Estimating Joint Treatment Effect from Marginal Experiments

based on:
Estimating Joint Treatment Effects by Combining Multiple Experiments, ICML-23

Yonghan Jung

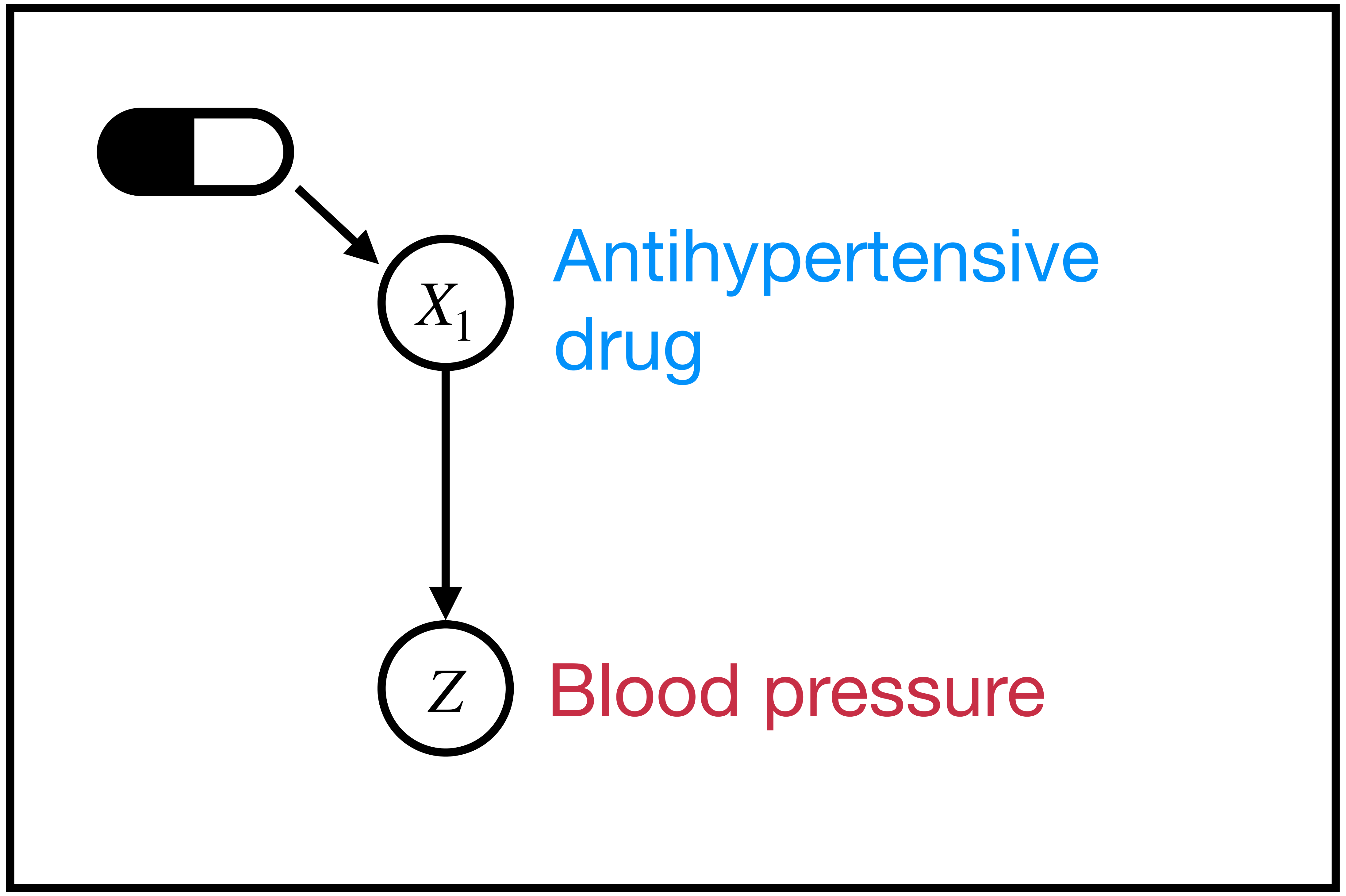
Dept. of Computer Science, Purdue University

yonghanjung.me

2023 Fall QM Seminar

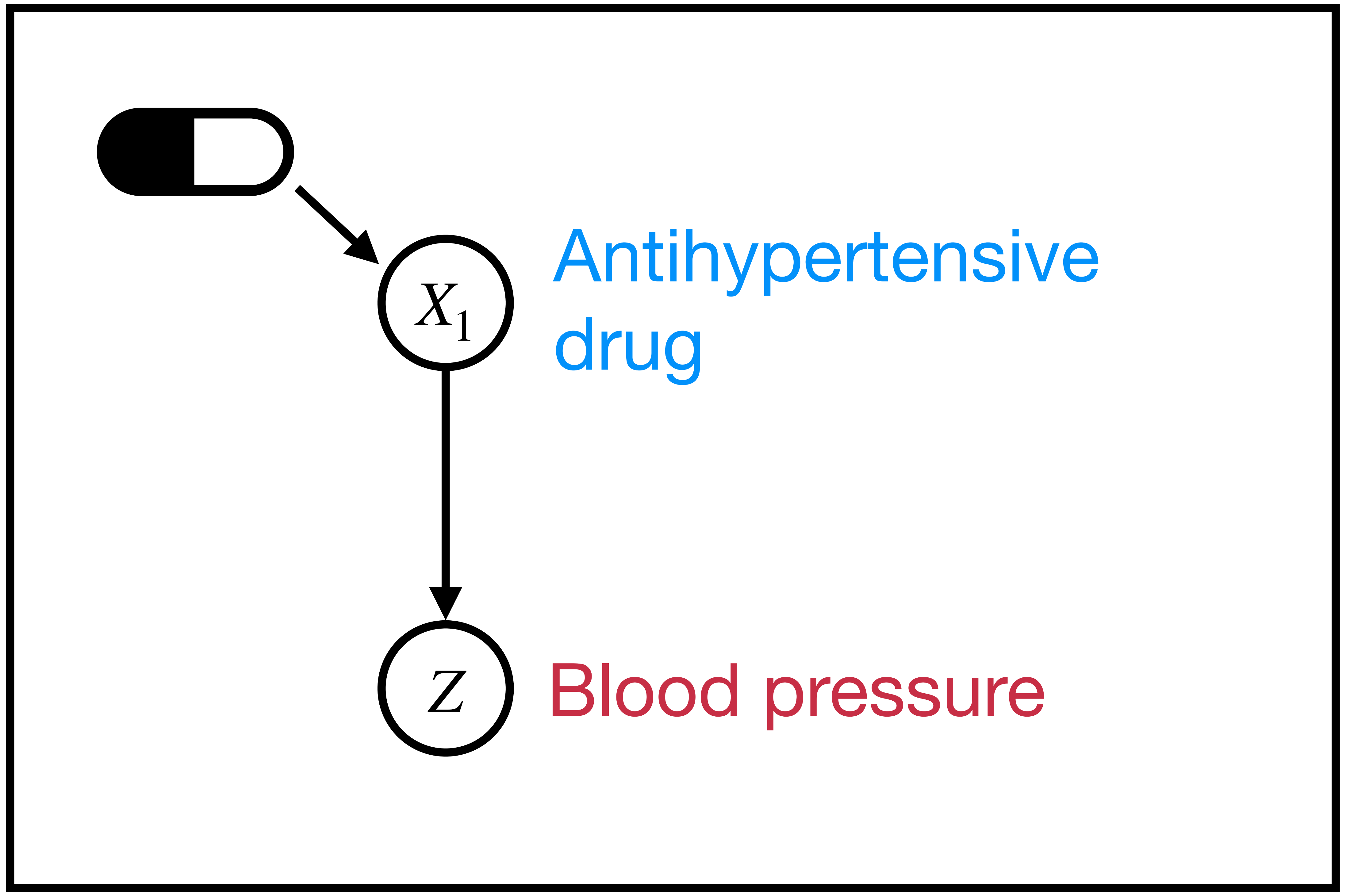
Treatment-Treatment interaction

Treatment-Treatment interaction



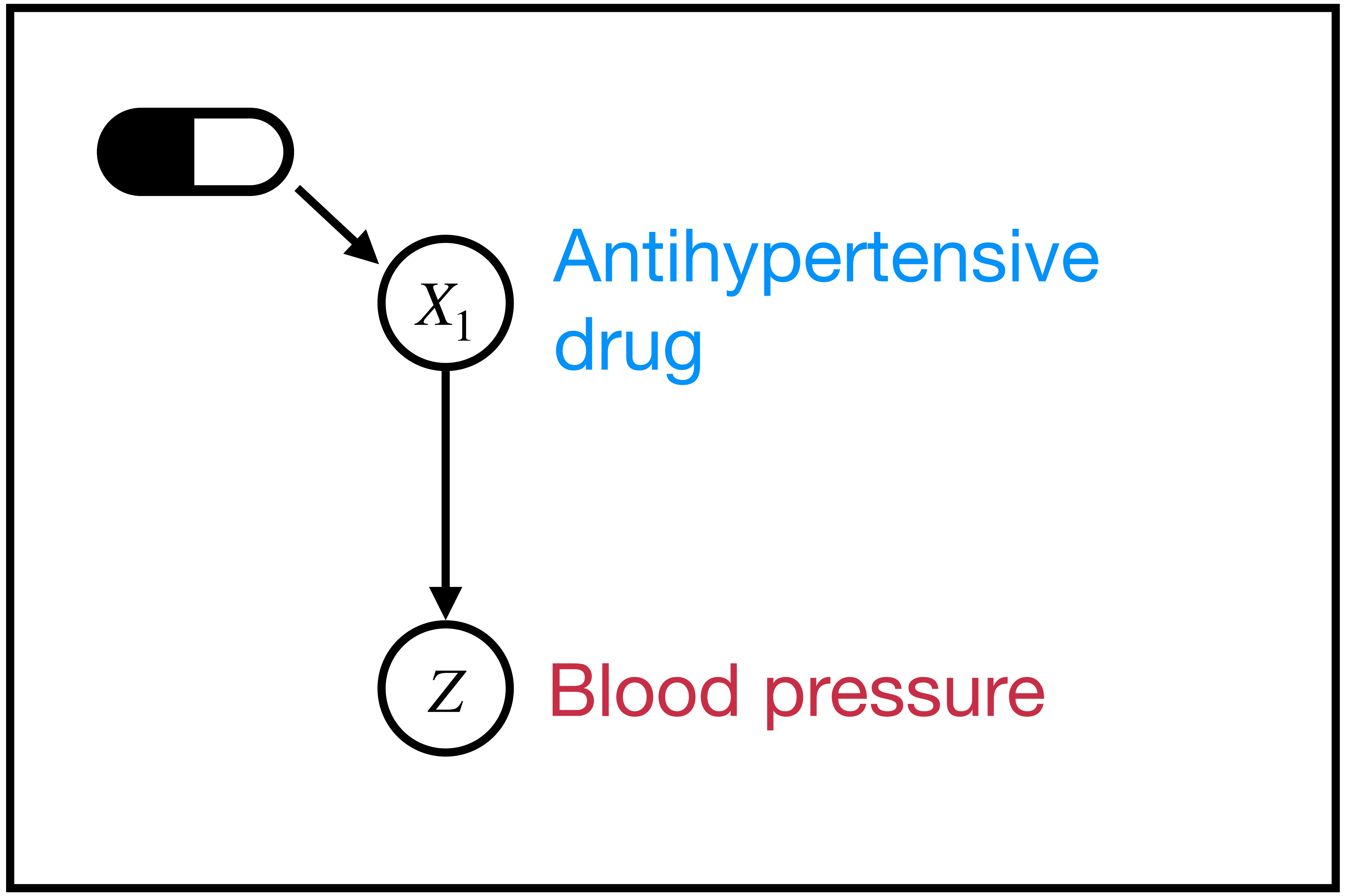
Experiment on X_1

Treatment-Treatment interaction

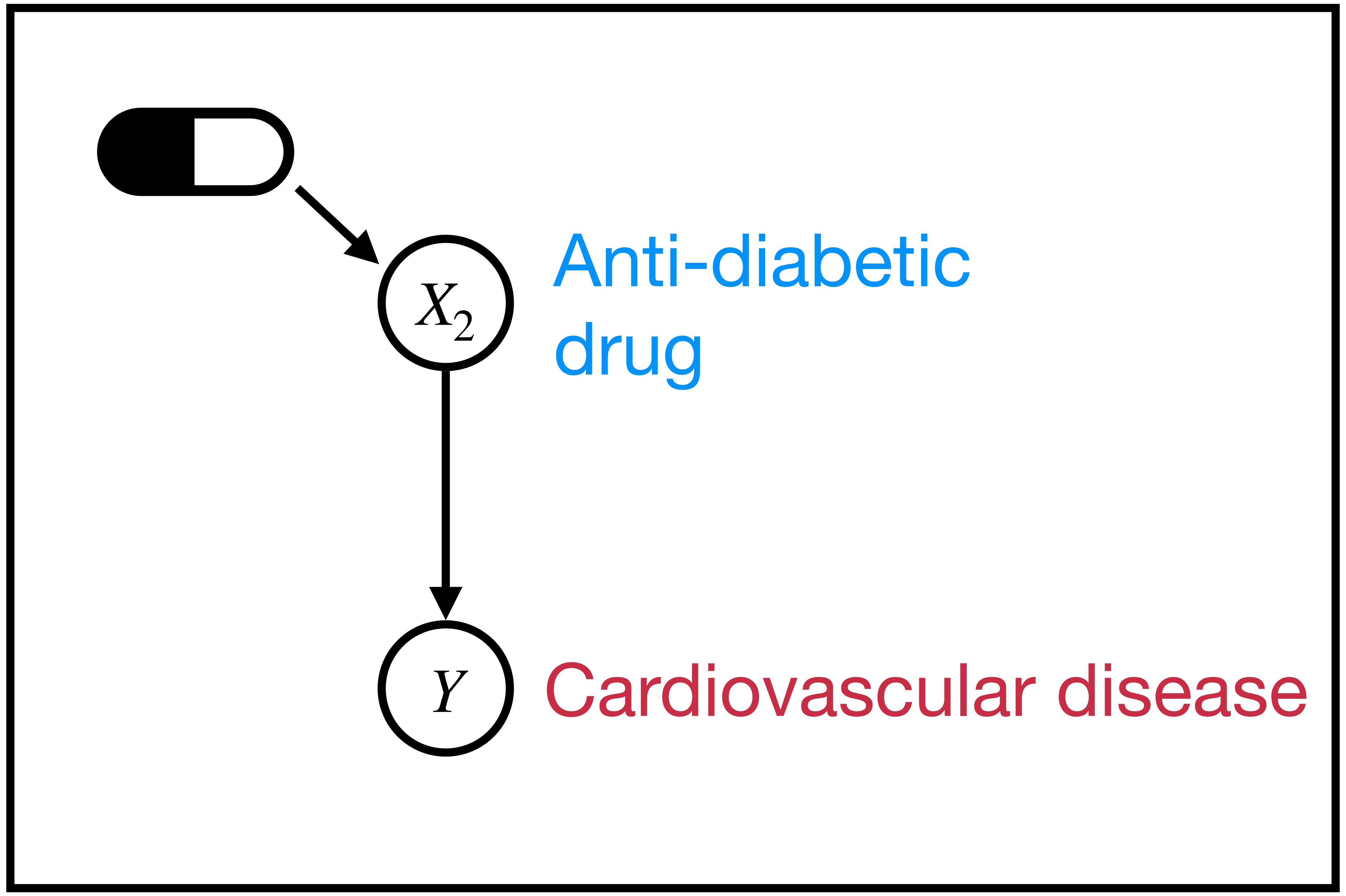


Experiment on X_1 😃

Treatment-Treatment interaction

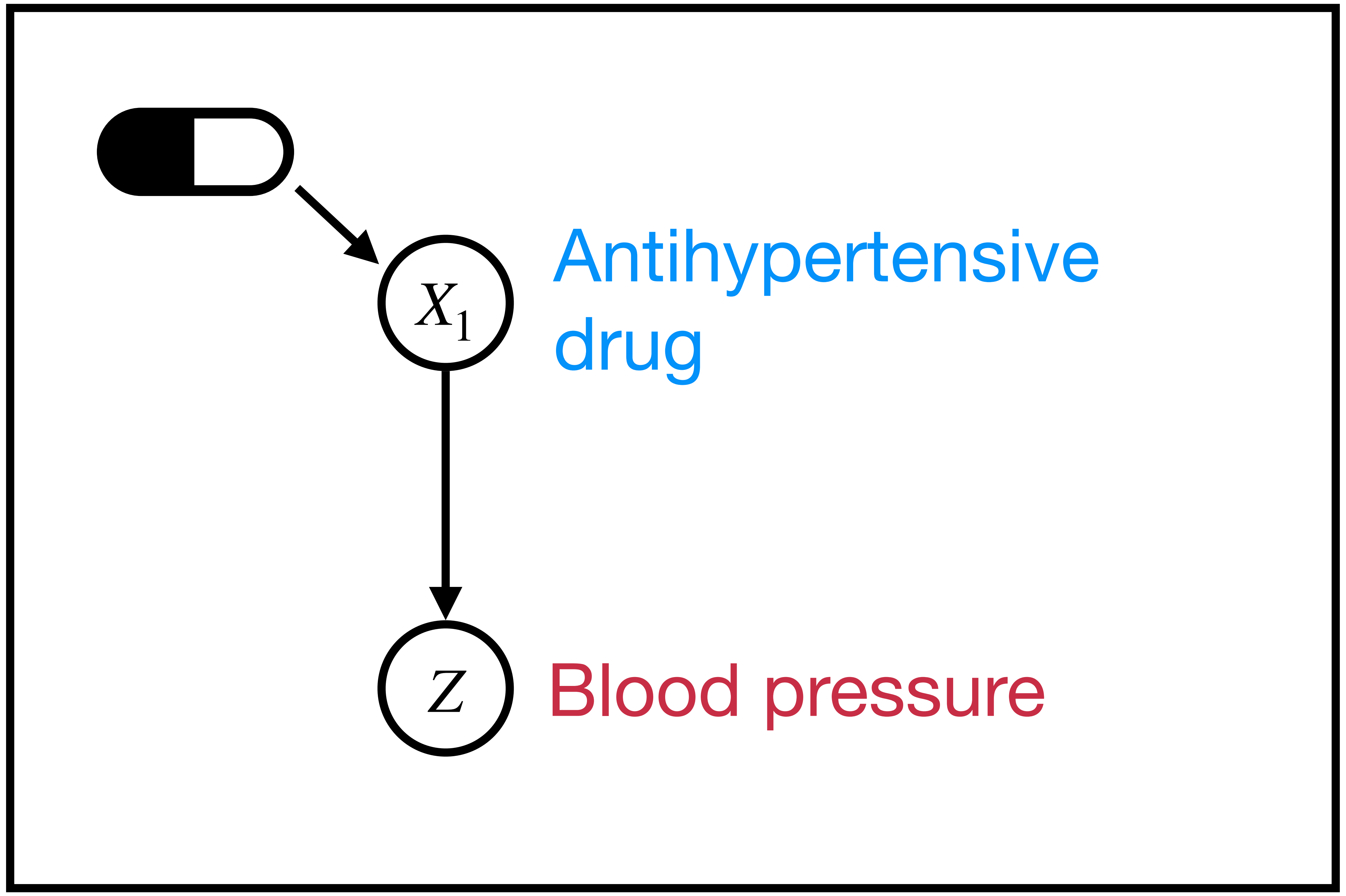


Experiment on X_1 😃

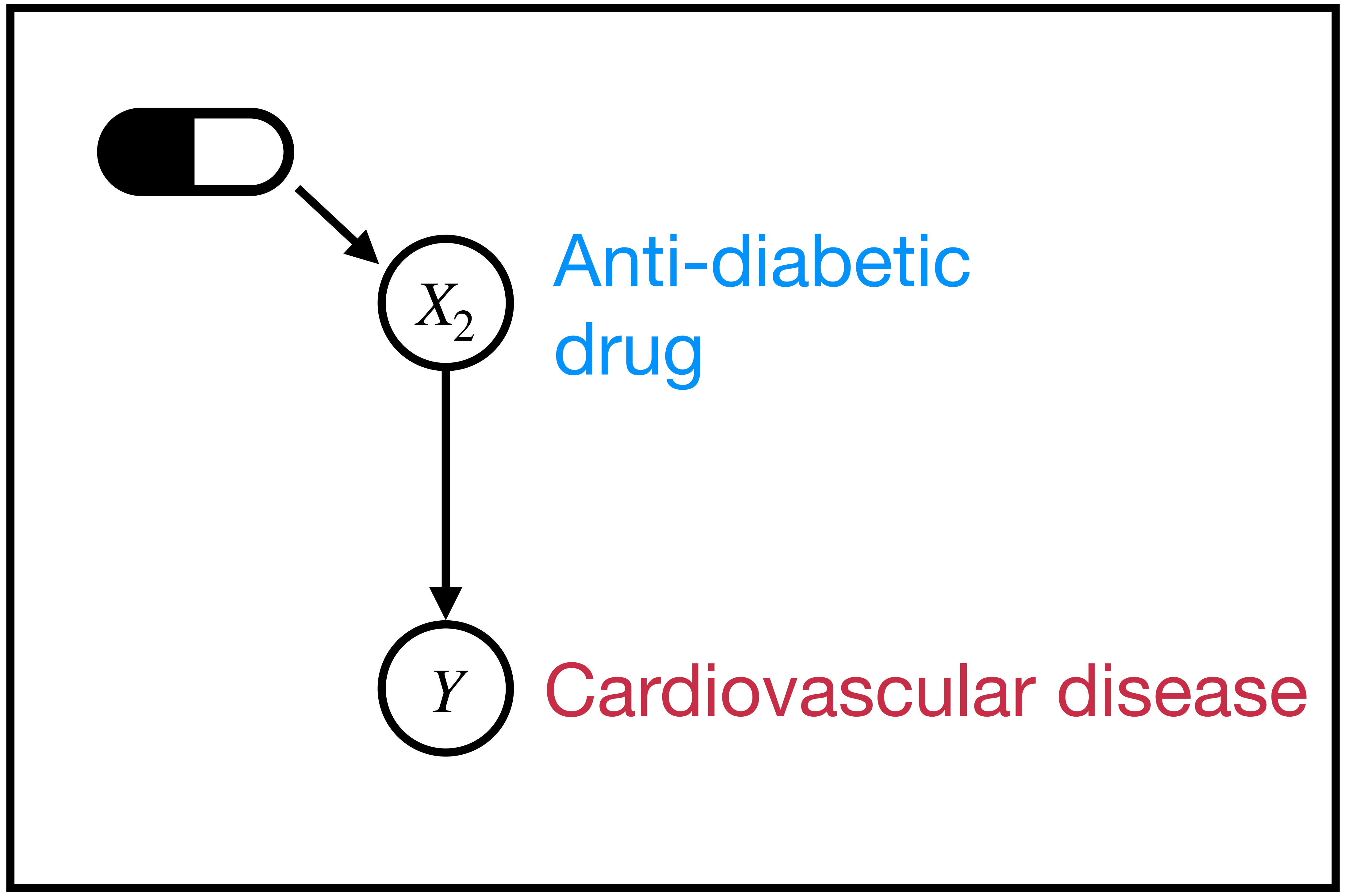


Experiment on X_2

Treatment-Treatment interaction



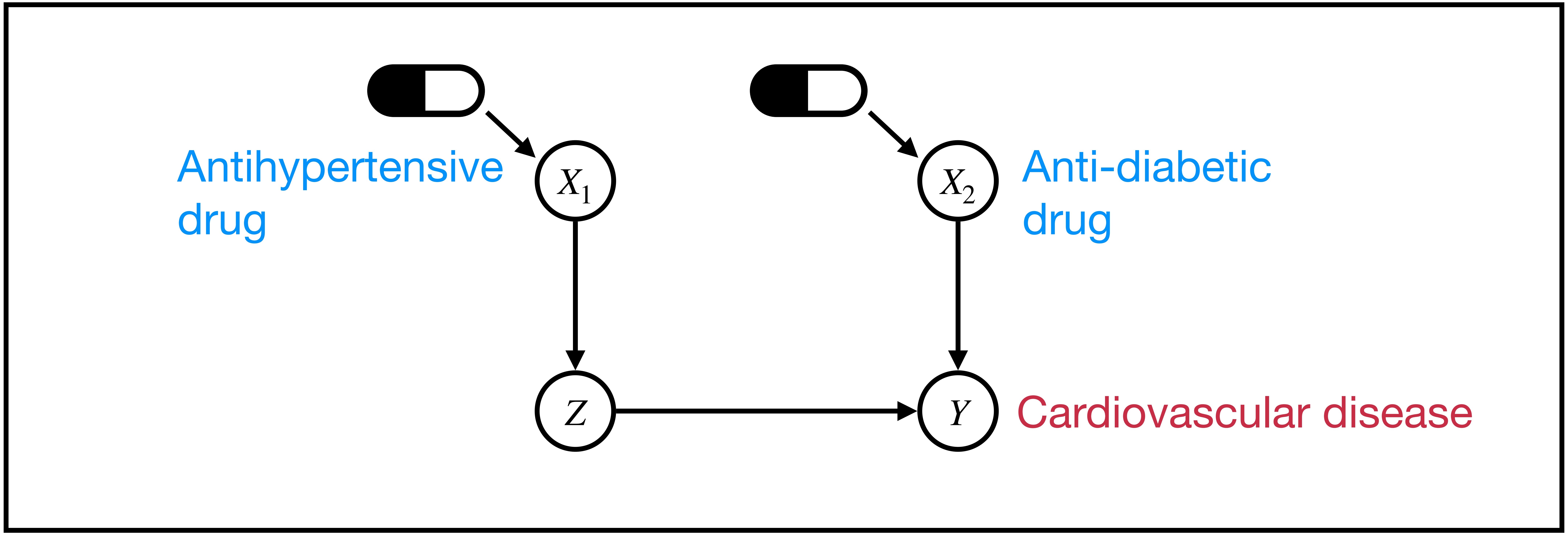
Experiment on X_1 😃



Experiment on X_2 😃

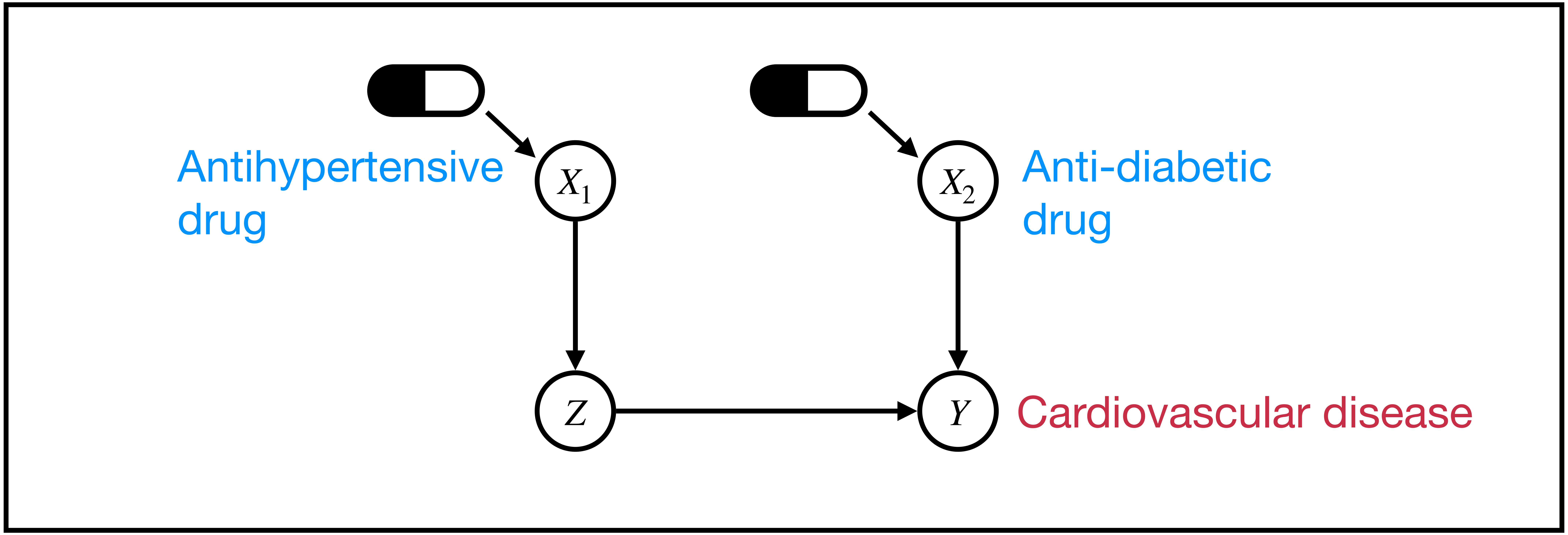
Treatment-Treatment interaction

Treatment-Treatment interaction



Experiment on X_1 and X_2

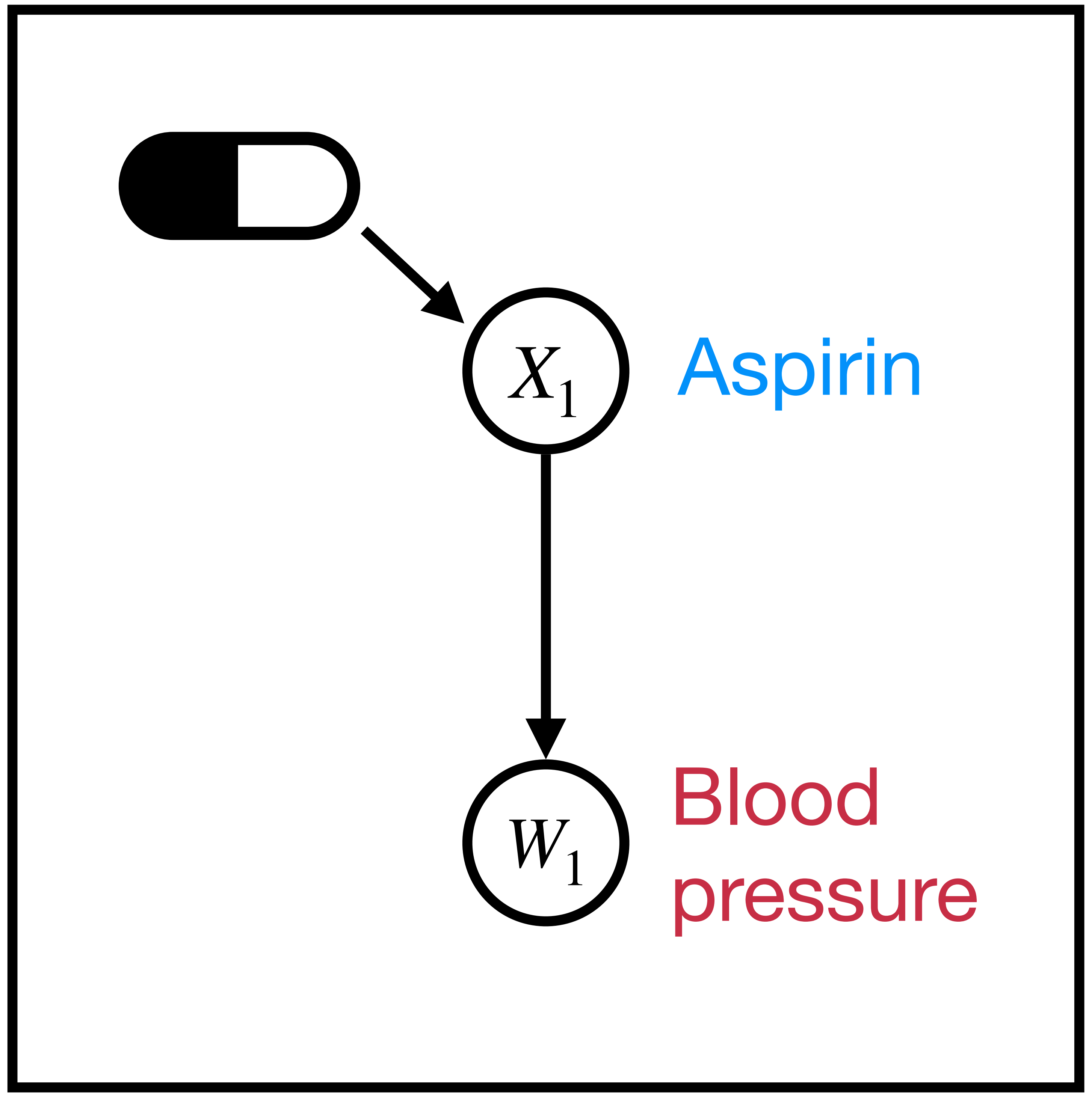
Treatment-Treatment interaction



Experiment on X_1 and X_2 😡

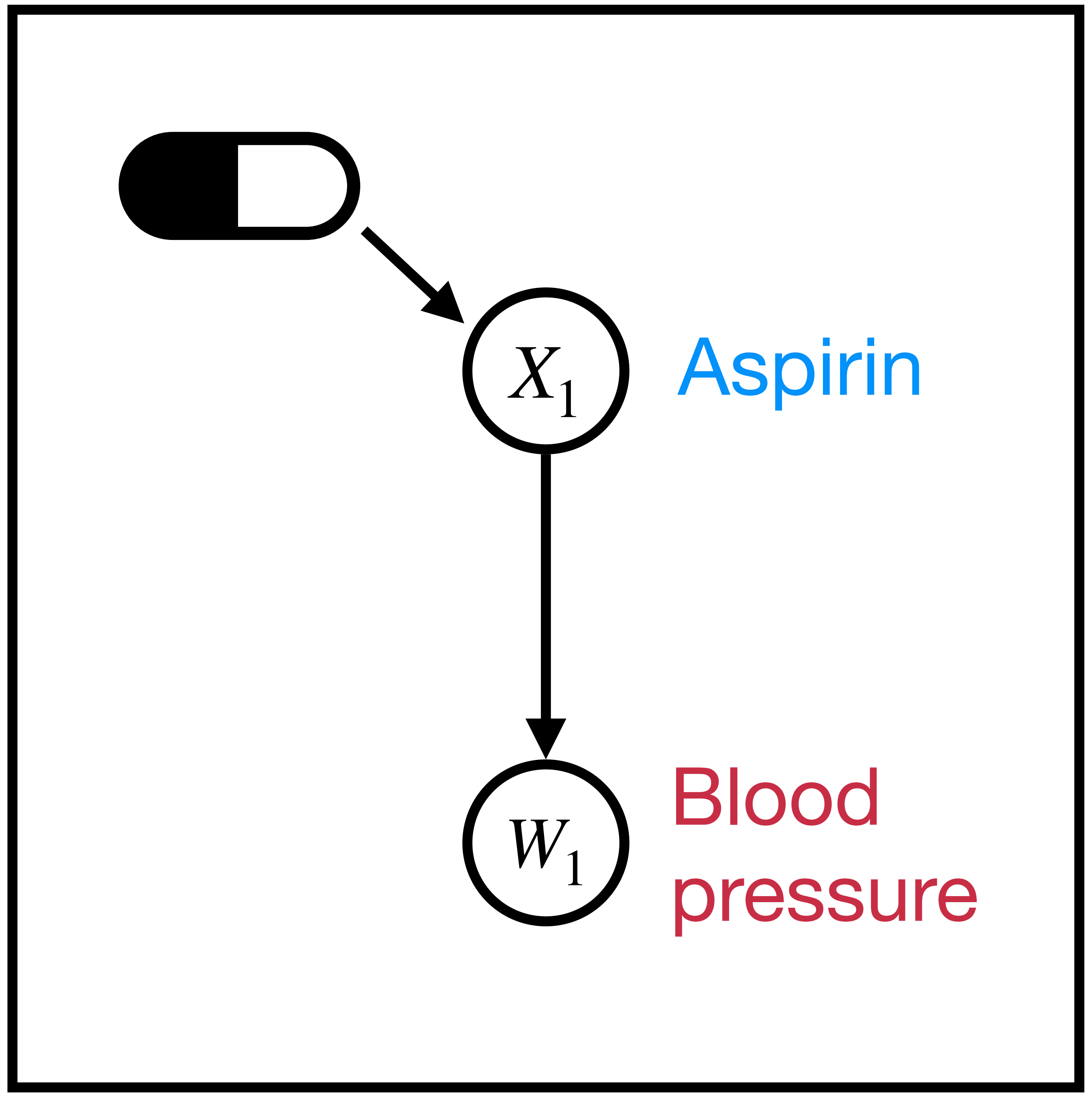
Multiple-Treatment interaction

Multiple-Treatment interaction

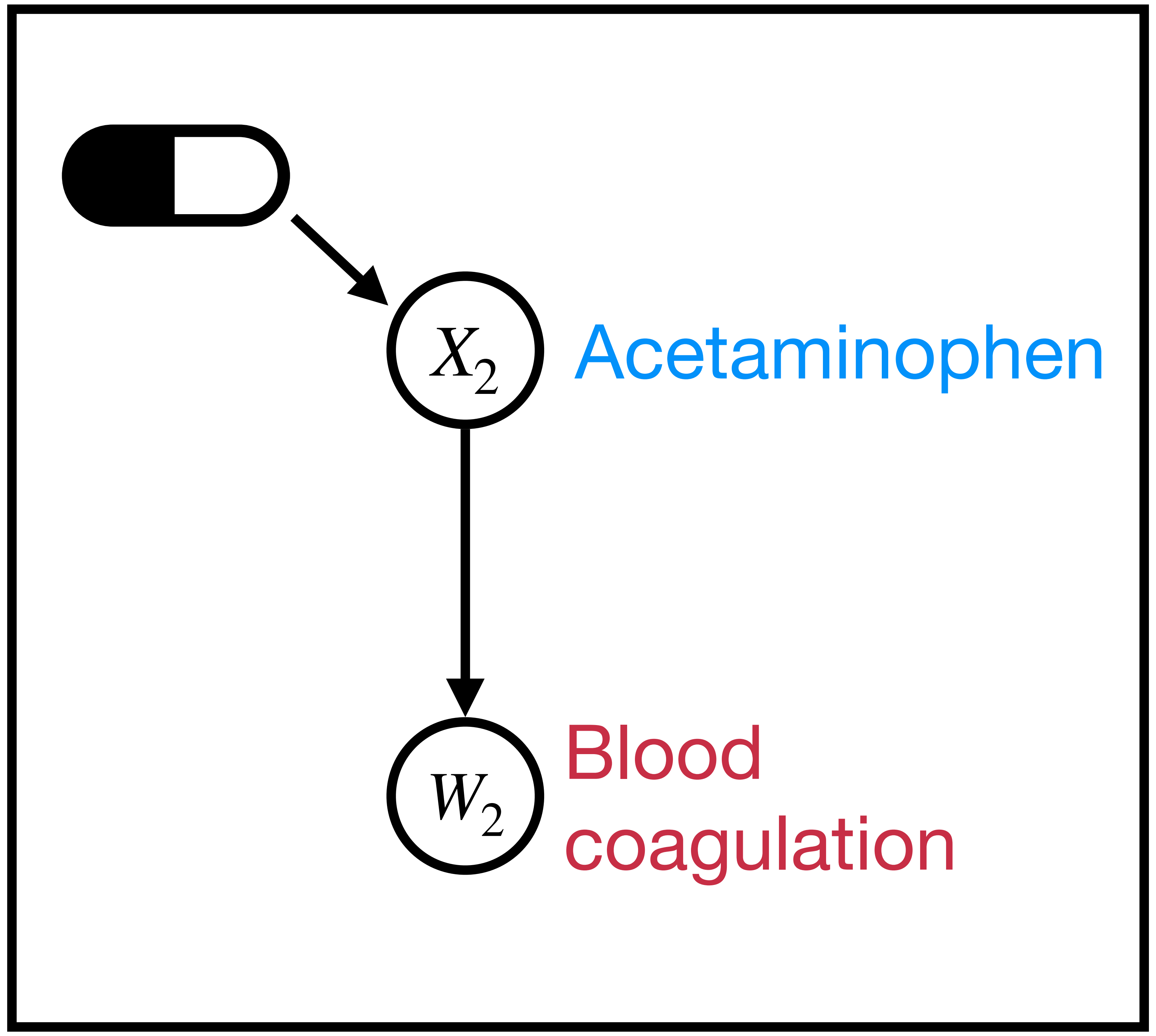


Experiment on X_1 😃

Multiple-Treatment interaction

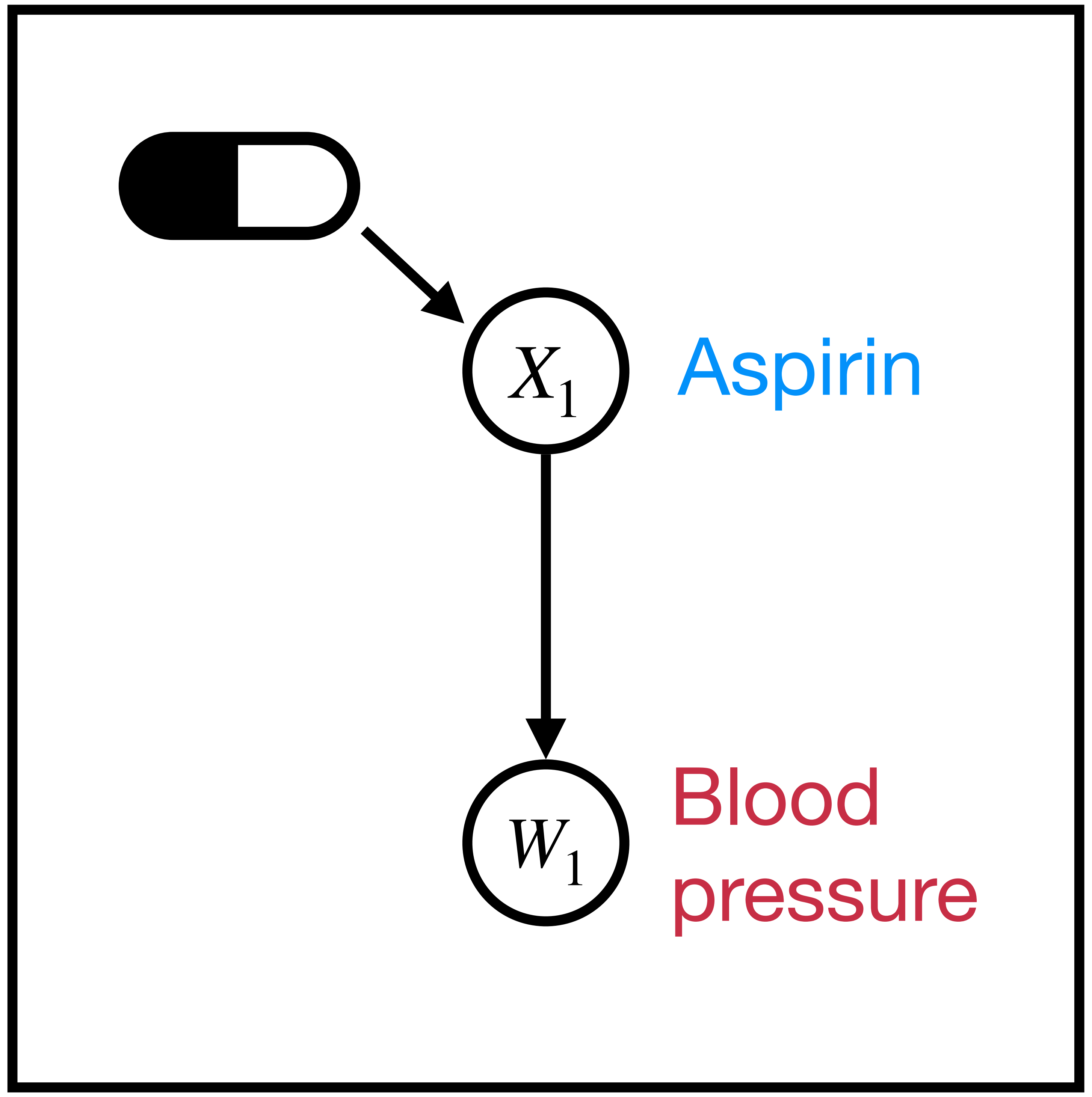


Experiment on X_1

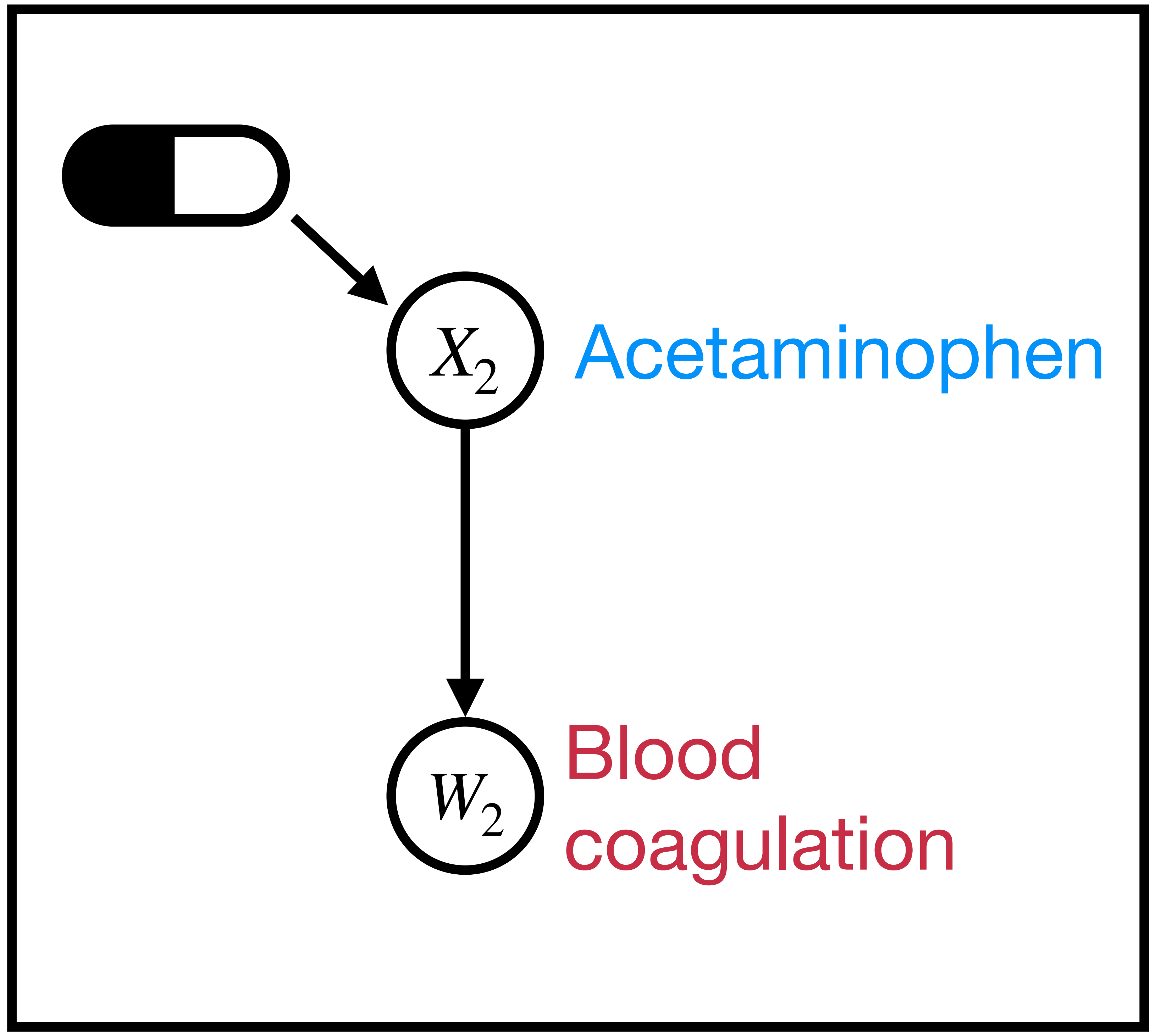


Experiment on X_2

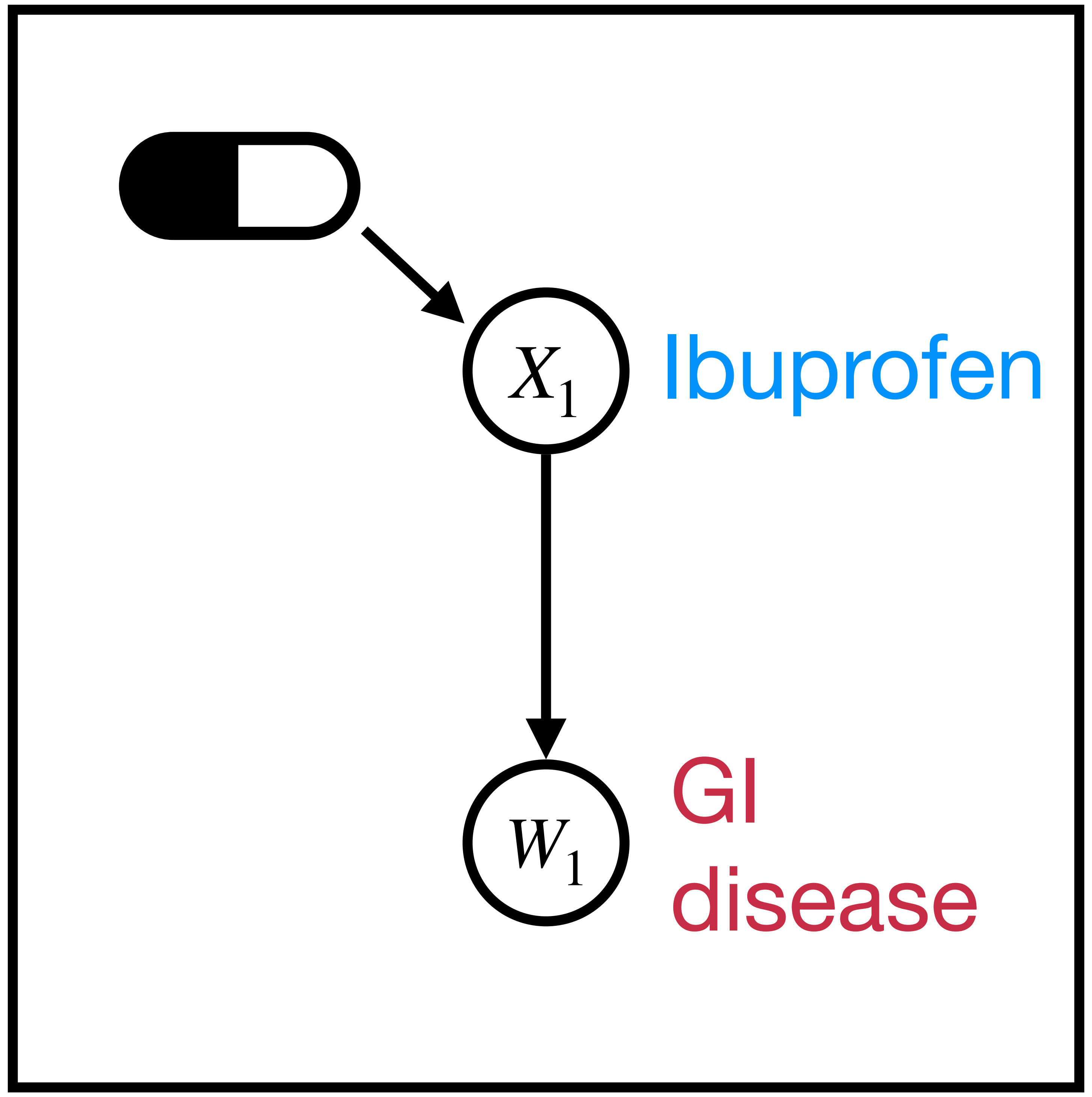
Multiple-Treatment interaction



Experiment on X_1 😃



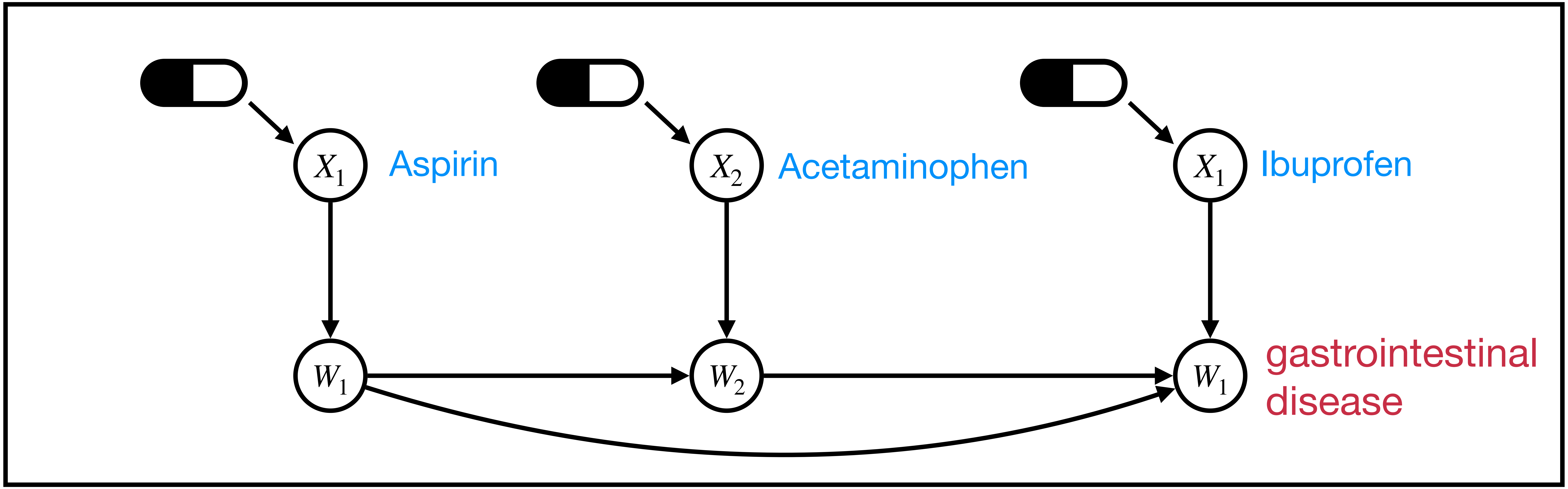
Experiment on X_2 😃



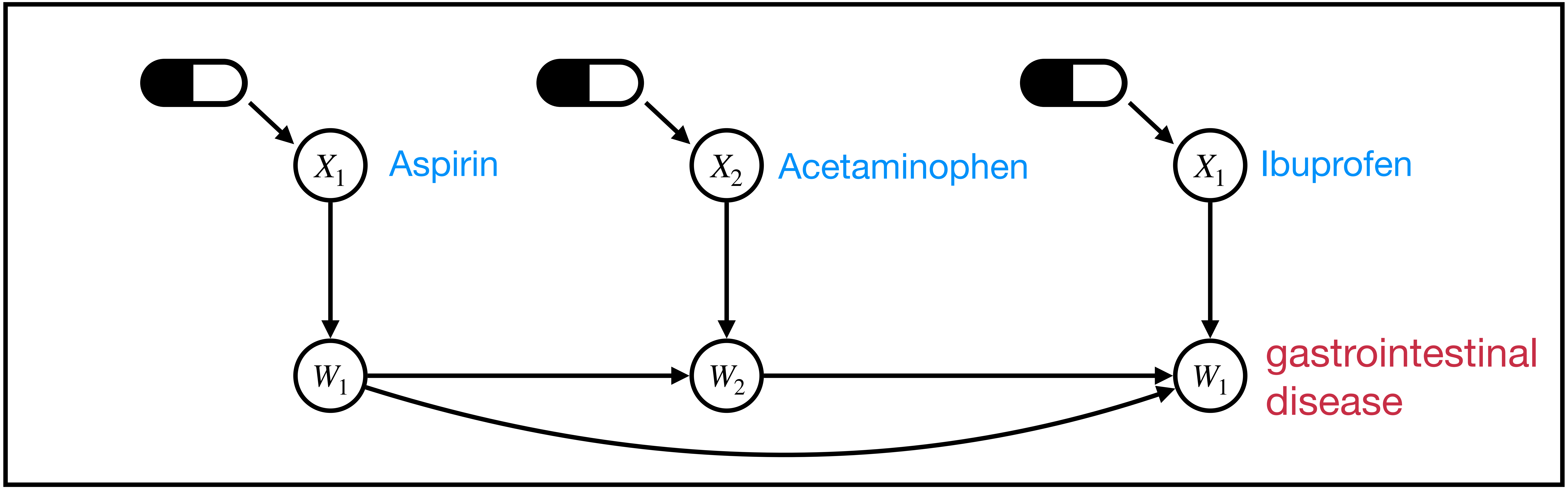
Experiment on X_3 😃

Multiple-Treatment interaction

Multiple-Treatment interaction

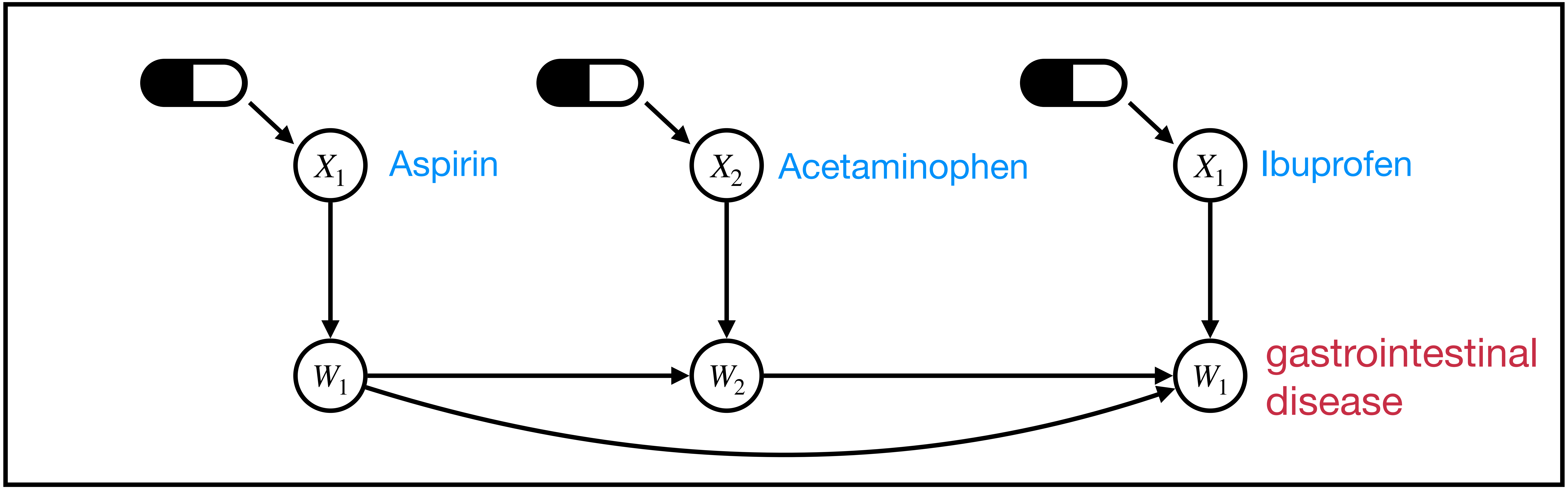


Multiple-Treatment interaction



Experiment on X_1, X_2, X_3

Multiple-Treatment interaction



Experiment on X_1, X_2, X_3 😡

Practical challenges

Practical challenges

1. Many randomized controlled trials are conducted to evaluate the effectiveness of marginal treatments, which involve intervening with a single treatment.

Practical challenges

1. Many randomized controlled trials are conducted to evaluate the effectiveness of marginal treatments, which involve intervening with a single treatment.

2. Trials for estimating the joint treatment effect can be costly, especially when multiple treatments are involved, as this requires a large number of participants.

Practical challenges

1. Many randomized controlled trials are conducted to evaluate the effectiveness of marginal treatments, which involve intervening with a single treatment.

2. Trials for estimating the joint treatment effect can be costly, especially when multiple treatments are involved, as this requires a large number of participants.

3. What about leveraging existing marginal experiments to estimate the joint treatment effect?

Practical challenges

1. Many randomized controlled trials are conducted to evaluate the effectiveness of marginal treatments, which involve intervening with a single treatment.

2. Trials for estimating the joint treatment effect can be costly, especially when multiple treatments are involved, as this requires a large number of participants.

3. What about leveraging existing marginal experiments to estimate the joint treatment effect?

This talk: Combining marginal experiments to estimate joint treatment etfects

Outline of this talk

1. Preliminary and Problem Setup
   (1) Structural Causal Model
2. Treatment-Treatment Interaction
   (1) Identification
   (2) Estimand
   (3) Estimation and Error Analysis
   (4) Simulation Results
3. Multiple-Treatment Interaction (+ Omitted Results)
4. Future directions & Summary

Outline of this talk

1. **Preliminary and Problem Setup**
   (1) Structural Causal Model
2. Treatment-Treatment Interaction
   (1) Identification
   (2) Estimand
   (3) Estimation and Error Analysis
   (4) Simulation Results
3. Multiple-Treatment Interaction (+ Omitted Results)
4. Future directions & Summary

Structural Causal Model: DGP

Structural Causal Model (SCM) $\langle \mathbf{V}, \mathbf{U}, \mathbf{F}, P(\mathbf{U}) \rangle$

Structural Causal Model: DGP

Structural Causal Model (SCM) $\langle \mathbf{V}, \mathbf{U}, \mathbf{F}, P(\mathbf{U}) \rangle$

- $\mathbf{V}$: A set of observable variables.
- $\mathbf{U}$: A set of latent variables.
- $\mathbf{F}$: A set of functions $\{f_{V_i}\}_{V_i \in \mathbf{V}}$ determining the value of $V_i \in \mathbf{V}$; i.e.,

  $V_i \leftarrow f_{v_i}(PA_{V_i}, U_{V_i})$ for some $PA_{V_i} \subseteq \mathbf{V}$ and $U_{V_i} \subseteq \mathbf{U}$.
- $P(\mathbf{U})$: A probability distribution for $\mathbf{U}$.

Causal Graphical Model

Causal Graphical Model

SCM

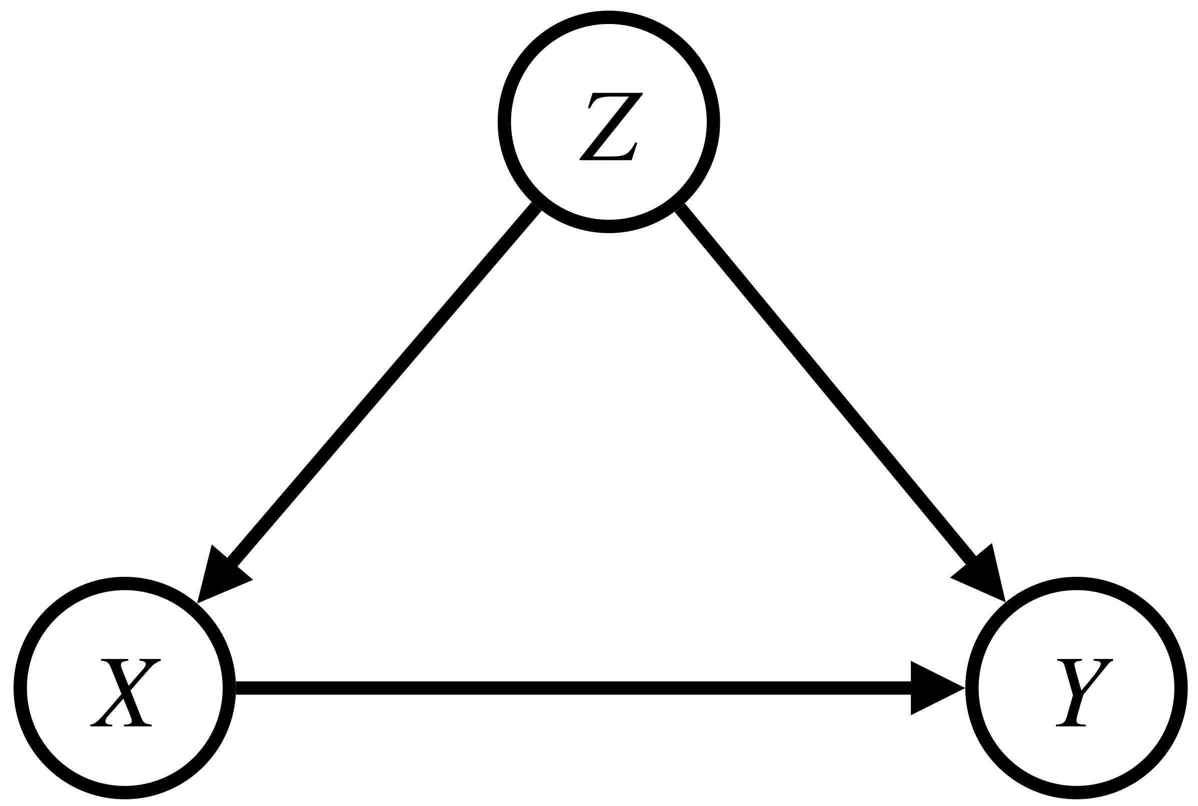
$$U_Z, U_X, U_Y \sim \text{normal}(0,1)$$
$$Z \leftarrow f_Z(U_Z)$$
$$X \leftarrow f_X(Z, U_X)$$
$$Y \leftarrow f_Y(X, Z, U_Y)$$

Causal Graphical Model

SCM

$$U_Z, U_X, U_Y \sim \text{normal}(0,1)$$
$$Z \leftarrow f_Z(U_Z)$$
$$X \leftarrow f_X(Z, U_X)$$
$$Y \leftarrow f_Y(X, Z, U_Y)$$

Graph

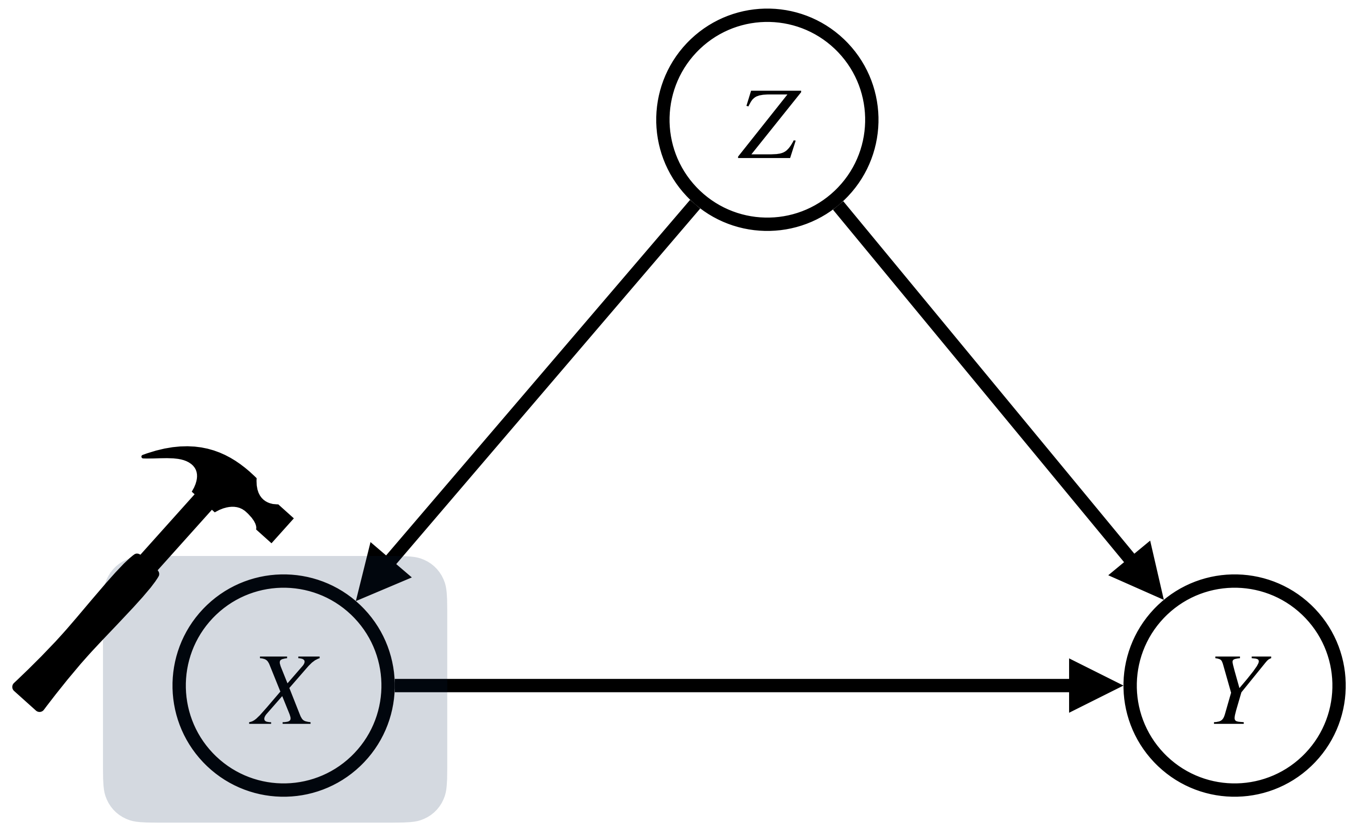


Causal Graphical Model: Intervention

SCM

$$U_Z, U_X, U_Y \sim \text{normal}(0,1)$$
$$Z \leftarrow f_Z(U_Z)$$
$$X \leftarrow f_X(Z, U_X)$$
$$Y \leftarrow f_Y(X, Z, U_Y)$$

Graph



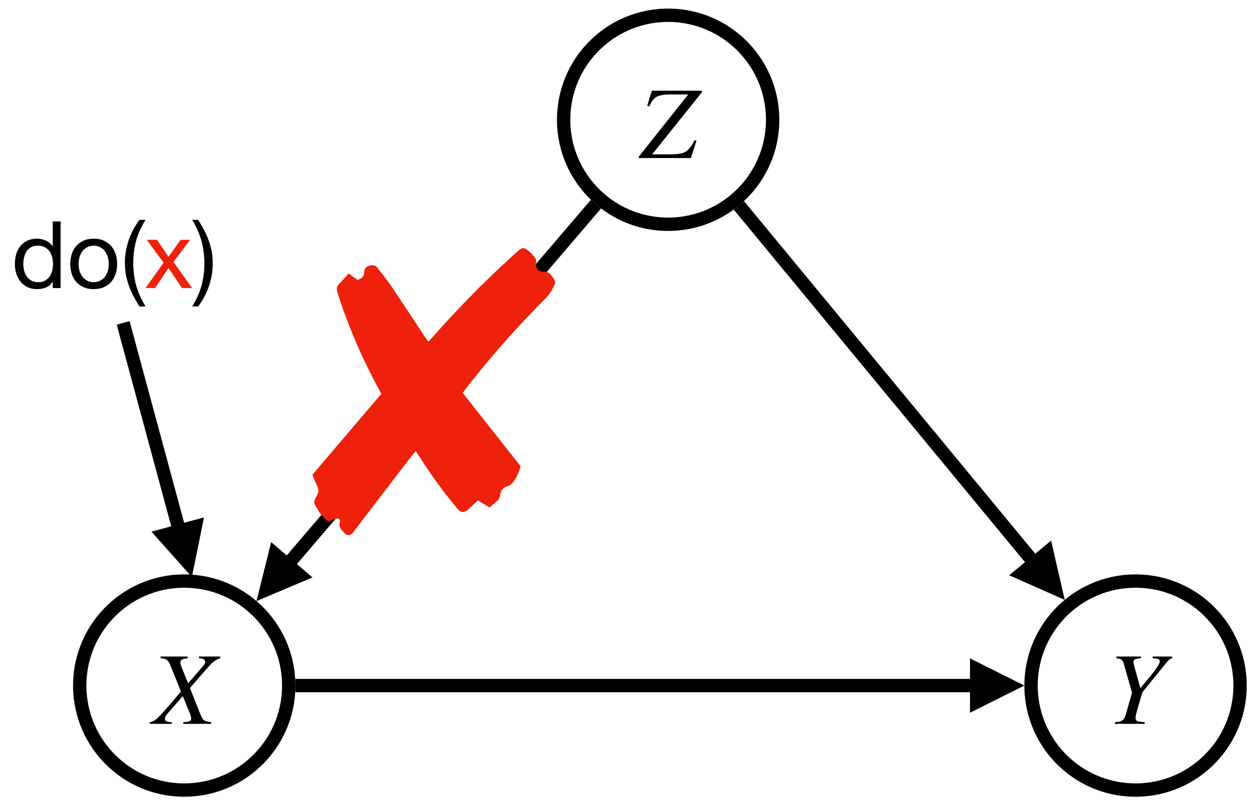
Causal Graphical Model: Intervention

SCM

$U_Z, U_X, U_Y \sim$ normal$(0,1)$
$Z \leftarrow f_Z(U_Z)$

$X \leftarrow x$ (= do(x))
$Y \leftarrow f_Y(x, Z, U_Y)$

Graph



Causal Graphical Model: Intervention

SCM

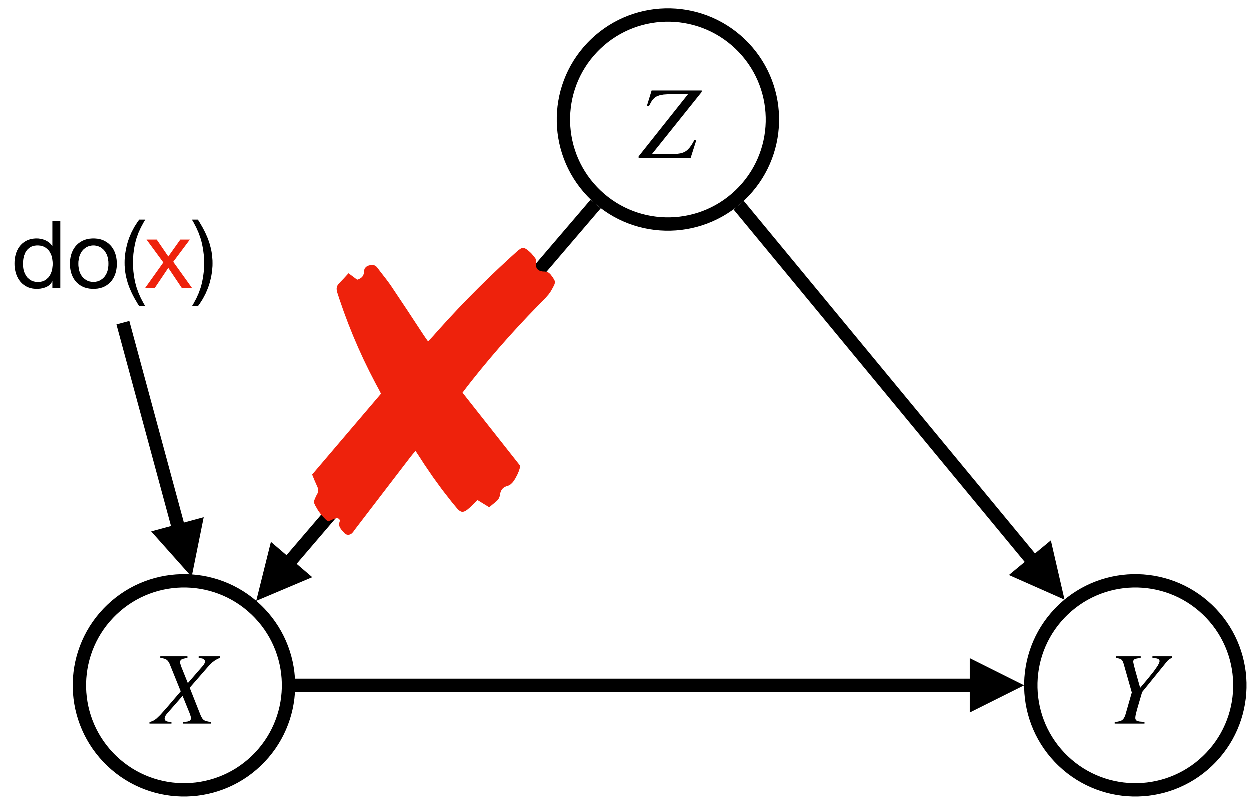
$U_Z, U_X, U_Y \sim \text{normal}(0,1)$

$Z \leftarrow f_Z(U_Z)$

$X \leftarrow x$ (= do(x))

$Y \leftarrow f_Y(x, Z, U_Y)$

Graph



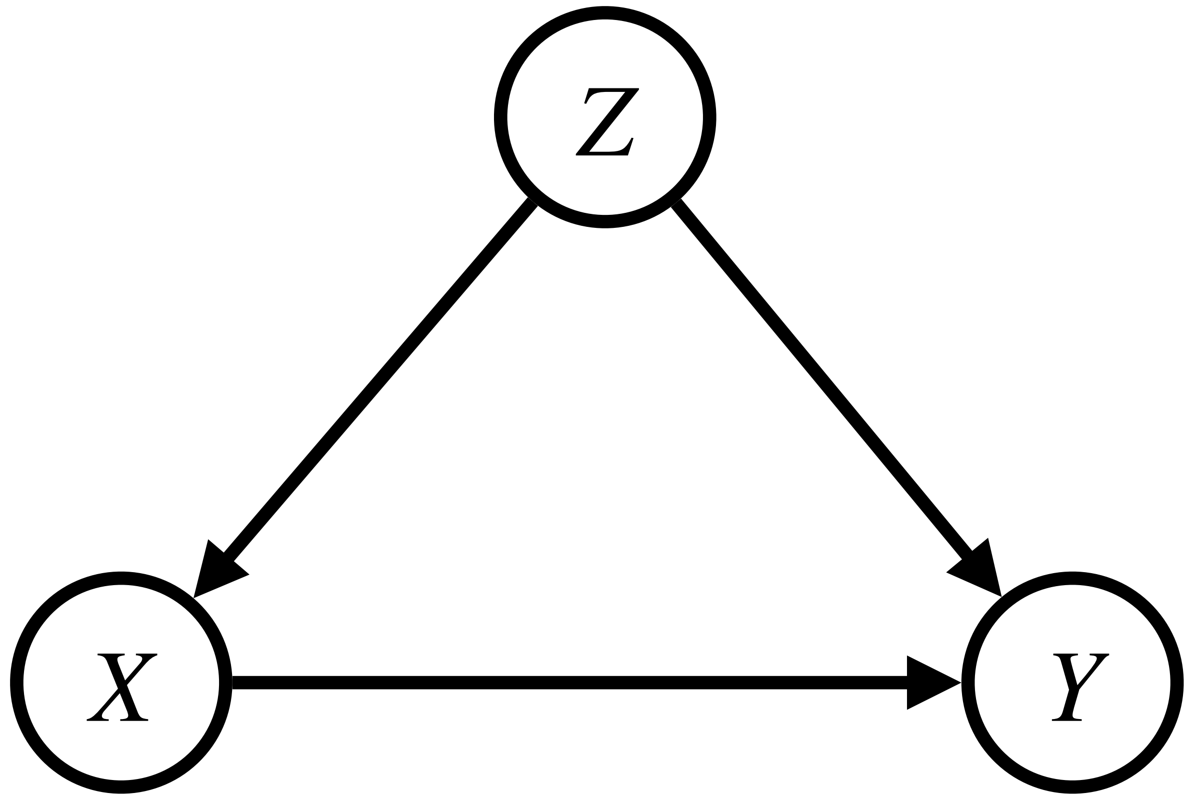
$G_{\overline{X}}$

Causal Graphical Model: Potential Outcome

SCM

$$U_Z, U_X, U_Y \sim \text{normal}(0,1)$$
$$Z \leftarrow f_Z(U_Z)$$
$$X \leftarrow f_X(Z, U_X)$$
$$Y \leftarrow f_Y(X, Z, U_Y)$$

Graph

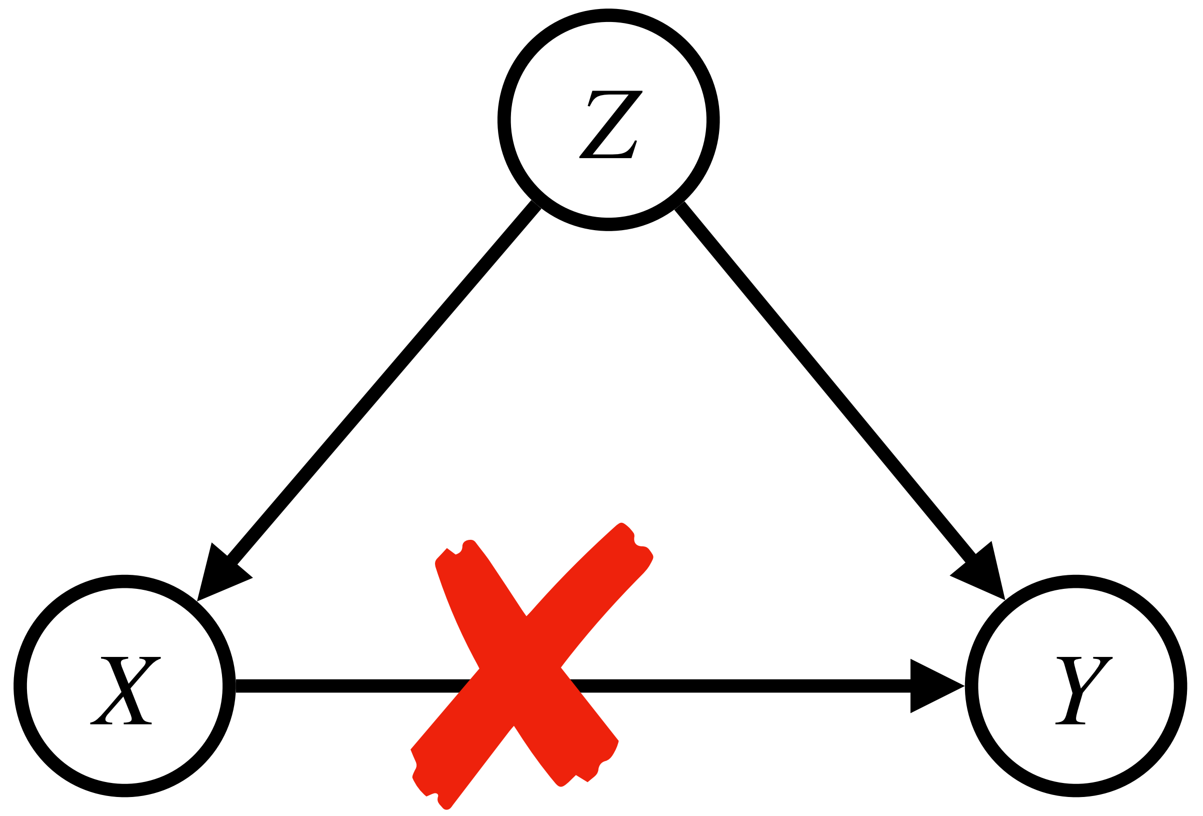


Causal Graphical Model: Potential Outcome

SCM

$$U_Z, U_X, U_Y \sim \text{normal}(0,1)$$
$$Z \leftarrow f_Z(U_Z)$$
$$X \leftarrow f_X(Z, U_X)$$
$$Y(x) \leftarrow f_Y(x, Z, U_Y)$$

Graph



Causal Graphical Model: Potential Outcome

SCM

$U_Z, U_X, U_Y \sim \text{normal}(0,1)$

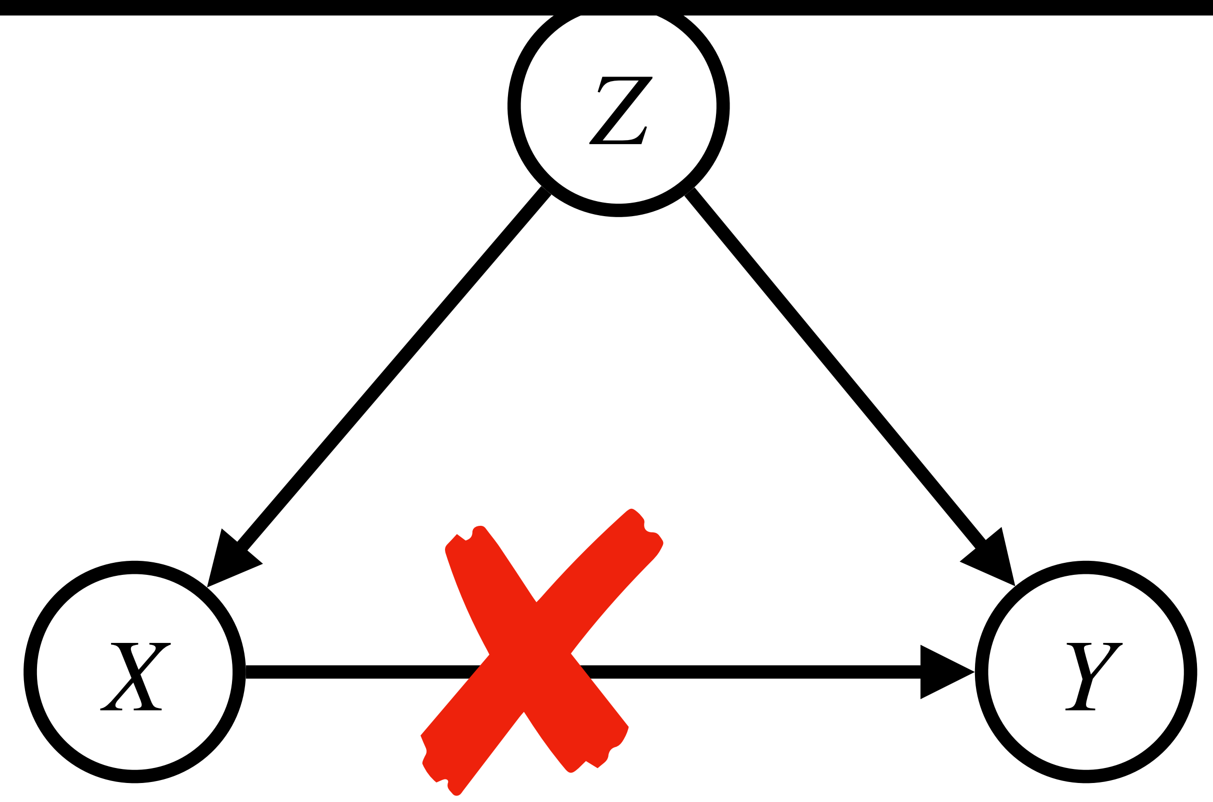
$Z \leftarrow f_Z(U_Z)$

$X \leftarrow f_X(Z, U_X)$

$Y(x) \leftarrow f_Y(x, Z, U_Y)$

Potential outcome $Y(x)$

Y if X had been fixed to x in the DGP

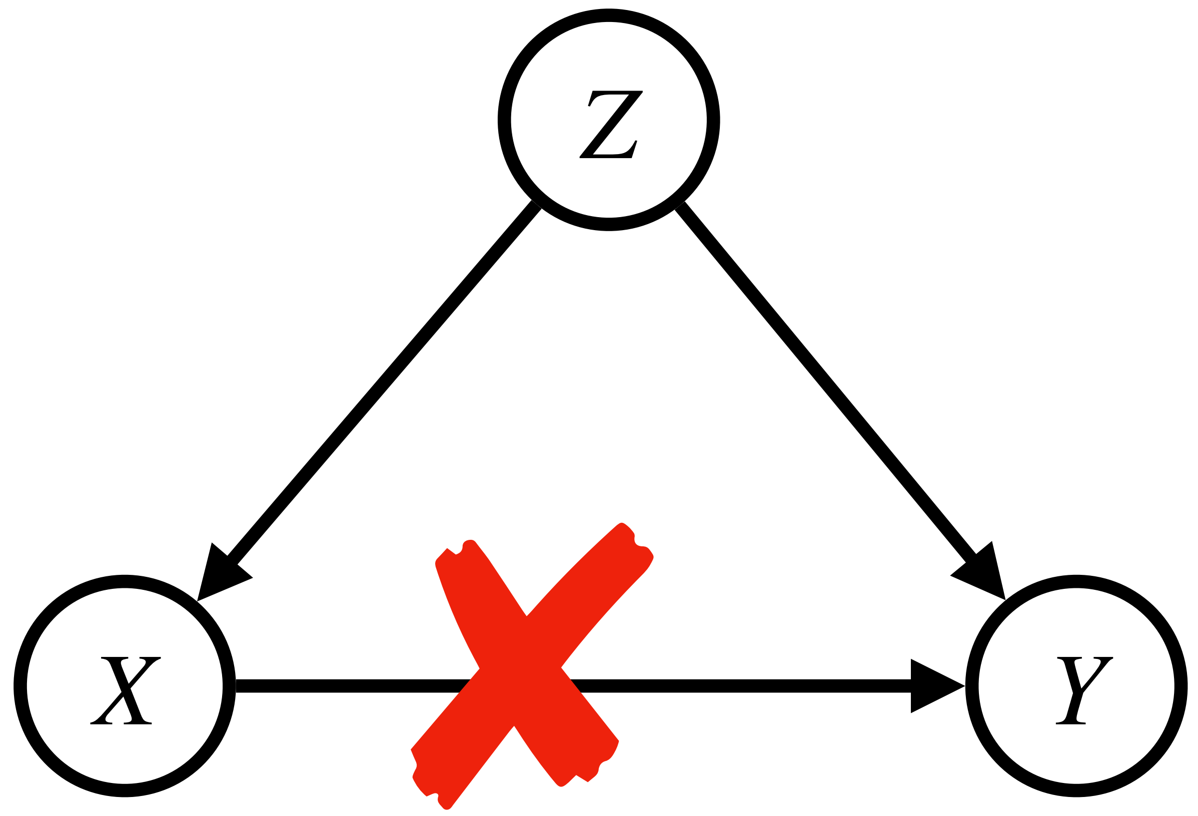


Causal Graphical Model: Potential Outcome

SCM

$$U_Z, U_X, U_Y \sim \text{normal}(0,1)$$
$$Z \leftarrow f_Z(U_Z)$$
$$X \leftarrow f_X(Z, U_X)$$
$$Y(x) \leftarrow f_Y(x, Z, U_Y)$$

Graph



$G_{\underline{X}}$

do-Calculus: Rule 1
conditional independence

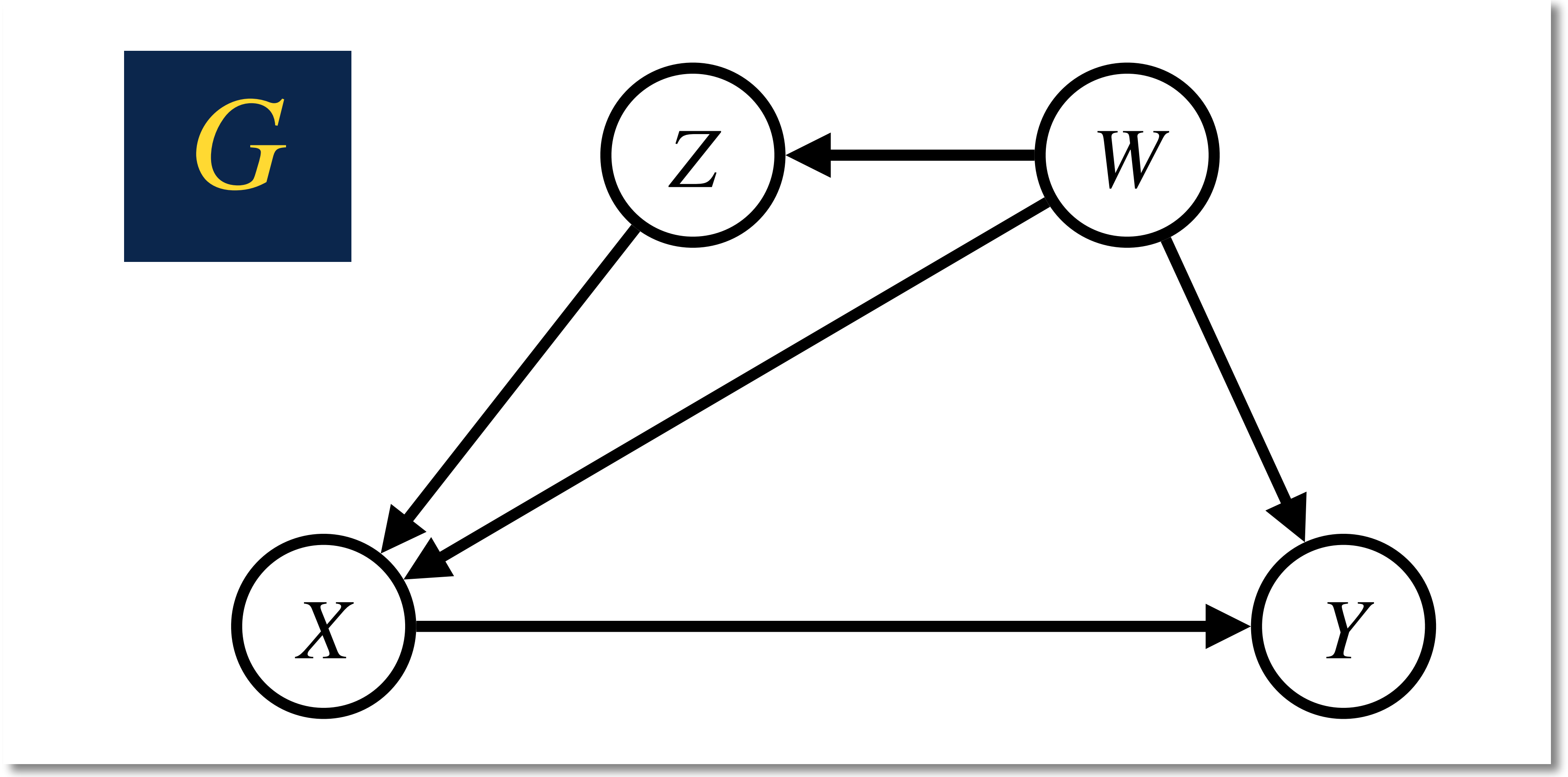
Conditional Independence after Intervention

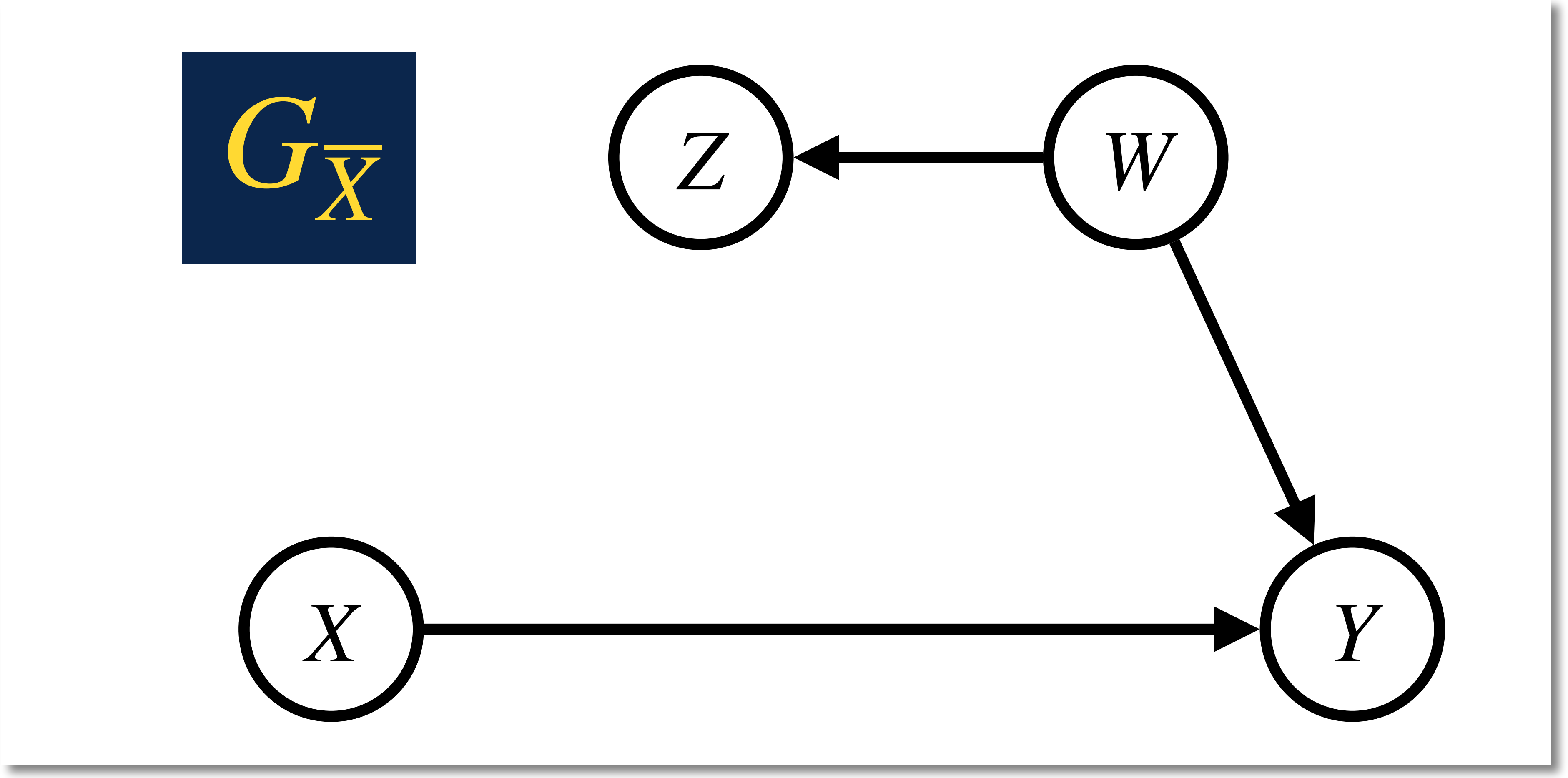
do-Calculus: Rule 1
conditional independence

Conditional Independence after Intervention

$$(Y \perp\!\!\!\perp Z \mid X, W)_{G_{\overline{X}}}$$

Z doesn't have an information regarding Y given W after intervening on X.



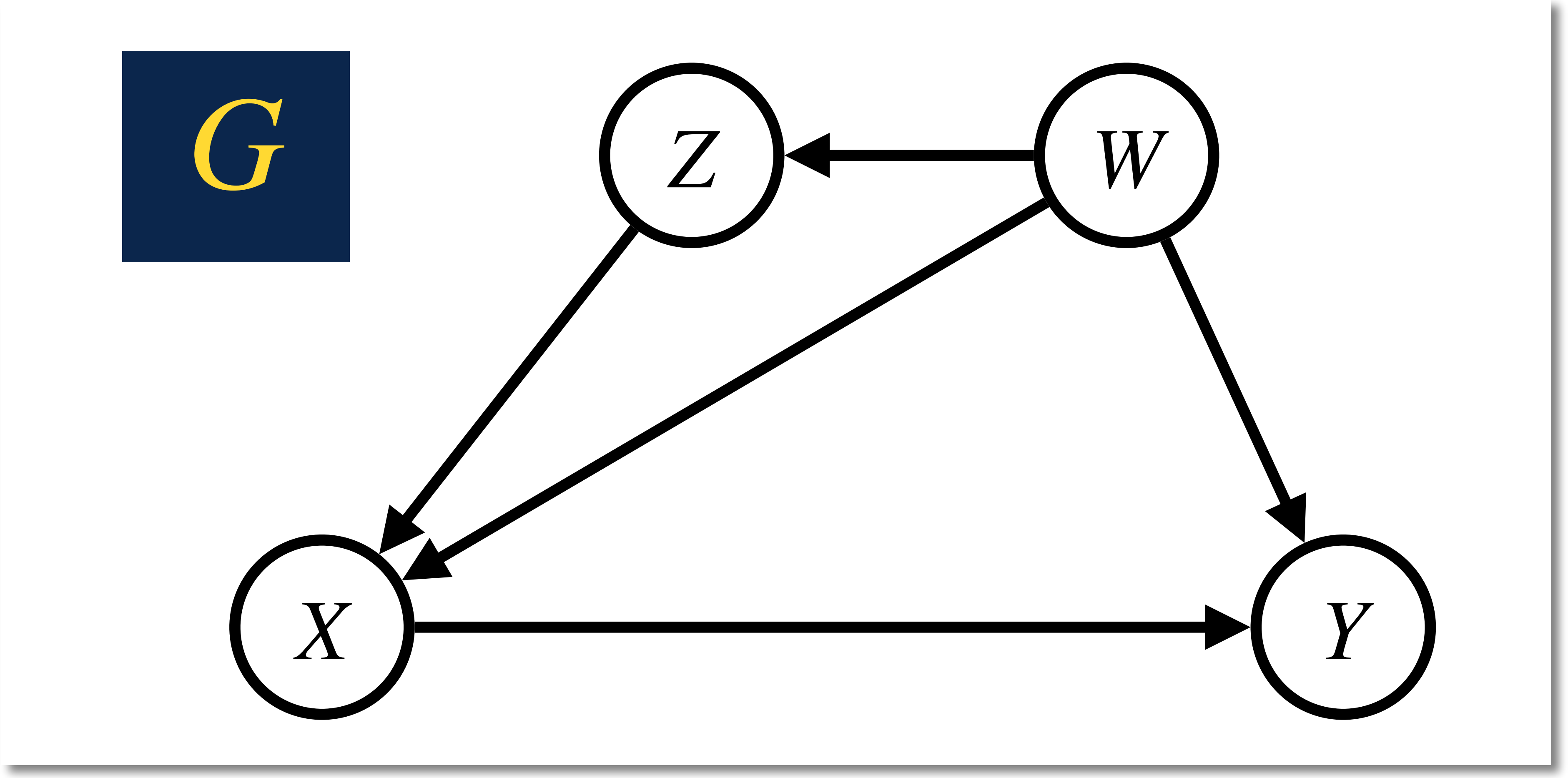


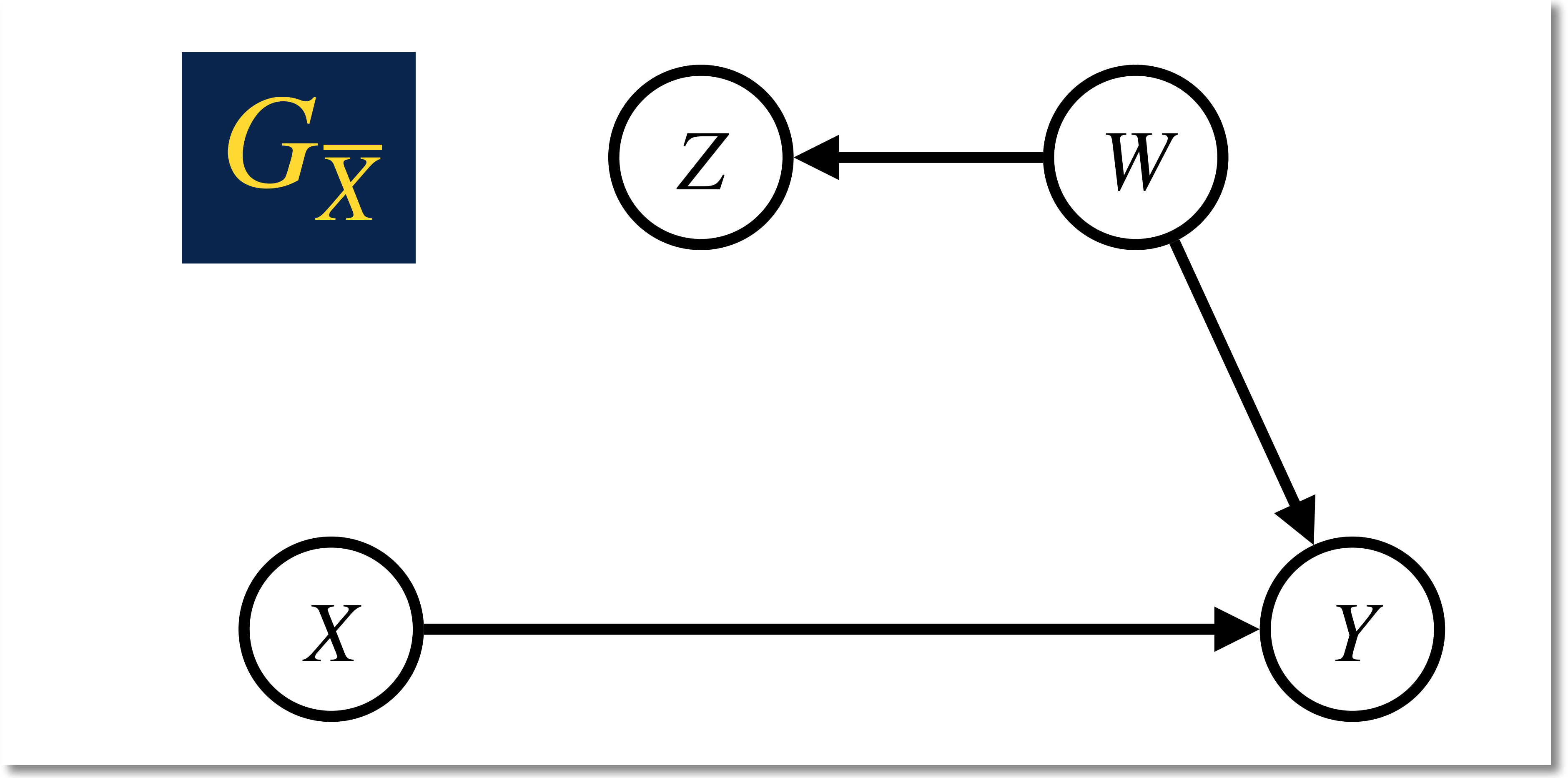
do-Calculus: Rule 1
conditional independence

Conditional Independence after Intervention

$$(Y \perp\!\!\!\perp Z \mid X, W)_{G_{\overline{X}}} \quad \Rightarrow \quad P(Y \mid do(x), z, w) = P(Y \mid do(x), w)$$

Z doesn't have an information regarding Y given W after intervening on X.





do-Calculus: Rule 2
No unmeasured confounders

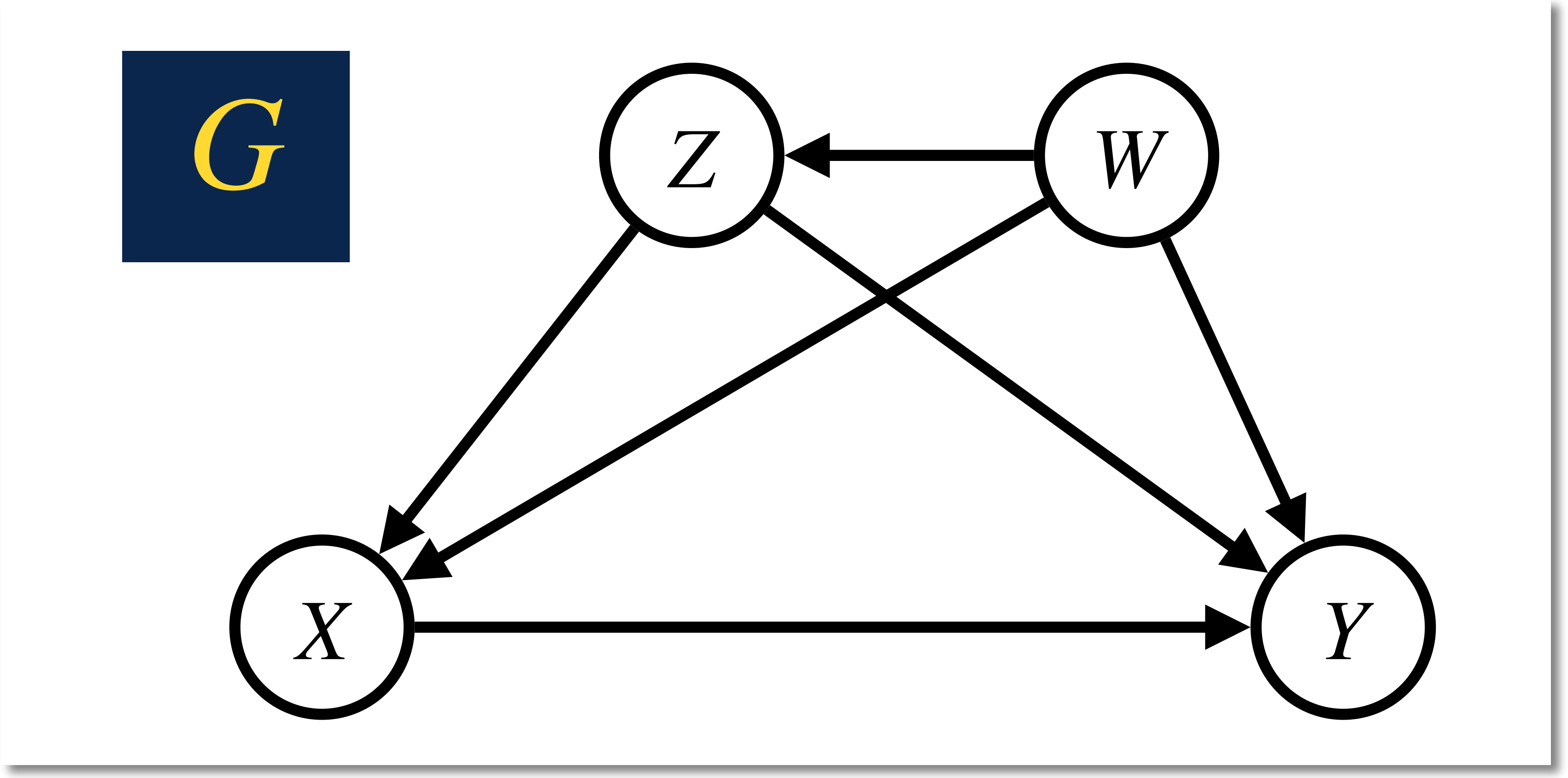
No unmeasured confounders

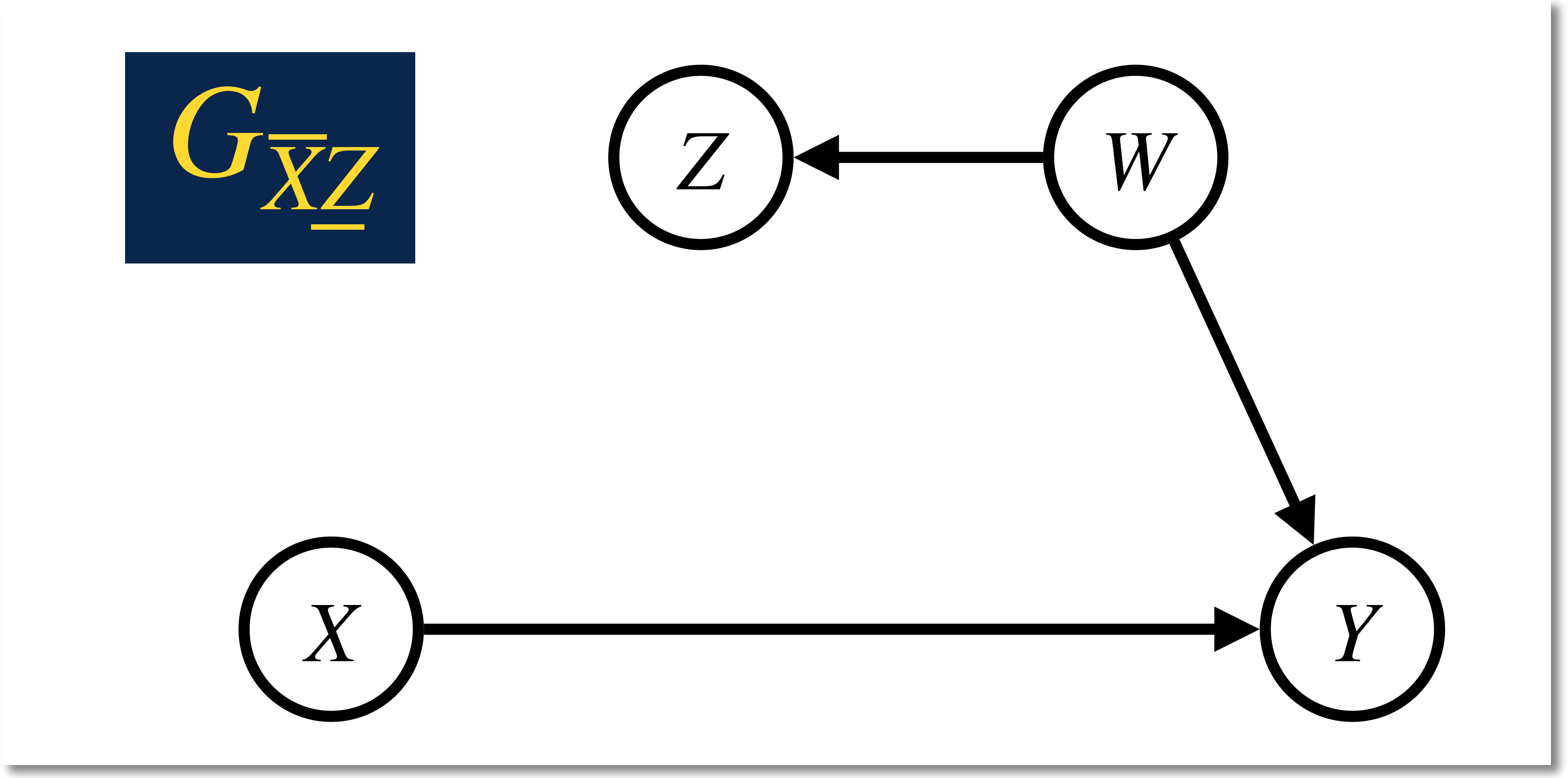
do-Calculus: Rule 2
No unmeasured confounders

No unmeasured confounders

$$(Y \perp\!\!\!\perp Z \mid X, W)_{G_{\overline{X}\underline{Z}}}$$

There are no unmeasured confounders between Y and Z conditioning on W.



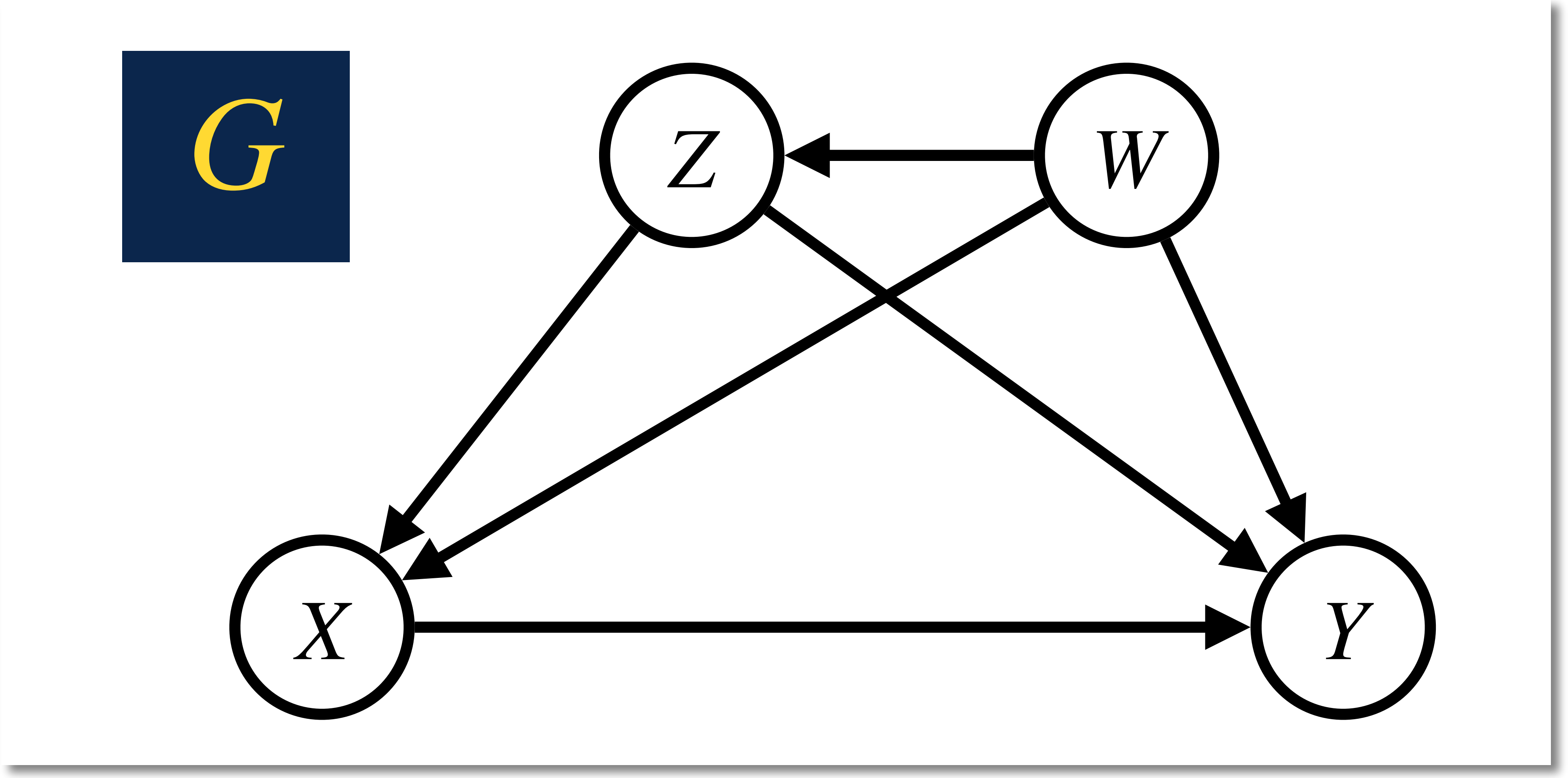


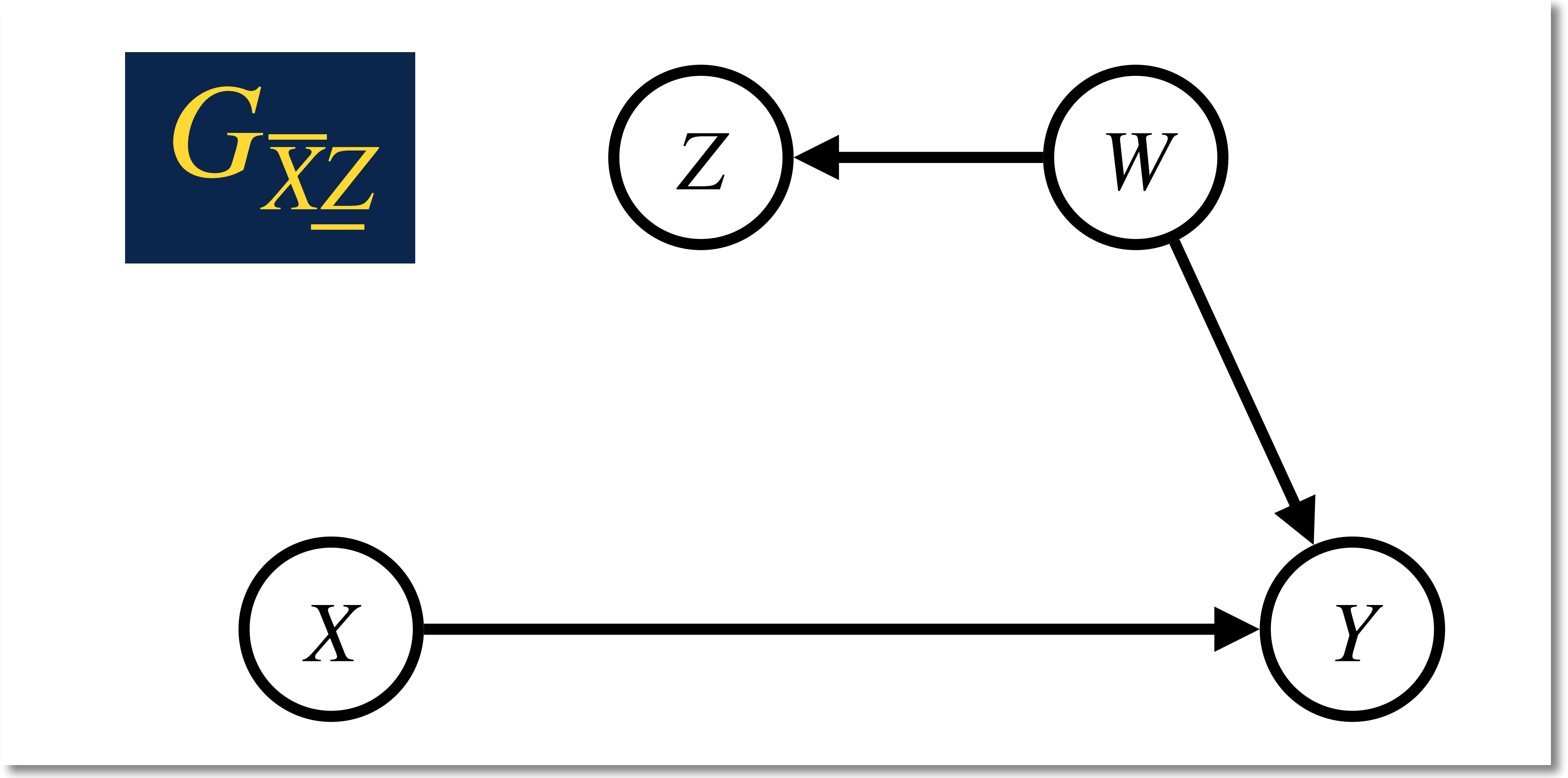
do-Calculus: Rule 2
No unmeasured confounders

No unmeasured confounders

$$(Y \perp\!\!\!\perp Z \mid X, W)_{G_{\overline{X}\underline{Z}}} \quad \Rightarrow \quad P(Y \mid do(x), do(z), w) = P(Y \mid do(x), z, w)$$

There are no unmeasured confounders between Y and Z conditioning on W.





do-Calculus: Rule 3
Intervention independence

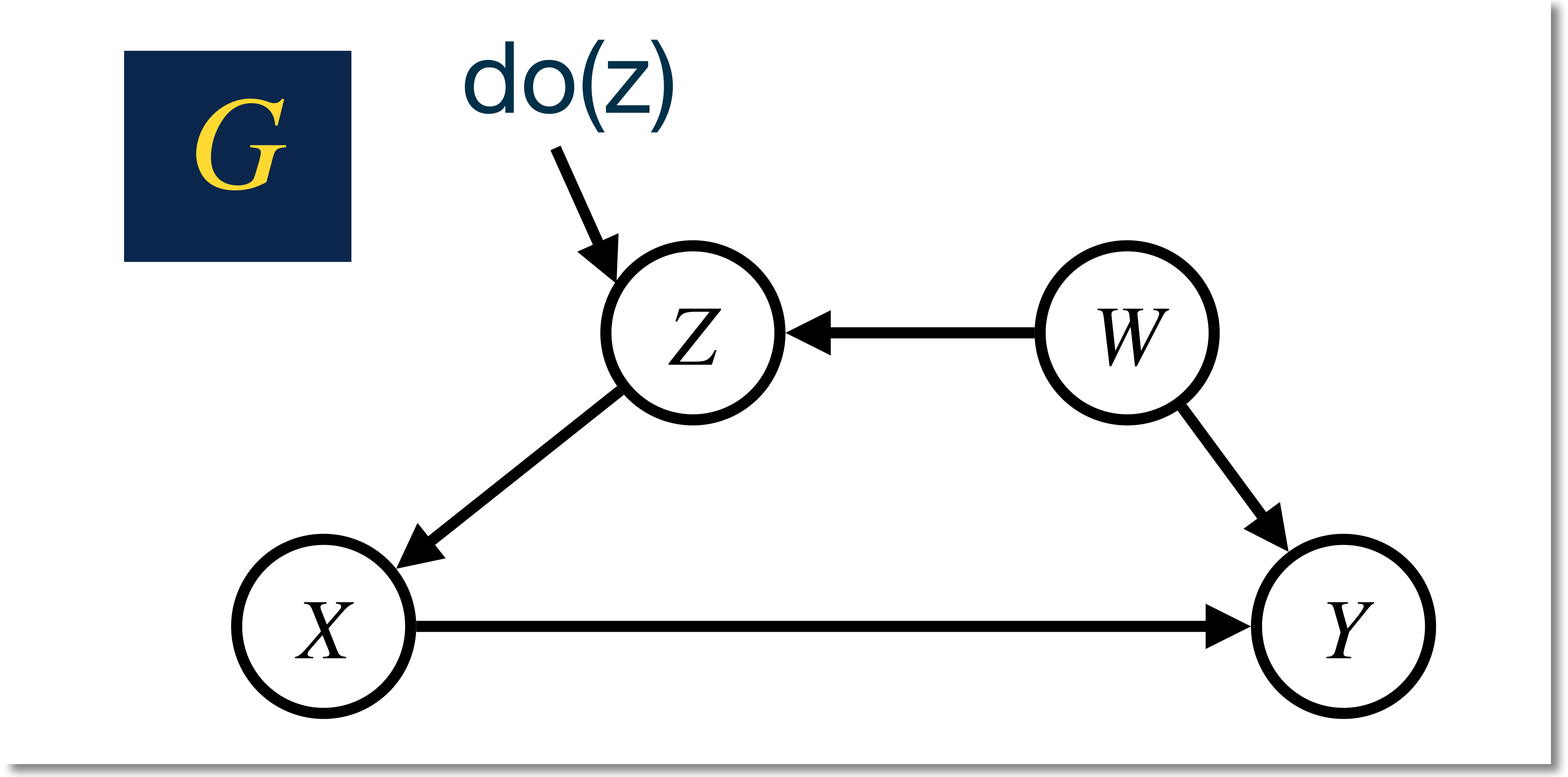
Intervention independence

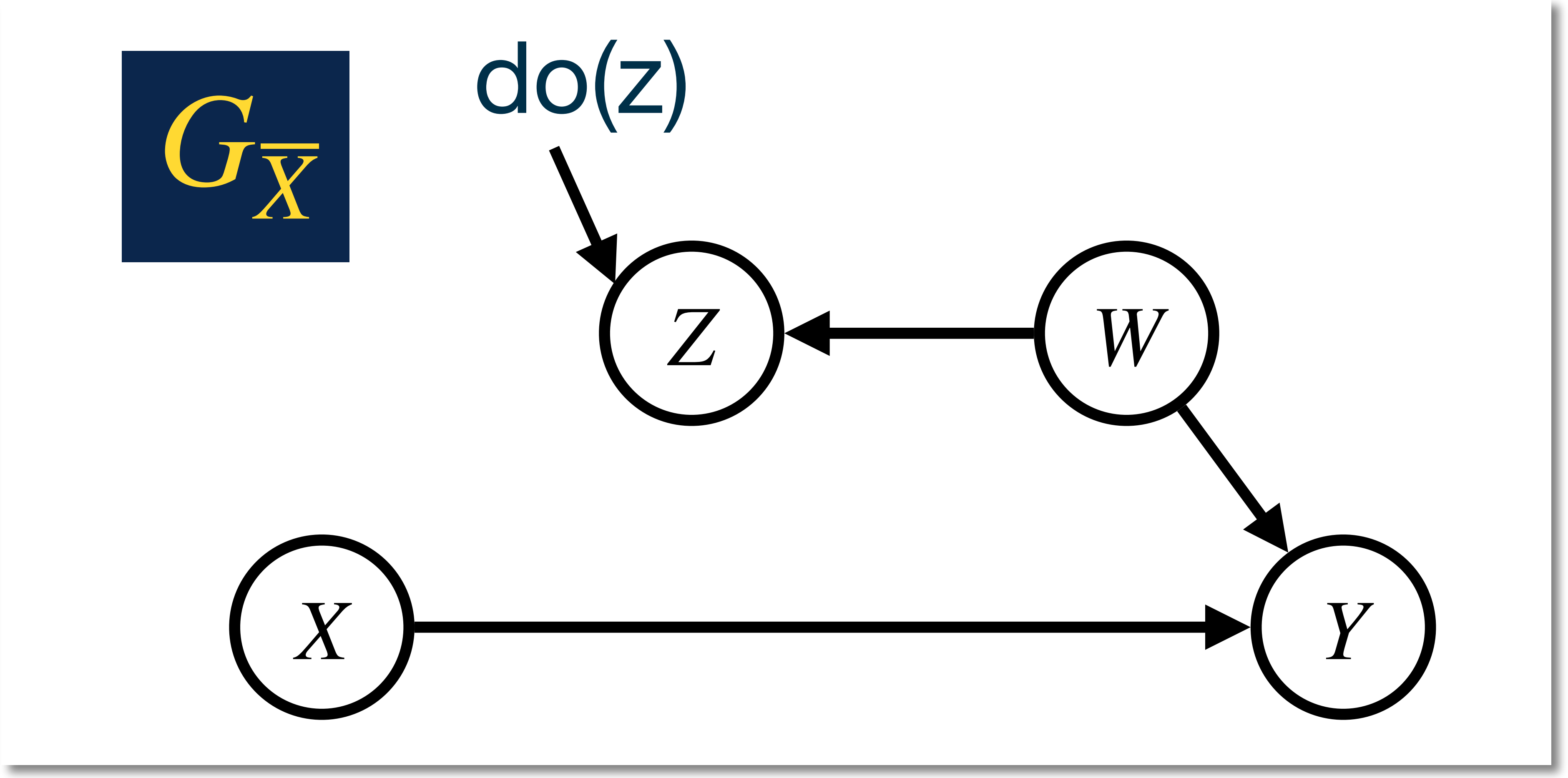
do-Calculus: Rule 3
Intervention independence

Intervention independence

$$(Y \perp\!\!\!\perp \mathrm{do}(z) \mid X, W)_{G_{\overline{X}}}$$

After intervening on X, intervening on Z has no information regarding Y given W



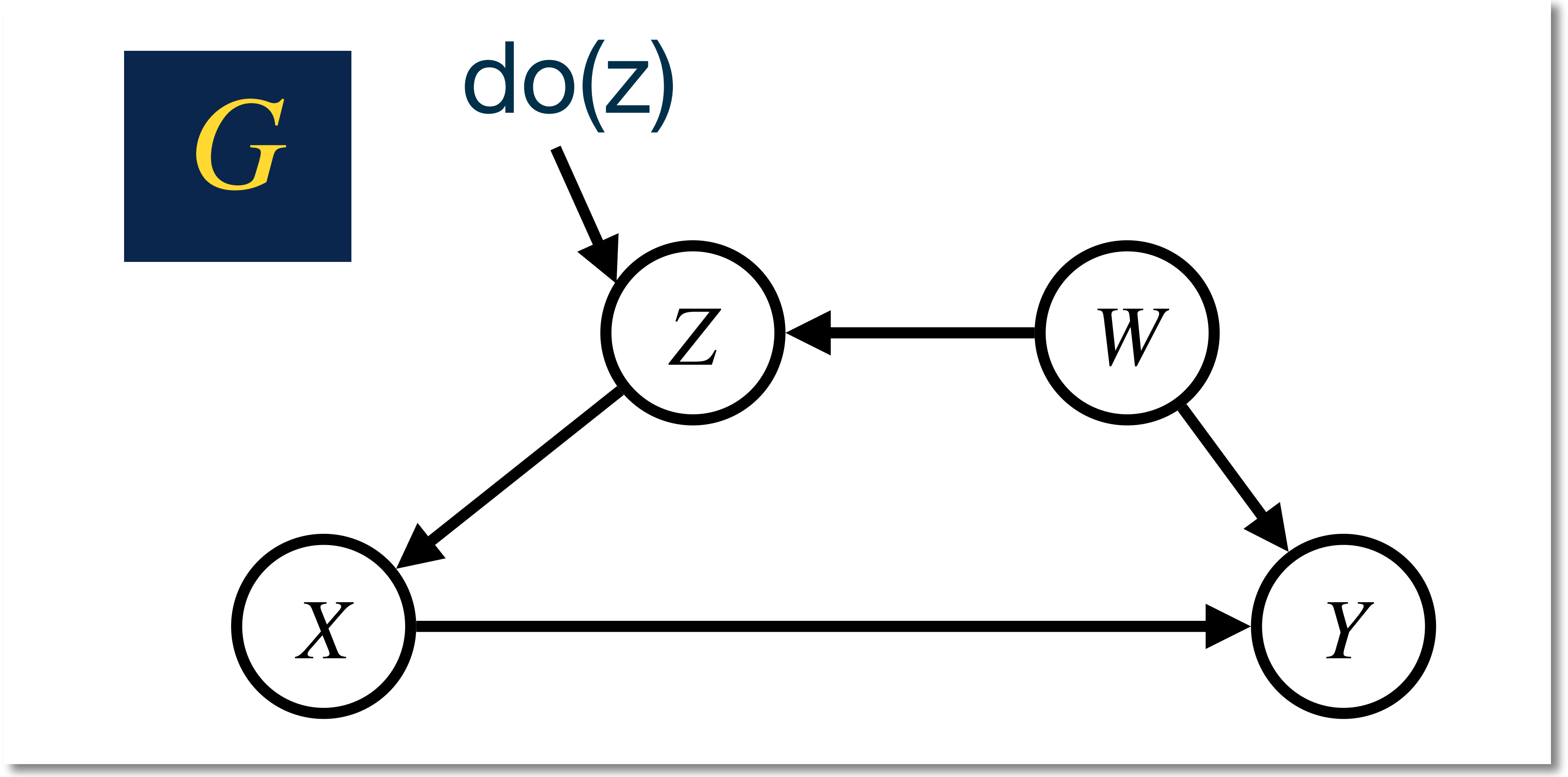


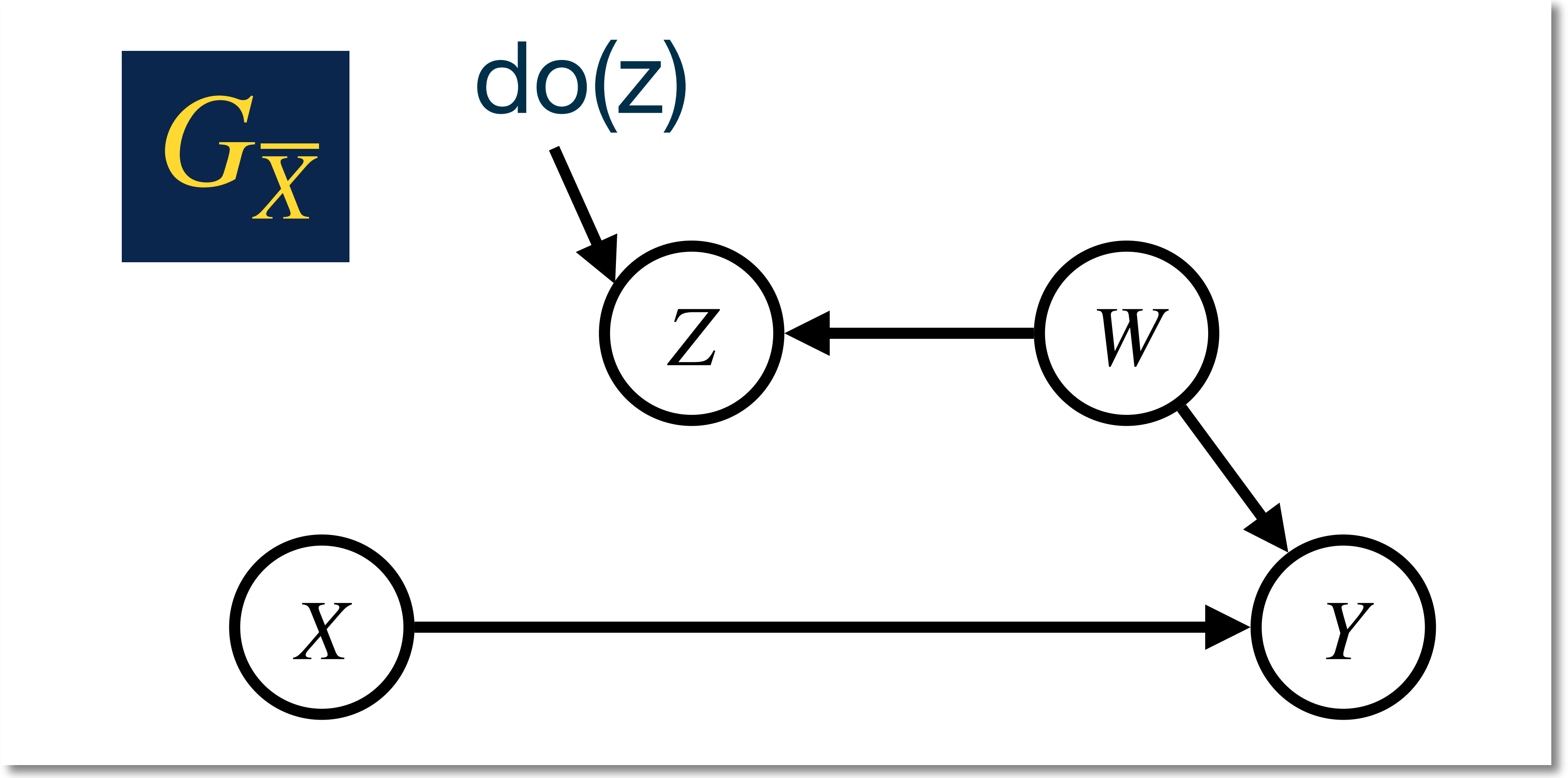
do-Calculus: Rule 3
Intervention independence

Intervention independence

$$(Y \perp\!\!\!\perp \text{do}(z) \,|\, X, W)_{G_{\overline{X}}} \quad \Rightarrow \quad P(Y \,|\, do(x), do(z), w) = P(Y \,|\, do(x), w)$$

After intervening on X, intervening on Z has no information regarding Y given W





do-Calculus in Identification

Pearl's do-Calculus: Graphical criterion for conditional independences of POs in Graphs

1. $(Y \perp\!\!\!\perp Z \,|\, X, W)_{G_{\overline{X}}} \Rightarrow P(Y \,|\, do(x), z, w) = P(Y \,|\, do(x), w)$

2. $(Y \perp\!\!\!\perp X \,|\, X, W)_{G_{\overline{X}\underline{Z}}} \Rightarrow P(Y \,|\, do(x), do(z), w) = P(Y \,|\, do(x), z, w)$

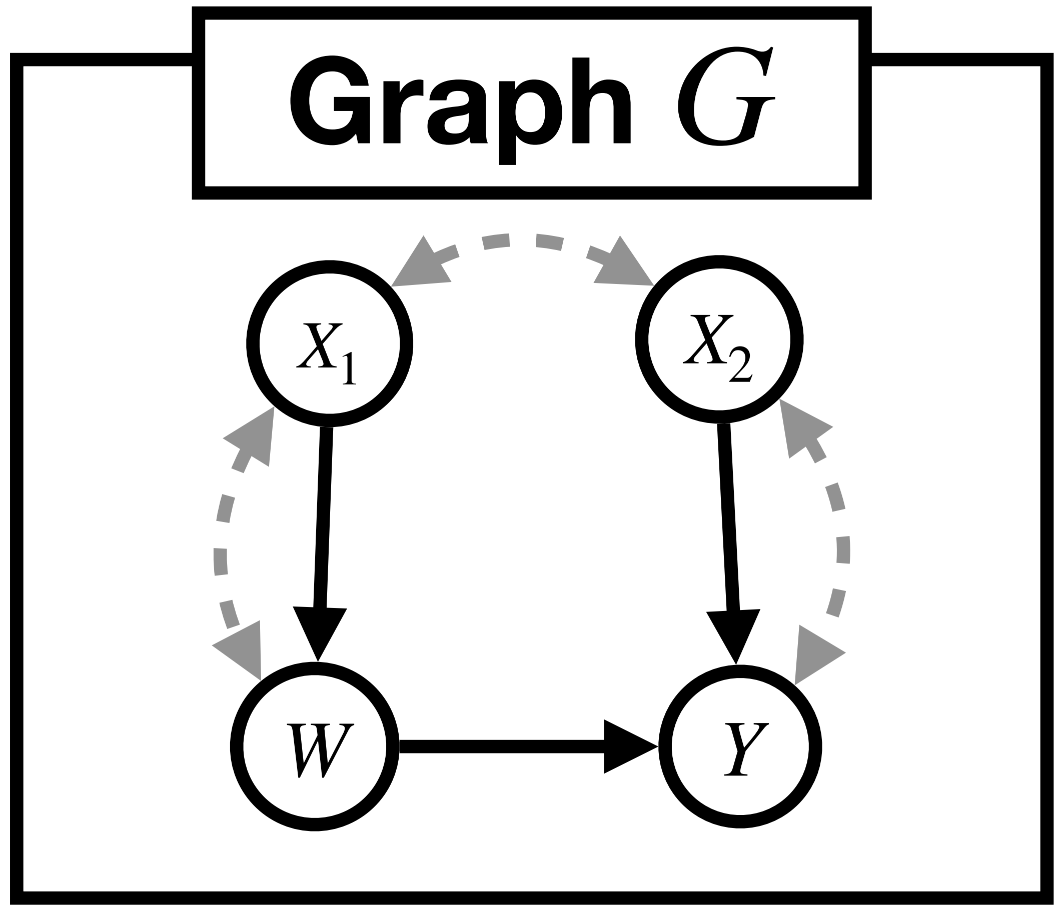
3. $(Y \perp\!\!\!\perp \text{intv}(Z) \,|\, X, W)_{G_{\overline{X}}} \Rightarrow P(Y \,|\, do(x), do(z), w) = P(Y \,|\, do(x), w)$

Outline of this talk

1. Preliminary and Problem Setup
    (1) Structural Causal Model
2. **Treatment-Treatment Interaction**
    (1) Identification
    (2) Estimand
    (3) Estimation and Error Analysis
    (4) Simulation Results
3. Multiple-Treatment Interaction (+ Omitted Results)
4. Future directions & Summary

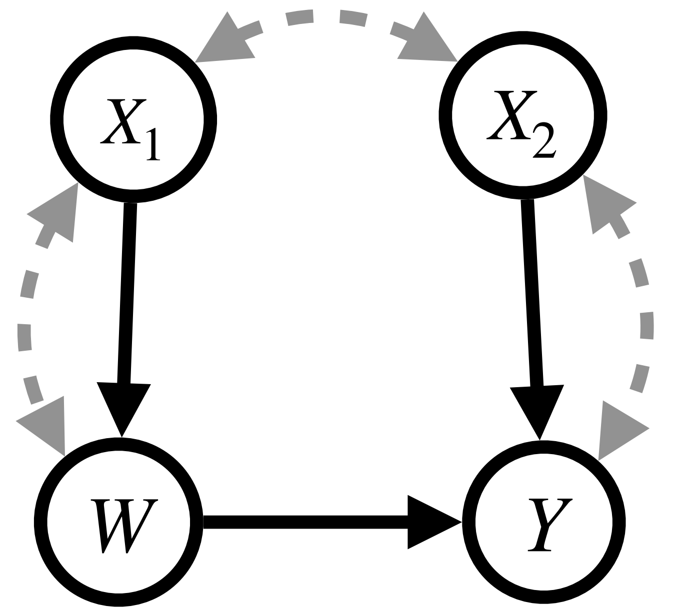
Treatment-Treatment Interaction

Treatment-Treatment Interaction



Treatment-Treatment Interaction

Graph G

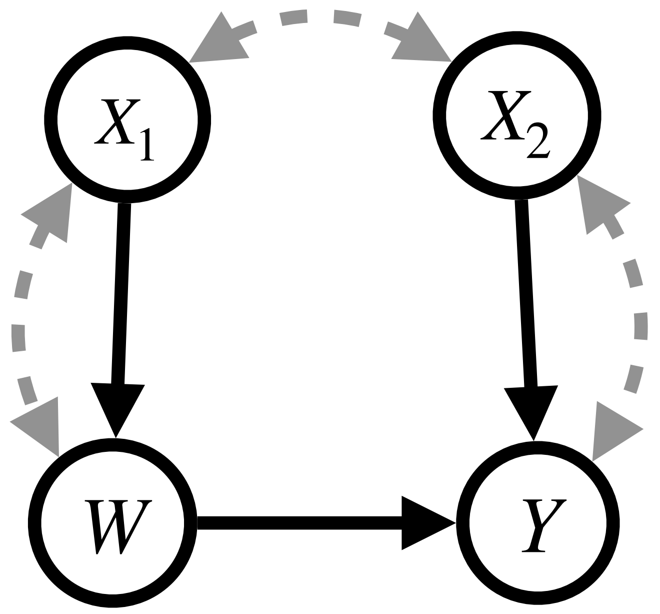


Dist. $\{P\}$

$$\{P\} := \{P(\mathbf{V} \mid do(x_1)), P(\mathbf{V} \mid do(x_2))\}$$

Treatment-Treatment Interaction

Graph G



Dist. $\{P\}$

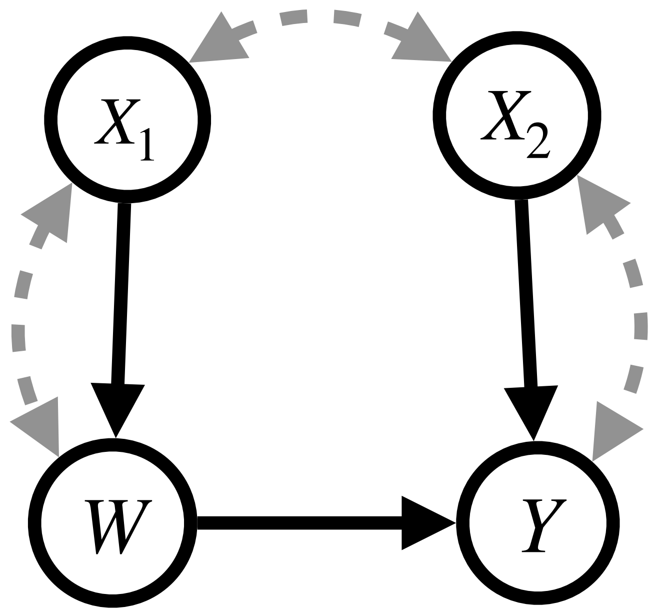
$$\{P\} := \{P(\mathbf{V} \mid do(x_1)), P(\mathbf{V} \mid do(x_2))\}$$

Query $\mathbb{Q}$

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

Treatment-Treatment Interaction

Graph G



Dist. $\{P\}$

$$\{P\} := \{P(\mathbf{V} \mid do(x_1)), P(\mathbf{V} \mid do(x_2))\}$$

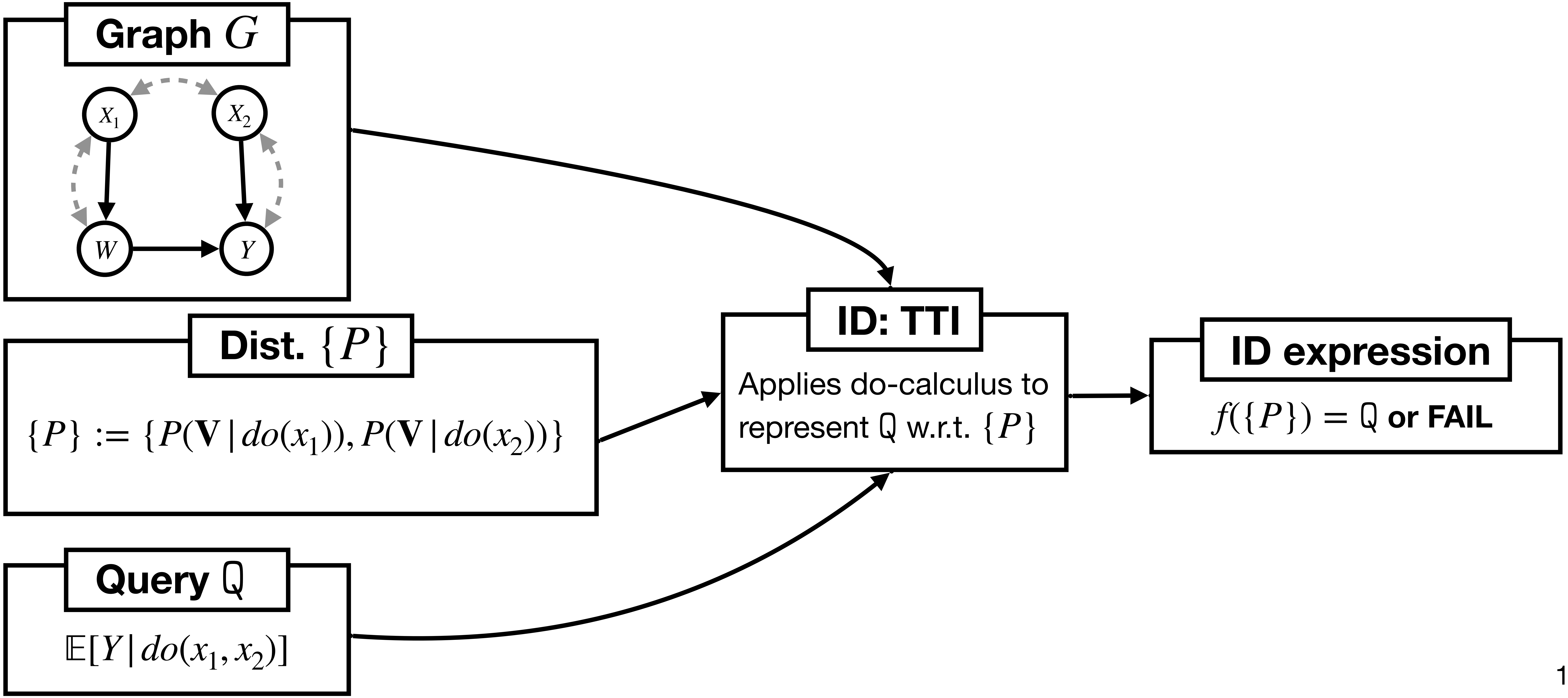
Query $\mathbb{Q}$

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

ID: TTI

Applies do-calculus to represent $\mathbb{Q}$ w.r.t. $\{P\}$

Treatment-Treatment Interaction



Sufficient ID Condition: TTI

Sufficient ID Condition: TTI

Identification Condition for Treatment-Treatment Interaction

Sufficient ID Condition: TTI

Identification Condition for Treatment-Treatment Interaction

There is a set of variables W s.t.

1. W is a pretreatment variable of the second treatment X_2.
2. There are no unmeasured confounders between X_1 and Y (other than W) in the post-treatment distribution $X_2 = x_2$

Sufficient ID Condition: TTI

Identification Condition for Treatment-Treatment Interaction

There is a set of variables W s.t.

1. W is a pretreatment variable of the second treatment X_2.
2. There are no unmeasured confounders between X_1 and Y (other than W) in the post-treatment distribution $X_2 = x_2$

Equivalently,

1. $(W \perp\!\!\!\perp X_2 \mid X_1)_{G_{\overline{X_1, X_2}}}$
2. $(Y \perp\!\!\!\perp X_1 \mid X_2, W)_{G_{\underline{X_1}\overline{X_2}}}$

Sufficient ID Condition: TTI

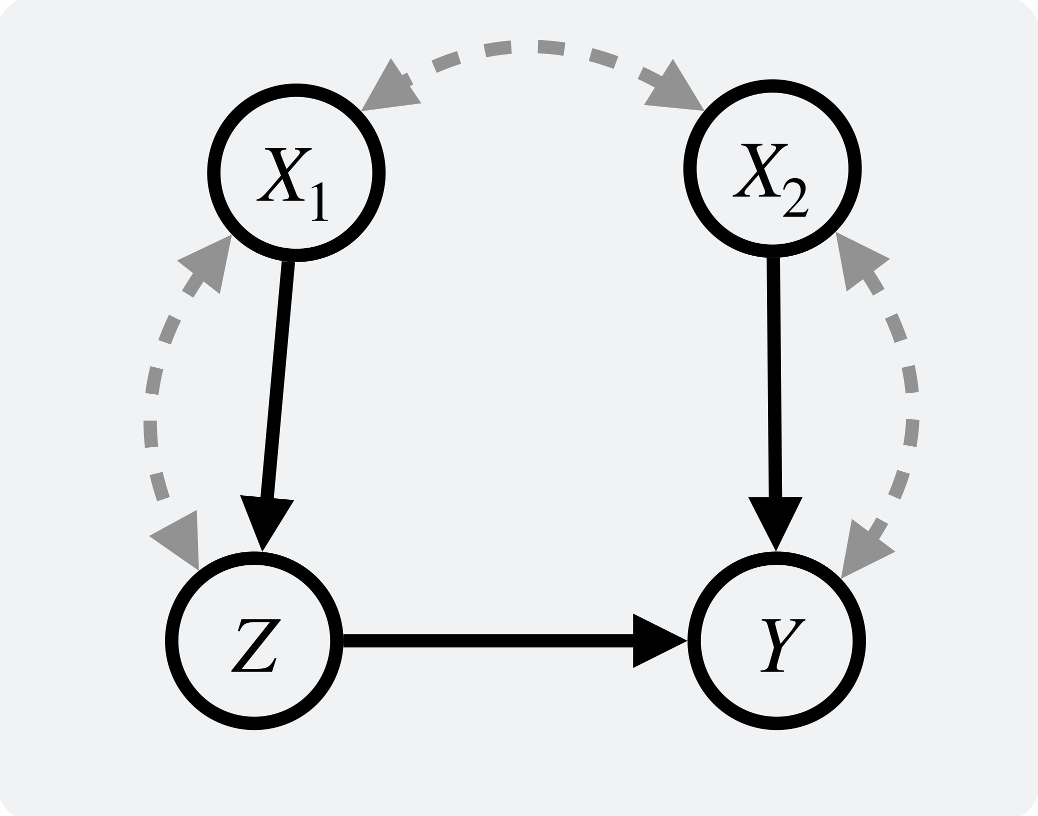
There is a set of variables W s.t.

1. W is a pretreatment variable of the second treatment X_2.
2. There are no unmeasured confounders between X_1 and Y other than W in the post-treatment distribution $X_2 = x_2$

Sufficient ID Condition: TTI

There is a set of variables W s.t.

1. W is a pretreatment variable of the second treatment X_2.
2. There are no unmeasured confounders between X_1 and Y other than W in the post-treatment distribution $X_2 = x_2$

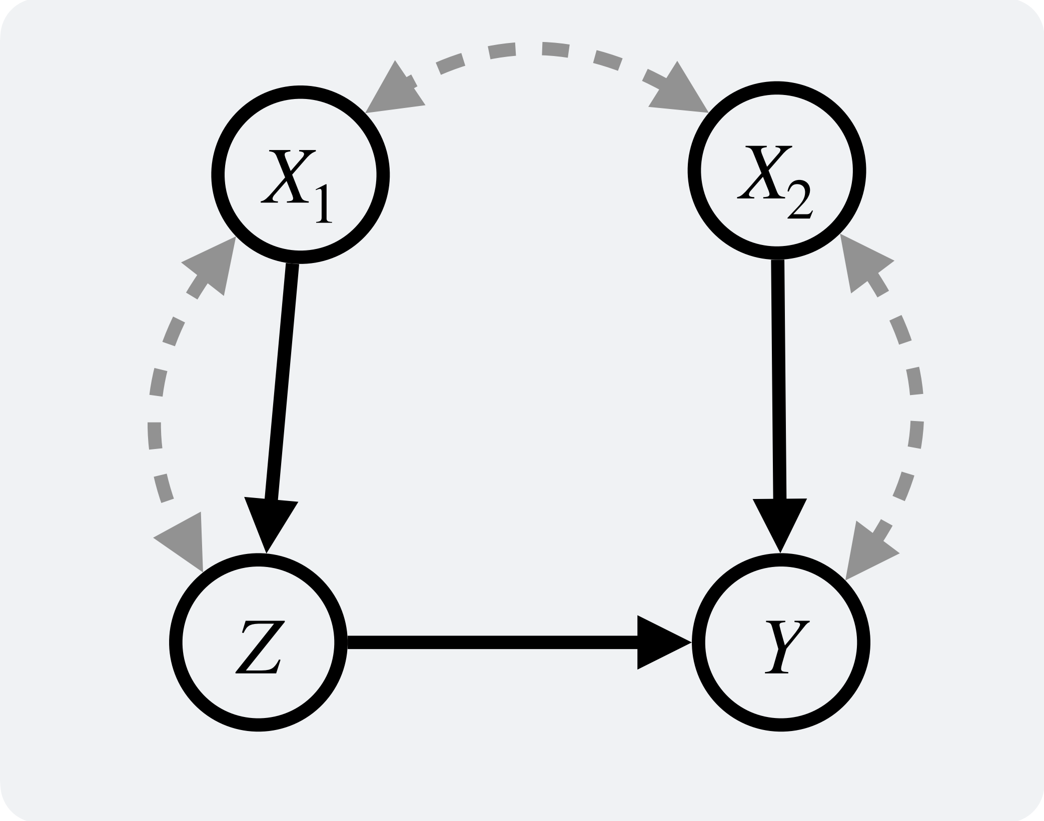


$$W = \{Z\}$$

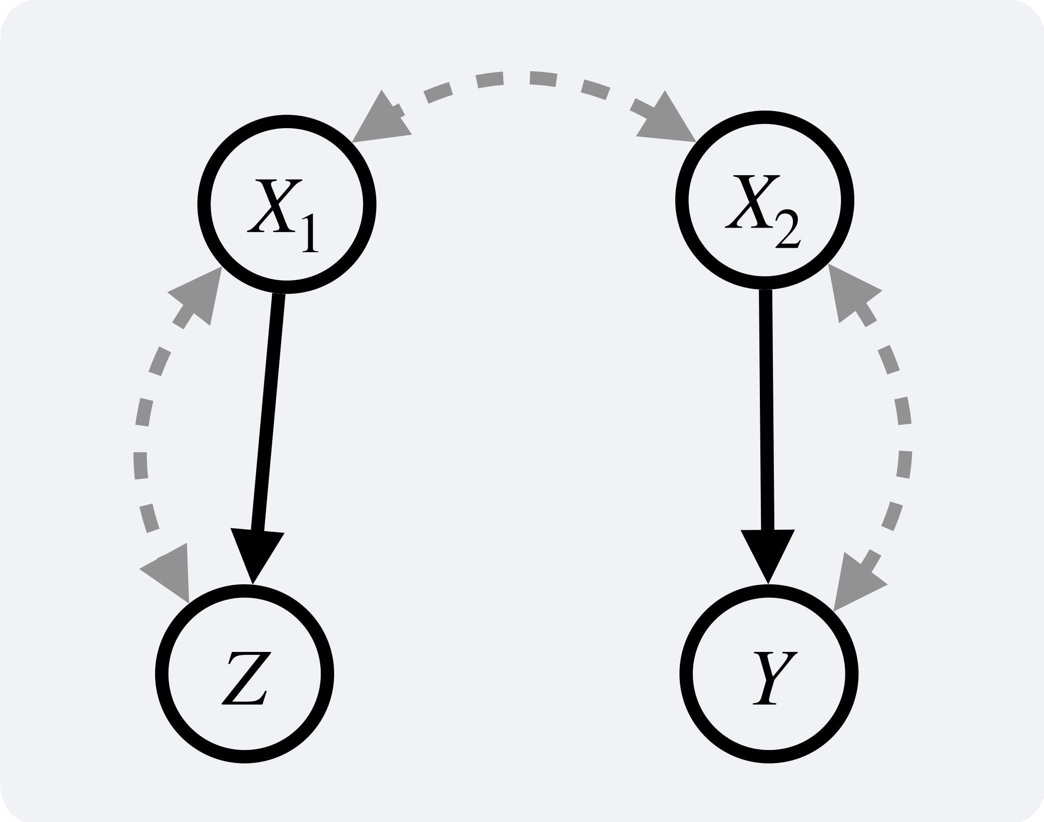
Sufficient ID Condition: TTI

There is a set of variables W s.t.

1. W is a pretreatment variable of the second treatment X_2.
2. There are no unmeasured confounders between X_1 and Y other than W in the post-treatment distribution $X_2 = x_2$



$W = \{Z\}$

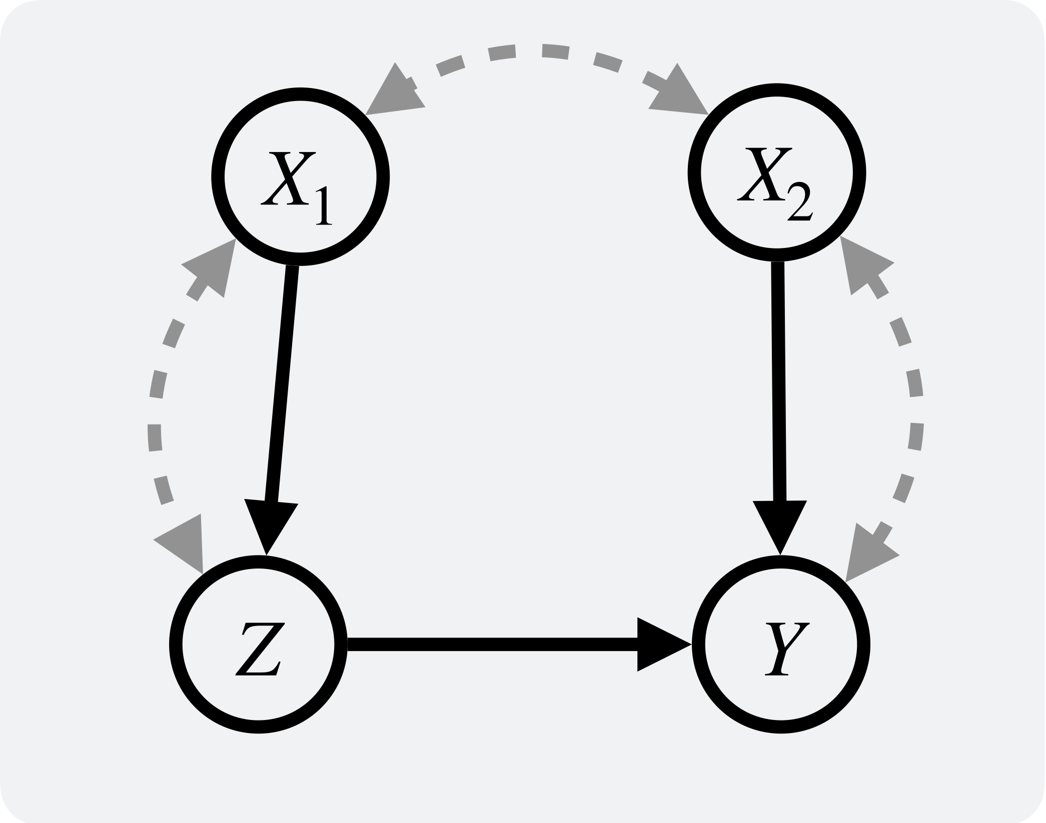


$W = \{Z\}$

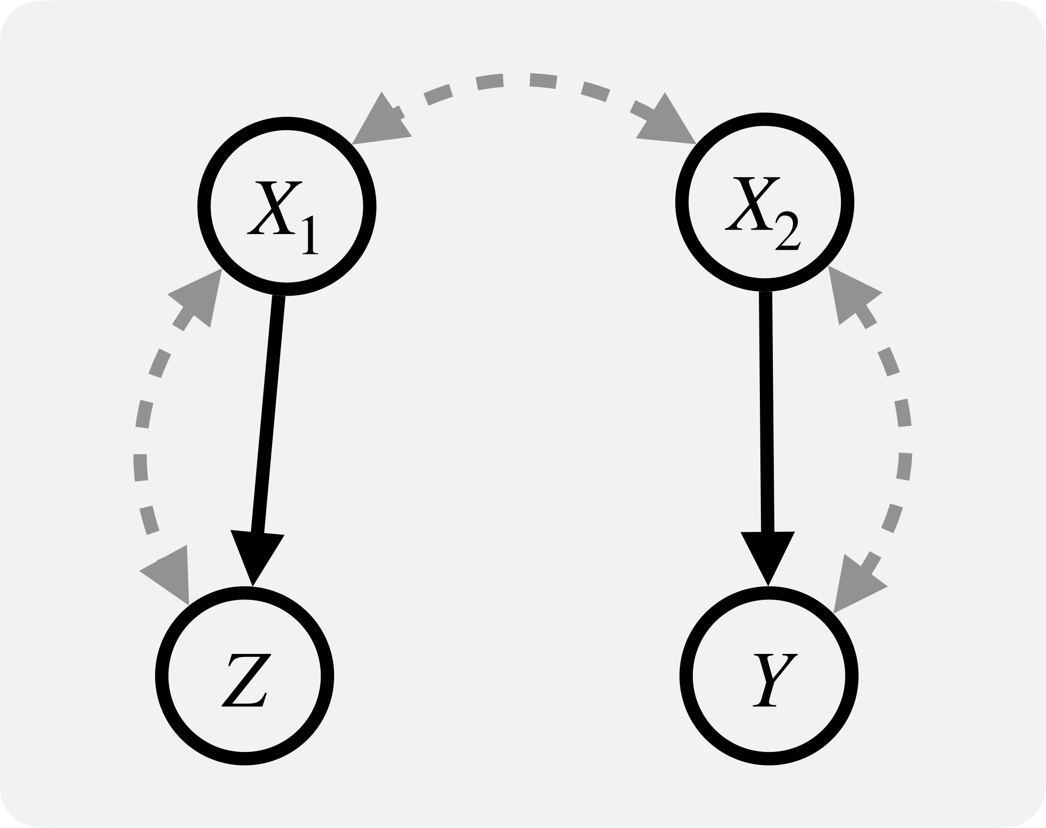
Sufficient ID Condition: TTI

There is a set of variables W s.t.

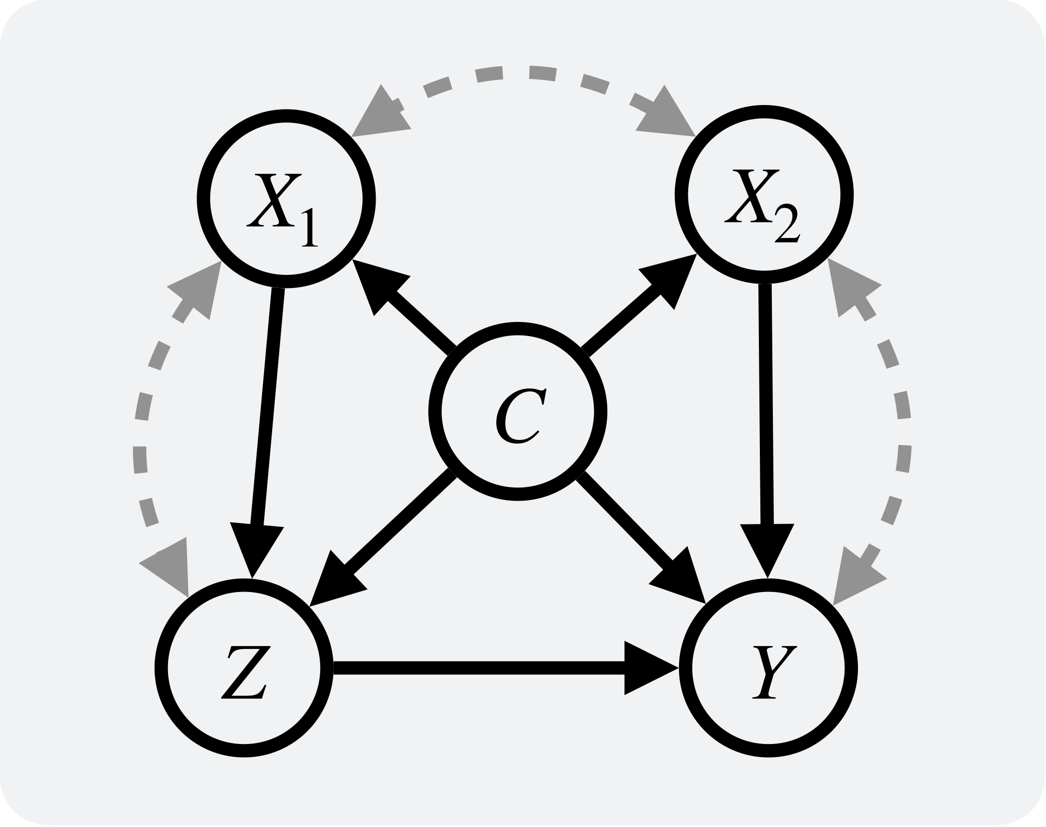
1. W is a pretreatment variable of the second treatment X_2.
2. There are no unmeasured confounders between X_1 and Y other than W in the post-treatment distribution $X_2 = x_2$



$W = \{Z\}$



$W = \{Z\}$



$W = \{Z, C\}$

Treatment-Treatment Interaction

Identification Condition for Treatment-Treatment Interaction

There is a set of variables W s.t.

1. W is a pretreatment variable of the second treatment X_2.
2. There are no unmeasured confounders between X_1 and Y in the post-treatment distribution $X_2 = x_2$

Treatment-Treatment Interaction

Identification Condition for Treatment-Treatment Interaction

There is a set of variables W s.t.

1. W is a pretreatment variable of the second treatment X_2.
2. There are no unmeasured confounders between X_1 and Y in the post-treatment distribution $X_2 = x_2$

Then, $\mathbb{E}[Y \mid do(x_1, x_2)]$ is identifiable from two experimental distributions $P(V \mid do(x_1))$ and $P(V \mid do(x_2))$ as follows:

$$\mathbb{E}[Y \mid do(x_1, x_2)] = \sum_{w \in \mathcal{W}} \mathbb{E}[Y \mid do(x_2), w, x_1] P(w \mid do(x_1)).$$

Treatment-Treatment Interaction

Identification Condition for Treatment-Treatment Interaction

There is a set of variables W s.t.

1. W is a pretreatment variable of the second treatment X_2.
2. There are no unmeasured confounders between X_1 and Y in the post-treatment distribution $X_2 = x_2$

Then, $\mathbb{E}[Y \mid do(x_1, x_2)]$ is identifiable from two experimental distributions $P(V \mid do(x_1))$ and $P(V \mid do(x_2))$ as follows:

$$\mathbb{E}[Y \mid do(x_1, x_2)] = \sum_{w \in \mathcal{W}} \underbrace{\mathbb{E}[Y \mid do(x_2), w, x_1]}_{\text{estimable from } P(V \mid do(x_2))} P(w \mid do(x_1)).$$

Treatment-Treatment Interaction

Identification Condition for Treatment-Treatment Interaction

There is a set of variables W s.t.

1. W is a pretreatment variable of the second treatment X_2.
2. There are no unmeasured confounders between X_1 and Y in the post-treatment distribution $X_2 = x_2$

Then, $\mathbb{E}[Y \mid do(x_1, x_2)]$ is identifiable from two experimental distributions $P(V \mid do(x_1))$ and $P(V \mid do(x_2))$ as follows:

$$\mathbb{E}[Y \mid do(x_1, x_2)] = \sum_{w \in \mathcal{W}} \mathbb{E}[Y \mid do(x_2), w, x_1] \underbrace{P(w \mid do(x_1))}_{\text{estimable from } P(V \mid do(x_1))}.$$

Treatment-Treatment Interaction: Intuition

Identification Condition for Treatment-Treatment Interaction

1. $(W \perp\!\!\!\perp X_2 \mid X_1)_{G_{\overline{X_1, X_2}}}$
2. $(Y \perp\!\!\!\perp X_1 \mid X_2, W)_{G_{\underline{X_1}\overline{X_2}}}$

Treatment-Treatment Interaction: Intuition

Identification Condition for Treatment-Treatment Interaction

1. $(W \perp\!\!\!\perp X_2 \mid X_1)_{G_{\overline{X_1, X_2}}}$
2. $(Y \perp\!\!\!\perp X_1 \mid X_2, W)_{G_{\underline{X_1}\overline{X_2}}}$

$$P(y \mid do(x_1, x_2)) = \sum_w P(y \mid do(x_1, x_2), w) P(w \mid do(x_1, x_2))$$

Treatment-Treatment Interaction: Intuition

Identification Condition for Treatment-Treatment Interaction

1. $(W \perp\!\!\!\perp X_2 \mid X_1)_{G_{\overline{X_1, X_2}}}$
2. $(Y \perp\!\!\!\perp X_1 \mid X_2, W)_{G_{\underline{X_1}\overline{X_2}}}$

$$P(y \mid do(x_1, x_2)) = \sum_w P(y \mid do(x_1, x_2), w) P(w \mid do(x_1, x_2))$$

$$= \sum_w P(y \mid do(x_1, x_2), w) P(w \mid do(x_1))$$

do-calc. 3 (ID condition 1)

Treatment-Treatment Interaction: Intuition

Identification Condition for Treatment-Treatment Interaction

1. $(W \perp\!\!\!\perp X_2 \mid X_1)_{G_{\overline{X_1, X_2}}}$
2. $(Y \perp\!\!\!\perp X_1 \mid X_2, W)_{G_{\underline{X_1}\overline{X_2}}}$

$$P(y \mid do(x_1, x_2)) = \sum_w P(y \mid do(x_1, x_2), w) P(w \mid do(x_1, x_2))$$

$$= \sum_w P(y \mid do(x_1, x_2), w) P(w \mid do(x_1))$$

do-calc. 3 (ID condition 1)

$$= \sum_w P(y \mid do(x_2), x_1, w) P(w \mid do(x_1))$$

do-calc. 2 (ID condition 2)

Outline of this talk

1. Preliminary and Problem Setup
   (1) Structural Causal Model
2. **Treatment-Treatment Interaction**
   (1) Identification
   (2) Estimand
   (3) Estimation and Error Analysis
   (4) Simulation Results
3. Multiple-Treatment Interaction (+ Omitted Results)
4. Future directions & Summary

Joint Treatment Effect Estimation : Regression

ID Expression of the Joint Treatment Effect:

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

Joint Treatment Effect Estimation : Regression

ID Expression of the Joint Treatment Effect:

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

Regression-based Representation

Joint Treatment Effect Estimation : Regression

ID Expression of the Joint Treatment Effect:

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

Regression-based Representation

$$\mu_0(W, X_1) := \mathbb{E}[Y \mid do(x_2), x_1, W]$$

$$\mathbb{E}[Y \mid do(x_1, x_2)] = \mathbb{E}[\mu_0(W, x_1) \mid do(x_1)]$$

Joint Treatment Effect Estimation : Regression

ID Expression of the Joint Treatment Effect:

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

Regression-based Representation

$$\mu_0(W, X_1) := \mathbb{E}[Y \mid do(x_2), x_1, W]$$

$$\mathbb{E}[Y \mid do(x_1, x_2)] = \mathbb{E}[\boxed{\mu_0(W, x_1)} \mid do(x_1)]$$

Estimable from $D_2 \sim P(\mathbf{V} \mid do(x_2))$

Joint Treatment Effect Estimation : Regression

ID Expression of the Joint Treatment Effect:

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

Regression-based Representation

$$\mu_0(W, X_1) := \mathbb{E}[Y \mid do(x_2), x_1, W]$$

$$\mathbb{E}[Y \mid do(x_1, x_2)] = \mathbb{E}[\mu_0(W, x_1) \mid do(x_1)]$$

Estimable from $D_1 \sim P(\mathbf{V} \mid do(x_1))$

Joint Treatment Effect Estimation : Regression

$$\mathbb{E}[Y \mid do(x_1, x_2)] = \mathbb{E}[\mu_0(W, x_1) \mid do(x_1)]$$

Joint Treatment Effect Estimation : Regression

$$\mathbb{E}[Y \mid do(x_1, x_2)] = \mathbb{E}[\mu_0(W, x_1) \mid do(x_1)]$$

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

Joint Treatment Effect Estimation : Regression

$$\mathbb{E}[Y \mid do(x_1, x_2)] = \mathbb{E}[\mu_0(W, x_1) \mid do(x_1)]$$

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

$$= \sum_w \mu_0(w, x_1) P(w \mid do(x_1))$$

Joint Treatment Effect Estimation : Regression

$$\mathbb{E}[Y \mid do(x_1, x_2)] = \mathbb{E}[\mu_0(W, x_1) \mid do(x_1)]$$

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

$$= \sum_w \mu_0(w, x_1) P(w \mid do(x_1))$$

$$= \mathbb{E}[\mu_0(W, x_1) \mid do(x_1)]$$

Joint Treatment Effect Estimation : Probability Weighting

ID Expression of the Joint Treatment Effect:

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

Joint Treatment Effect Estimation : Probability Weighting

ID Expression of the Joint Treatment Effect:

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

Probability Weighting-based Representation

$$\pi_0(W, X_1) := \frac{P(W \mid do(x_1))}{P(W, X_1 \mid do(x_2))}$$

Joint Treatment Effect Estimation : Probability Weighting

ID Expression of the Joint Treatment Effect:

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

Probability Weighting-based Representation

$$\pi_0(W, X_1) := \frac{P(W \mid do(x_1))}{P(W, X_1 \mid do(x_2))}$$

$$\mathbb{E}[Y \mid do(x_1, x_2)] = \mathbb{E}[\pi_0(W, X_1) I_{x_1}(X_1) Y \mid do(x_2)]$$

Joint Treatment Effect Estimation : Probability Weighting

ID Expression of the Joint Treatment Effect:

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

Probability Weighting-based Representation

$$\pi_0(W, X_1) := \frac{P(W \mid do(x_1))}{P(W, X_1 \mid do(x_2))}$$

$$\mathbb{E}[Y \mid do(x_1, x_2)] = \mathbb{E}[\pi_0(W, X_1) I_{x_1}(X_1) Y \mid do(x_2)]$$

Estimable from $D_1 \sim P(\mathbf{V} \mid do(x_1))$

Joint Treatment Effect Estimation : Probability Weighting

ID Expression of the Joint Treatment Effect:

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

Probability Weighting-based Representation

$$\pi_0(W, X_1) := \frac{P(W \mid do(x_1))}{P(W, X_1 \mid do(x_2))}$$

$$\mathbb{E}[Y \mid do(x_1, x_2)] = \mathbb{E}[\pi_0(W, X_1) I_{x_1}(X_1) Y \mid do(x_2)]$$

Estimable from $D_2 \sim P(\mathbf{V} \mid do(x_2))$

Joint Treatment Effect Estimation : Probability Weighting

$$\mathbb{E}[Y \mid do(x_1, x_2)] = \mathbb{E}[\pi_0(W, X_1) I_{x_1}(X_1) Y \mid do(x_2)]$$

where $\pi_0(W, X_1) := \dfrac{P(W \mid do(x_1))}{P(W, X_1 \mid do(x_2))}$

Joint Treatment Effect Estimation : Probability Weighting

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

$$\mathbb{E}[Y|do(x_1, x_2)] = \mathbb{E}[\pi_0(W, X_1) I_{x_1}(X_1) Y | do(x_2)]$$

where $\pi_0(W, X_1) := \dfrac{P(W|do(x_1))}{P(W, X_1|do(x_2))}$

Joint Treatment Effect Estimation : Probability Weighting

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

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

$$\mathbb{E}[Y \mid do(x_1, x_2)] = \mathbb{E}[\pi_0(W, X_1) I_{x_1}(X_1) Y \mid do(x_2)]$$

where $\pi_0(W, X_1) := \dfrac{P(W \mid do(x_1))}{P(W, X_1 \mid do(x_2))}$

Joint Treatment Effect Estimation : Probability Weighting

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

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

$$= \sum_{w,y} y P(y \mid do(x_2), x_1, w) P(w \mid do(x_1))$$

$$\mathbb{E}[Y \mid do(x_1, x_2)] = \mathbb{E}[\pi_0(W, X_1) I_{x_1}(X_1) Y \mid do(x_2)]$$

where $\pi_0(W, X_1) := \dfrac{P(W \mid do(x_1))}{P(W, X_1 \mid do(x_2))}$

Joint Treatment Effect Estimation : Probability Weighting

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

$$= \sum_{w} \mathbb{E}[Y \,|\, do(x_2), x_1, w] P(w \,|\, do(x_1))$$

$$= \sum_{w,y} y P(y \,|\, do(x_2), x_1, w) P(w \,|\, do(x_1))$$

$$= \sum_{w,y} y \frac{P(y, x_1, w \,|\, do(x_2))}{P(x_1, w \,|\, do(x_2))} P(w \,|\, do(x_1))$$

$$\mathbb{E}[Y \,|\, do(x_1, x_2)] = \mathbb{E}[\pi_0(W, X_1) I_{x_1}(X_1) Y \,|\, do(x_2)]$$

where $\pi_0(W, X_1) := \dfrac{P(W \,|\, do(x_1))}{P(W, X_1 \,|\, do(x_2))}$

Joint Treatment Effect Estimation : Probability Weighting

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

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

$$= \sum_{w,y} y P(y \mid do(x_2), x_1, w) P(w \mid do(x_1))$$

$$= \sum_{w,y} y \frac{P(y, x_1, w \mid do(x_2))}{P(x_1, w \mid do(x_2))} P(w \mid do(x_1))$$

$$= \sum_{w,y,x_1'} \underbrace{\frac{P(w \mid do(x_1))}{P(x_1', w \mid do(x_2))}}_{=\pi_0(w, x_1')} I_{x_1}(x_1') y P(y, x_1', w \mid do(x_2))$$

$$\mathbb{E}[Y \mid do(x_1, x_2)] = \mathbb{E}[\pi_0(W, X_1) I_{x_1}(X_1) Y \mid do(x_2)]$$

where $\pi_0(W, X_1) := \dfrac{P(W \mid do(x_1))}{P(W, X_1 \mid do(x_2))}$

Joint Treatment Effect Estimation : Probability Weighting

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

$$= \sum\nolimits_{w} \mathbb{E}[Y \mid do(x_2), x_1, w] P(w \mid do(x_1))$$

$$= \sum\nolimits_{w,y} y P(y \mid do(x_2), x_1, w) P(w \mid do(x_1))$$

$$= \sum\nolimits_{w,y} y \frac{P(y, x_1, w \mid do(x_2))}{P(x_1, w \mid do(x_2))} P(w \mid do(x_1))$$

$$= \sum\nolimits_{w,y,x_1'} \underbrace{\frac{P(w \mid do(x_1))}{P(x_1', w \mid do(x_2))}}_{=\pi_0(w, x_1')} I_{x_1}(x_1') y P(y, x_1', w \mid do(x_2)) \quad = \mathbb{E}[\pi_0(W, X_1) I_{x_1}(X_1) Y \mid do(x_2)]$$

$$\mathbb{E}[Y \mid do(x_1, x_2)] = \mathbb{E}[\pi_0(W, X_1) I_{x_1}(X_1) Y \mid do(x_2)]$$

where $\pi_0(W, X_1) := \dfrac{P(W \mid do(x_1))}{P(W, X_1 \mid do(x_2))}$

Joint Treatment Effect Estimation : DR Estimand

$$\mathbb{E}[Y \mid do(x_1, x_2)] = \begin{cases} \mathbb{E}[\mu_0(W, x_1) \mid do(x_1)] \\ \\ \mathbb{E}[\pi_0(W, X_1) I_{x_1}(X_1) Y \mid do(x_2)] \end{cases}$$

Joint Treatment Effect Estimation : DR Estimand

$$\mathbb{E}[Y \mid do(x_1, x_2)] = \begin{cases} \mathbb{E}[\mu_0(W, x_1) \mid do(x_1)] \\ \mathbb{E}[\pi_0(W, X_1) I_{x_1}(X_1) Y \mid do(x_2)] \end{cases}$$

$$\mathbb{E}[Y \mid do(x_1, x_2)] = \mathbb{E}[\mu_0(W, x_1) \mid do(x_1)]$$

Joint Treatment Effect Estimation : DR Estimand

$$\mathbb{E}[Y \mid do(x_1, x_2)] = \begin{cases} \mathbb{E}[\mu_0(W, x_1) \mid do(x_1)] \\ \mathbb{E}[\pi_0(W, X_1) I_{x_1}(X_1) Y \mid do(x_2)] \end{cases}$$

$$\mathbb{E}[Y \mid do(x_1, x_2)] = \mathbb{E}[\mu_0(W, x_1) \mid do(x_1)]$$

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

Joint Treatment Effect Estimation : DR Estimand

$$\mathbb{E}[Y \mid do(x_1, x_2)] = \begin{cases} \mathbb{E}[\mu_0(W, x_1) \mid do(x_1)] \\ \mathbb{E}[\pi_0(W, X_1) I_{x_1}(X_1) Y \mid do(x_2)] \end{cases}$$

$$\mathbb{E}[Y \mid do(x_1, x_2)] = \mathbb{E}[\mu_0(W, x_1) \mid do(x_1)]$$

$$+ \mathbb{E}[\pi_0(W, X_1) I_{x_1}(X_1) Y \mid do(x_2)]$$

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

Joint Treatment Effect Estimation : DR Estimand

$$\mathbb{E}[Y \mid do(x_1, x_2)] = \begin{cases} \mathbb{E}[\mu_0(W, x_1) \mid do(x_1)] \\ \\ \mathbb{E}[\pi_0(W, X_1) I_{x_1}(X_1) Y \mid do(x_2)] \end{cases}$$

$$\mathbb{E}[Y \mid do(x_1, x_2)] = \mathbb{E}[\mu_0(W, x_1) \mid do(x_1)]$$

$$+ \mathbb{E}[\pi_0(W, X_1) I_{x_1}(X_1) Y \mid do(x_2)]$$

$$- \mathbb{E}[\pi_0(W, X_1) I_{x_1}(X_1) \mu_0(W, X_1) \mid do(x_2)]$$

Bias correcting term

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

Joint Treatment Effect Estimation : DR Estimand

$$\mathbb{E}[Y \mid do(x_1, x_2)] = \begin{cases} \mathbb{E}[\mu_0(W, x_1) \mid do(x_1)] \\ \\ \mathbb{E}[\pi_0(W, X_1) I_{x_1}(X_1) Y \mid do(x_2)] \end{cases}$$

DR Representation

$$\mathbb{E}[Y \mid do(x_1, x_2)] = \mathbb{E}\left[\pi_0 I_{x_1}(X_1)(Y - \mu_0) \middle| do(x_2) \right] + \mathbb{E}\left[\mu_0(W, x_1) \middle| do(x_1)) \right]$$

Joint Treatment Effect Estimation : DR Estimand

Joint Treatment Effect Estimation : DR Estimand

$$\mathbb{E}[Y \mid do(x_1, x_2)] := \Psi(\pi_0, \mu_0) = \mathbb{E}\left[\pi_0 I_{x_1}(X_1)(Y - \mu_0) \middle| do(x_2)\right] + \mathbb{E}\left[\mu_0(W, x_1) \middle| do(x_1))\right]$$

Joint Treatment Effect Estimation : DR Estimand

$$\mathbb{E}[Y \mid do(x_1, x_2)] := \Psi(\pi_0, \mu_0) = \mathbb{E}\left[\pi_0 I_{x_1}(X_1)(Y - \mu_0) \middle| do(x_2)\right] + \mathbb{E}\left[\mu_0(W, x_1) \middle| do(x_1)\right]$$

Doubly Robustness of DR representation

Joint Treatment Effect Estimation : DR Estimand

$$\mathbb{E}[Y \mid do(x_1, x_2)] := \Psi(\pi_0, \mu_0) = \mathbb{E}\left[\pi_0 I_{x_1}(X_1)(Y - \mu_0) \middle| do(x_2)\right] + \mathbb{E}\left[\mu_0(W, x_1) \middle| do(x_1)\right]$$

Doubly Robustness of DR representation

The DR representation above $\Psi(\pi_0, \mu_0)$ is unbiased when π_0 or μ_0 is misspecified; i.e., for arbitrary $\pi, \mu \in L_2$,

$$\Psi(\pi_0, \mu_0) = \Psi(\pi, \mu_0) = \Psi(\pi_0, \mu)$$

Outline of this talk

1. Preliminary and Problem Setup
   (1) Structural Causal Model
2. **Treatment-Treatment Interaction**
   (1) Identification
   (2) Estimand
   (3) Estimation and Error Analysis
   (4) Simulation Results
3. Multiple-Treatment Interaction (+ Omitted Results)
4. Future directions & Summary

DML Estimator

$$\mathbb{E}[Y \mid do(x_1, x_2)] = \mathbb{E}[\pi_0 I_{x_1}(X_1)(Y - \mu_0) \mid do(x_2)] + \mathbb{E}[\mu_0(W, x_1) \mid do(x_1))]$$

DML Estimator

$$\mathbb{E}[Y \mid do(x_1, x_2)] = \mathbb{E}[\pi_0 I_{x_1}(X_1)(Y - \mu_0) \mid do(x_2)] + \mathbb{E}[\mu_0(W, x_1) \mid do(x_1))]$$

DML estimator T^{dml}

DML Estimator

$$\mathbb{E}[Y \mid do(x_1, x_2)] = \mathbb{E}[\pi_0 I_{x_1}(X_1)(Y - \mu_0) \mid do(x_2)] + \mathbb{E}[\mu_0(W, x_1) \mid do(x_1))]$$

DML estimator T^{dml}

1. Randomly split the sample $D_1 := D_{1,a} \cup D_{1,b}$ and $D_2 := D_{2,a} \cup D_{2,b}$.

DML Estimator

$$\mathbb{E}[Y \mid do(x_1, x_2)] = \mathbb{E}[\pi_0 I_{x_1}(X_1)(Y - \mu_0) \mid do(x_2)] + \mathbb{E}[\mu_0(W, x_1) \mid do(x_1))]$$

DML estimator T^{dml}

1. Randomly split the sample $D_1 := D_{1,a} \cup D_{1,b}$ and $D_2 := D_{2,a} \cup D_{2,b}$.
2. Estimate (μ^a, π^a) for $\mu_0(W, X_1) := \mathbb{E}[Y \mid do(x_2), X_1, W]$ and $\pi_0(W, X_1) := P(W \mid do(x_1))/P(W, X_1 \mid do(x_2))$ using $D_{1,a}, D_{2,a}$.

DML Estimator

$$\mathbb{E}[Y \mid do(x_1, x_2)] = \mathbb{E}[\pi_0 I_{x_1}(X_1)(Y - \mu_0) \mid do(x_2)] + \mathbb{E}[\mu_0(W, x_1) \mid do(x_1))]$$

DML estimator T^{dml}

1. Randomly split the sample $D_1 := D_{1,a} \cup D_{1,b}$ and $D_2 := D_{2,a} \cup D_{2,b}$.
2. Estimate (μ^a, π^a) for $\mu_0(W, X_1) := \mathbb{E}[Y \mid do(x_2), X_1, W]$ and $\pi_0(W, X_1) := P(W \mid do(x_1))/P(W, X_1 \mid do(x_2))$ using $D_{1,a}, D_{2,a}$.
3. Evaluate $T^b := \mathbb{E}_{D_{2,b}}[\pi^a I_{x_1}(X_1)\{Y - \mu^a(W, X_1)\}] + \mathbb{E}_{D_{1,b}}[\mu^a(W, x_1)]$ using $D_{1,b}, D_{2,b}$.

DML Estimator

$$\mathbb{E}[Y \mid do(x_1, x_2)] = \mathbb{E}[\pi_0 I_{x_1}(X_1)(Y - \mu_0) \mid do(x_2)] + \mathbb{E}[\mu_0(W, x_1) \mid do(x_1))]$$

DML estimator T^{dml}

1. Randomly split the sample $D_1 := D_{1,a} \cup D_{1,b}$ and $D_2 := D_{2,a} \cup D_{2,b}$.
2. Estimate (μ^a, π^a) for $\mu_0(W, X_1) := \mathbb{E}[Y \mid do(x_2), X_1, W]$ and $\pi_0(W, X_1) := P(W \mid do(x_1))/P(W, X_1 \mid do(x_2))$ using $D_{1,a}, D_{2,a}$.
3. Evaluate $T^b := \mathbb{E}_{D_{2,b}}[\pi^a I_{x_1}(X_1)\{Y - \mu^a(W, X_1)\}] + \mathbb{E}_{D_{1,b}}[\mu^a(W, x_1)]$ using $D_{1,b}, D_{2,b}$.
4. Repeat 2-3 by switching $D_{1,a}, D_{2,a}$ and $D_{1,b}, D_{2,b}$ and obtain T^a with (μ^b, π^b).

DML Estimator

$$\mathbb{E}[Y \mid do(x_1, x_2)] = \mathbb{E}[\pi_0 I_{x_1}(X_1)(Y - \mu_0) \mid do(x_2)] + \mathbb{E}[\mu_0(W, x_1) \mid do(x_1))]$$

DML estimator T^{dml}

1. Randomly split the sample $D_1 := D_{1,a} \cup D_{1,b}$ and $D_2 := D_{2,a} \cup D_{2,b}$.
2. Estimate (μ^a, π^a) for $\mu_0(W, X_1) := \mathbb{E}[Y \mid do(x_2), X_1, W]$ and $\pi_0(W, X_1) := P(W \mid do(x_1))/P(W, X_1 \mid do(x_2))$ using $D_{1,a}, D_{2,a}$.
3. Evaluate $T^b := \mathbb{E}_{D_{2,b}}[\pi^a I_{x_1}(X_1)\{Y - \mu^a(W, X_1)\}] + \mathbb{E}_{D_{1,b}}[\mu^a(W, x_1)]$ using $D_{1,b}, D_{2,b}$.
4. Repeat 2-3 by switching $D_{1,a}, D_{2,a}$ and $D_{1,b}, D_{2,b}$ and obtain T^a with (μ^b, π^b).
5. $T^{dml} := (T^a + T^b)/2$.

Estimating Density Ratio

$$\pi_0(W, X_1) := \frac{P(W \mid do(x_1))}{P(W, X_1 \mid do(x_2))}$$

Estimating Density Ratio

$$\pi_0(W, X_1) := \frac{P(W \mid do(x_1))}{P(W, X_1 \mid do(x_2))}$$

Classification-based Density Ratio Estimation

Estimating Density Ratio

$$\pi_0(W, X_1) := \frac{P(W \mid do(x_1))}{P(W, X_1 \mid do(x_2))}$$

Classification-based Density Ratio Estimation

1. For each sample $(W_i, X_{1,i})$,
   1. set $\lambda_i = 0$ if $(W_i, X_{1,i}) \in D_1 \sim P(\mathbf{V} \mid do(x_1))$.
   2. Set $\lambda_i = 1$ if $(W_i, X_{1,i}) \in D_2 \sim P(\mathbf{V} \mid do(x_2))$

Estimating Density Ratio

$$\pi_0(W, X_1) := \frac{P(W \mid do(x_1))}{P(W, X_1 \mid do(x_2))}$$

Classification-based Density Ratio Estimation

1. For each sample $(W_i, X_{1,i})$,
   1. set $\lambda_i = 0$ if $(W_i, X_{1,i}) \in D_1 \sim P(\mathbf{V} \mid do(x_1))$.
   2. Set $\lambda_i = 1$ if $(W_i, X_{1,i}) \in D_2 \sim P(\mathbf{V} \mid do(x_2))$
2. Augment (W_i, X_i, λ_i). Then,

$$\pi_0(W, X_1) = \frac{P(\lambda = 1)}{P(\lambda = 0)} \frac{P(\lambda = 0 \mid W, X_1)}{P(\lambda = 1 \mid W, X_1)}$$

Finite Sample Error Analysis

Finite Sample Error Analysis

Finite Sample Error Analysis

Finite Sample Error Analysis

With a probability $1 - 4\epsilon$

$$T^{dml} - \mathbb{E}[Y \,|\, do(x_1, x_2)] \leq (a) + (b),$$

Finite Sample Error Analysis

Finite Sample Error Analysis

With a probability $1 - 4\epsilon$

$$T^{dml} - \mathbb{E}[Y \,|\, do(x_1, x_2)] \leq (a) + (b),$$

$$(a) := \frac{1}{2} \sum_{k \in \{a,b\}} \frac{\sqrt{2}}{\sqrt{n_2 \epsilon}} \left(\sqrt{\mathbb{V}_{P_2}[\pi_0(Y - \mu_0)]} + \|\pi^k(Y - \mu^k) - \pi_0(Y - \mu_0)\|_{P_2} \right) + \frac{\sqrt{2}}{\sqrt{n_1 \epsilon}} \left(\sqrt{\mathbb{V}_{P_1}[\mu_0]} + \|\mu^k - \mu_0\|_{P_1} \right)$$

where $n_1 := |D_1|$, $n_2 := |D_2|$, $\|\|_P$ is a $L_2(P)$ norm.

Finite Sample Error Analysis

Finite Sample Error Analysis

With a probability $1 - 4\epsilon$

$$T^{dml} - \mathbb{E}[Y \mid do(x_1, x_2)] \leq \boxed{(a)} + \boxed{(b)},$$

$$(a) := \frac{1}{2} \sum_{k \in \{a,b\}} \frac{\sqrt{2}}{\sqrt{n_2 \epsilon}} \left(\sqrt{\mathbb{V}_{P_2}[\pi_0(Y - \mu_0)]} + \|\pi^k(Y - \mu^k) - \pi_0(Y - \mu_0)\|_{P_2} \right) + \frac{\sqrt{2}}{\sqrt{n_1 \epsilon}} \left(\sqrt{\mathbb{V}_{P_1}[\mu_0]} + \|\mu^k - \mu_0\|_{P_1} \right)$$

$$(b) := \sum_{k \in \{a,b\}} \mathbb{E}_{P_2}[\{\mu^k - \mu_0\}\{\pi_0 - \pi^k\}]$$

where $n_1 := |D_1|$, $n_2 := |D_2|$, $\|\|_P$ is a $L_2(P)$ norm.

Asymptotic Error Analysis

Asymptotic Error Analysis

Asymptotic Error Analysis

Asymptotic Error Analysis

Asymptotic Error Analysis

Assume

1. $||\pi^k - \pi_0||_{P_2}||\mu^k - \mu_0||_{P_2} = o_{P_2}(1), k \in \{a, b\}$; i.e., π^k, μ^k converges to the true parameters.

Asymptotic Error Analysis

Asymptotic Error Analysis

Assume

1. $\|\pi^k - \pi_0\|_{P_2} \|\mu^k - \mu_0\|_{P_2} = o_{P_2}(1), k \in \{a, b\}$; i.e., π^k, μ^k converges to the true parameters.
2. $\|\pi^k(Y - \mu^k) - \pi_0(Y - \mu_0)\|_{P_2} = O_P(1)$; i.e., $\pi^k(Y - \mu^k) - \pi_0(Y - \mu_0)$ is bounded.

Asymptotic Error Analysis

Asymptotic Error Analysis

Assume

1. $\|\pi^k - \pi_0\|_{P_2}\|\mu^k - \mu_0\|_{P_2} = o_{P_2}(1), k \in \{a, b\}$; i.e., π^k, μ^k converges to the true parameters.
2. $\|\pi^k(Y - \mu^k) - \pi_0(Y - \mu_0)\|_{P_2} = O_P(1)$; i.e., $\pi^k(Y - \mu^k) - \pi_0(Y - \mu_0)$ is bounded.

$$T^{dml} - \mathbb{E}[Y \mid do(x_1, x_2)] = (R_1 + R_2) + \sum_{k \in \{a,b\}} O_{P_2}\left(\|\pi^k - \pi_0\|_{P_2}\|\mu^k - \mu_0\|_{P_2}\right)$$

Asymptotic Error Analysis

Asymptotic Error Analysis

Assume

1. $\|\pi^k - \pi_0\|_{P_2}\|\mu^k - \mu_0\|_{P_2} = o_{P_2}(1), k \in \{a, b\}$; i.e., π^k, μ^k converges to the true parameters.
2. $\|\pi^k(Y - \mu^k) - \pi_0(Y - \mu_0)\|_{P_2} = O_P(1)$; i.e., $\pi^k(Y - \mu^k) - \pi_0(Y - \mu_0)$ is bounded.

$$T^{dml} - \mathbb{E}[Y \mid do(x_1, x_2)] = (R_1 + R_2) + \sum_{k \in \{a,b\}} O_{P_2}\left(\|\pi^k - \pi_0\|_{P_2}\|\mu^k - \mu_0\|_{P_2}\right)$$

where R_i (for $i \in \{1,2\}$) is a random variable s.t. $\sqrt{n_i}R_i \xrightarrow{d} Z_i \sim$ normal$(0,\sigma_i^2)$, where $\sigma_1^2 := \mathbb{V}_{P_1}[\mu_0(W, x_1)]$ and $\sigma_0^2 := \mathbb{V}_{P_2}[\pi_0(Y - \mu_0)]$.

Robustness of DML estimator

$$T^{dml} - \mathbb{E}[Y \mid do(x_1, x_2)] = (R_1 + R_2) + \sum_{k \in \{a,b\}} O_{P_2}\left(\|\pi^k - \pi_0\|_{P_2} \|\mu^k - \mu_0\|_{P_2} \right)$$

Robustness of DML estimator

$$T^{dml} - \mathbb{E}[Y \mid do(x_1, x_2)] = (R_1 + R_2) + \sum_{k \in \{a,b\}} O_{P_2}\left(\|\pi^k - \pi_0\|_{P_2} \|\mu^k - \mu_0\|_{P_2} \right)$$

Doubly Robustness & Debiasedness

Robustness of DML estimator

$$T^{dml} - \mathbb{E}[Y \mid do(x_1, x_2)] = (R_1 + R_2) + \sum_{k \in \{a,b\}} O_{P_2}\left(\|\pi^k - \pi_0\|_{P_2} \|\mu^k - \mu_0\|_{P_2} \right)$$

Doubly Robustness & Debiasedness

1. **Doubly Robustness (DR):** T^{dml} converges to $\mathbb{E}[Y \mid do(x_1, x_2)]$ at $n^{-1/2}$-rate (where $n := \min(n_1, n_2)$) if $\pi = \pi_0$ or $\mu = \mu_0$.

Robustness of DML estimator

$$T^{dml} - \mathbb{E}[Y \mid do(x_1, x_2)] = (R_1 + R_2) + \sum_{k \in \{a,b\}} O_{P_2}\left(\|\pi^k - \pi_0\|_{P_2} \|\mu^k - \mu_0\|_{P_2} \right)$$

Doubly Robustness & Debiasedness

1. **Doubly Robustness (DR):** T^{dml} converges to $\mathbb{E}[Y \mid do(x_1, x_2)]$ at $n^{-1/2}$-rate (where $n := \min(n_1, n_2)$) if $\pi = \pi_0$ or $\mu = \mu_0$.

2. **Debiasedness (DB):** T^{dml} converges to $\mathbb{E}[Y \mid do(x_1, x_2)]$ at $n^{-1/2}$-rate if π, μ converges to π_0, μ_0 at least at $n^{-1/4}$ rate.

Efficiency of DML estimator

$$T^{dml} - \mathbb{E}[Y \mid do(x_1, x_2)] = (R_1 + R_2) + \sum_{k \in \{a,b\}} O_{P_2}\left(\|\pi^k - \pi_0\|_{P_2} \|\mu^k - \mu_0\|_{P_2} \right)$$

Efficiency of DML estimator

$$T^{dml} - \mathbb{E}[Y \mid do(x_1, x_2)] = (R_1 + R_2) + \sum_{k \in \{a,b\}} O_{P_2}\left(\|\pi^k - \pi_0\|_{P_2} \|\mu^k - \mu_0\|_{P_2} \right)$$

CAN and Efficiency

Efficiency of DML estimator

$$T^{dml} - \mathbb{E}[Y \mid do(x_1, x_2)] = \boxed{(R_1 + R_2)} + \sum_{k \in \{a,b\}} O_{P_2}\left(\|\pi^k - \pi_0\|_{P_2} \|\mu^k - \mu_0\|_{P_2} \right)$$

CAN and Efficiency

1. **Consistency and Asymptotic Normality (CAN):** Under the {DR,DB} conditions, T^{dml} achieves consistency and asymptotic normality (CAN).

Efficiency of DML estimator

$$T^{dml} - \mathbb{E}[Y \mid do(x_1, x_2)] = \boxed{(R_1 + R_2)} + \sum_{k \in \{a,b\}} O_{P_2}\left(\|\pi^k - \pi_0\|_{P_2} \|\mu^k - \mu_0\|_{P_2} \right)$$

CAN and Efficiency

1. **Consistency and Asymptotic Normality (CAN):** Under the {DR,DB} conditions, T^{dml} achieves consistency and asymptotic normality (CAN).

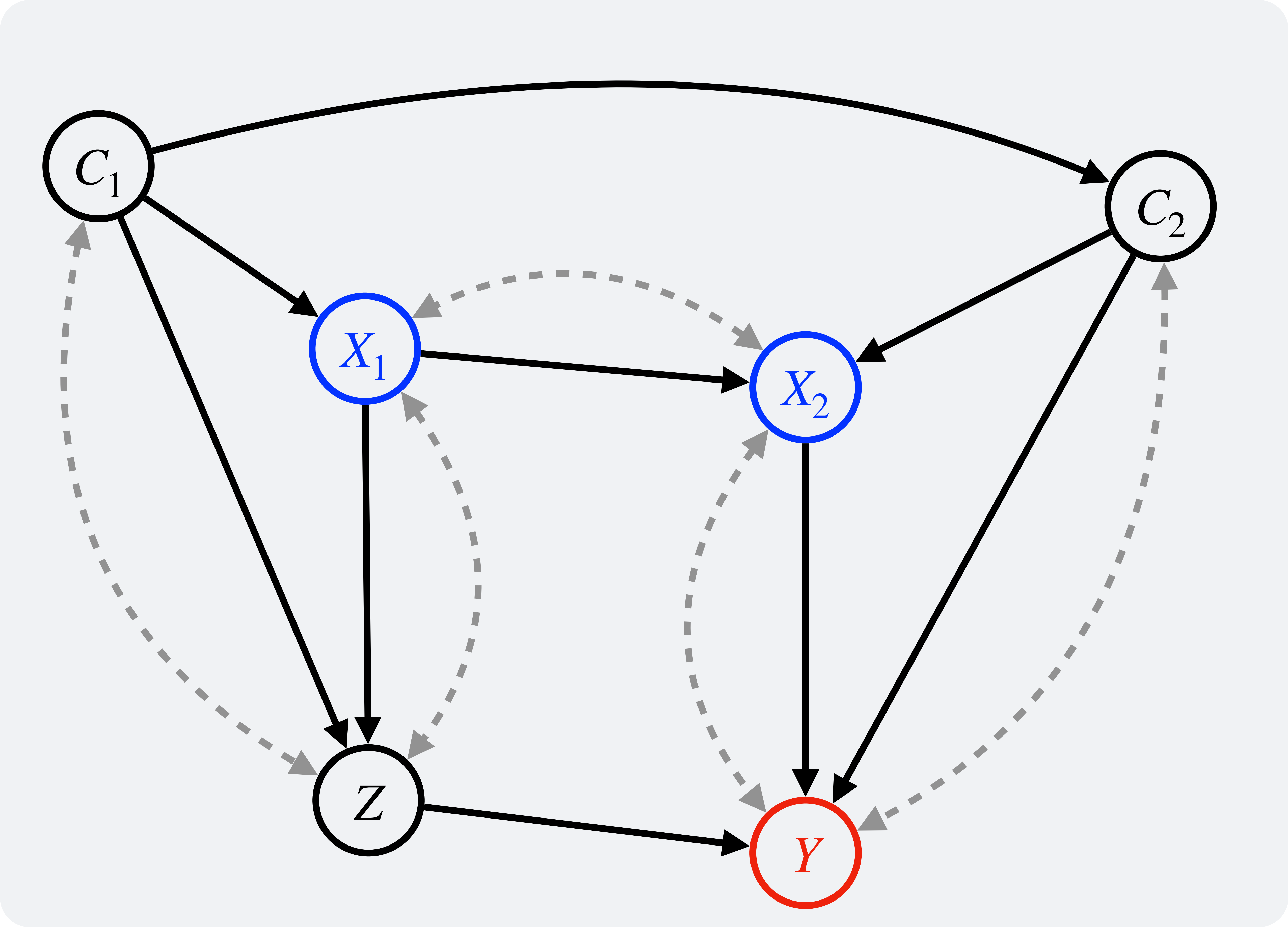
2. **Statistical Efficiency:** Under the {DR,DB} conditions, T^{dml} achieves the nonparametric efficiency bound (i.e., achieves the minimum variance).

Outline of this talk

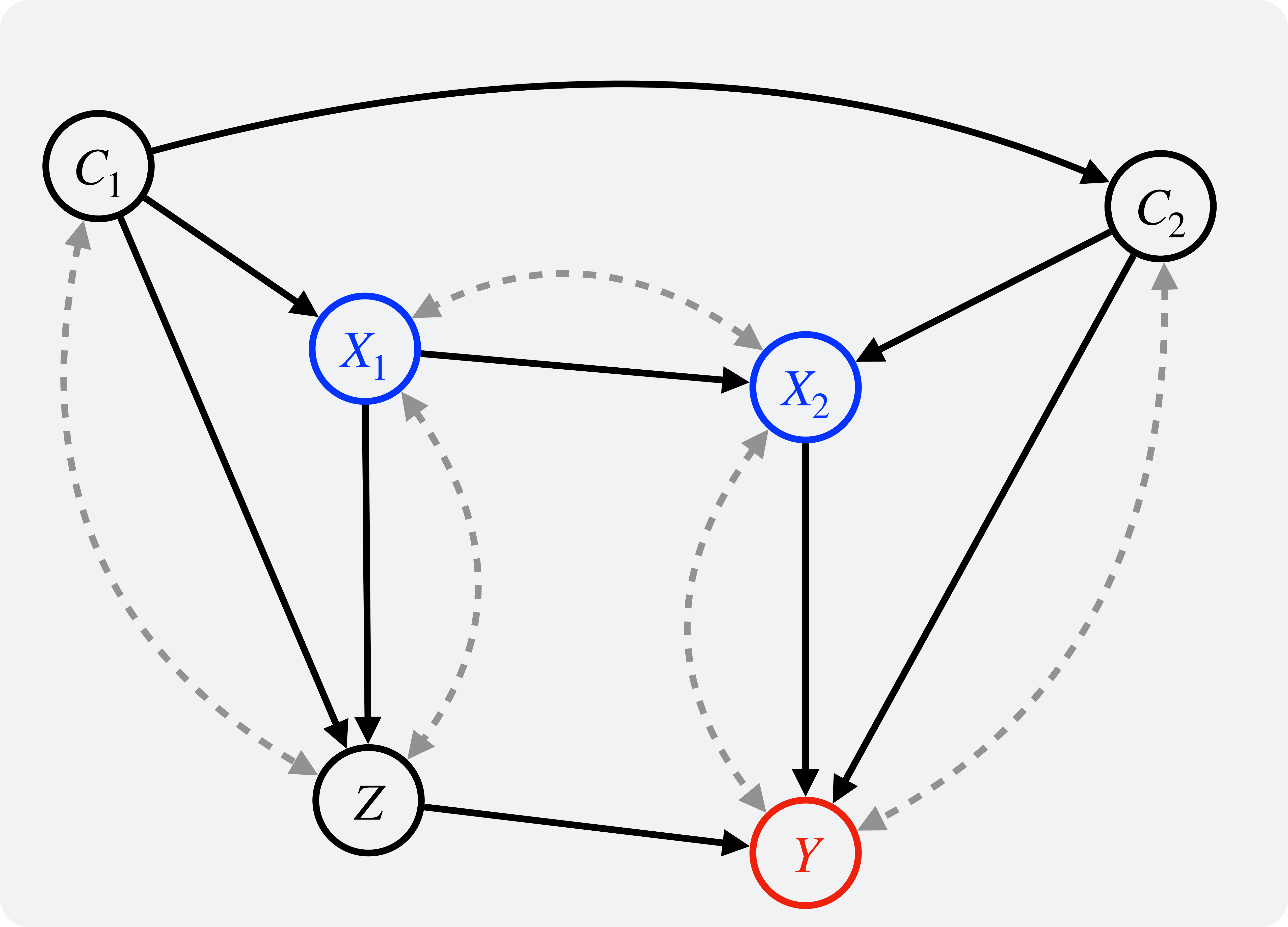
1. Preliminary and Problem Setup
   (1) Structural Causal Model
2. **Treatment-Treatment Interaction**
   (1) Identification
   (2) Estimand
   (3) Estimation and Error Analysis
   (4) Simulation Results
3. Multiple-Treatment Interaction (+ Omitted Results)
4. Future directions & Summary

Simulations: Synthetic Dataset

Simulations: Synthetic Dataset

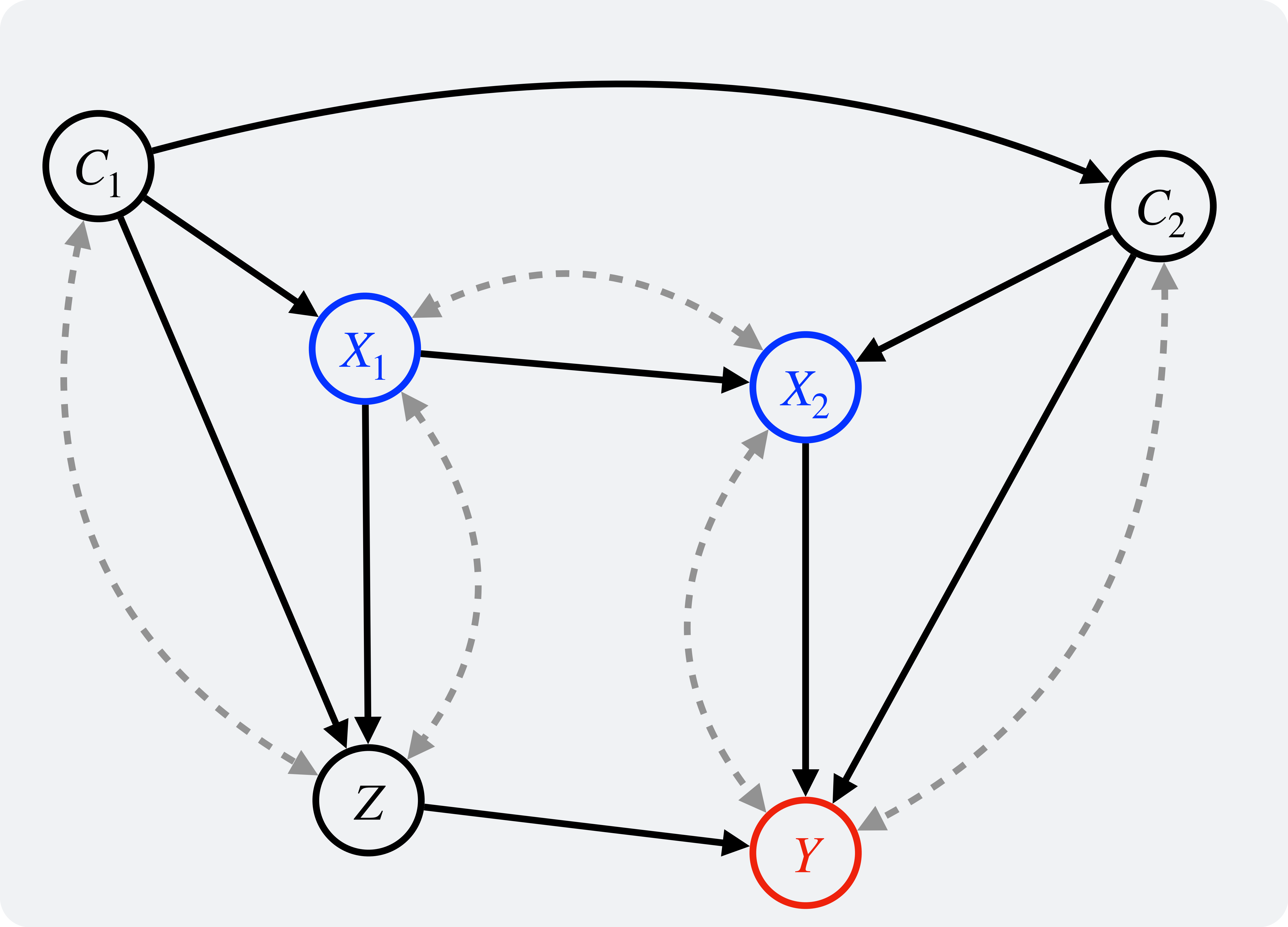


Simulations: Synthetic Dataset



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

Simulations: Synthetic Dataset



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

Input:

1. $D_1 \overset{iid}{\sim} P(C_1, C_2, X_2, Z, Y \mid do(x_1))$
2. $D_2 \overset{iid}{\sim} P(C_1, C_2, X_1, Z, Y \mid do(x_2))$

Simulation Scenario

$$\mathbb{E}[Y \,|\, do(x_1, x_2)] = \mathbb{E}[\pi_0 I_{x_1}(X_1)(Y - \mu_0) \,|\, do(x_2)] + \mathbb{E}[\mu_0(W, x_1) \,|\, do(x_1))]$$

Simulation Scenario

$$\mathbb{E}[Y \mid do(x_1, x_2)] = \mathbb{E}[\pi_0 I_{x_1}(X_1)(Y - \mu_0) \mid do(x_2)] + \mathbb{E}[\mu_0(W, x_1) \mid do(x_1))]$$

Scenario 1. Regular learning environment.

Simulation Scenario

$$\mathbb{E}[Y \mid do(x_1, x_2)] = \mathbb{E}[\pi_0 I_{x_1}(X_1)(Y - \mu_0) \mid do(x_2)] + \mathbb{E}[\mu_0(W, x_1) \mid do(x_1))]$$

Scenario 1. Regular learning environment.

Scenario 2. Adding the "convergence noise" $\epsilon \sim$ normal(a_n, b_n^2) to the estimated nuisance π, μ, where $a_n, b_n = O_P(n^{-1/4})$; i.e., $\pi \leftarrow \pi + \epsilon$ and $\mu \leftarrow \mu + \epsilon$. This forces π, μ converges at $n^{-1/4}$ rate to highlight the fast convergence of the estimator T^{dml}.

Simulation Scenario

$$\mathbb{E}[Y \,|\, do(x_1, x_2)] = \mathbb{E}[\pi_0 I_{x_1}(X_1)(Y - \mu_0) \,|\, do(x_2)] + \mathbb{E}[\mu_0(W, x_1) \,|\, do(x_1))]$$

Scenario 1. Regular learning environment.

Scenario 2. Adding the “convergence noise” $\epsilon \sim$ normal(a_n, b_n^2) to the estimated nuisance π, μ, where $a_n, b_n = O_P(n^{-1/4})$; i.e., $\pi \leftarrow \pi + \epsilon$ and $\mu \leftarrow \mu + \epsilon$. This forces π, μ converges at $n^{-1/4}$ rate to highlight the fast convergence of the estimator T^{dml}.

Scenario 3. π is wrongly estimated .

Simulation Scenario

$$\mathbb{E}[Y \,|\, do(x_1, x_2)] = \mathbb{E}[\pi_0 I_{x_1}(X_1)(Y - \mu_0) \,|\, do(x_2)] + \mathbb{E}[\mu_0(W, x_1) \,|\, do(x_1))]$$

Scenario 1. Regular learning environment.

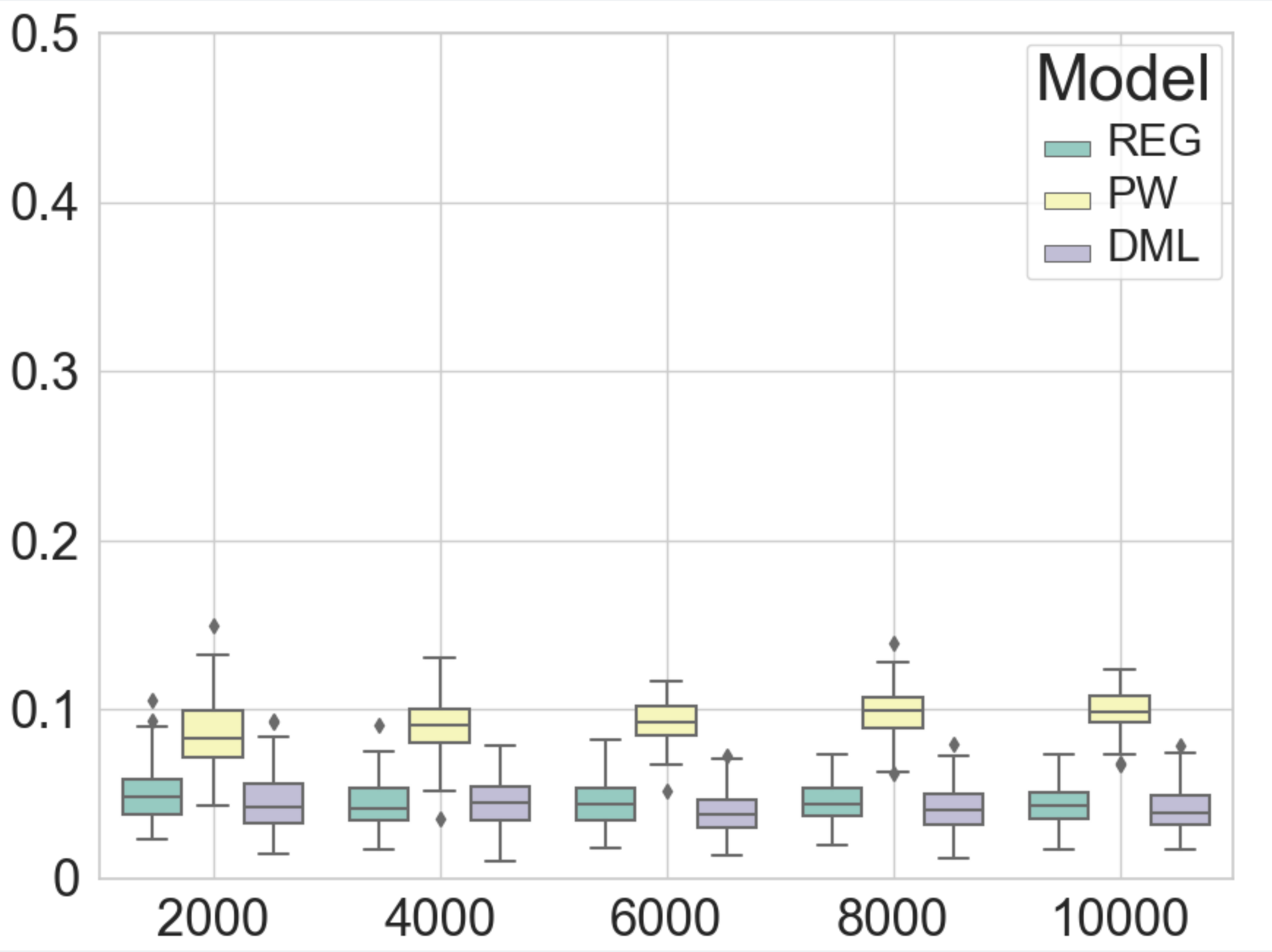
Scenario 2. Adding the "convergence noise" $\epsilon \sim$ normal(a_n, b_n^2) to the estimated nuisance π, μ, where $a_n, b_n = O_P(n^{-1/4})$; i.e., $\pi \leftarrow \pi + \epsilon$ and $\mu \leftarrow \mu + \epsilon$. This forces π, μ converges at $n^{-1/4}$ rate to highlight the fast convergence of the estimator T^{dml}.

Scenario 3. π is wrongly estimated .

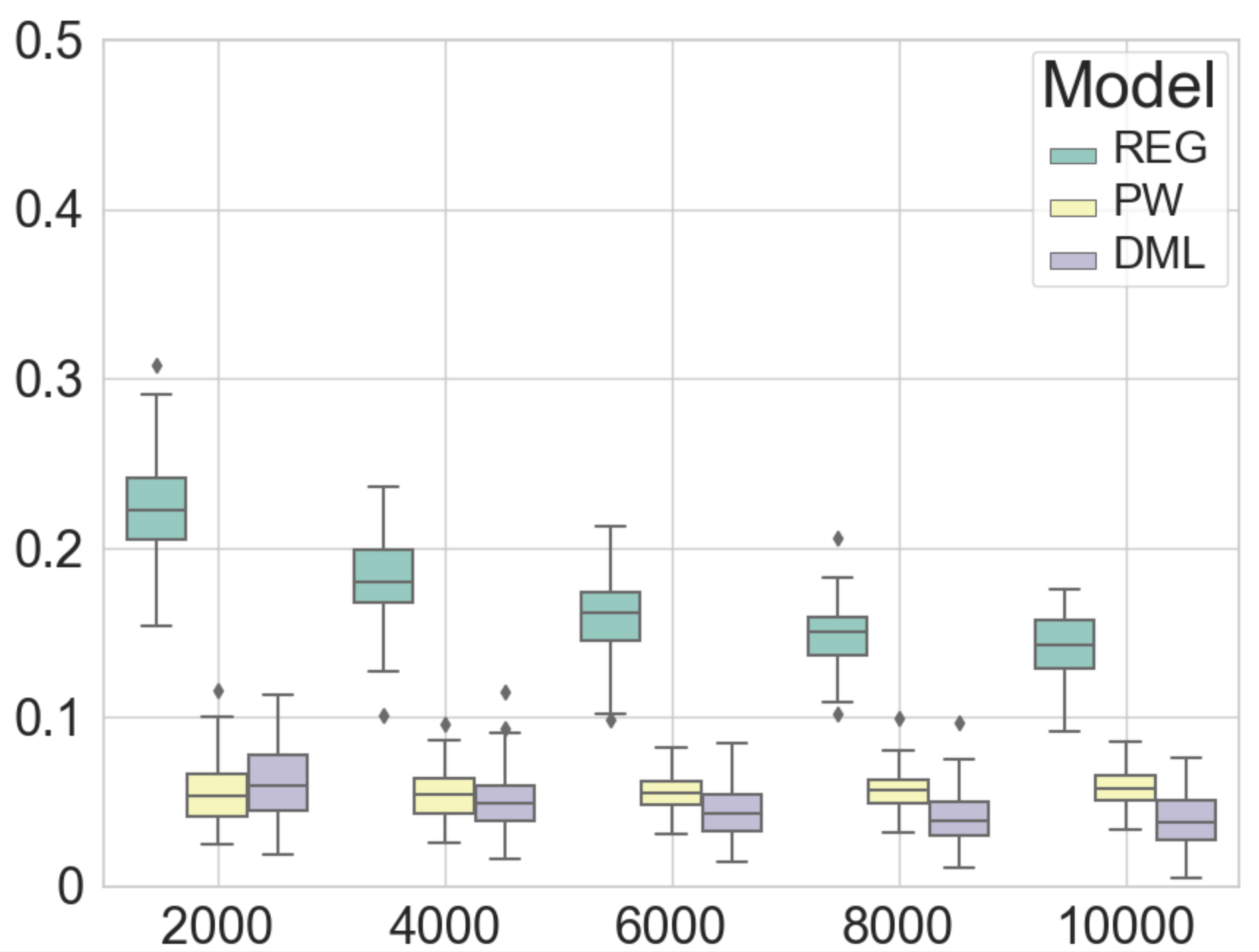
Scenario 4. μ is wrongly estimated .

Simulation Results

Scenario 1

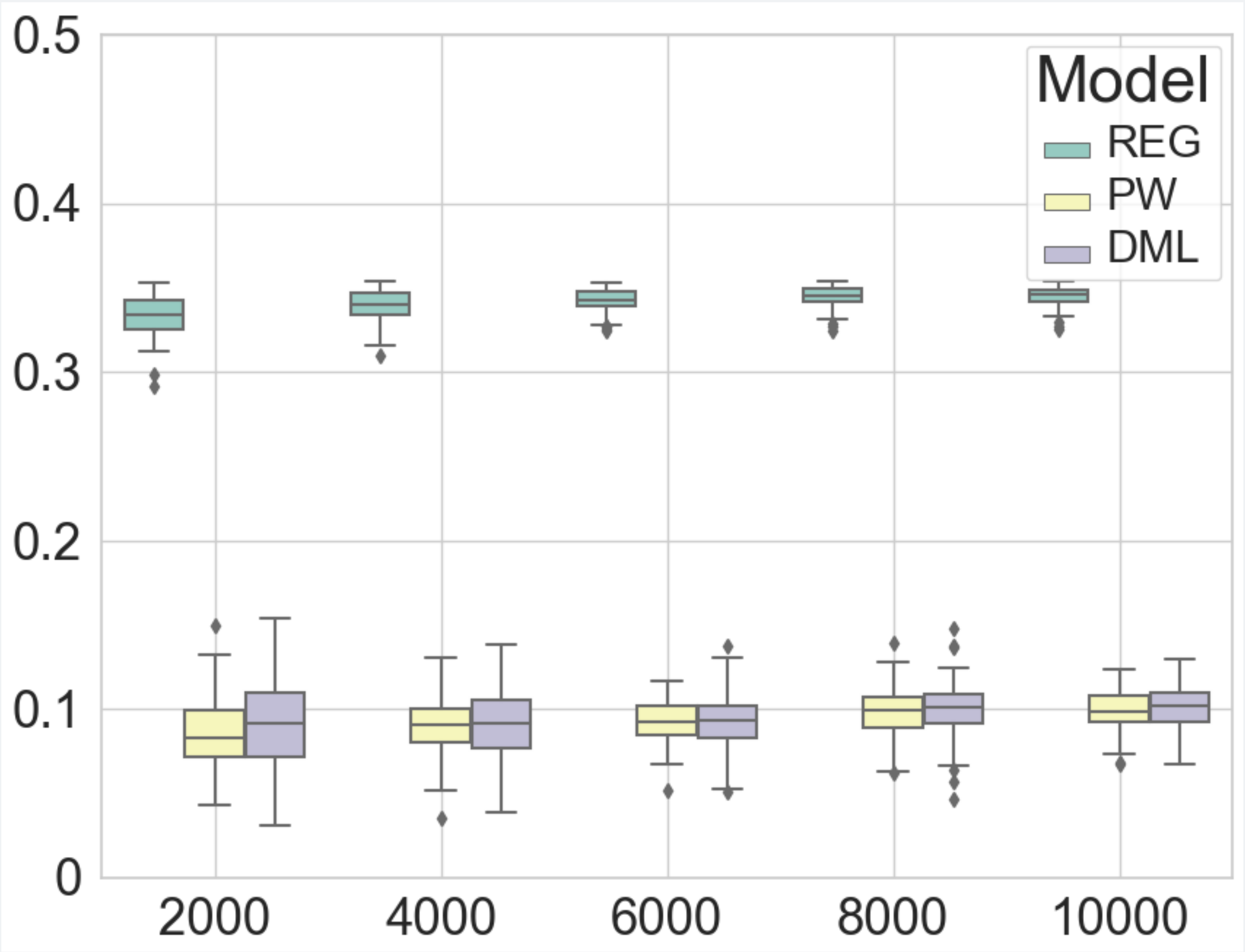


Scenario 2

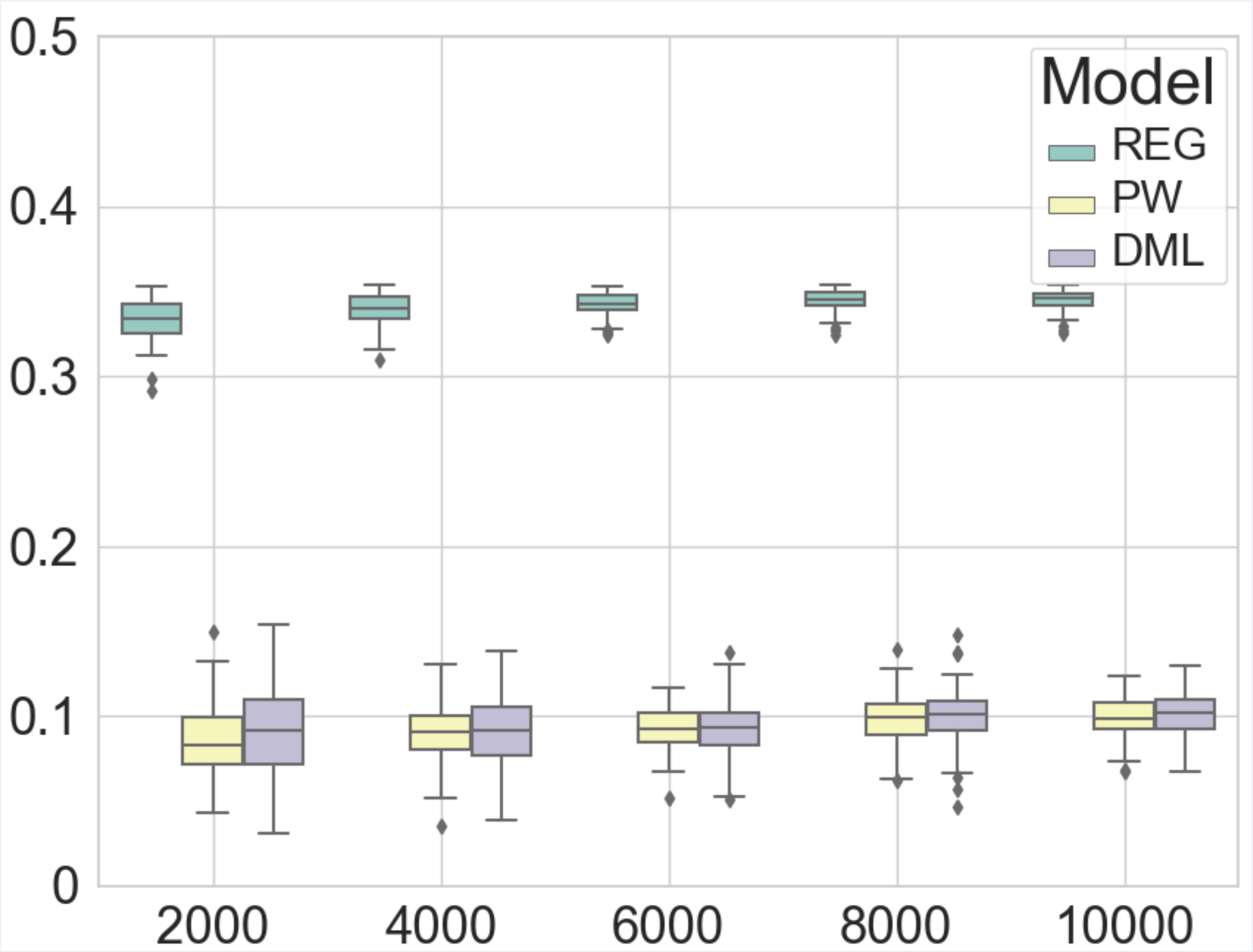


Simulation Results

Scenario 3



Scenario 4



Simulations: Project STAR

Simulations: Project STAR

1. The Tennessee Student/Teacher Achievement Ratio (STAR) project was a randomized controlled trial that assigned students to different student/teacher ratios of "small," "medium," and "large" through randomization. The outcome measure was the testing score.

Simulations: Project STAR

1. The Tennessee Student/Teacher Achievement Ratio (STAR) project was a randomized controlled trial that assigned students to different student/teacher ratios of "small," "medium," and "large" through randomization. The outcome measure was the testing score.

2. The study was longitudinal, with students randomized from Pre-K to the 3rd grade.

Simulations: Project STAR

1. The Tennessee Student/Teacher Achievement Ratio (STAR) project was a randomized controlled trial that assigned students to different student/teacher ratios of "small," "medium," and "large" through randomization. The outcome measure was the testing score.

2. The study was longitudinal, with students randomized from Pre-K to the 3rd grade.

3. We only focus on the student/teacher ratio for the pre-K students (X_1) and the third-grade students (X_2) as well as their testing scores at the third grade (Y).

Simulations: Project STAR

1. The Tennessee Student/Teacher Achievement Ratio (STAR) project was a randomized controlled trial that assigned students to different student/teacher ratios of "small," "medium," and "large" through randomization. The outcome measure was the testing score.

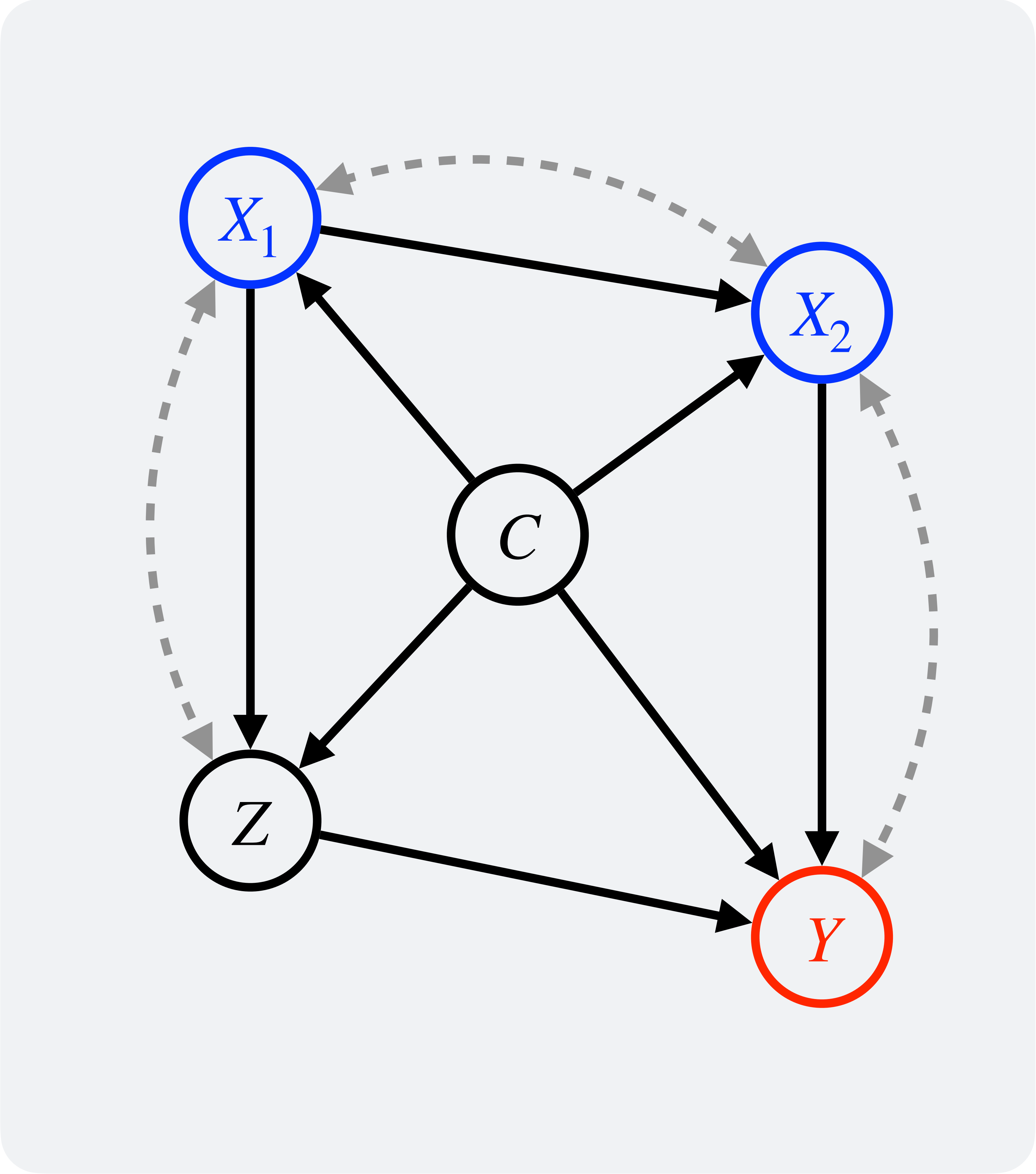
2. The study was longitudinal, with students randomized from Pre-K to the 3rd grade.

3. We only focus on the student/teacher ratio for the pre-K students (X_1) and the third-grade students (X_2) as well as their testing scores at the third grade (Y).

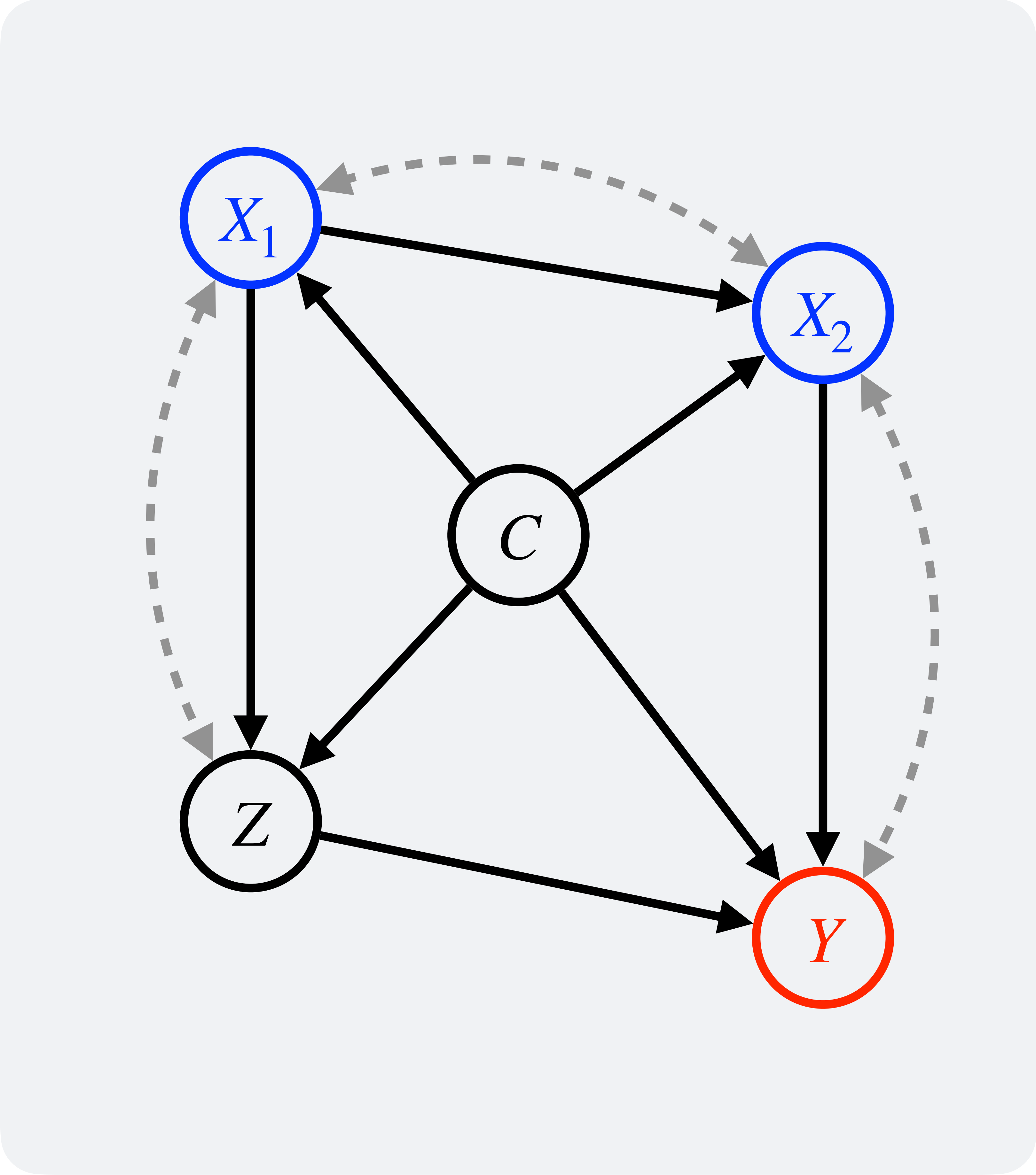
4. We created confounders to to ensure that the data aligns with our graphical settings.

Data Generating Process for STAR

Data Generating Process for STAR



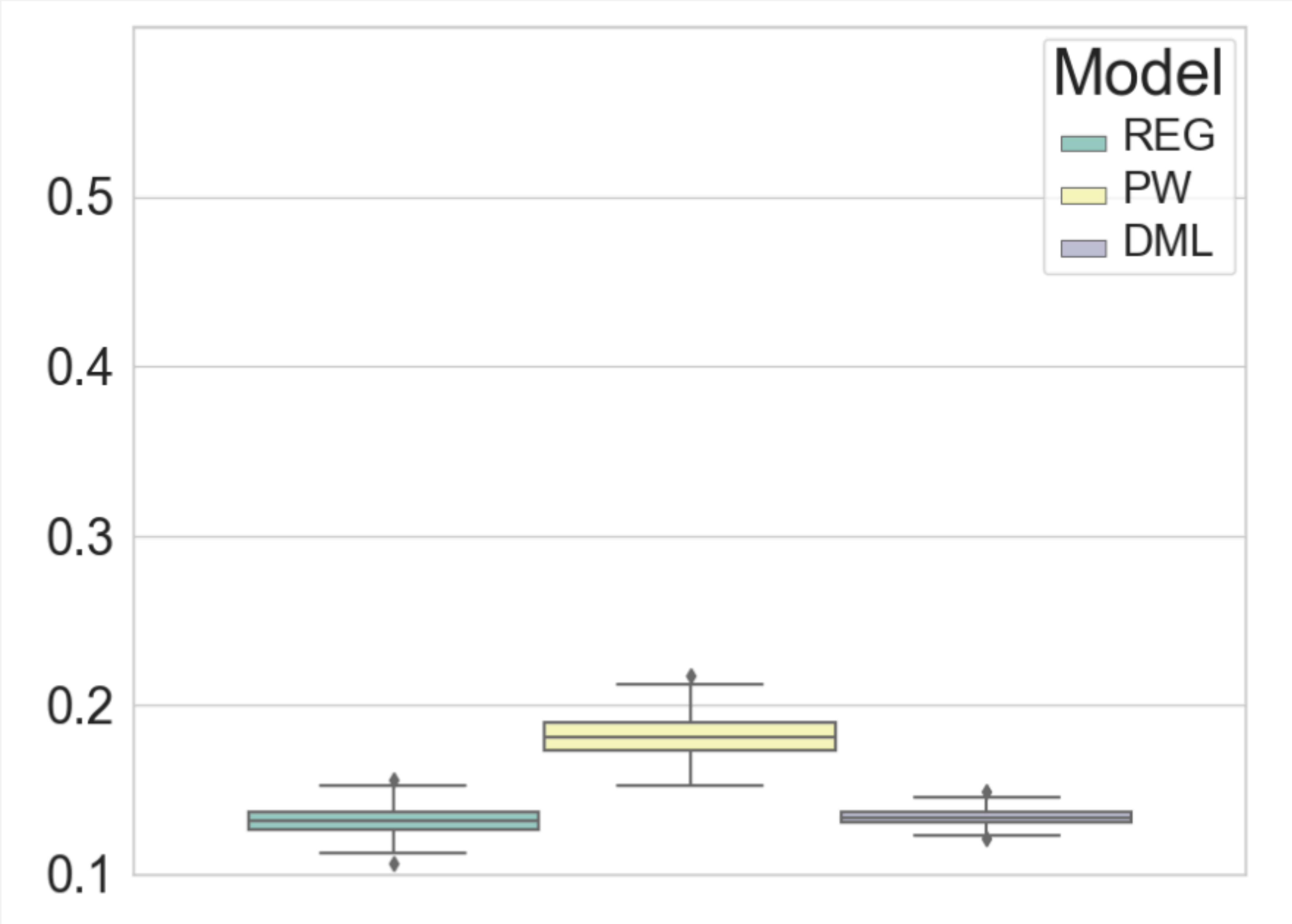
Data Generating Process for STAR



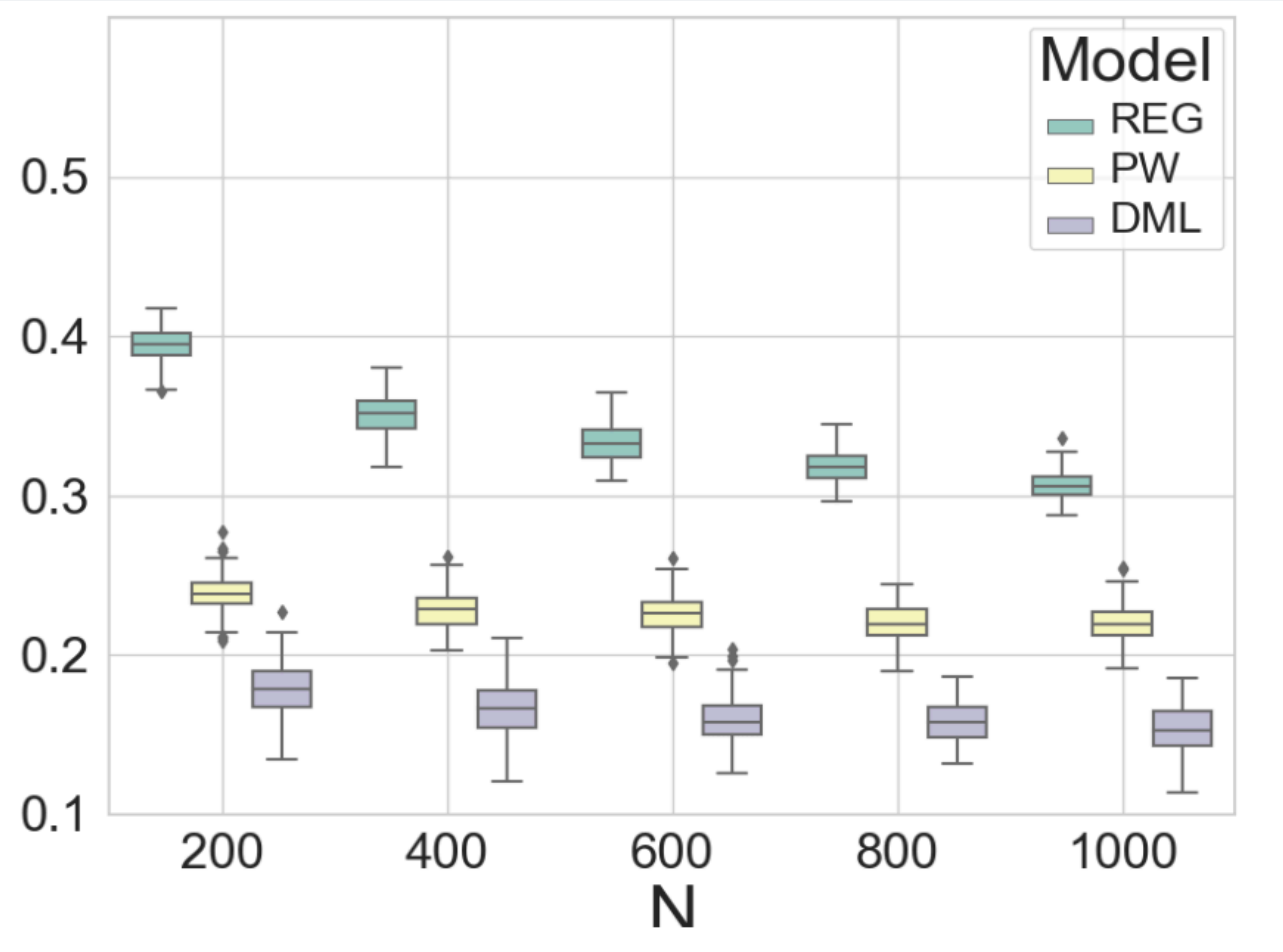
- X_1: The student/teacher ratio for the pre-K students.
- X_2: The student/teacher ratio for the 3rd grade students.
- Y: The testing score of the third grade students.
- C: Baseline covariates (socioeconomic factors, teacher's education, etc.)
- Z: The testing score of the pre-K students.

Simulation Results: STAR

Scenario 1

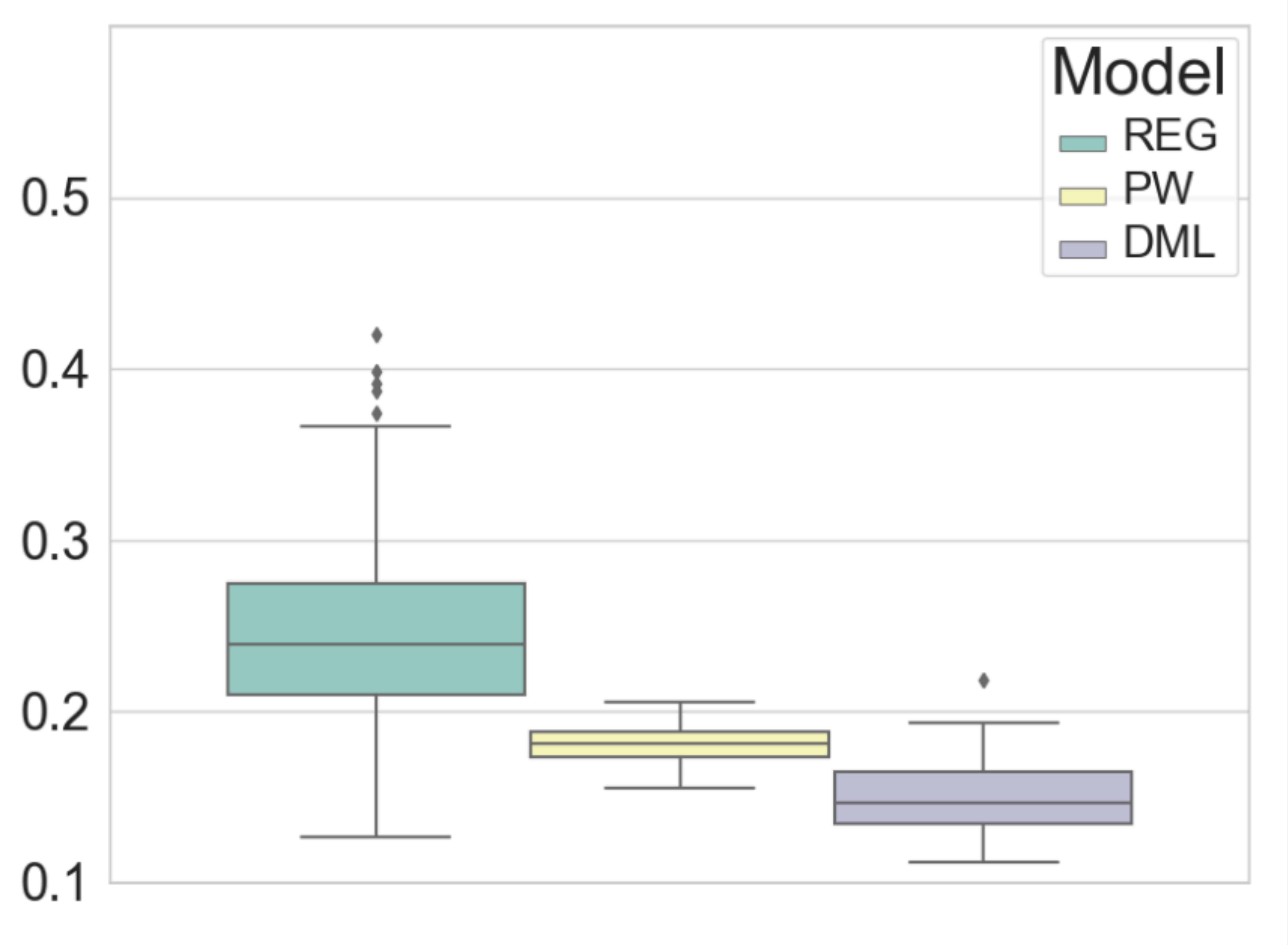


Scenario 2

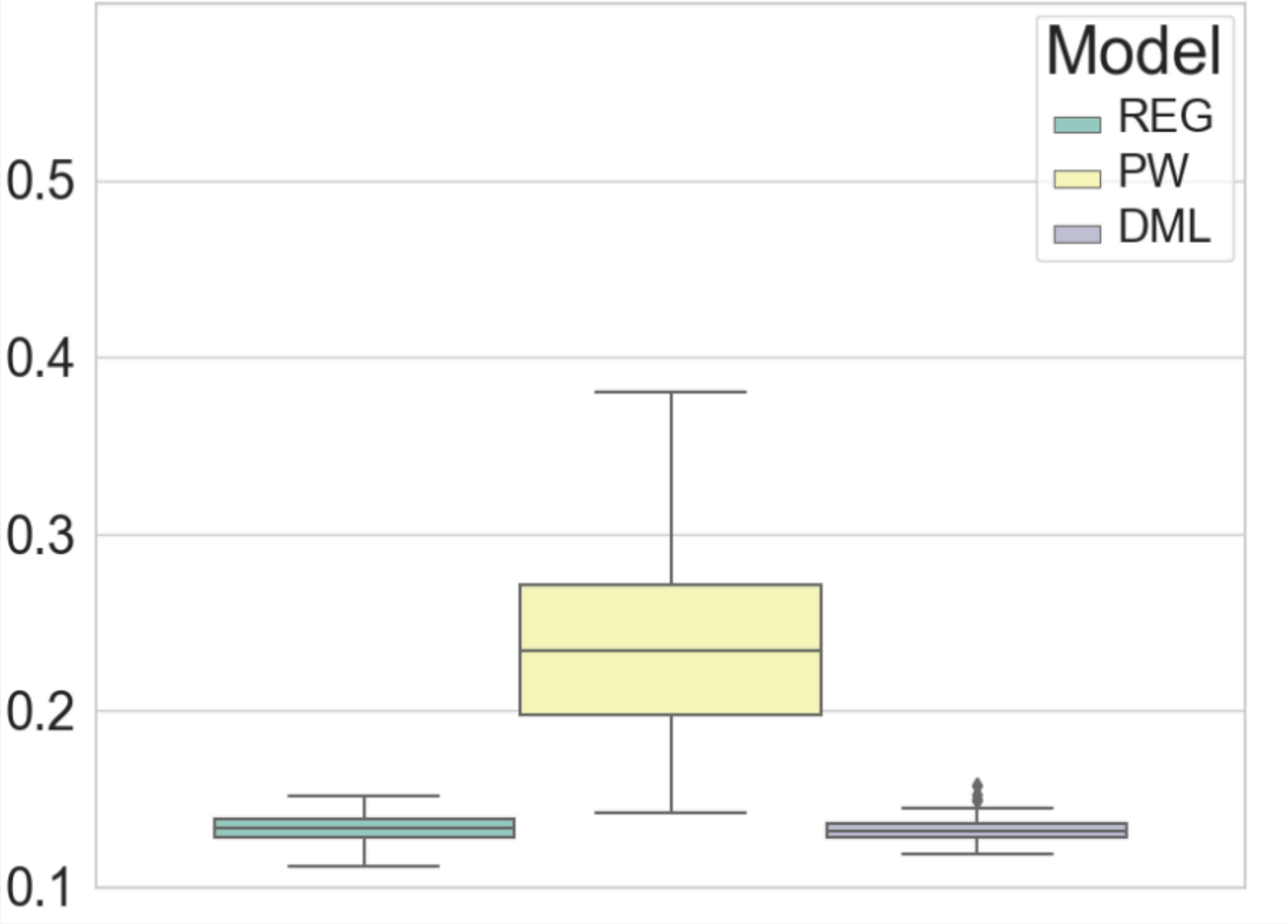


Simulation Results: STAR

Scenario 3



Scenario 4

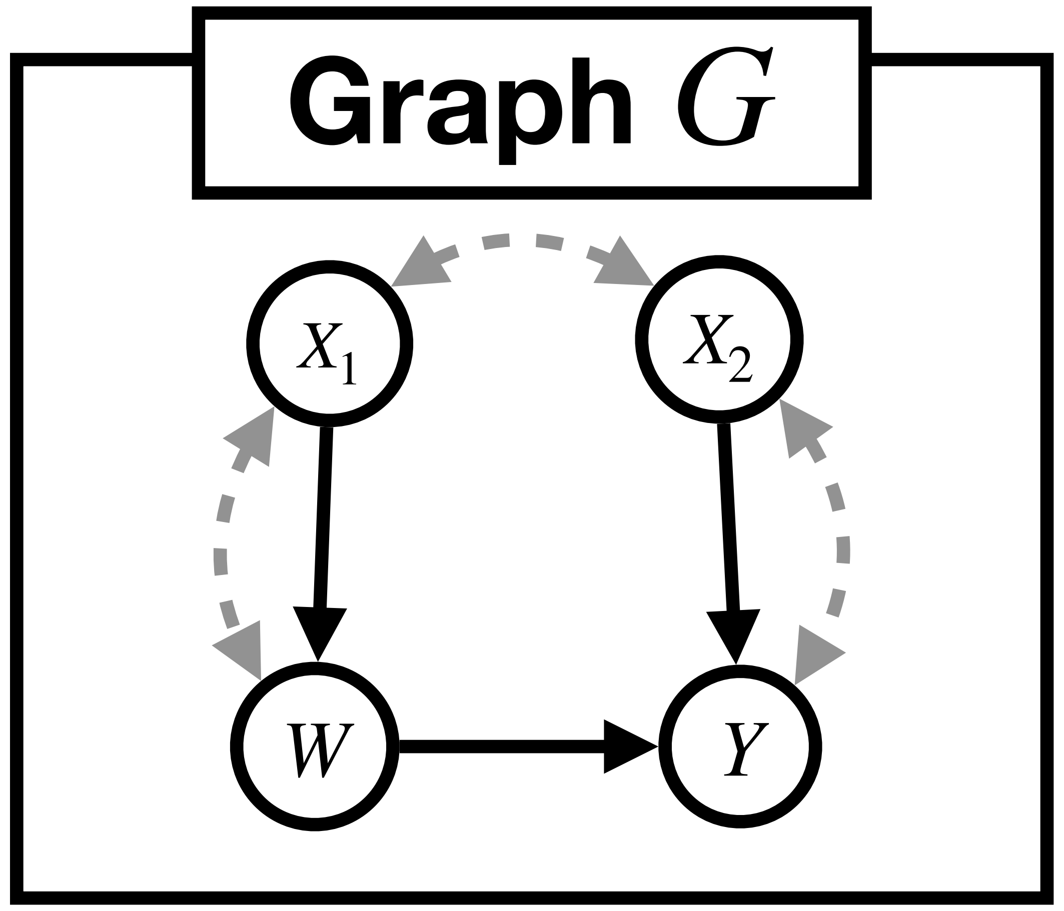


Outline of this talk

1. Preliminary and Problem Setup
   (1) Structural Causal Model
2. Treatment-Treatment Interaction
   (1) Identification
   (2) Estimand
   (3) Estimation and Error Analysis
   (4) Simulation Results
3. **Multiple-Treatment Interaction (+ Omitted Results)**
4. Future directions

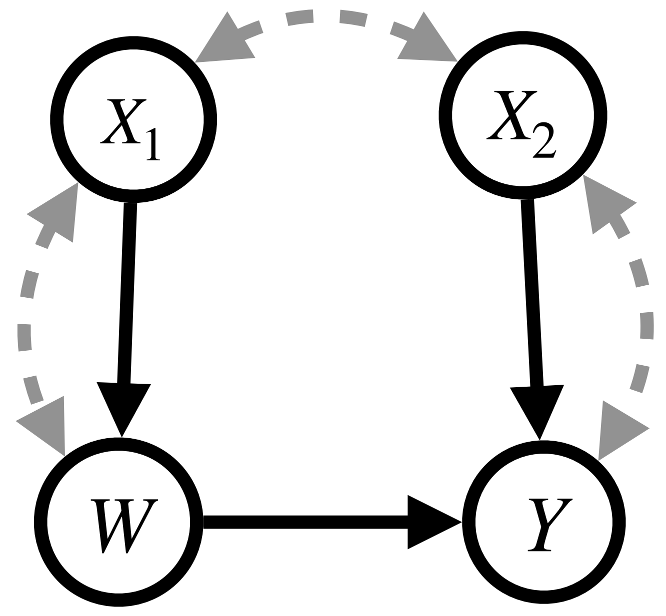
Multiple Treatment Interaction

Multiple Treatment Interaction



Multiple Treatment Interaction

Graph G

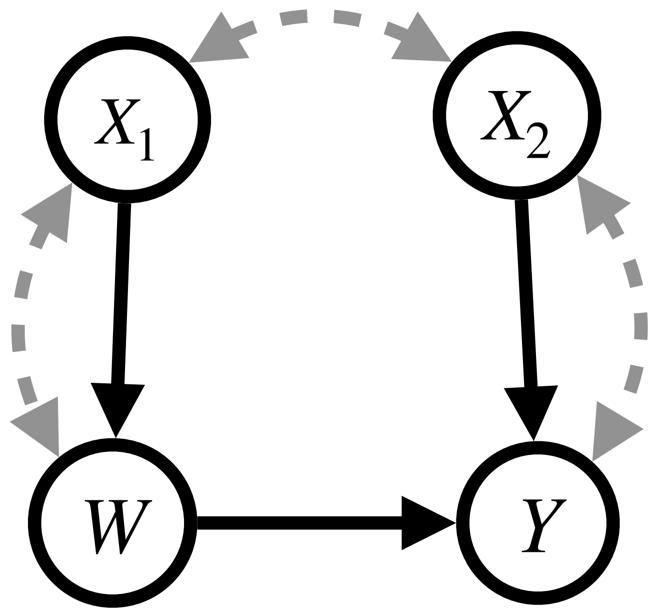


Dist. $\{P\}$

$\{P(\mathbf{V} \mid do(x_i)\}$
for $i = 1, \cdots, m$

Multiple Treatment Interaction

Graph G



Dist. $\{P\}$

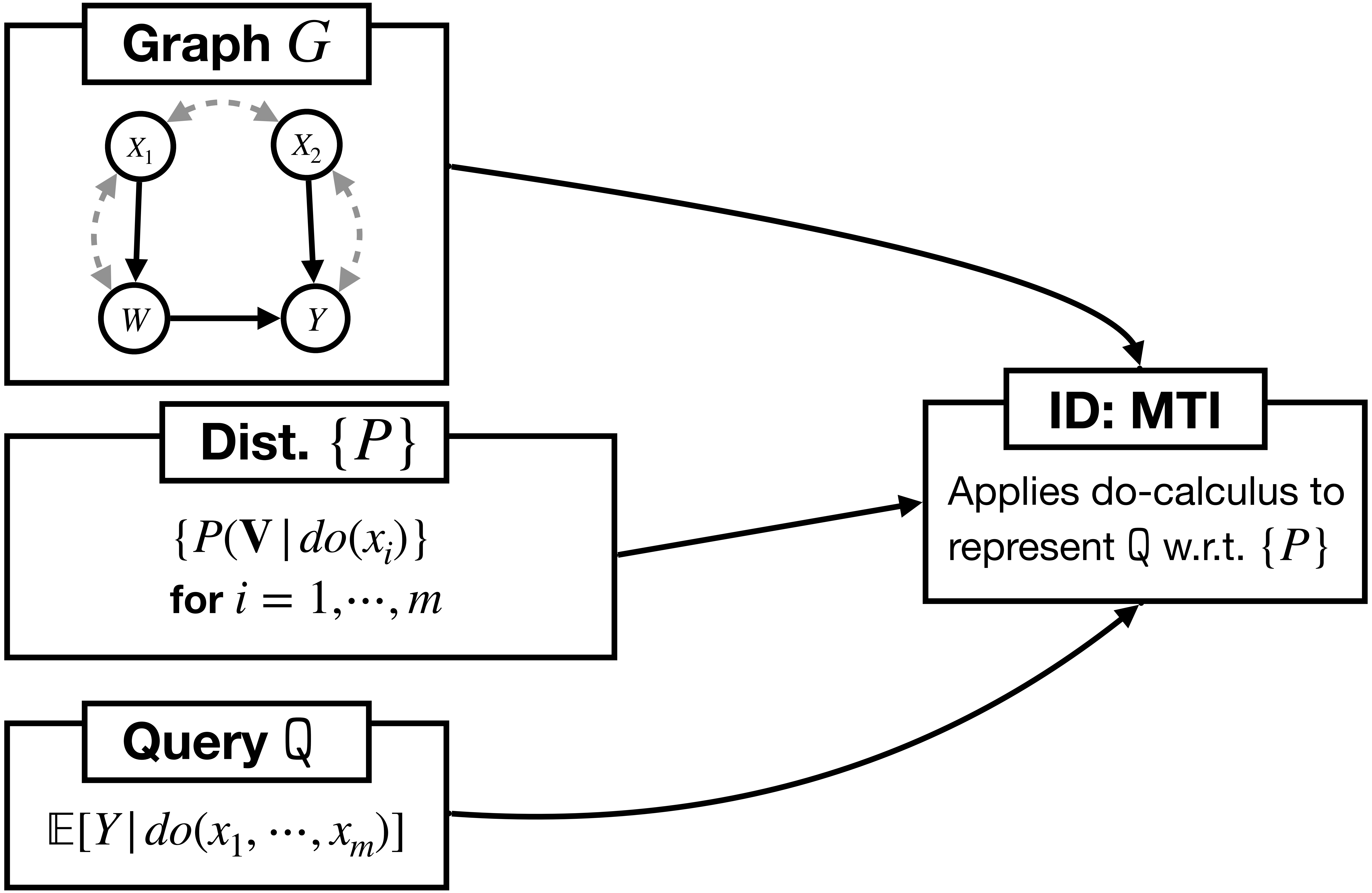
$\{P(\mathbf{V} \mid do(x_i)\}$

for $i = 1, \cdots, m$

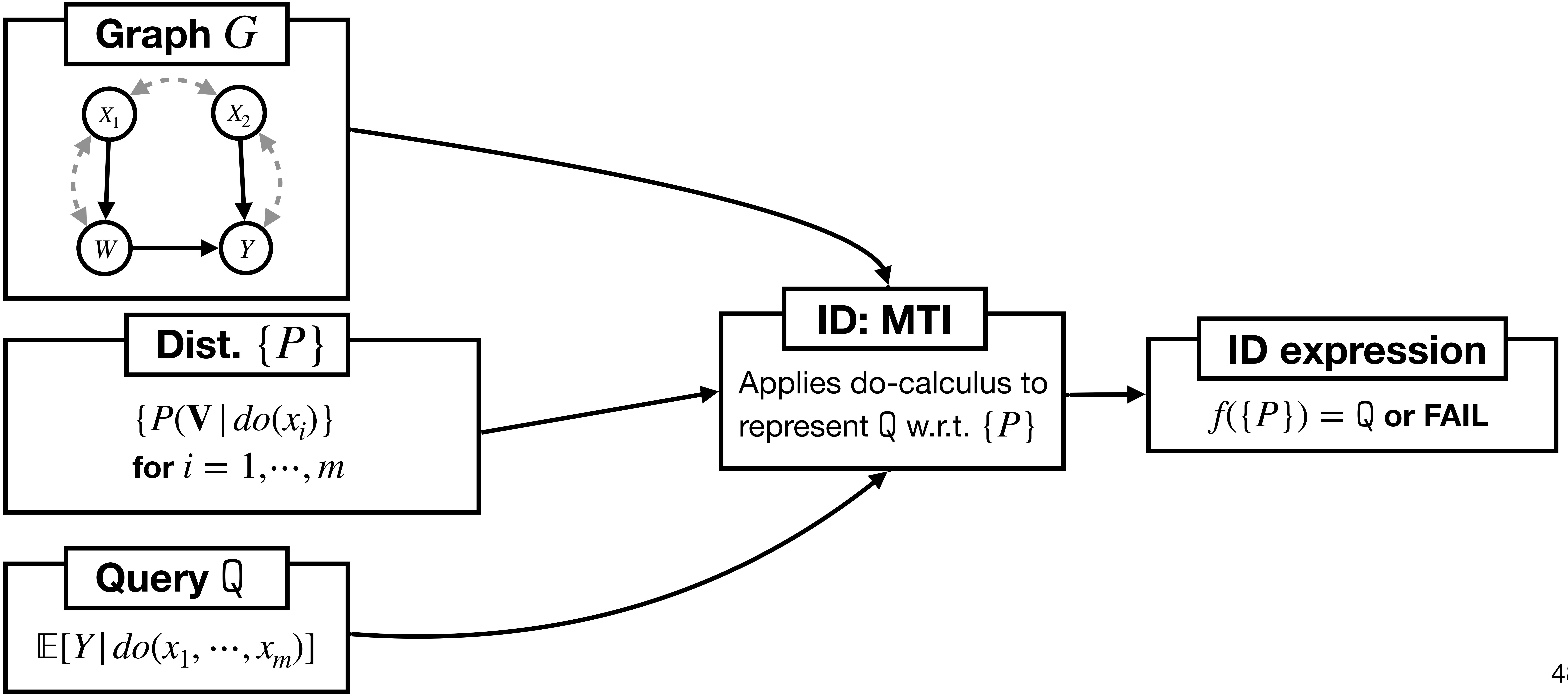
Query $\mathbb{Q}$

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

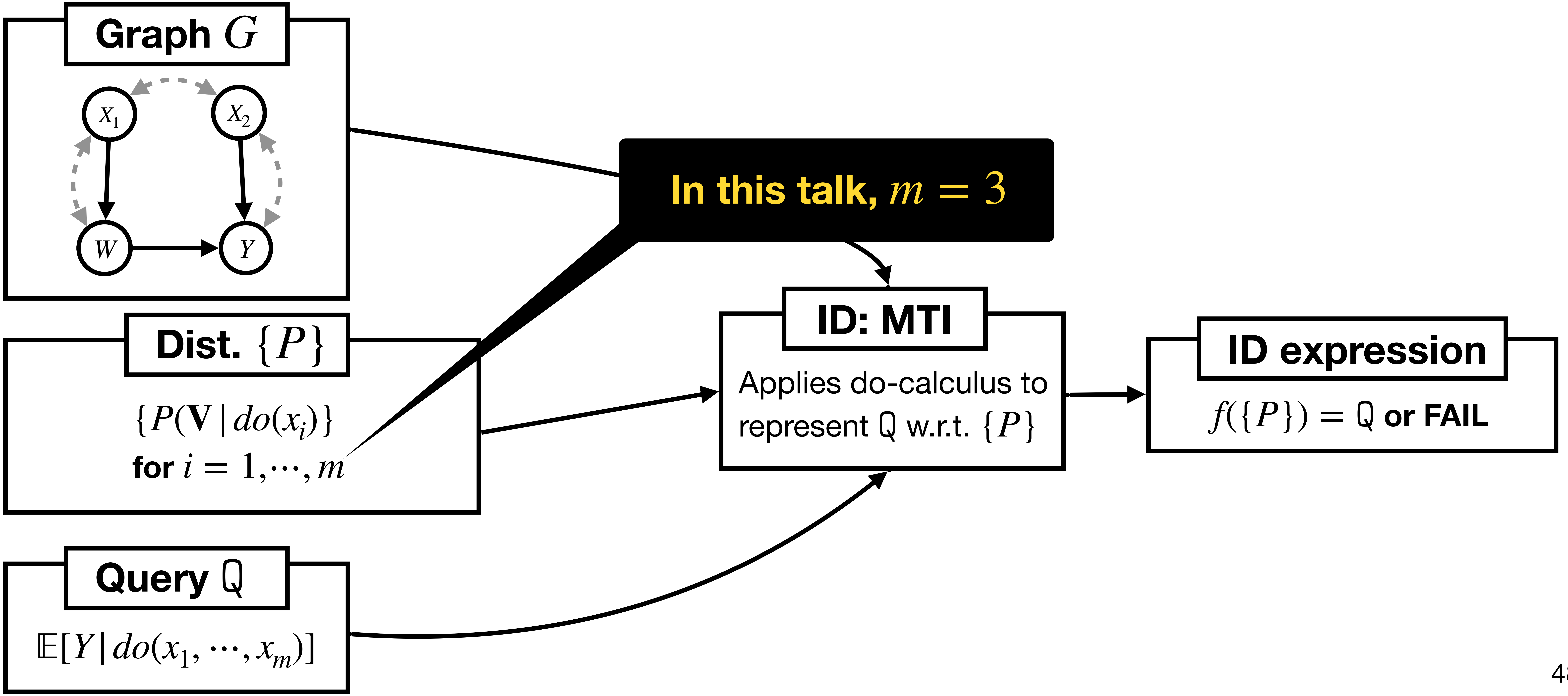
Multiple Treatment Interaction



Multiple Treatment Interaction



Multiple Treatment Interaction



Sufficient ID Condition: MTI

Sufficient ID Condition: MTI

Identification Condition for Multiple Treatment Interaction

Sufficient ID Condition: MTI

Identification Condition for Multiple Treatment Interaction

There is a set of variables W_1, W_2 s.t.

1. W_1 is a pretreatment variable of the treatments $\{X_2, X_3\}$; W_2 is a pretreatment variable of the second treatment X_3.

Sufficient ID Condition: MTI

Identification Condition for Multiple Treatment Interaction

There is a set of variables W_1, W_2 s.t.

1. W_1 is a pretreatment variable of the treatments $\{X_2, X_3\}$; W_2 is a pretreatment variable of the second treatment X_3.

2. There are no unmeasured confounders between (Y, X_1) given W_1 in the post-treatment distribution $X_2 = x_2, X_3 = x_3$.

Sufficient ID Condition: MTI

Identification Condition for Multiple Treatment Interaction

There is a set of variables W_1, W_2 s.t.

1. W_1 is a pretreatment variable of the treatments $\{X_2, X_3\}$; W_2 is a pretreatment variable of the second treatment X_3.

2. There are no unmeasured confounders between (Y, X_1) given W_1 in the post-treatment distribution $X_2 = x_2, X_3 = x_3$.

3. There are no unmeasured confounders between (Y, X_2) given W_2, X_1, W_1 in the post-treatment distribution $X_3 = x_3$.

Sufficient ID Condition: MTI

Sufficient ID Condition: MTI

Equivalent Condition for Multiple Treatment Interaction

Sufficient ID Condition: MTI

Equivalent Condition for Multiple Treatment Interaction

There is a set of variables W_1, W_2 s.t.

1. W_1 is a pretreatment variable of the treatments $\{X_2, X_3\}$; W_2 is a pretreatment variable of the second treatment X_3.
2. $(Y \perp\!\!\!\perp X_1 \mid W_1)_{G_{\underline{X_1}\overline{X_2, X_3}}}$; $(Y \perp\!\!\!\perp X_2 \mid W_2, W_1, X_1)_{G_{\underline{X_2}\overline{X_3}}}$

Multiple Treatment Interaction: Examples

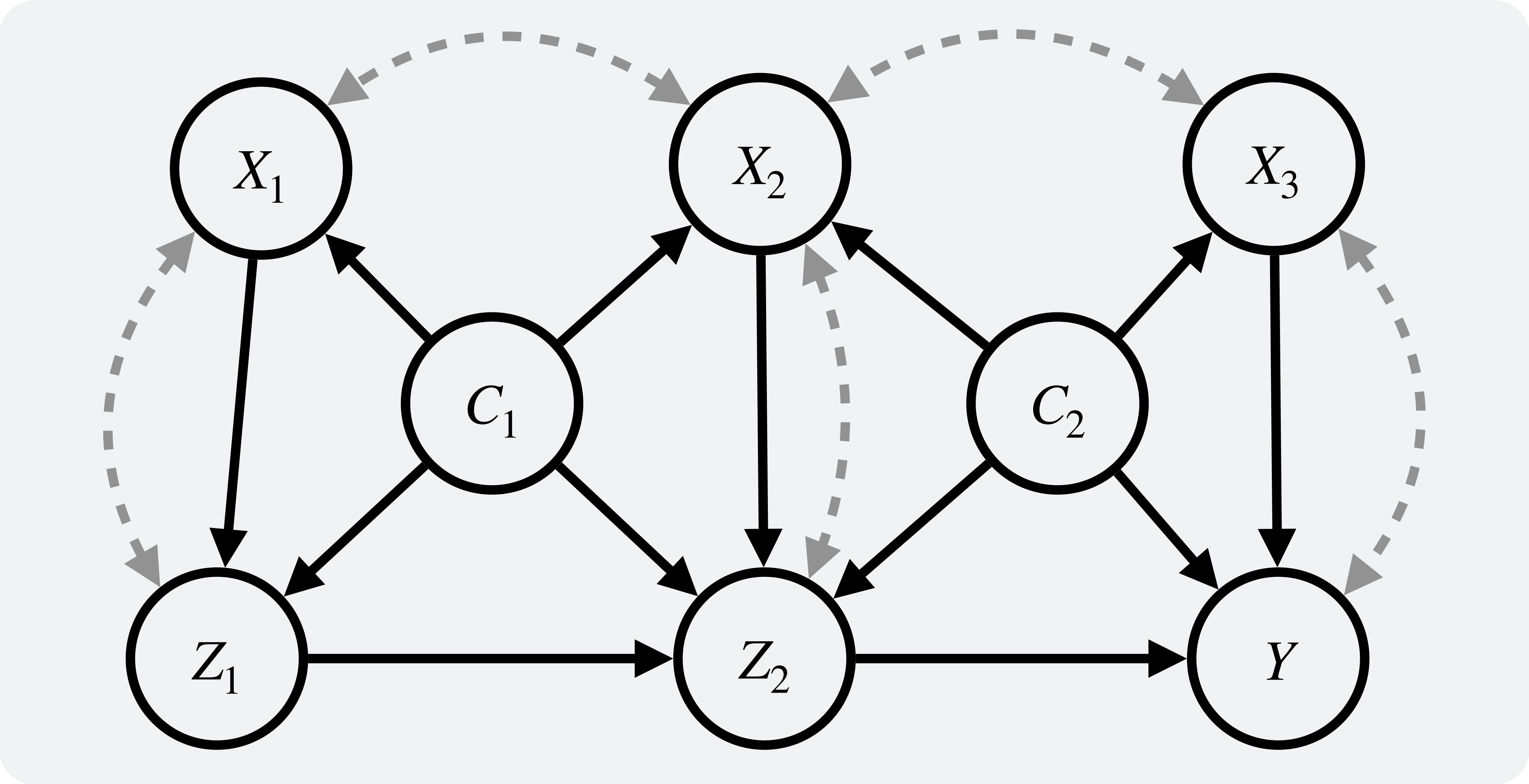
There is a set of variables W_1, W_2 s.t.

1. W_1 is a pretreatment variable of the treatments $\{X_2, X_3\}$; W_2 is a pretreatment variable of the second treatment X_3.
2. $(Y \perp\!\!\!\perp X_1 \mid W_1)_{G_{\underline{X_1}\overline{X_2, X_3}}}$; $(Y \perp\!\!\!\perp X_2 \mid W_2, W_1, X_1)_{G_{\underline{X_2}\overline{X_3}}}$

Multiple Treatment Interaction: Examples

There is a set of variables W_1, W_2 s.t.

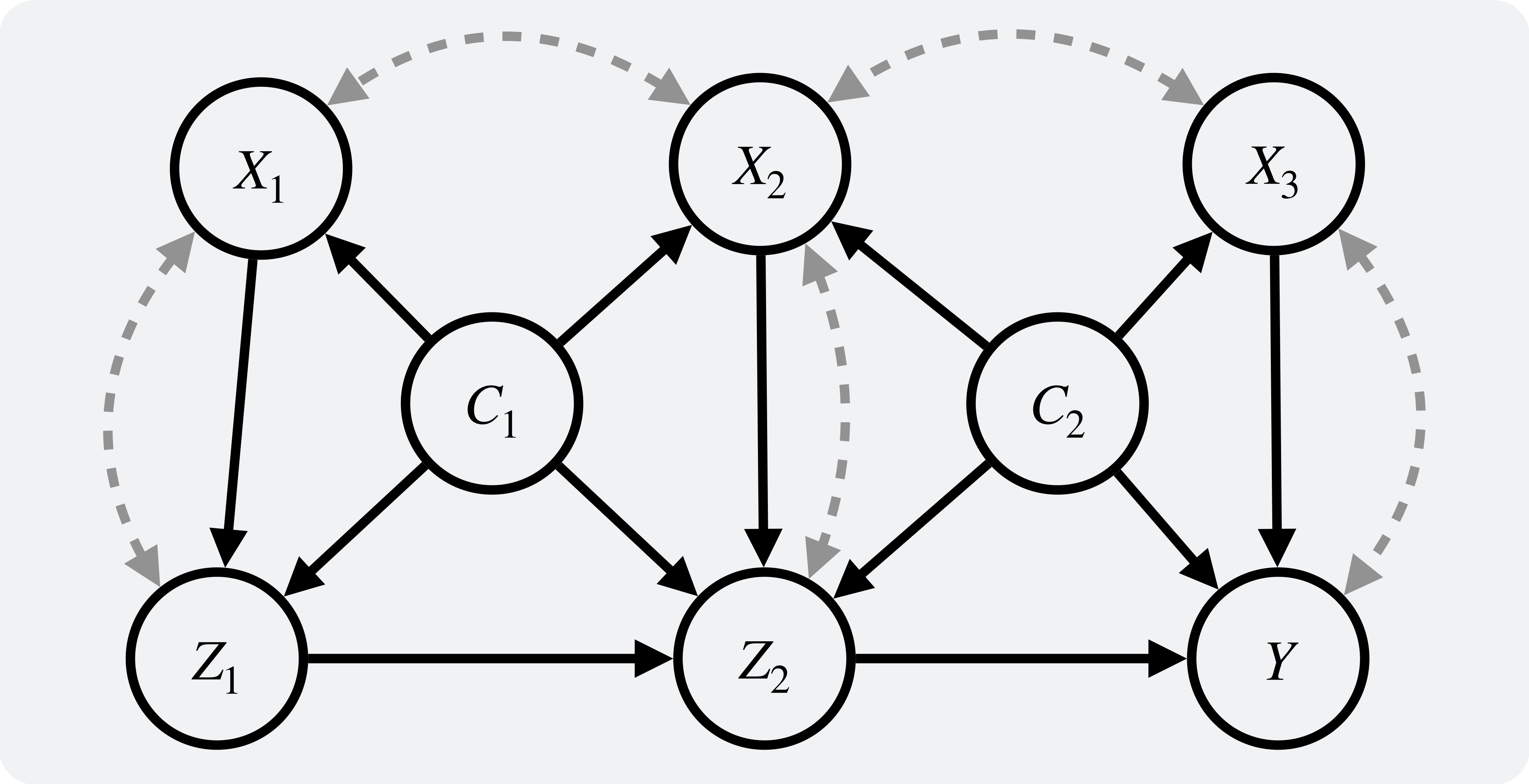
1. W_1 is a pretreatment variable of the treatments $\{X_2, X_3\}$; W_2 is a pretreatment variable of the second treatment X_3.
2. $(Y \perp\!\!\!\perp X_1 \mid W_1)_{G_{\underline{X_1}\overline{X_2, X_3}}}$; $(Y \perp\!\!\!\perp X_2 \mid W_2, W_1, X_1)_{G_{\underline{X_2}\overline{X_3}}}$



Multiple Treatment Interaction: Examples

There is a set of variables W_1, W_2 s.t.

1. W_1 is a pretreatment variable of the treatments $\{X_2, X_3\}$; W_2 is a pretreatment variable of the second treatment X_3.
2. $(Y \perp\!\!\!\perp X_1 \mid W_1)_{G_{\underline{X_1}\overline{X_2, X_3}}}$; $(Y \perp\!\!\!\perp X_2 \mid W_2, W_1, X_1)_{G_{\underline{X_2}\overline{X_3}}}$



$W_1 = \{Z_1, C_1\}, W_2 = \{Z_2, C_2\}$

Multiple-Treatment Interaction

Identification Condition for Multiple Treatment Interaction

1. W_1 is a pretreatment variable of the treatments $\{X_2, X_3\}$; W_2 is a pretreatment variable of the second treatment X_3.
2. $(Y \perp\!\!\!\perp X_1 \,|\, W_1)_{G_{\underline{X_1}\overline{X_2, X_3}}}$; $(Y \perp\!\!\!\perp X_2 \,|\, W_2, W_1, X_1)_{G_{\underline{X_2}\overline{X_3}}}$

Multiple-Treatment Interaction

Identification Condition for Multiple Treatment Interaction

1. W_1 is a pretreatment variable of the treatments $\{X_2, X_3\}$; W_2 is a pretreatment variable of the second treatment X_3.
2. $(Y \perp\!\!\!\perp X_1 \mid W_1)_{G_{\underline{X_1}\overline{X_2,X_3}}}$; $(Y \perp\!\!\!\perp X_2 \mid W_2, W_1, X_1)_{G_{\underline{X_2}\overline{X_3}}}$

Then, $\mathbb{E}[Y \mid do(x_1, x_2, x_3)]$ is identifiable from three experimental distributions $\{P(\cdot \mid do(x_1)), P(\cdot \mid do(x_2)), P(\cdot \mid do(x_3))\}$ and $P(\cdot \mid do(x_2))$ as follows:

$$\mathbb{E}[Y \mid do(\mathbf{x})] = \sum_{w_1, w_2 \in \mathscr{W}_1, \mathscr{W}_2} \mathbb{E}[Y \mid do(x_3), w_1, w_2, x_1, x_2] P(w_2 \mid do(x_2), w_1, x_1) P(w_1 \mid do(x_1))$$

Multiple-Treatment Interaction

Identification Condition for Multiple Treatment Interaction

1. W_1 is a pretreatment variable of the treatments $\{X_2, X_3\}$; W_2 is a pretreatment variable of the second treatment X_3.
2. $(Y \perp\!\!\!\perp X_1 \mid W_1)_{G_{\underline{X_1}\overline{X_2, X_3}}}$; $(Y \perp\!\!\!\perp X_2 \mid W_2, W_1, X_1)_{G_{\underline{X_2}\overline{X_3}}}$

Then, $\mathbb{E}[Y \mid do(x_1, x_2, x_3)]$ is identifiable from three experimental distributions $\{P(\cdot \mid do(x_1)), P(\cdot \mid do(x_2)), P(\cdot \mid do(x_3))\}$ and $P(\cdot \mid do(x_2))$ as follows:

$$\mathbb{E}[Y \mid do(\mathbf{x})] = \sum_{w_1, w_2 \in \mathscr{W}_1, \mathscr{W}_2} \mathbb{E}[Y \mid do(x_3), w_1, w_2, x_1, x_2] P(w_2 \mid do(x_2), w_1, x_1) P(w_1 \mid do(x_1))$$

Estimable from $D_3 \sim P(\mathbf{V} \mid do(x_3))$

Multiple-Treatment Interaction

Identification Condition for Multiple Treatment Interaction

1. W_1 is a pretreatment variable of the treatments $\{X_2, X_3\}$; W_2 is a pretreatment variable of the second treatment X_3.
2. $(Y \perp\!\!\!\perp X_1 \mid W_1)_{G_{\underline{X_1}\overline{X_2, X_3}}}$; $(Y \perp\!\!\!\perp X_2 \mid W_2, W_1, X_1)_{G_{\underline{X_2}\overline{X_3}}}$

Then, $\mathbb{E}[Y \mid do(x_1, x_2, x_3)]$ is identifiable from three experimental distributions $\{P(\cdot \mid do(x_1)), P(\cdot \mid do(x_2)), P(\cdot \mid do(x_3))\}$ and $P(\cdot \mid do(x_2))$ as follows:

$$\mathbb{E}[Y \mid do(\mathbf{x})] = \sum_{w_1, w_2 \in \mathscr{W}_1, \mathscr{W}_2} \mathbb{E}[Y \mid do(x_3), w_1, w_2, x_1, x_2] \boxed{P(w_2 \mid do(x_2), w_1, x_1)} P(w_1 \mid do(x_1))$$

Estimable from $D_2 \sim P(\mathbf{V} \mid do(x_2))$

Multiple-Treatment Interaction

Identification Condition for Multiple Treatment Interaction

1. W_1 is a pretreatment variable of the treatments $\{X_2, X_3\}$; W_2 is a pretreatment variable of the second treatment X_3.
2. $(Y \perp\!\!\!\perp X_1 \mid W_1)_{G_{\underline{X_1}\overline{X_2, X_3}}}$; $(Y \perp\!\!\!\perp X_2 \mid W_2, W_1, X_1)_{G_{\underline{X_2}\overline{X_3}}}$

Then, $\mathbb{E}[Y \mid do(x_1, x_2, x_3)]$ is identifiable from three experimental distributions $\{P(\cdot \mid do(x_1)), P(\cdot \mid do(x_2)), P(\cdot \mid do(x_3))\}$ and $P(\cdot \mid do(x_2))$ as follows:

$$\mathbb{E}[Y \mid do(\mathbf{x})] = \sum_{w_1, w_2 \in \mathscr{W}_1, \mathscr{W}_2} \mathbb{E}[Y \mid do(x_3), w_1, w_2, x_1, x_2] P(w_2 \mid do(x_2), w_1, x_1) P(w_1 \mid do(x_1))$$

Estimable from $D_1 \sim P(\mathbf{V} \mid do(x_1))$

Joint Treatment Effect Estimation : Regression

$$\mathbb{E}[Y \mid do(\mathbf{x})] = \sum_{w_1, w_2 \in \mathscr{W}_1, \mathscr{W}_2} \mathbb{E}[Y \mid do(x_3), w_1, w_2, x_1, x_2] P(w_2 \mid do(x_2), w_1, x_1) P(w_1 \mid do(x_1))$$

Joint Treatment Effect Estimation : Regression

$$\mathbb{E}[Y \mid do(\mathbf{x})] = \sum_{w_1, w_2 \in \mathscr{W}_1, \mathscr{W}_2} \mathbb{E}[Y \mid do(x_3), w_1, w_2, x_1, x_2] P(w_2 \mid do(x_2), w_1, x_1) P(w_1 \mid do(x_1))$$

Regression-based Representation

Joint Treatment Effect Estimation : Regression

$$\mathbb{E}[Y \mid do(\mathbf{x})] = \sum_{w_1, w_2 \in \mathcal{W}_1, \mathcal{W}_2} \mathbb{E}[Y \mid do(x_3), w_1, w_2, x_1, x_2] P(w_2 \mid do(x_2), w_1, x_1) P(w_1 \mid do(x_1))$$

Regression-based Representation

$$\mu_0^2(W_1, W_2, X_1, X_2) := \mathbb{E}[Y \mid do(x_3), W_1, W_2, X_1, X_2]$$

Joint Treatment Effect Estimation : Regression

$$\mathbb{E}[Y \mid do(\mathbf{x})] = \sum_{w_1, w_2 \in \mathscr{W}_1, \mathscr{W}_2} \mathbb{E}[Y \mid do(x_3), w_1, w_2, x_1, x_2] P(w_2 \mid do(x_2), w_1, x_1) P(w_1 \mid do(x_1))$$

Regression-based Representation

$$\mu_0^2(W_1, W_2, X_1, X_2) := \mathbb{E}[Y \mid do(x_3), W_1, W_2, X_1, X_2]$$

Estimable from $D_3 \sim P(\mathbf{V} \mid do(x_3))$

Joint Treatment Effect Estimation : Regression

$$\mathbb{E}[Y \mid do(\mathbf{x})] = \sum_{w_1, w_2 \in \mathscr{W}_1, \mathscr{W}_2} \mathbb{E}[Y \mid do(x_3), w_1, w_2, x_1, x_2] P(w_2 \mid do(x_2), w_1, x_1) P(w_1 \mid do(x_1))$$

Regression-based Representation

$$\mu_0^2(W_1, W_2, X_1, X_2) := \mathbb{E}[Y \mid do(x_3), W_1, W_2, X_1, X_2]$$

Estimable from $D_3 \sim P(\mathbf{V} \mid do(x_3))$

$$\mu_0^1(W_1, X_1) := \mathbb{E}[\mu_0^2(W_1, W_2, X_1, x_2) \mid do(x_2), W_1, X_1]$$

Joint Treatment Effect Estimation : Regression

$$\mathbb{E}[Y \mid do(\mathbf{x})] = \sum_{w_1, w_2 \in \mathscr{W}_1, \mathscr{W}_2} \mathbb{E}[Y \mid do(x_3), w_1, w_2, x_1, x_2] P(w_2 \mid do(x_2), w_1, x_1) P(w_1 \mid do(x_1))$$

Regression-based Representation

$\mu_0^2(W_1, W_2, X_1, X_2) := \mathbb{E}[Y \mid do(x_3), W_1, W_2, X_1, X_2]$ Estimable from $D_3 \sim P(\mathbf{V} \mid do(x_3))$

$\mu_0^1(W_1, X_1) := \mathbb{E}[\mu_0^2(W_1, W_2, X_1, x_2) \mid do(x_2), W_1, X_1]$ Estimable from $D_2 \sim P(\mathbf{V} \mid do(x_2))$

Joint Treatment Effect Estimation : Regression

$$\mathbb{E}[Y \mid do(\mathbf{x})] = \sum_{w_1, w_2 \in \mathscr{W}_1, \mathscr{W}_2} \mathbb{E}[Y \mid do(x_3), w_1, w_2, x_1, x_2] P(w_2 \mid do(x_2), w_1, x_1) P(w_1 \mid do(x_1))$$

Regression-based Representation

$$\mu_0^2(W_1, W_2, X_1, X_2) := \mathbb{E}[Y \mid do(x_3), W_1, W_2, X_1, X_2]$$ Estimable from $D_3 \sim P(\mathbf{V} \mid do(x_3))$

$$\mu_0^1(W_1, X_1) := \mathbb{E}[\mu_0^2(W_1, W_2, X_1, x_2) \mid do(x_2), W_1, X_1]$$ Estimable from $D_2 \sim P(\mathbf{V} \mid do(x_2))$

$$\mathbb{E}[Y \mid do(x_1, x_2, x_3)] = \mathbb{E}[\mu_0^1(W_1, x_1) \mid do(x_1)]$$

Joint Treatment Effect Estimation : Regression

$$\mathbb{E}[Y \mid do(\mathbf{x})] = \sum_{w_1, w_2 \in \mathscr{W}_1, \mathscr{W}_2} \mathbb{E}[Y \mid do(x_3), w_1, w_2, x_1, x_2] P(w_2 \mid do(x_2), w_1, x_1) P(w_1 \mid do(x_1))$$

Regression-based Representation

$$\mu_0^2(W_1, W_2, X_1, X_2) := \mathbb{E}[Y \mid do(x_3), W_1, W_2, X_1, X_2]$$ Estimable from $D_3 \sim P(\mathbf{V} \mid do(x_3))$

$$\mu_0^1(W_1, X_1) := \mathbb{E}[\mu_0^2(W_1, W_2, X_1, x_2) \mid do(x_2), W_1, X_1]$$ Estimable from $D_2 \sim P(\mathbf{V} \mid do(x_2))$

$$\mathbb{E}[Y \mid do(x_1, x_2, x_3)] = \mathbb{E}[\mu_0^1(W_1, x_1) \mid do(x_1)]$$ Estimable from $D_1 \sim P(\mathbf{V} \mid do(x_1))$

Joint Treatment Effect Estimation : Probability Weighting

$$\mathbb{E}[Y|do(\mathbf{x})] = \sum_{w_1, w_2 \in \mathcal{W}_1, \mathcal{W}_2} \mathbb{E}[Y|do(x_3), w_1, w_2, x_1, x_2] P(w_2|do(x_2), w_1, x_1) P(w_1|do(x_1))$$

Joint Treatment Effect Estimation : Probability Weighting

$$\mathbb{E}[Y \mid do(\mathbf{x})] = \sum_{w_1, w_2 \in \mathscr{W}_1, \mathscr{W}_2} \mathbb{E}[Y \mid do(x_3), w_1, w_2, x_1, x_2] P(w_2 \mid do(x_2), w_1, x_1) P(w_1 \mid do(x_1))$$

Probability Weighting-based Representation

Joint Treatment Effect Estimation : Probability Weighting

$$\mathbb{E}[Y \mid do(\mathbf{x})] = \sum_{w_1, w_2 \in \mathcal{W}_1, \mathcal{W}_2} \mathbb{E}[Y \mid do(x_3), w_1, w_2, x_1, x_2] P(w_2 \mid do(x_2), w_1, x_1) P(w_1 \mid do(x_1))$$

Probability Weighting-based Representation

$$\pi_0^2(W_1, W_2, X_1, X_2) := \frac{P(W_1, W_2, X_1 \mid do(x_2))}{P(W_1, W_2, X_1, X_2 \mid do(x_3))}$$

Joint Treatment Effect Estimation : Probability Weighting

$$\mathbb{E}[Y \mid do(\mathbf{x})] = \sum_{w_1, w_2 \in \mathscr{W}_1, \mathscr{W}_2} \mathbb{E}[Y \mid do(x_3), w_1, w_2, x_1, x_2] P(w_2 \mid do(x_2), w_1, x_1) P(w_1 \mid do(x_1))$$

Probability Weighting-based Representation

$$\pi_0^2(W_1, W_2, X_1, X_2) := \frac{P(W_1, W_2, X_1 \mid do(x_2))}{P(W_1, W_2, X_1, X_2 \mid do(x_3))}$$

$$\pi_0^1(W_1, X_1) := \frac{P(W_1 \mid do(x_1))}{P(W_1, X_1 \mid do(x_2))}$$

Joint Treatment Effect Estimation : Probability Weighting

$$\mathbb{E}[Y|do(\mathbf{x})] = \sum_{w_1,w_2 \in \mathscr{W}_1,\mathscr{W}_2} \mathbb{E}[Y|do(x_3), w_1, w_2, x_1, x_2]P(w_2|do(x_2), w_1, x_1)P(w_1|do(x_1))$$

Probability Weighting-based Representation

$$\pi_0^2(W_1, W_2, X_1, X_2) := \frac{P(W_1, W_2, X_1 | do(x_2))}{P(W_1, W_2, X_1, X_2 | do(x_3))}$$

$$\pi_0^1(W_1, X_1) := \frac{P(W_1 | do(x_1))}{P(W_1, X_1 | do(x_2))}$$

$$\mathbb{E}[Y|do(x_1, x_2, x_3)] = \mathbb{E}[\pi_0^1 \pi_0^2 I_{x_1,x_2}(X_1, X_2) Y | do(x_3)]$$

Joint Treatment Effect Estimation : Probability Weighting

$$\mathbb{E}[Y \mid do(\mathbf{x})] = \sum_{w_1, w_2 \in \mathscr{W}_1, \mathscr{W}_2} \mathbb{E}[Y \mid do(x_3), w_1, w_2, x_1, x_2] P(w_2 \mid do(x_2), w_1, x_1) P(w_1 \mid do(x_1))$$

Probability Weighting-based Representation

$$\pi_0^2(W_1, W_2, X_1, X_2) := \frac{P(W_1, W_2, X_1 \mid do(x_2))}{P(W_1, W_2, X_1, X_2 \mid do(x_3))}$$

$$\pi_0^1(W_1, X_1) := \frac{P(W_1 \mid do(x_1))}{P(W_1, X_1 \mid do(x_2))}$$

$$\mathbb{E}[Y \mid do(x_1, x_2, x_3)] = \mathbb{E}[\pi_0^1 \pi_0^2 I_{x_1, x_2}(X_1, X_2) Y \mid do(x_3)]$$

Estimable from $D_3 \sim P(\mathbf{V} \mid do(x_3))$

Joint Treatment Effect Estimation : Probability Weighting

$$\mathbb{E}[Y \mid do(\mathbf{x})] = \sum_{w_1, w_2 \in \mathscr{W}_1, \mathscr{W}_2} \mathbb{E}[Y \mid do(x_3), w_1, w_2, x_1, x_2] P(w_2 \mid do(x_2), w_1, x_1) P(w_1 \mid do(x_1))$$

Probability Weighting-based Representation

$$\pi_0^2(W_1, W_2, X_1, X_2) := \frac{P(W_1, W_2, X_1 \mid do(x_2))}{P(W_1, W_2, X_1, X_2 \mid do(x_3))}$$

$$\pi_0^1(W_1, X_1) := \frac{P(W_1 \mid do(x_1))}{P(W_1, X_1 \mid do(x_2))}$$

Estimable from $D_2 \sim P(\mathbf{V} \mid do(x_2))$

$$\mathbb{E}[Y \mid do(x_1, x_2, x_3)] = \mathbb{E}[\pi_0^1 \pi_0^2 I_{x_1, x_2}(X_1, X_2) Y \mid do(x_3)]$$

Joint Treatment Effect Estimation : Probability Weighting

$$\mathbb{E}[Y \mid do(\mathbf{x})] = \sum_{w_1, w_2 \in \mathscr{W}_1, \mathscr{W}_2} \mathbb{E}[Y \mid do(x_3), w_1, w_2, x_1, x_2] P(w_2 \mid do(x_2), w_1, x_1) P(w_1 \mid do(x_1))$$

Probability Weighting-based Representation

$$\pi_0^2(W_1, W_2, X_1, X_2) := \frac{P(W_1, W_2, X_1 \mid do(x_2))}{P(W_1, W_2, X_1, X_2 \mid do(x_3))}$$

$$\pi_0^1(W_1, X_1) := \frac{P(W_1 \mid do(x_1))}{P(W_1, X_1 \mid do(x_2))}$$

Estimable from $D_1 \sim P(\mathbf{V} \mid do(x_1))$

$$\mathbb{E}[Y \mid do(x_1, x_2, x_3)] = \mathbb{E}[\pi_0^1 \pi_0^2 I_{x_1, x_2}(X_1, X_2) Y \mid do(x_3)]$$

Joint Treatment Effect Estimation : DR Estimand

Joint Treatment Effect Estimation : DR Estimand

DR Representation

Joint Treatment Effect Estimation : DR Estimand

DR Representation

$$\mathbb{E}[Y \,|\, do(\mathbf{x})] = \mathbb{E}[\pi_0^2 \pi_0^1 I_{x_1}(X_1) I_{x_2}(X_2)(Y - \mu_0^2(\mathbf{W}, \mathbf{X})) \,|\, do(x_3)]$$

$$+\mathbb{E}[\pi_0^1 I_{x_1}(X_1)(\mu_0^2(\mathbf{W}, X_1, x_2) - \mu_0^1(W_1, X_1)) \,|\, do(x_2)]$$

$$+\mathbb{E}[\mu_0^1(W_1, x_1) \,|\, do(x_1))]$$

Joint Treatment Effect Estimation : DR Estimand

DR Representation

$$\mathbb{E}[Y \mid do(\mathbf{x})] = \mathbb{E}[\pi_0^2 \pi_0^1 I_{x_1}(X_1) I_{x_2}(X_2)(Y - \mu_0^2(\mathbf{W}, \mathbf{X})) \mid do(x_3)]$$

$$+\mathbb{E}[\pi_0^1 I_{x_1}(X_1)(\mu_0^2(\mathbf{W}, X_1, x_2) - \mu_0^1(W_1, X_1)) \mid do(x_2)]$$

$$+\mathbb{E}[\mu_0^1(W_1, x_1) \mid do(x_1))]$$

$$\mathbf{X} := (X_1, X_2)$$
$$\mathbf{W} := (W_1, W_2)$$

Asymptotic Error Analysis

Asymptotic Error Analysis

Asymptotic Error Analysis

Asymptotic Error Analysis

Asymptotic Error Analysis

Assume

1. $\|\pi_k^2 - \pi_0^2\|_{P_3}\|\mu^2 - \mu_0^2\|_{P_3} = o_{P_3}(1)$ $\|\pi_k^1 - \pi_0^1\|_{P_2}\|\mu^1 - \mu_0^1\|_{P_2} = o_{P_2}(1), k \in \{a, b\}$; i.e., $\pi_k^2, \mu_k^2, \pi_k^1, \mu_k^1$ converges to the true parameters.

Asymptotic Error Analysis

Asymptotic Error Analysis

Assume

1. $\|\pi_k^2 - \pi_0^2\|_{P_3}\|\mu^2 - \mu_0^2\|_{P_3} = o_{P_3}(1)$ $\|\pi_k^1 - \pi_0^1\|_{P_2}\|\mu^1 - \mu_0^1\|_{P_2} = o_{P_2}(1), k \in \{a, b\}$; i.e., $\pi_k^2, \mu_k^2, \pi_k^1, \mu_k^1$ converges to the true parameters.

2. $\|\pi_k^1\pi_k^2(Y - \mu_k^2) - \pi_0^1\pi_0^2(Y - \mu_0^2)\|_{P_3} = O_{P_3}(1)$; i.e., $\pi_k^1\pi_k^2(Y - \mu_k^2) - \pi_0^1\pi_0^2(Y - \mu_0^2)$ is bounded.

Asymptotic Error Analysis

Asymptotic Error Analysis

Assume

1. $\|\pi_k^2 - \pi_0^2\|_{P_3}\|\mu^2 - \mu_0^2\|_{P_3} = o_{P_3}(1)$ $\|\pi_k^1 - \pi_0^1\|_{P_2}\|\mu^1 - \mu_0^1\|_{P_2} = o_{P_2}(1), k \in \{a, b\}$; i.e., $\pi_k^2, \mu_k^2, \pi_k^1, \mu_k^1$ converges to the true parameters.

2. $\|\pi_k^1\pi_k^2(Y - \mu_k^2) - \pi_0^1\pi_0^2(Y - \mu_0^2)\|_{P_3} = O_{P_3}(1)$; i.e., $\pi_k^1\pi_k^2(Y - \mu_k^2) - \pi_0^1\pi_0^2(Y - \mu_0^2)$ is bounded.

3. $\|\pi_k^1(\mu_k^2 - \mu_k^1) - \pi_0^1(\mu_k^2 - \mu_0^1)\|_{P_2} = O_{P_2}(1)$; i.e., $\pi_k^1(\mu_k^2 - \mu_k^1) - \pi_0^1(\mu_k^2 - \mu_0^1)$ is bounded.

Asymptotic Error Analysis

Asymptotic Error Analysis

Asymptotic Error Analysis

Asymptotic Error Analysis

$$T^{dml} - \mathbb{E}[Y \mid do(\mathbf{x})] = (R_1 + R_2 + R_3)$$

$$+ \sum_{k \in \{a,b\}} O_{P_3}\left(\|\pi_k^2 - \pi_0^2\|_{P_3} \|\mu_k^2 - \mu_0^2\|_{P_3} \right)$$

$$+ \sum_{k \in \{a,b\}} O_{P_2}\left(\|\pi_k^1 - \pi_0^1\|_{P_2} \|\mu_k^1 - \mu_0^1\|_{P_2} \right)$$

Asymptotic Error Analysis

Asymptotic Error Analysis

$$
\begin{aligned}
T^{dml} - \mathbb{E}[Y \mid do(\mathbf{x})] = (R_1 + R_2 + R_3) \\
+ \sum_{k \in \{a,b\}} O_{P_3}\left(\|\pi_k^2 - \pi_0^2\|_{P_3} \|\mu_k^2 - \mu_0^2\|_{P_3} \right) \\
+ \sum_{k \in \{a,b\}} O_{P_2}\left(\|\pi_k^1 - \pi_0^1\|_{P_2} \|\mu_k^1 - \mu_0^1\|_{P_2} \right)
\end{aligned}
$$

where R_i (for $i \in \{1,2,3\}$) is a random variable s.t. $\sqrt{n_i} R_i \xrightarrow{d} Z_i \sim \text{normal}(0, \sigma_i^2)$, where $\sigma_1^2 := \mathbb{V}_{P_1}[\mu_0^2]$, $\sigma_2^2 := \mathbb{V}_{P_2}[\pi_0^1(\mu^2 - \mu_0^1)]$, and $\sigma_3^2 := \mathbb{V}_{P_2}[\pi_0^1 \pi_0^2 (Y - \mu_0^2)]$.

Robustness of DML estimator

$$T^{dml} - \mathbb{E}[Y \mid do(\mathbf{x})] = (R_1 + R_2 + R_3) \; + \sum_{k \in \{a,b\}} O_{P_3}\left(\|\pi_k^2 - \pi_0^2\|_{P_3} \|\mu_k^2 - \mu_0^2\|_{P_3} \right) + \sum_{k \in \{a,b\}} O_{P_2}\left(\|\pi_k^1 - \pi_0^1\|_{P_2} \|\mu_k^1 - \mu_0^1\|_{P_2} \right)$$

Robustness of DML estimator

$$T^{dml} - \mathbb{E}[Y \mid do(\mathbf{x})] = (R_1 + R_2 + R_3) \ + \sum_{k \in \{a,b\}} O_{P_3}\left(\|\pi_k^2 - \pi_0^2\|_{P_3} \|\mu_k^2 - \mu_0^2\|_{P_3} \right) + \sum_{k \in \{a,b\}} O_{P_2}\left(\|\pi_k^1 - \pi_0^1\|_{P_2} \|\mu_k^1 - \mu_0^1\|_{P_2} \right)$$

Doubly Robustness & Debiasedness

Robustness of DML estimator

$$T^{dml} - \mathbb{E}[Y \mid do(\mathbf{x})] = (R_1 + R_2 + R_3) + \sum_{k \in \{a,b\}} O_{P_3}\left(\|\pi_k^2 - \pi_0^2\|_{P_3} \|\mu_k^2 - \mu_0^2\|_{P_3} \right) + \sum_{k \in \{a,b\}} O_{P_2}\left(\|\pi_k^1 - \pi_0^1\|_{P_2} \|\mu_k^1 - \mu_0^1\|_{P_2} \right)$$

Doubly Robustness & Debiasedness

1. **Doubly Robustness (DR):** T^{dml} converges to $\mathbb{E}[Y \mid do(\mathbf{x})]$ at $n^{-1/2}$-rate (where $n := \min(n_1, n_2, n_3)$) if $\pi_k^1 = \pi_0^1$ or $\mu_k^1 = \mu_0^1$, and $\pi_k^2 = \pi_0^2$ or $\mu_k^2 = \mu_0^2$.

Robustness of DML estimator

$$T^{dml} - \mathbb{E}[Y \mid do(\mathbf{x})] = (R_1 + R_2 + R_3) \ + \sum_{k \in \{a,b\}} O_{P_3}\left(\|\pi_k^2 - \pi_0^2\|_{P_3} \|\mu_k^2 - \mu_0^2\|_{P_3} \right) + \sum_{k \in \{a,b\}} O_{P_2}\left(\|\pi_k^1 - \pi_0^1\|_{P_2} \|\mu_k^1 - \mu_0^1\|_{P_2} \right)$$

Doubly Robustness & Debiasedness

1. **Doubly Robustness (DR):** T^{dml} converges to $\mathbb{E}[Y \mid do(\mathbf{x})]$ at $n^{-1/2}$-rate (where $n := \min(n_1, n_2, n_3)$) if $\pi_k^1 = \pi_0^1$ or $\mu_k^1 = \mu_0^1$, and $\pi_k^2 = \pi_0^2$ or $\mu_k^2 = \mu_0^2$.

2. **Debiasedness (DB):** T^{dml} converges to $\mathbb{E}[Y \mid do(x_1, x_2)]$ at $n^{-1/2}$-rate if $\pi^2, \pi^1, \mu^2, \mu^1$ converges at least at $n^{-1/4}$ rate.

Efficiency of DML estimator

$$T^{dml} - \mathbb{E}[Y \mid do(\mathbf{x})] = (R_1 + R_2 + R_3) \ + \sum_{k \in \{a,b\}} O_{P_3}\left(\|\pi_k^2 - \pi_0^2\|_{P_3} \|\mu_k^2 - \mu_0^2\|_{P_3} \right) + \sum_{k \in \{a,b\}} O_{P_2}\left(\|\pi_k^1 - \pi_0^1\|_{P_2} \|\mu_k^1 - \mu_0^1\|_{P_2} \right)$$

Efficiency of DML estimator

$$T^{dml} - \mathbb{E}[Y \mid do(\mathbf{x})] = (R_1 + R_2 + R_3) + \sum_{k \in \{a,b\}} O_{P_3}\left(\|\pi_k^2 - \pi_0^2\|_{P_3} \|\mu_k^2 - \mu_0^2\|_{P_3} \right) + \sum_{k \in \{a,b\}} O_{P_2}\left(\|\pi_k^1 - \pi_0^1\|_{P_2} \|\mu_k^1 - \mu_0^1\|_{P_2} \right)$$

CAN and Efficiency

Efficiency of DML estimator

$$T^{dml} - \mathbb{E}[Y \mid do(\mathbf{x})] = (R_1 + R_2 + R_3) + \sum_{k \in \{a,b\}} O_{P_3}\left(\|\pi_k^2 - \pi_0^2\|_{P_3} \|\mu_k^2 - \mu_0^2\|_{P_3} \right) + \sum_{k \in \{a,b\}} O_{P_2}\left(\|\pi_k^1 - \pi_0^1\|_{P_2} \|\mu_k^1 - \mu_0^1\|_{P_2} \right)$$

CAN and Efficiency

1. **Consistency and Asymptotic Normality (CAN):** Under the {DR,DB} conditions, T^{dml} achieves consistency and asymptotic normality (CAN).

Efficiency of DML estimator

$$T^{dml} - \mathbb{E}[Y \mid do(\mathbf{x})] = (R_1 + R_2 + R_3) + \sum_{k \in \{a,b\}} O_{P_3}\left(\|\pi_k^2 - \pi_0^2\|_{P_3} \|\mu_k^2 - \mu_0^2\|_{P_3} \right) + \sum_{k \in \{a,b\}} O_{P_2}\left(\|\pi_k^1 - \pi_0^1\|_{P_2} \|\mu_k^1 - \mu_0^1\|_{P_2} \right)$$

CAN and Efficiency

1. **Consistency and Asymptotic Normality (CAN):** Under the {DR,DB} conditions, T^{dml} achieves consistency and asymptotic normality (CAN).

2. **Statistical Efficiency:** Under the {DR,DB} conditions, T^{dml} achieves the nonparametric efficiency bound (i.e., achieves the minimum variance).

Omitted Results

Omitted Results

Question: When input distributions are arbitrary interventional/observational distributions?

Omitted Results

Question: When input distributions are arbitrary interventional/observational distributions?

$$\mathbb{P} := \{P(V \mid do(z_1)), \cdots, P(V \mid do(z_p))\}$$

$$\mathtt{Q} := \mathbb{E}[Y \mid do(x_1, \cdots, x_m)]$$

Omitted Results

Question: When input distributions are arbitrary interventional/observational distributions?

$$\mathbb{P} := \{P(V \mid do(z_1)), \cdots, P(V \mid do(z_p))\}$$

$$\mathsf{Q} := \mathbb{E}[Y \mid do(x_1, \cdots, x_m)]$$

* Z_i can be any set of variables including the empty set.
* Generalization of the case of $p = m$ and $Z_1 = X_1, \cdots, Z_p = X_m$.

Outline of this talk

1. Preliminary and Problem Setup
   (1) Structural Causal Model
2. Treatment-Treatment Interaction
   (1) Identification
   (2) Estimand
   (3) Estimation and Error Analysis
   (4) Simulation Results
3. Multiple-Treatment Interaction (+ Omitted Results)
4. **Future directions & Summary**

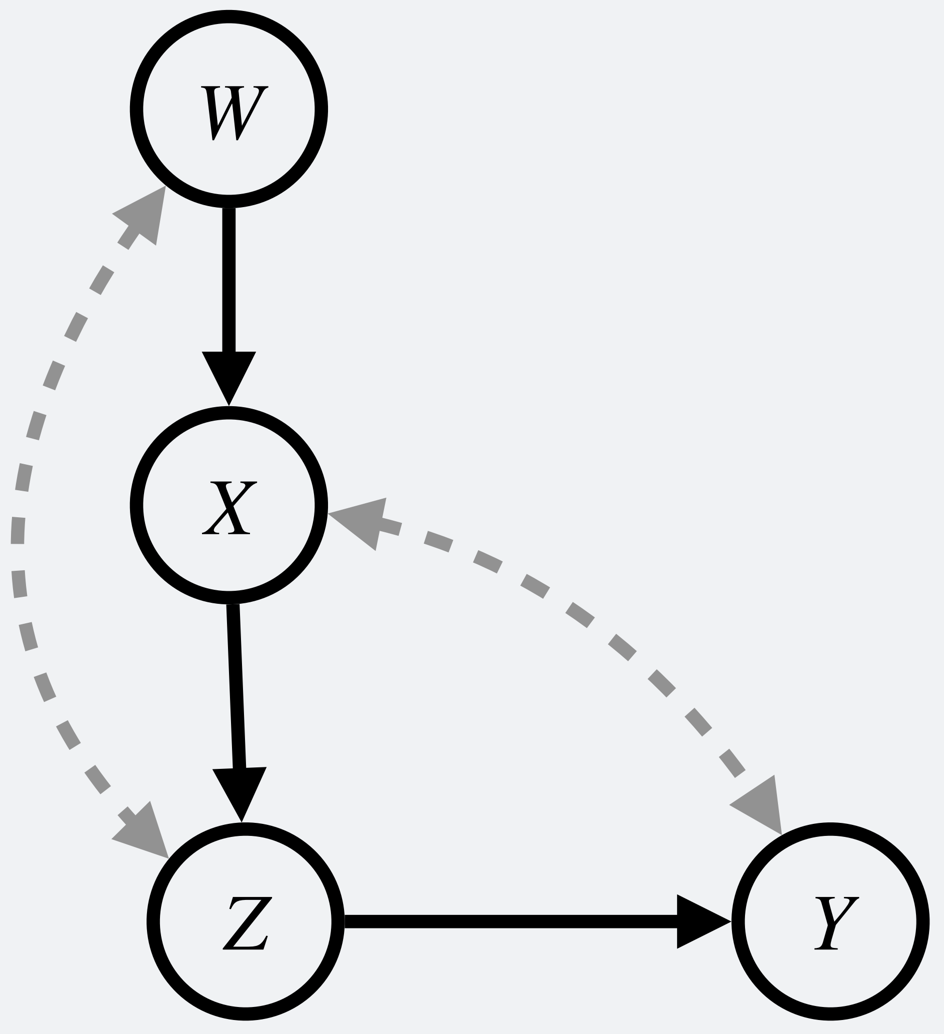
Future Research Direction 1

Future Research Direction 1

Scenario: When identification expression is not a covariate adjustment…

Future Research Direction 1

Scenario: When identification expression is not a covariate adjustment…



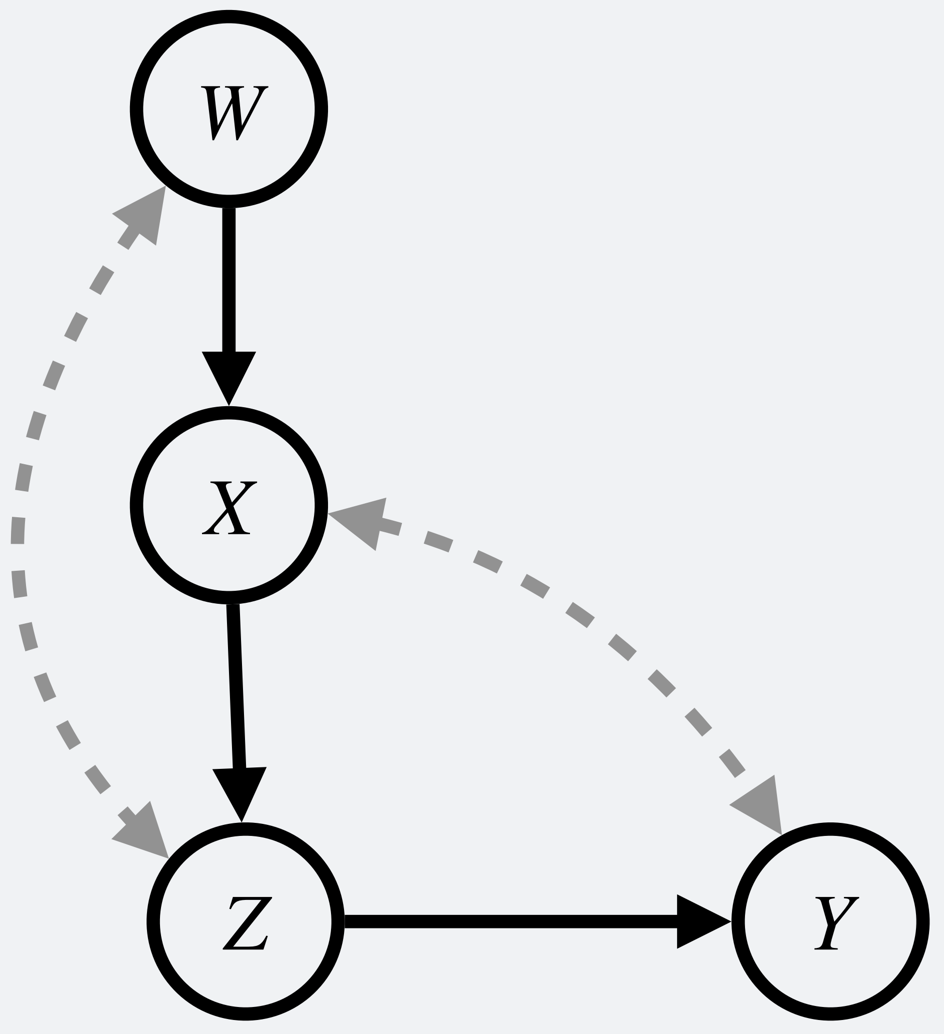
$$\mathbb{P} := \{P(V \mid do(z), P(V)\}$$

$$\mathtt{Q} := \mathbb{E}[Y \mid do(x)]$$

$$\mathtt{id} := \sum_{z \in \mathscr{Z}} \mathbb{E}[Y \mid do(z)] \sum_{w \in \mathscr{W}} P(z \mid x, w) P(w)$$

Future Research Direction 1

Scenario: When identification expression is not a covariate adjustment…



$$\mathbb{P} := \{P(V \mid do(z), P(V)\}$$

$$\mathtt{Q} := \mathbb{E}[Y \mid do(x)]$$

$$\mathtt{id} := \sum_{z \in \mathscr{Z}} \mathbb{E}[Y \mid do(z)] \sum_{w \in \mathscr{W}} P(z \mid x, w) P(w)$$

Estimating Causal Effects Identifiable from a Combination of Observations and Experiments, NeurIPS-23

Future Research Direction 2

Future Research Direction 2

Scenario: When input (source) distributions are different from the target distribution?

Future Research Direction 2

Scenario: When input (source) distributions are different from the target distribution?

$$\mathbb{P} := \{P(V \mid do(z_1), S = 1), \cdots, P(V \mid do(z_p), S = p)\}$$

$$\mathbb{Q} := \mathbb{E}[Y \mid do(x_1, \cdots, x_m), S = 0]$$

Summary

Summary

Question: Can we estimate the joint treatment effect from marginal experiments?

Summary

Question: Can we estimate the joint treatment effect from marginal experiments?

Answer: In general, not identifiable… 😞

Summary

Question: Can we estimate the joint treatment effect from marginal experiments?

Answer: In general, not identifiable… 😞

In this study,

Summary

Question: Can we estimate the joint treatment effect from marginal experiments?

Answer: In general, not identifiable… 😞

In this study,

1. Sufficient graphical criterion for identifying joint treatment effect $\mathbb{E}[Y \mid do(x_1, \cdots, x_m)]$ from marginal distributions $\mathbb{P} := \{P(V \mid do(x_1)), \cdots, P(V \mid do(x_m))\}$.

Summary

Question: Can we estimate the joint treatment effect from marginal experiments?

Answer: In general, not identifiable… 😞

In this study,

1. Sufficient graphical criterion for identifying joint treatment effect $\mathbb{E}[Y \mid do(x_1, \cdots, x_m)]$ from marginal distributions $\mathbb{P} := \{P(V \mid do(x_1)), \cdots, P(V \mid do(x_m))\}$.
2. Double/debiased ML-based estimator, which exhibits fast convergence, doubly robustness, and efficiency.