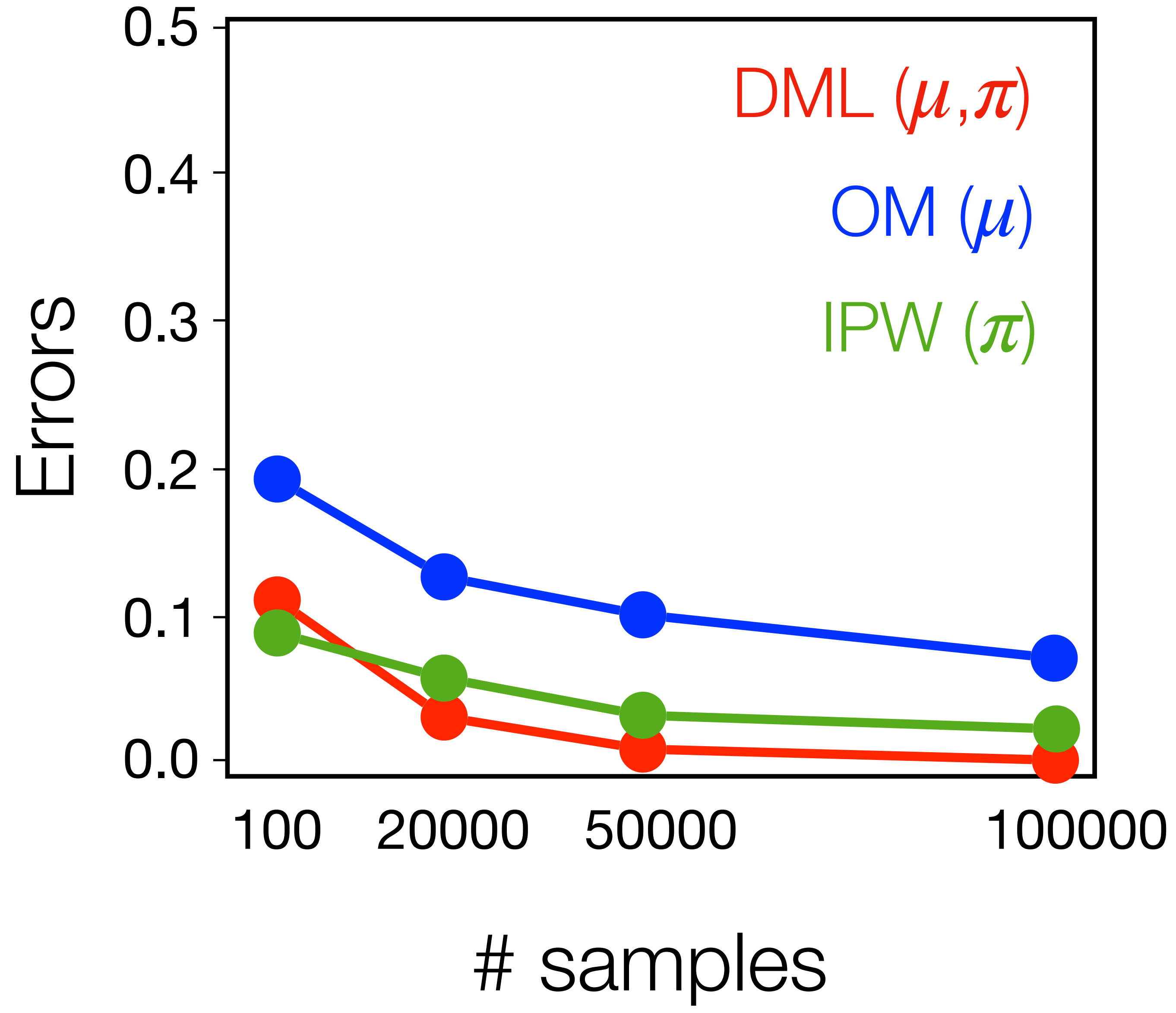
$\hat{\pi}$ misspecified ($\hat{\pi} \neq \pi$)



DML-ID - Simulation (Sec. 3.5)

Fast Convergence

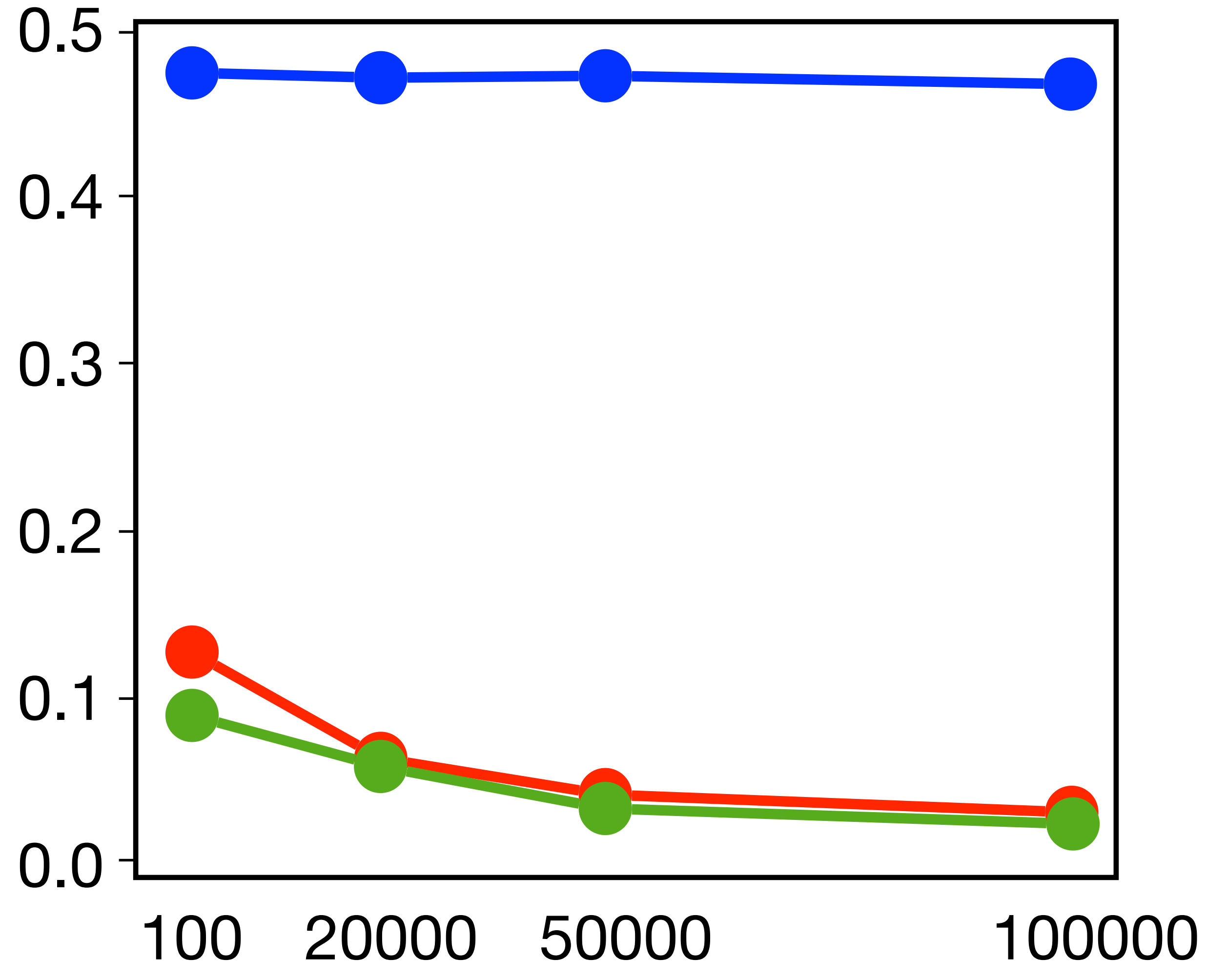
$(\hat{\boldsymbol{\mu}},\hat{\boldsymbol{\pi}}) \rightarrow (\boldsymbol{\mu}_0, \boldsymbol{\pi}_0)$ slowly



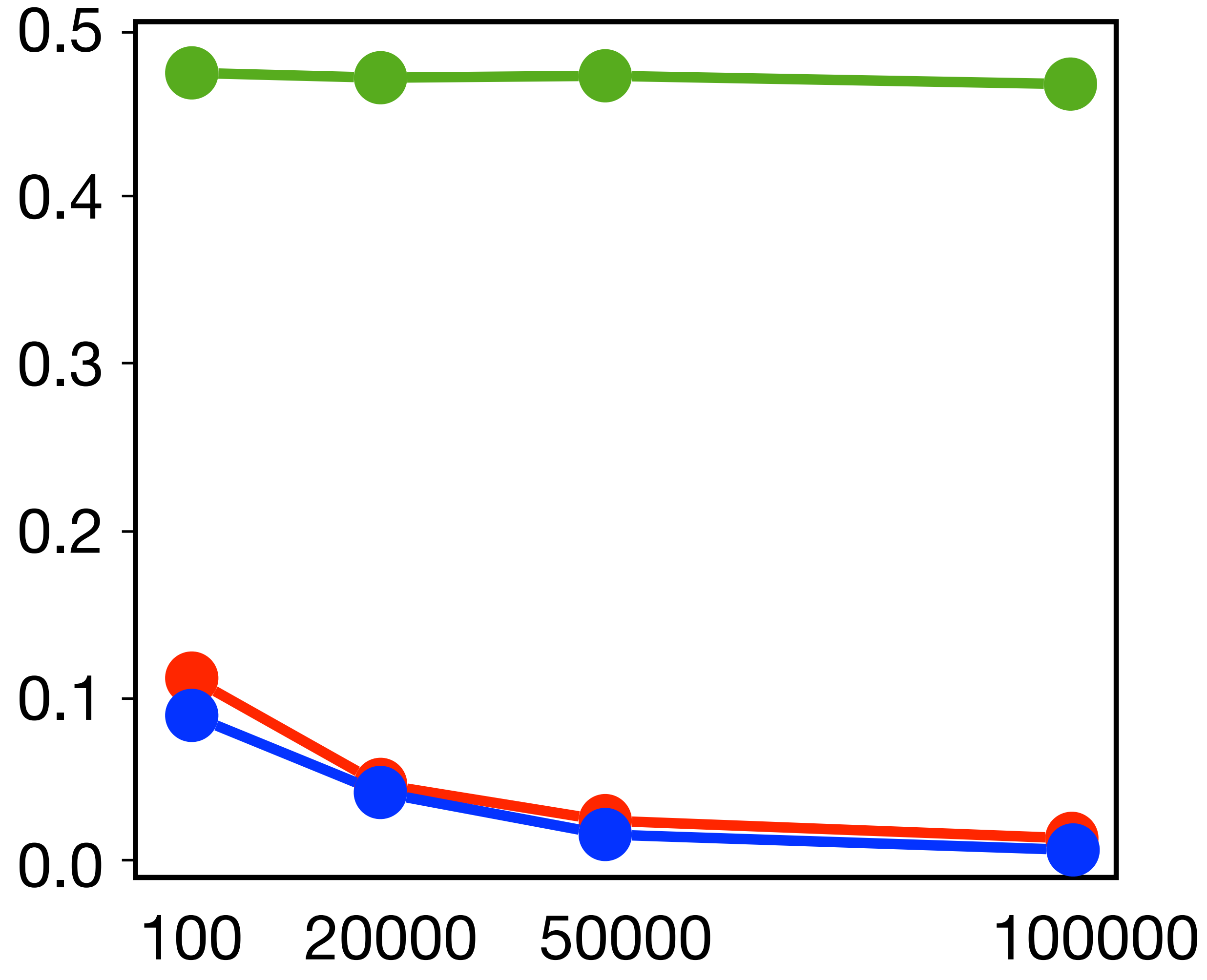
DML-ID converges fast, even when $(\hat{\boldsymbol{\mu}},\hat{\boldsymbol{\pi}})$ converge slowly

Double Robustness

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



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



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

DML-ID - Random (Sec. 3.5)

DML-ID - Random (Sec. 3.5)

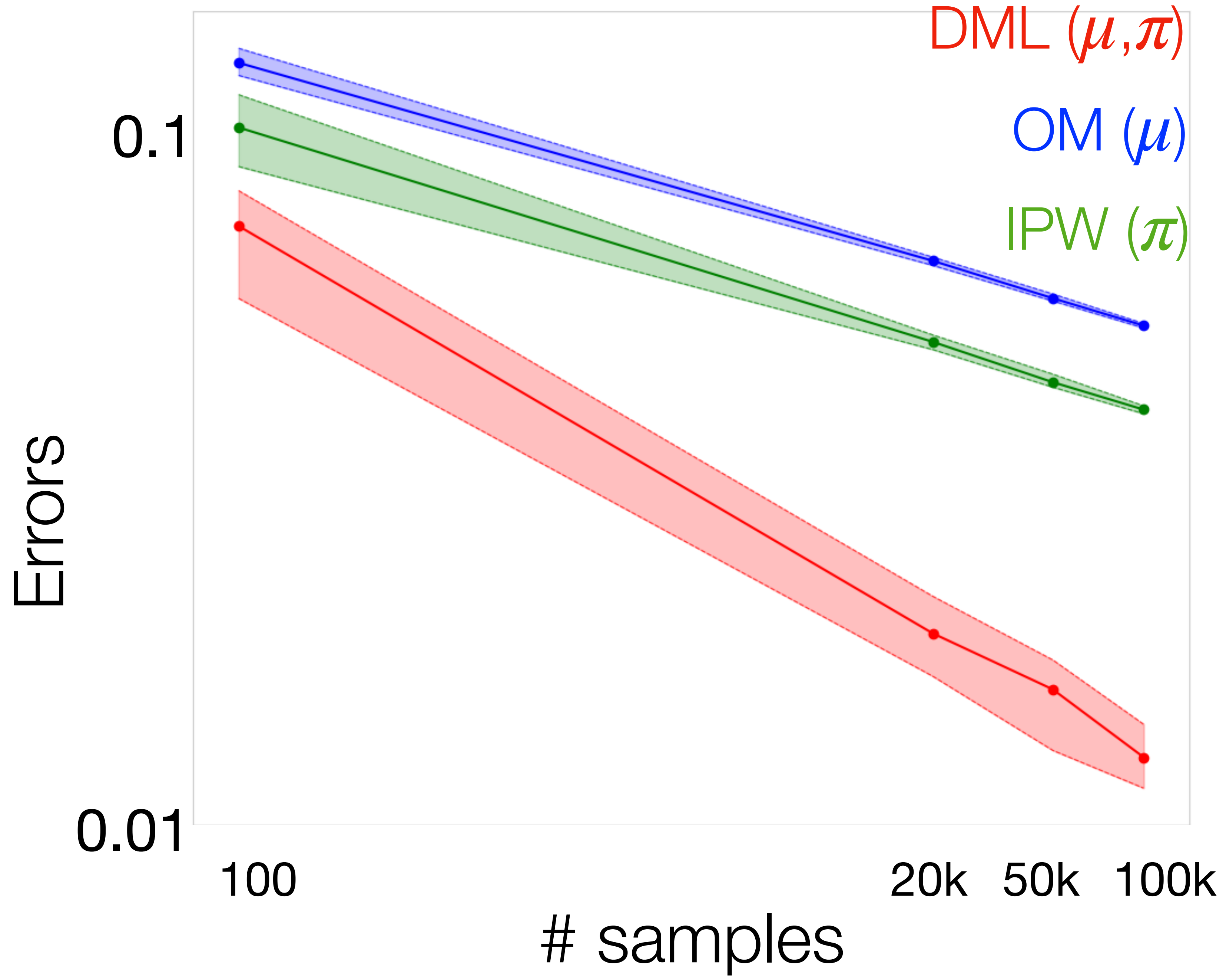
Performed simulations for 100 random graphs.

DML-ID - Random (Sec. 3.5)

Performed simulations for 100 random graphs.

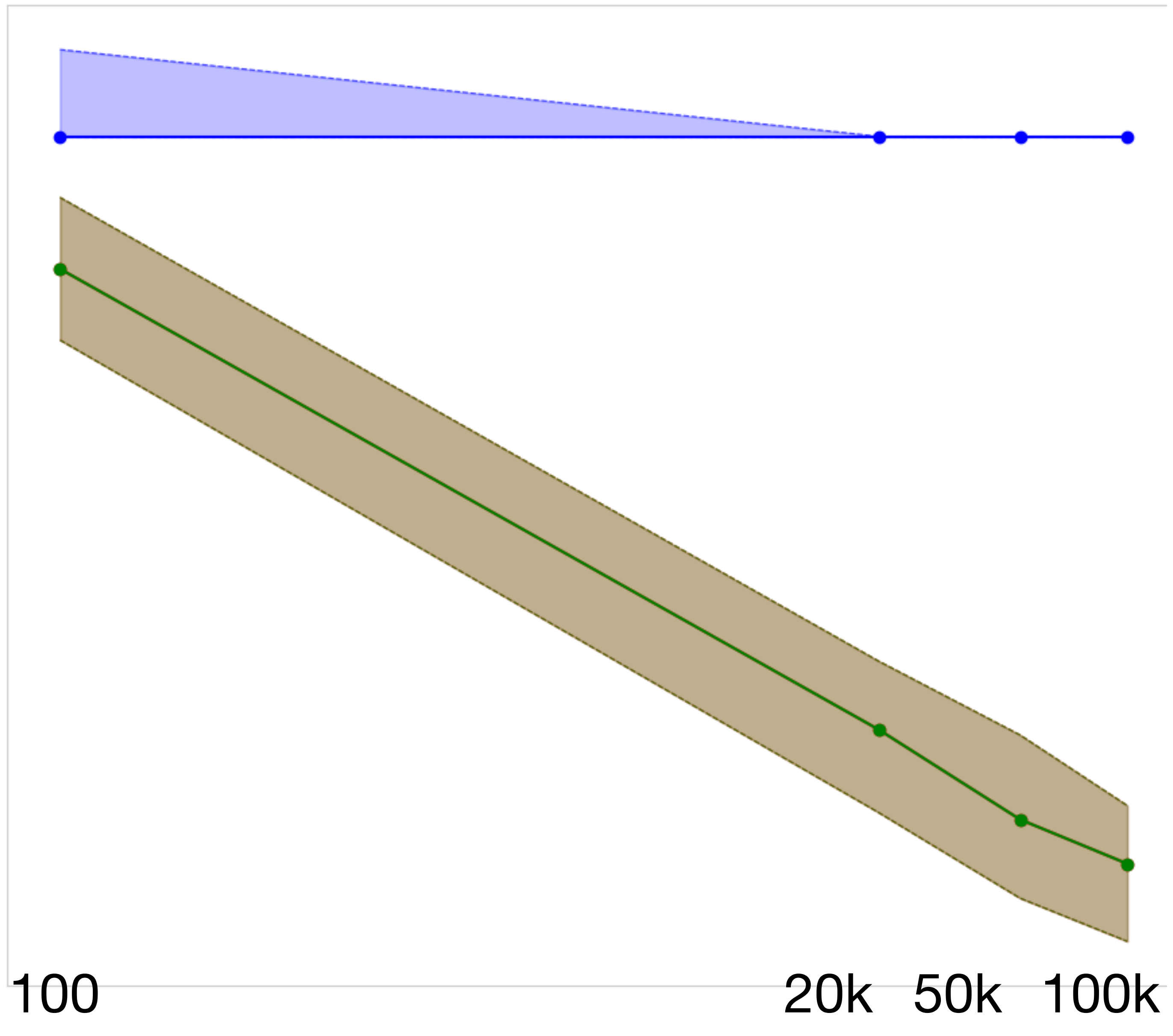
Fast Convergence

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

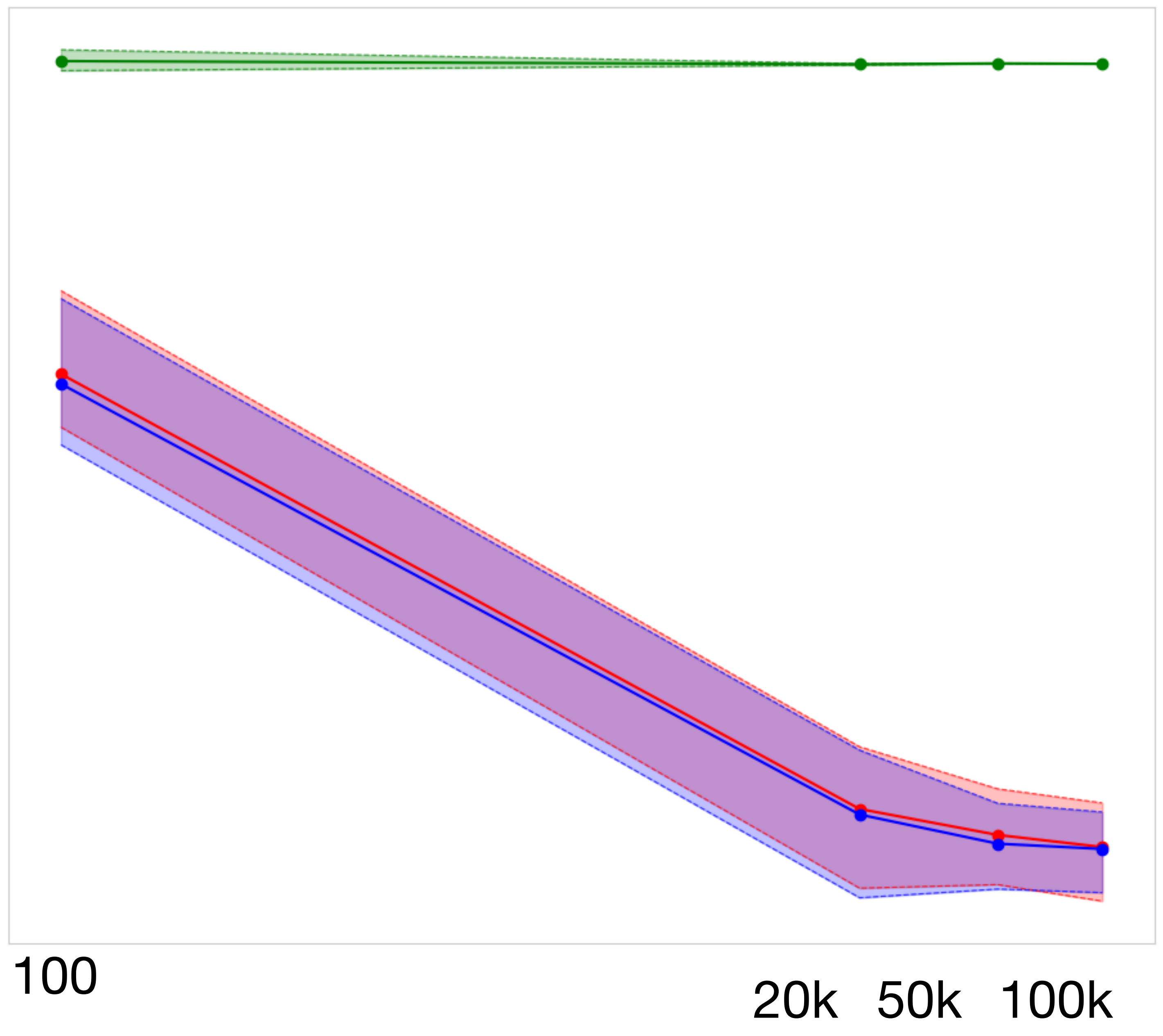


Double Robustness

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



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



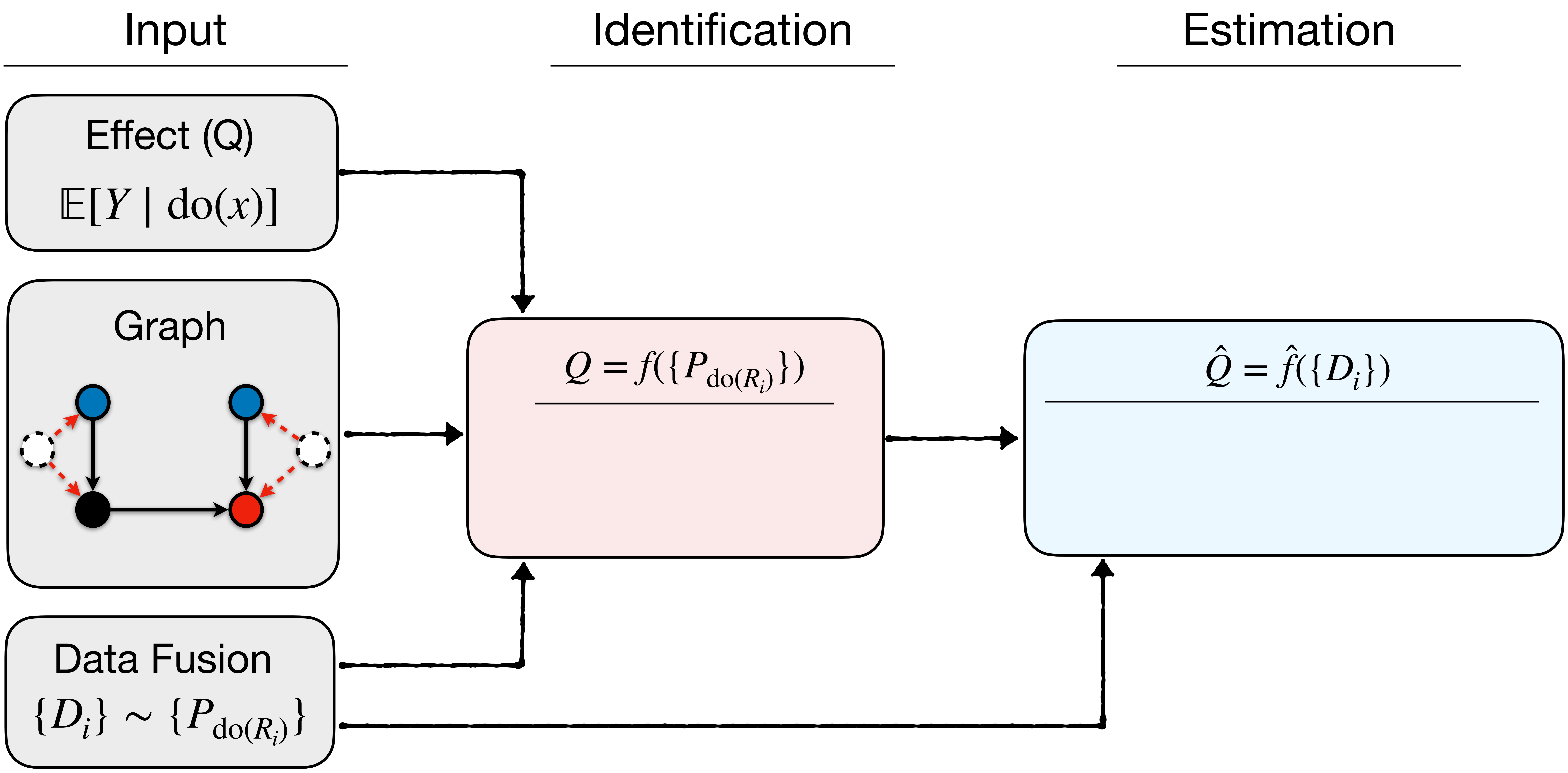
Talk Outline

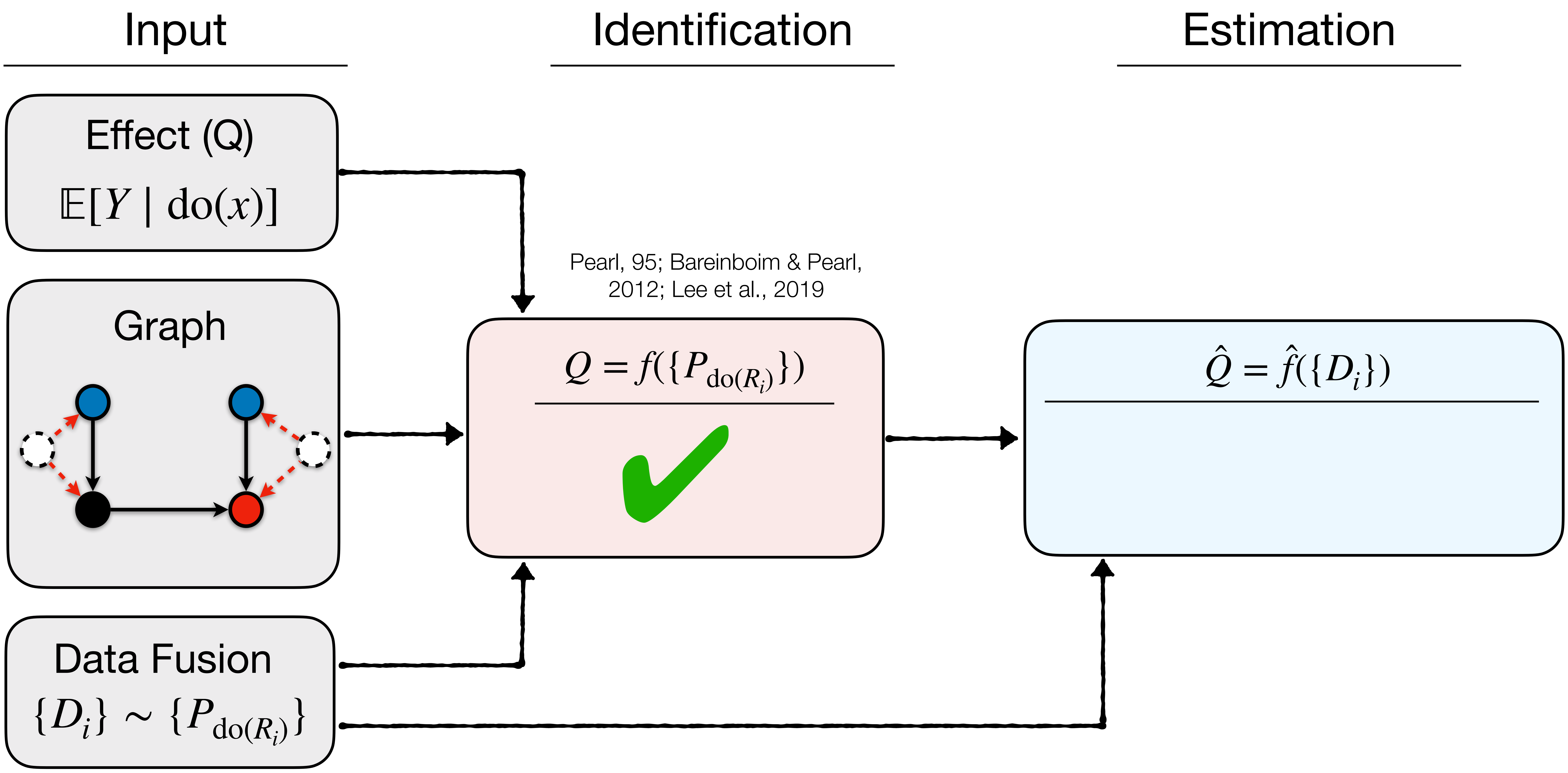
1. **Ch.3** Estimating causal effects from observations
2. **Ch.4** Estimating causal effects from data fusion
3. **Ch.5** Unified causal effect estimation method
4. Conclusion

Talk Outline

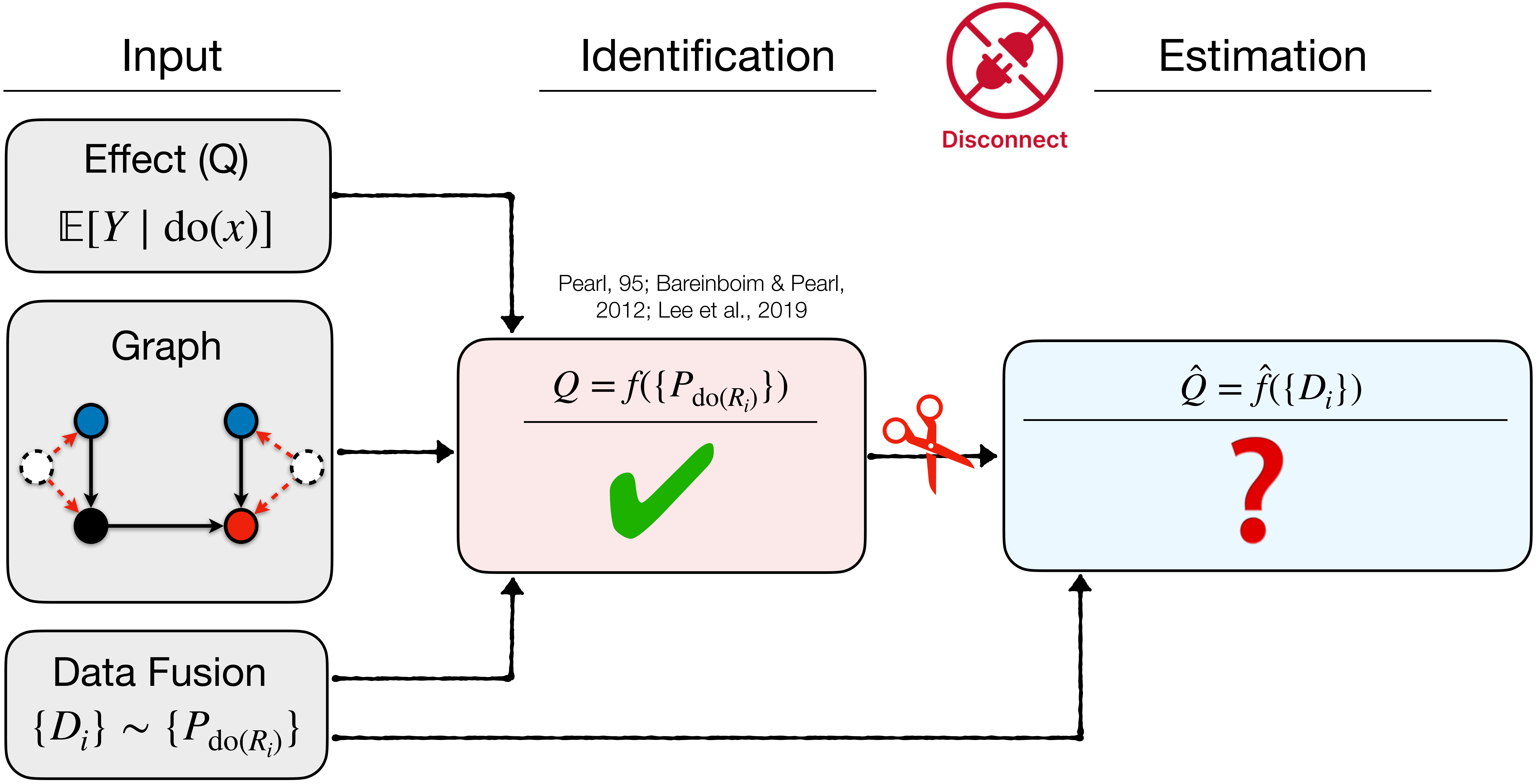
✔

➡ ❷ **Ch.4** Estimating causal effects from data fusion

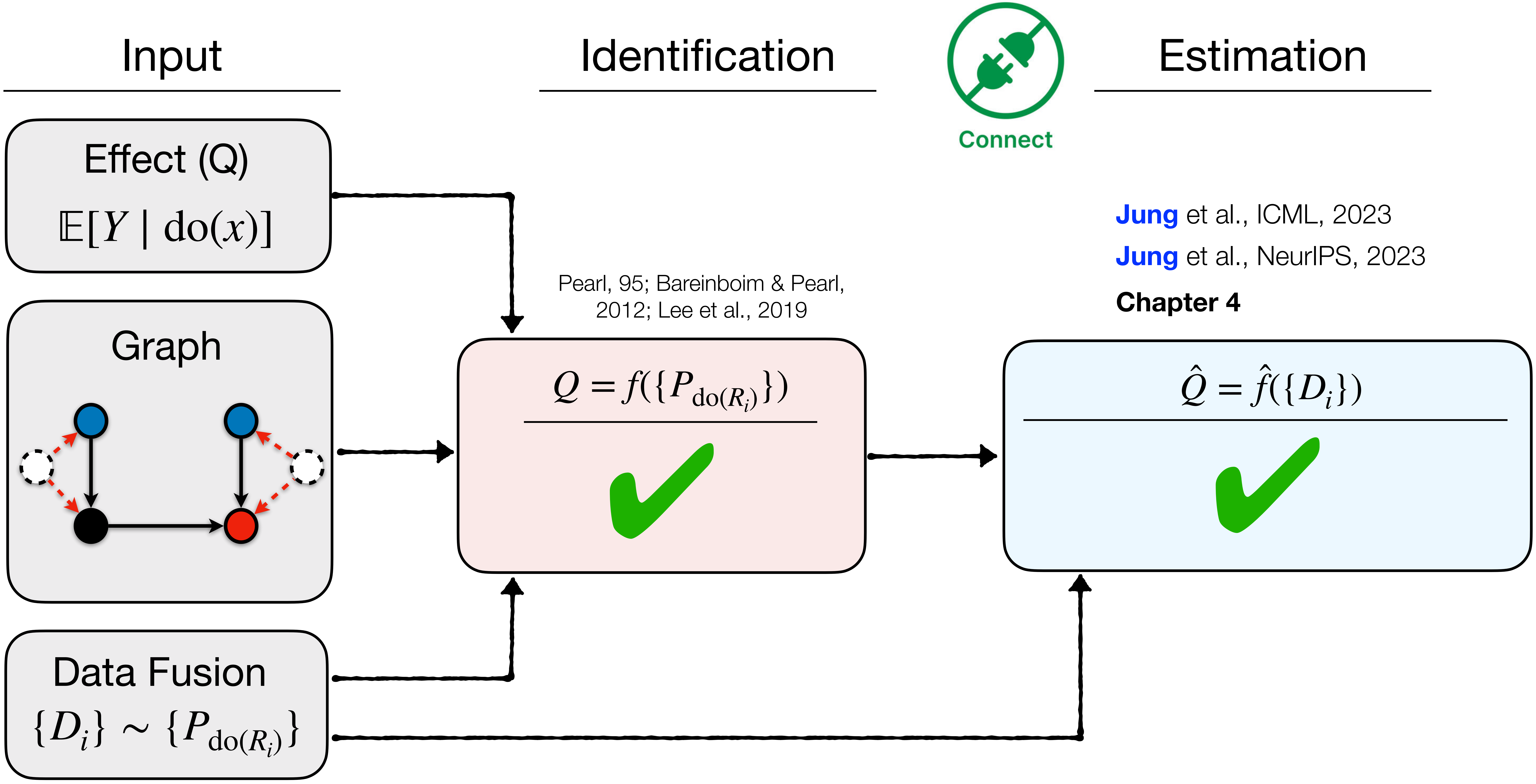

Input
Identification
Estimation
Effect (Q)
$\mathbb{E}[Y \mid do(x)]$
Graph
Data Fusion
$\{D_i\} \sim \{P_{do(R_i)}\}$
$Q = f(\{P_{do(R_i)}\})$
$\hat{Q} = \hat{f}(\{D_i\})$



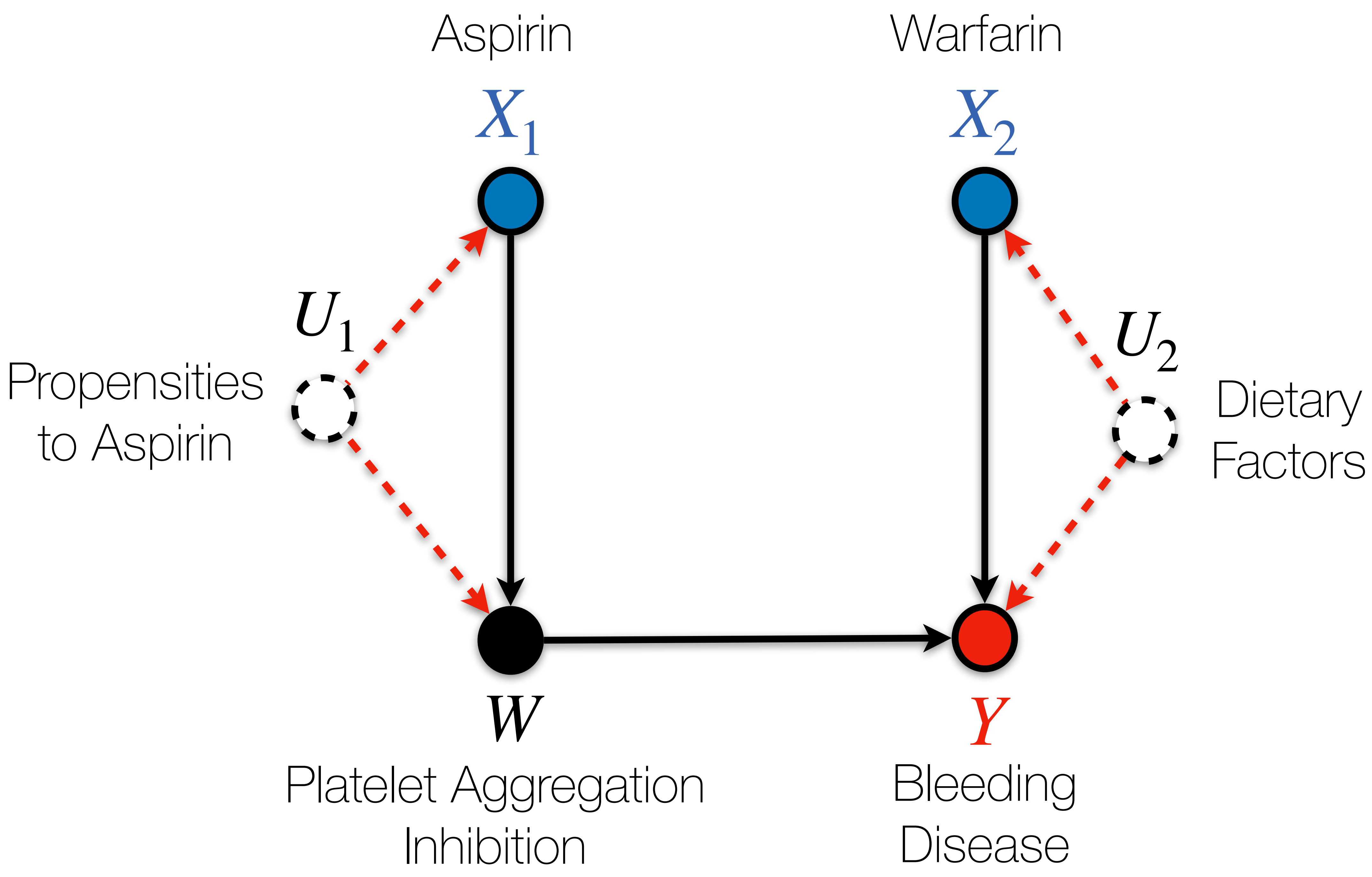
Input
Identification
Estimation
Effect (Q)
$\mathbb{E}[Y \mid \mathrm{do}(x)]$
Graph
Data Fusion
$\{D_i\} \sim \{P_{\mathrm{do}(R_i)}\}$
Pearl, 95; Bareinboim & Pearl, 2012; Lee et al., 2019
$Q = f(\{P_{\mathrm{do}(R_i)}\})$
$\hat{Q} = \hat{f}(\{D_i\})$



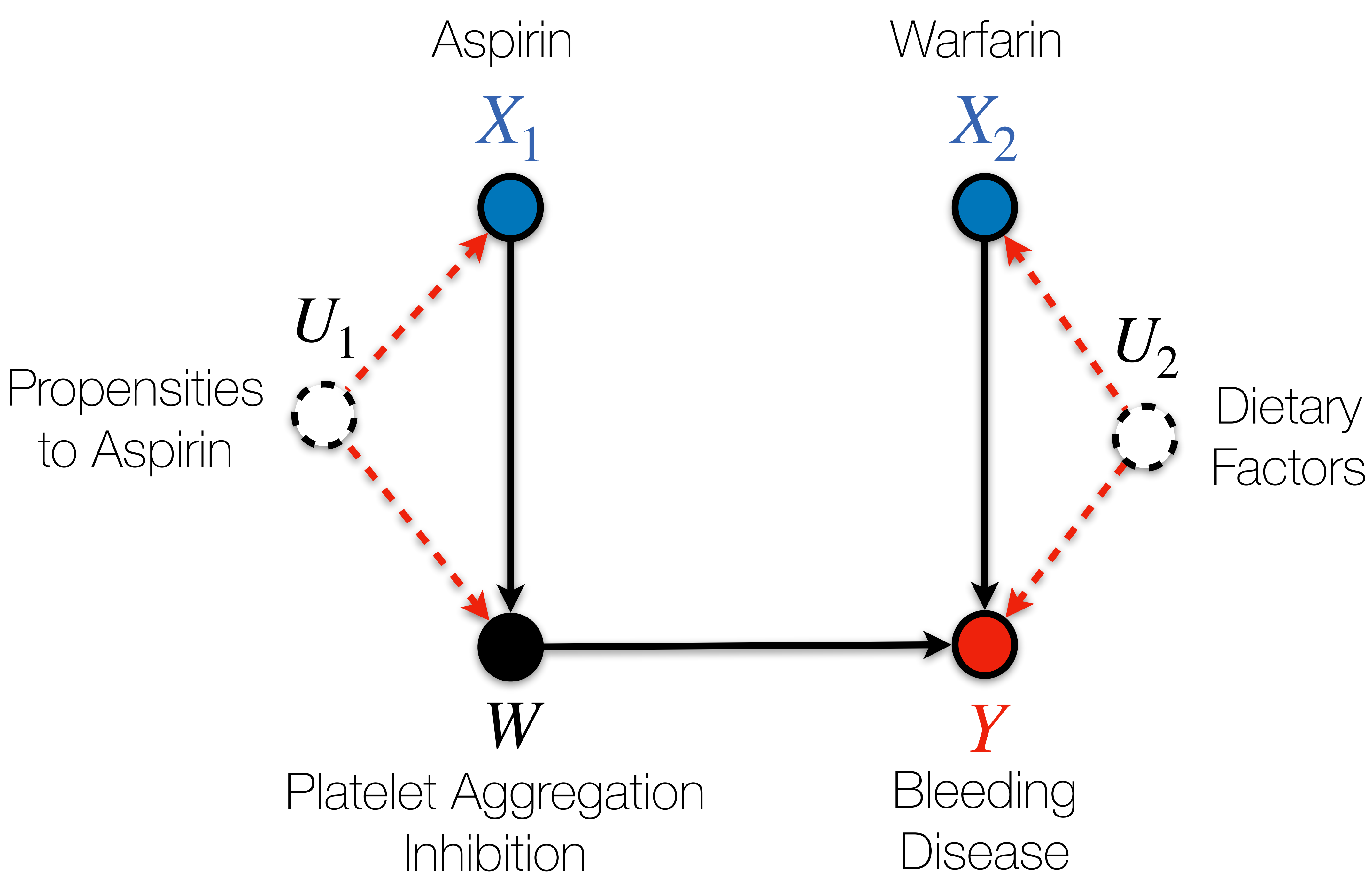
Input
Identification
Disconnect
Estimation
Effect (Q)
$\mathbb{E}[Y \mid do(x)]$
Graph
Data Fusion
$\{D_i\} \sim \{P_{do(R_i)}\}$
Pearl, 95; Bareinboim & Pearl, 2012; Lee et al., 2019
$Q = f(\{P_{do(R_i)}\})$
$\hat{Q} = \hat{f}(\{D_i\})$


Input
Identification
Connect
Estimation
Effect (Q)
$\mathbb{E}[Y \mid do(x)]$
Graph
Data Fusion
$\{D_i\} \sim \{P_{do(R_i)}\}$
Pearl, 95; Bareinboim & Pearl, 2012; Lee et al., 2019
$Q = f(\{P_{do(R_i)}\})$
Jung et al., ICML, 2023
Jung et al., NeurIPS, 2023
Chapter 4
$\hat{Q} = \hat{f}(\{D_i\})$

Motivation: Joint Treatment Effect Estimation

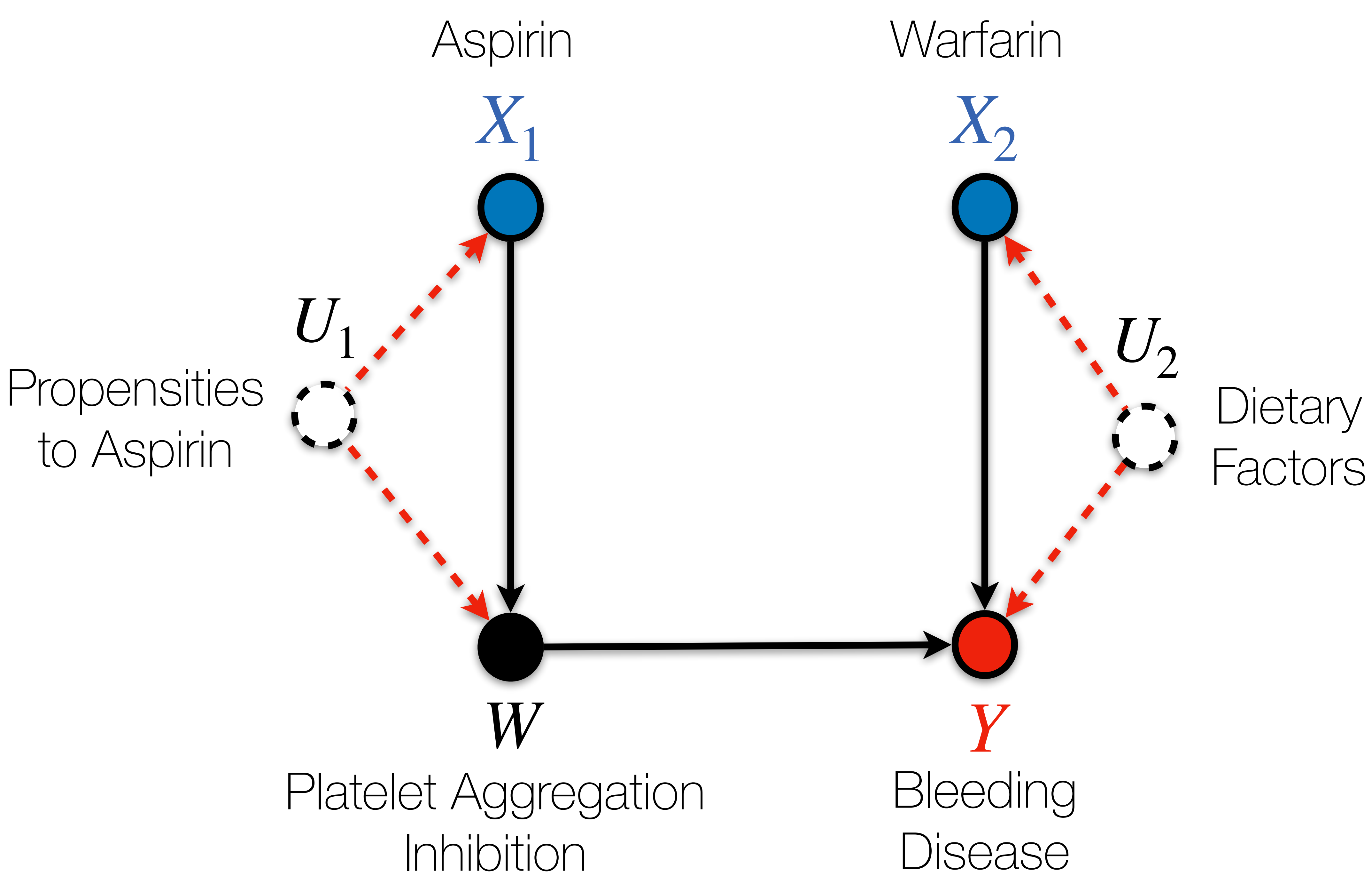


Motivation: Joint Treatment Effect Estimation



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

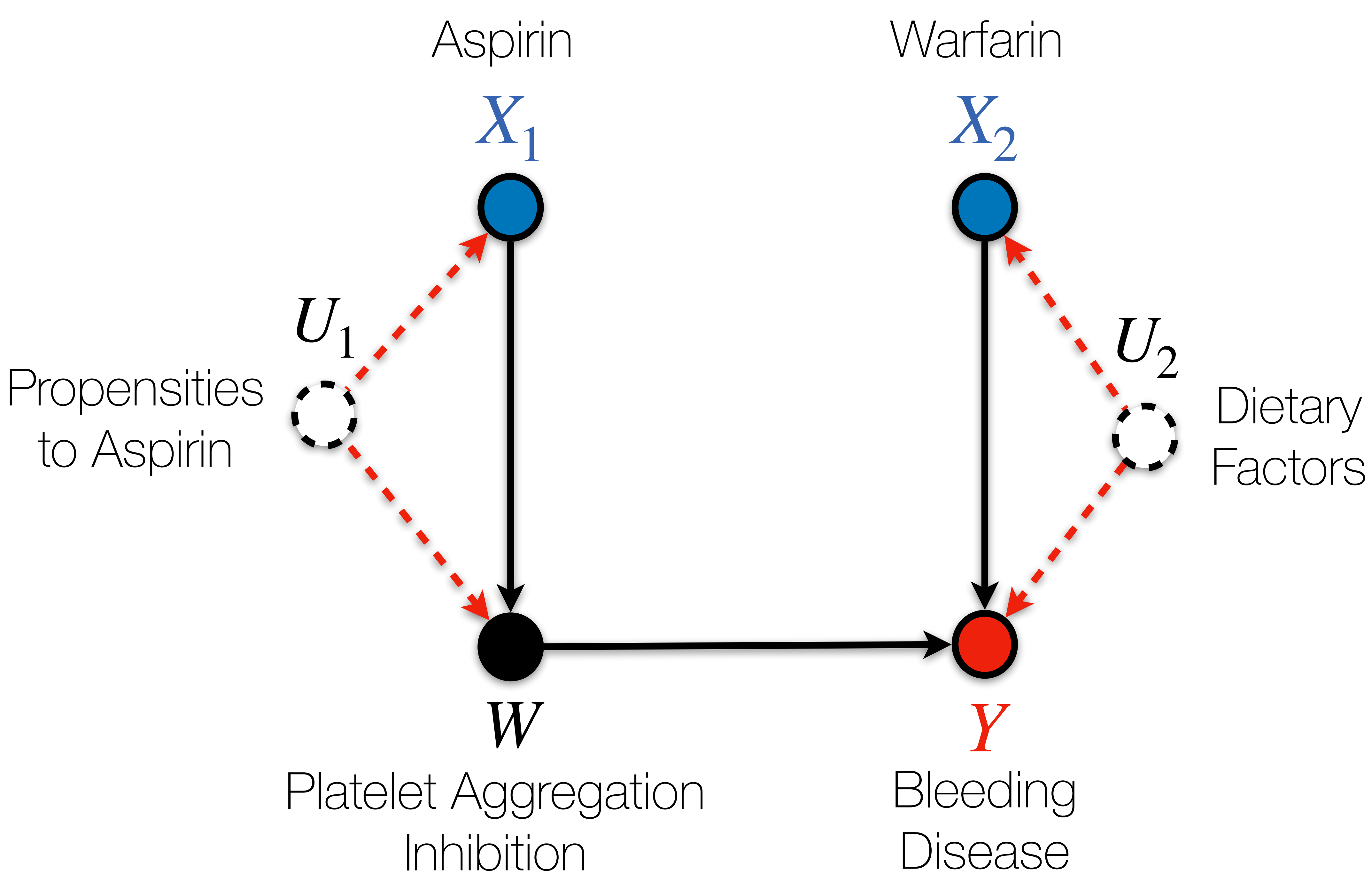
Motivation: Joint Treatment Effect Estimation



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

- BD is not applicable

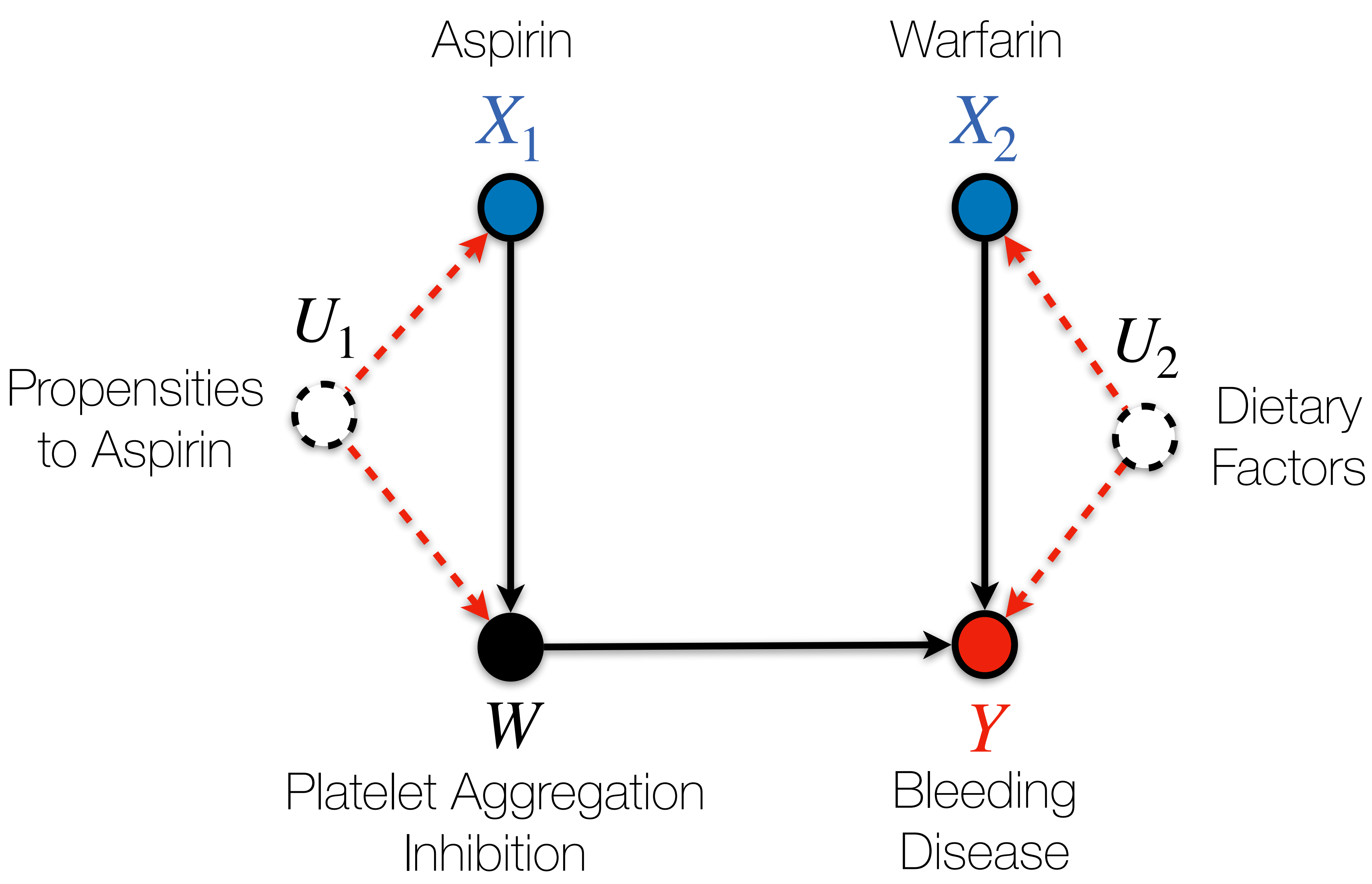
Motivation: Joint Treatment Effect Estimation



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

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

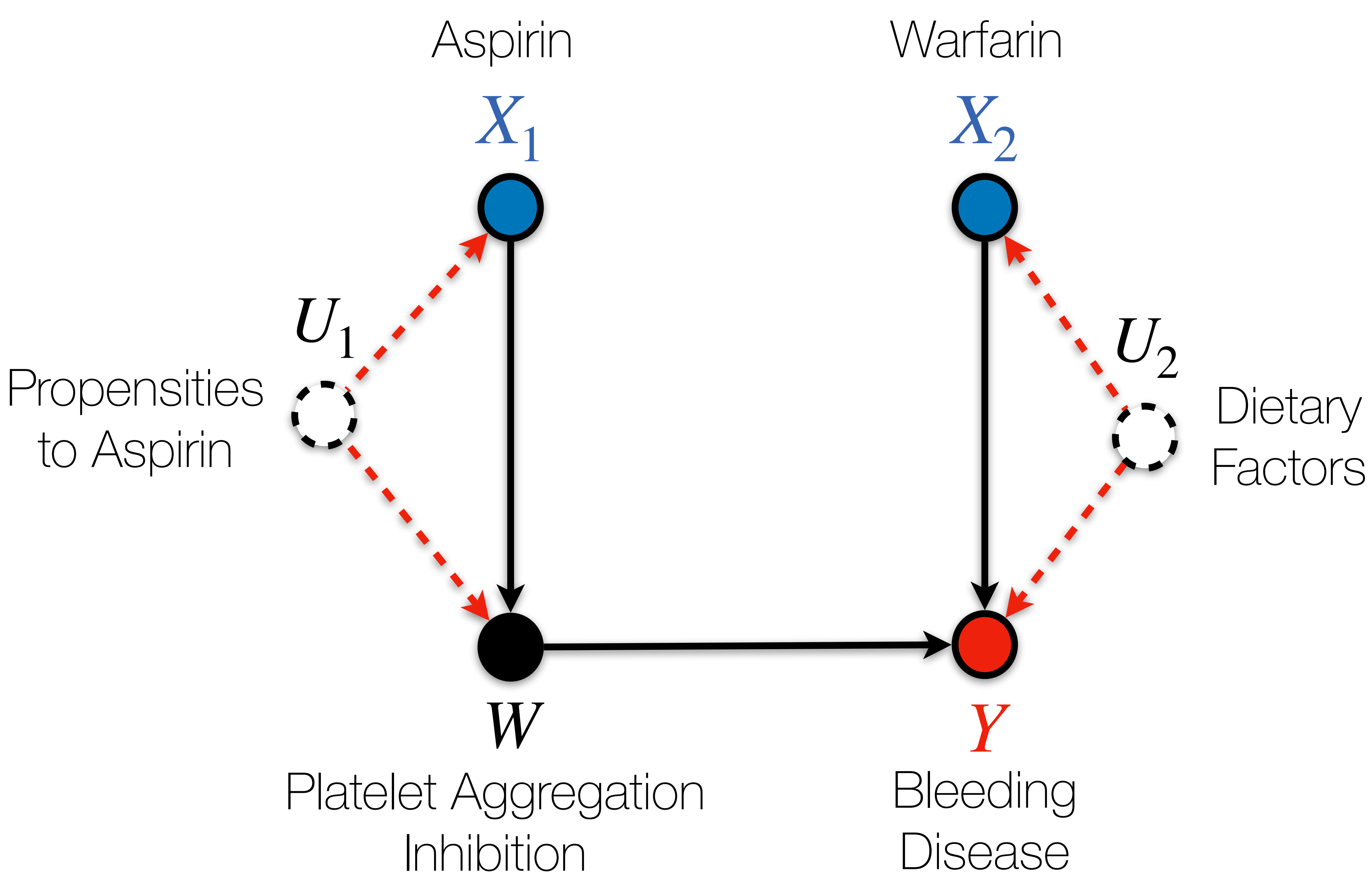
Motivation: Joint Treatment Effect Estimation



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

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

(Def. 39) BD Criterion for Joint Treatment Effect (BD^+))

A set $\mathbf{Z}$ satisfies the *BD criterion from marginal experiments* $P_{\text{do}(\mathbf{x}_1)}$ *and* $P_{\text{do}(\mathbf{x}_2)}$ 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 $\mathcal{G}$; and
2. $\mathbf{Z}$ blocks every spurious path between $\mathbf{X}_1$ and $\mathbf{Y}$ in the experiment $\text{do}(\mathbf{X}_2)$

Joint Treatment Effect Identification

(Def. 39) BD Criterion for Joint Treatment Effect (BD$^+$))

A set $\mathbf{Z}$ satisfies the *BD criterion from marginal experiments* $P_{\text{do}(\mathbf{x}_1)}$ *and* $P_{\text{do}(\mathbf{x}_2)}$ 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 $\mathcal{G}$; and
2. $\mathbf{Z}$ blocks every spurious path between $\mathbf{X}_1$ and $\mathbf{Y}$ in the experiment $\text{do}(\mathbf{X}_2)$

(Theorem 17)

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

Joint Treatment Effect Identification

(Def. 39) BD Criterion for Joint Treatment Effect (BD$^+$)

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

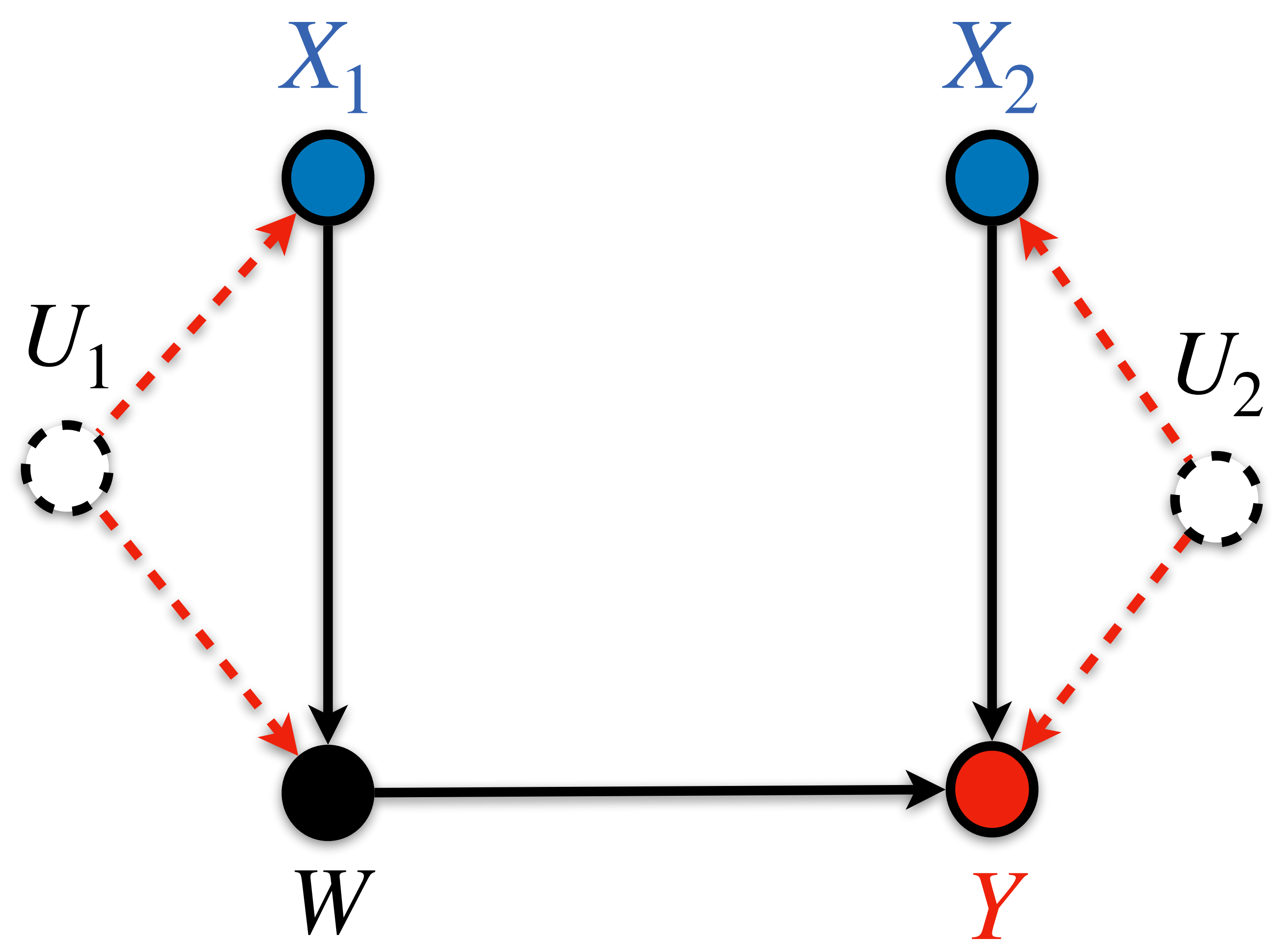
1. $\mathbf{Z}$ is not a descendent of $\mathbf{X}_2$ in $\mathcal{G}$; and
2. $\mathbf{Z}$ blocks every spurious path between $\mathbf{X}_1$ and $\mathbf{Y}$ in the experiment $\text{do}(\mathbf{X}_2)$

(Theorem 17)

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

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 $\text{do}(\mathbf{X}_2)$



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

Parametrization of BD+ (Sec. 4.2.2)

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

Parametrization of BD+ (Sec. 4.2.2)

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

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

Parametrization of BD+ (Sec. 4.2.2)

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

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

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

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

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

Parametrization of BD+ (Sec. 4.2.2)

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

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

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

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

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

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

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

Parametrization of BD+ (Sec. 4.2.2)

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

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

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

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

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

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

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

"Double Robustness"

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

Doubly Robust Estimator for BD+

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

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

Doubly Robust Estimator for BD+

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

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

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

Doubly Robust Estimator for BD+

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

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

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

$$= \mathbb{E}_{do(x_2)}[\hat{\pi}\{Y-\hat{\mu}\}] + \mathbb{E}_{do(x_1)}[\hat{\mu}(x, C)]$$

Doubly Robust Estimator for BD+

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

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

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

$$= \mathbb{E}_{do(x_2)}[\hat{\pi}\{Y-\hat{\mu}\}] + \mathbb{E}_{do(x_1)}[\hat{\mu}(x, C)]$$

DML-BD+ (Def. 46)

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

Robustness of DML-BD+

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

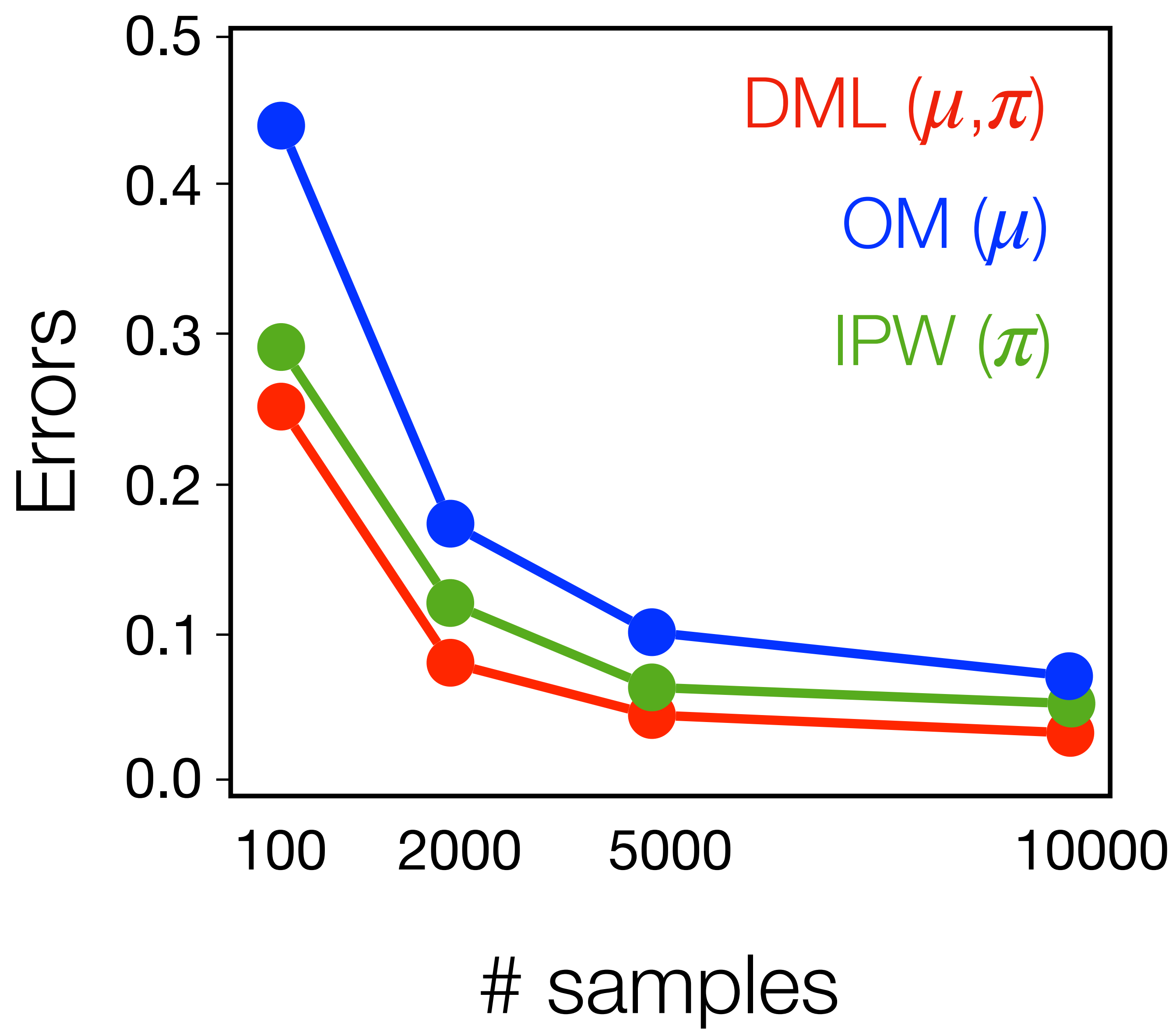
- **Double Robustness:** Error $= 0$ if either $\hat{\mu} = \mu$ or $\hat{\pi} = \pi$
- **Fast Convergence:** Error $\to 0$ *fast* even when $\hat{\mu} \to \mu$ and $\hat{\pi} \to \pi$ *slowly*.

Simulation: DML-BD+

Simulation: DML-BD+

Fast Convergence

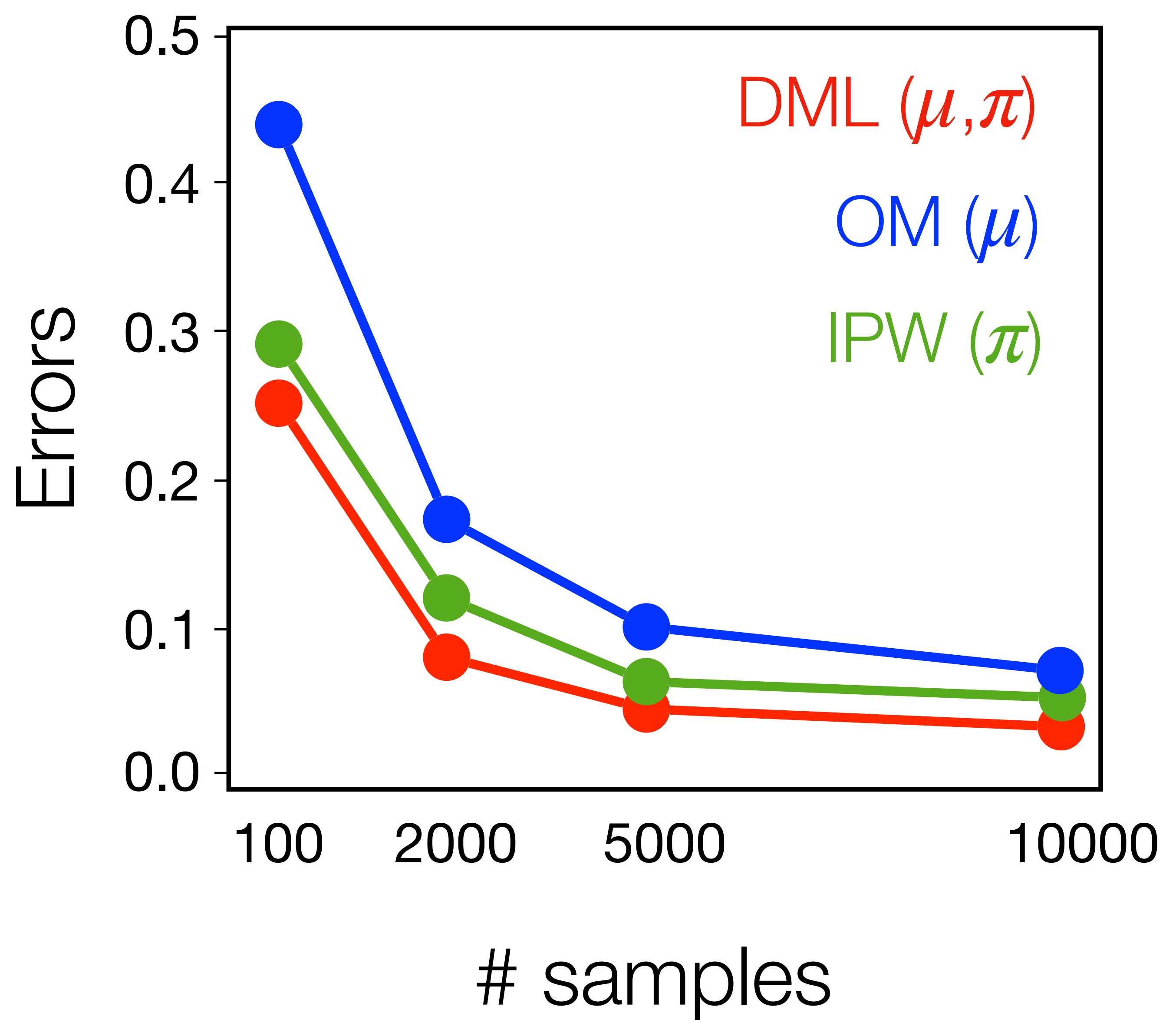
$(\hat{\boldsymbol{\mu}}, \hat{\boldsymbol{\pi}}) \rightarrow (\boldsymbol{\mu}_0, \boldsymbol{\pi}_0)$ slowly



Simulation: DML-BD+

Fast Convergence

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

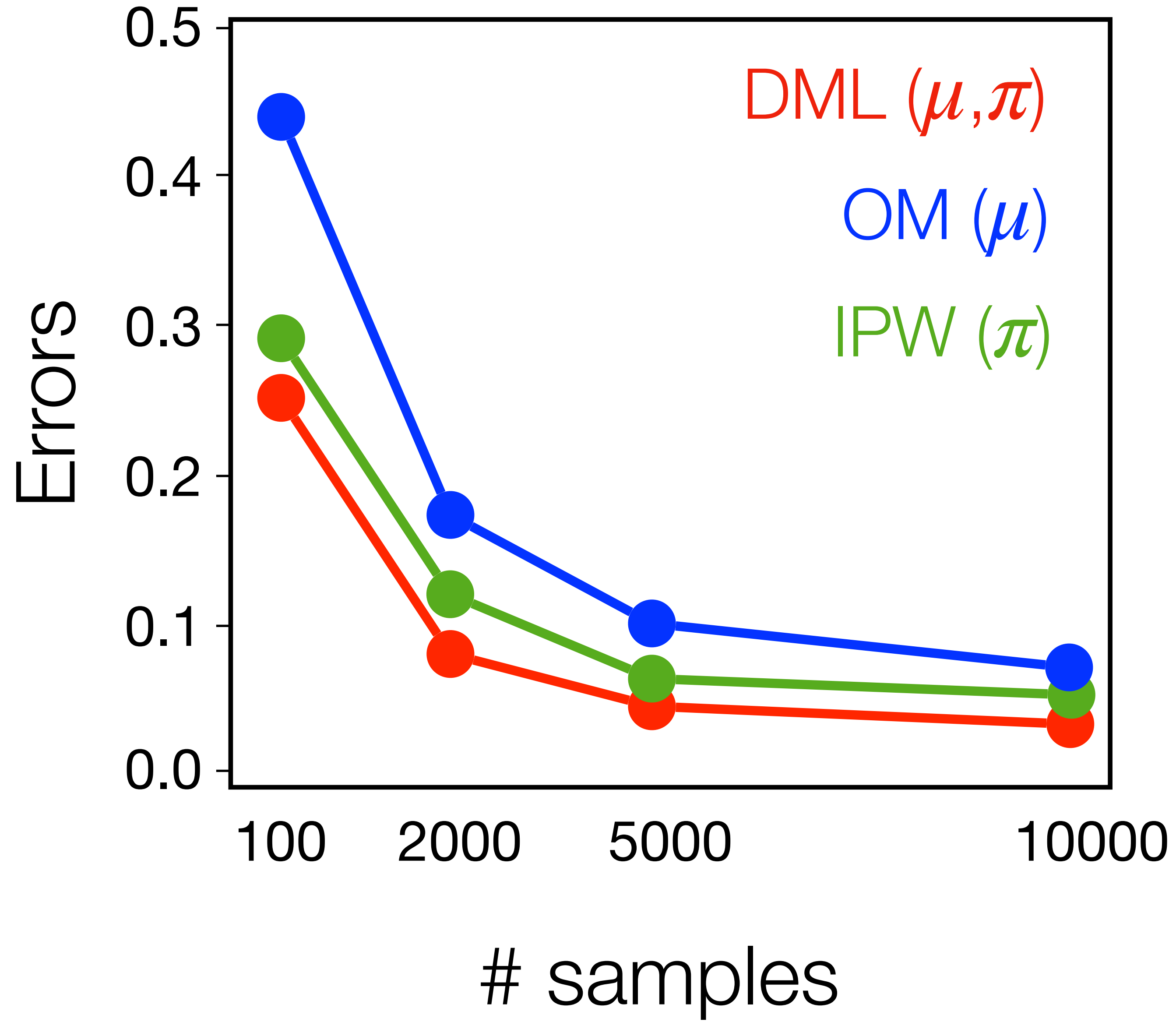


DML-BD+ converges fast, even when $(\hat{\boldsymbol{\mu}}, \hat{\boldsymbol{\pi}})$ converge slowly

Simulation: DML-BD+

Fast Convergence

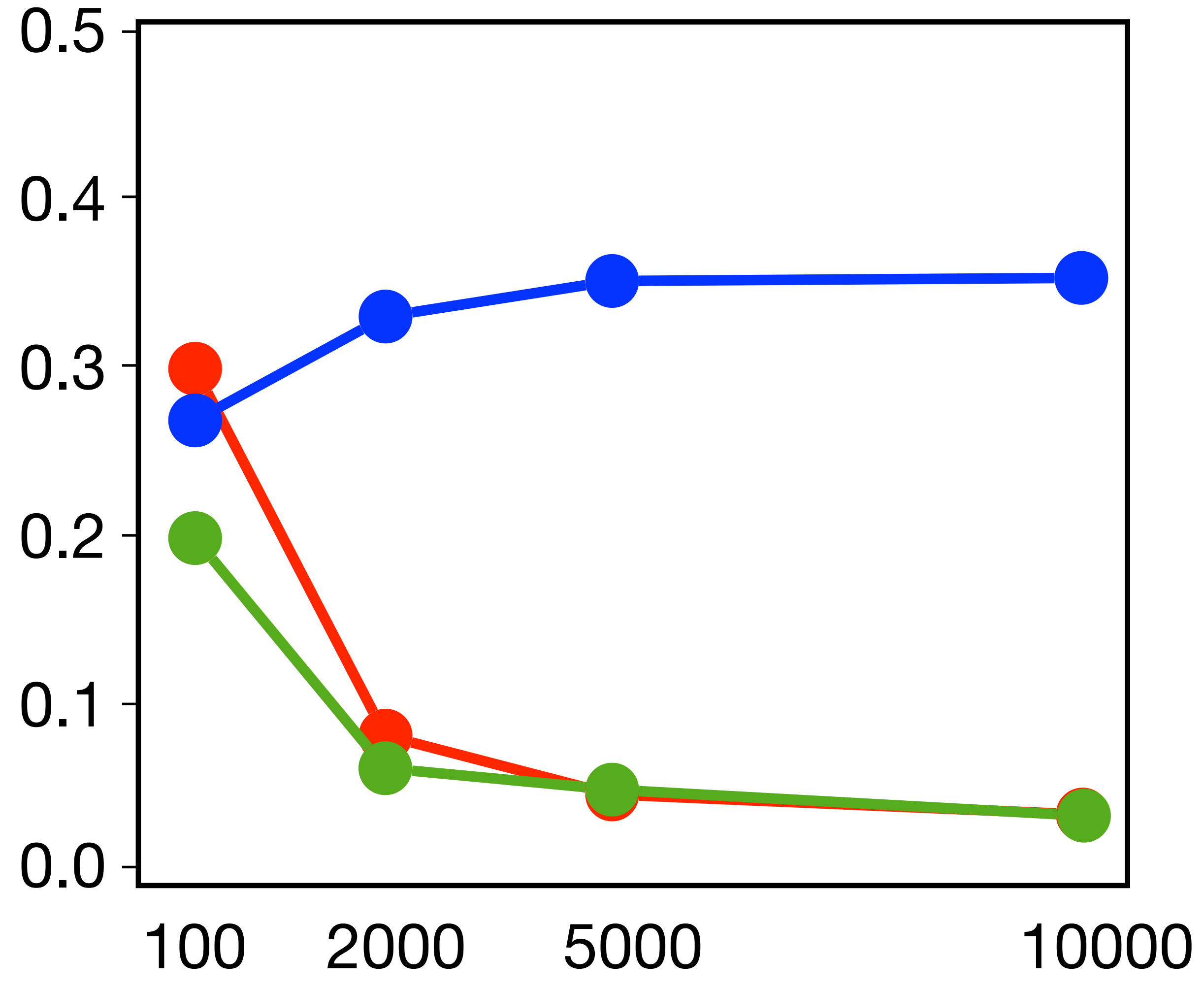
$(\hat{\mu},\hat{\pi}) \rightarrow (\mu_0, \pi_0)$ slowly



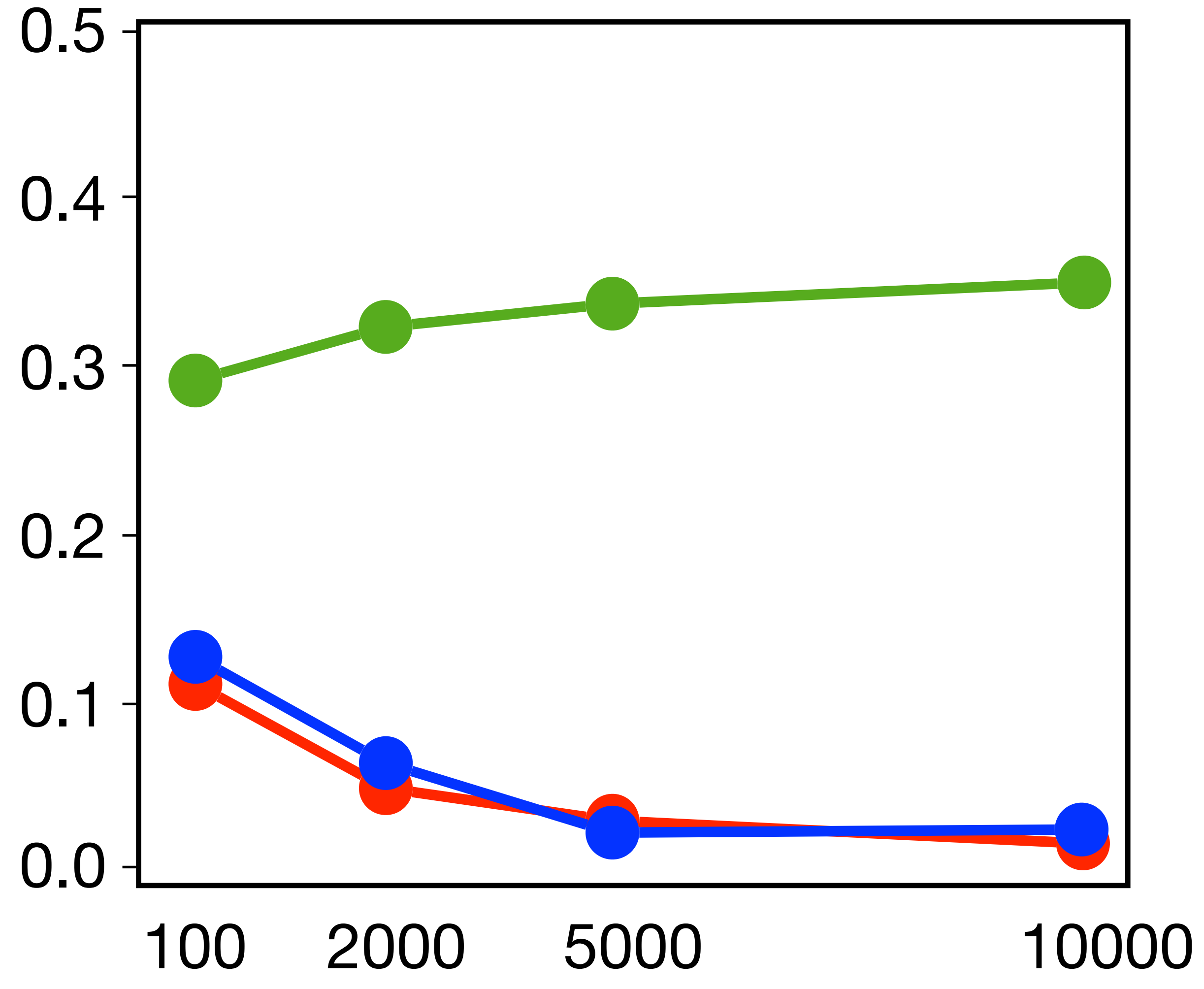
DML-BD+ converges fast, even when $(\hat{\mu},\hat{\pi})$ converge slowly

Double Robustness

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



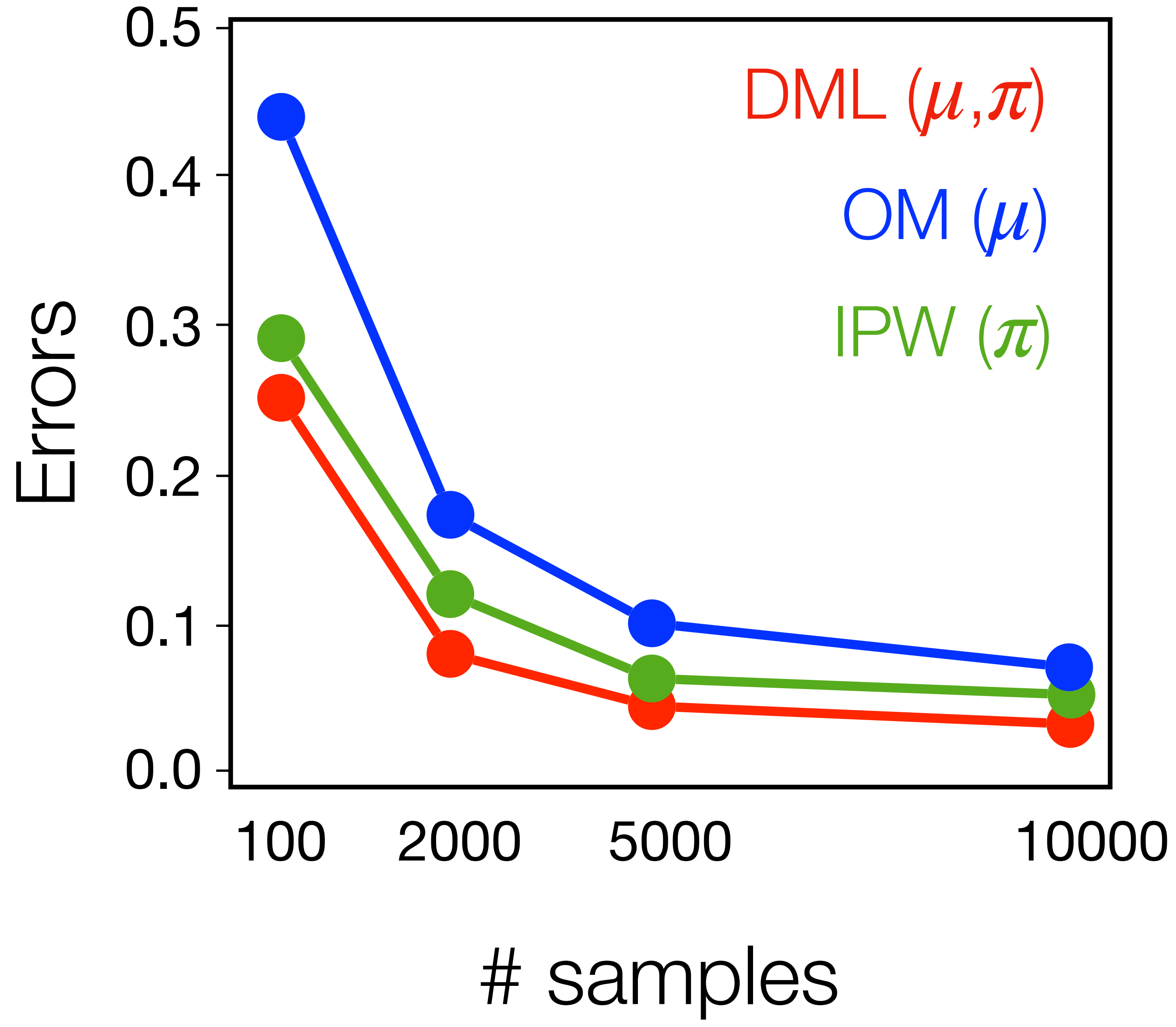
$\hat{\pi}$ misspecified ($\hat{\pi} \neq \pi$)



Simulation: DML-BD+

Fast Convergence

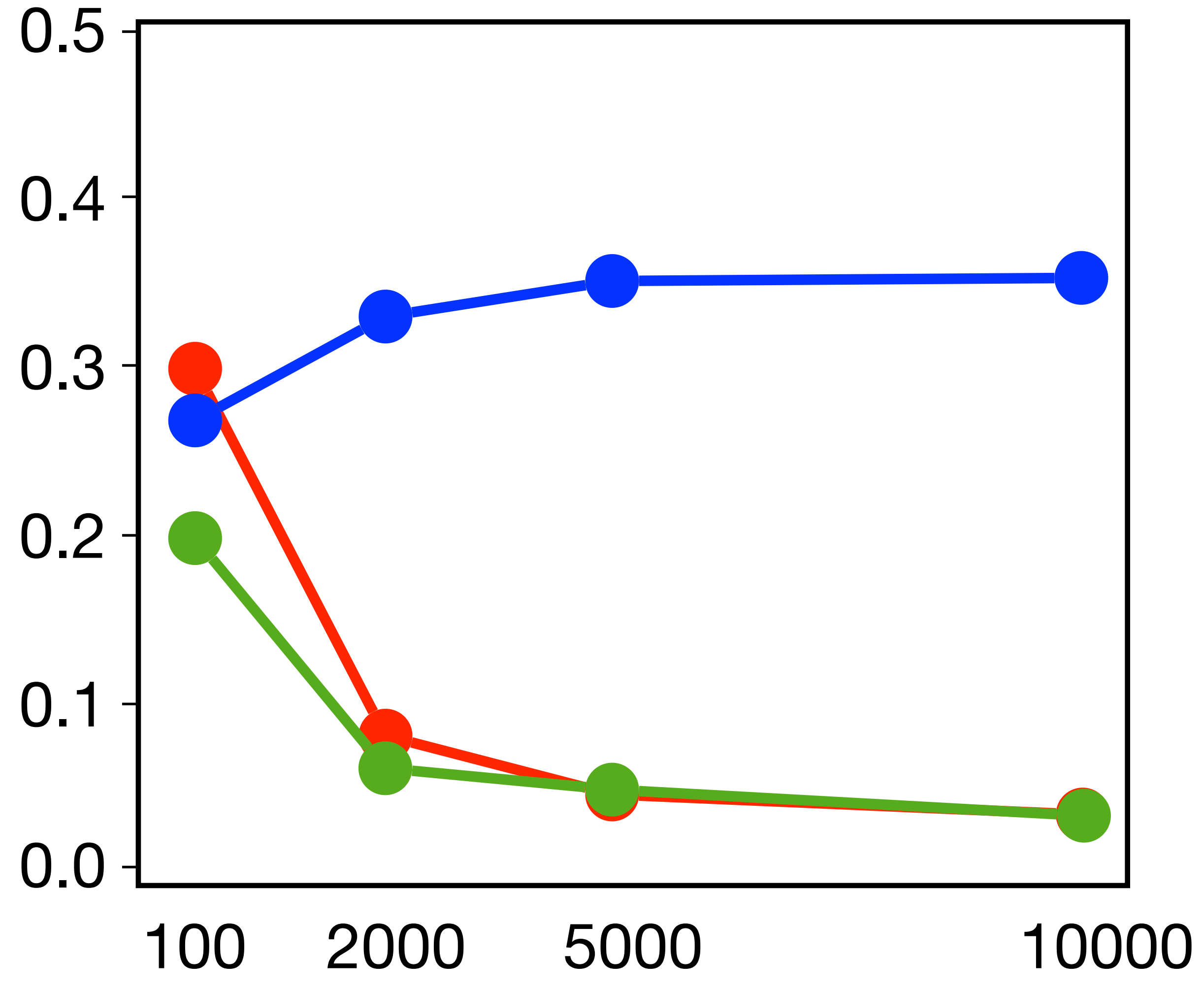
$(\hat{\boldsymbol{\mu}}, \hat{\boldsymbol{\pi}}) \to (\boldsymbol{\mu}_0, \boldsymbol{\pi}_0)$ slowly



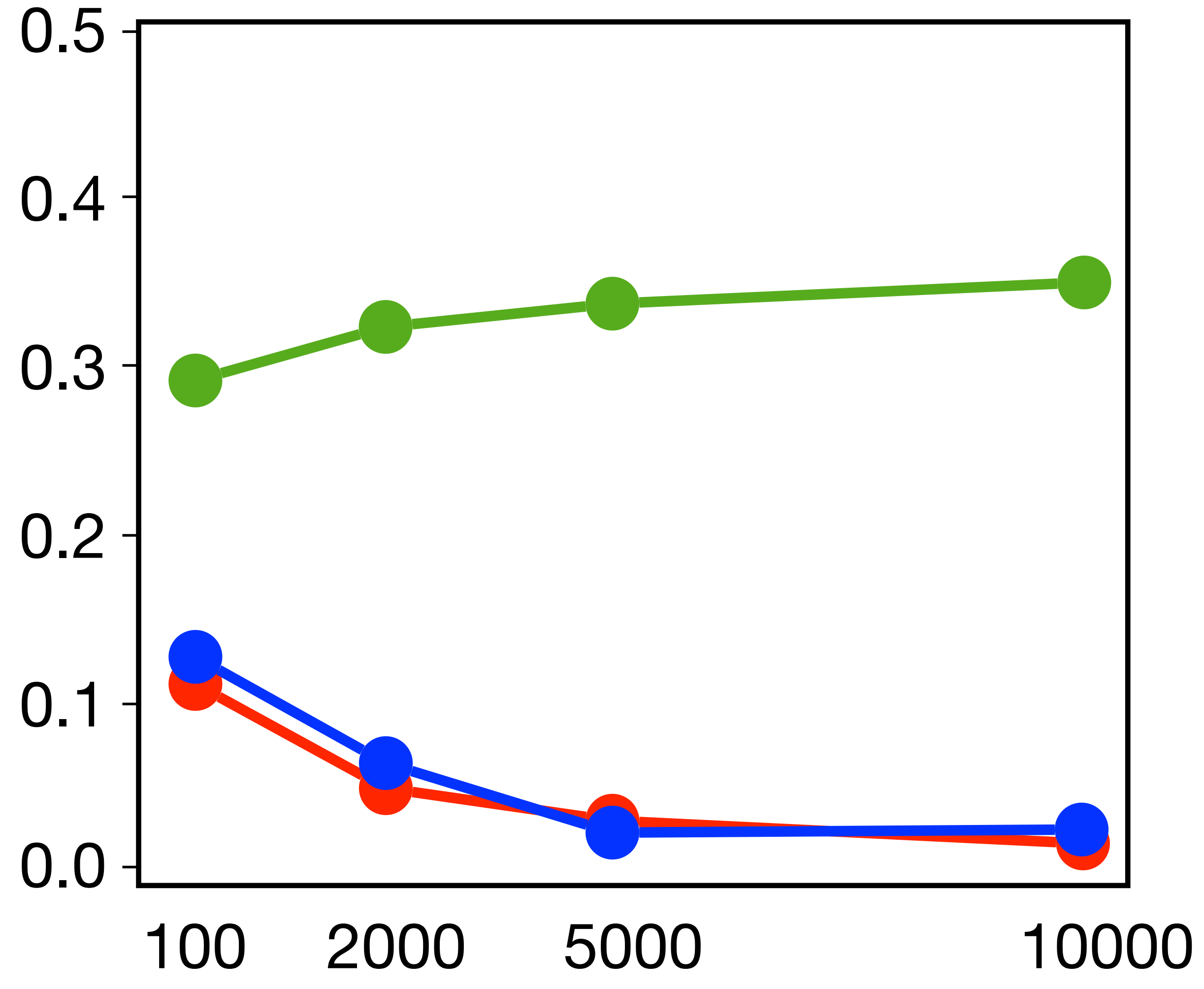
DML-BD+ converges fast, even when $(\hat{\boldsymbol{\mu}}, \hat{\boldsymbol{\pi}})$ converge slowly

Double Robustness

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



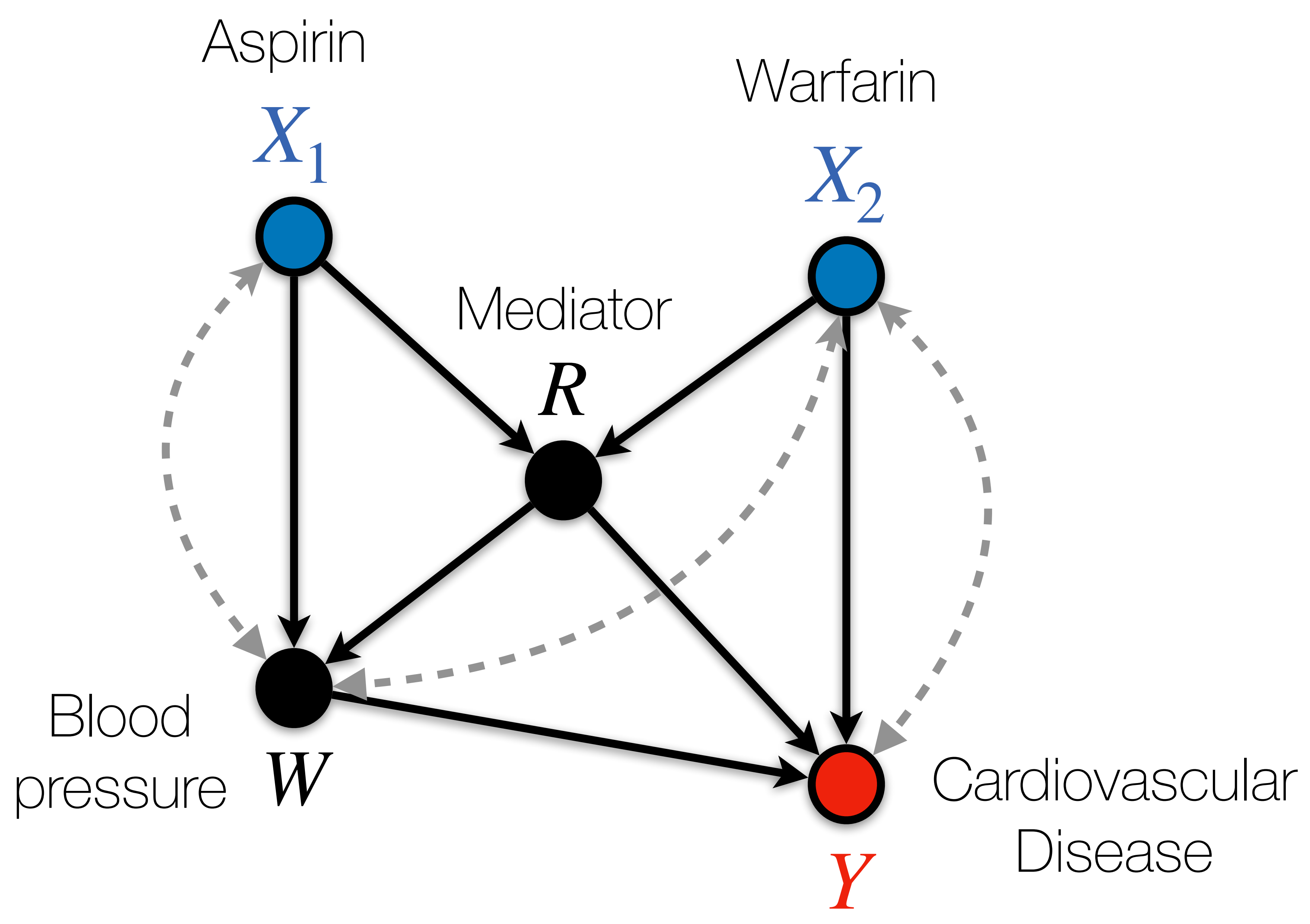
$\hat{\boldsymbol{\pi}}$ misspecified ($\hat{\boldsymbol{\pi}} \neq \boldsymbol{\pi}$)



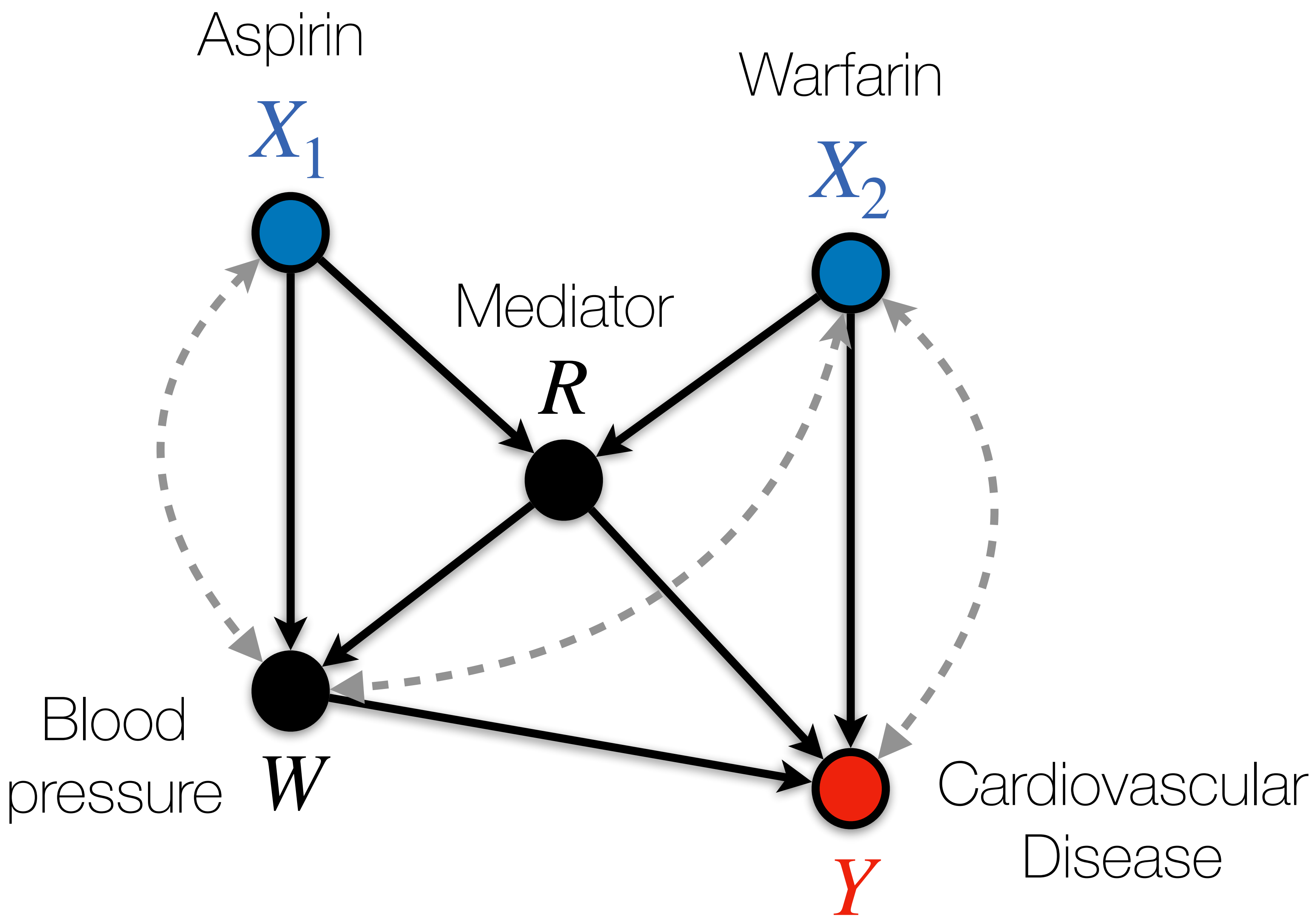
DML-BD+ converges to the true causal effect even when $\hat{\boldsymbol{\mu}}$ or $\hat{\boldsymbol{\pi}}$ are misspecified.

Example where BD^+ Fails

Example where BD+ Fails

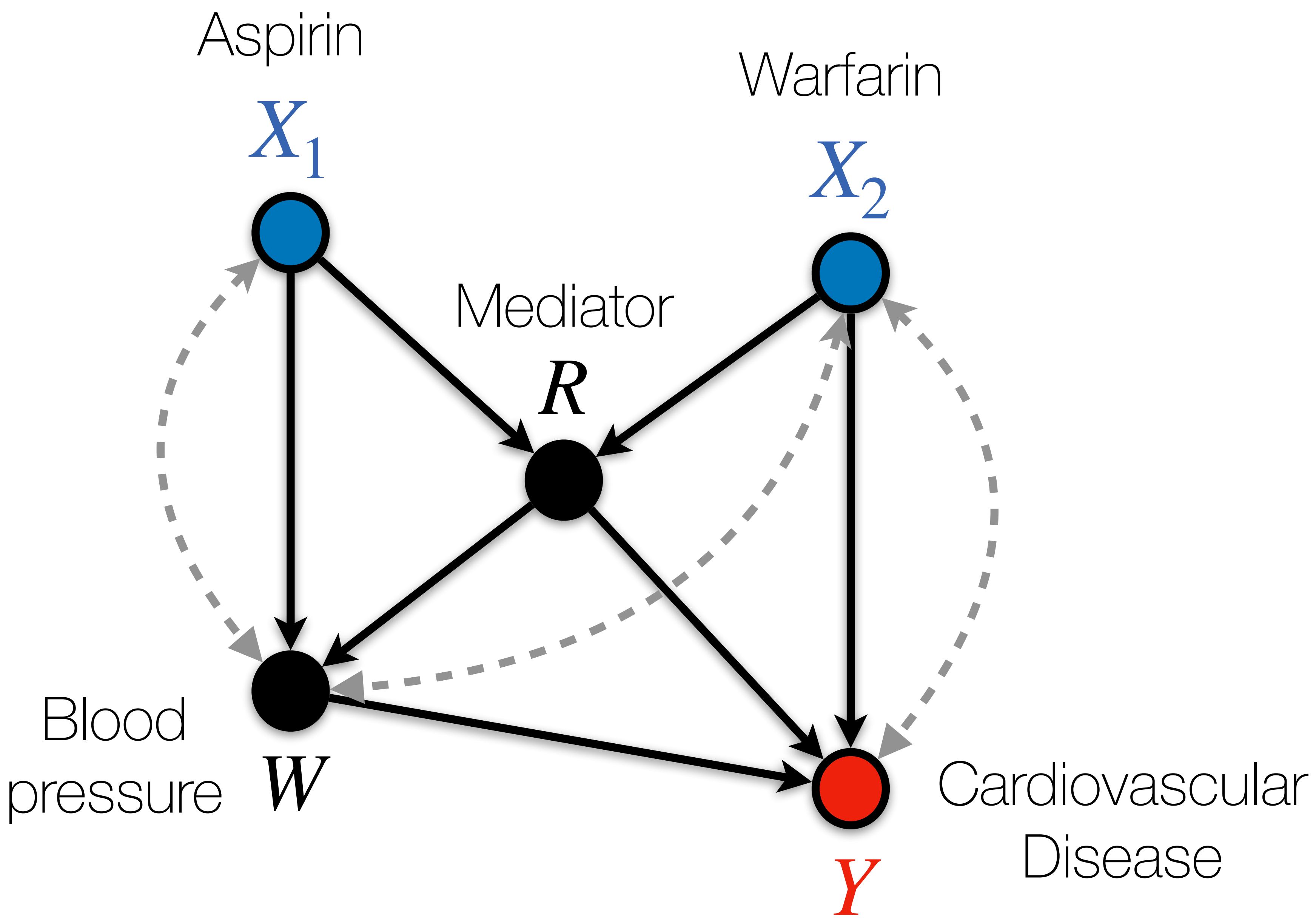


Example where BD+ Fails



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

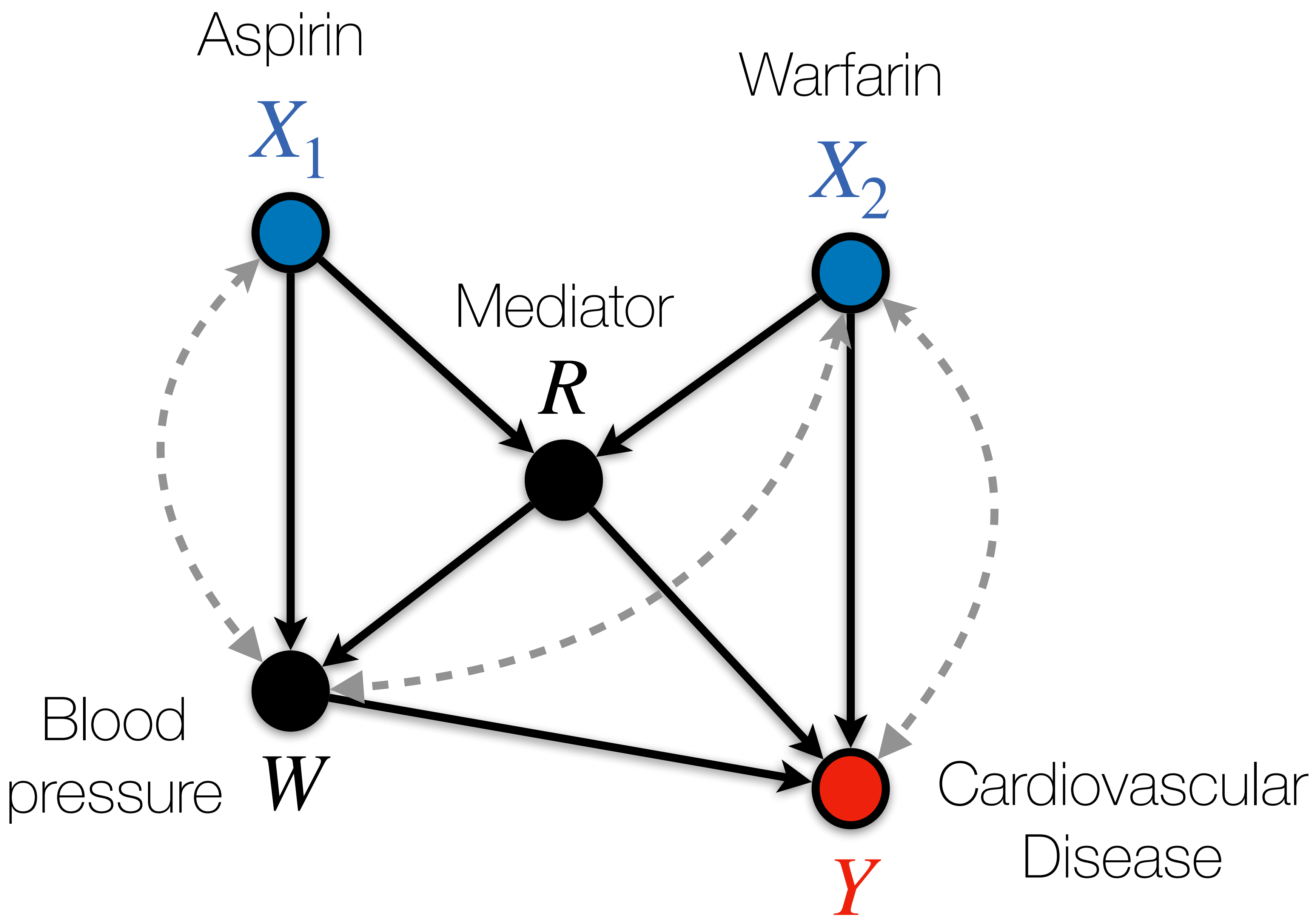
Example where BD+ Fails



✗ BD+ fails

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

Example where BD+ Fails



X BD+ fails

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

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

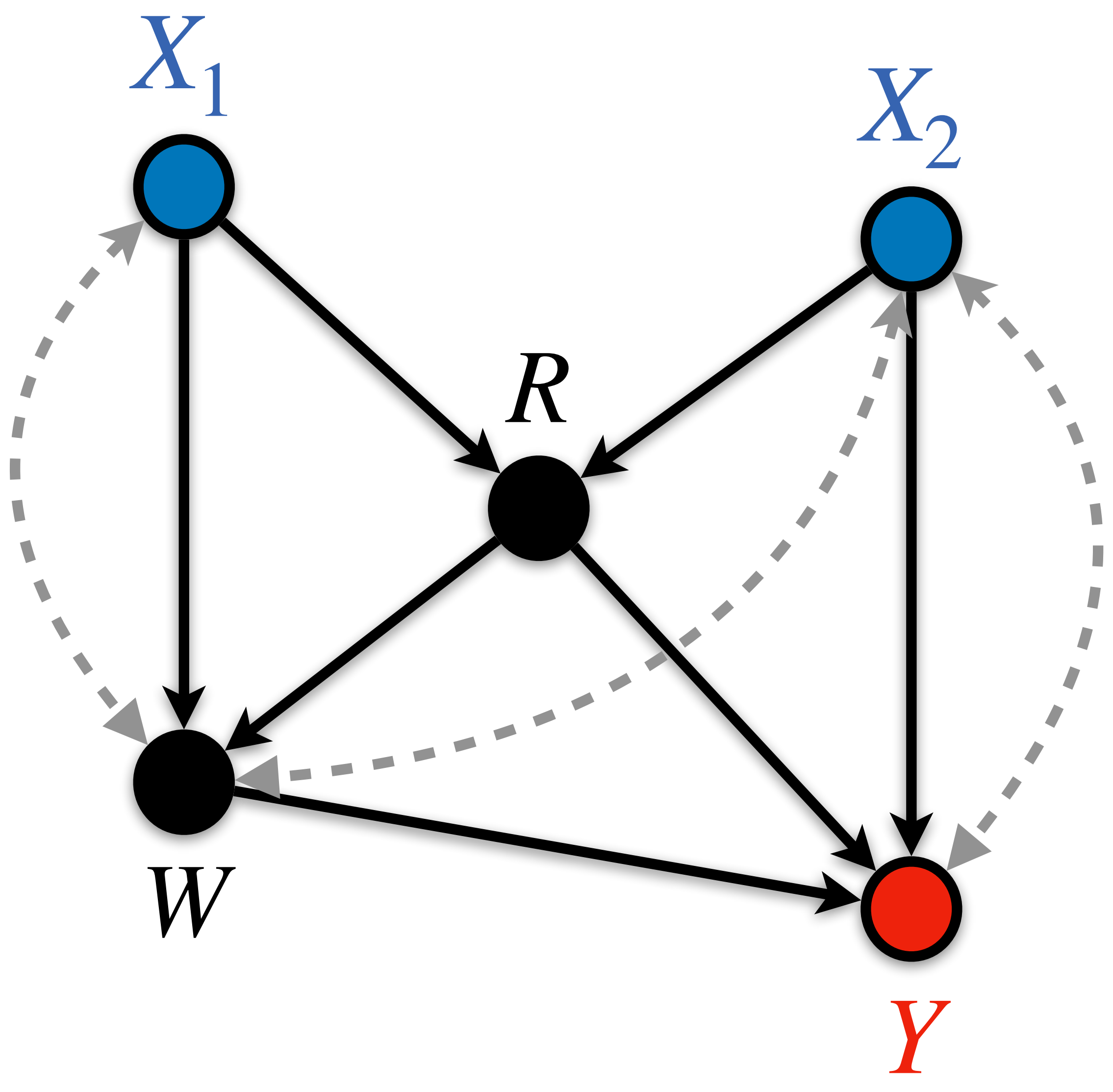
Background: General Identification from Data Fusion

General Identification (gID, Algo 5)

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, Algo 5)

Bareinboim and Pearl, 2012; Lee et al. 2019

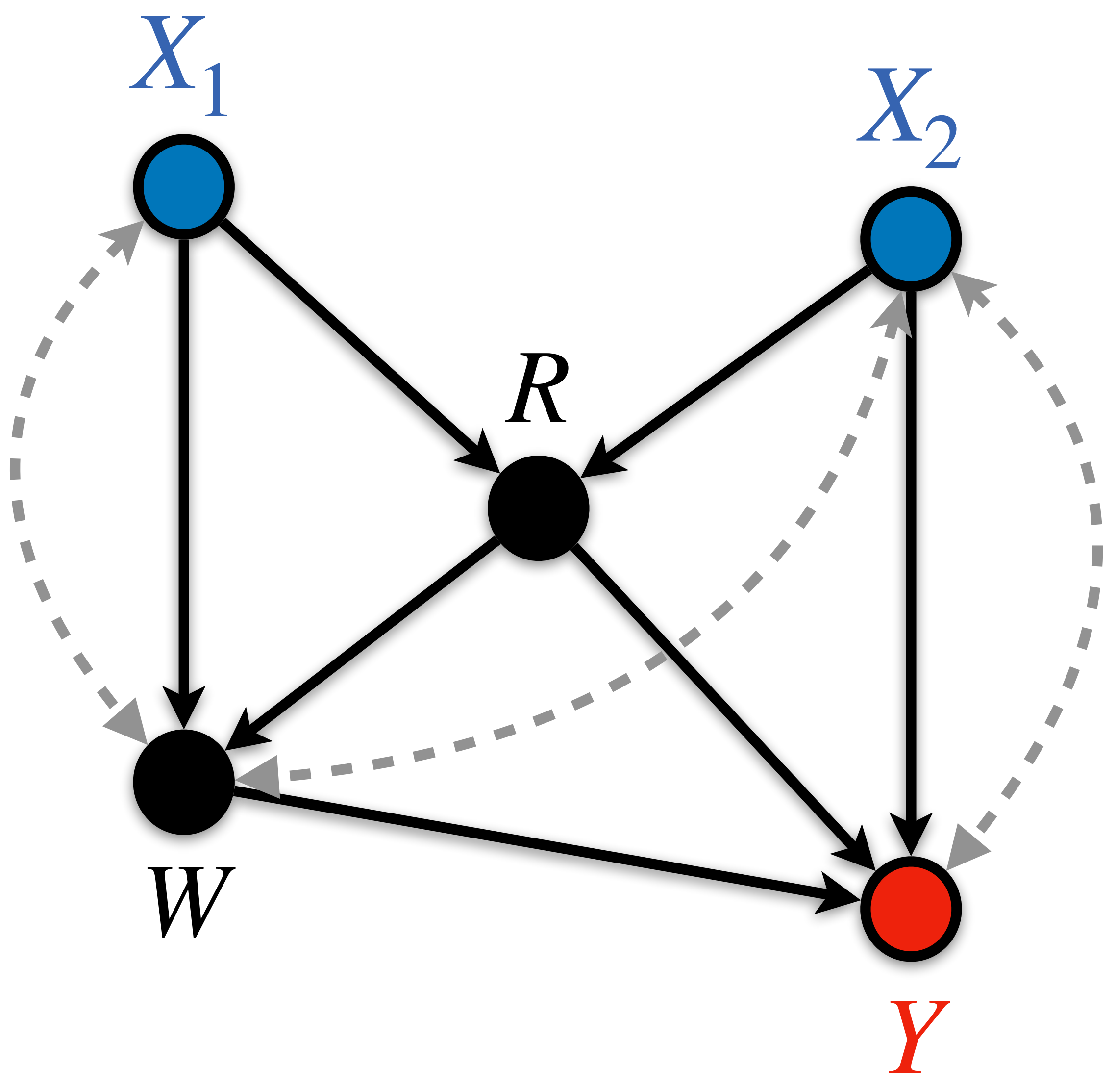
- spanning a *tree* from available distributions $\{P_{\text{do}(\mathbf{r}_i)}(\mathbf{V})\}_{\mathbf{R}_i \subseteq \mathbf{V}}$
- to reach to causal distribution $P(\mathbf{Y} \mid \text{do}(\mathbf{X}))$
- through factorization & marginalization of distributions

Available distributions

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

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

Background: General Identification from Data Fusion



General Identification (gID, Algo 5)

Bareinboim and Pearl, 2012; Lee et al. 2019

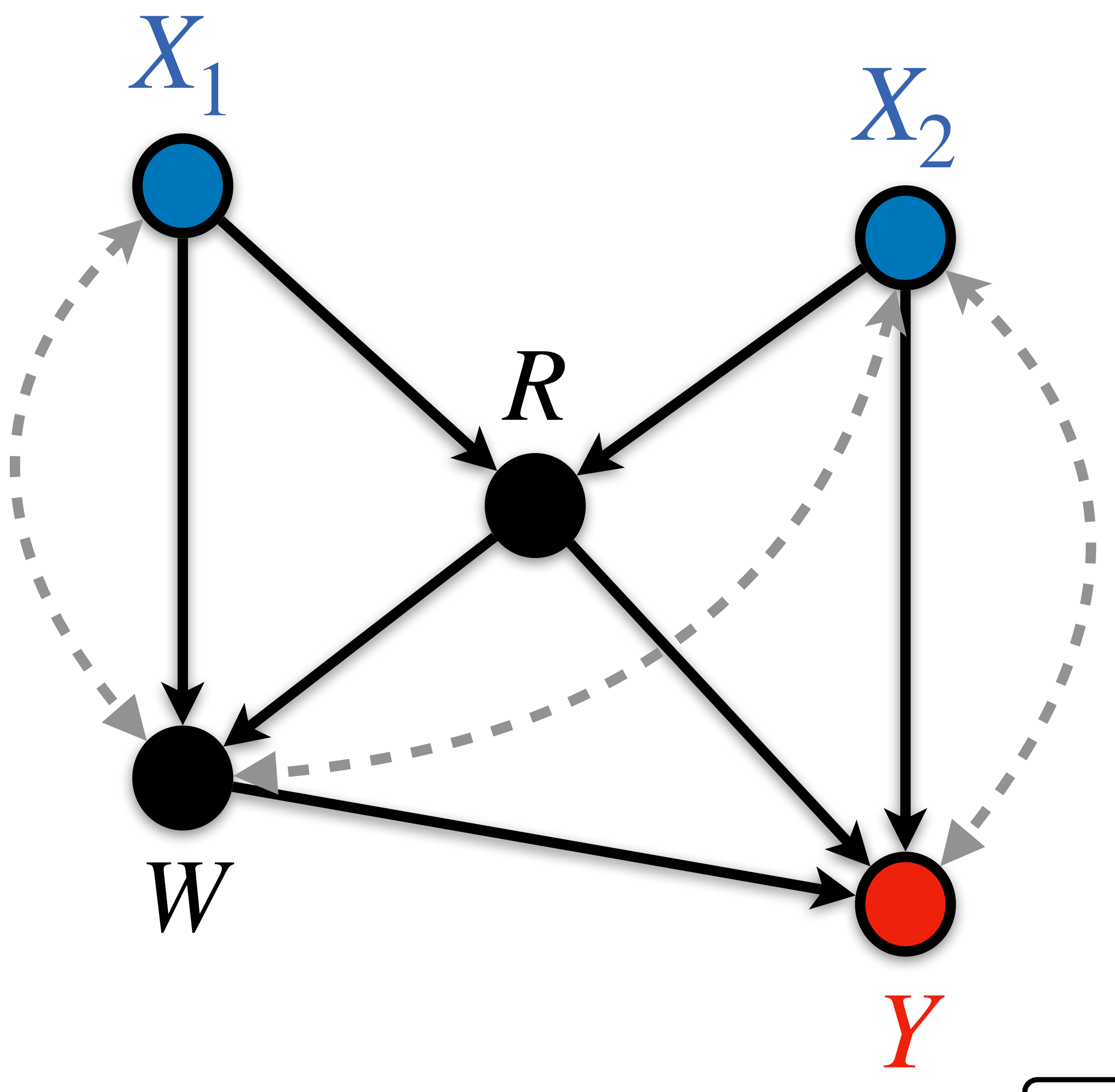
- spanning a *tree* from available distributions $\{P_{\mathrm{do}(\mathbf{r}_i)}(\mathbf{V})\}_{\mathbf{R}_i \subseteq \mathbf{V}}$
- to reach to causal distribution $P(\mathbf{Y} \mid \mathrm{do}(\mathbf{X}))$
- through factorization & marginalization of distributions

$P_{\mathrm{do}(x_1)}(RWX_2Y)$ —Factorization→ $P_{\mathrm{do}(x_1,r)}(WX_2Y)$ —Marginalization→ $P_{\mathrm{do}(x_1,r)}(W)$

$P_{\mathrm{do}(x_1)}(RWX_2Y)$ —Factorization→ $P_{\mathrm{do}(x_1,x_2)}(R)$

$P_{\mathrm{do}(x_2)}(RWX_1Y)$ —Factorization→ $P_{\mathrm{do}(w,r,x_2)}(Y)$

Background: General Identification from Data Fusion



General Identification (gID, Algo 5)

Bareinboim and Pearl, 2012; Lee et al. 2019

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

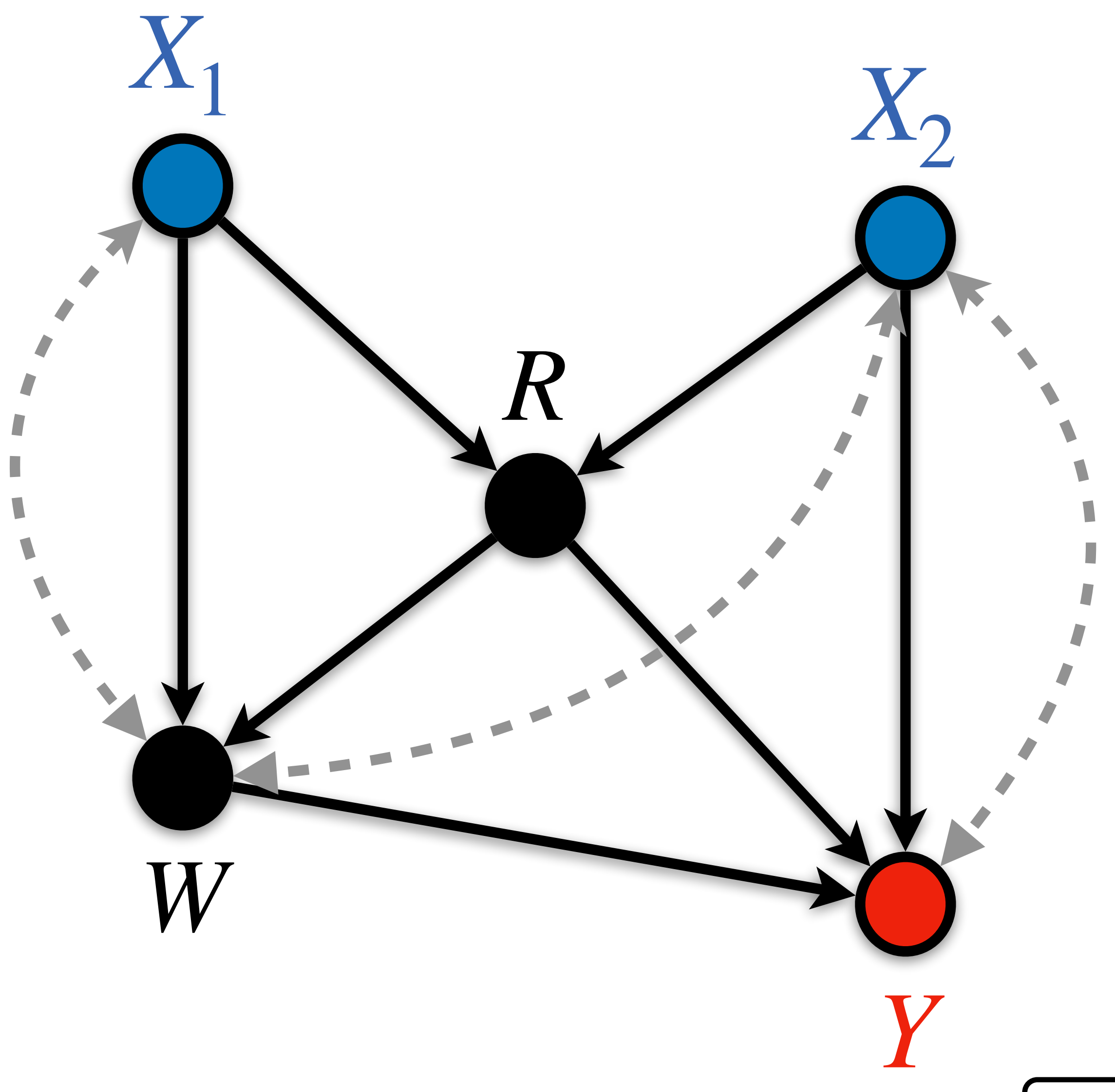
$P_{do(x_1)}(RWX_2Y)$ —Factorization→ $P_{do(x_1,r)}(WX_2Y)$ [$P_{do(x_1)}(y \mid wrx_2)P_{do(x_1)}(w \mid rx_2)P_{do(x_1)}(x_2)$]

$P_{do(x_1,r)}(WX_2Y)$ —Marginalization→ $P_{do(x_1,r)}(W)$ [$\sum_{x_2'} P_{do(x_1)}(w \mid r, x_2')P_{do(x_1)}(x_2')$]

$P_{do(x_1)}(RWX_2Y)$ —Factorization→ $P_{do(x_1,x_2)}(R)$ [$P_{do(x_1)}(r \mid x_2)$]

$P_{do(x_2)}(RWX_1Y)$ —Factorization→ $P_{do(w,r,x_2)}(Y)$ [$P_{do(x_2)}(y \mid rwx_1)$]

Background: General Identification from Data Fusion



General Identification (gID, Algo 5)

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

$P_{\mathrm{do}(x_1)}(RWX_2Y)$ —Factorization→ $P_{\mathrm{do}(x_1,r)}(WX_2Y)$

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

$P_{\mathrm{do}(x_1,r)}(WX_2Y)$ —Marginalization→ $P_{\mathrm{do}(x_1,r)}(W)$

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

$P_{\mathrm{do}(x_1)}(RWX_2Y)$ —Factorization→ $P_{\mathrm{do}(x_1,x_2)}(R)$

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

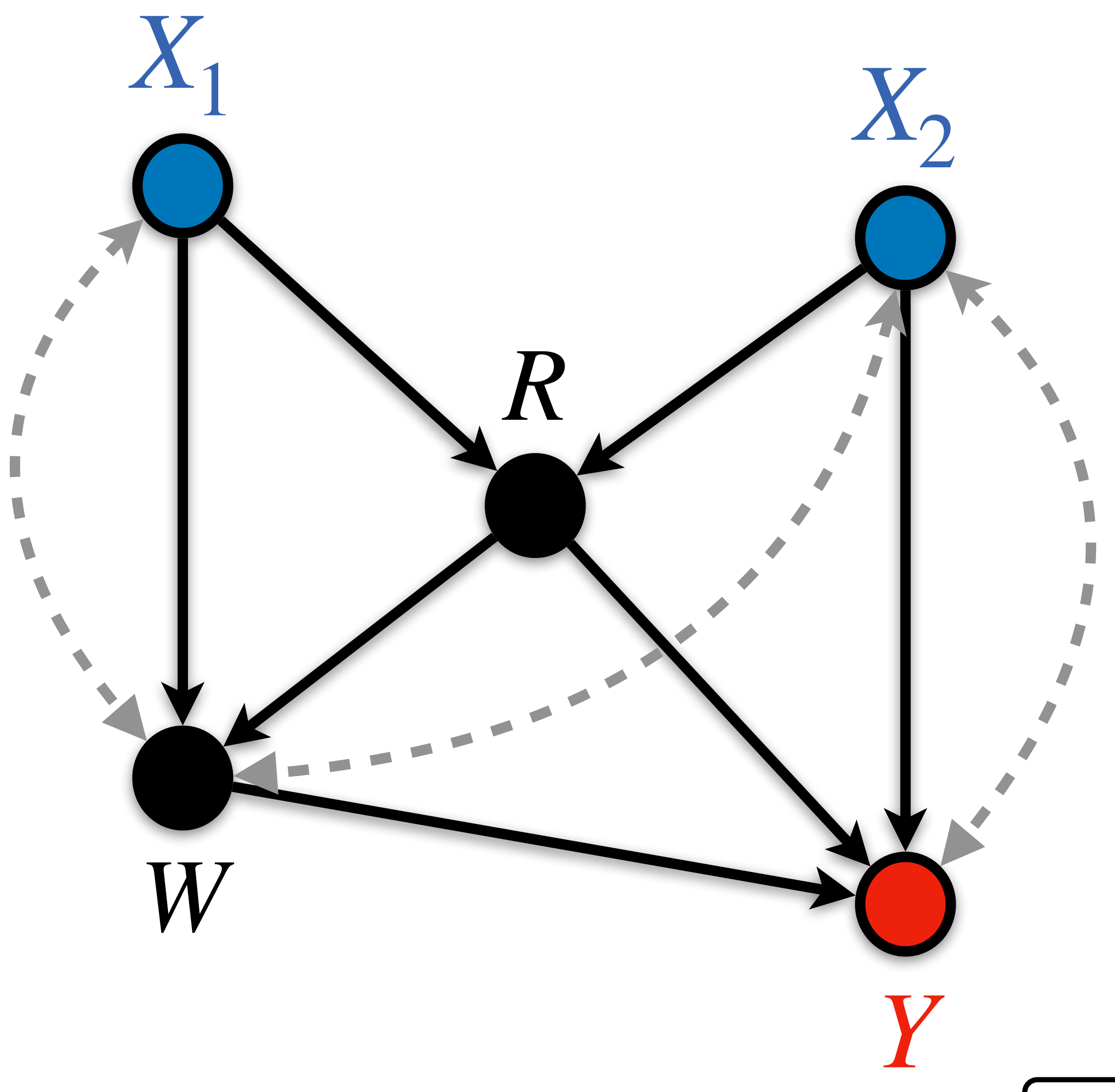
$P_{\mathrm{do}(x_2)}(RWX_1Y)$ —Factorization→ $P_{\mathrm{do}(w,r,x_2)}(Y)$

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

$P_{\mathrm{do}(x_1,r)}(W)$ —Factorization→ $P_{\mathrm{do}(x_1,x_2)}(RWY)$; $P_{\mathrm{do}(x_1,x_2)}(R)$ → $P_{\mathrm{do}(x_1,x_2)}(RWY)$; $P_{\mathrm{do}(w,r,x_2)}(Y)$ → $P_{\mathrm{do}(x_1,x_2)}(RWY)$

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

Background: General Identification from Data Fusion

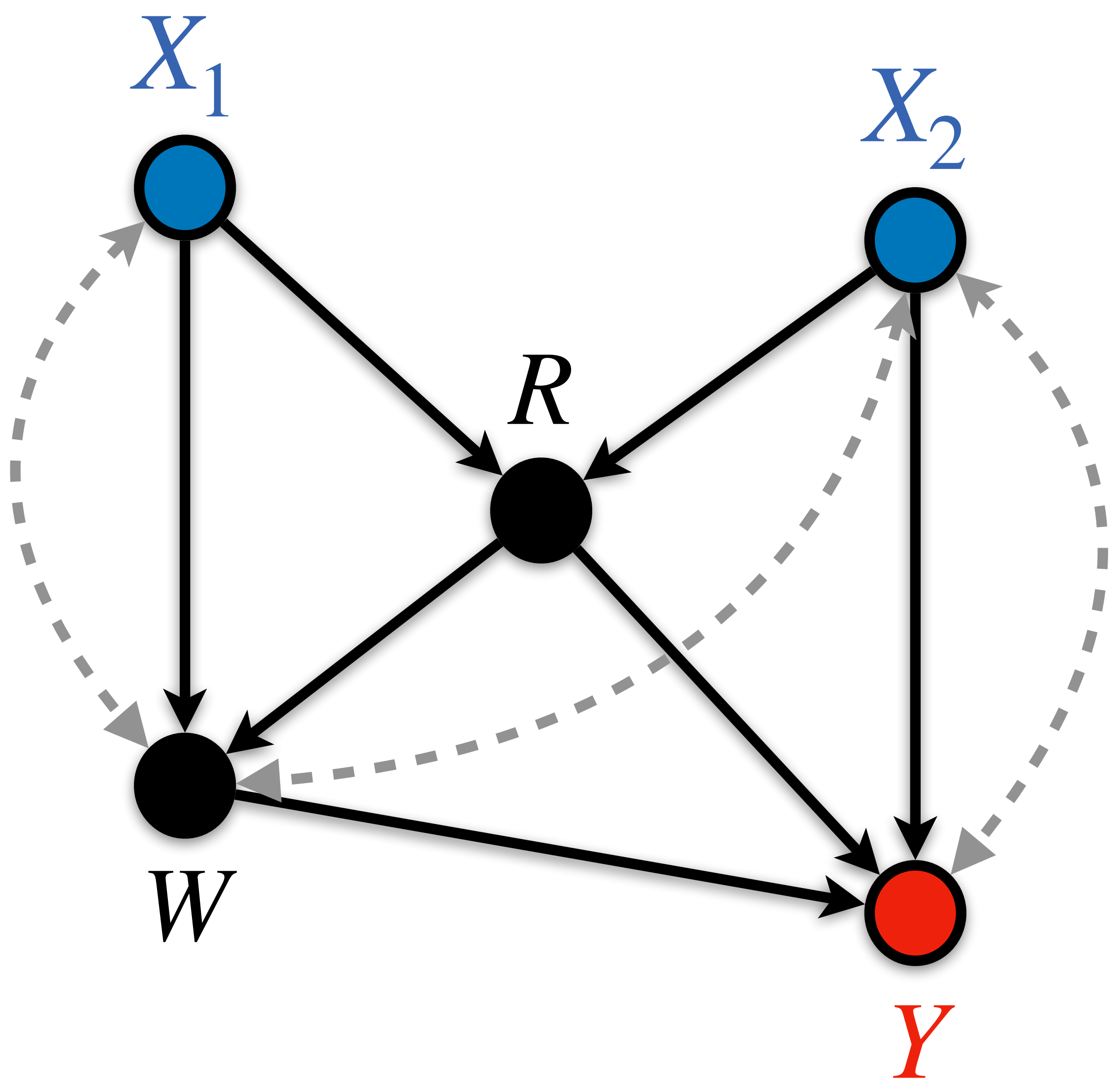


General Identification (gID, Algo 5)

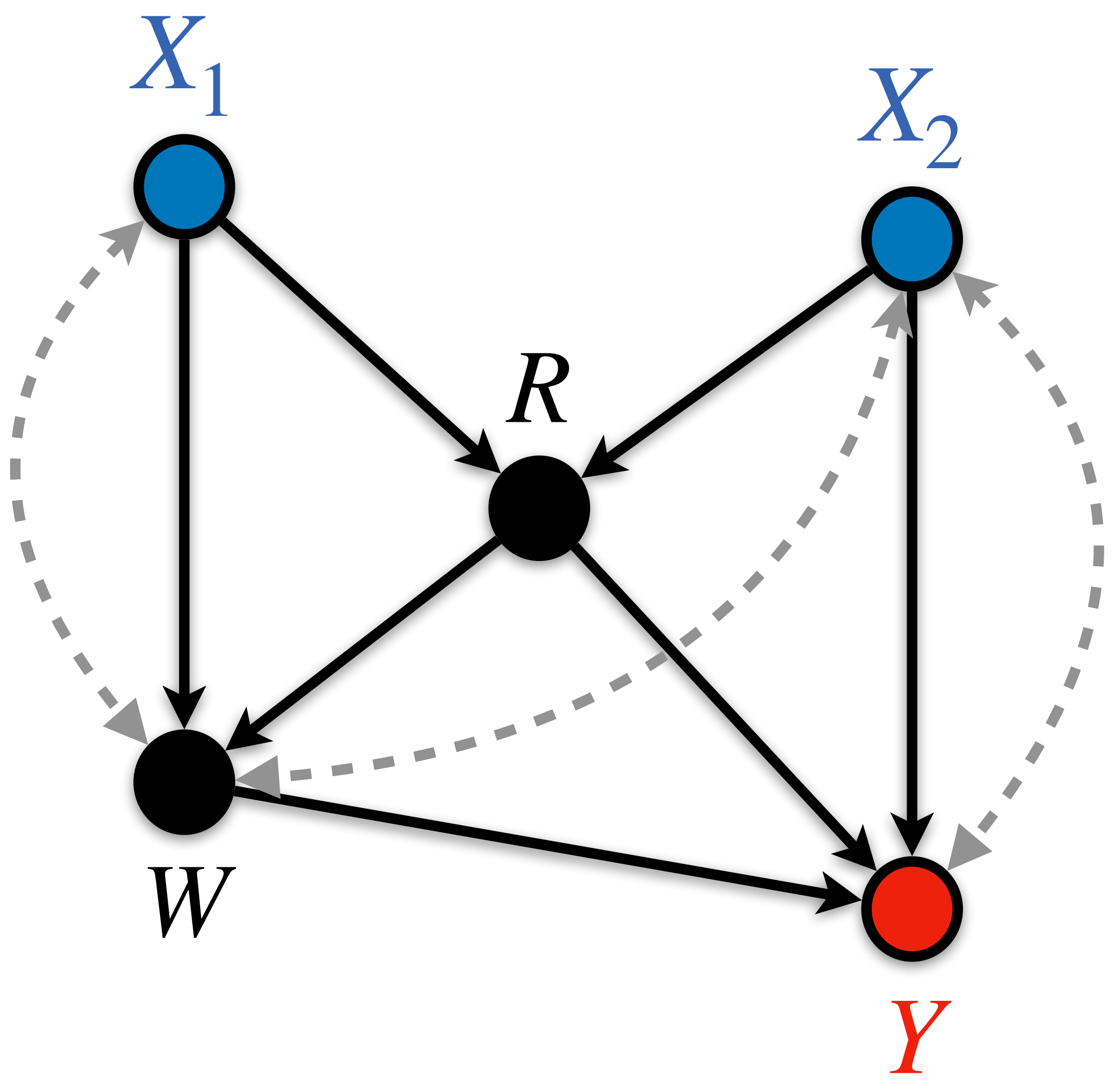
Bareinboim and Pearl, 2012; Lee et al. 2019

- 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

Causal effects as a function of BD⁺



Causal effects as a function of BD+



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

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

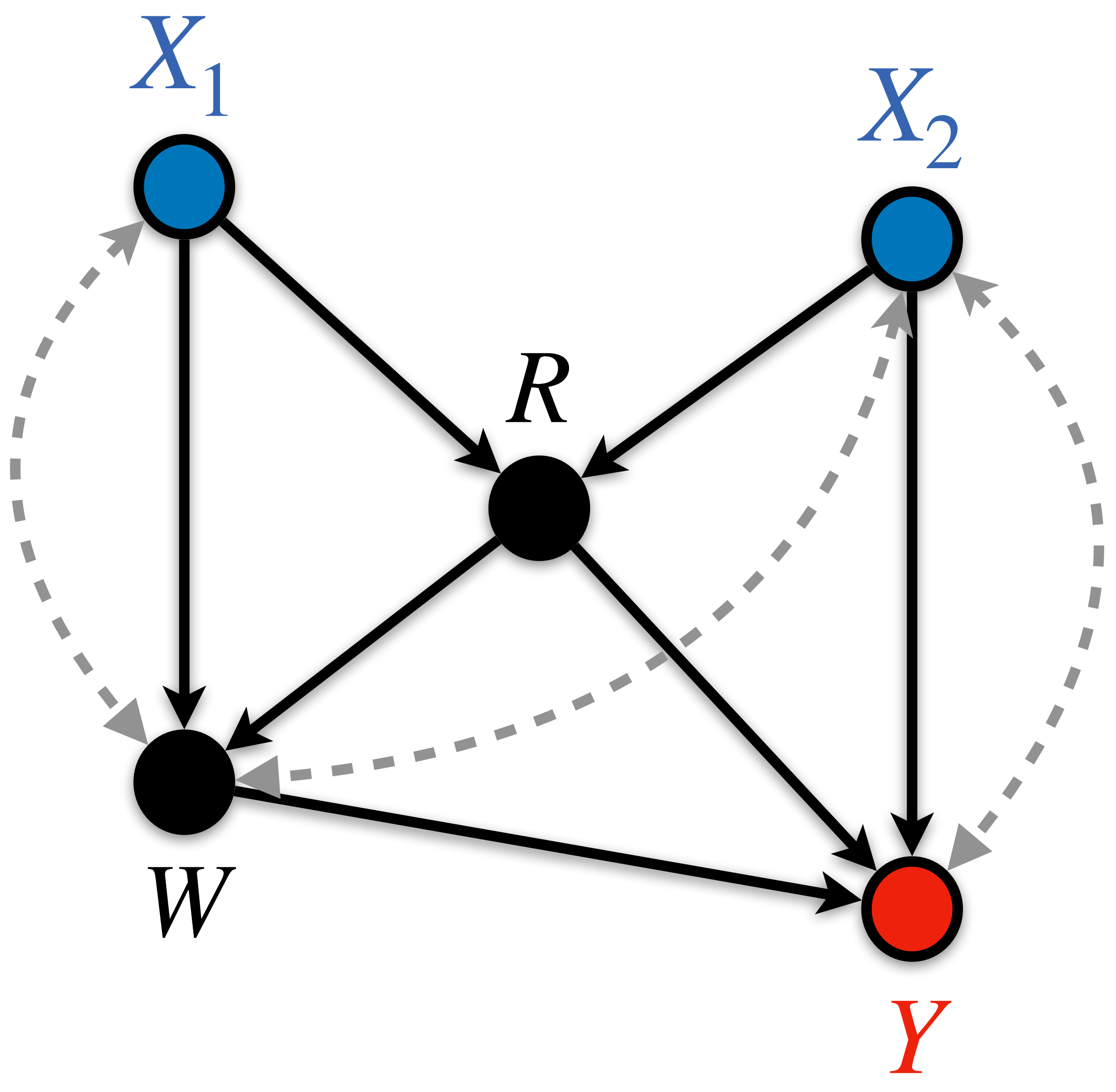
$P_{\text{do}(x_2)}(RWX_1Y) \rightarrow P_{\text{do}(w,r,x_2)}(Y)$

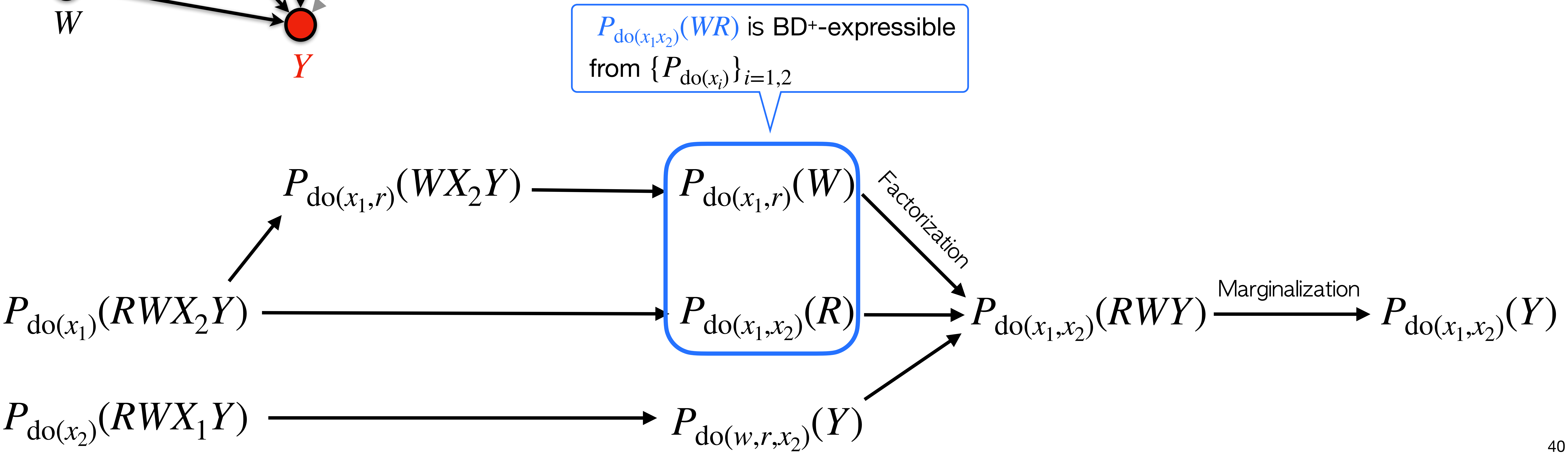
$P_{\text{do}(x_1)}(RWX_2Y) \rightarrow P_{\text{do}(x_1,r)}(WX_2Y)$

[Factorization] $P_{\text{do}(x_1,r)}(W),\ P_{\text{do}(x_1,x_2)}(R),\ P_{\text{do}(w,r,x_2)}(Y) \rightarrow P_{\text{do}(x_1,x_2)}(RWY)$

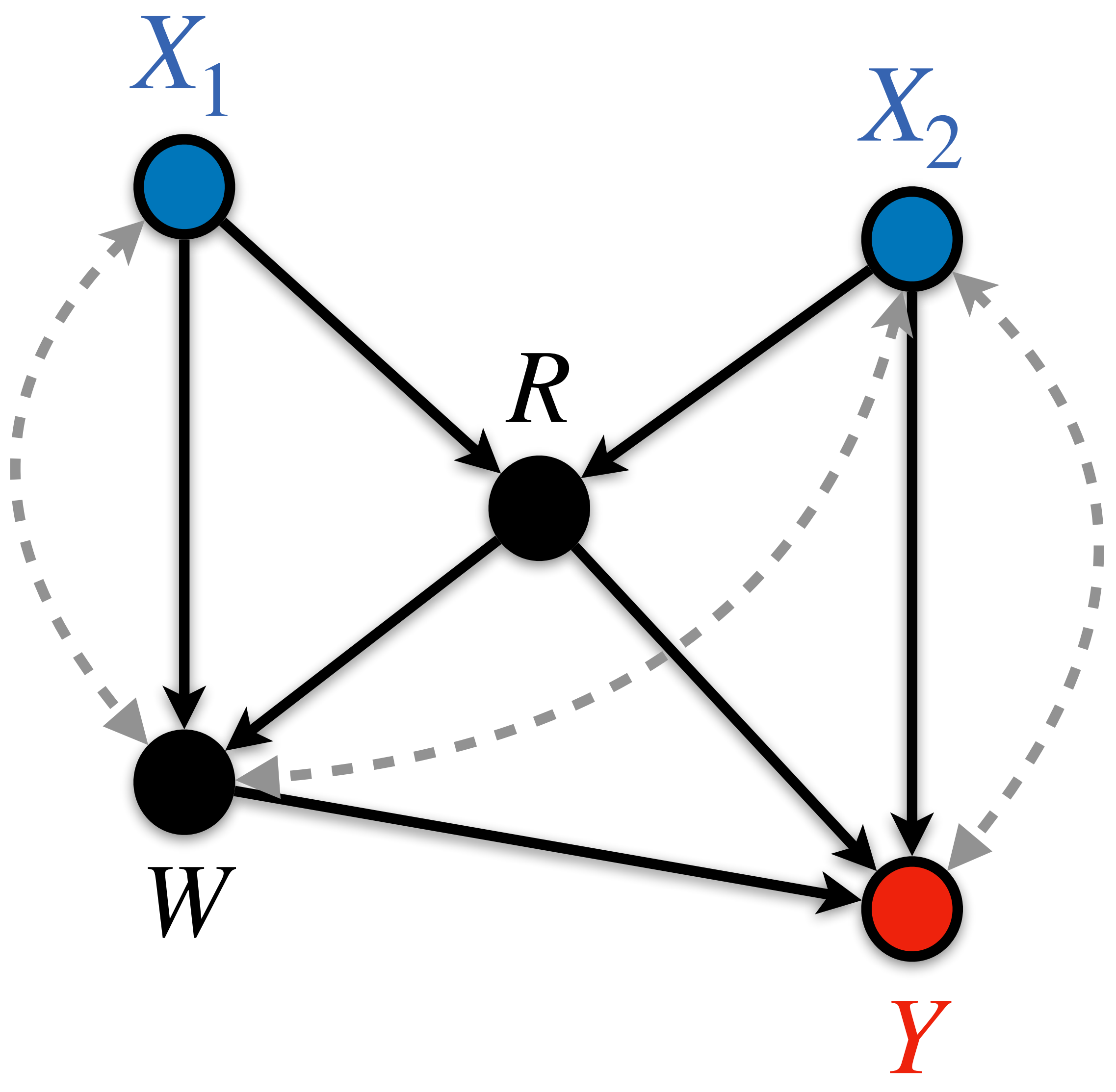
[Marginalization] $P_{\text{do}(x_1,x_2)}(RWY) \rightarrow P_{\text{do}(x_1,x_2)}(Y)$

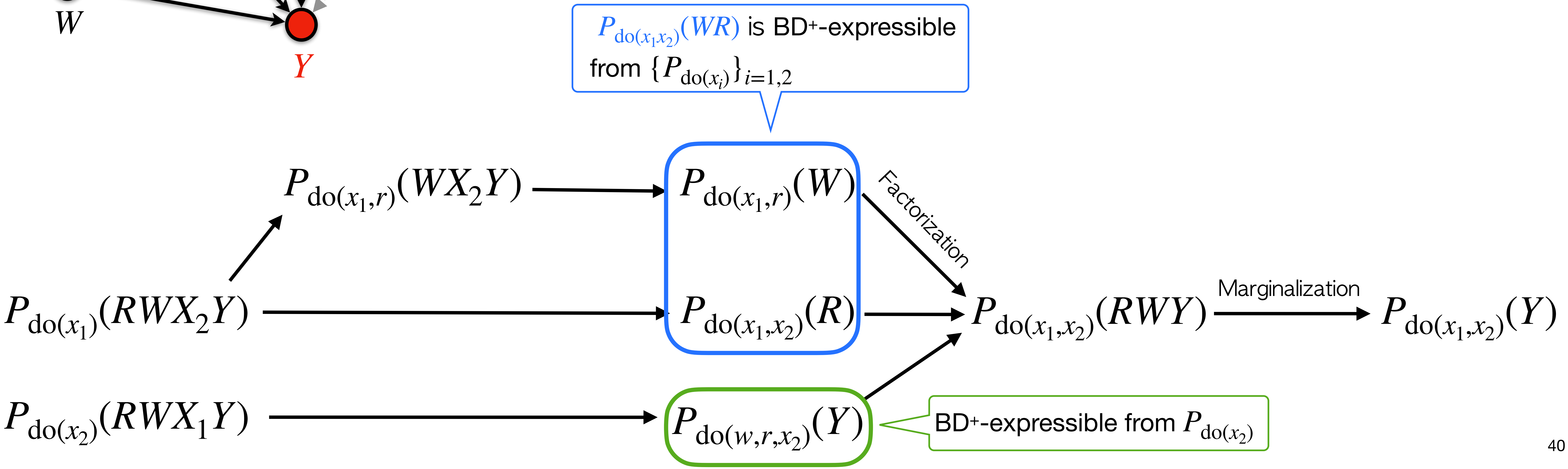
Causal effects as a function of BD+



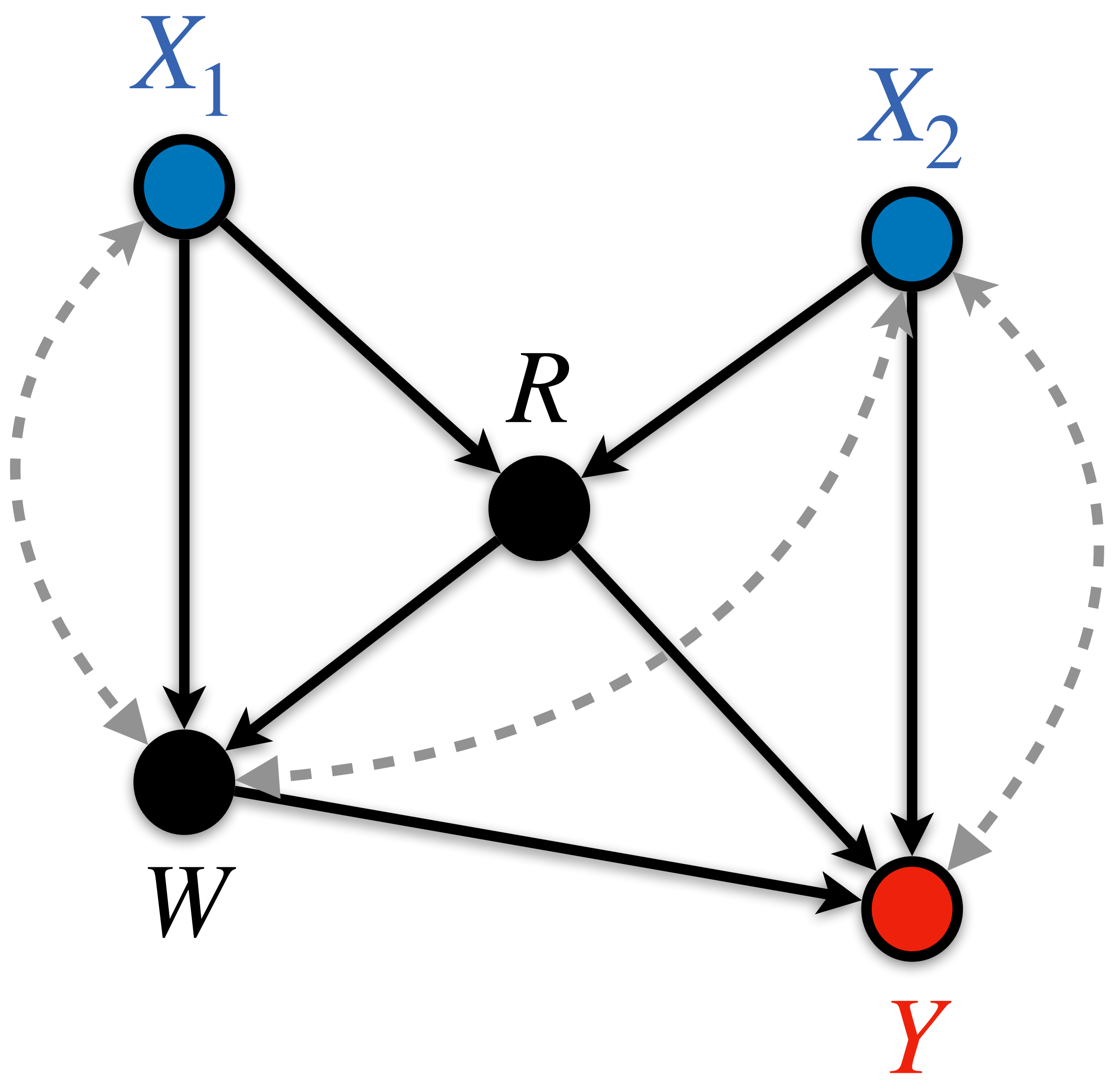


Causal effects as a function of BD+





Causal effects as a function of BD+



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

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

$P_{do(x_1)}(RWX_2Y) \rightarrow P_{do(x_1,r)}(WX_2Y) \rightarrow P_{do(x_1,r)}(W)$

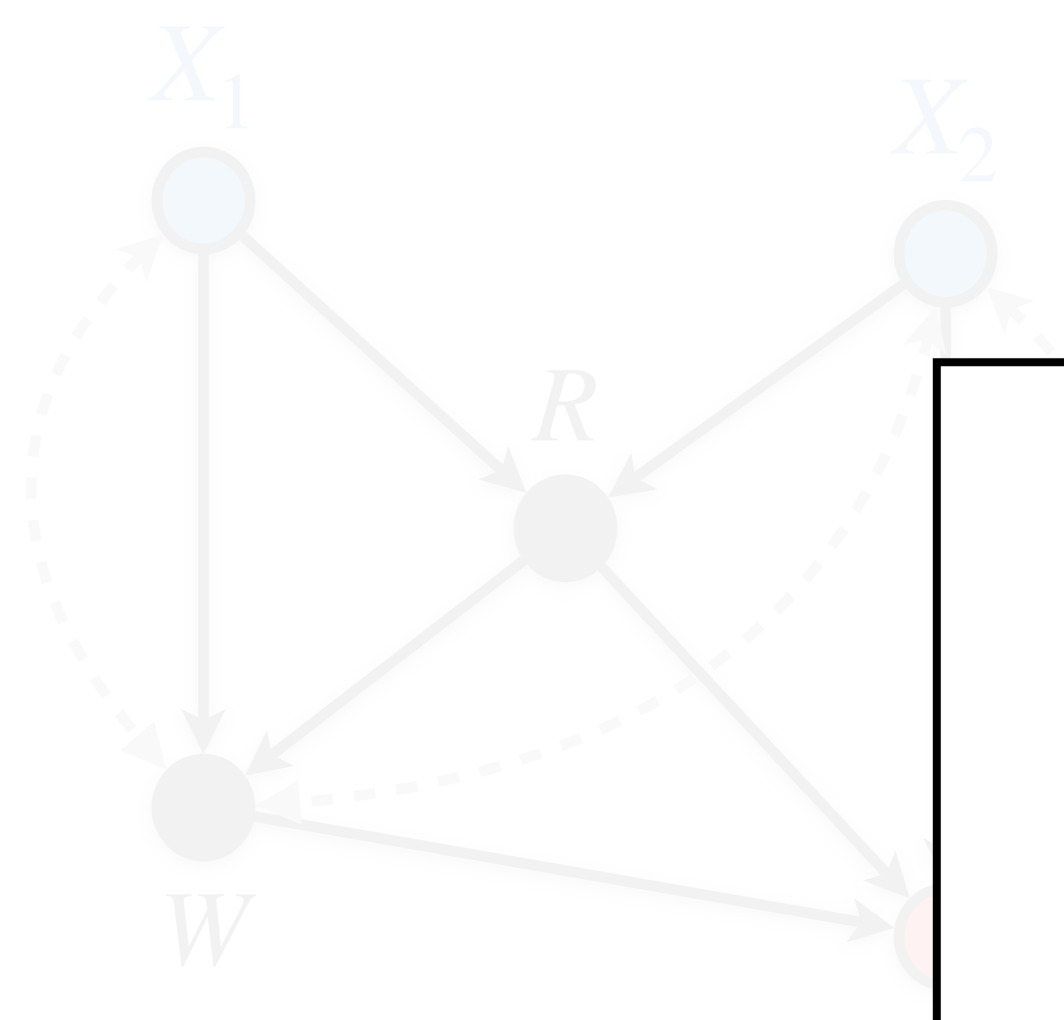
$P_{do(x_1)}(RWX_2Y) \rightarrow P_{do(x_1,x_2)}(R)$

$P_{do(x_2)}(RWX_1Y) \rightarrow P_{do(w,r,x_2)}(Y)$ — BD+-expressible from $P_{do(x_2)}$

Factorization: $P_{do(x_1,r)}(W),\ P_{do(x_1,x_2)}(R),\ P_{do(w,r,x_2)}(Y) \rightarrow P_{do(x_1,x_2)}(RWY)$

Marginalization: $P_{do(x_1,x_2)}(RWY) \rightarrow P_{do(x_1,x_2)}(Y)$

Causal effects as a function of BD+



Theorem 26

The followings are equivalent:

1. $P(\mathbf{y} \mid \mathrm{do}(\mathbf{x}))$ is identifiable from $(\mathcal{G}, \{P_{\mathrm{do}(\mathbf{r}_i)}\})$
2. $P(\mathbf{y} \mid \mathrm{do}(\mathbf{x}))$ is expressible as a ***function of BD+s*** through AdmissibleGID (Algo 6)

DML-gID: Estimator for Causal Effects from Fusion

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

DML-gID: Estimator for Causal Effects from Fusion

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

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

"DML-gID" (Def 49)

DML-gID: Estimator for Causal Effects from Fusion

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

DML-BD+ DML-BD+ ⋯ DML-BD+

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

"DML-gID" (Def 49)

Robustness of DML-gID

Theorem 27

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

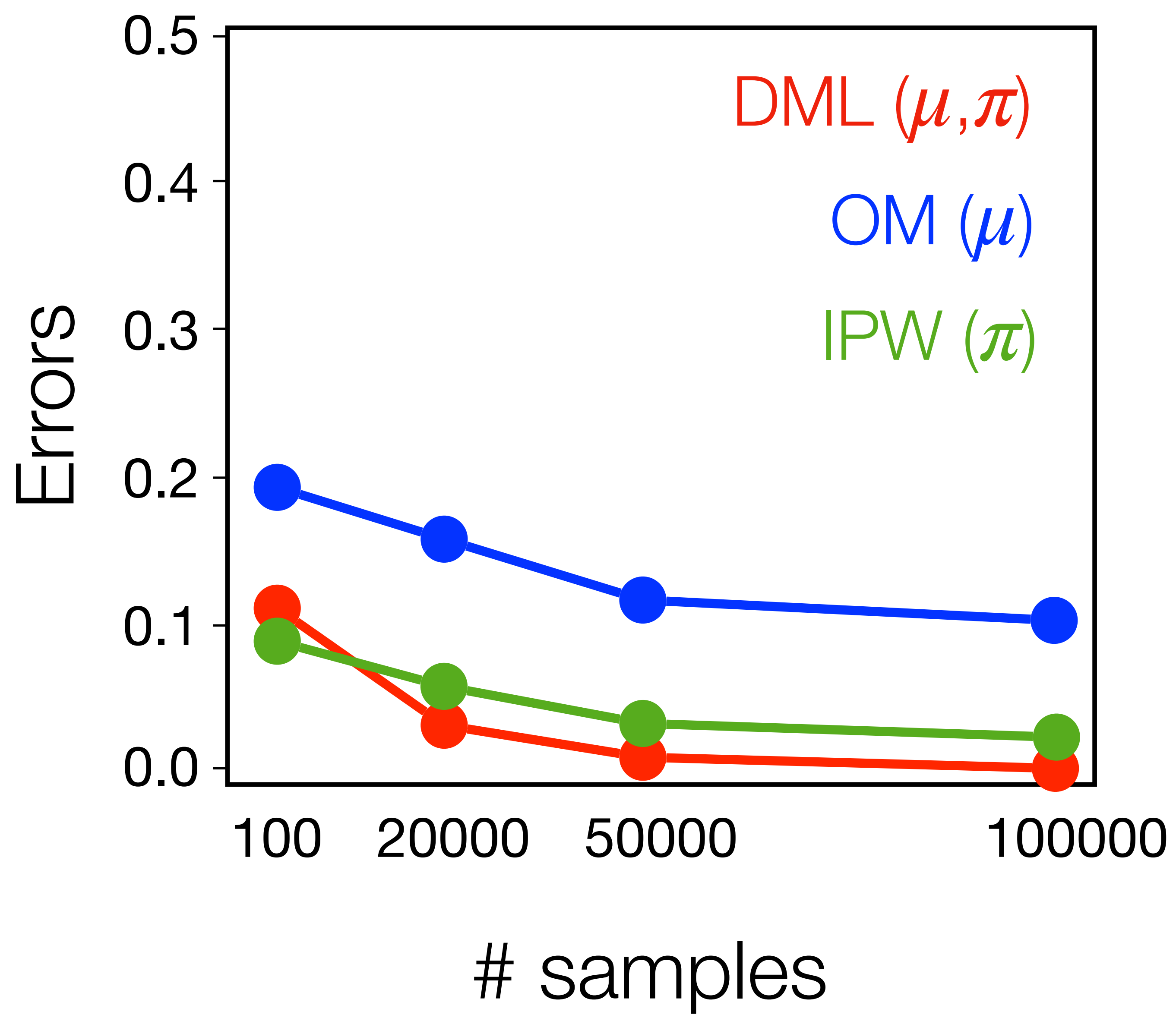
- **Double Robustness:** Error $= 0$ if either $\hat{\mu}_i = \mu_i$ or $\hat{\pi}_i = \pi_i$ for all $i = 1, \cdots, m$.
- **Fast Convergence:** Error $\rightarrow 0$ *fast* even when $\hat{\mu}_i \rightarrow \mu_i$ and $\hat{\pi}_i \rightarrow \pi_i$ *slow*.

DML-gID - Simulation (Sec. 4.6)

DML-gID - Simulation (Sec. 4.6)

Fast Convergence

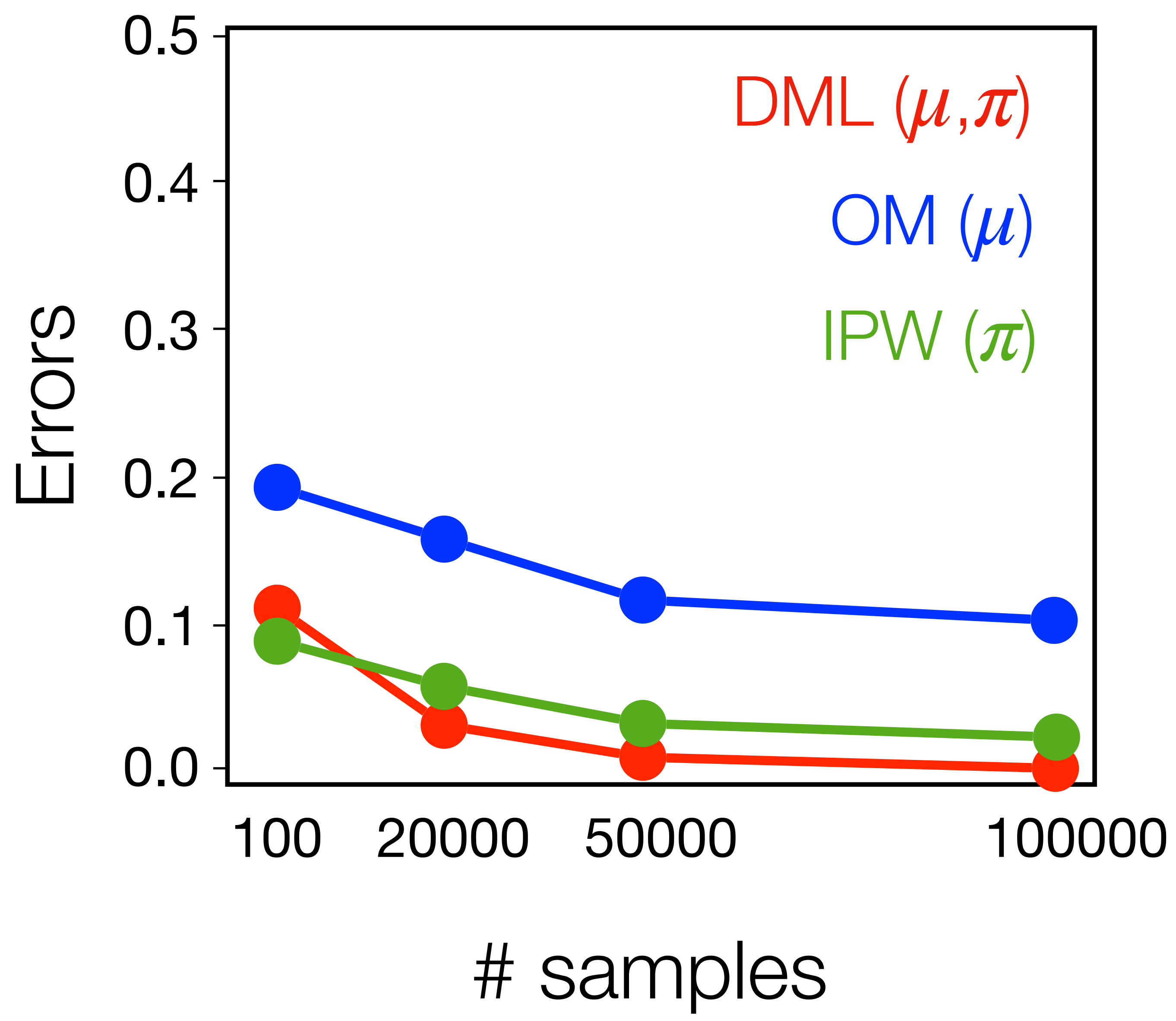
$(\hat{\boldsymbol{\mu}}, \hat{\boldsymbol{\pi}}) \rightarrow (\boldsymbol{\mu}_0, \boldsymbol{\pi}_0)$ slowly



DML-gID - Simulation (Sec. 4.6)

Fast Convergence

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

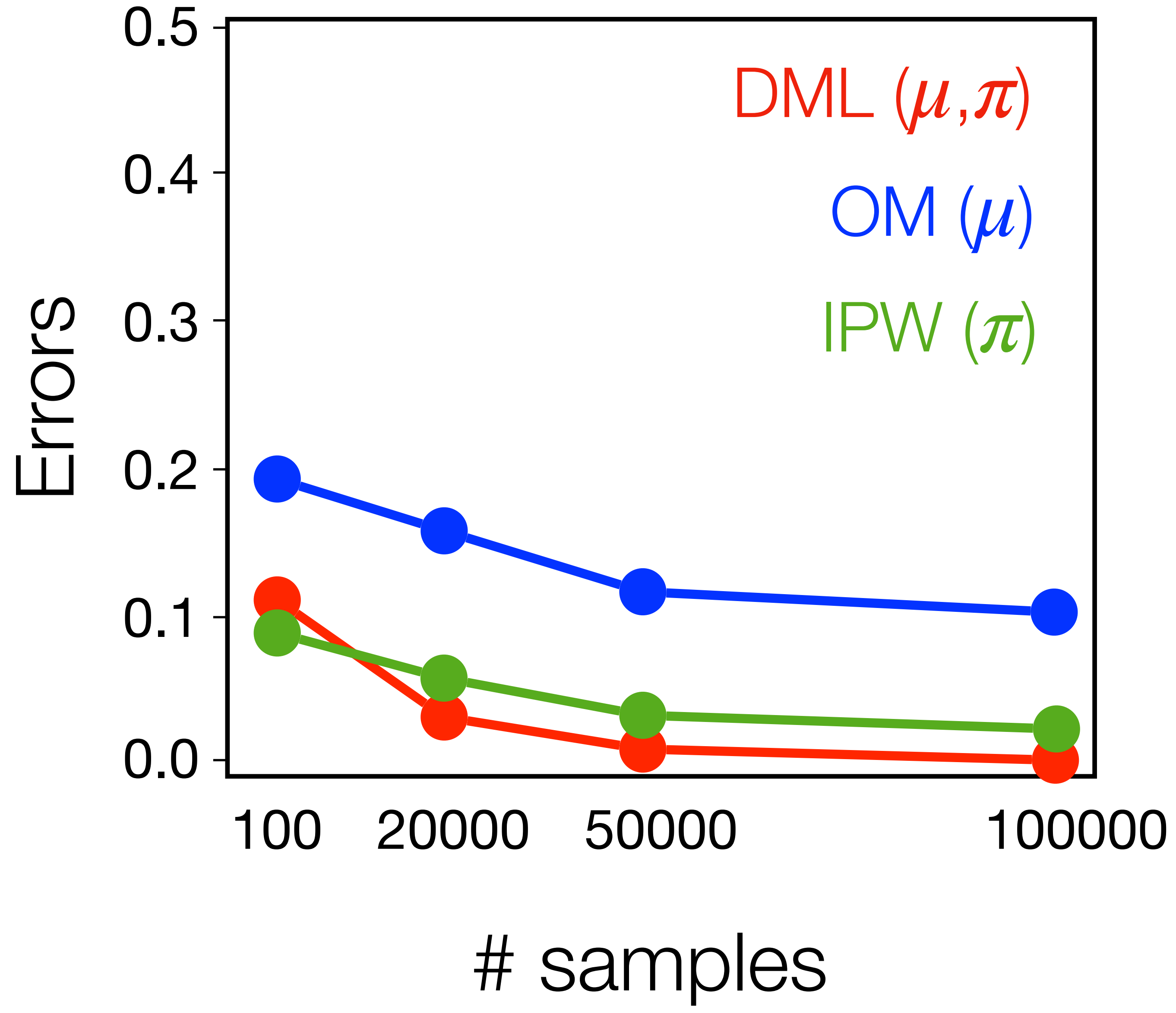


DML-gID converges fast, even when $(\hat{\boldsymbol{\mu}}, \hat{\boldsymbol{\pi}})$ converge slowly

DML-gID - Simulation (Sec. 4.6)

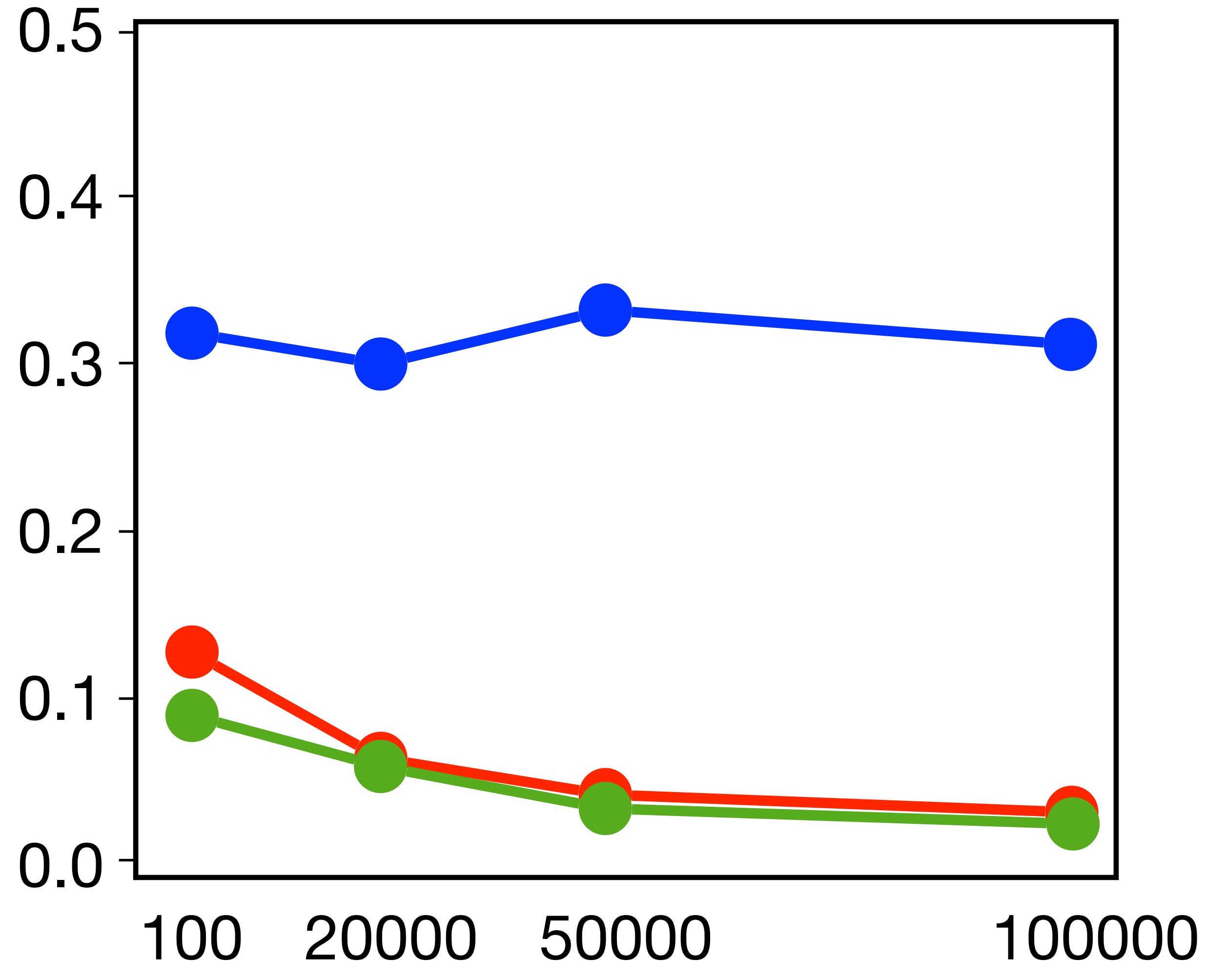
Fast Convergence

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

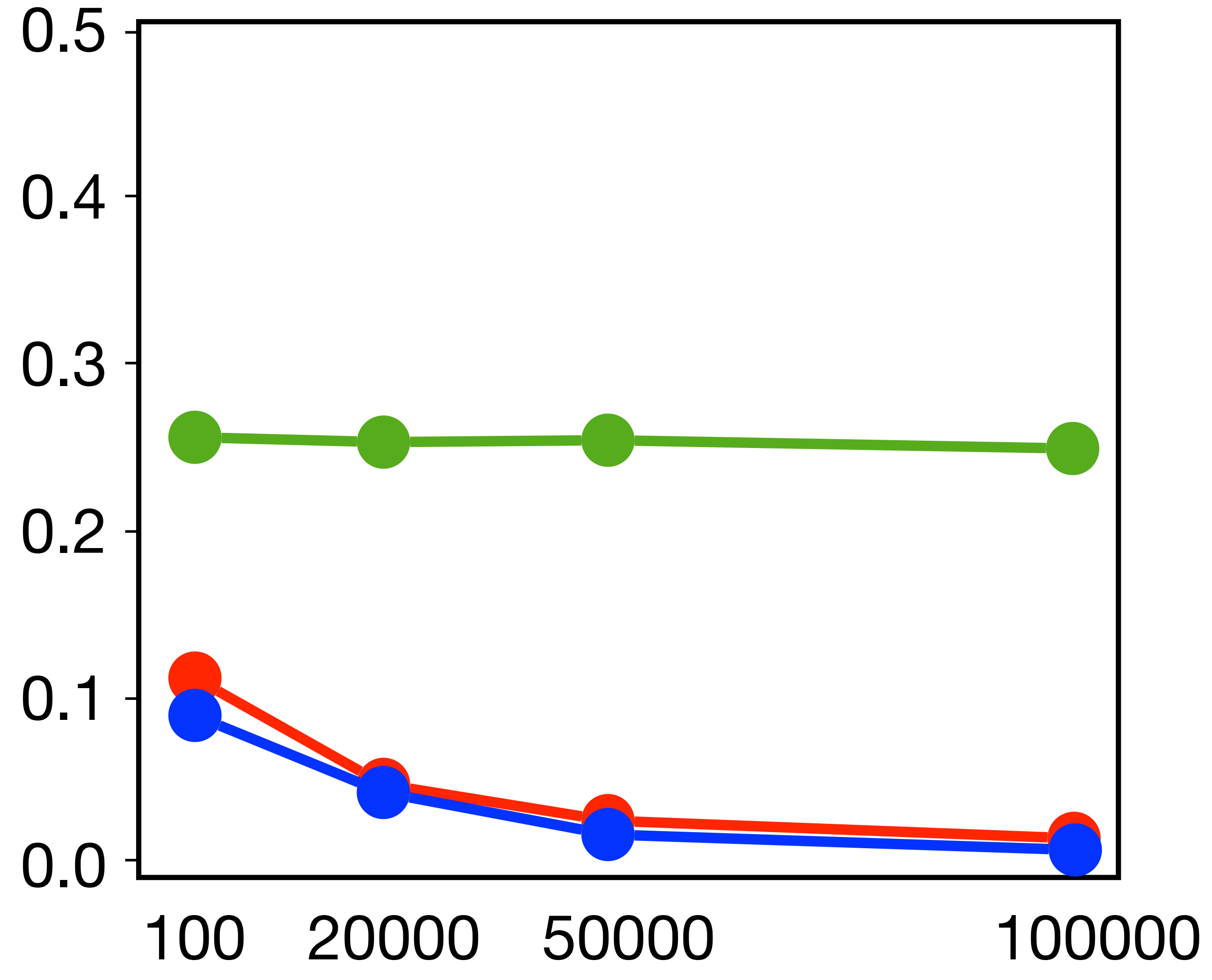


Double Robustness

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



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

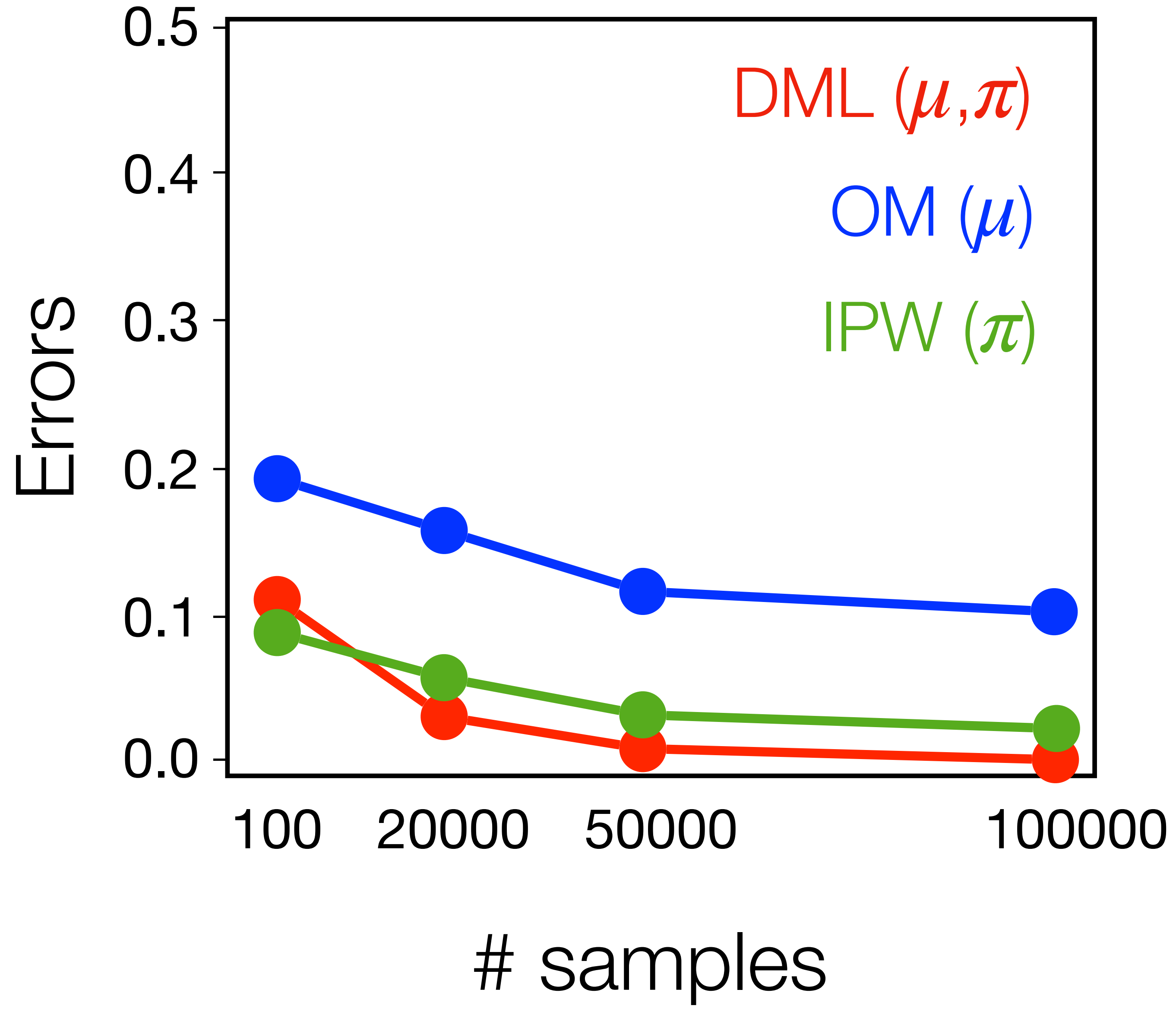


DML-gID converges fast, even when $(\hat{\mu},\hat{\pi})$ converge slowly

DML-gID - Simulation (Sec. 4.6)

Fast Convergence

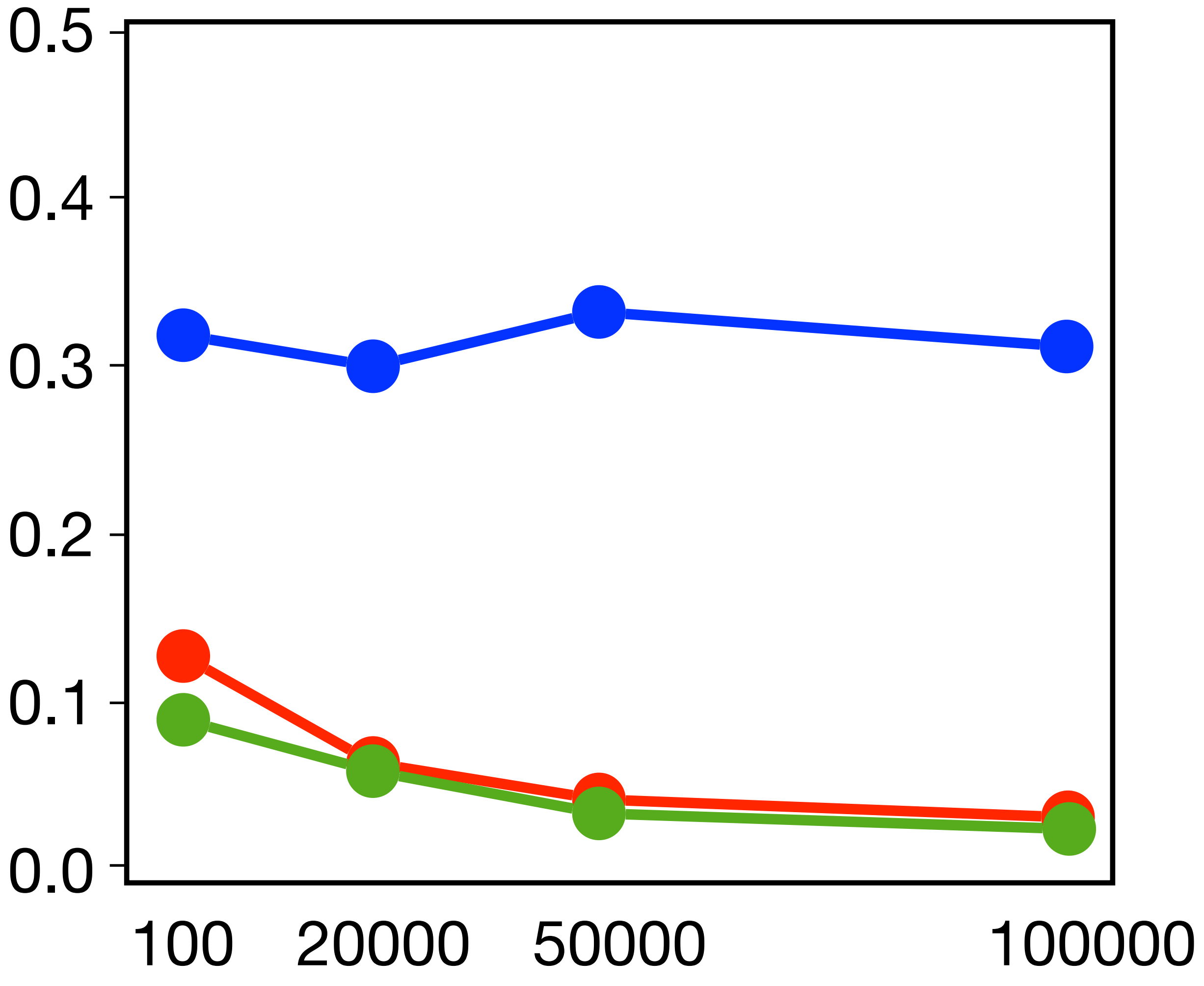
$(\hat{\mu},\hat{\pi}) \rightarrow (\mu_0, \pi_0)$ slowly



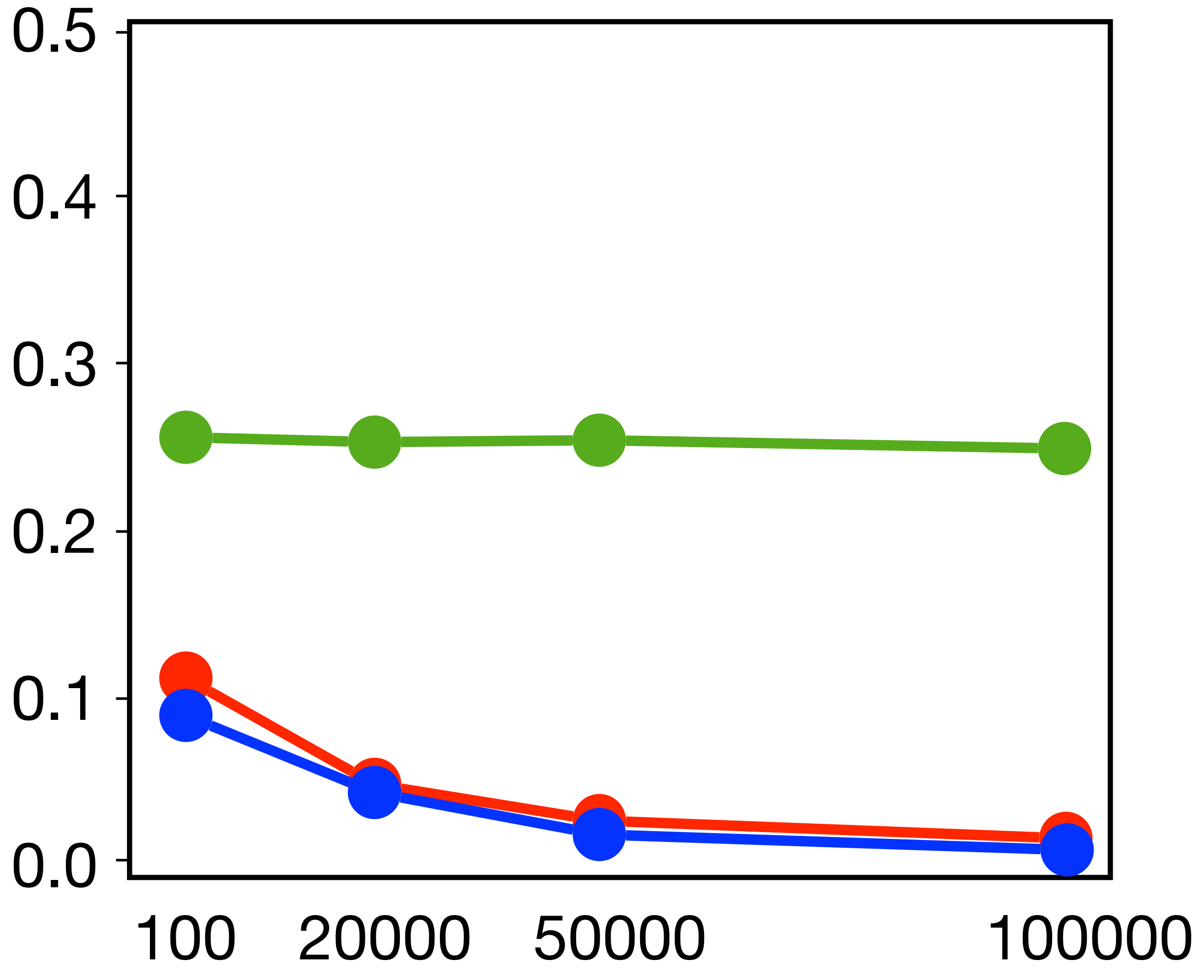
DML-gID converges fast, even when $(\hat{\mu},\hat{\pi})$ converge slowly

Double Robustness

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



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



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

Talk Outline

1. **Ch.3** Estimating causal effects from observations
2. **Ch.4** Estimating causal effects from data fusion
3. **Ch.5** Unified causal effect estimation method
4. Conclusion

Talk Outline

✔

➡ ❸ **Ch.5** Unified causal effect estimation method

Towards More General Causal Inference Queries

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

Towards More General Causal Inference Queries

Fairness Analysis

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

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

Towards More General Causal Inference Queries

Offline Policy Evaluation

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

Recovery rate of a drug dosage policy given baseline characteristics

Towards More General Causal Inference Queries

Joint Treatment Effect

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

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

Towards More General Causal Inference Queries

Counterfactual

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

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

Towards More General Causal Inference Queries

Missing Data

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

The effect of a treatment identifiable from missing data

Towards More General Causal Inference Queries

Domain Transfer

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

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

Towards More General Causal Inference Queries

Towards More General Causal Inference Queries

Fairness Analysis

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

Domain Transfer

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

Offline Policy Evaluation

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

Missing Data

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

Joint Treatment Effect

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

Counterfactual

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

Towards More General Causal Inference Queries

Fairness Analysis

Estimator 1

Domain Transfer

Estimator 6

Offline Policy Evaluation

Estimator 2

Missing Data

Estimator 5

Joint Treatment Effect

Estimator 3

Counterfactual

Estimator 4

Towards More General Causal Inference Queries

Jung et al., NeurIPS 2024

Chapter 5

Unified Covariate Adjustment (UCA)

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

Kernel Policy Product & Unified Covariate Adjustment

Kernel Policy Product & Unified Covariate Adjustment

Kernel Policy Product (Def. 50)

$$P_{m+1}(Y \mid \mathbf{S}^Z_m)\prod_{i=1}^{m}\sigma_i(\mathbf{X}_i \mid \mathbf{S}^X_i)P_i(\mathbf{Z}_i \mid \mathbf{S}^Z_{i-1})$$

Kernel Policy Product & Unified Covariate Adjustment

Kernel Policy Product (Def. 50)

Arbitrary prob. kernel
(distributions)

$$P_{m+1}(Y \mid \mathbf{S}^{Z}_{m}) \prod_{i=1}^{m} \sigma_i(\mathbf{X}_i \mid \mathbf{S}^{X}_{i}) P_i(\mathbf{Z}_i \mid \mathbf{S}^{Z}_{i-1})$$

Kernel Policy Product & Unified Covariate Adjustment

Kernel Policy Product (Def. 50)

Arbitrary prob. kernel (distributions)

Arbitrary vectors

$$P_{m+1}(Y \mid \mathbf{S}^{Z}_{m})\prod_{i=1}^{m} \sigma_i(\mathbf{X}_i \mid \mathbf{S}^{X}_{i})P_i(\mathbf{Z}_i \mid \mathbf{S}^{Z}_{i-1})$$

Kernel Policy Product & Unified Covariate Adjustment

Kernel Policy Product (Def. 50)

Arbitrary prob. kernel (distributions)

Arbitrary vectors

$$P_{m+1}(Y \mid \mathbf{S}^{Z}_{m}) \prod_{i=1}^{m} \sigma_i(\mathbf{X}_i \mid \mathbf{S}^{X}_{i}) P_i(\mathbf{Z}_i \mid \mathbf{S}^{Z}_{i-1})$$

outcome

Policies

Covariates

Kernel Policy Product & Unified Covariate Adjustment

Kernel Policy Product (Def. 50)

Arbitrary prob. kernel (distributions)

Arbitrary vectors

$$P_{m+1}(Y \mid \mathbf{S}^{Z}_{m}) \prod_{i=1}^{m} \sigma_i(\mathbf{X}_i \mid \mathbf{S}^{X}_{i}) P_i(\mathbf{Z}_i \mid \mathbf{S}^{Z}_{i-1})$$

outcome

Policies

Covariates

Unified Covariate Adjustment (Def. 51)

Expectation of Y over the KPP

Canonical Example of UCA

arbitrary distributions

outcome

arbitrary policy of treatments

arbitrary distributions

set of variables

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

Canonical Example of UCA

arbitrary distributions — outcome — arbitrary policy of treatments — arbitrary distributions — set of variables

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

UCA

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

→

BD+

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

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

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

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

$Z \leftarrow \{W\}$

Canonical Example of UCA

arbitrary distributions
outcome
arbitrary policy of treatments
arbitrary distributions
set of variables

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

UCA

BD+

Theorem 28

UCA can represent ***any*** causal effects expressible as a sum of products of arbitrary conditional distributions (*Kernel-Policy Product*), by choosing $Z, P_1, P_2, \sigma(\cdot \mid \cdot)$ properly.

Parameterization for UCA (Def. 54)

$$\psi_0 \triangleq \sum_{x,c} \mathbb{E}_{P_2}[Y \mid x, z]\sigma(x \mid z)P_1(z)$$

Parameterization for UCA (Def. 54)

$$\psi_0 \triangleq \sum_{x,c} \mathbb{E}_{P_2}[Y \mid x, z]\sigma(x \mid z)P_1(z)$$

$$\mu(\mathbf{X},\mathbf{Z}) \triangleq \mathbb{E}_{P_2}[Y \mid \mathbf{X}, \mathbf{Z}]$$

Parameterization for UCA (Def. 54)

$$\psi_0 \triangleq \sum_{x,c} \mathbb{E}_{P_2}[Y \mid x, z]\sigma(x \mid z)P_1(z)$$

$$\mu(\mathbf{X},\mathbf{Z}) \triangleq \mathbb{E}_{P_2}[Y \mid \mathbf{X}, \mathbf{Z}]$$

$$\mathbb{E}_{P_1}[\mathbb{E}_{\sigma(\mathbf{X}|\mathbf{Z})}[\mu(\mathbf{X},\mathbf{Z})]]$$

$$= \sum_{\mathbf{z},\mathbf{x}} \mu(\mathbf{x},\mathbf{z})\sigma_X(\mathbf{x}|\mathbf{z})P_1(\mathbf{z})$$

$$= \psi_0$$

Parameterization for UCA (Def. 54)

$$\psi_0 \triangleq \sum_{x,c} \mathbb{E}_{P_2}[Y \mid x, z]\sigma(x \mid z)P_1(z)$$

$$\mu(\mathbf{X},\mathbf{Z}) \triangleq \mathbb{E}_{P_2}[Y \mid \mathbf{X}, \mathbf{Z}]$$

$$\mathbb{E}_{P_1}[\mathbb{E}_{\sigma(\mathbf{X}|\mathbf{Z})}[\mu(\mathbf{X},\mathbf{Z})]]$$

$$= \sum_{\mathbf{z},\mathbf{x}} \mu(\mathbf{x},\mathbf{z})\sigma_X(\mathbf{x}|\mathbf{z})P_1(\mathbf{z})$$

$$= \psi_0$$

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

$$\mathbb{E}_{P_2}[\pi(\mathbf{X},\mathbf{Z}) \times \mu(\mathbf{X},\mathbf{Z})] = \mathbb{E}_{P_1}[\mathbb{E}_{\sigma_X}[\mu(\mathbf{X},\mathbf{Z})]]$$

Parameterization for UCA (Def. 54)

$$\psi_0 \triangleq \sum_{x,c} \mathbb{E}_{P_2}[Y \mid x, z]\sigma(x \mid z)P_1(z)$$

$$\mu(\mathbf{X},\mathbf{Z}) \triangleq \mathbb{E}_{P_2}[Y \mid \mathbf{X}, \mathbf{Z}]$$

$$\mathbb{E}_{P_1}[\mathbb{E}_{\sigma(\mathbf{X}|\mathbf{Z})}[\mu(\mathbf{X},\mathbf{Z})]]$$

$$= \sum_{\mathbf{z},\mathbf{x}} \mu(\mathbf{x},\mathbf{z})\sigma_X(\mathbf{x}|\mathbf{z})P_1(\mathbf{z})$$

$$= \psi_0$$

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

$$\mathbb{E}_{P_2}[\pi(\mathbf{X},\mathbf{Z}) \times \mu(\mathbf{X},\mathbf{Z})] = \mathbb{E}_{P_1}[\mathbb{E}_{\sigma_X}[\mu(\mathbf{X},\mathbf{Z})]]$$

$$\mathbb{E}_{P_2}[\pi(\mathbf{X},\mathbf{Z}) \times Y]$$

$$= \mathbb{E}_{P_2}[\pi(\mathbf{X},\mathbf{Z}) \times \mu(\mathbf{X},\mathbf{Z})]$$

$$= \mathbb{E}_{P_1}[\mathbb{E}_{\sigma_X}[\mu(\mathbf{X},\mathbf{Z})]]$$

$$= \psi_0$$

Doubly Robust Estimator for UCA

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

Doubly Robust Estimator for UCA

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

"Double Robustness"

$$?(\hat{\mu}, \hat{\pi}) - \mathbb{E}_{P_2}[\mu \times \pi] = \mathbb{E}_{P_2}[\{\hat{\mu} - \mu\} \times \{\pi - \hat{\pi}\}]$$

Doubly Robust Estimator for UCA

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

$$?(\hat{\mu}, \hat{\pi}) = \mathbb{E}_{P_2}[\{\hat{\mu} - \mu\} \times \{\pi - \hat{\pi}\}] + \mathbb{E}_{P_2}[\mu \times \pi]$$

Doubly Robust Estimator for UCA

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

$$?(\hat{\mu}, \hat{\pi}) = \mathbb{E}_{P_2}[\{\hat{\mu} - \mu\} \times \{\pi - \hat{\pi}\}] + \mathbb{E}_{P_2}[\mu \times \pi]$$

$$= \mathbb{E}_{P_2}[\hat{\pi}\{\mu - \hat{\mu}\} + \pi\hat{\mu}]$$

Doubly Robust Estimator for UCA

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

$$\begin{aligned} \textbf{?}(\hat{\mu}, \hat{\pi}) &= \mathbb{E}_{P_2}[\{\hat{\mu} - \mu\} \times \{\pi - \hat{\pi}\}] + \mathbb{E}_{P_2}[\mu \times \pi] \\ &= \mathbb{E}_{P_2}[\hat{\pi}\{\mu - \hat{\mu}\} + \pi\hat{\mu}] \\ &= \mathbb{E}_{P_2}[\hat{\pi}\{Y - \hat{\mu}\}] + \mathbb{E}_{P_1}[\mathbb{E}_{\sigma_X}[\hat{\mu}]] \end{aligned}$$

Doubly Robust Estimator for UCA

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

$$\textbf{?}(\hat{\mu}, \hat{\pi}) = \mathbb{E}_{P_2}[\{\hat{\mu} - \mu\} \times \{\pi - \hat{\pi}\}] + \mathbb{E}_{P_2}[\mu \times \pi]$$

$$= \mathbb{E}_{P_2}[\hat{\pi}\{\mu - \hat{\mu}\} + \pi\hat{\mu}]$$

$$= \mathbb{E}_{P_2}[\hat{\pi}\{Y - \hat{\mu}\}] + \mathbb{E}_{P_1}[\mathbb{E}_{\sigma_X}[\hat{\mu}]]$$

DML-UCA (Double Machine Learning estimator for UCA)

$$\widehat{\text{UCA}}(\hat{\mu}, \hat{\pi}) \triangleq \mathbb{E}_{P_2}[\hat{\pi}\{Y - \hat{\mu}\}] + \mathbb{E}_{P_1}[\mathbb{E}_{\sigma_X}[\hat{\mu}]]$$

Robustness Property of DML-UCA

Theorem 33

$$\text{Error}(\text{DML-UCA}, \psi_0) = \sum_{i=1}^{m} \text{Error}(\hat{\mu}_i, \mu_i) \times \text{Error}(\hat{\pi}_i, \pi_i)$$

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

Simulation: DML-UCA

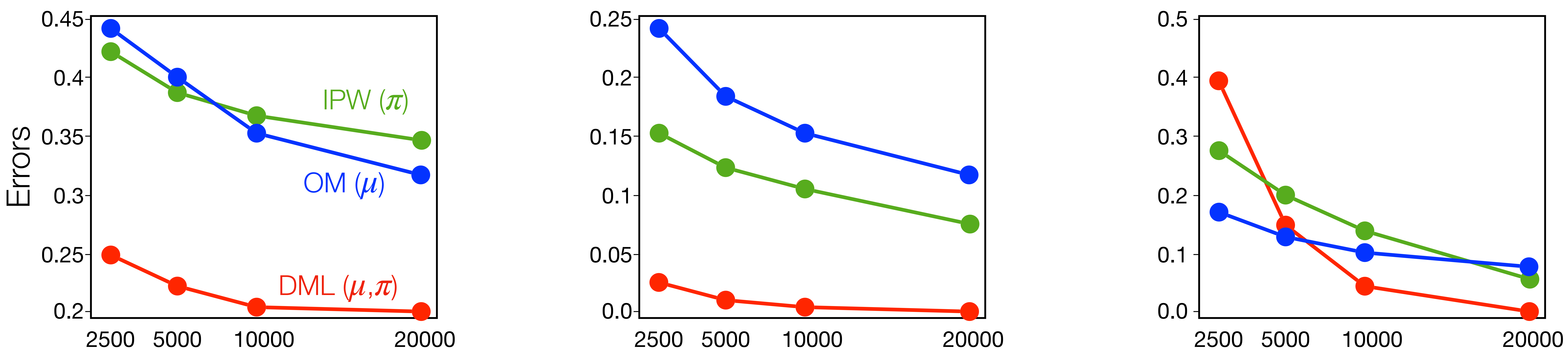
Simulation: DML-UCA

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

Simulation: DML-UCA

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

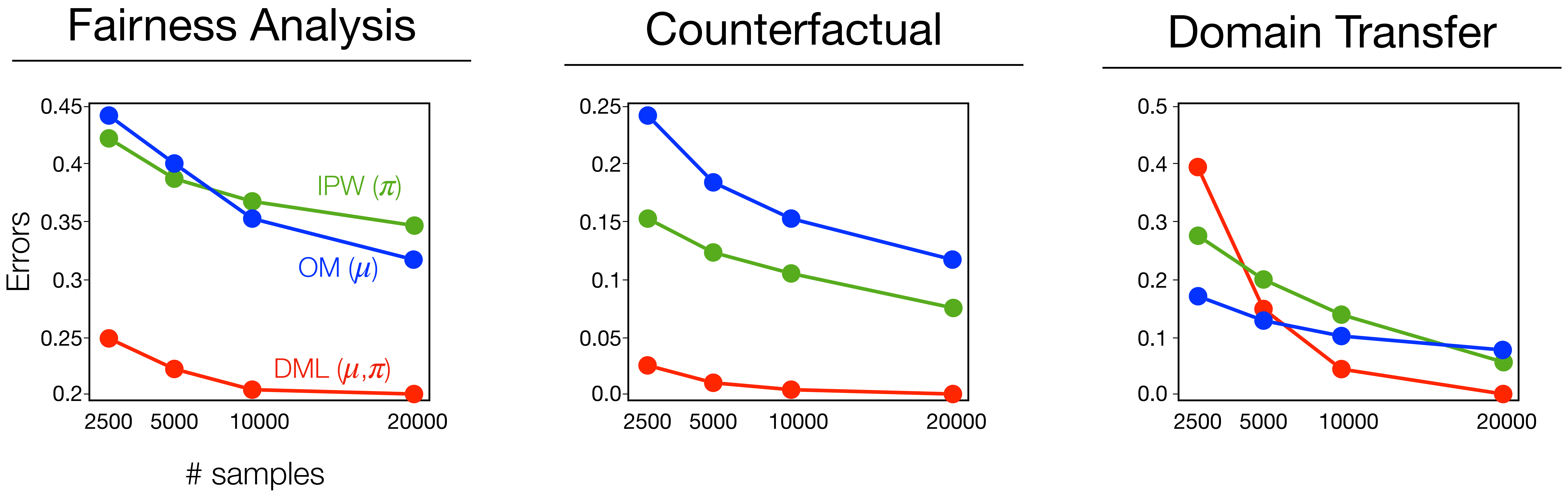
Fairness Analysis | Counterfactual | Domain Transfer



samples

Simulation: DML-UCA

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



DML-UCA converges fast even when $(\hat{\boldsymbol{\mu}}, \hat{\boldsymbol{\pi}}) \to (\boldsymbol{\mu}_0, \boldsymbol{\pi}_0)$ slowly

Talk Outline

1. **Ch.3** Estimating causal effects from observations
2. **Ch.4** Estimating causal effects from data fusion
3. **Ch.5** Unified causal effect estimation method
4. Conclusion

Talk Outline

✔

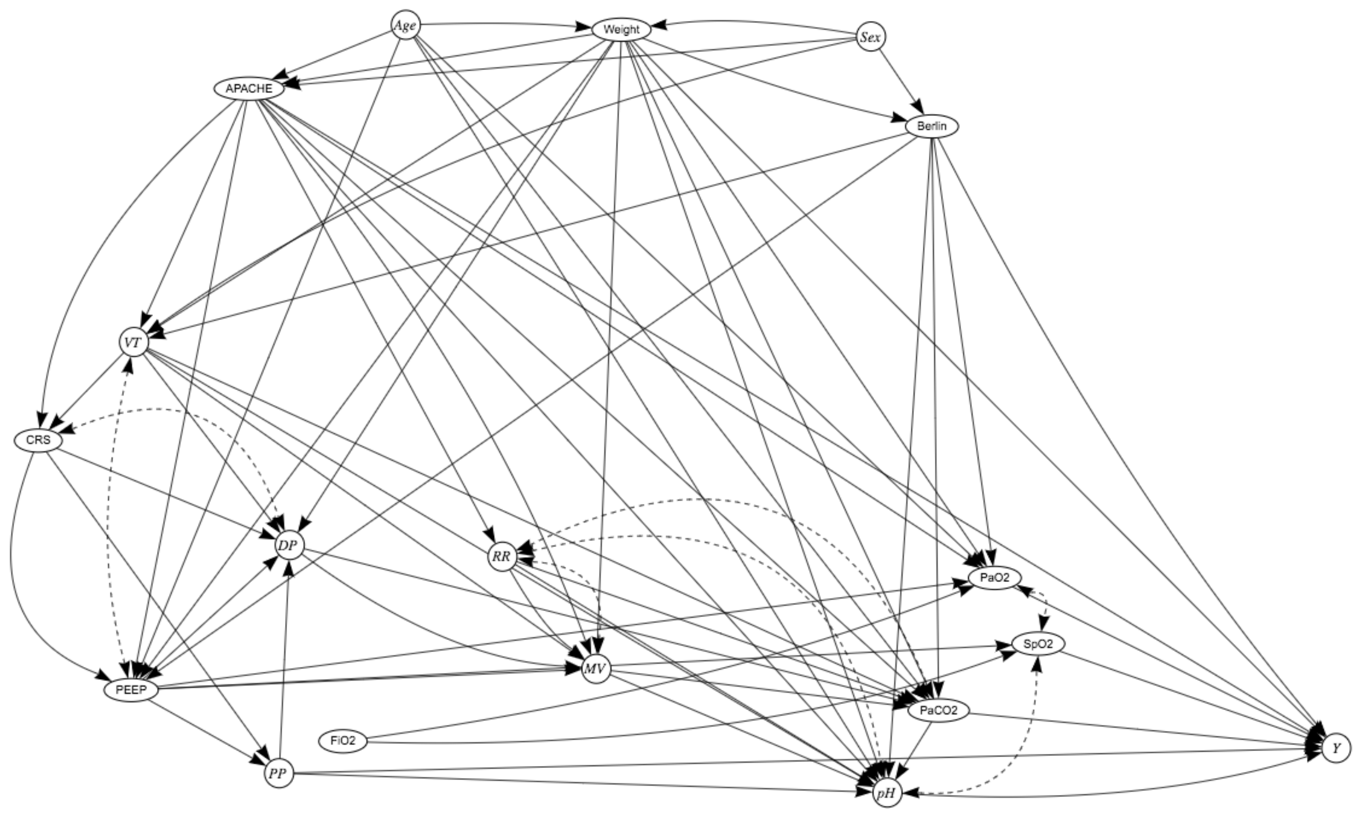
4 Conclusion

This Talk: Estimating Causal Effects

This Talk: Estimating Causal Effects

Tasks

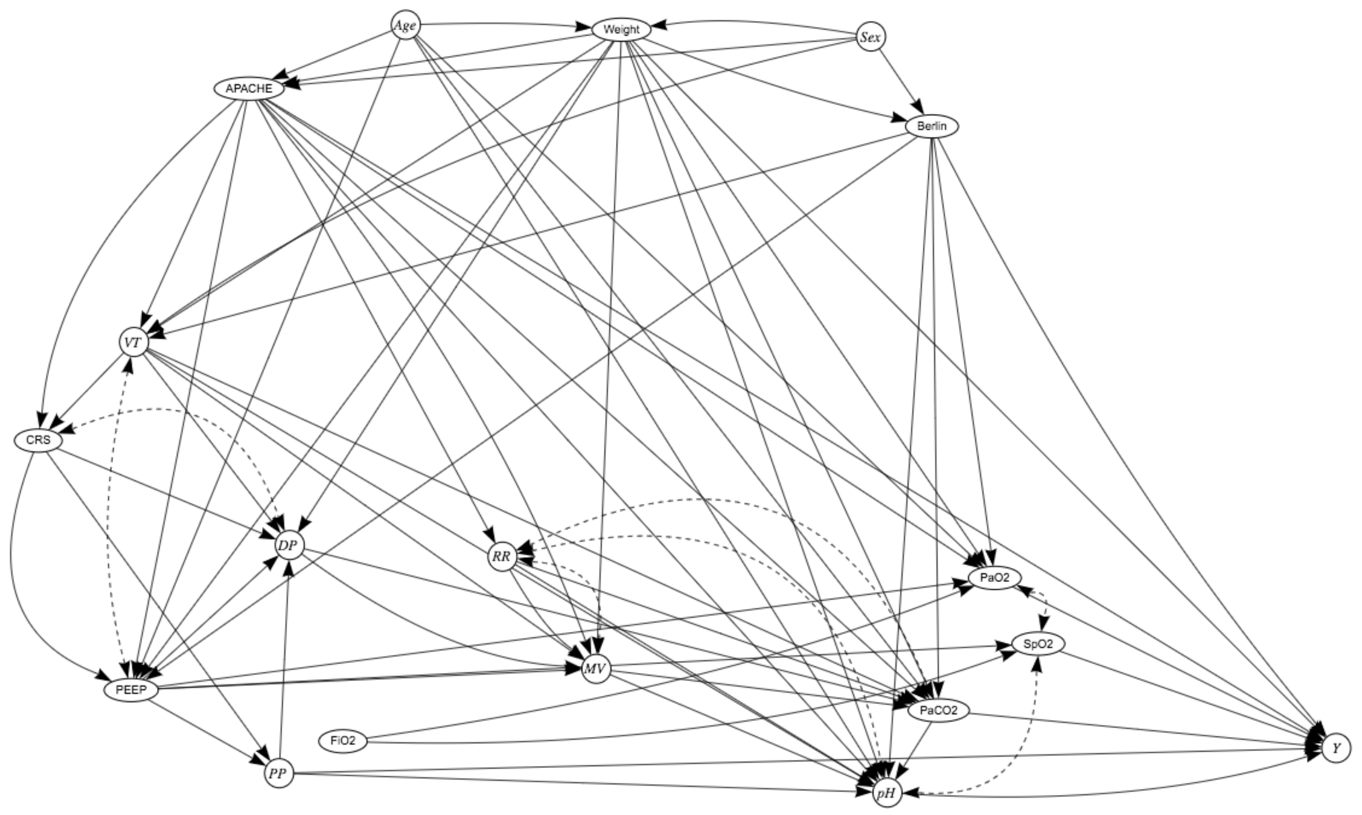
1. From Observation



This Talk: Estimating Causal Effects

Tasks

1. From Observation



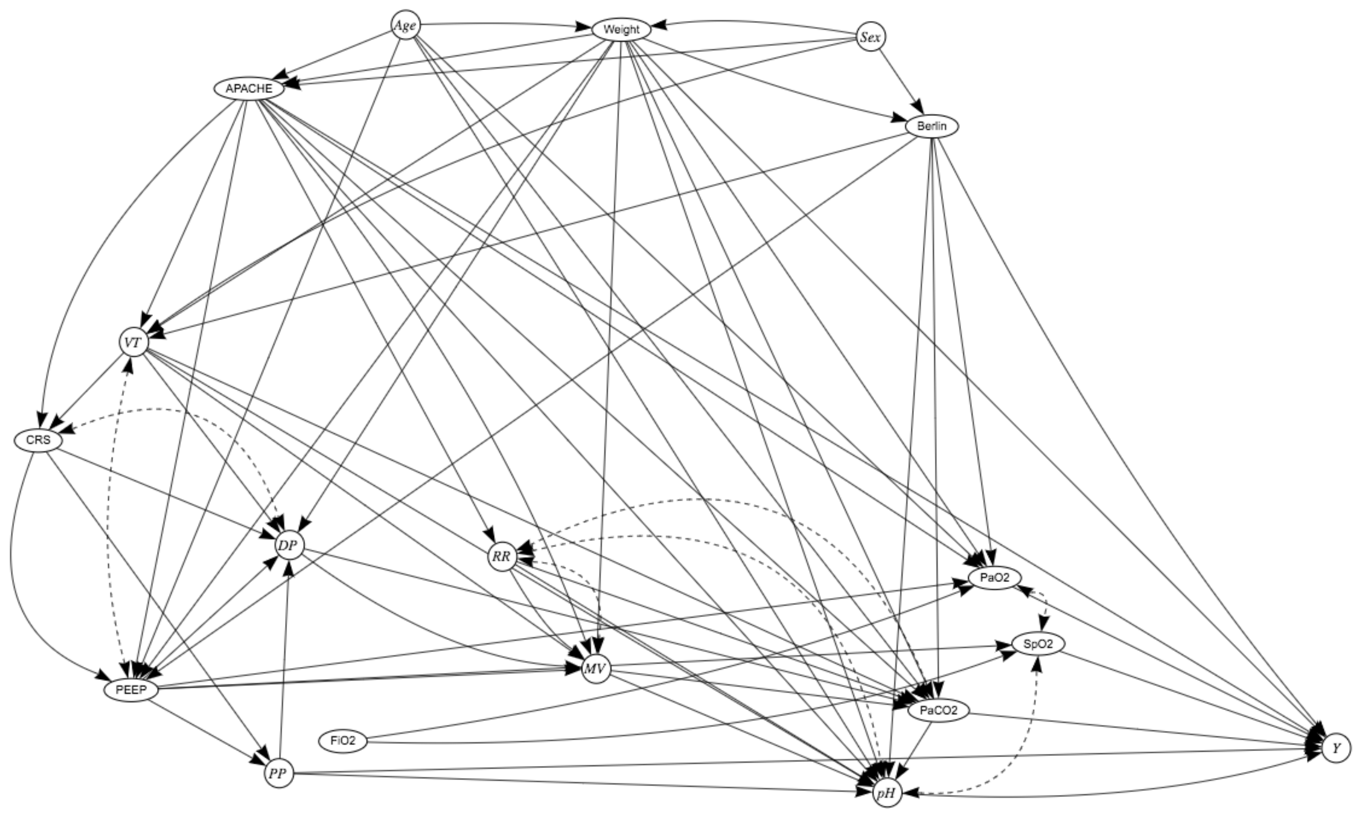
Solution

DML-ID

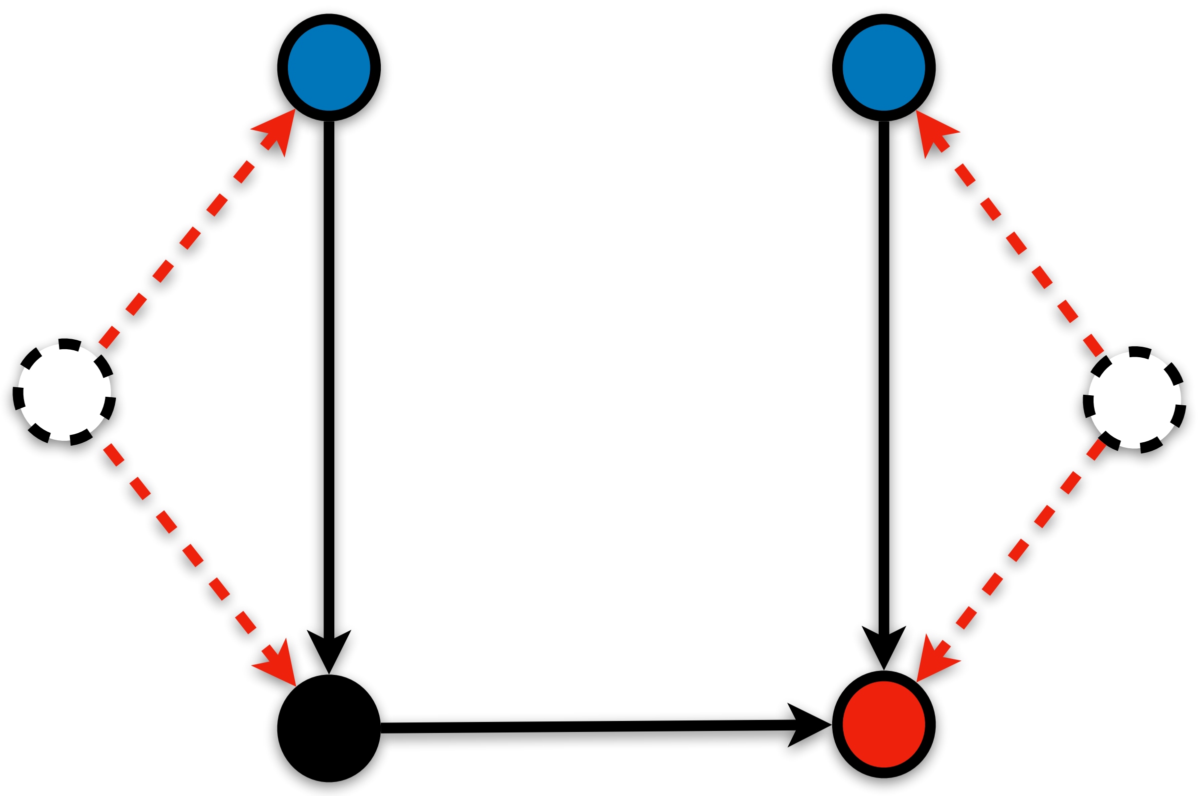
This Talk: Estimating Causal Effects

Tasks

1. From Observation



2. From Data Fusion



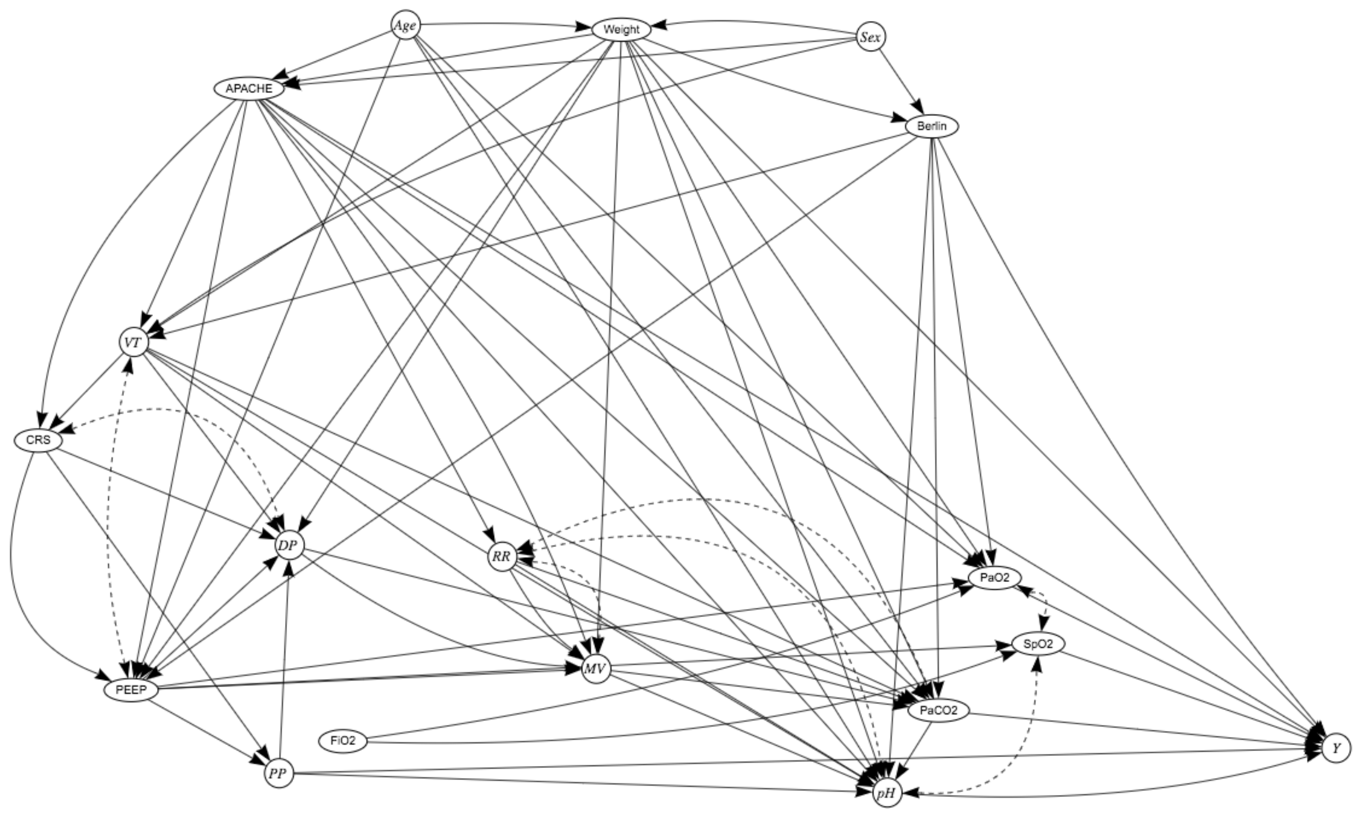
Solution

DML-ID

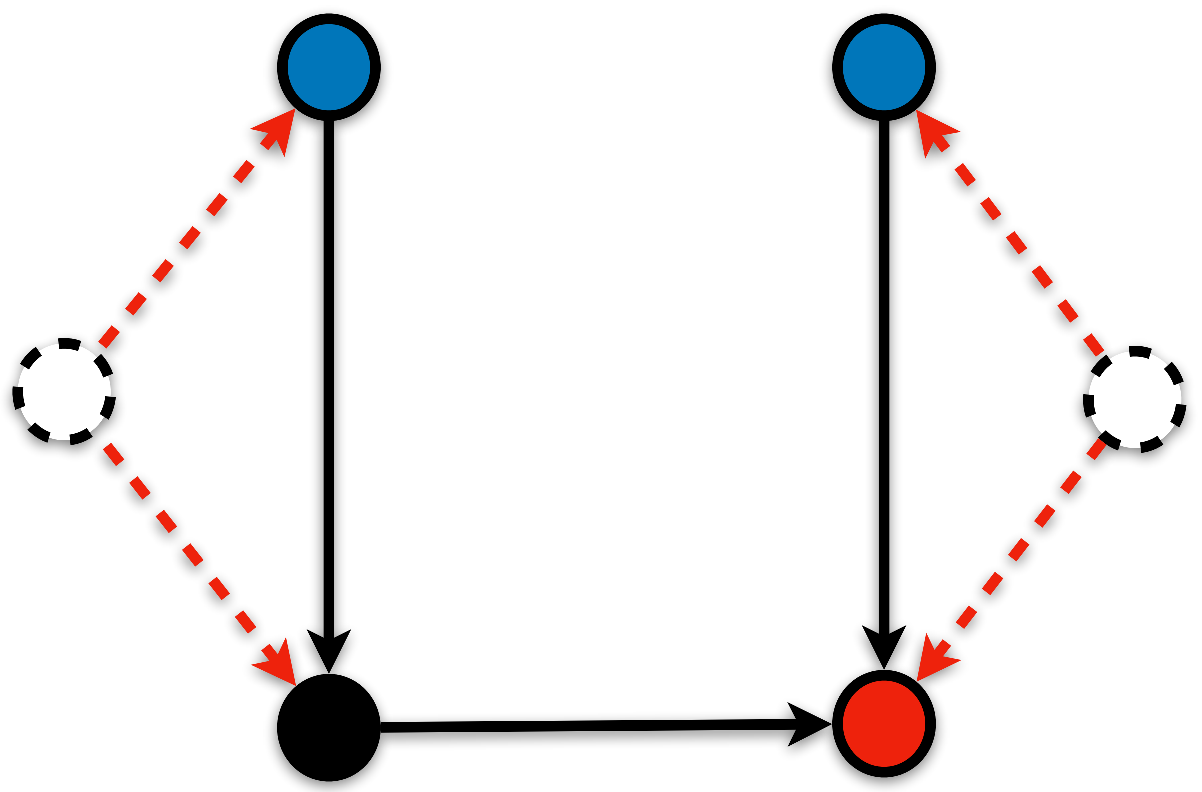
This Talk: Estimating Causal Effects

Tasks

1. From Observation



2. From Data Fusion



Solution

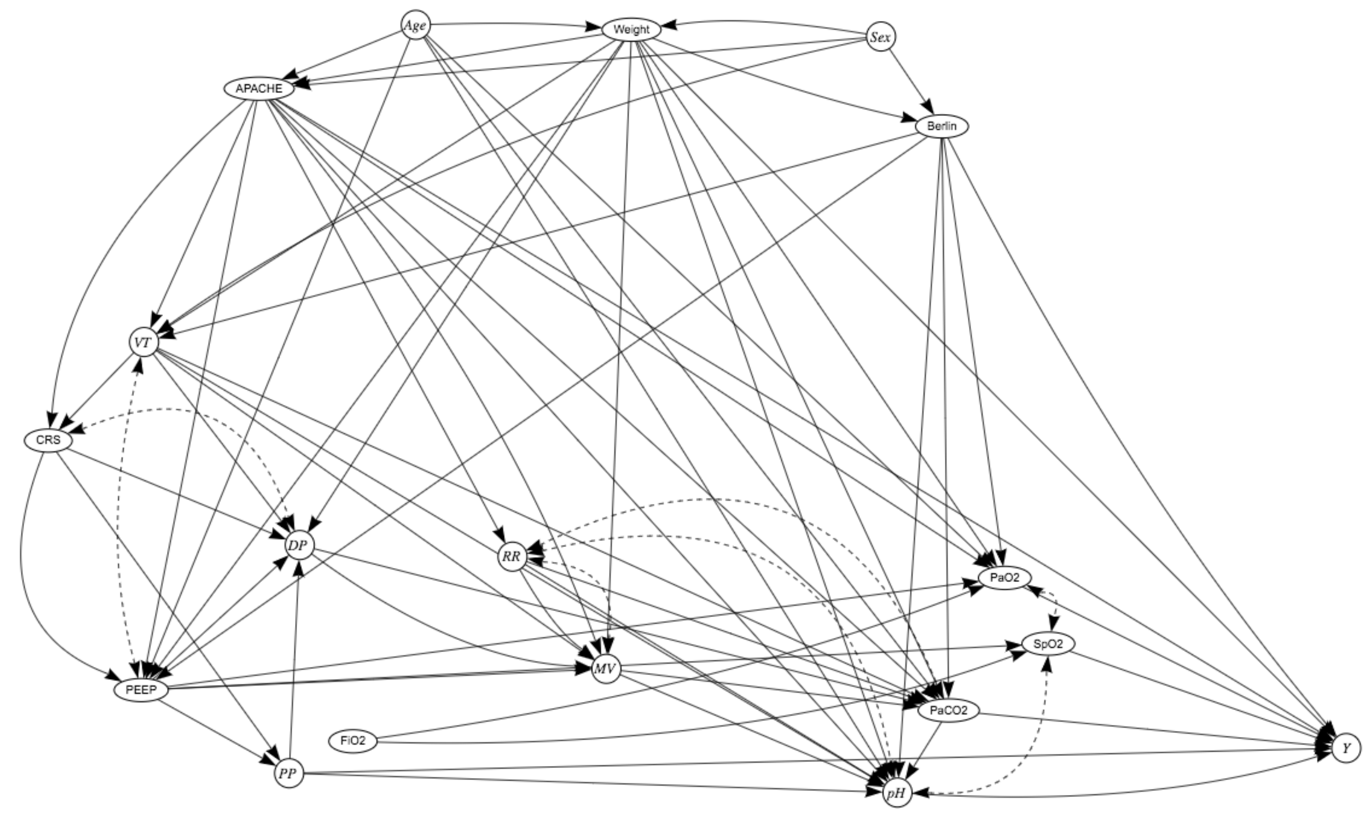
DML-ID

- DML-BD⁺
- DML-gID

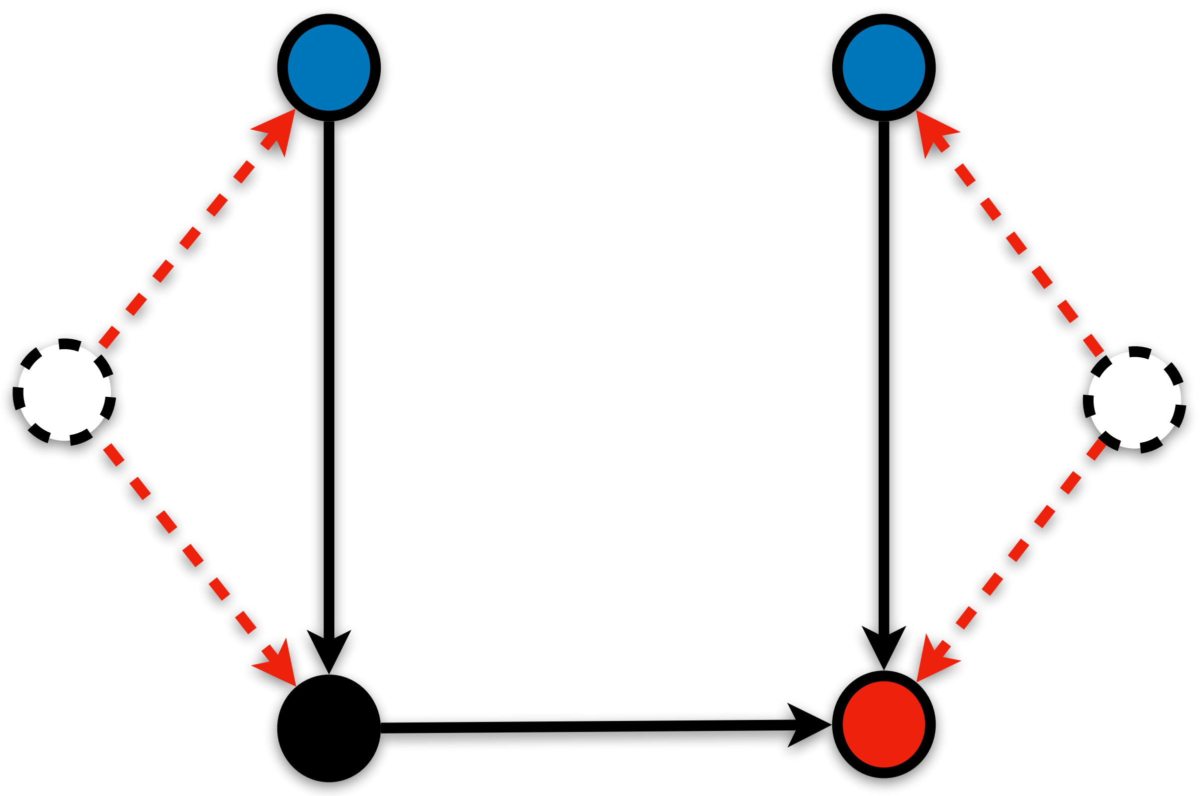
This Talk: Estimating Causal Effects

Tasks

1. From Observation



2. From Data Fusion



3. Unified Estimation

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

Off-policy evaluation $\mathbb{E}[Y_{\tau(X|C)}]$

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

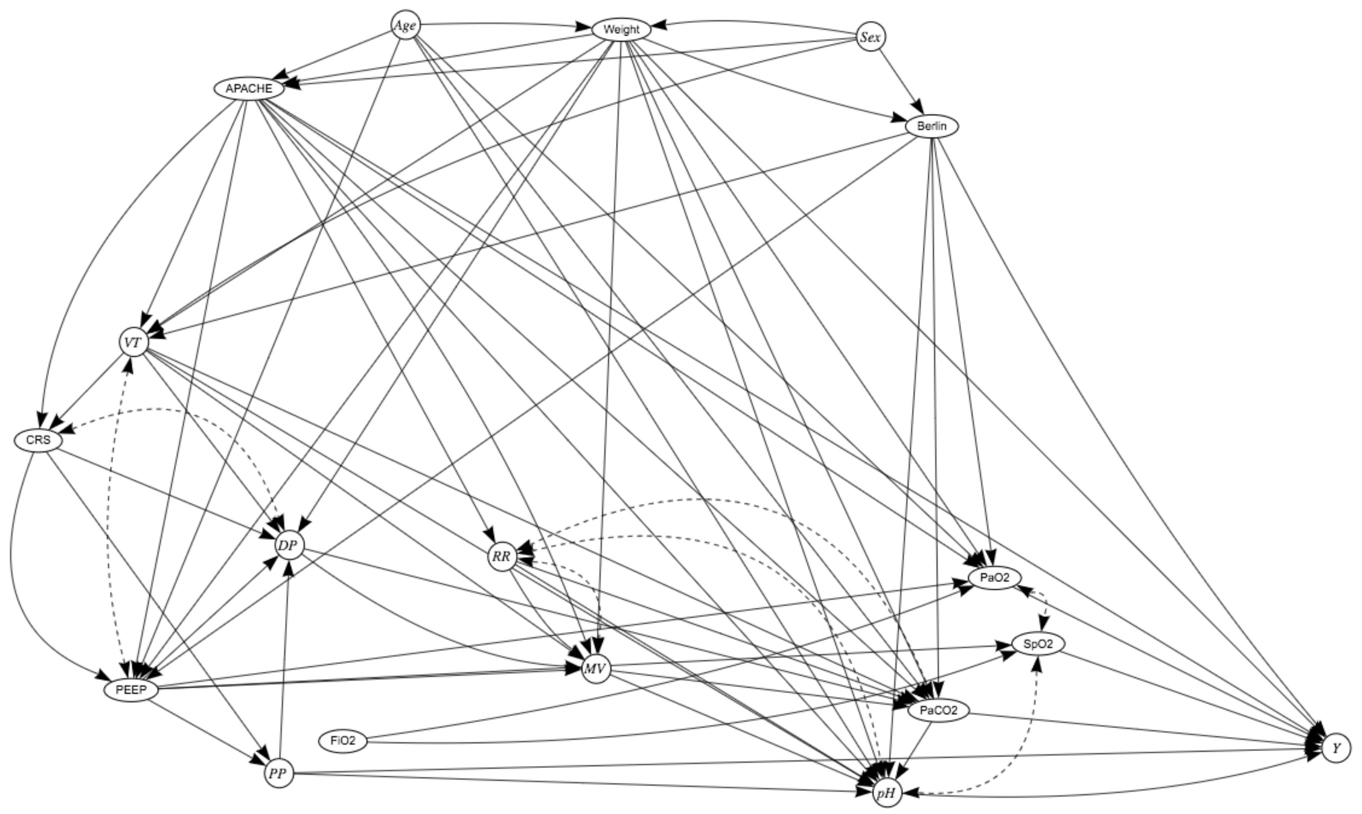
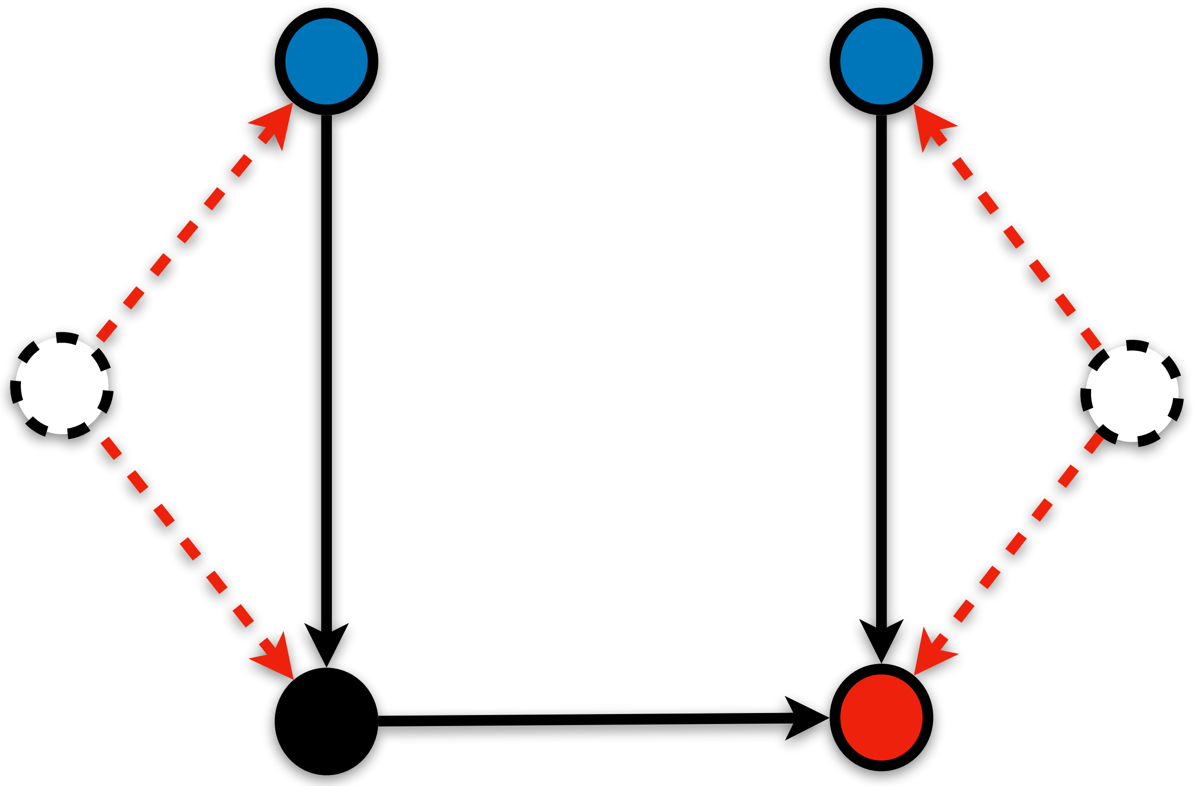
...

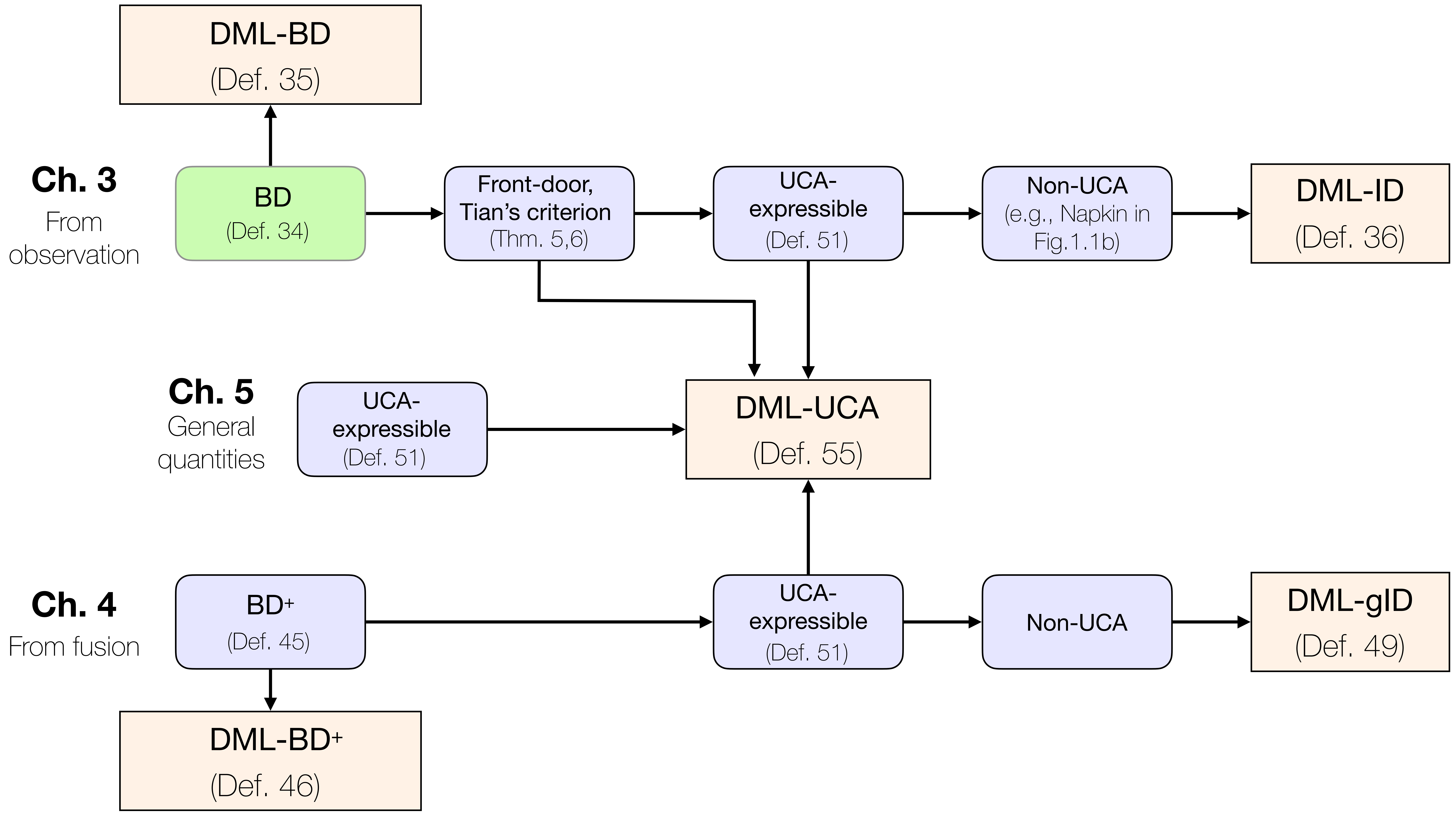
Solution

DML-ID

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

This Talk: Estimating Causal Effects

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



DML-BD
(Def. 35)
Ch. 3
From observation
BD
(Def. 34)
Front-door, Tian's criterion
(Thm. 5,6)
UCA-expressible
(Def. 51)
Non-UCA
(e.g., Napkin in Fig.1.1b)
DML-ID
(Def. 36)
Ch. 5
General quantities
UCA-expressible
(Def. 51)
DML-UCA
(Def. 55)
Ch. 4
From fusion
BD+
(Def. 45)
UCA-expressible
(Def. 51)
Non-UCA
DML-gID
(Def. 49)
DML-BD+
(Def. 46)

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.

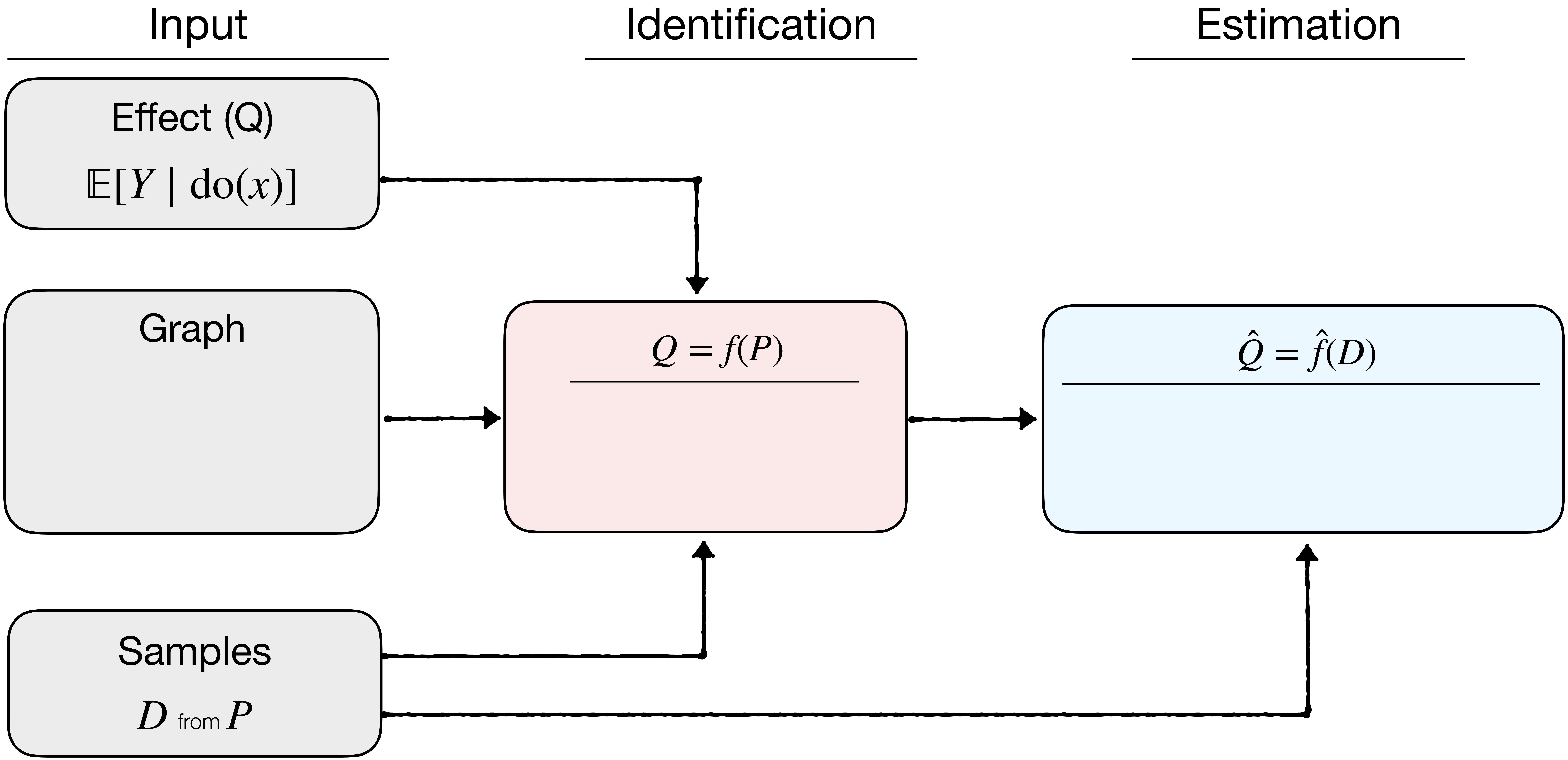
I kindly ask that you complete these by **June 12** to meet the PhD completion deadline for my next job appointment — Assistant Professor at UIUC's School of Information Sciences.

Logistics

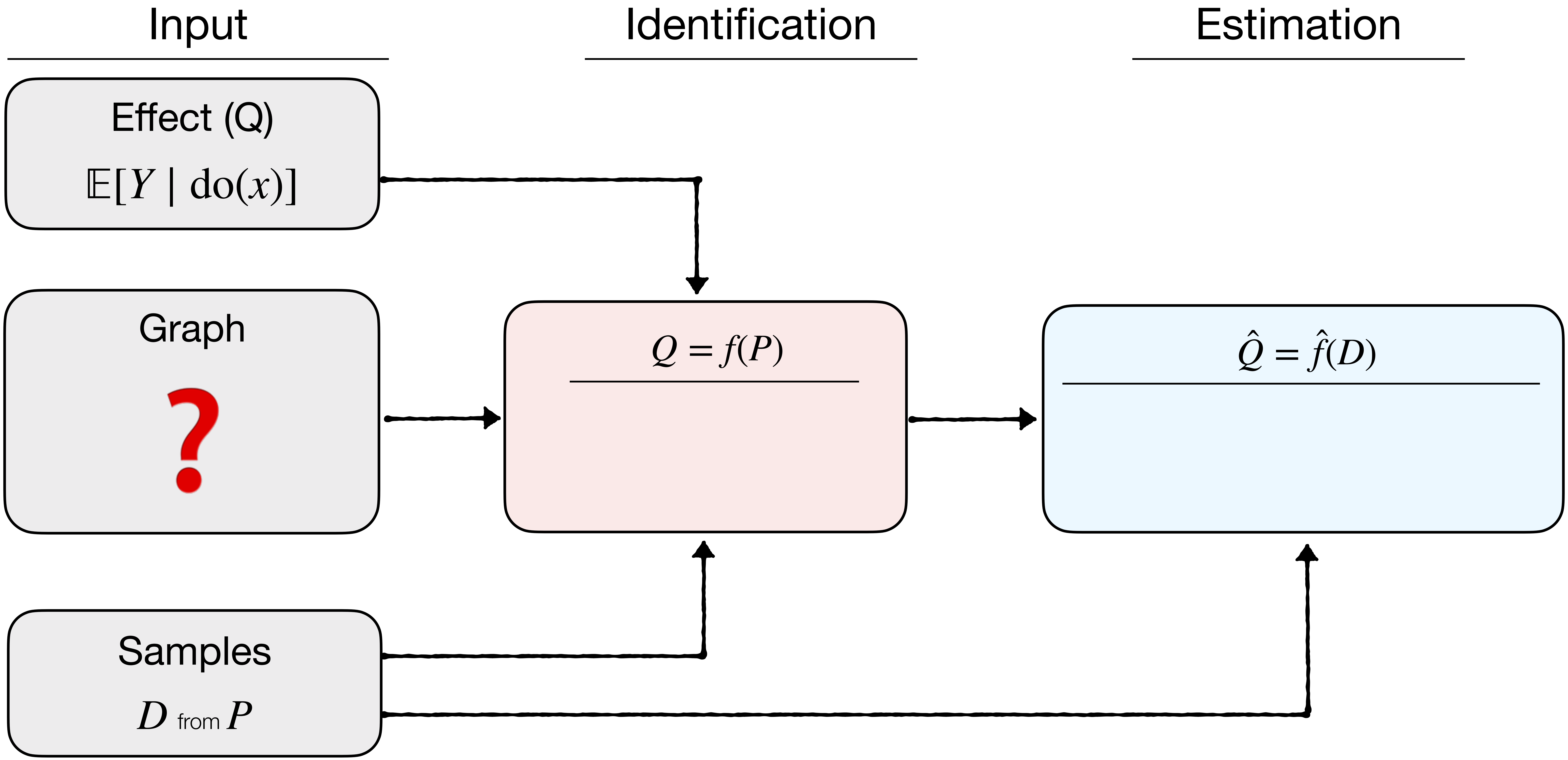
- Please initiate and sign "Form 11: Report of the Final Examination"
- Please approve the "Form 9: Electronic Thesis Acceptance Form (ETAF)" after reviewing the thesis.

Omitted Works

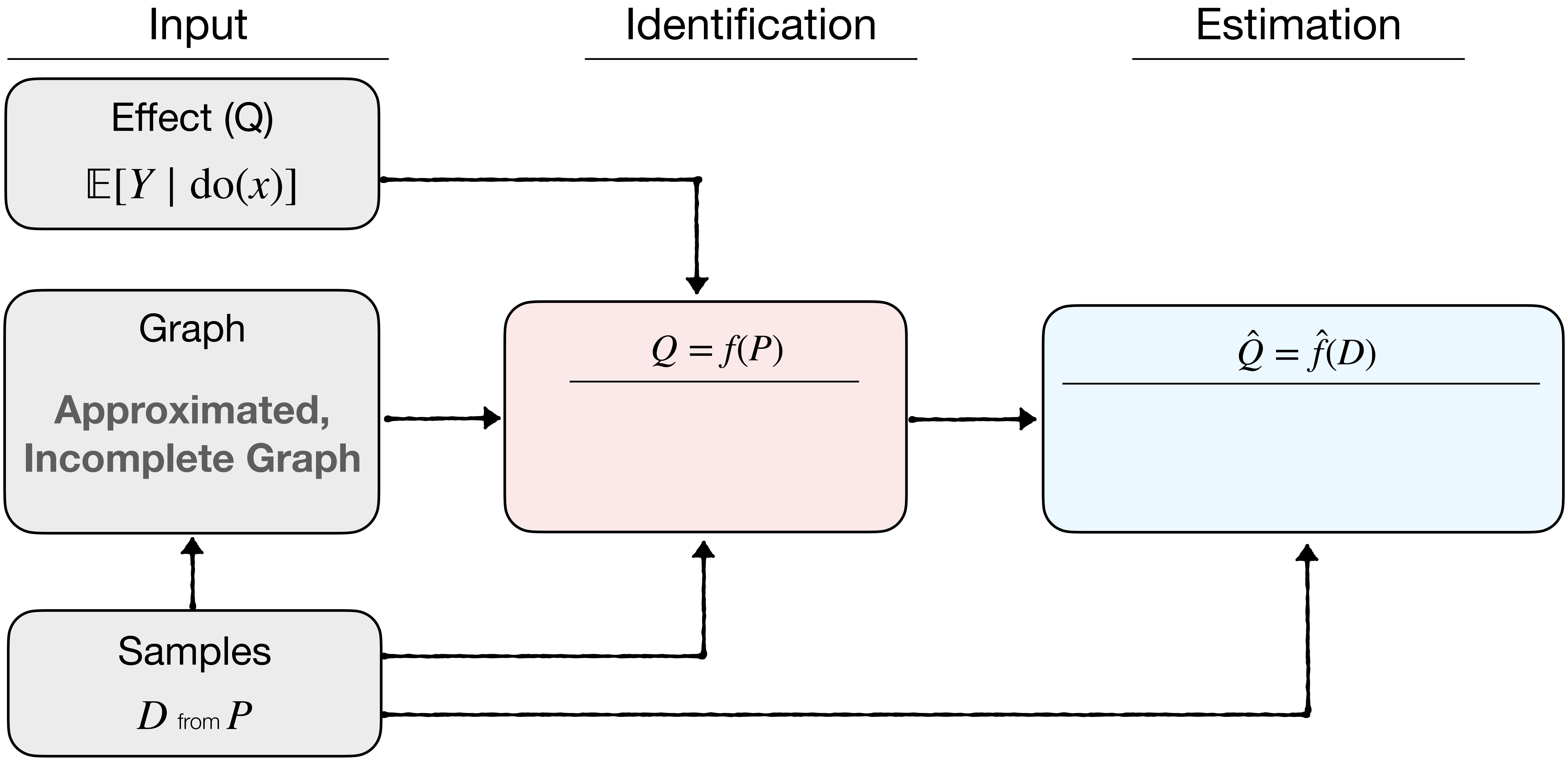
Other Work 1: Causal inference Without Graphs



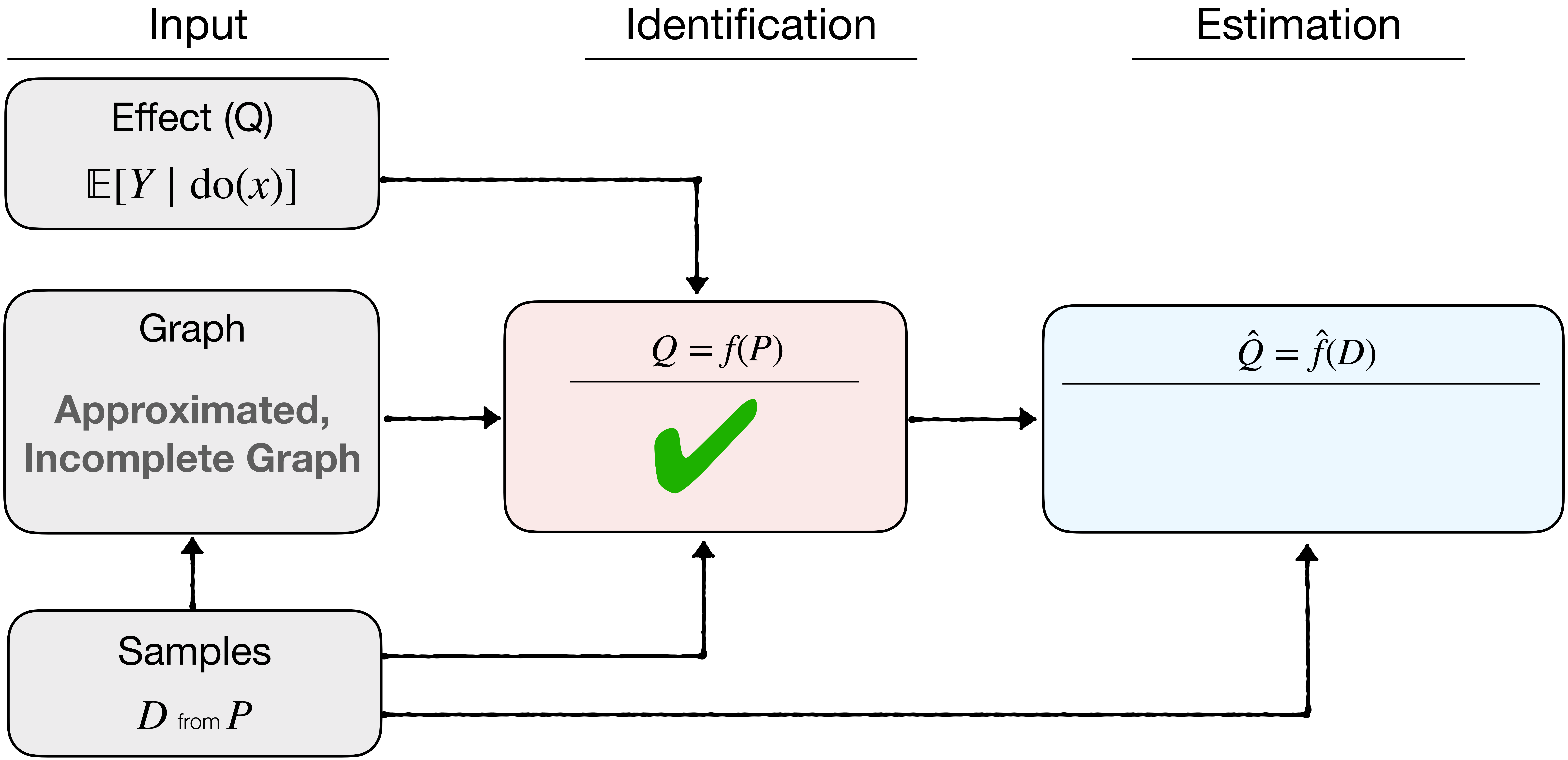
Other Work 1: Causal inference Without Graphs



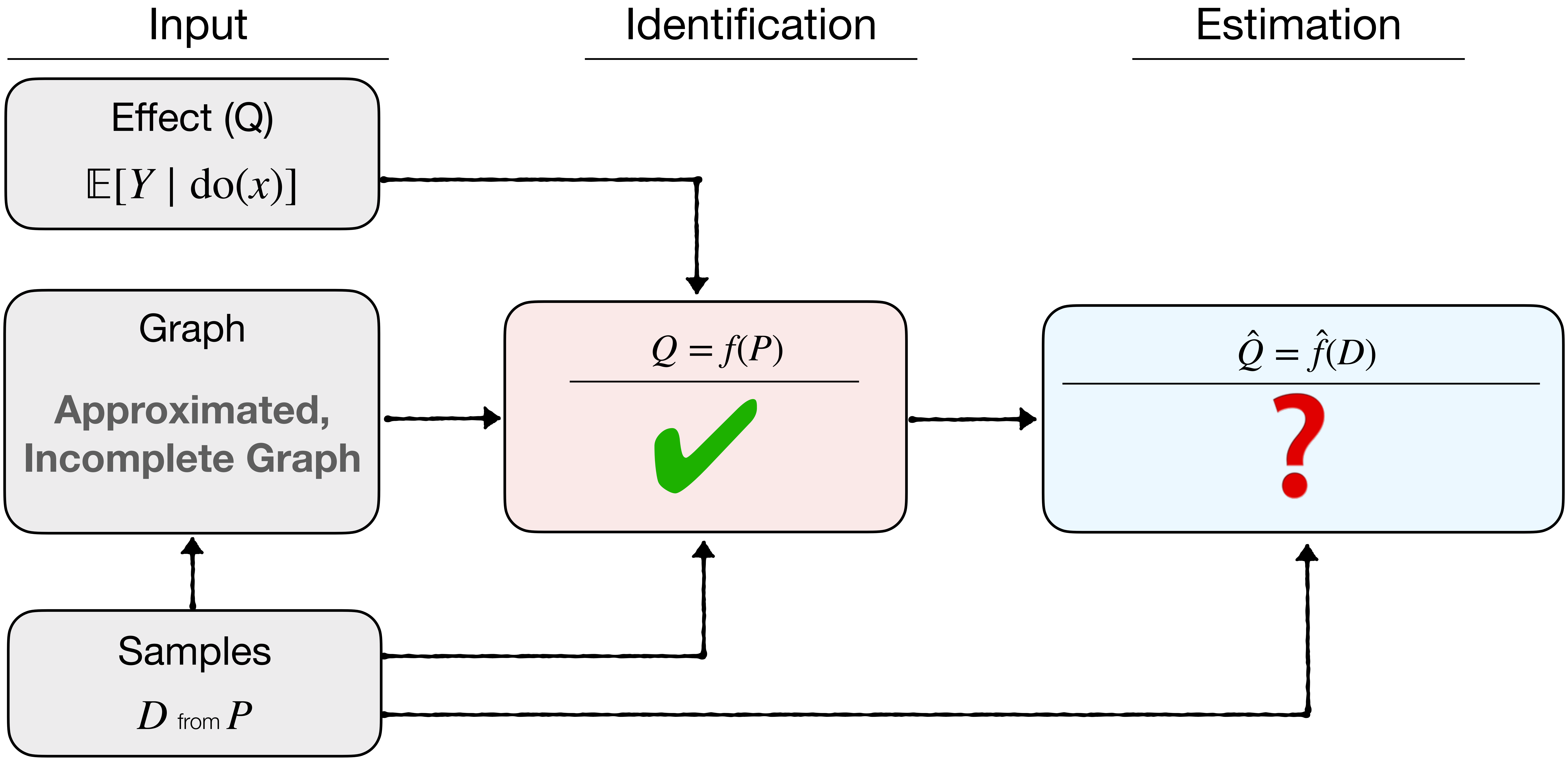
Other Work 1: Causal inference Without Graphs



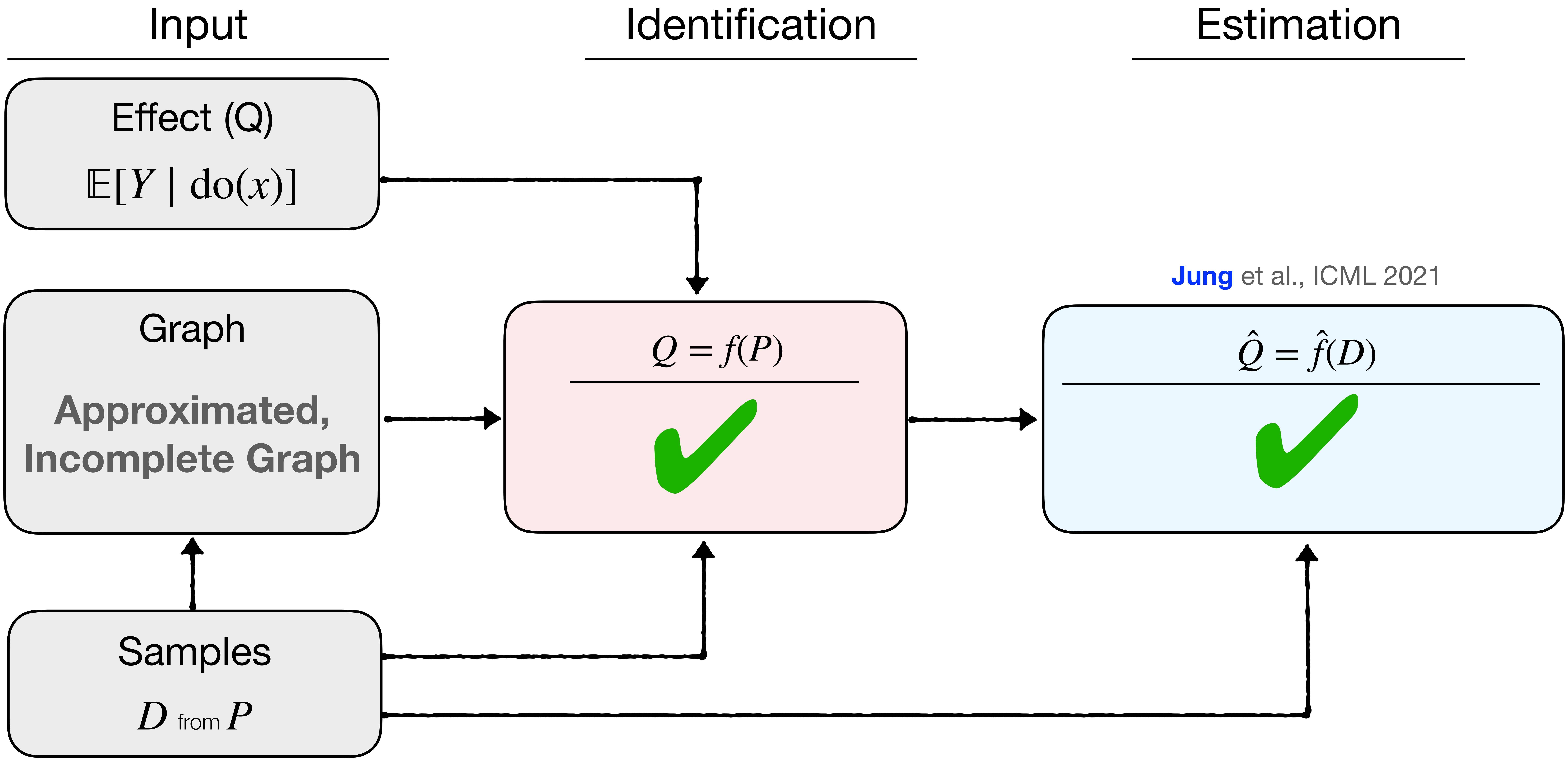
Other Work 1: Causal inference Without Graphs



Other Work 1: Causal inference Without Graphs



Other Work 1: Causal inference Without Graphs



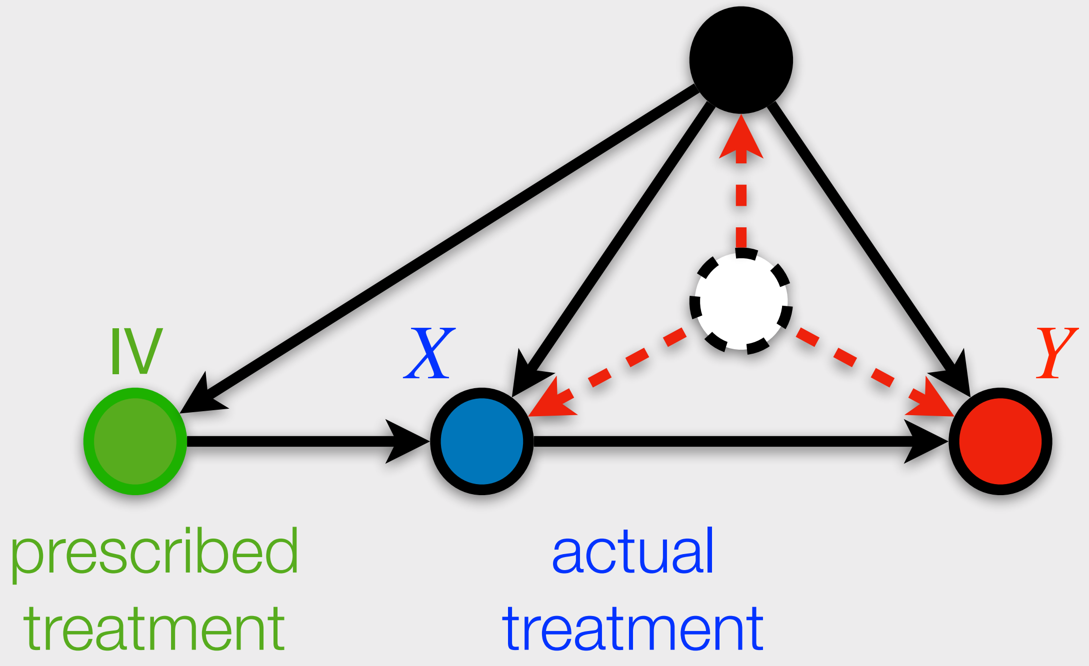
Other Work 2: Instrumental Variable (IV) Analysis

Input | Identification | Estimation

Effect (Q)

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

Graph



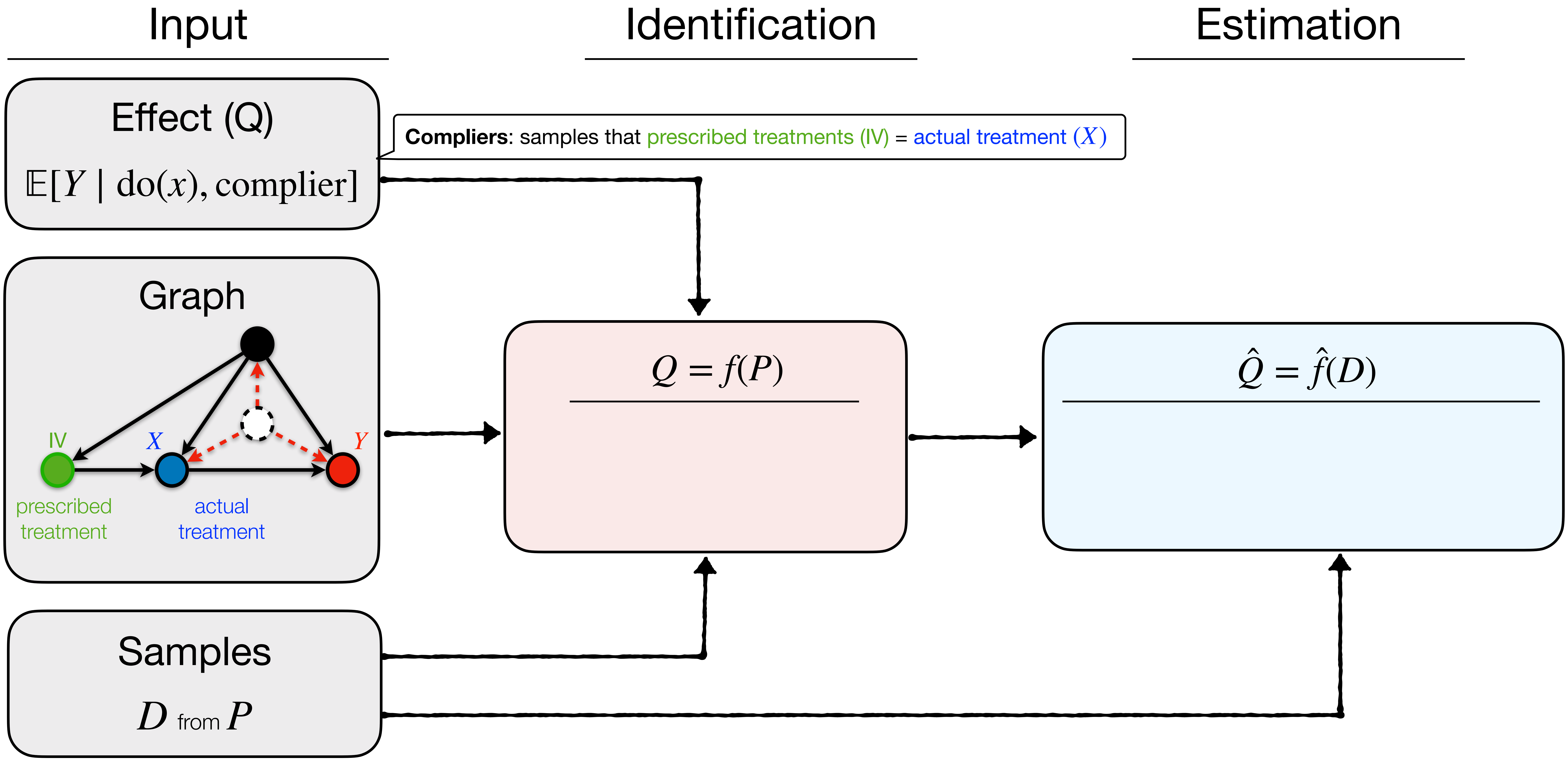
Samples

D from P

$Q = f(P)$

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

Other Work 2: Instrumental Variable (IV) Analysis



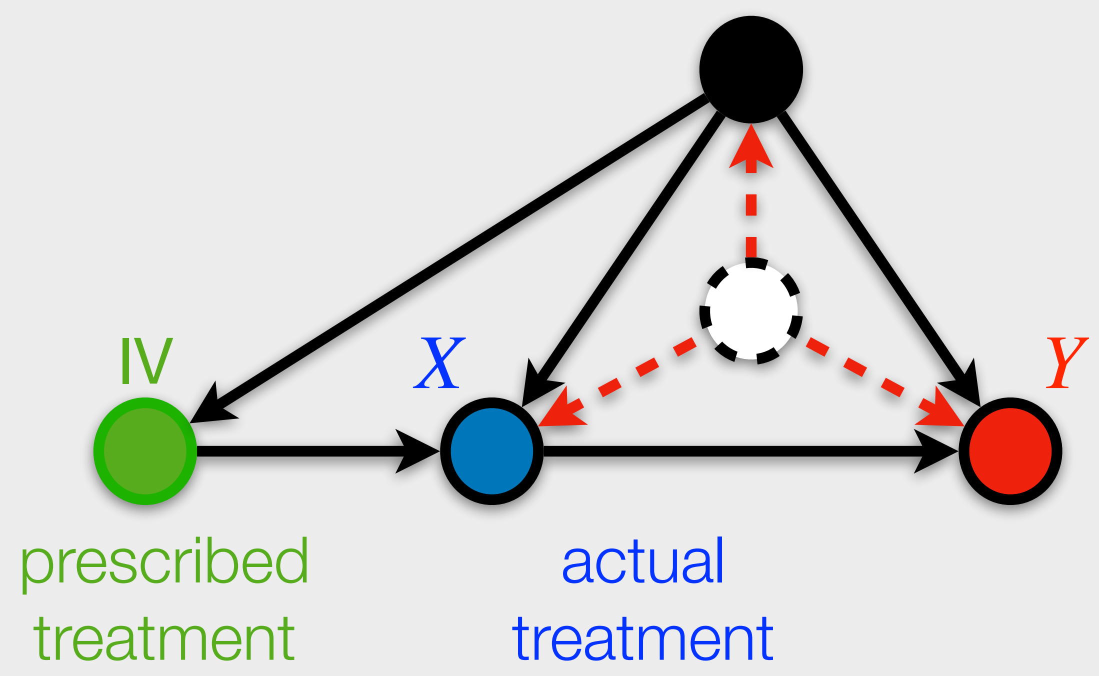
Other Work 2: Instrumental Variable (IV) Analysis

Input

Effect (Q)

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

Graph



Samples

D from P

Identification

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

$Q = f(P)$

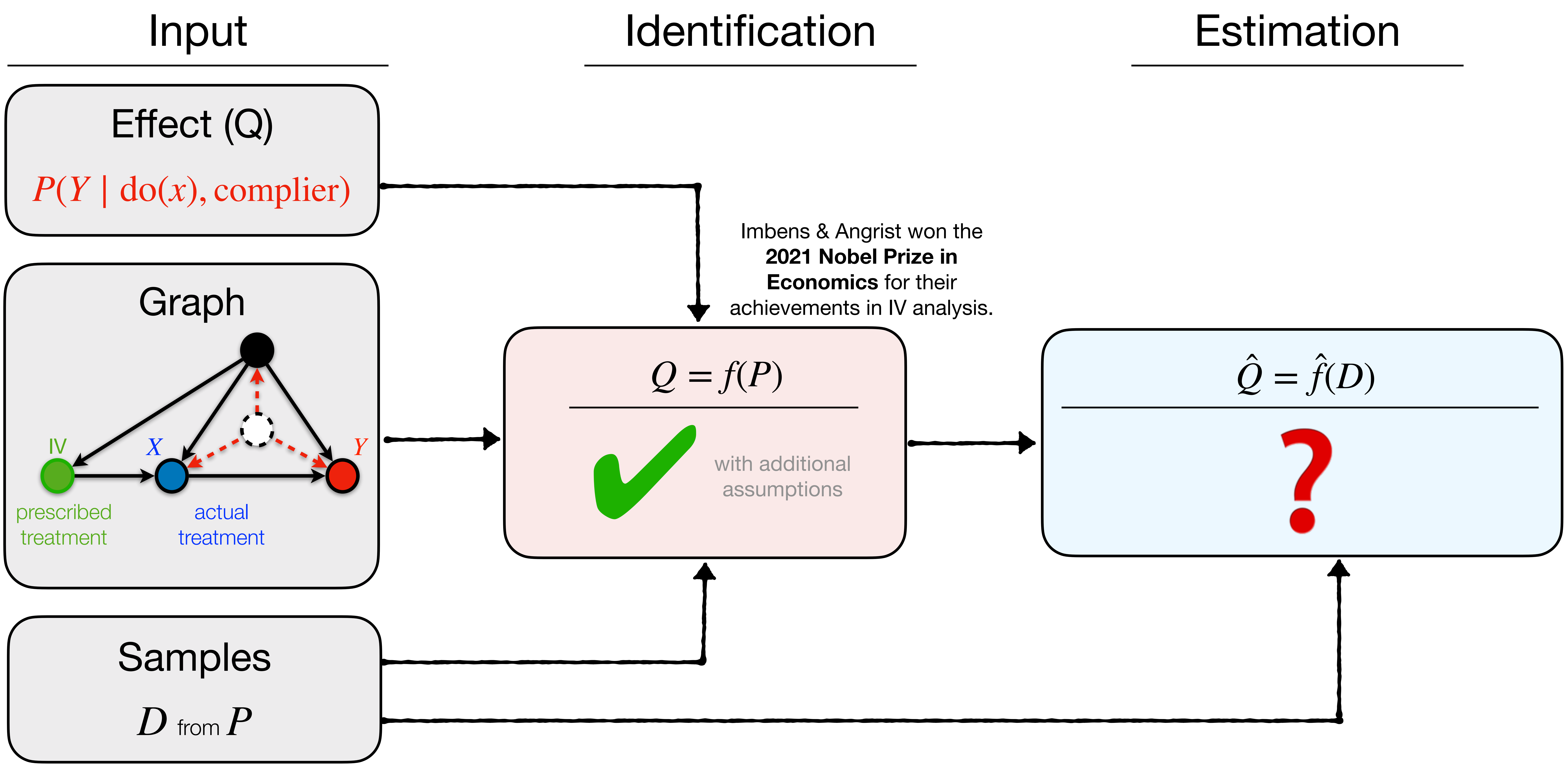
✔ with additional assumptions

Estimation

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

✔

Other Work 2: Instrumental Variable (IV) Analysis



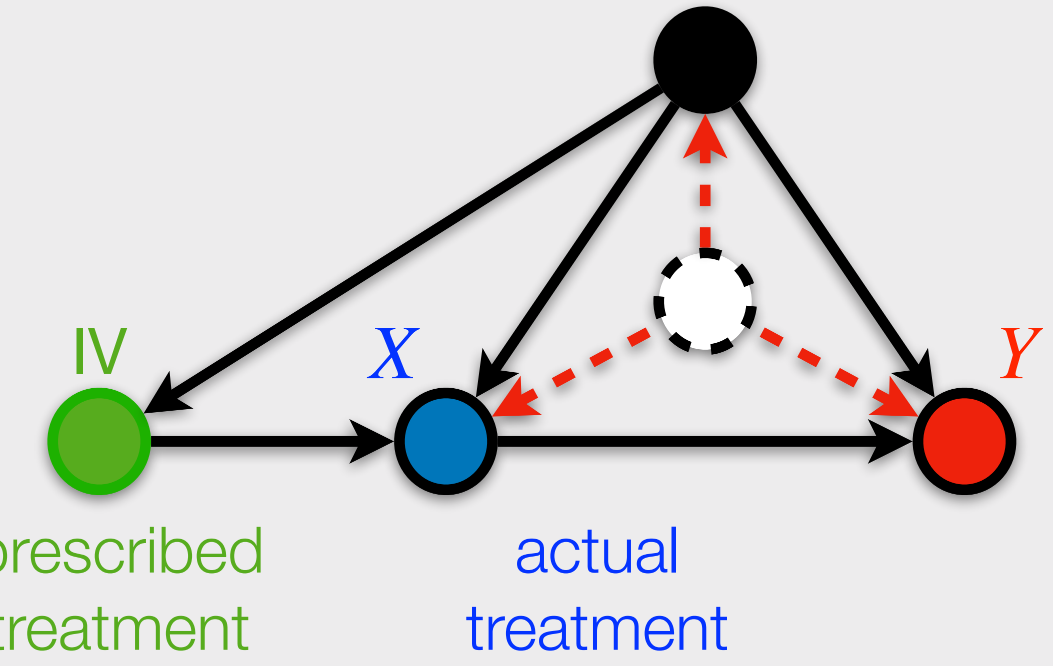
Other Work 2: Instrumental Variable (IV) Analysis

Input

Effect (Q)

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

Graph



Samples

D from P

Identification

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

$$Q = f(P)$$

✔ with additional assumptions

Estimation

Jung et al., NeurIPS 2021 (Spotlight)

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

✔

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)

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

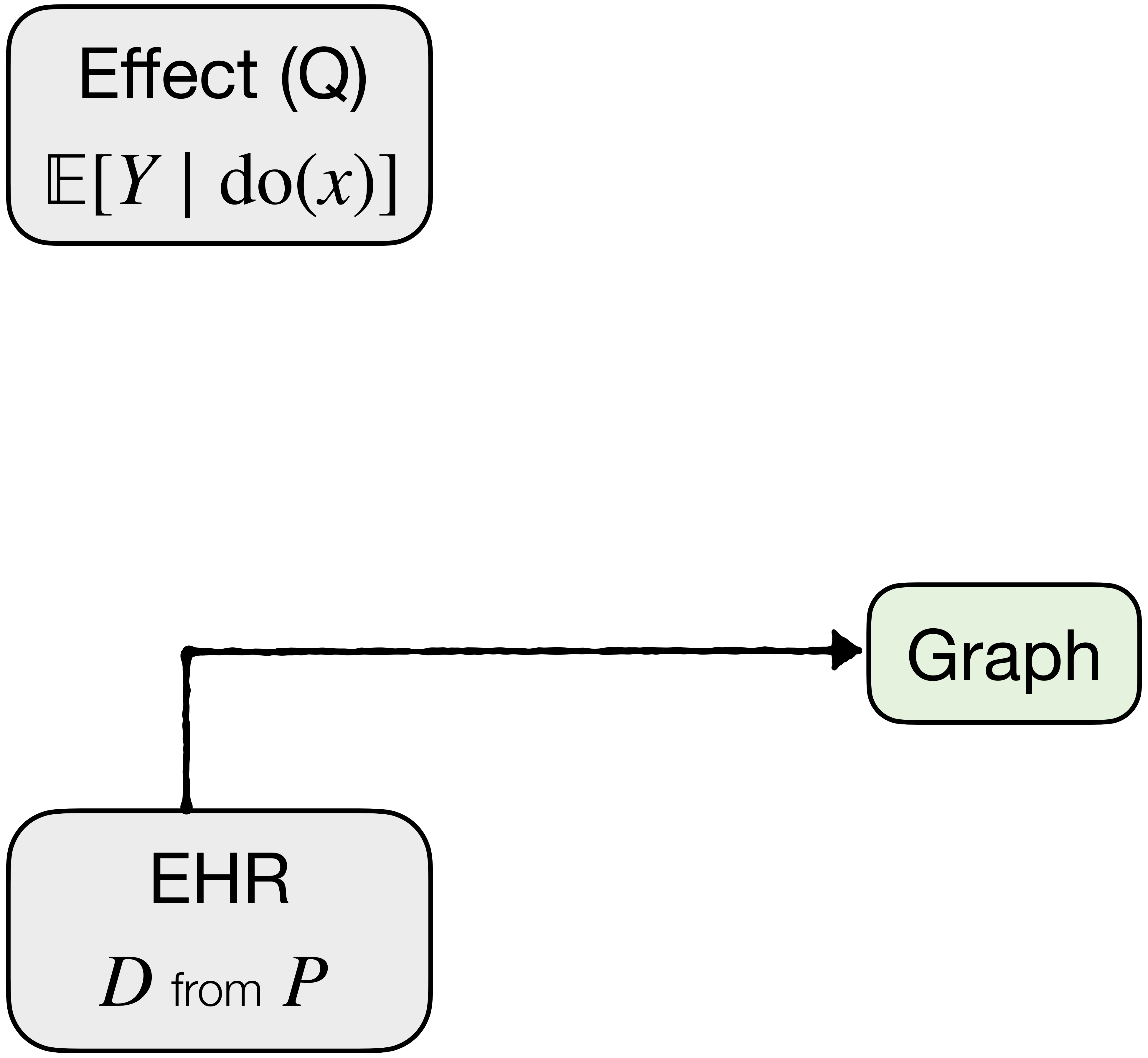
EHR

D from P

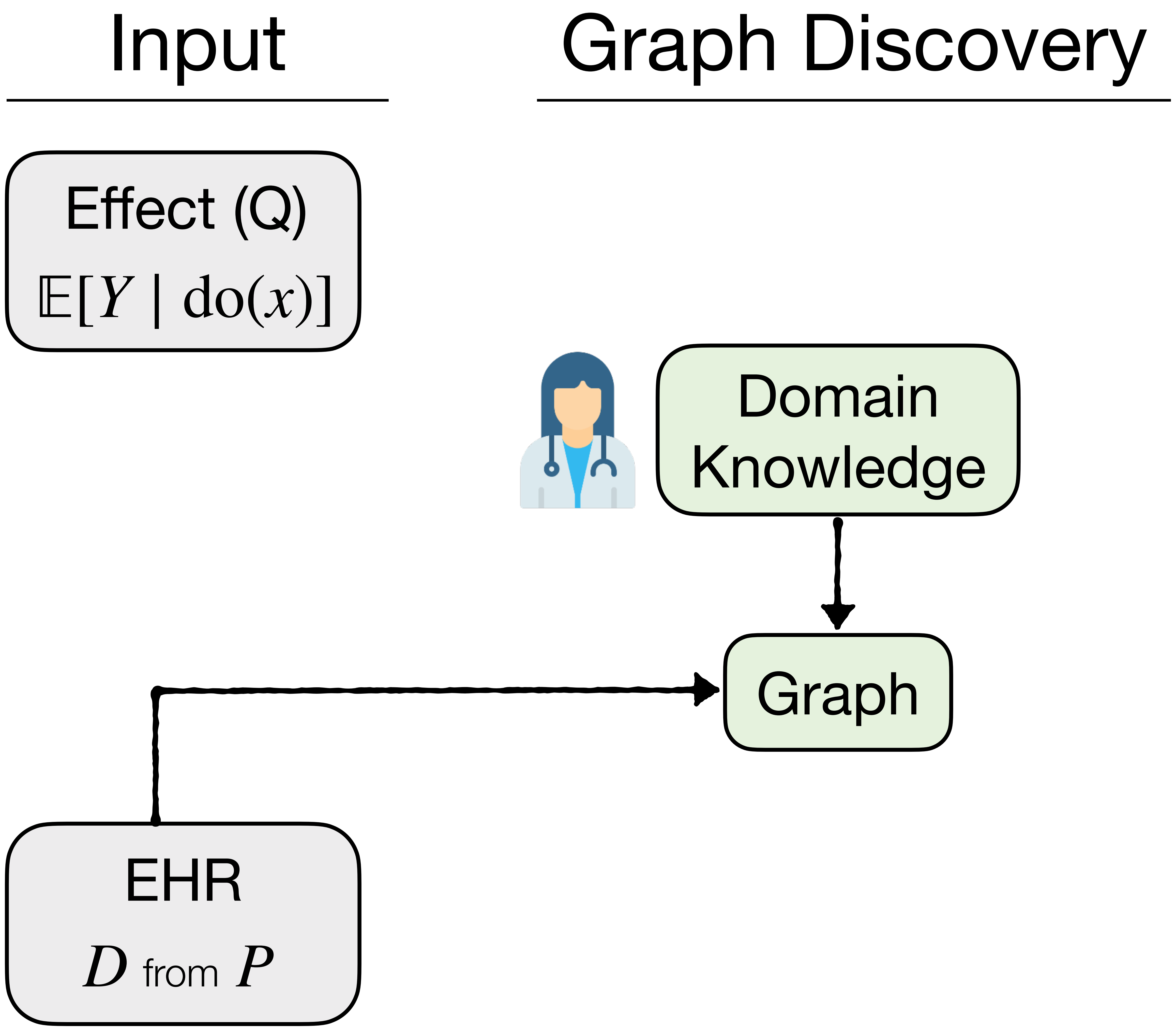
Application 1. Emulating RCT from EHR

Input

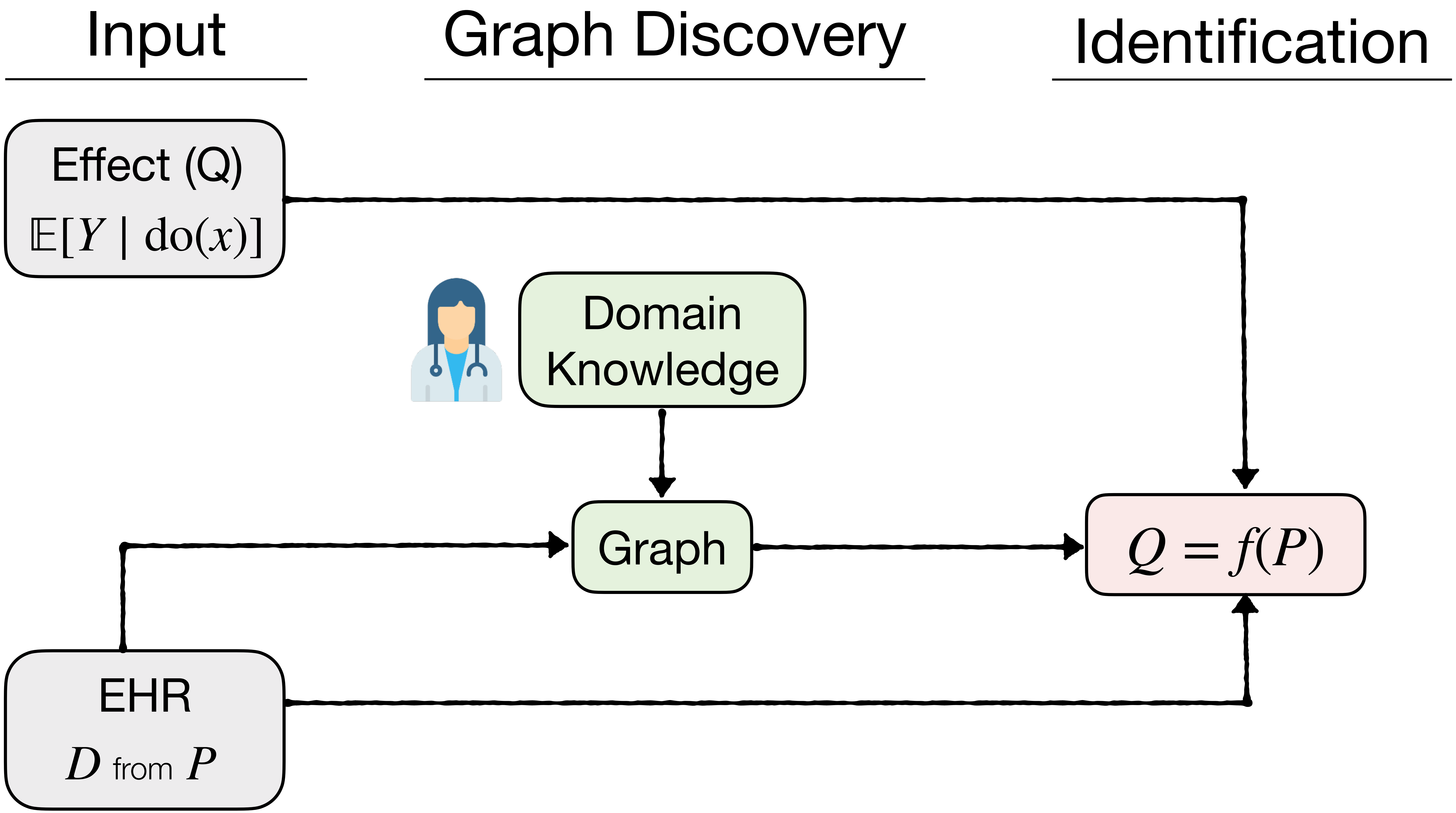
Graph Discovery



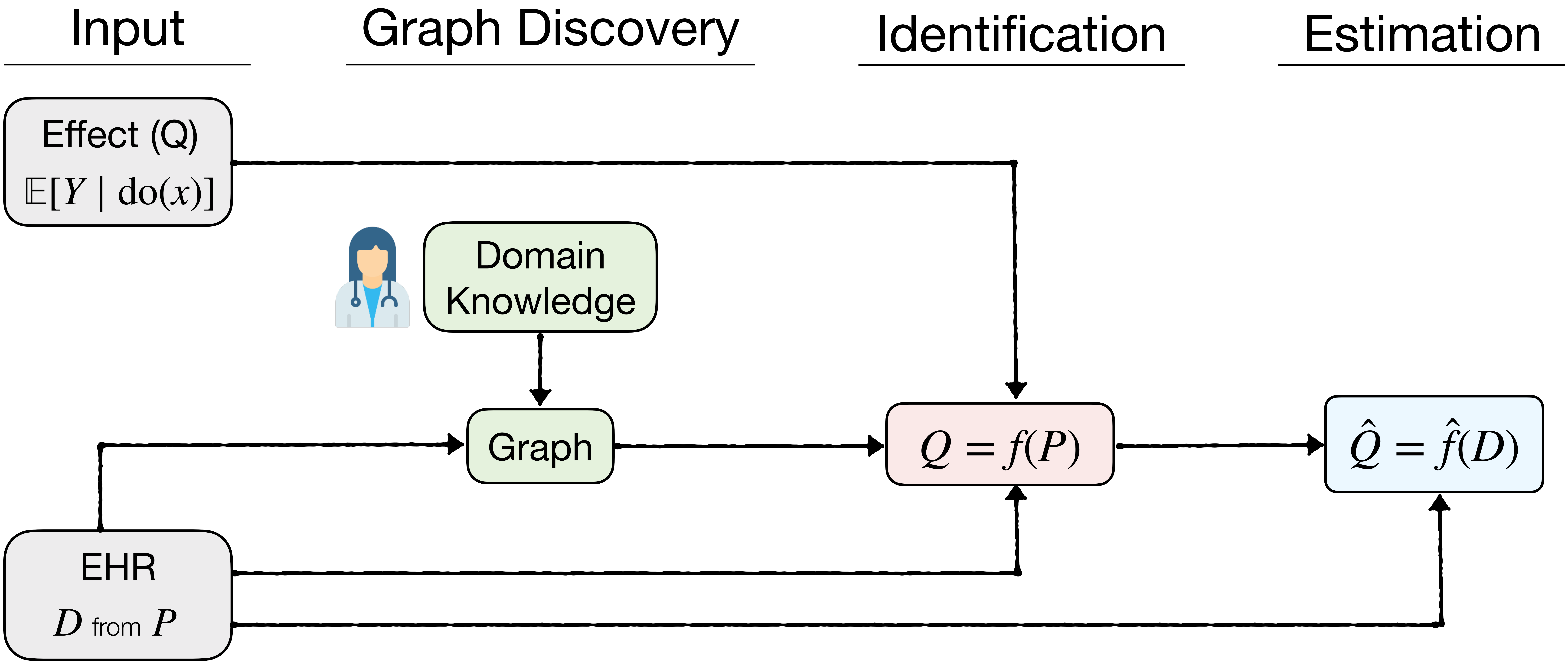
Application 1. Emulating RCT from EHR



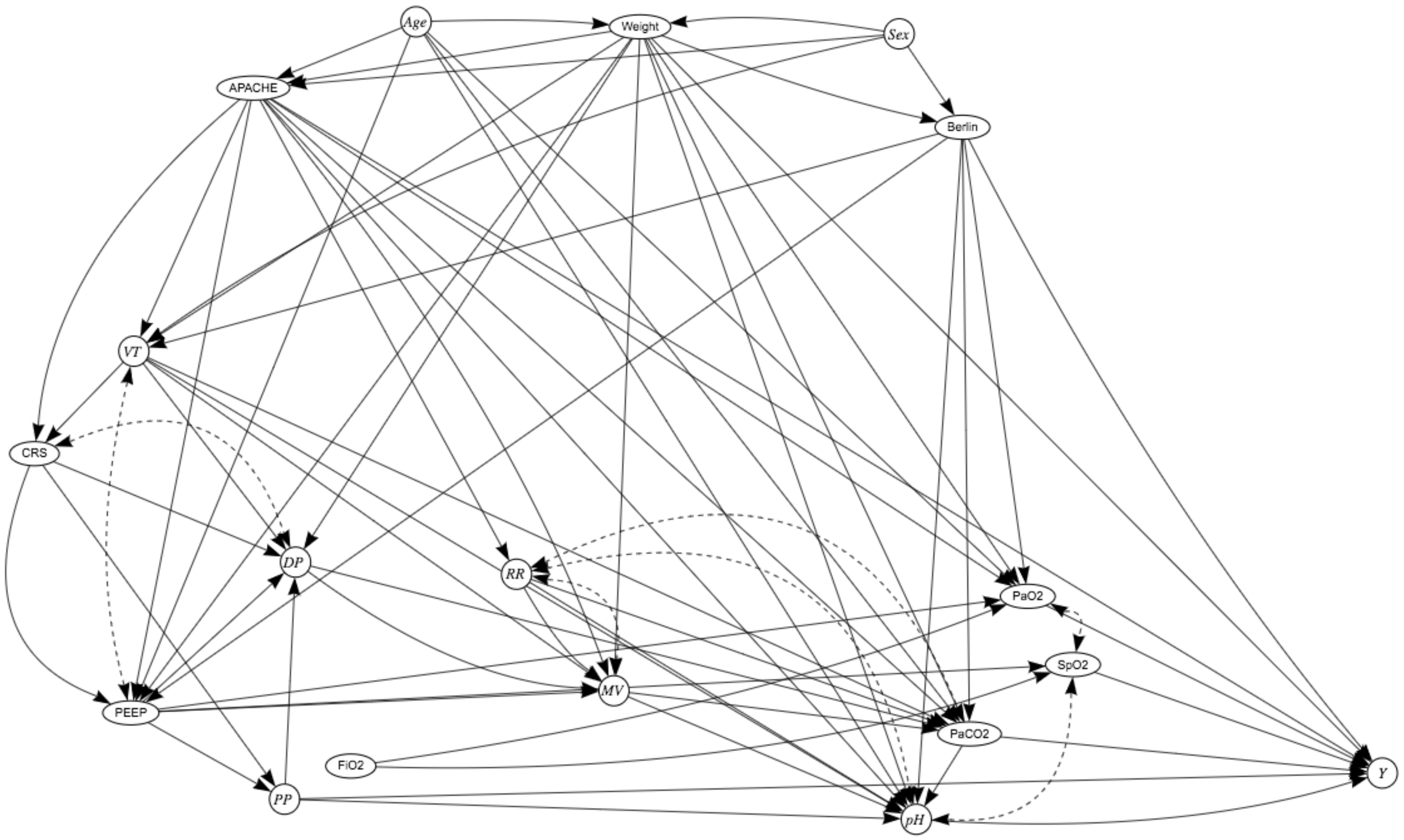
Application 1. Emulating RCT from EHR



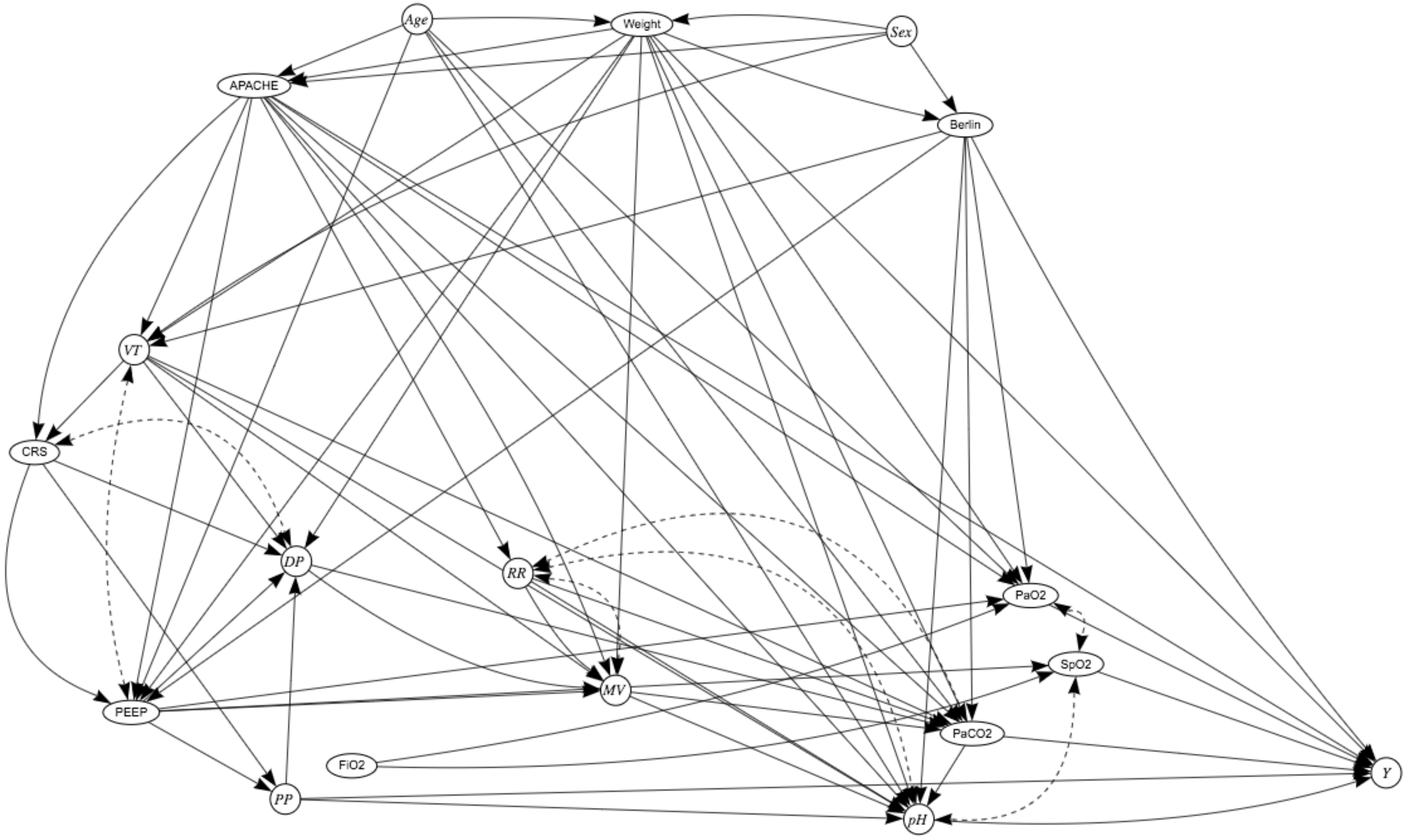
Application 1. Emulating RCT from EHR



Application 1. Emulating RCT from EHR



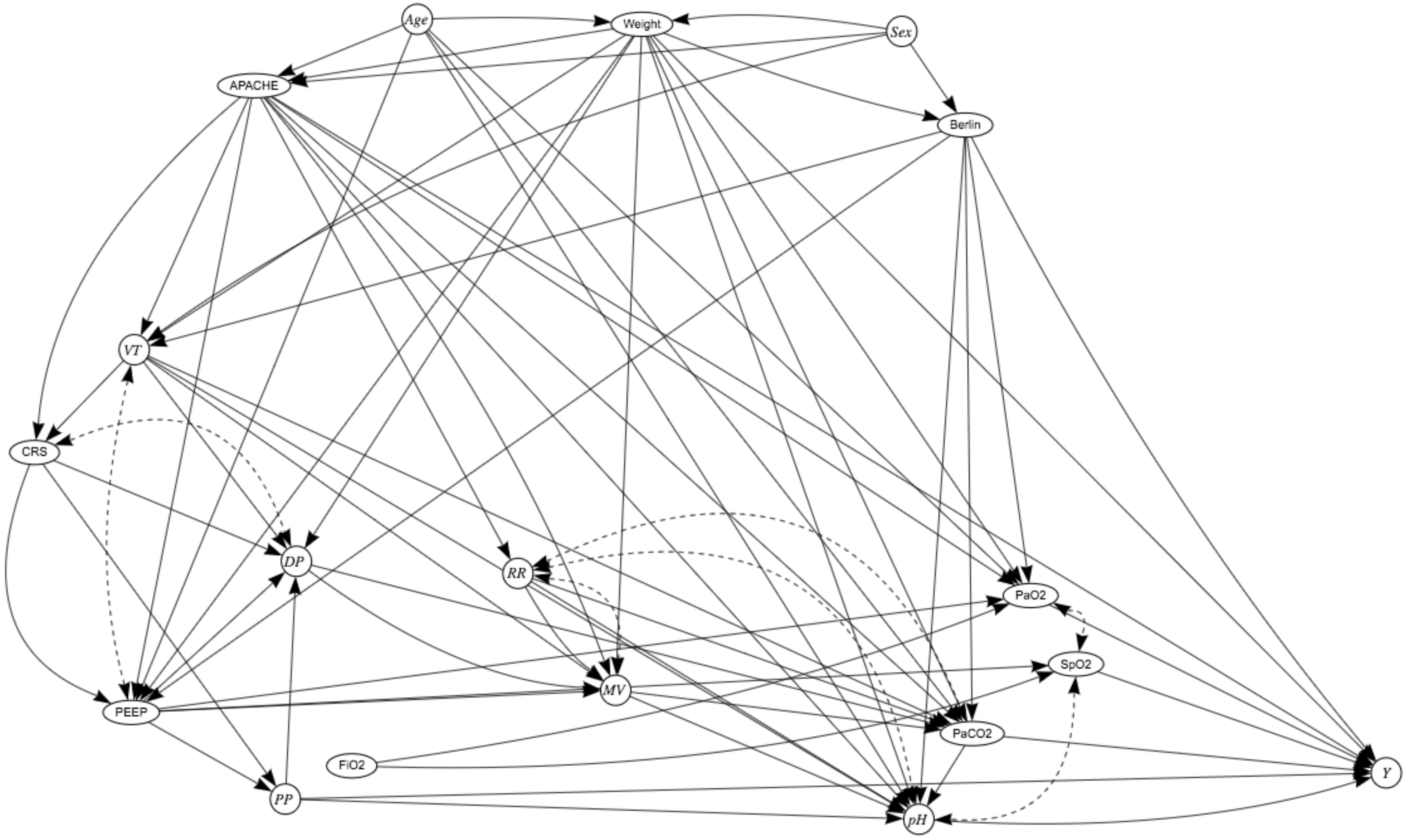
Application 1. Emulating RCT from EHR



Causal graph on Acute Respiratory Distress Syndrome (ARDS)

Application 1. Emulating RCT from EHR

Jung et al., American Thoracic Society, 2018



Causal graph on Acute Respiratory Distress Syndrome (ARDS)

Result

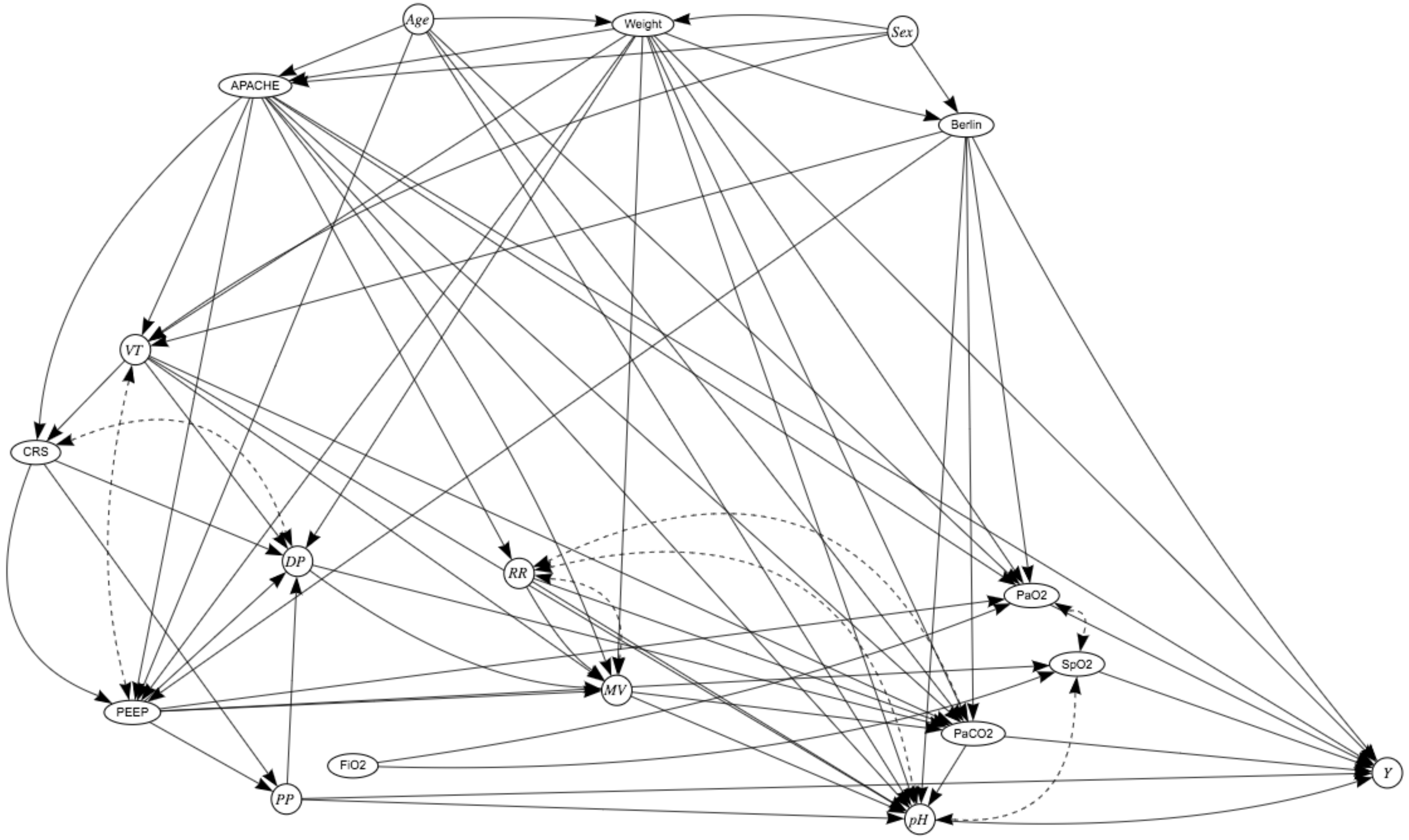
For seminal RCTs,

Our treatment recommendation

= Trials' treatment recommendation

Application 1. Emulating RCT from EHR

Jung et al., American Thoracic Society, 2018



Causal graph on Acute Respiratory Distress Syndrome (ARDS)

Result

For seminal RCTs,

Our treatment recommendation

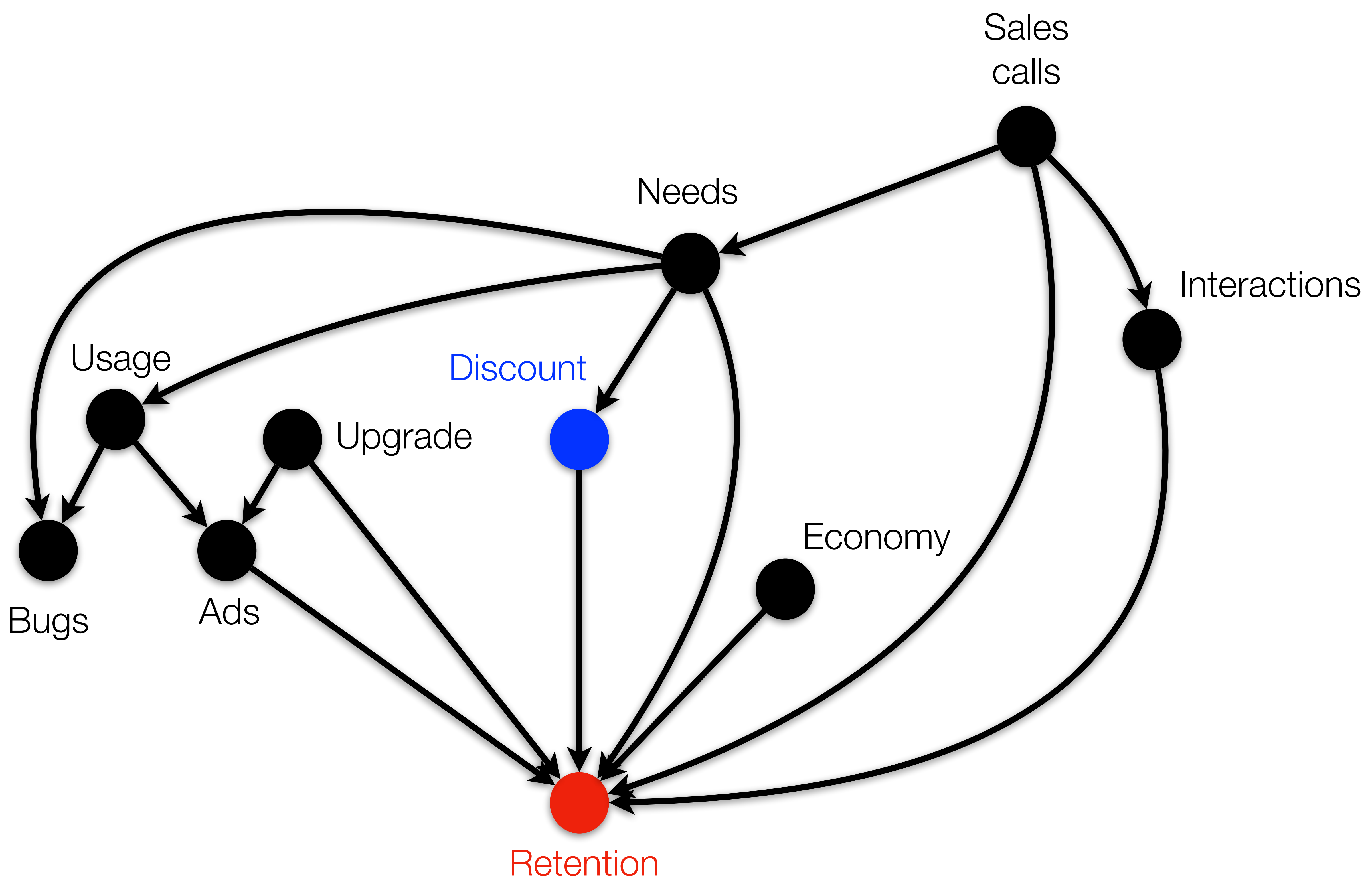
= Trials' treatment recommendation

Impact

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

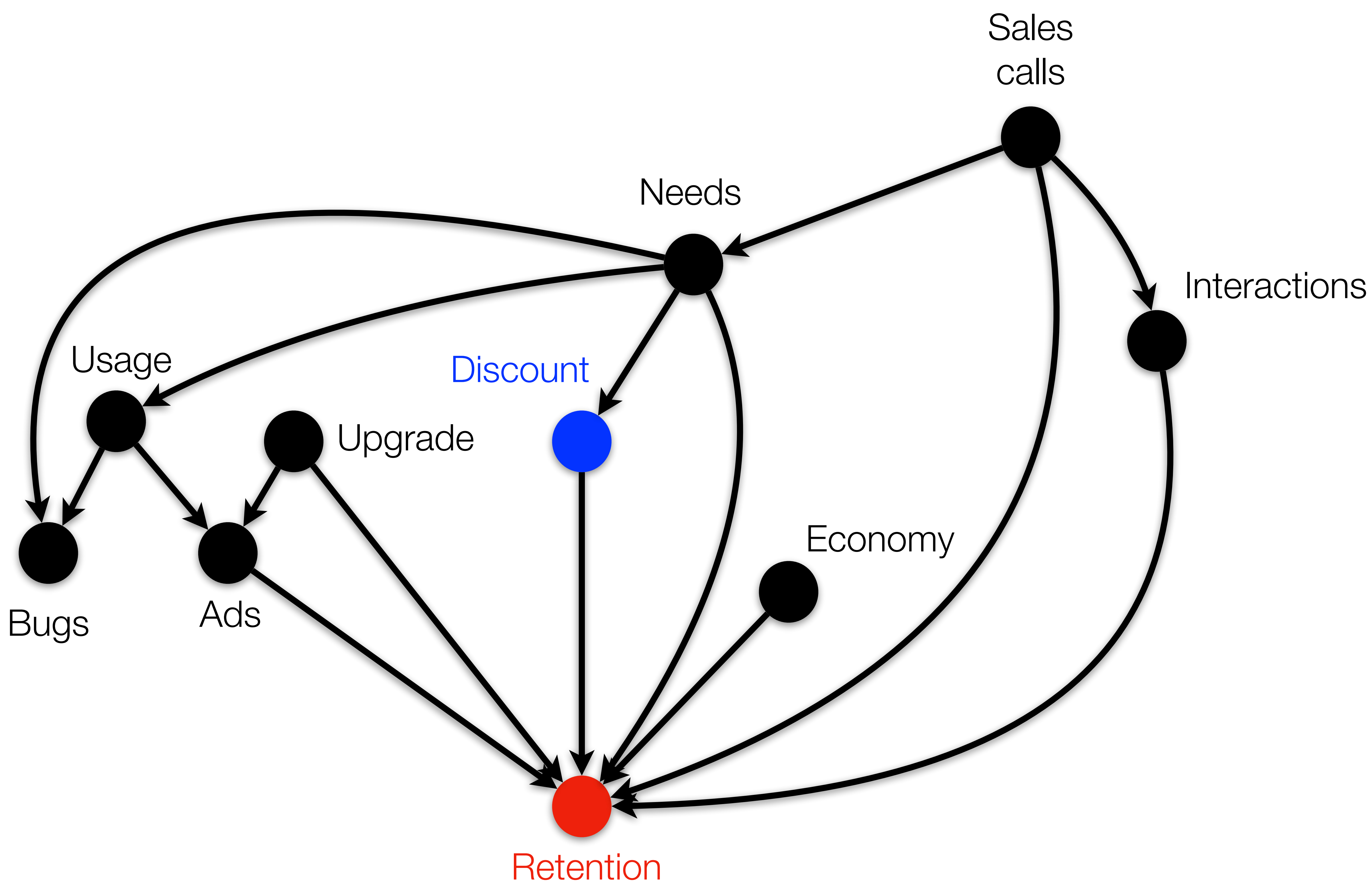
Application 2. Explainable AI

Application 2. Explainable AI



Contribution of Discount to the Retention?

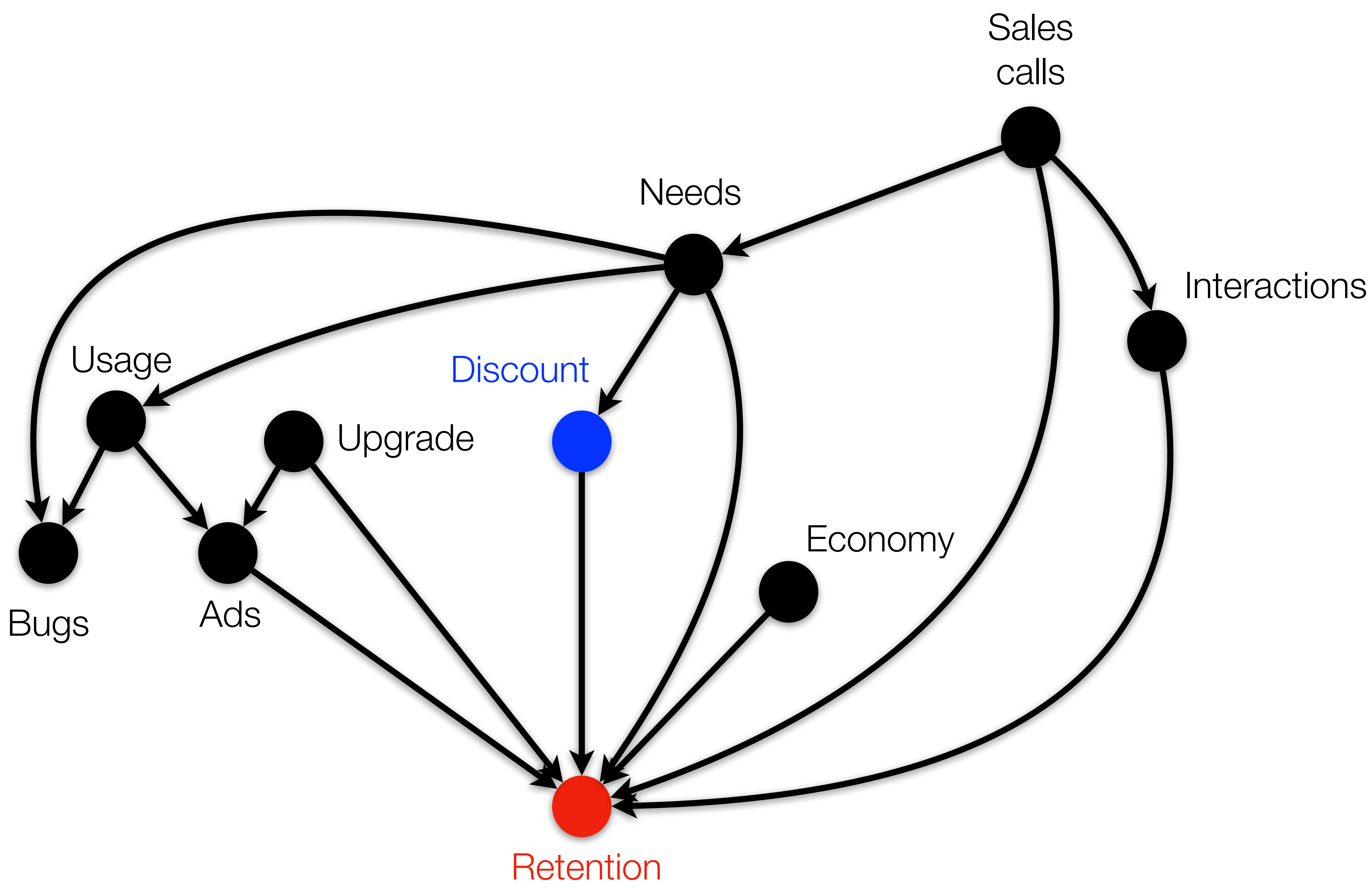
Application 2. Explainable AI



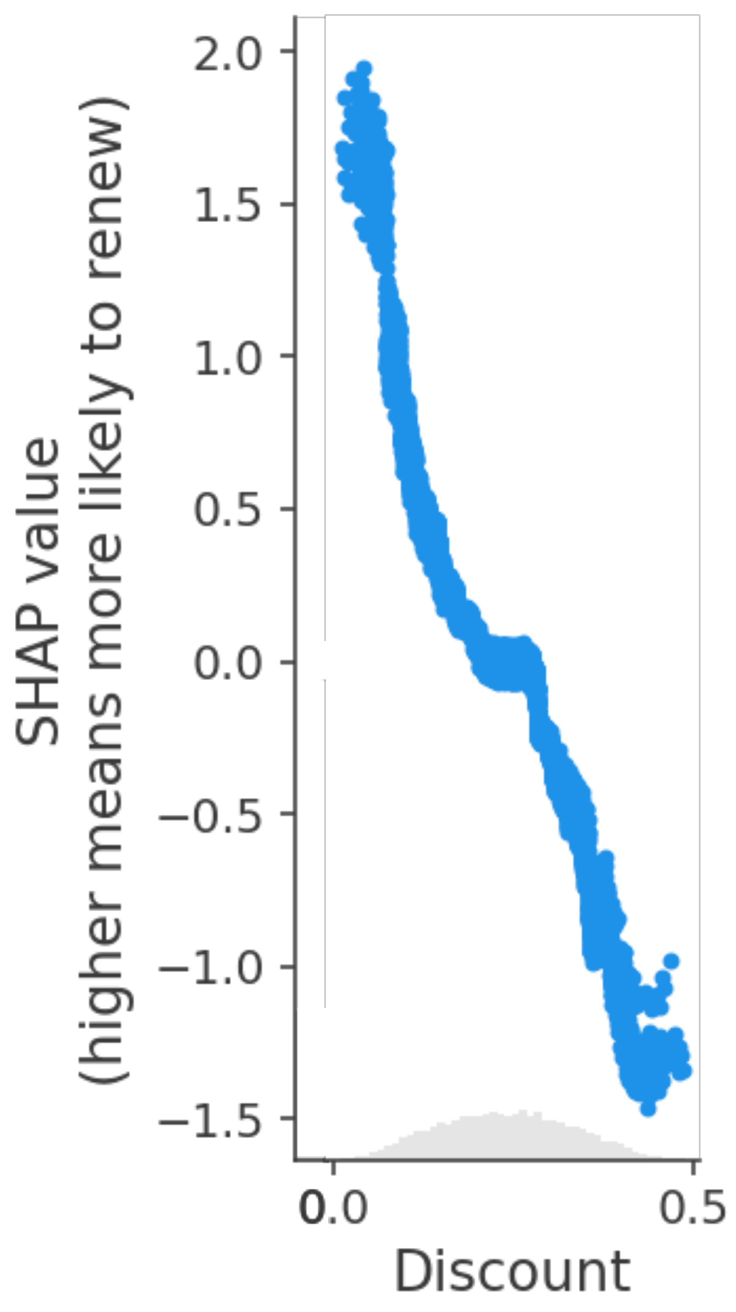
Contribution of Discount to the Retention?

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

Application 2. Explainable AI

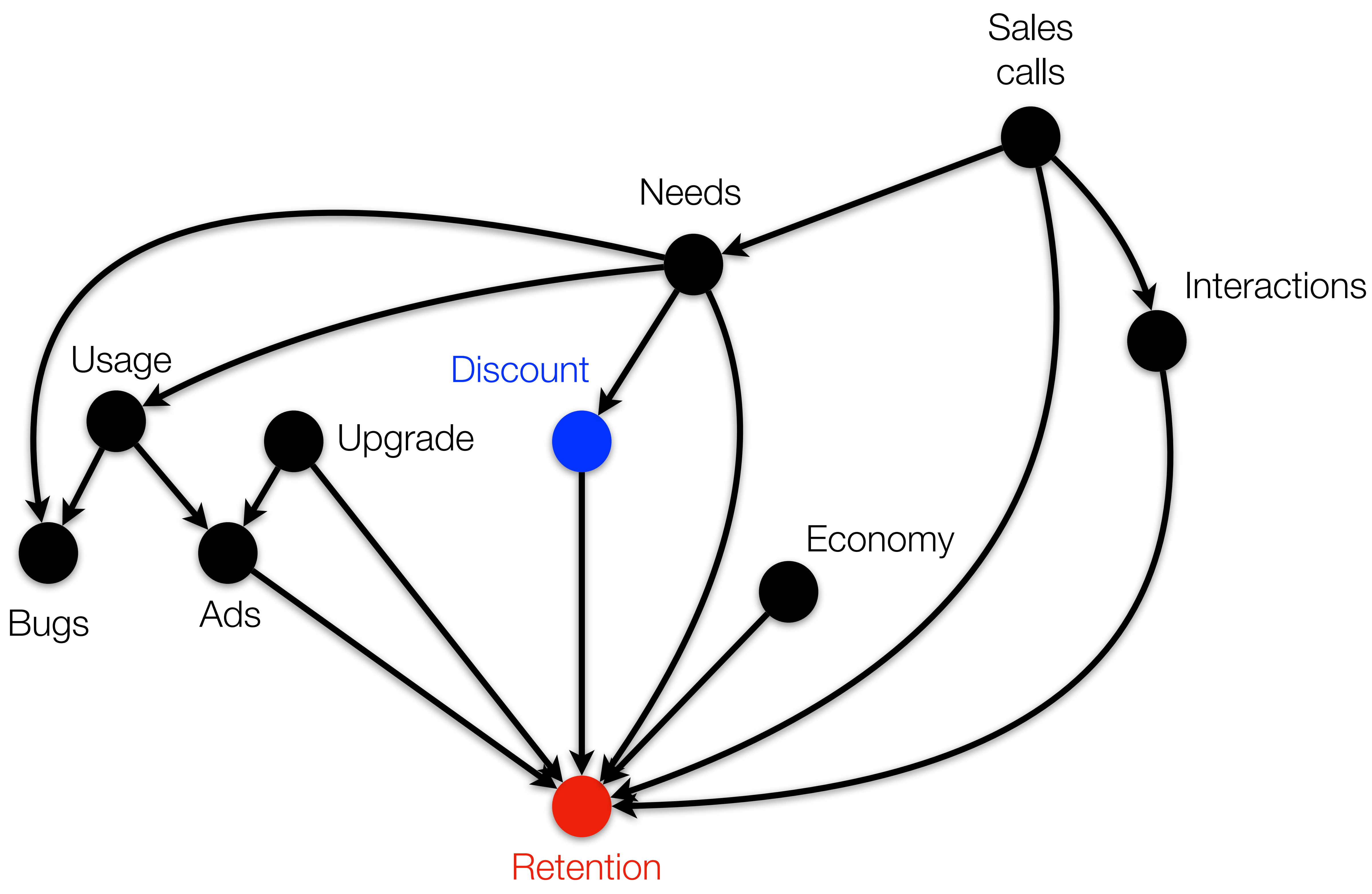


Contribution of Discount to the Retention?

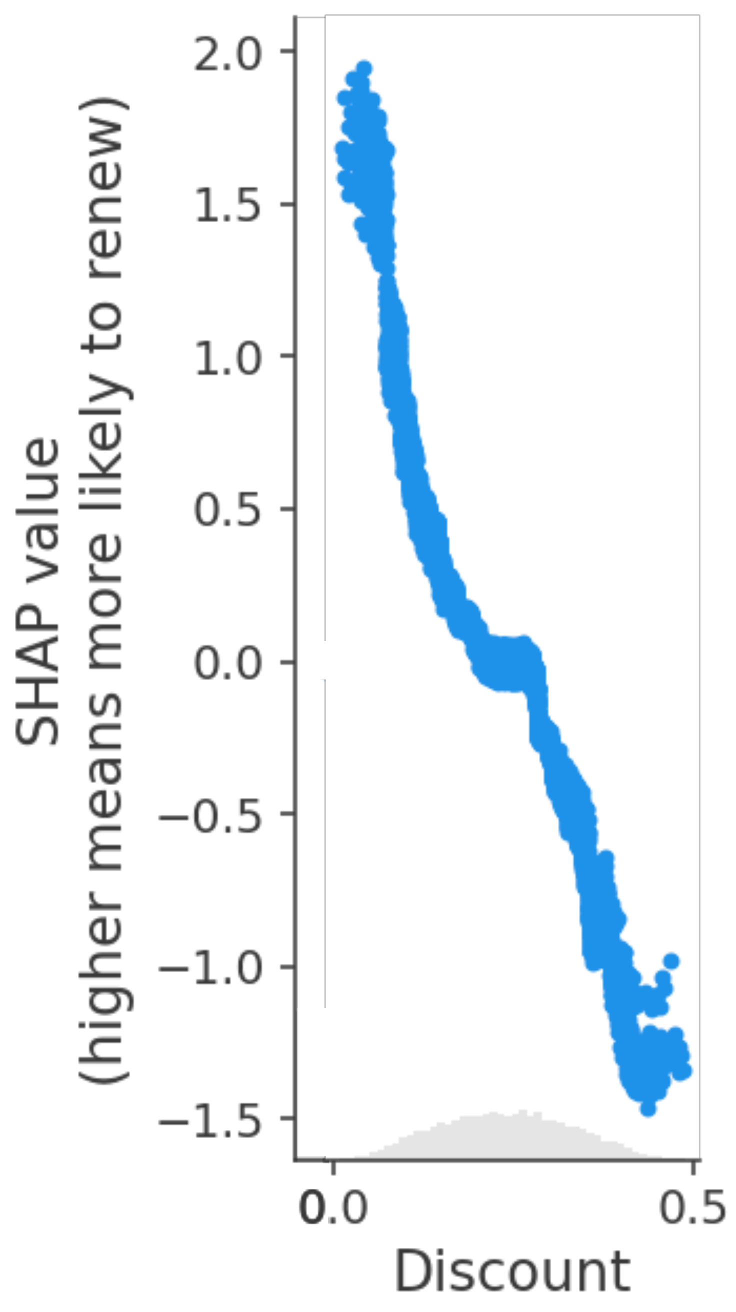


- SHAP value: one of the most cited measure for the feature importance

Application 2. Explainable AI

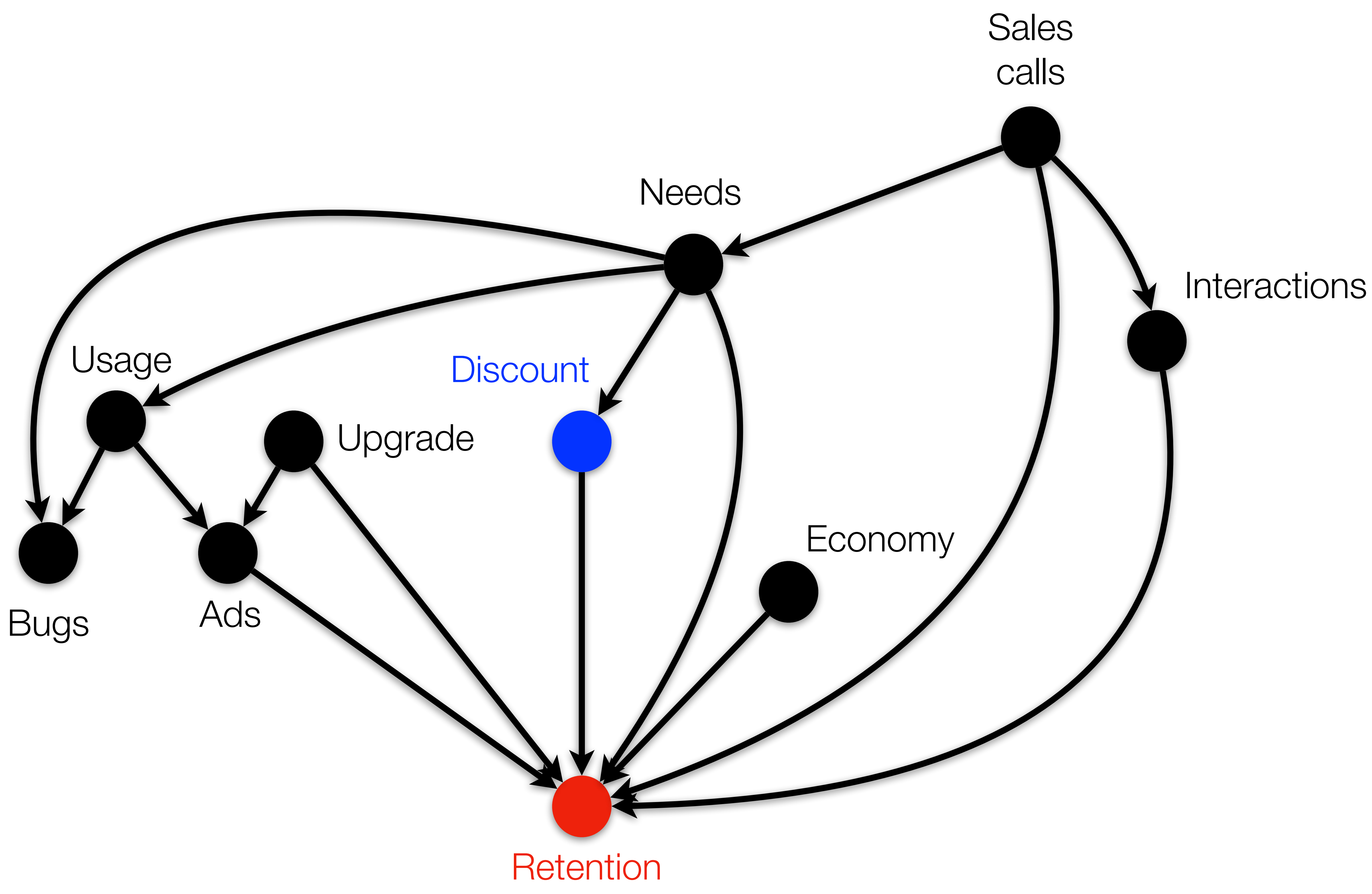


Contribution of Discount to the Retention?

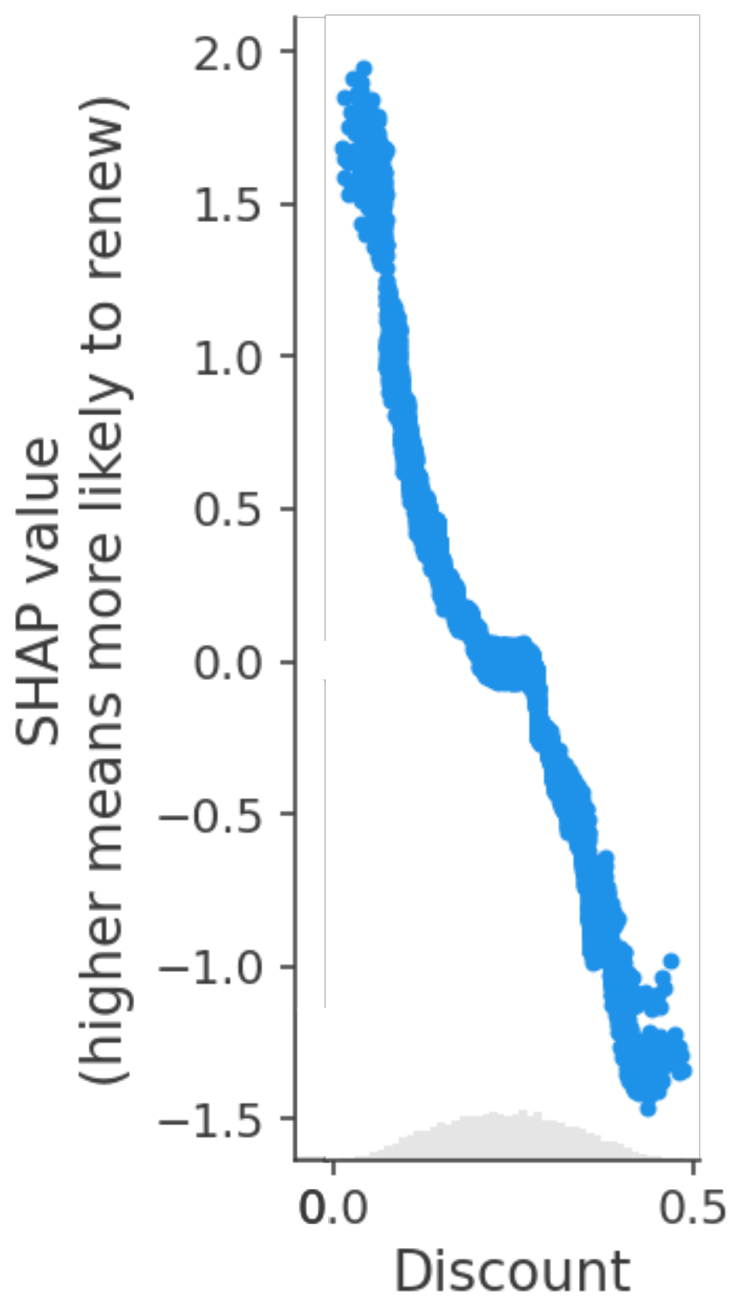


- SHAP value: one of the most cited measure for the feature importance
- *Larger* discounts contribute *less* to retention?

Application 2. Explainable AI

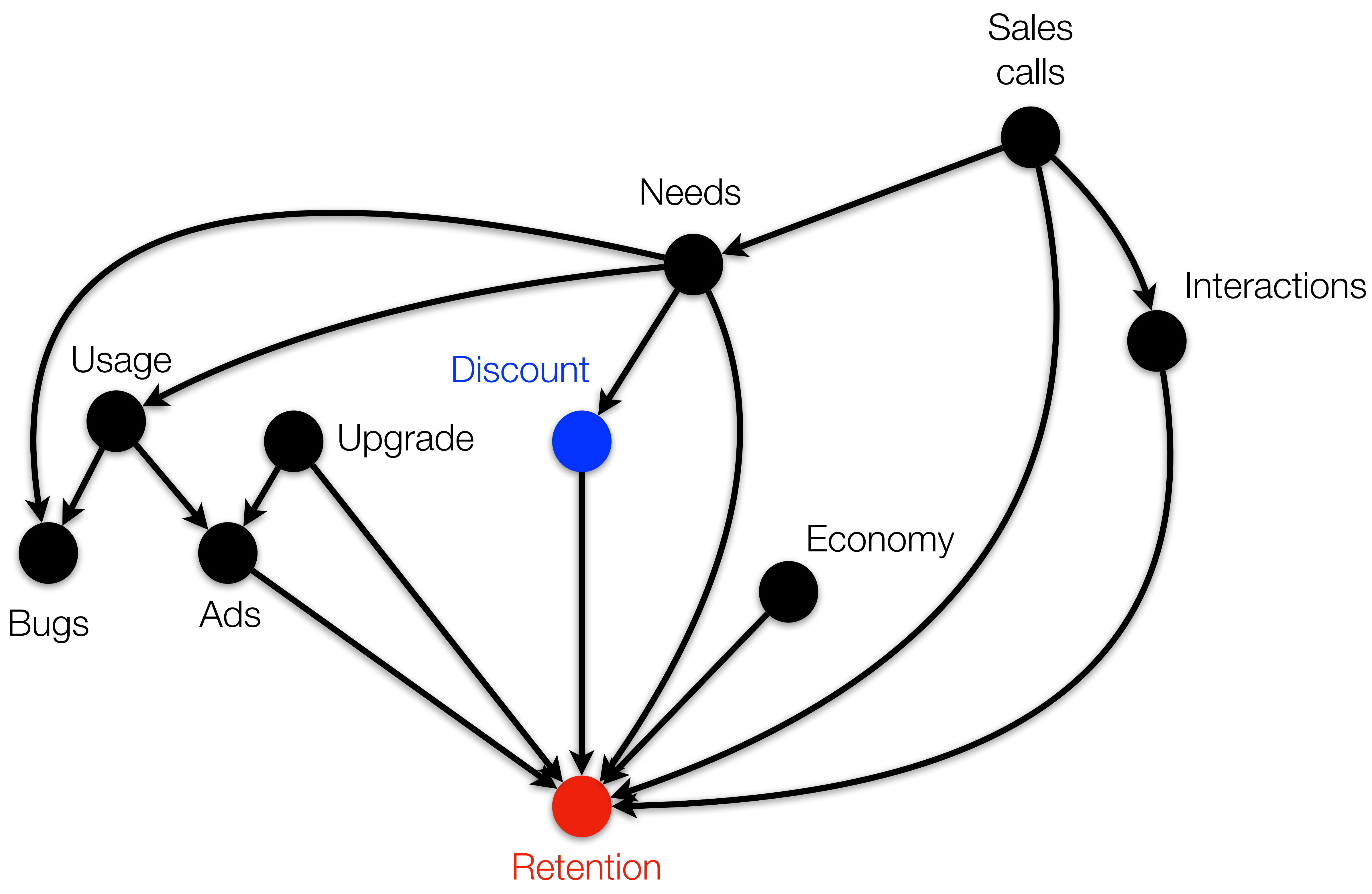


Contribution of Discount to the Retention?

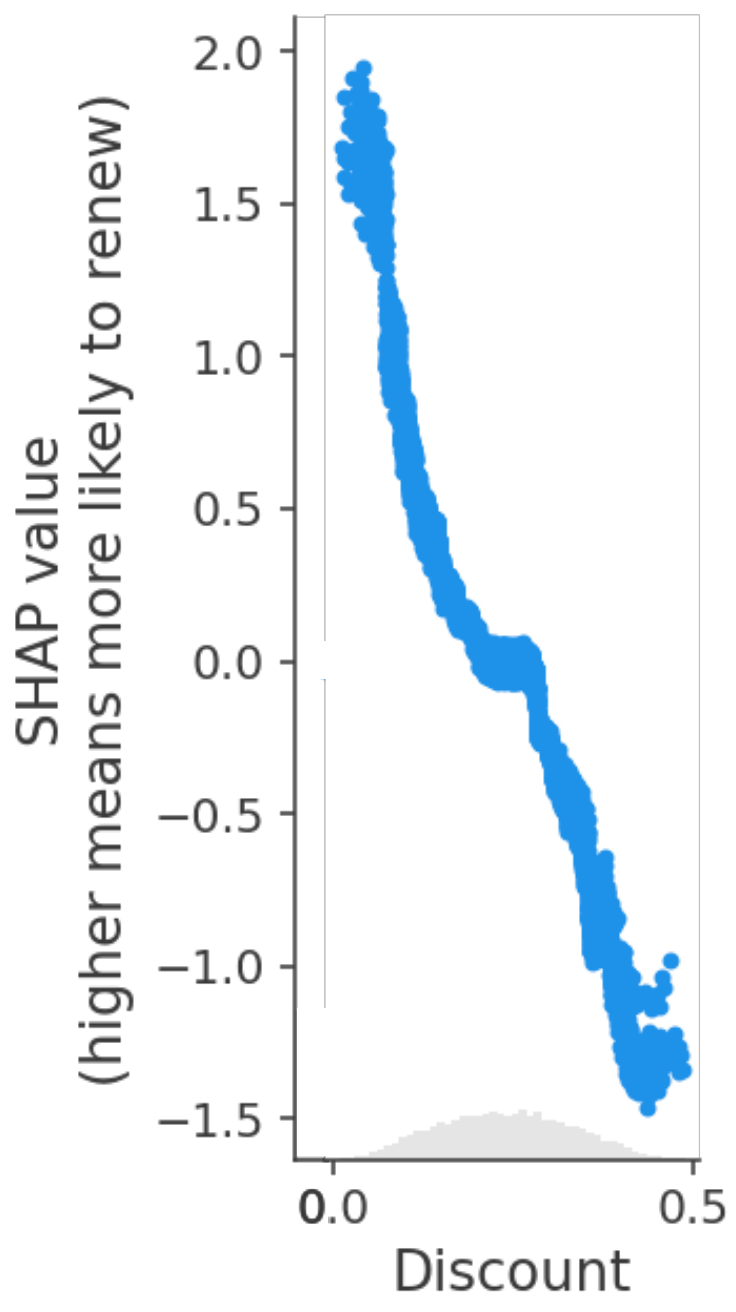


- SHAP value: one of the most cited measure for the feature importance
- *Larger* discounts contribute *less* to retention?
- Mismatch with human intuition is due to computing the importance based on correlation (e.g. $\mathbb{E}[\text{retention}|\text{discount}]$)

Application 2. Explainable AI



Contribution of Discount to the Retention?



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

Causality-based feature importance measure is essential

do-Shapley: Causality-based Feature Attribution

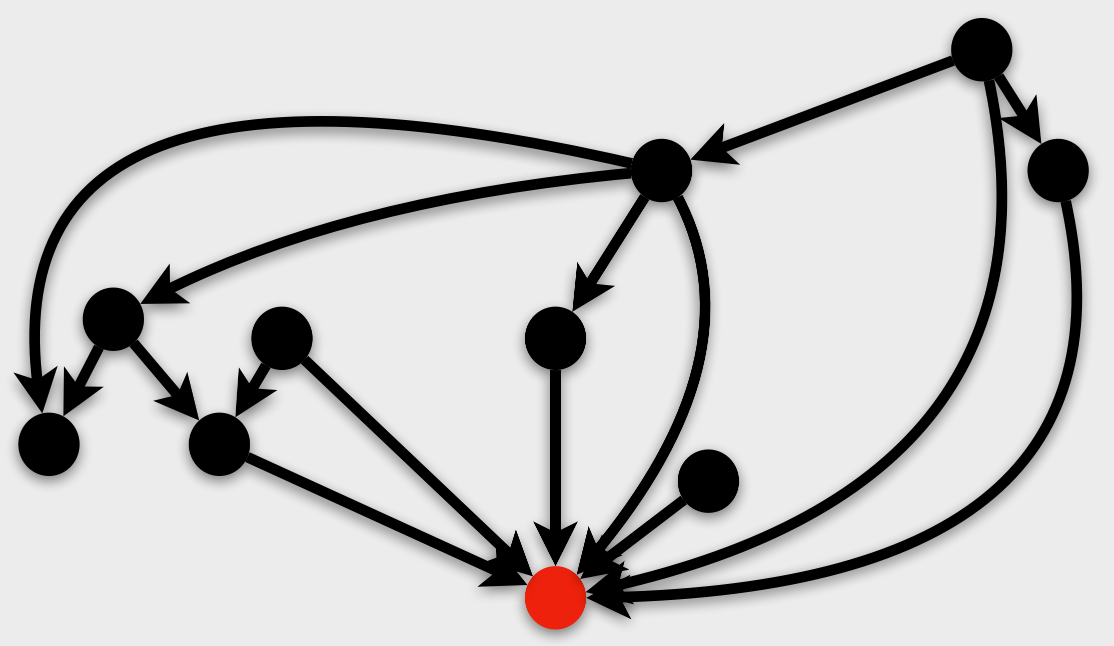
Jung et al.,ICML 2022

do-Shapley: Causality-based Feature Attribution

Jung et al.,ICML 2022

Input

Graph



Data

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

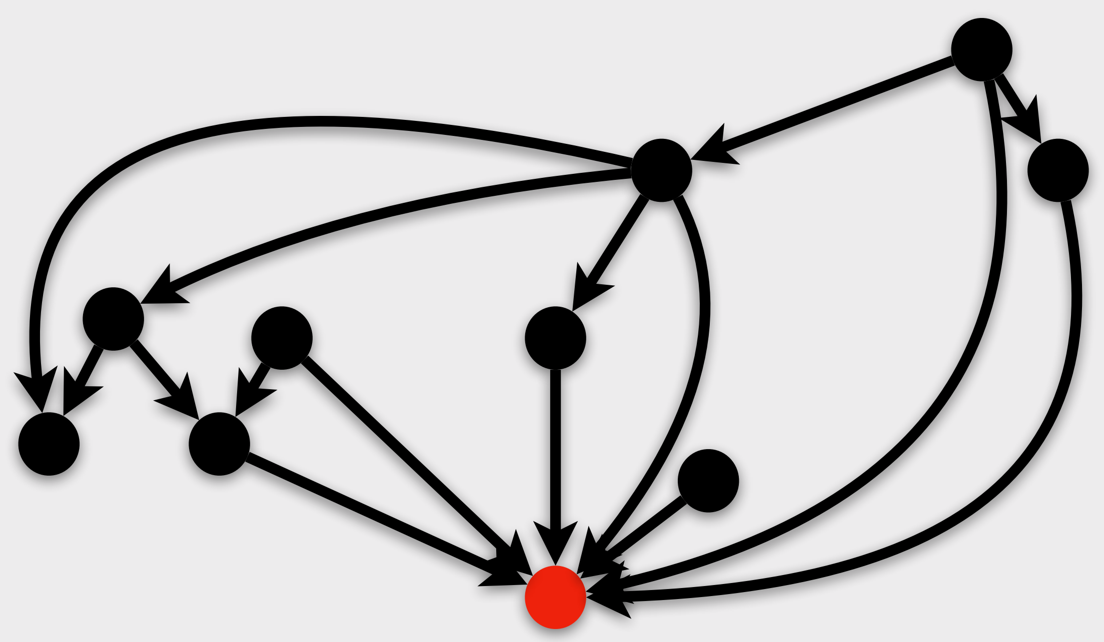
do-Shapley: Causality-based Feature Attribution

Jung et al.,ICML 2022

Input

Attribution

Graph



do-Shapley

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

Causality-based feature attribution

Data

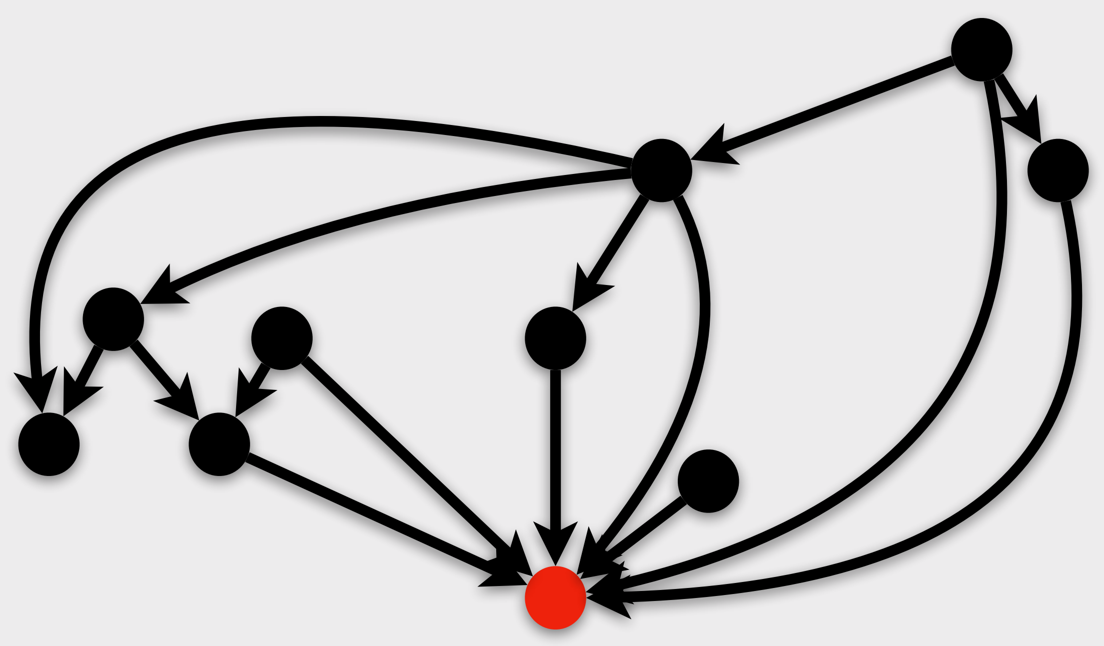
- Input: $(X_1, \cdots, X_m)$
- output: $f(X_1, \cdots, X_m)$

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

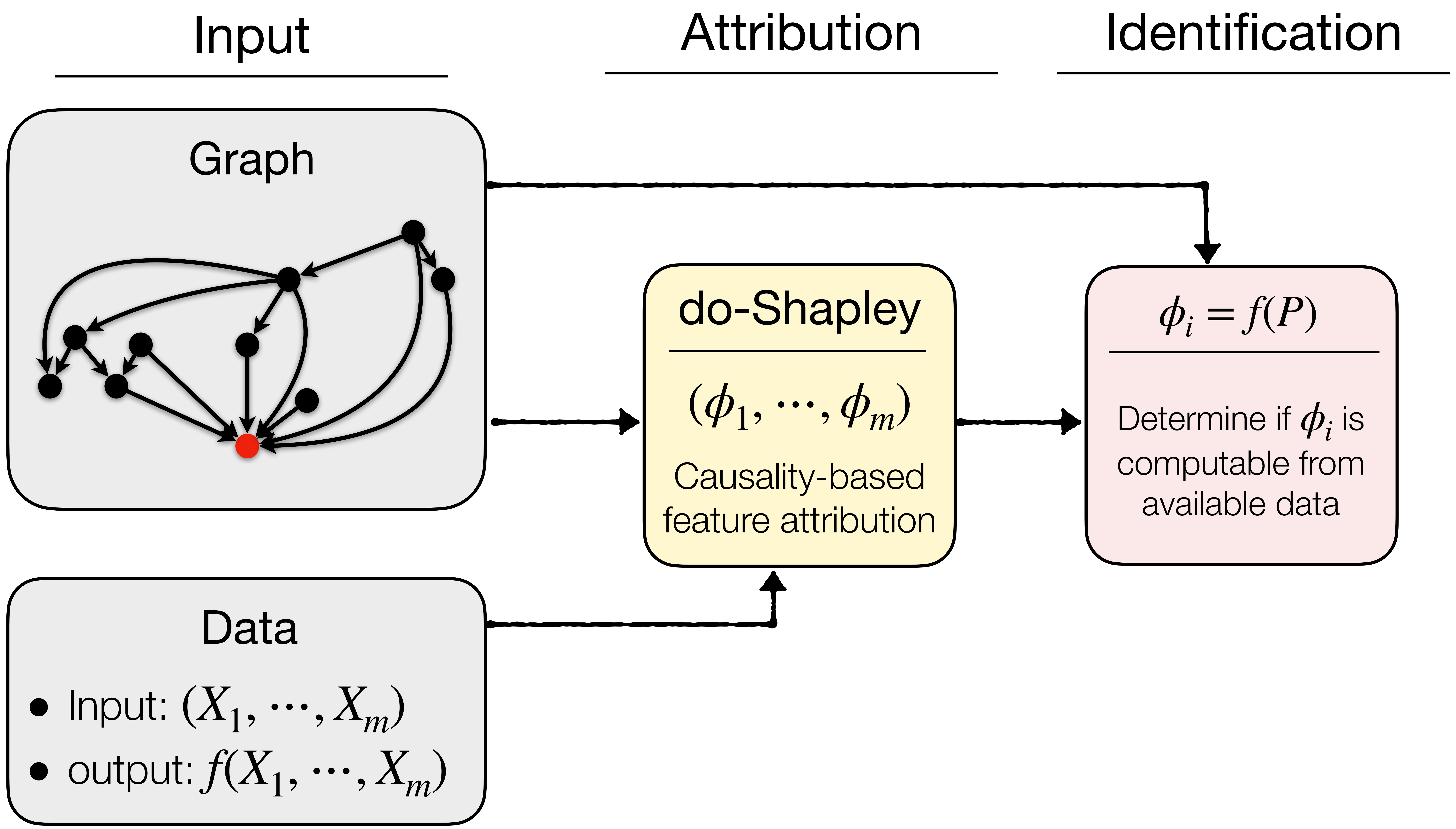
$(\phi_1, \cdots, \phi_m)$

Causality-based feature attribution

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

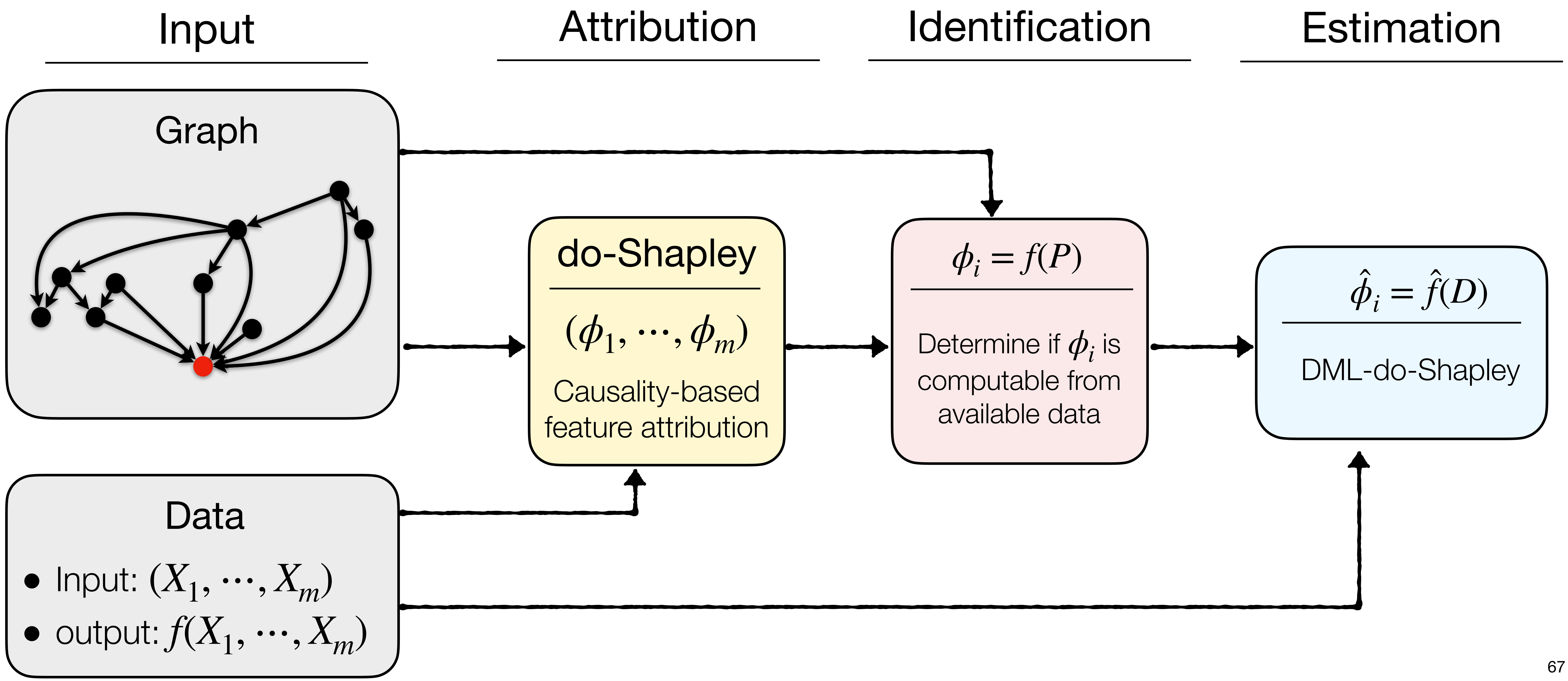
do-Shapley: Causality-based Feature Attribution

Jung et al.,ICML 2022



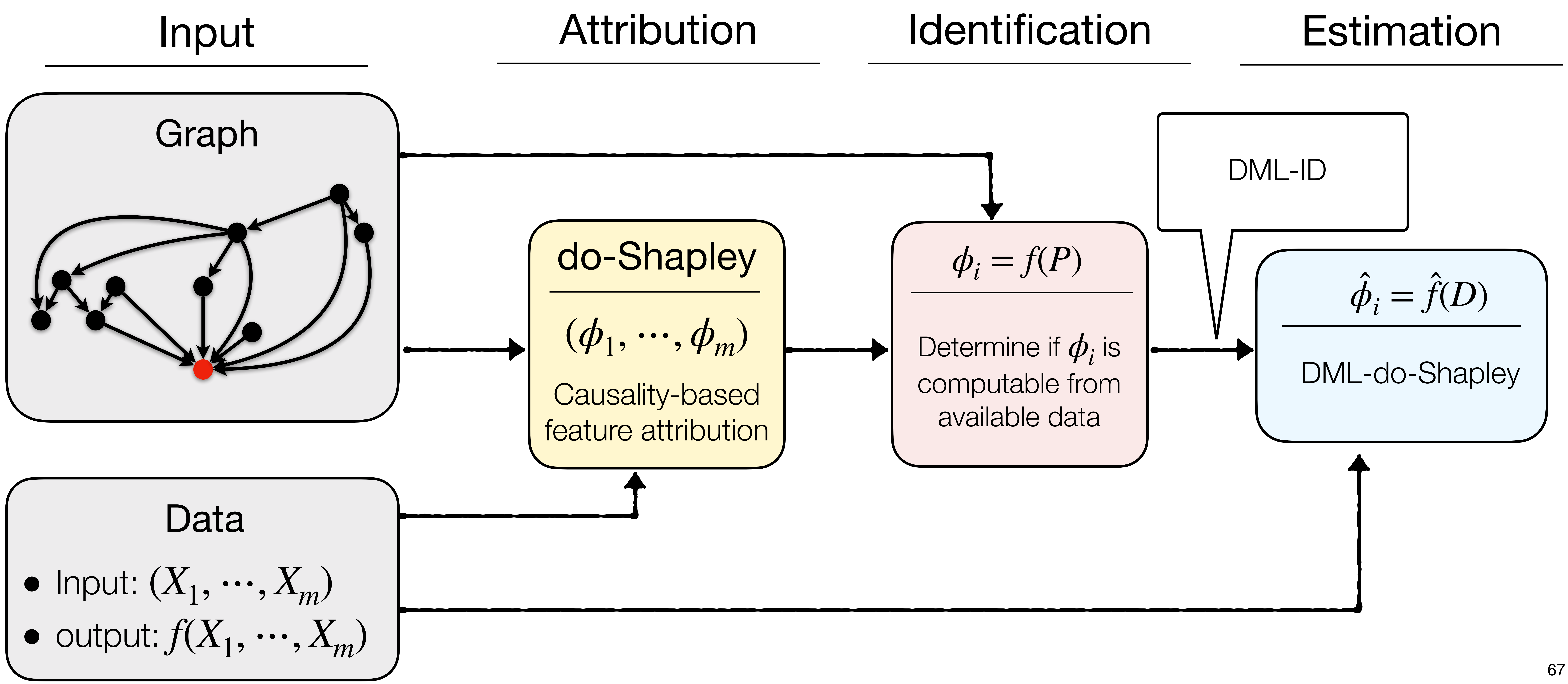
do-Shapley: Causality-based Feature Attribution

Jung et al.,ICML 2022



do-Shapley: Causality-based Feature Attribution

Jung et al.,ICML 2022



Simulation: Better Interpretability

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

Simulation: Better Interpretability

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

Simulation: Better Interpretability

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

Impact on Explainable AI

Impact on Explainable AI

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

Impact on Explainable AI

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

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

Impact on Explainable AI

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

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

Impact on Explainable AI

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

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

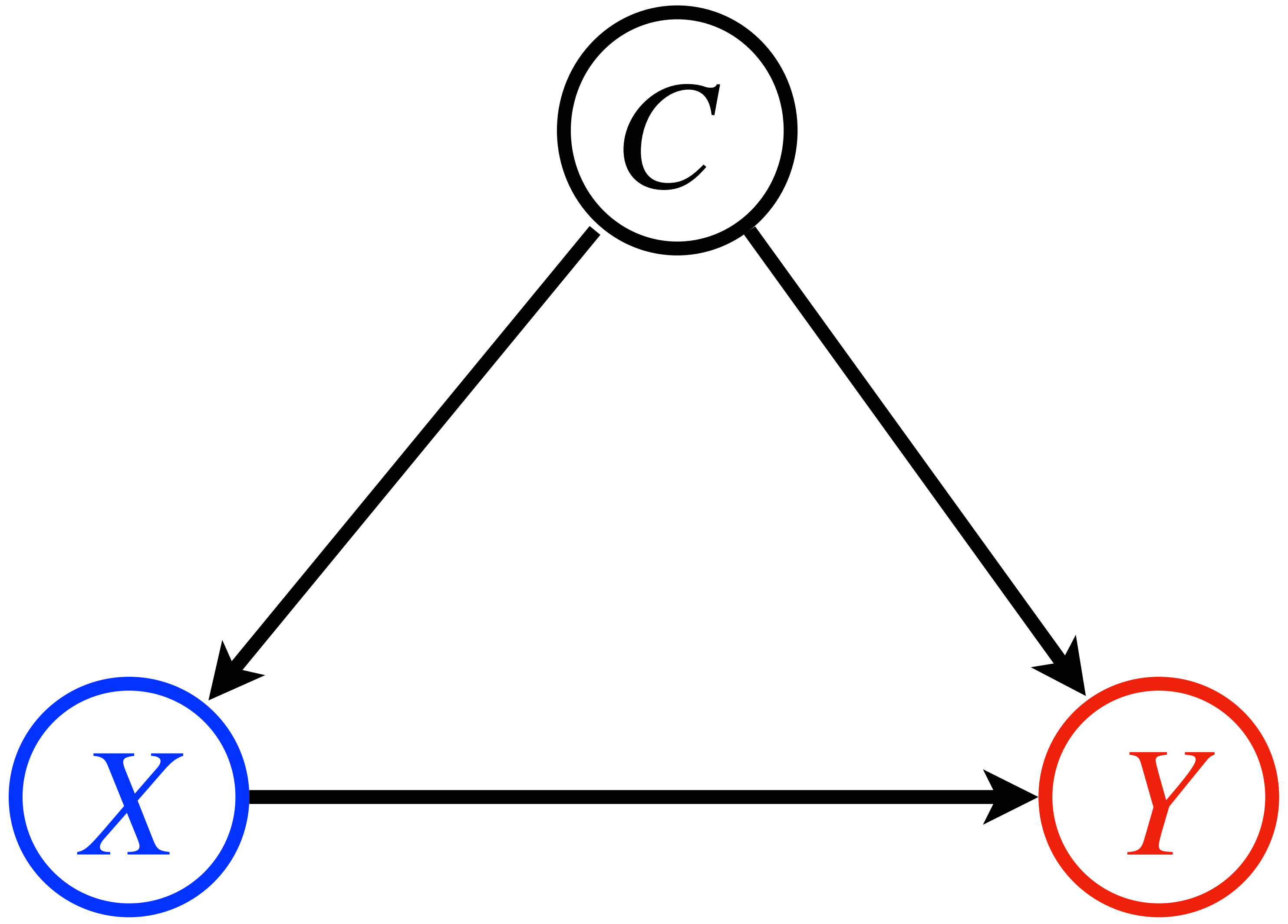
Impact on Explainable AI

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

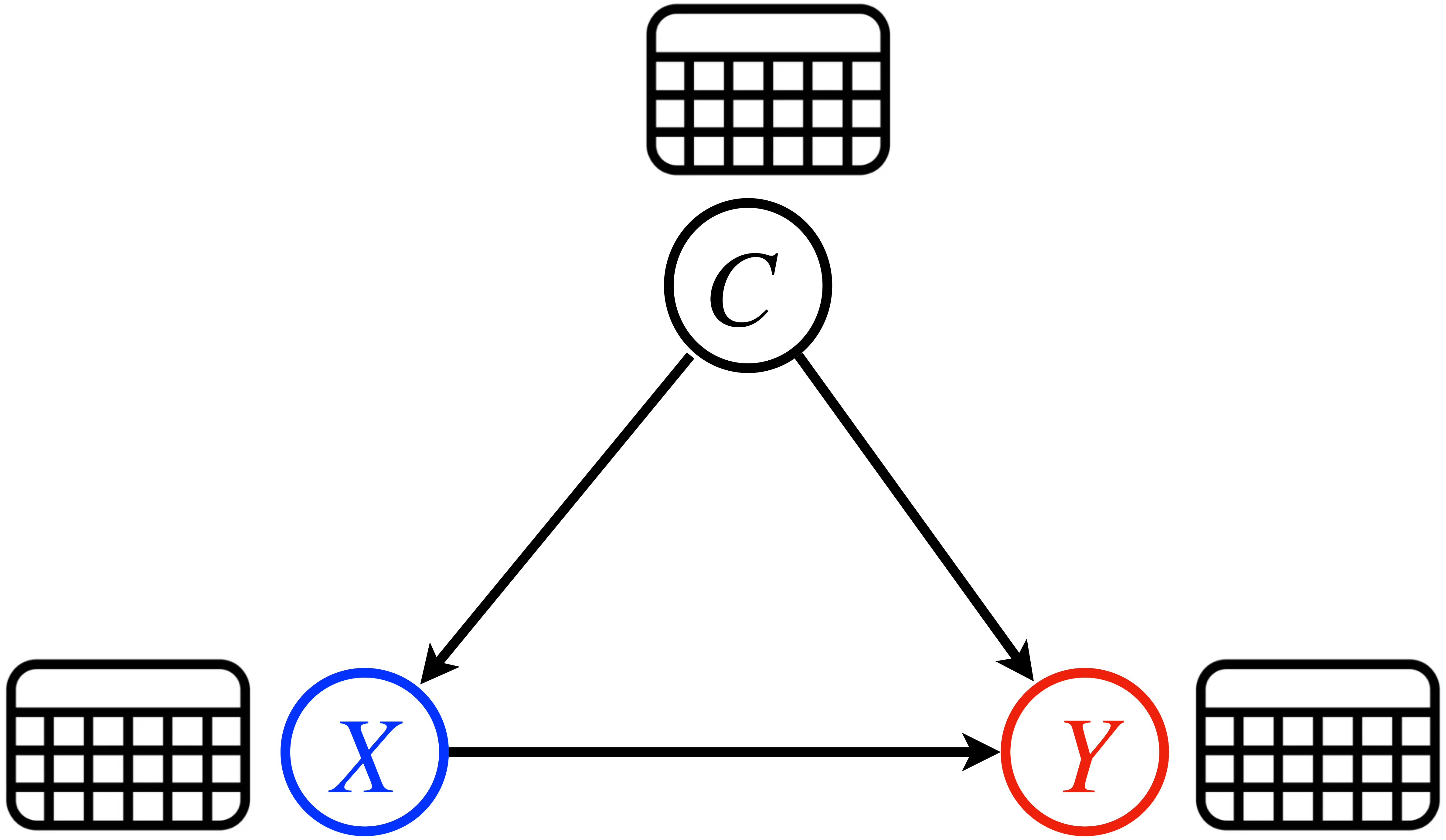
- Two features receive equal contributions whenever their causal effects are the same.
- Feature's contribution = 0 if it has no causal effect
- Feature contributions closely approximate their causal effects on the outcome
- The sum of feature contributions = The outcome $f(X_1, \cdots, X_m)$

Future Direction

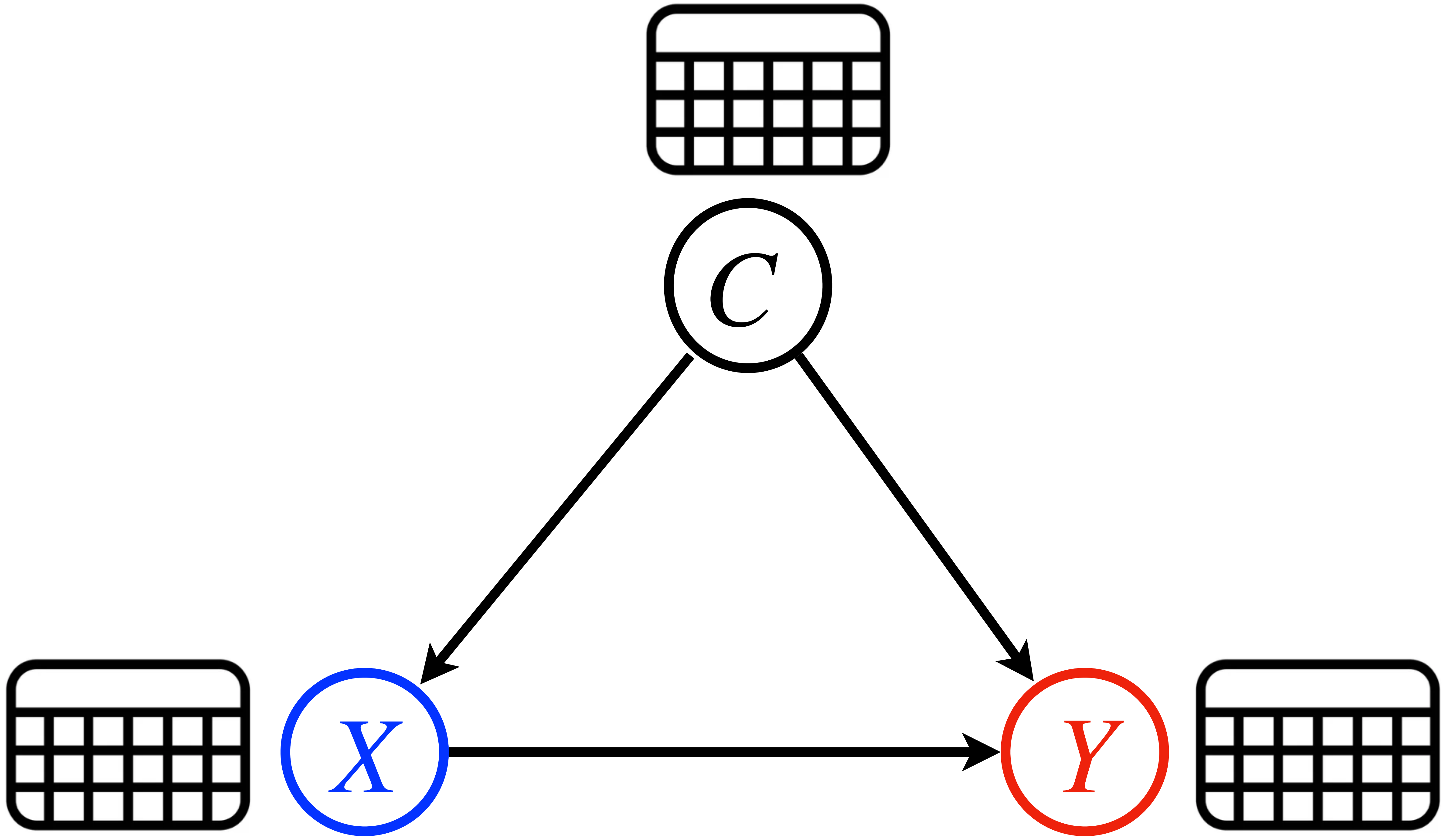
Future 1: Inference with Multi-modal Data



Future 1: Inference with Multi-modal Data

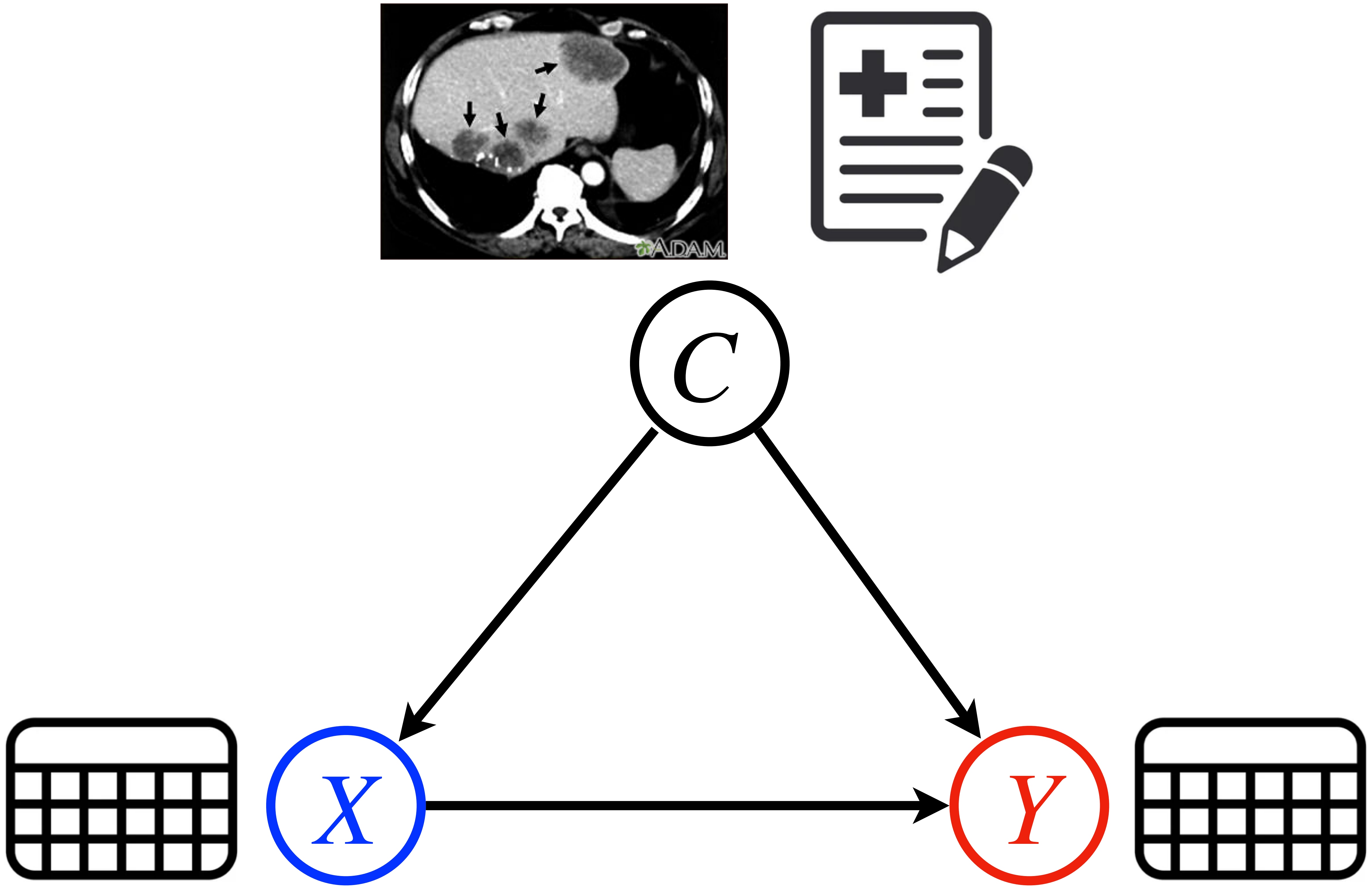


Future 1: Inference with Multi-modal Data



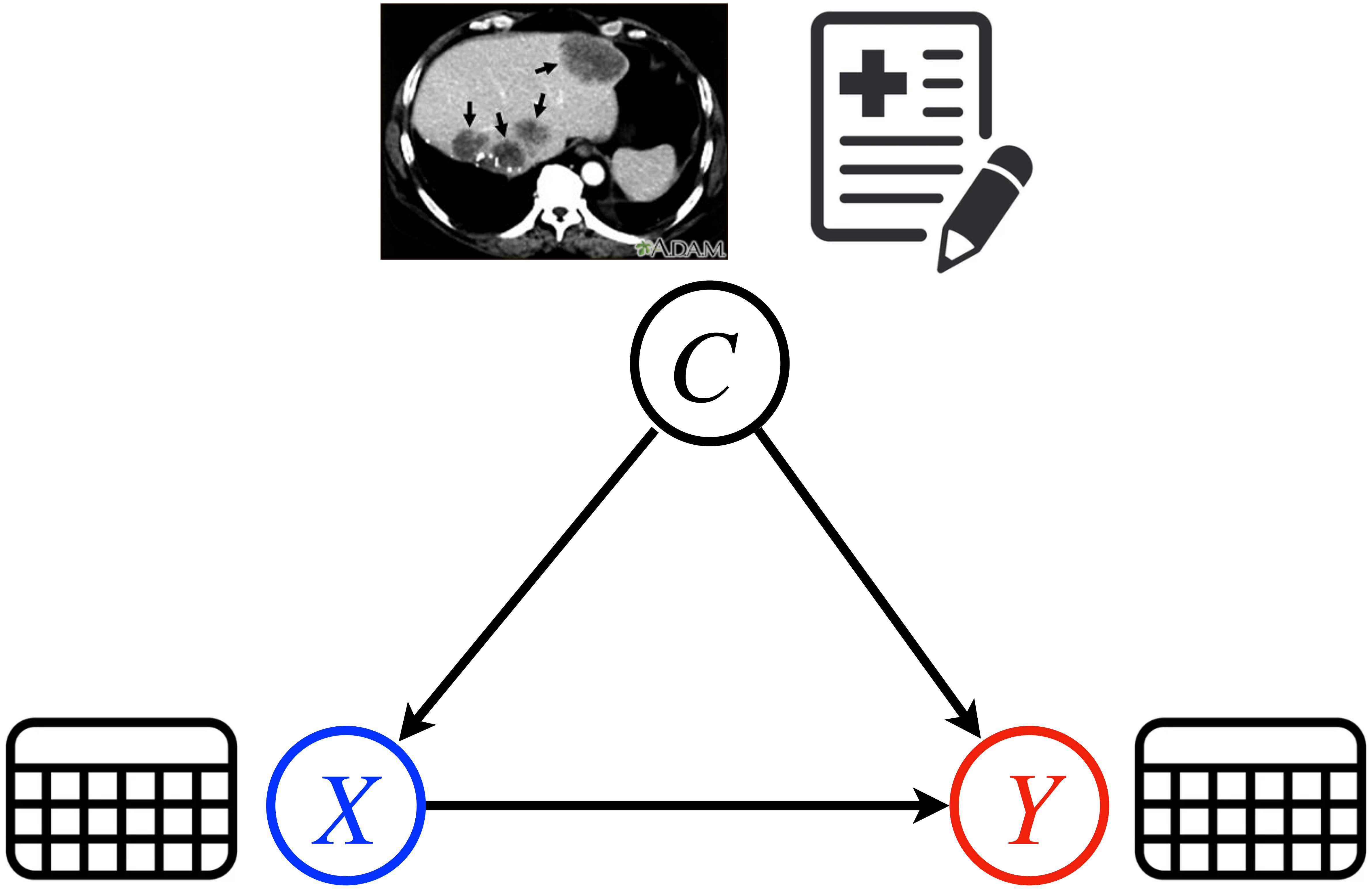
$$\mathbb{E}[Y \mid \text{do}(x)] = \sum_c \mathbb{E}[Y \mid x, c]\, P(c)$$

Future 1: Inference with Multi-modal Data



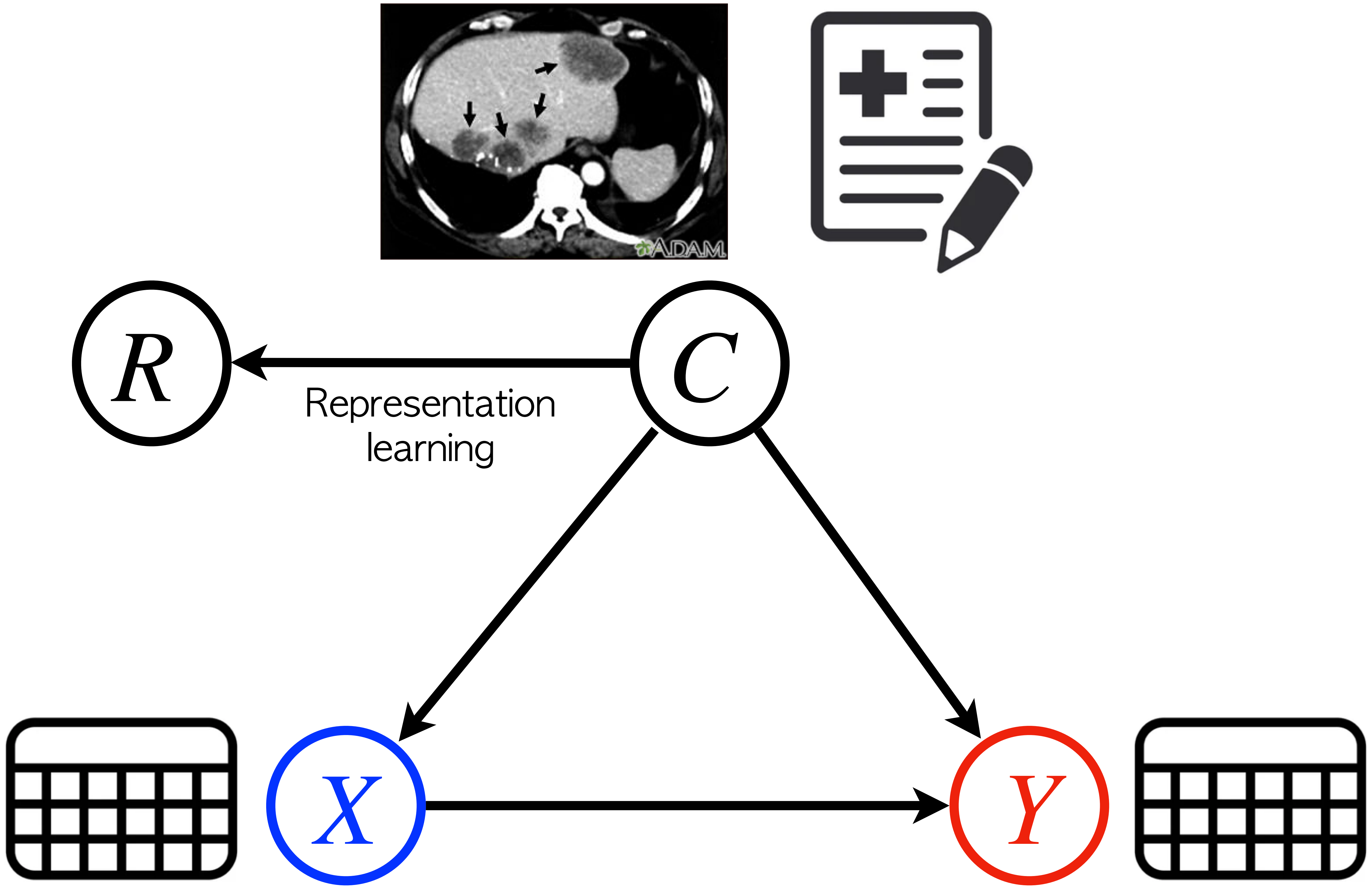
$$\mathbb{E}[Y \mid \mathrm{do}(x)] = \sum_c \mathbb{E}[Y \mid x, c]\, P(c)$$

Future 1: Inference with Multi-modal Data



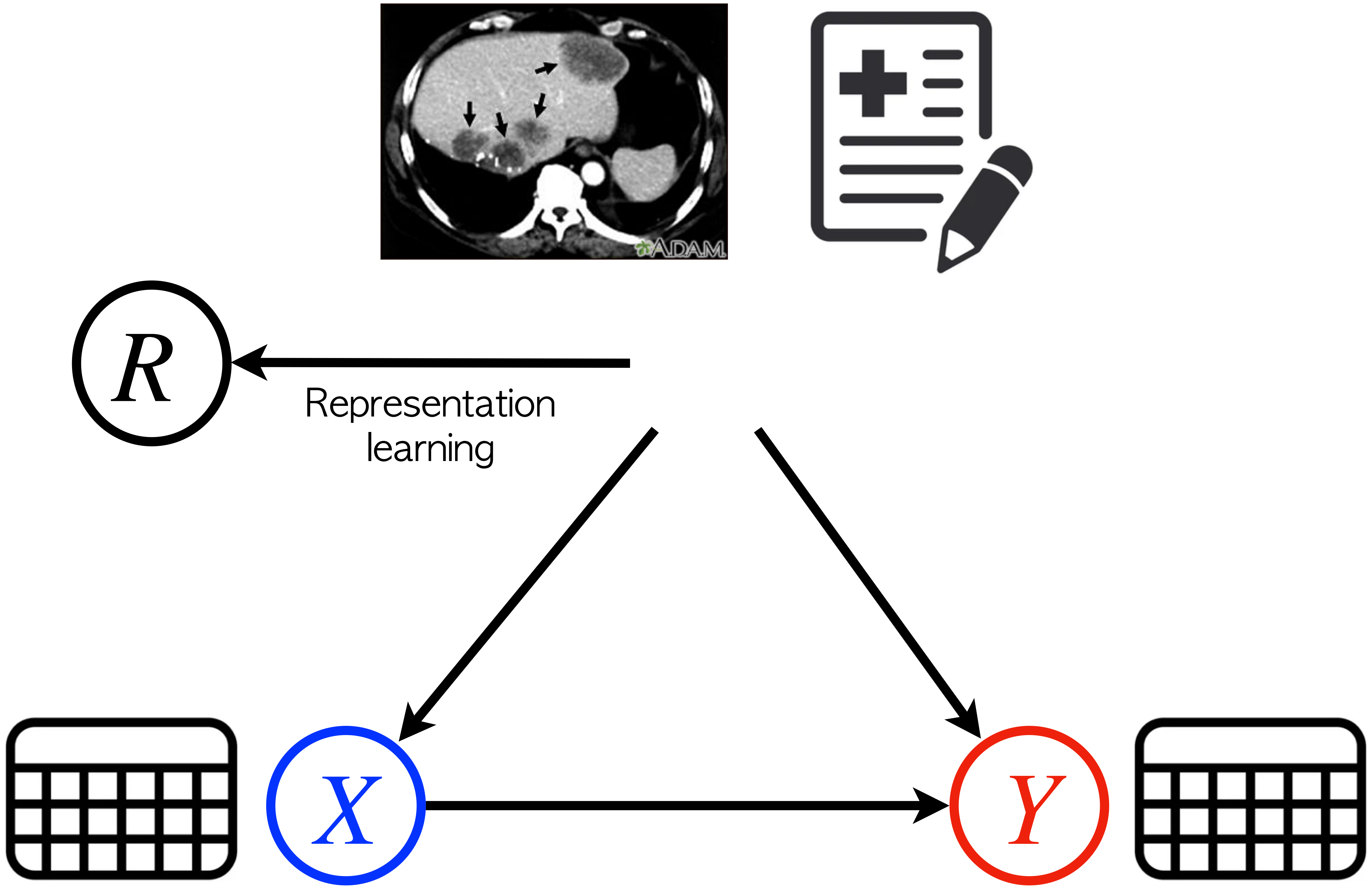
$$\mathbb{E}[Y \mid \mathrm{do}(x)] = \quad ?$$

Future 1: Inference with Multi-modal Data



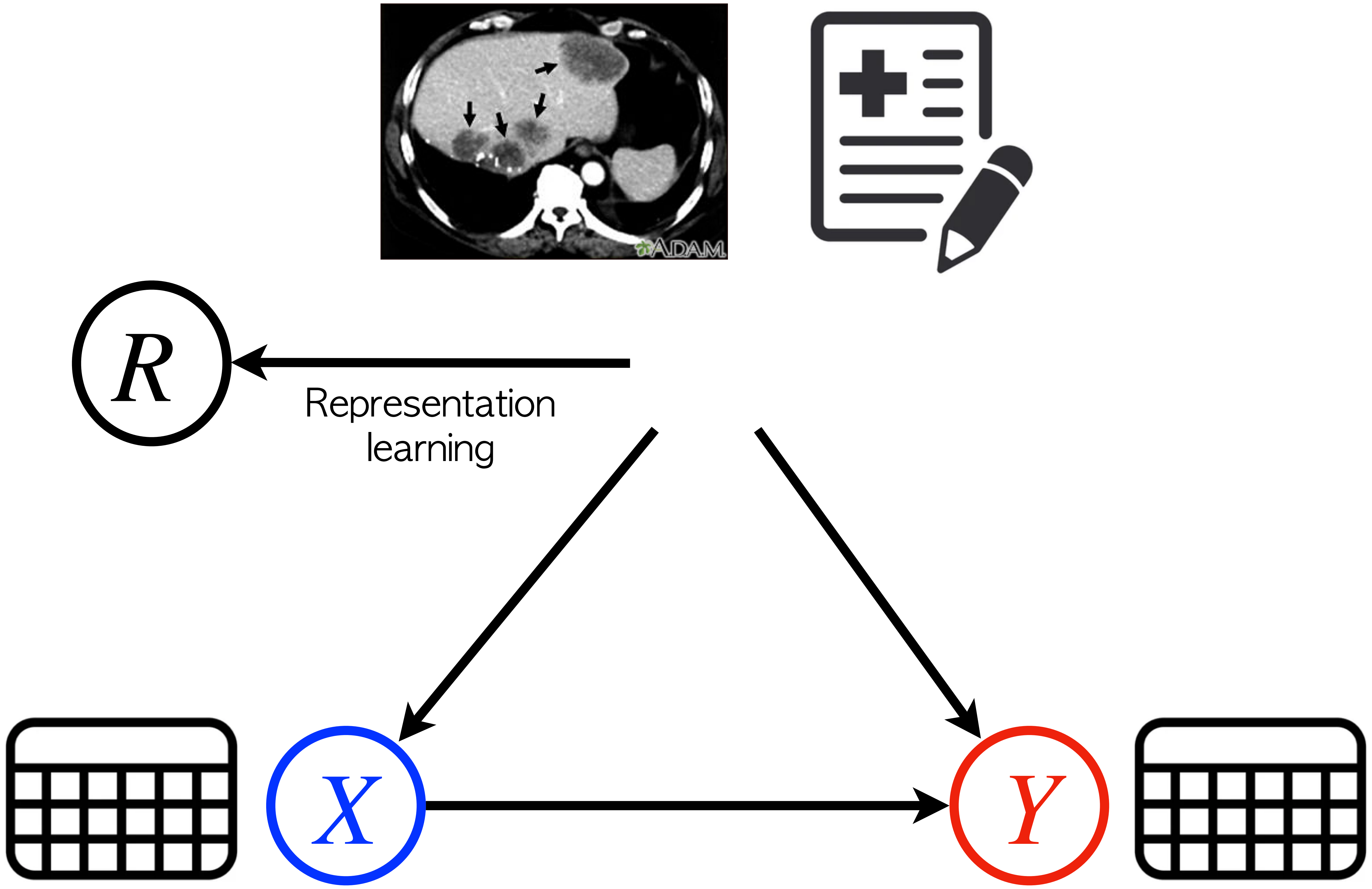
$$\mathbb{E}[Y \mid \mathrm{do}(x)] = \quad ?$$

Future 1: Inference with Multi-modal Data



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

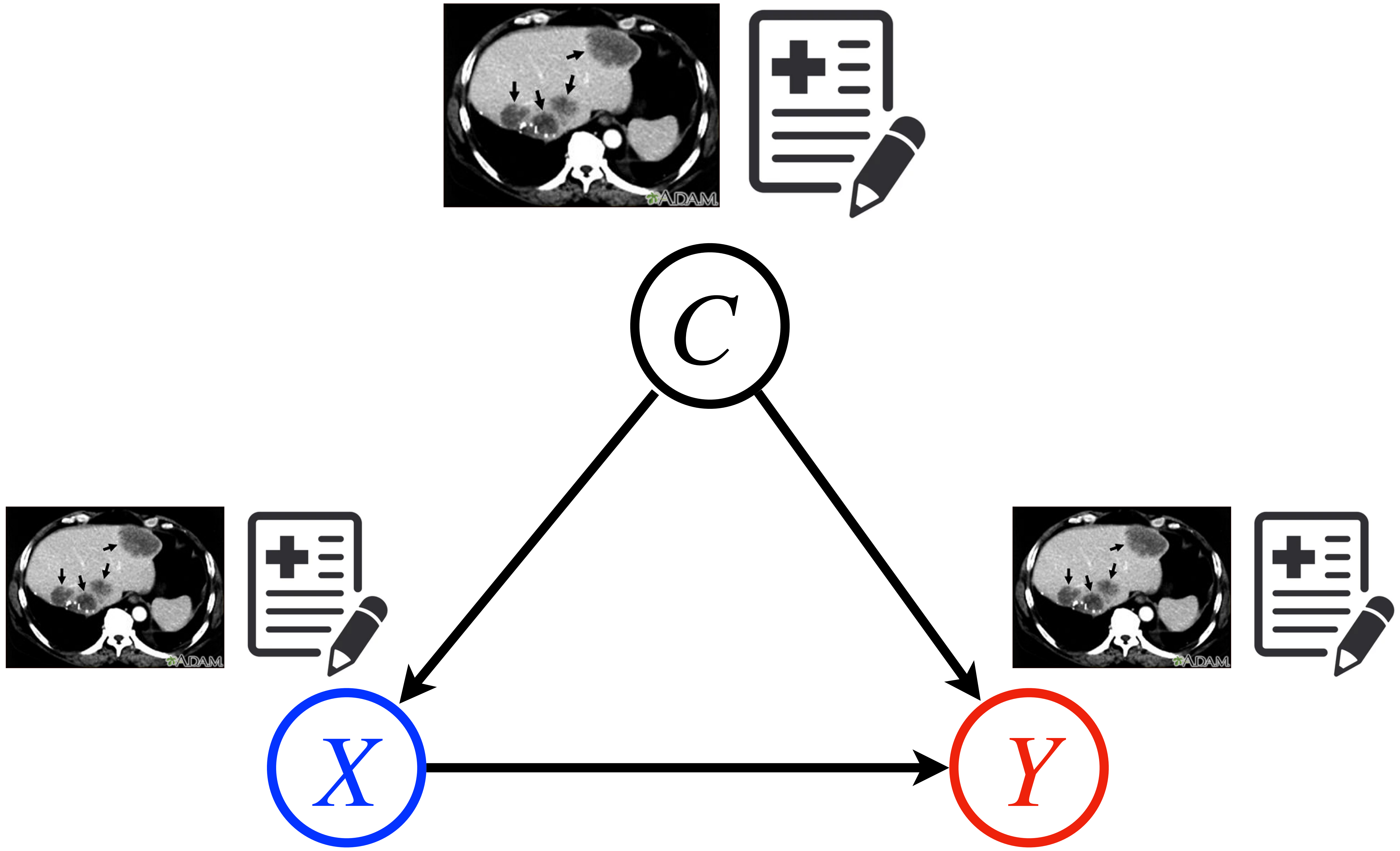
Future 1: Inference with Multi-modal Data



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

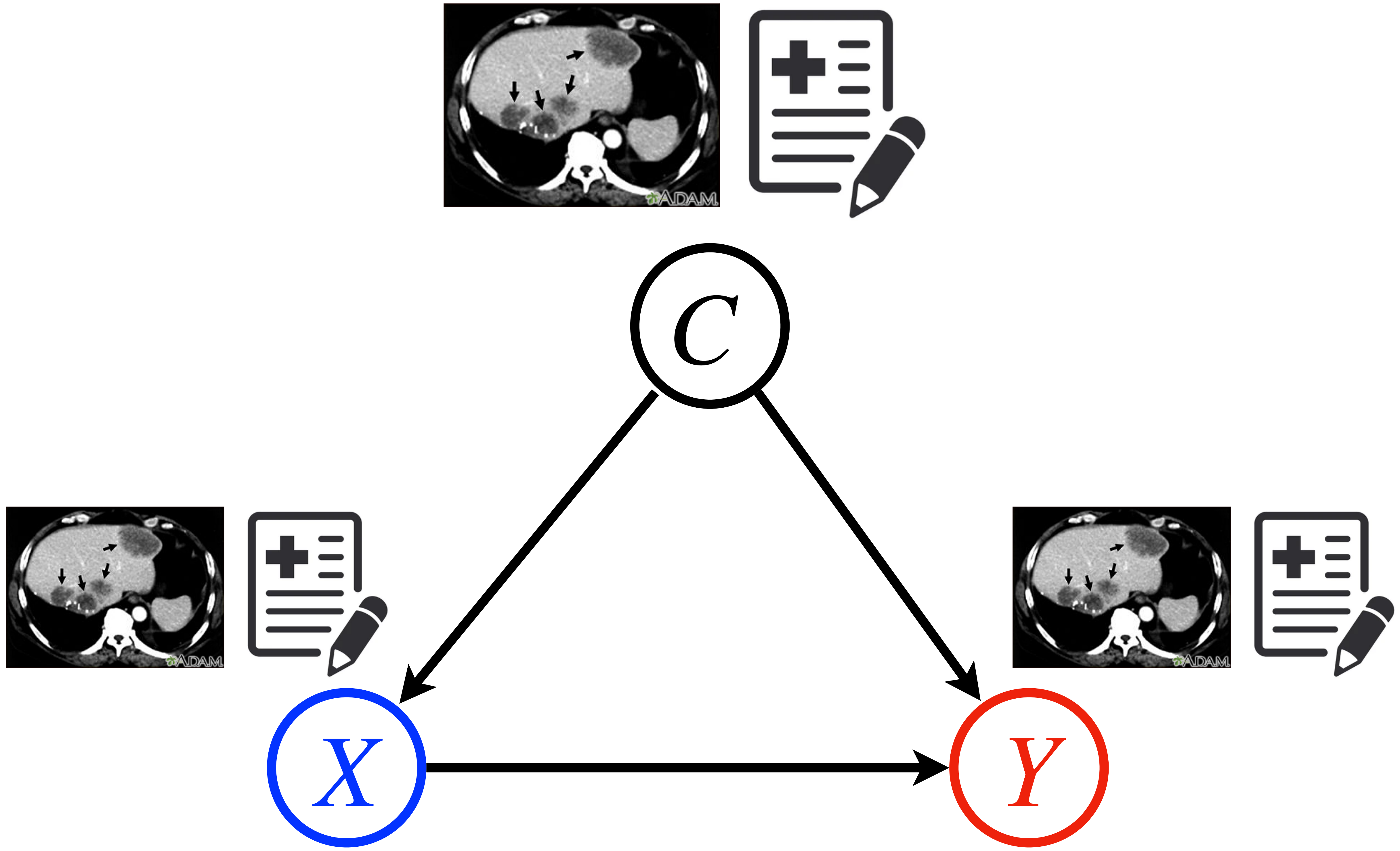
R doesn't satisfy the BD criterion

Future 1: Inference with Multi-modal Data



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

Future 1: Inference with Multi-modal Data



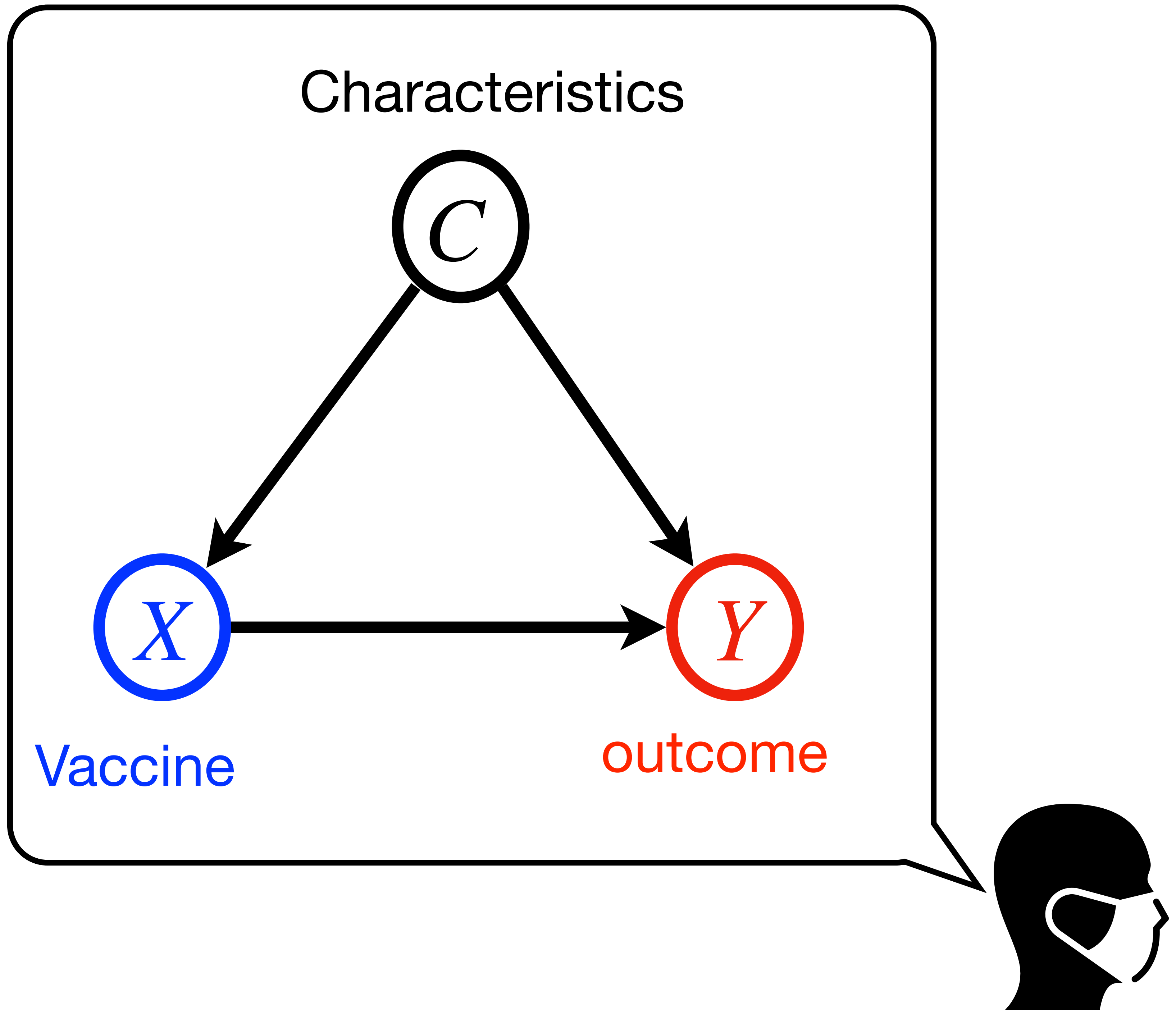
$$\mathbb{E}[Y \mid do(x)] = \text{?}$$

Approach

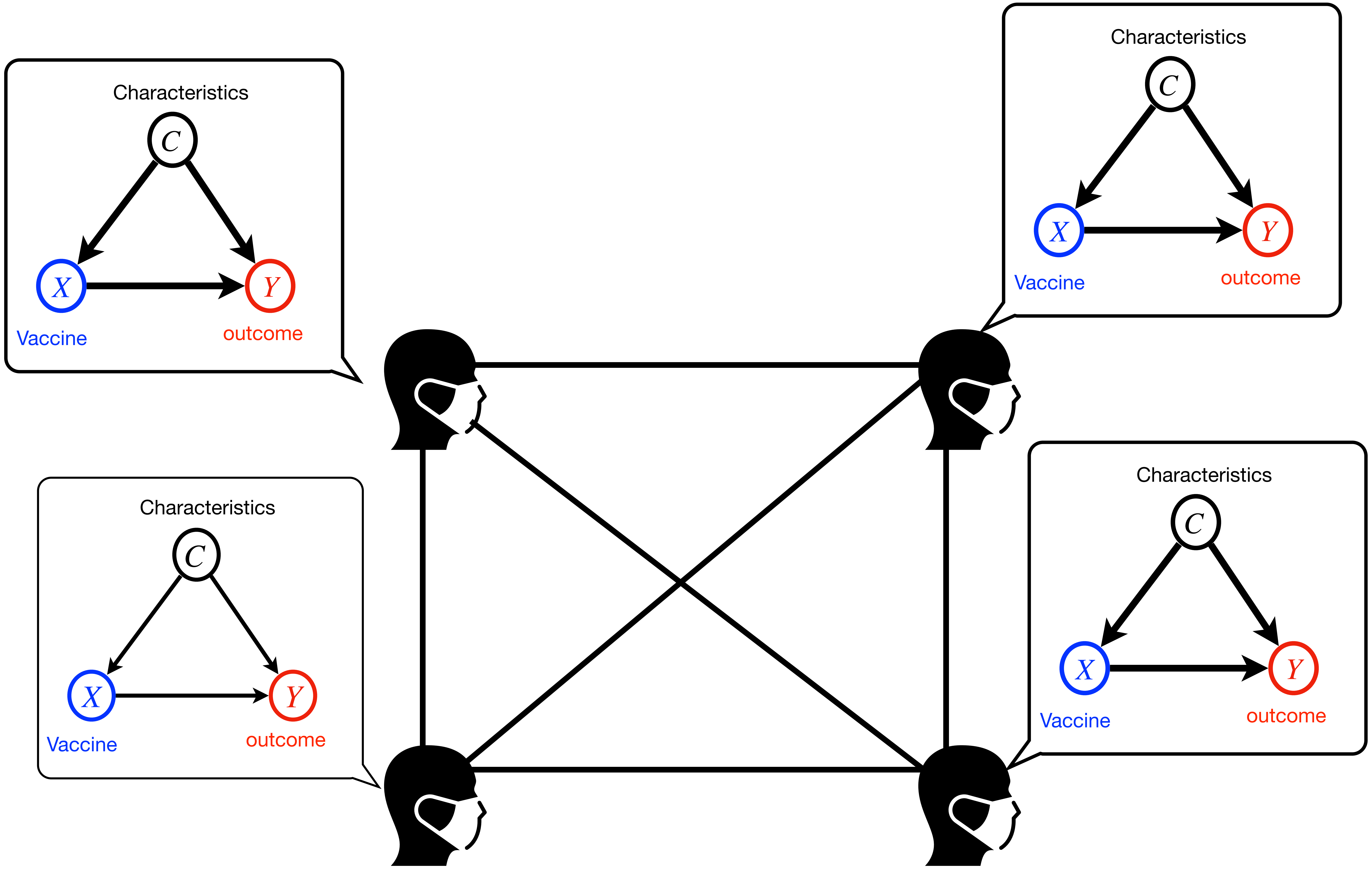
- Representation learning taking account of causal dependencies
- New causal inference methods that allows us to use existing representation learning models

Future 2: Causal Inference with Spatiotemporal Data

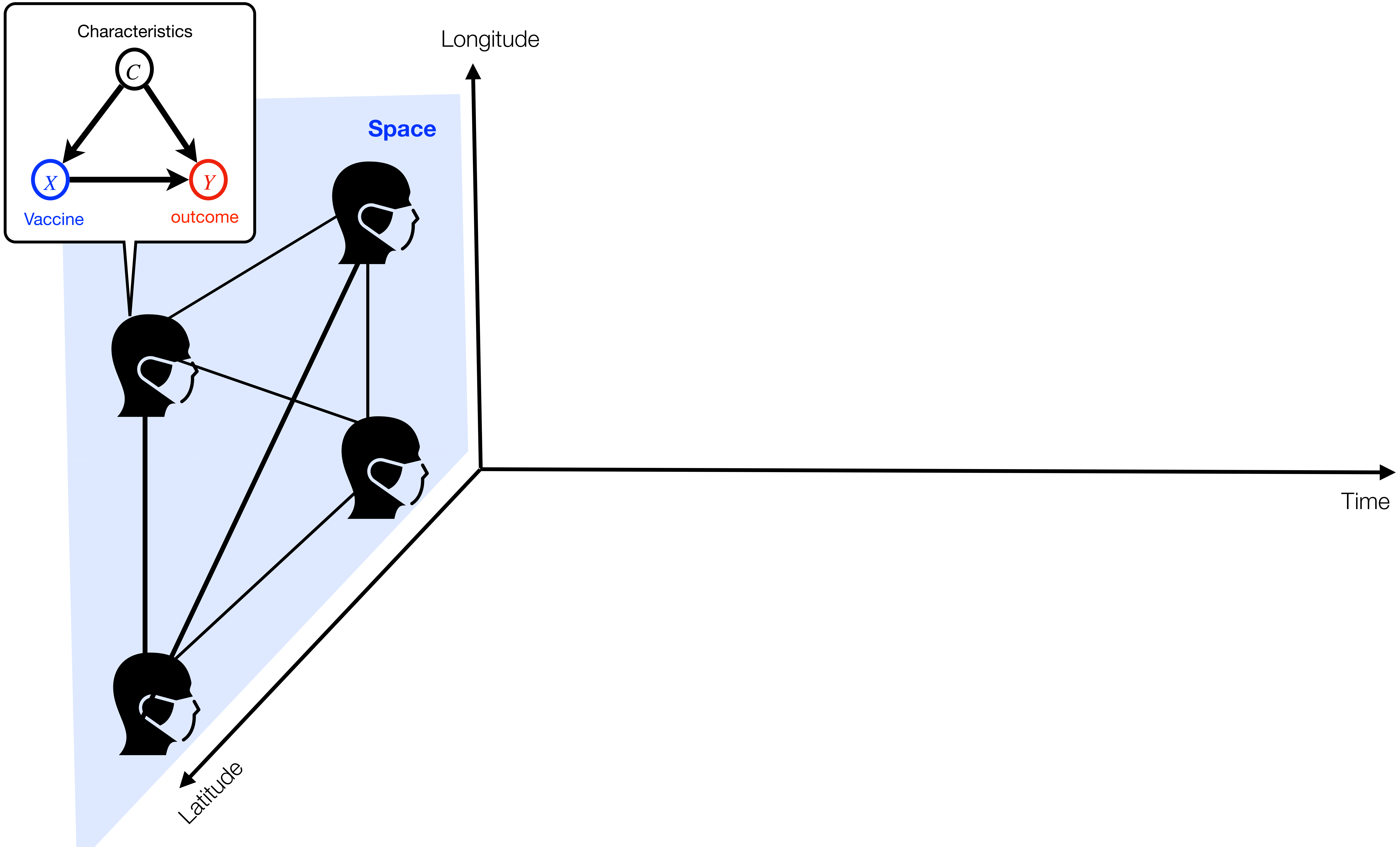
Future 2: Causal Inference with Spatiotemporal Data



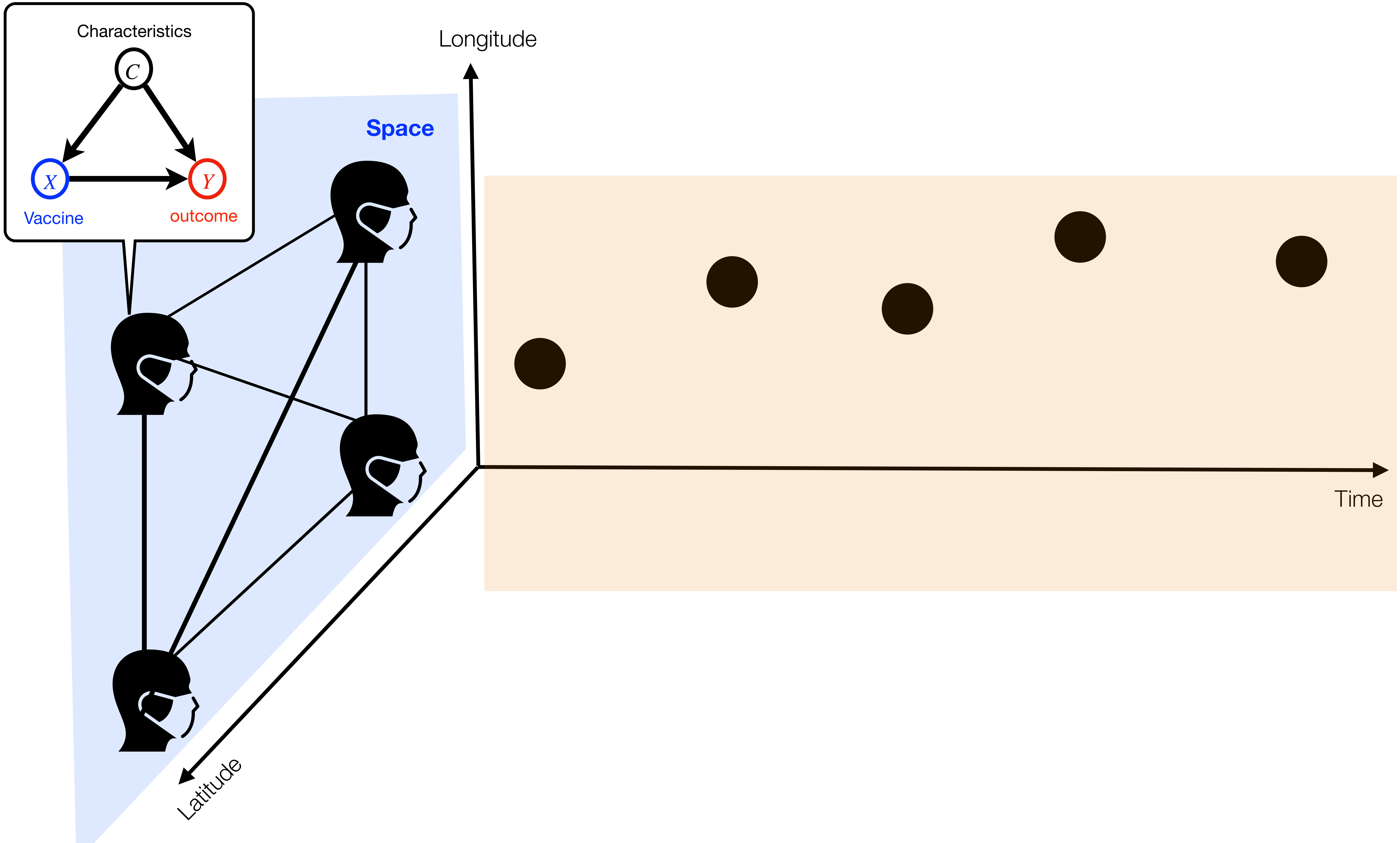
Future 2: Causal Inference with Spatiotemporal Data



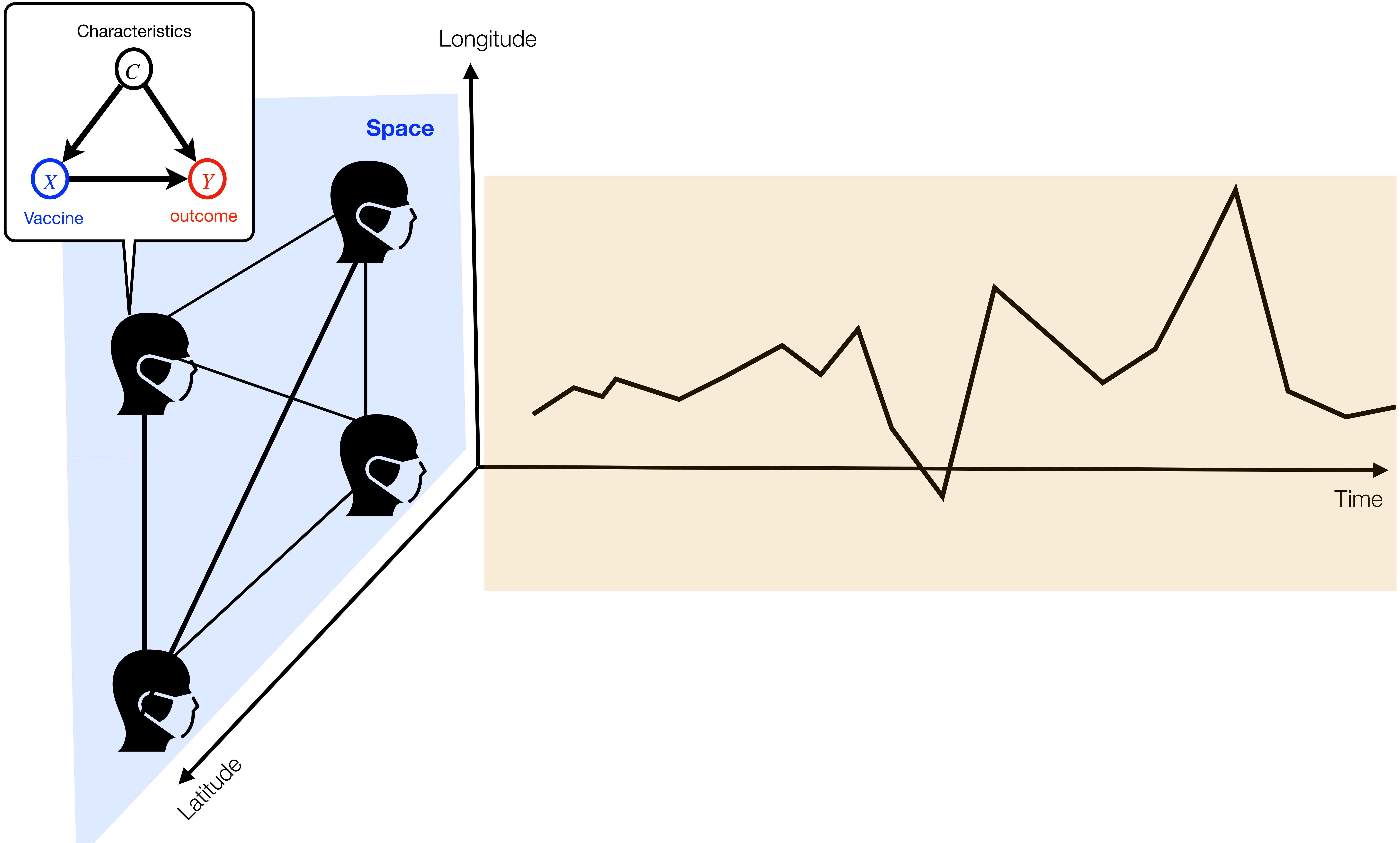
Future 2: Causal Inference with Spatiotemporal Data



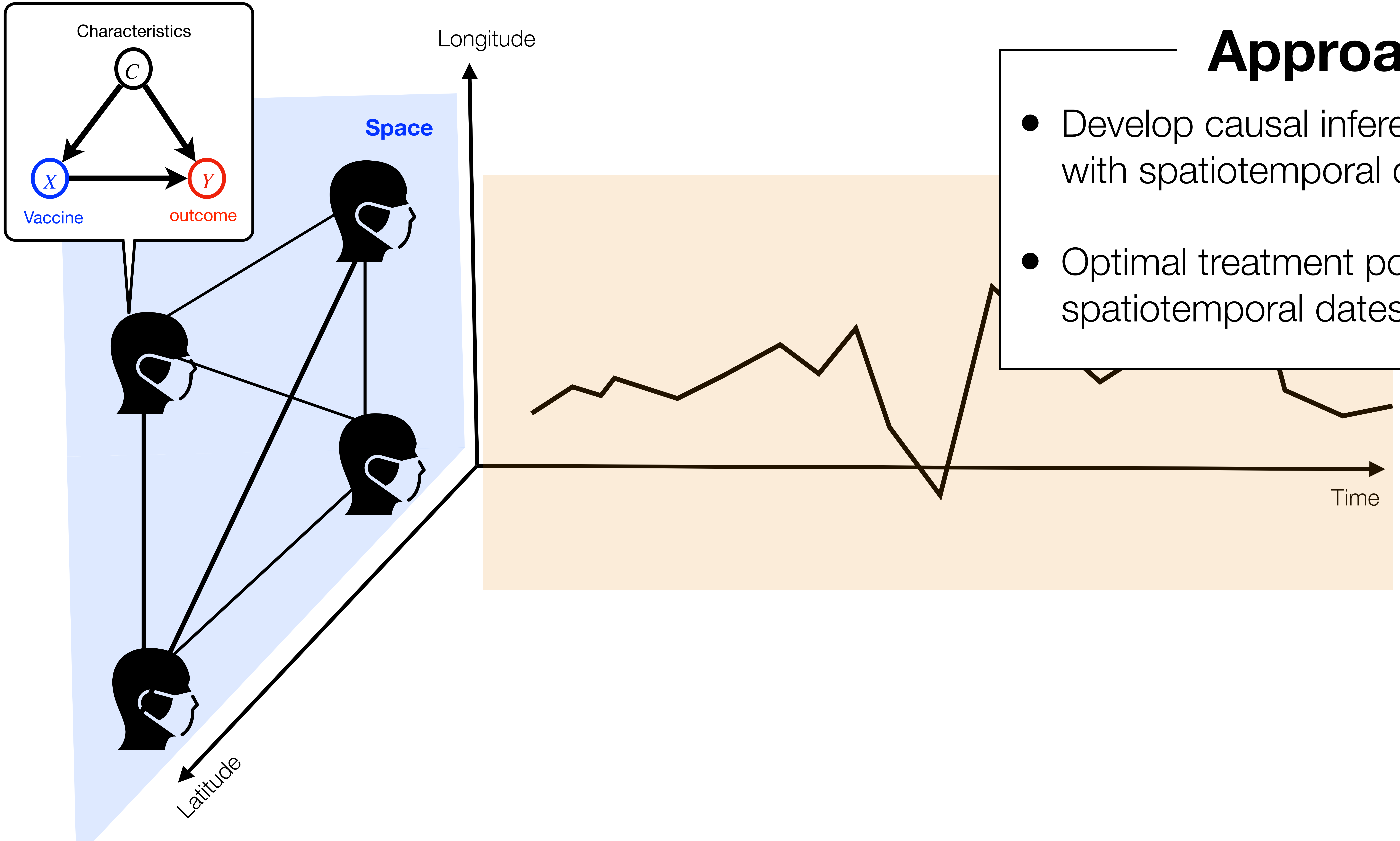
Future 2: Causal Inference with Spatiotemporal Data



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

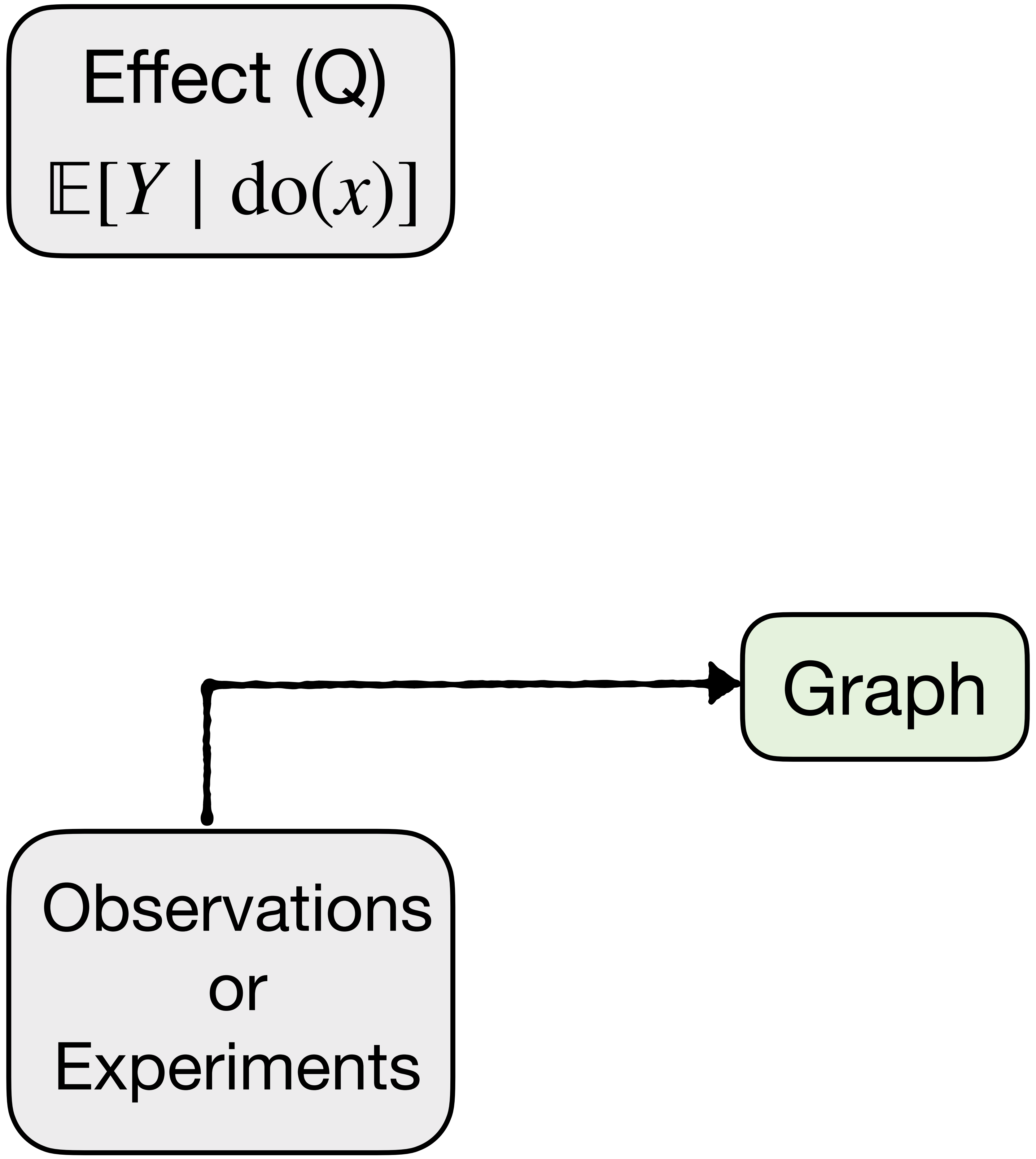
Future 3: Causal Inference Loop with Uncertainty

Effect (Q)

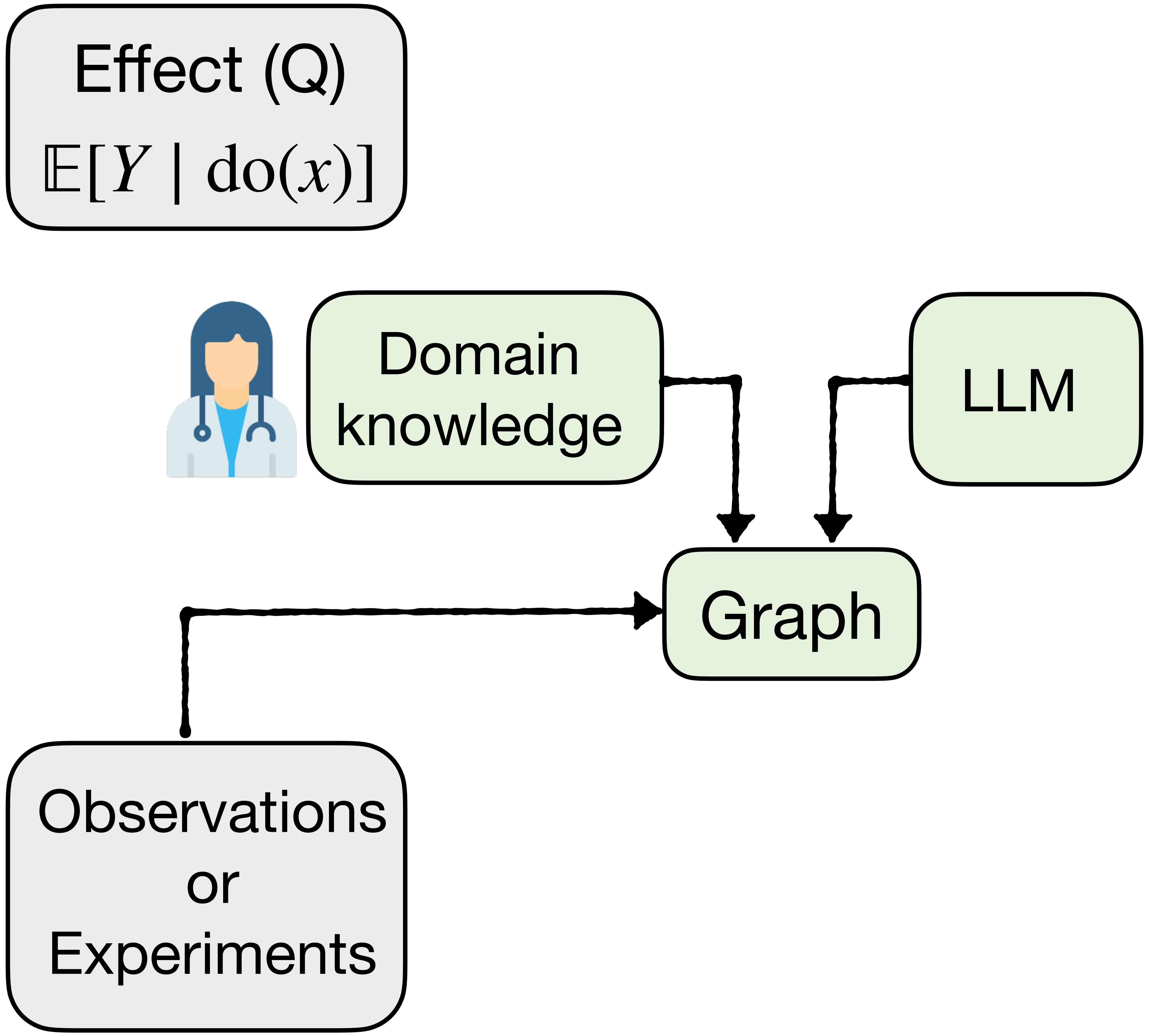
$\mathbb{E}[Y \mid \mathrm{do}(x)]$

Observations
or
Experiments

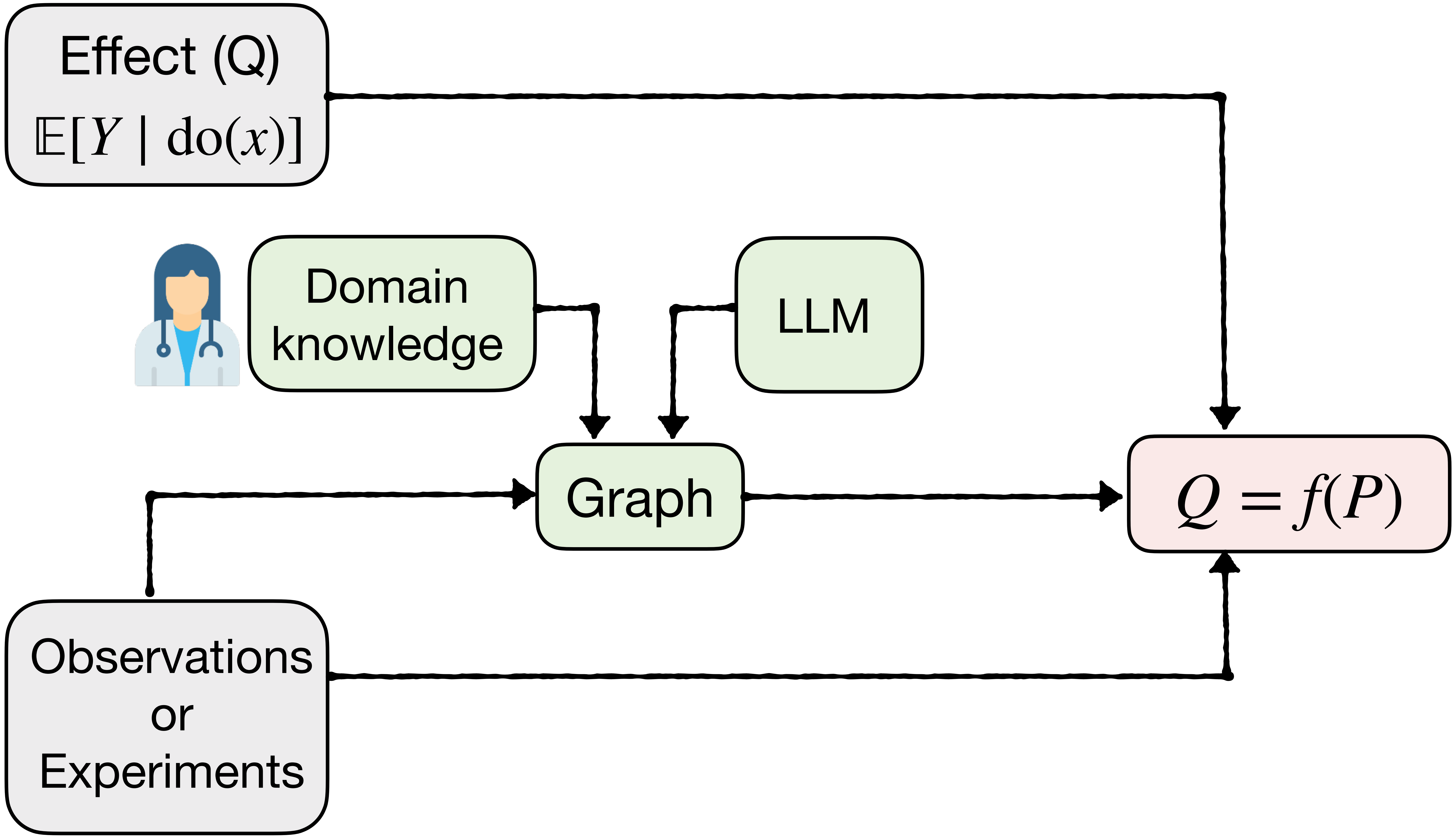
Future 3: Causal Inference Loop with Uncertainty



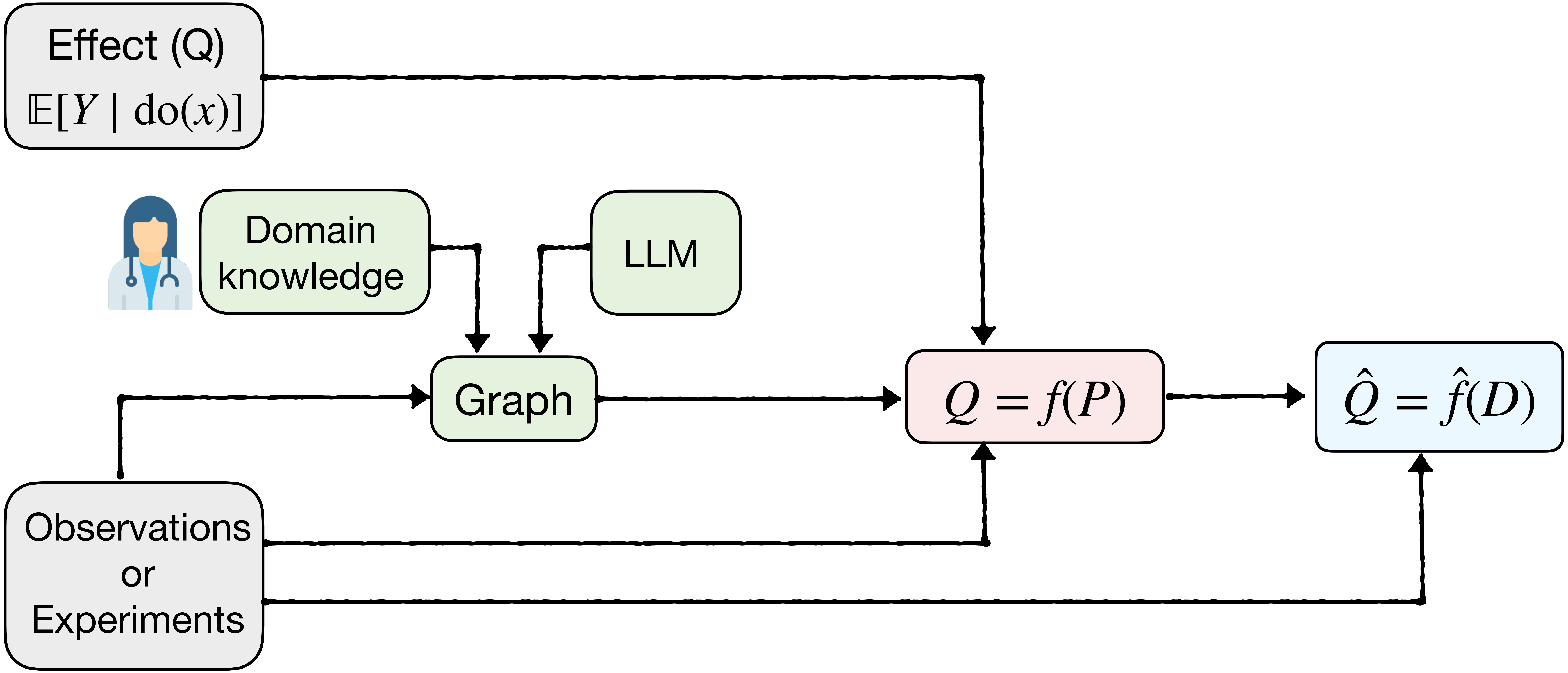
Future 3: Causal Inference Loop with Uncertainty



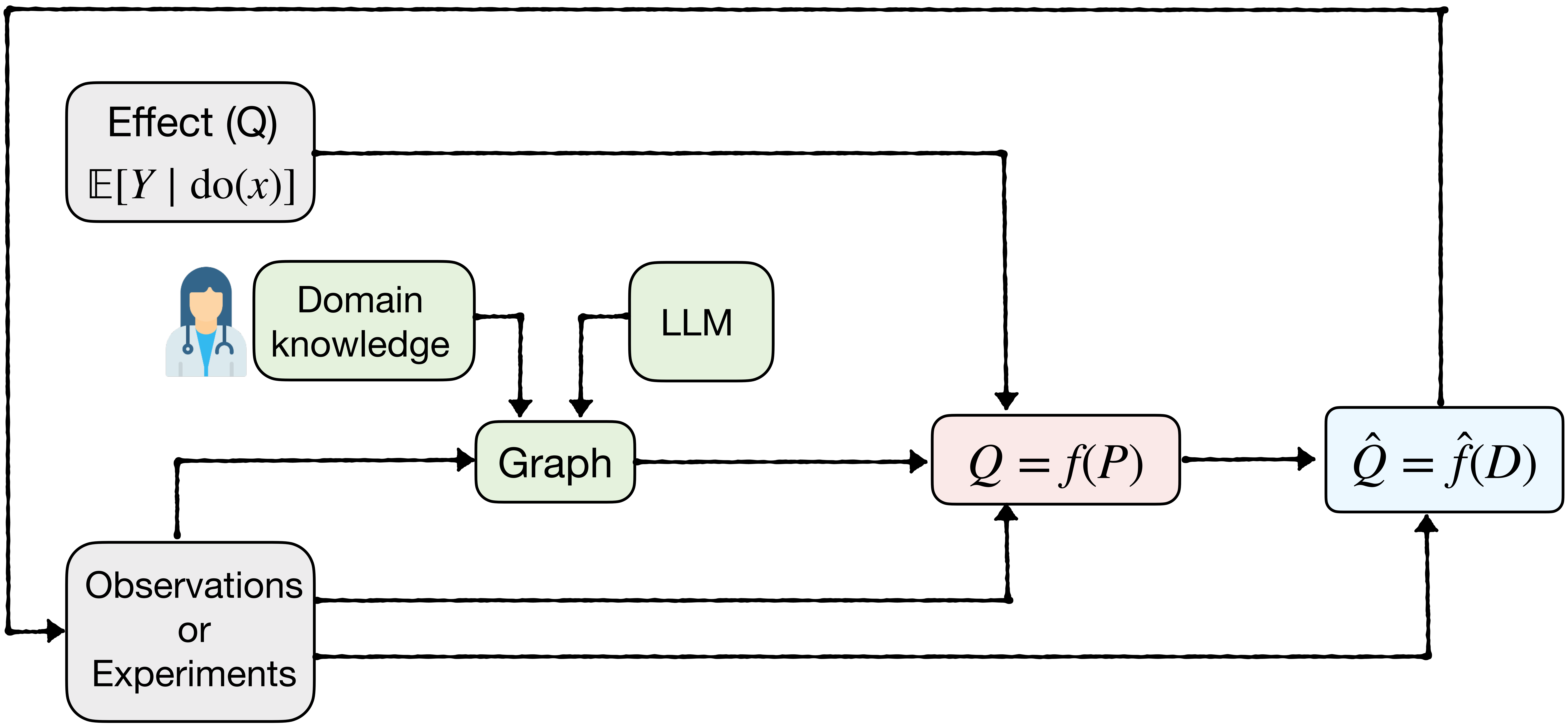
Future 3: Causal Inference Loop with Uncertainty



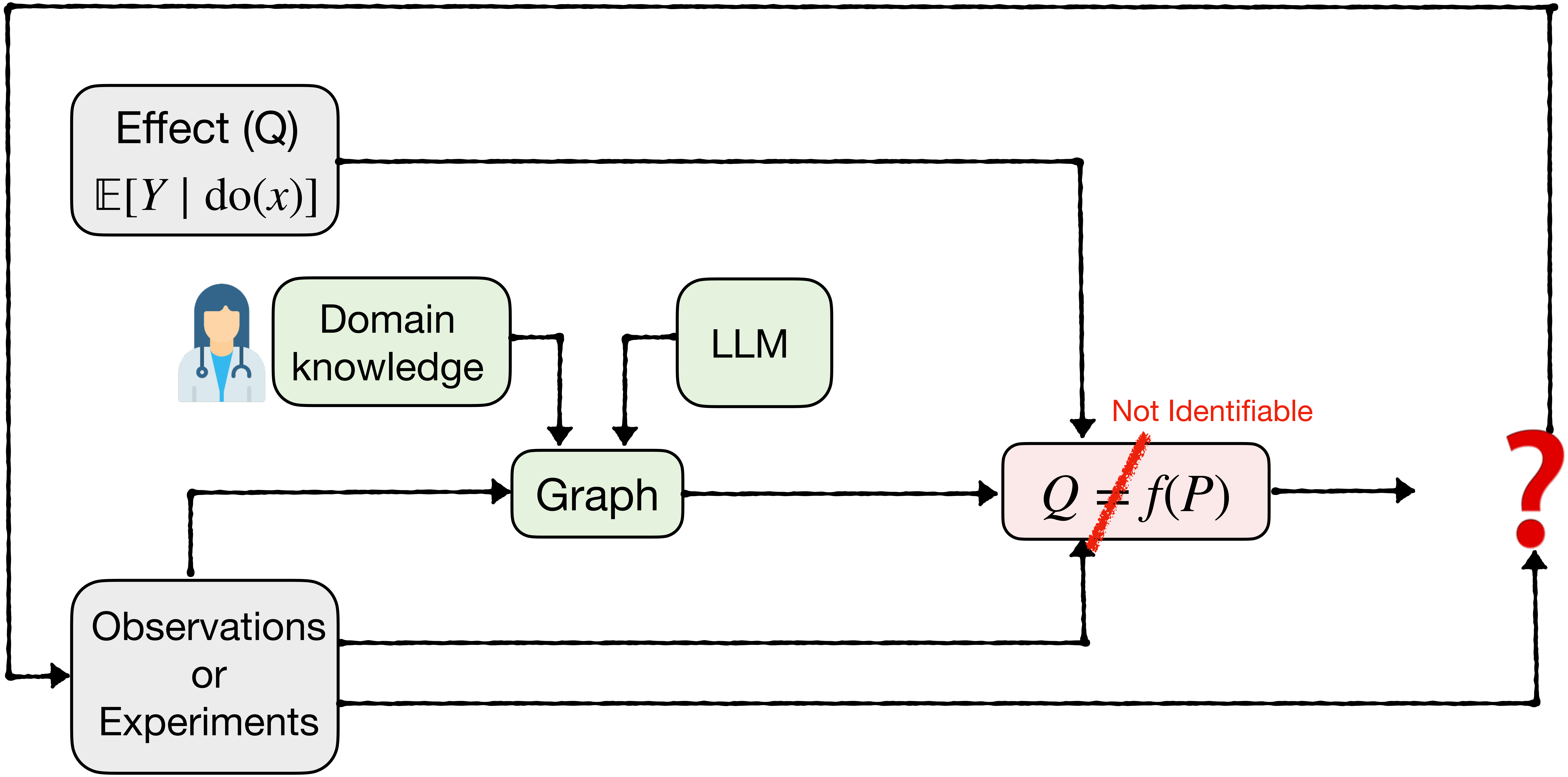
Future 3: Causal Inference Loop with Uncertainty



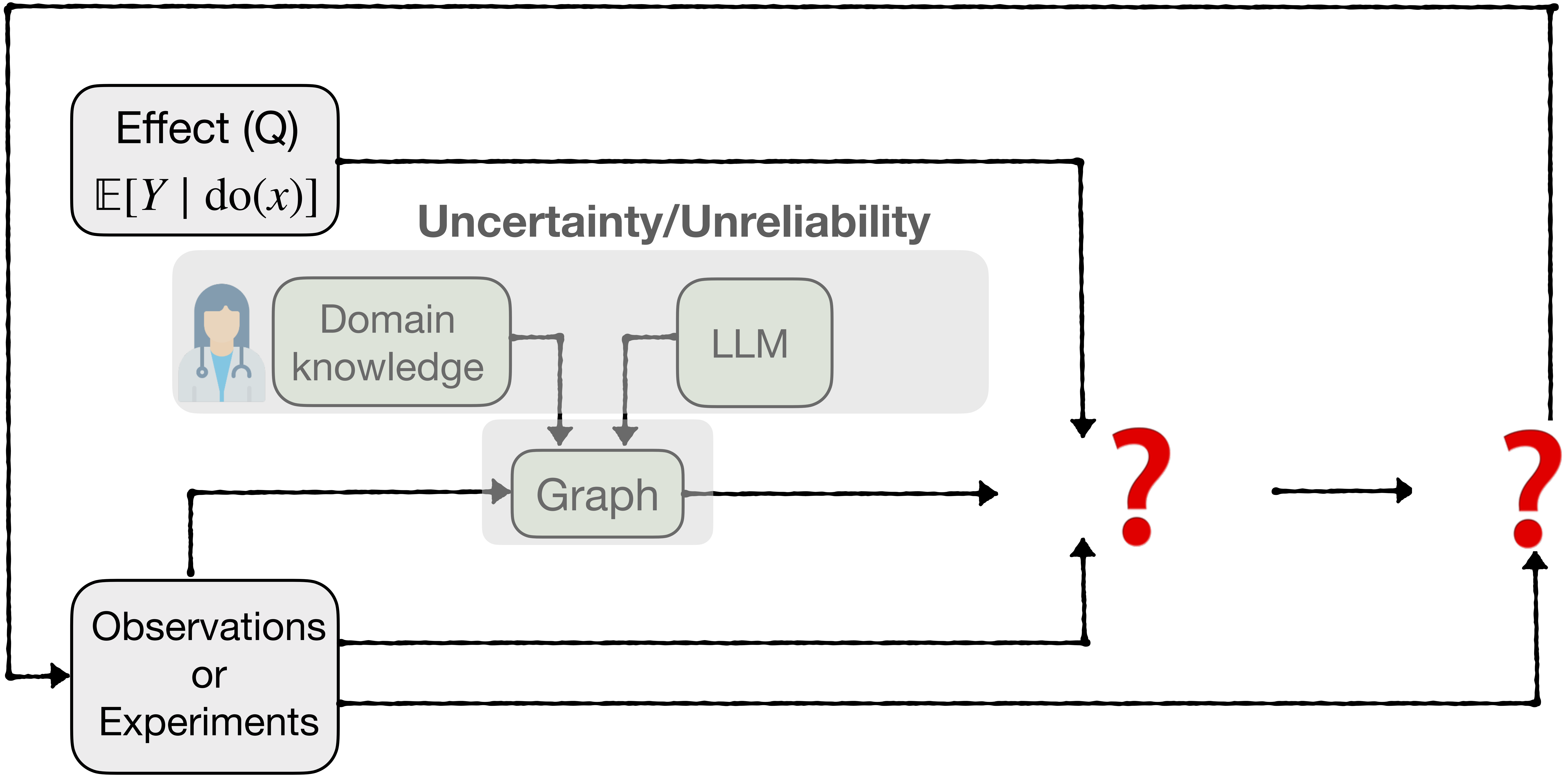
Future 3: Causal Inference Loop with Uncertainty



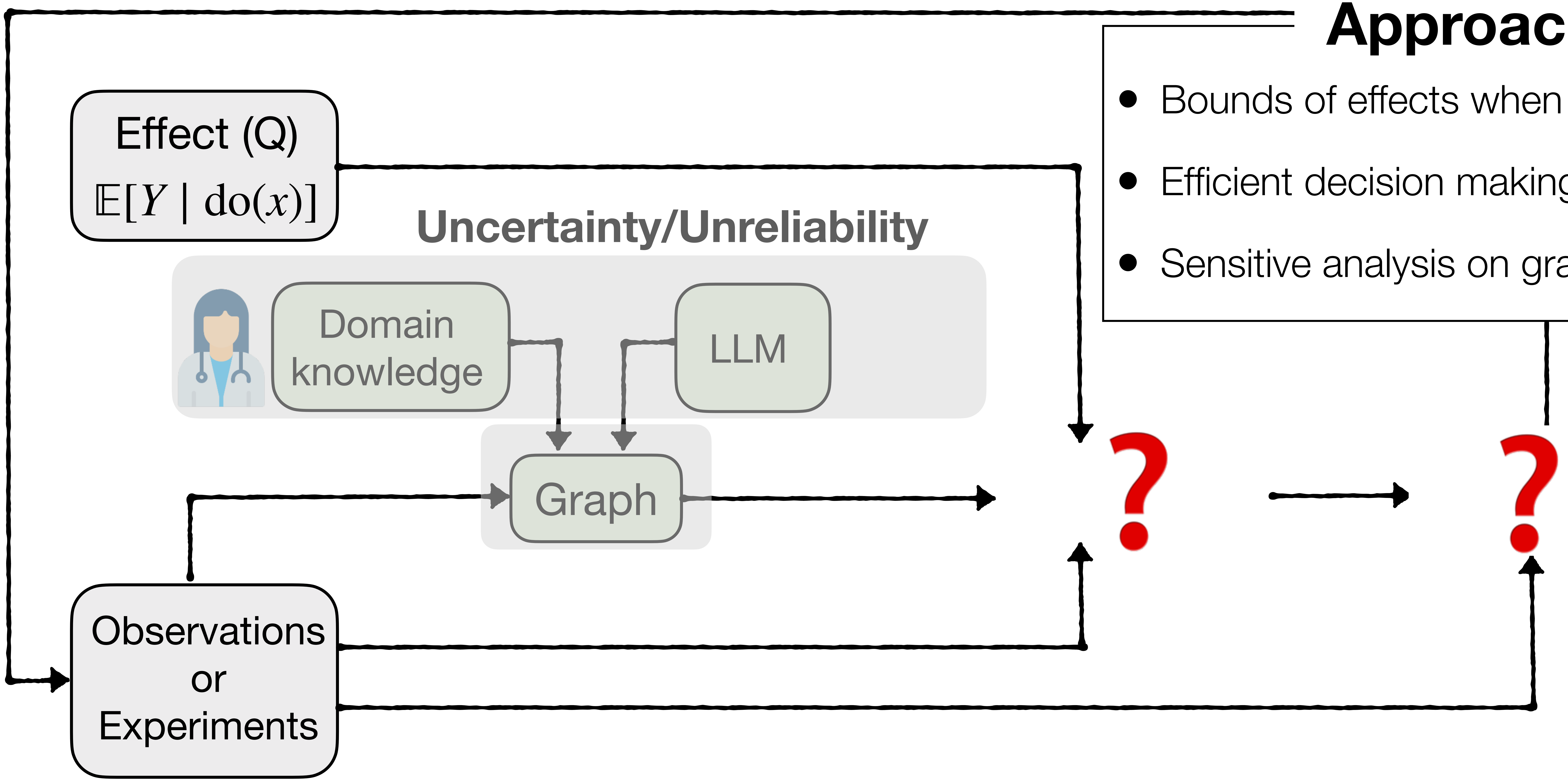
Future 3: Causal Inference Loop with Uncertainty



Future 3: Causal Inference Loop with Uncertainty



Future 3: Causal Inference Loop with Uncertainty



Approach

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