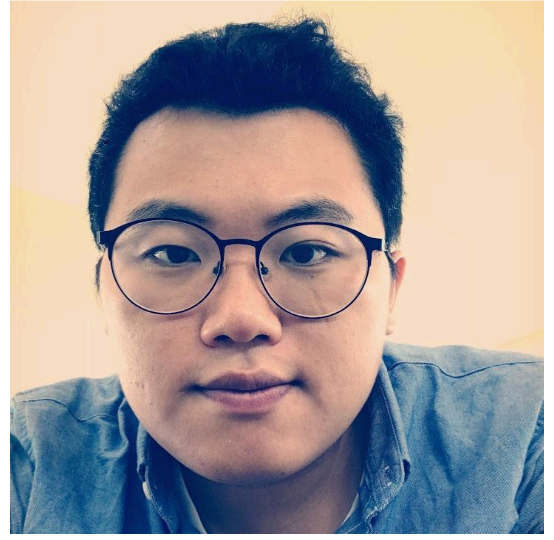
NeurIPS-20-12276

Learning Causal Effects via Weighted Empirical Risk Minimization

Yonghan Jung
PURDUE UNIVERSITY

Jin Tian
IOWA STATE UNIVERSITY

Elias Bareinboim
COLUMBIA UNIVERSITY IN THE CITY OF NEW YORK

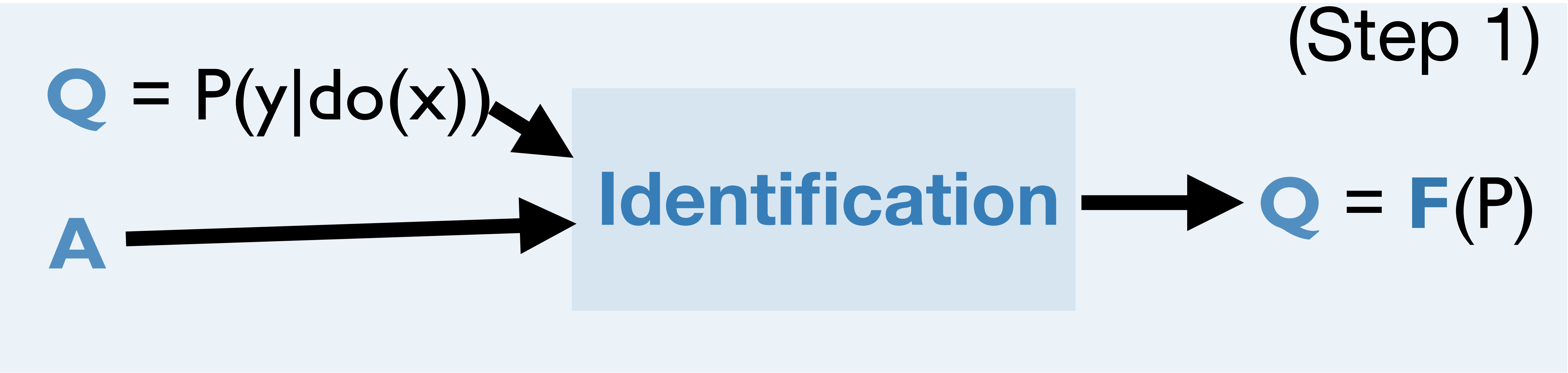


Learning causal effect: 2-step procedures

Learning causal effect: 2-step procedures

Step 1.
(Identification)

Represent the causal query Q=P(y|do(x)) as a function of P (i.e., Q=F(P)) from causal assumption A.



Learning causal effect: 2-step procedures

Step 1. (Identification)

Represent the causal query Q=P(y|do(x)) as a function of P (i.e., Q=F(P)) from causal assumption A.

Step 2. (Estimation)

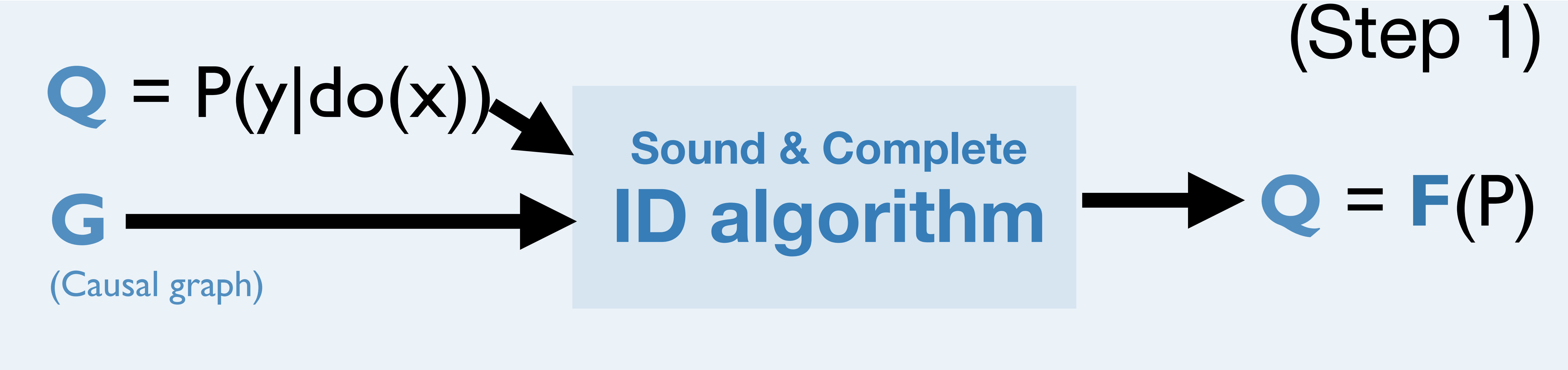
Estimate the identified estimand Q=F(P) from finite samples D

Current status of causal inference

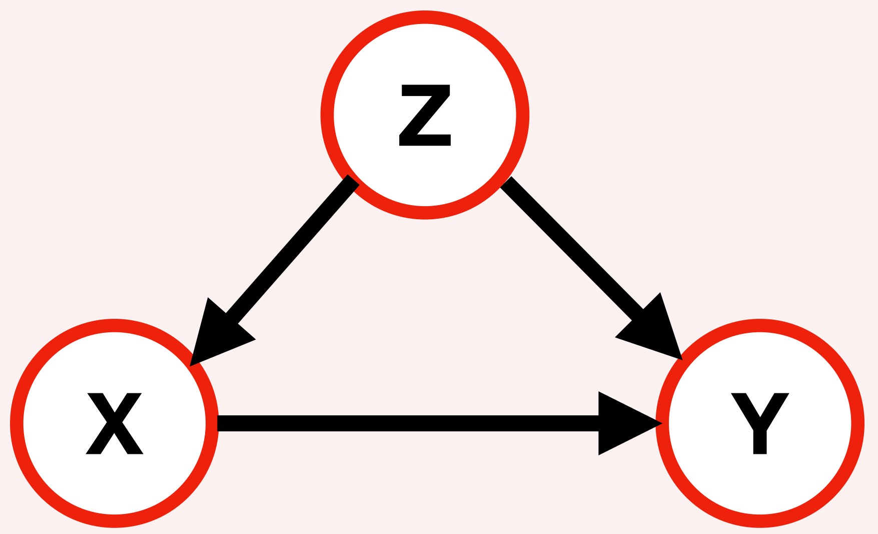
Current status of causal inference

Strength

- We have a sound and complete identification algorithm — A procedure for determining whether a causal query Q is could be represented as a functional of P (i.e., Q=F(P)) from the causal assumption.



Current status of causal infe



Ignorability assumption

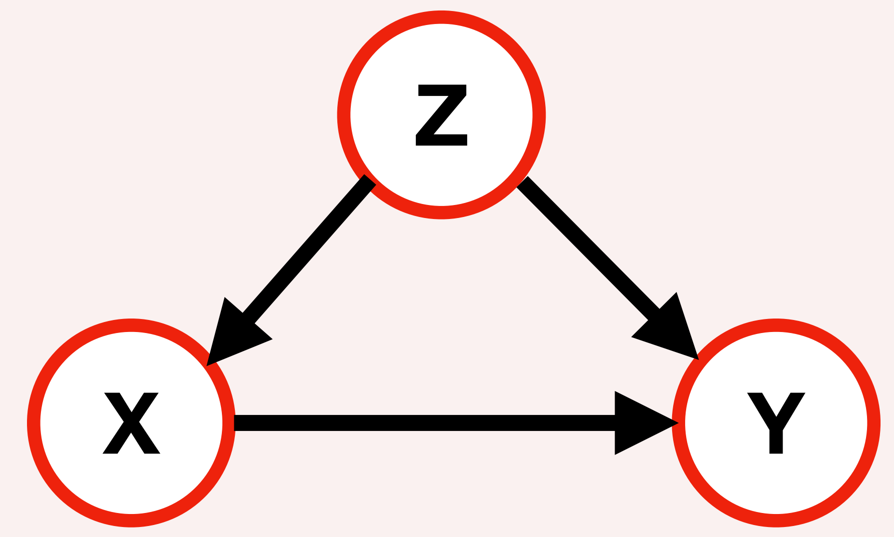
Strength

- We have a sound and complete identification algorithm — A procedure for determining whether a causal query Q is could be represented as a functional of P (i.e., Q=F(P)) from the causal assumption.

Weakness

- Estimation has been mainly done on 'Ignorability' assumption.

Current status of causal infe



Ignorability assumption

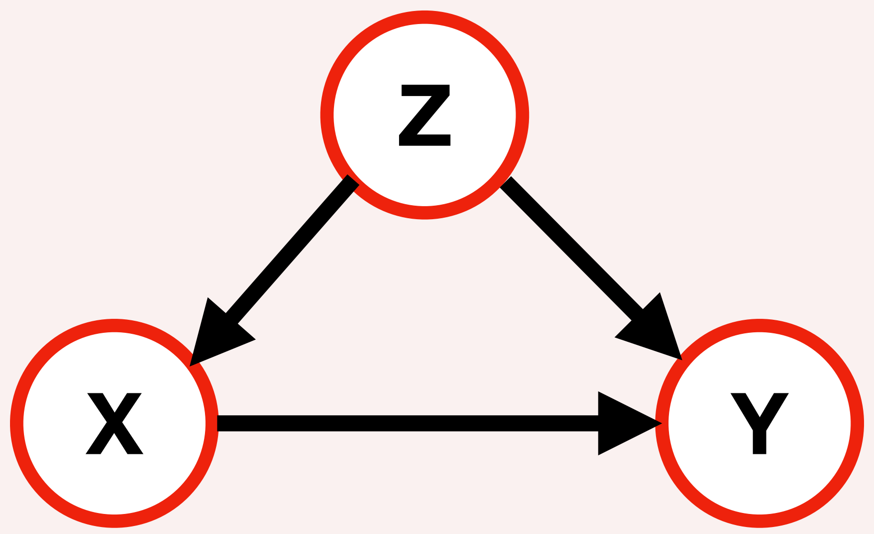
Strength

- We have a sound and complete identification algorithm — A procedure for determining whether a causal query Q is could be represented as a functional of P (i.e., Q=F(P)) from the causal assumption.

Weakness

- Estimation has been mainly done on 'Ignorability' assumption.
- For the general identifiable estimand, it's unclear how to estimate the causal effect sample & time-efficiently.

Current status of causal infe



Ignorability assumption

Strength

- We have a sound and complete identification algorithm — A procedure for determining whether a causal query Q is could be represented as a functional of P (i.e., Q=F(P)) from the causal assumption.

Weakness

- Estimation has been mainly done on 'Ignorability' assumption.
- For the general identifiable estimand, it's unclear how to estimate the causal effect sample & time-efficiently.

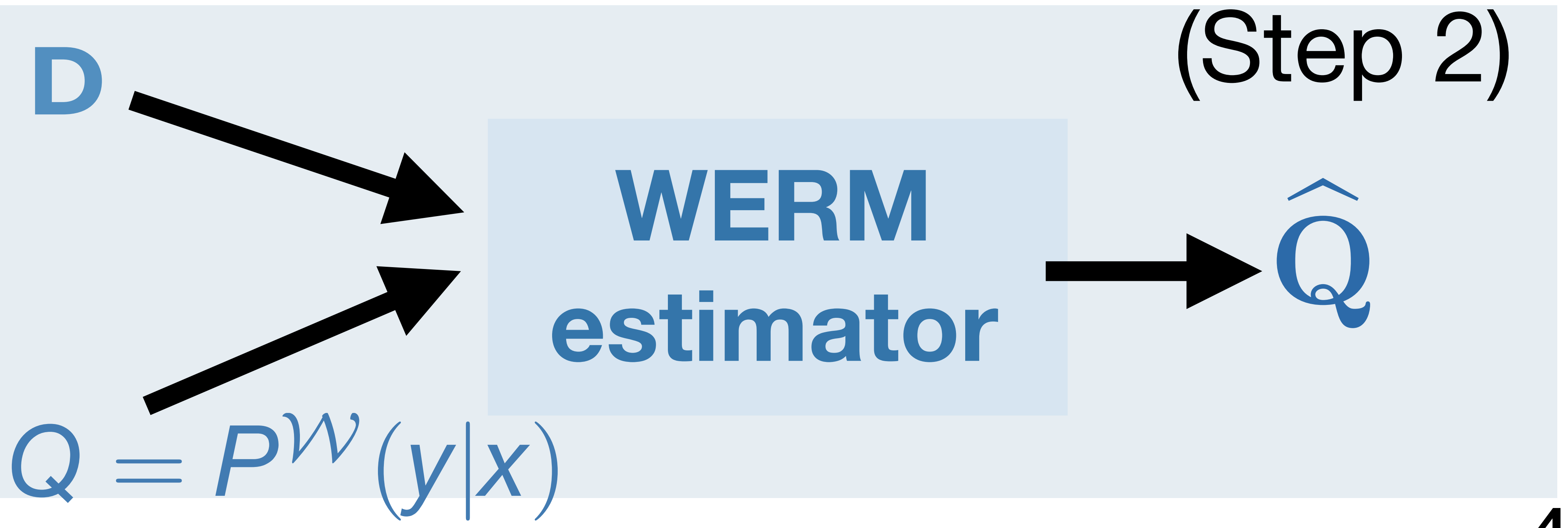
Learning causal effect: (Weighted) ERM

Learning causal effect: (Weighted) ERM

Strength

- When Q=F(P) is given as ***weighed distribution*** (e.g., *inverse probability weights (IPW) or importance sampling)*, then the empirical risk minimization (ERM) based estimators have been established.

Counterfactual Risk Minimization
Swaminathan and Joachims (2015)
Re-weighted Risk Minimization
Johansson (2018, 2020)



(e.g., W=P(x)/P(x|z) under ignorability assumption)

Learning causal effect: (Weighted) ERM

Weakness

- When Q=P(y|do(x)) is identifiable, but the identified estimand is given arbitrarily, how to use the WERM based estimators is unclear.

Strength

- When Q=F(P) is given as ***weighed distribution*** (e.g., *inverse probability weights (IPW) or importance sampling)*, then the empirical risk minimization (ERM) based estimators have been established.

Counterfactual Risk Minimization
Swaminathan and Joachims (2015)
Re-weighted Risk Minimization
Johansson (2018, 2020)

Learning causal effect: (Weighted) ERM

Weakness

- When Q=P(y|do(x)) is identifiable, but the identified estimand is given arbitrarily, how to use the WERM based estimators is unclear.

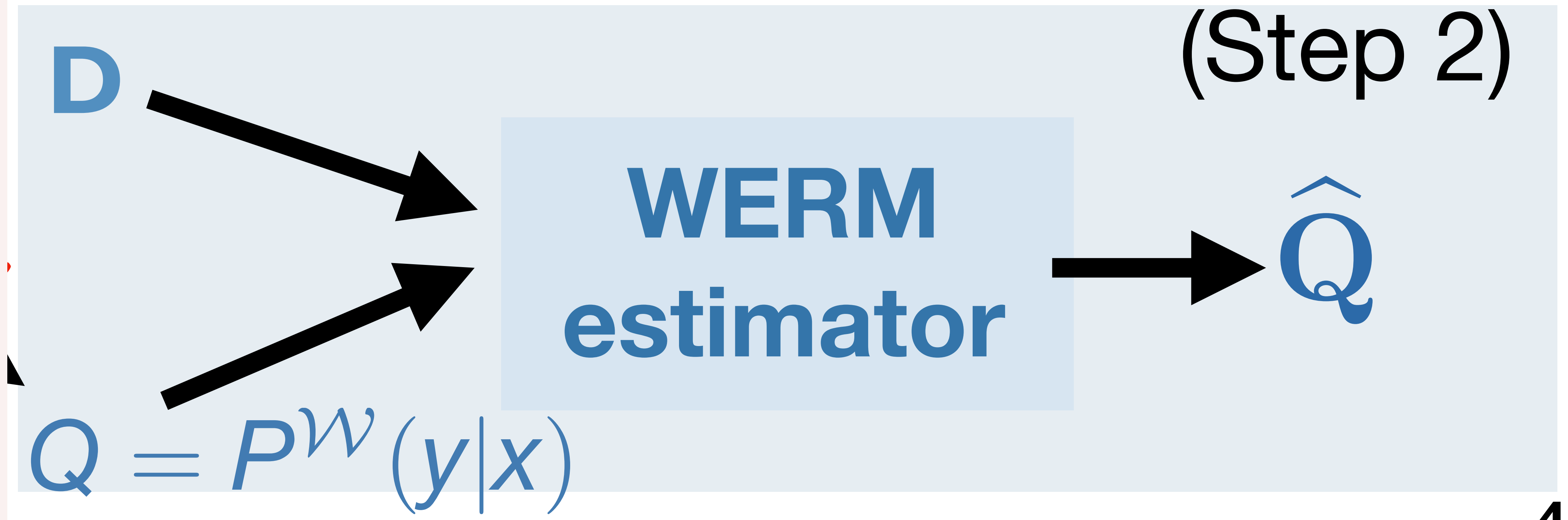
?

Strength

- When Q=F(P) is given as ***weighed distribution*** (e.g., *inverse probability weights (IPW) or importance sampling)*, then the empirical risk minimization (ERM) based estimators have been established.

Counterfactual Risk Minimization
Swaminathan and Joachims (2015)

Re-weighted Risk Minimization
Johansson (2018, 2020)

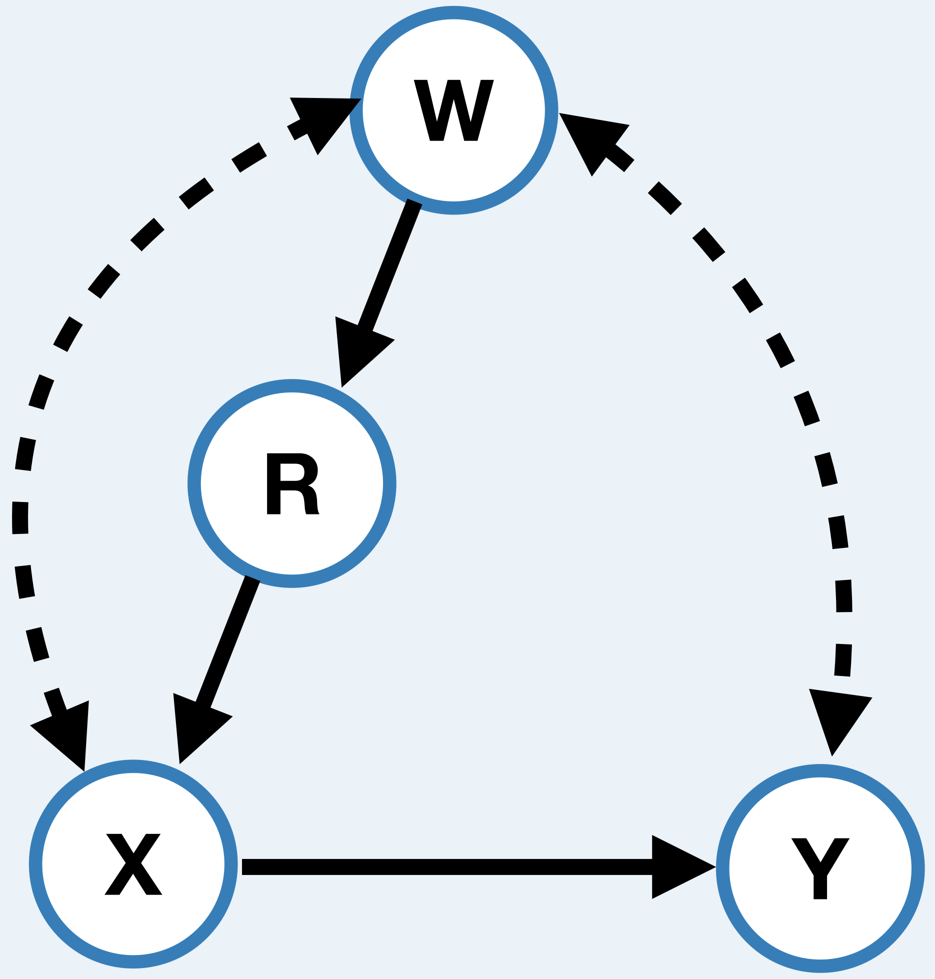


(e.g., W=P(x)/P(x|z) under ignorability assumption)

Connection b/w non-igno. and WERM

Connection b/w non-igno. and WERM

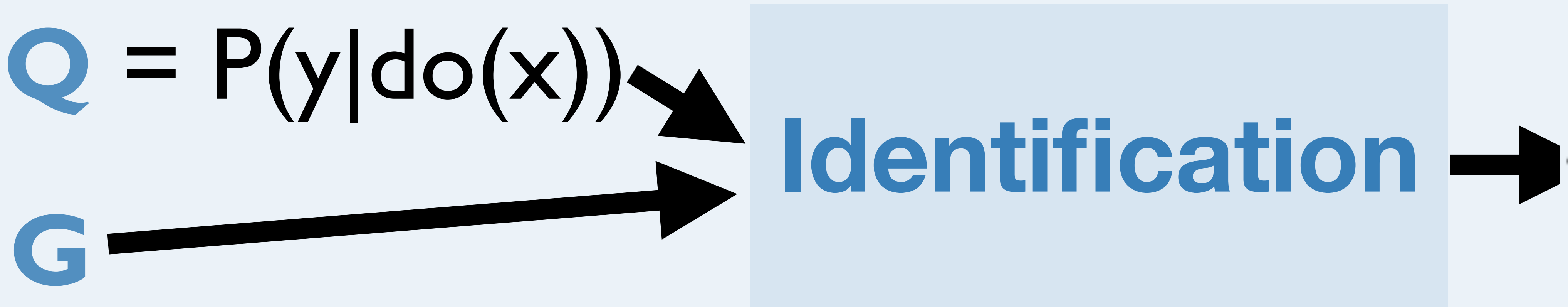
Practical Scenario on Non-igno. case (Example 1)



$$P(y|do(x)) = \frac{\sum_w P(y,x|r,w)P(w)}{\sum_w P(x|r,w)P(w)}$$

- The causal effect is identifiable, but the estimand is not a typical input for the WERM estimator.

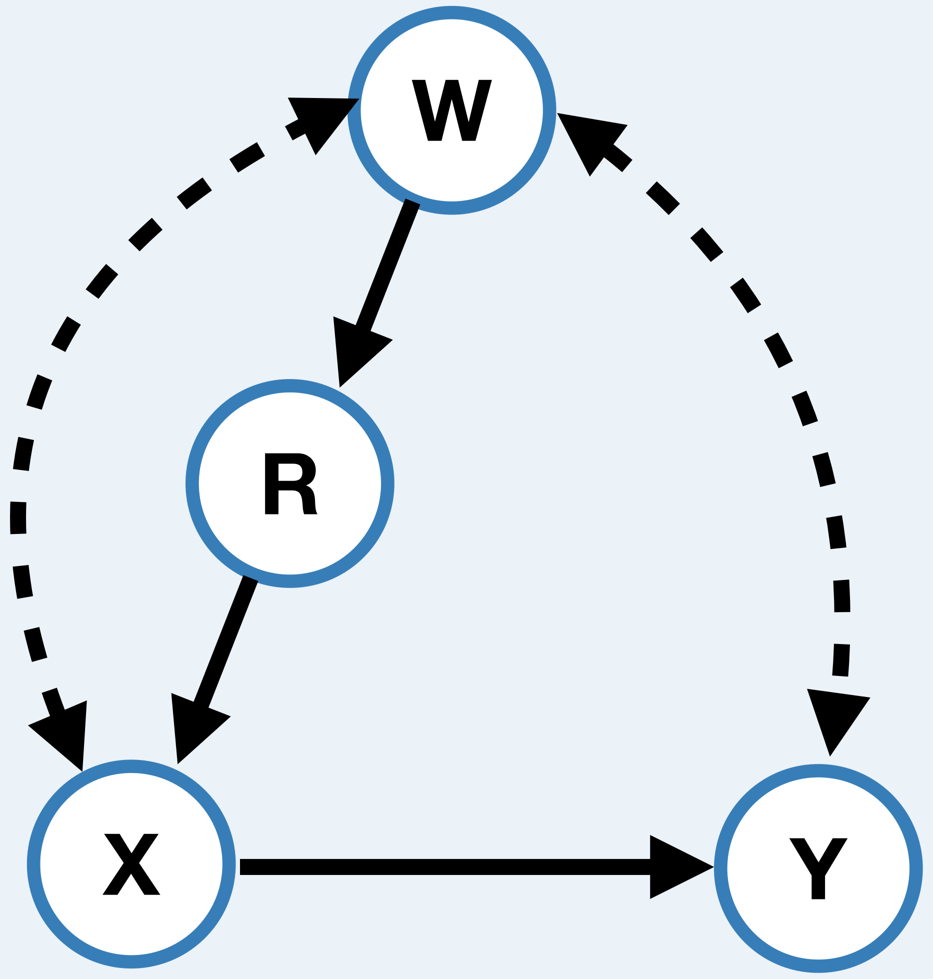
(Step 1)



$$\frac{\sum_w P(y,x|r,w)P(w)}{\sum_w P(x|r,w)P(w)}$$

Connection b/w non-igno. and WERM

Practical Scenario on Non-igno. case (Example 1)



$$P(y|do(x)) = \frac{\sum_w P(y,x|r,w)P(w)}{\sum_w P(x|r,w)P(w)}$$

- The causal effect is identifiable, but the estimand is not a typical input for the WERM estimator.
- However, we can represent the query Q=P(y|do(x)) as a weighted distribution
- Taking W=P(r)/P(r|w). Then,

$$P(y|do(x)) = P^{\mathcal{W}}(y|x,r)$$

End-to-end solution to causal inference

End-to-end solution to causal inference

Contribution Filling the bridge between causal effect identification and the estimation

End-to-end solution to causal inference

Contribution Filling the bridge between causal effect identification and the estimation

Q = P(y|do(x)) **G**

End-to-end solution to causal inference

Contribution Filling the bridge between causal effect identification and the estimation

Q = P(y|do(x)) **G**

Identification

$(W^*, R) = \mathbf{wID}(Q, P, G)$

(Algo. 1)

such that

$P(y|do(x)) = P^{W^*}(y|r)$

End-to-end solution to causal inference

Contribution Filling the bridge between causal effect identification and the estimation

Q $= P(y|do(x))$

G

Identification

$(W^*, R) =$ **wID**(Q, P, G)

(Algo. 1)

such that

$$P(y|do(x)) = P^{W^*}(y|r)$$

Estimation (Sec. 4)

- Generalization bound (Prop. 1)
- Structural Risk Minimization (Thm. 2)

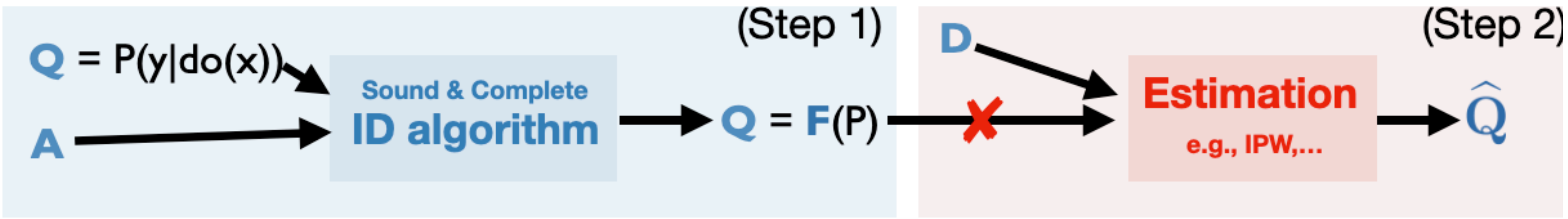
Summary

Summary

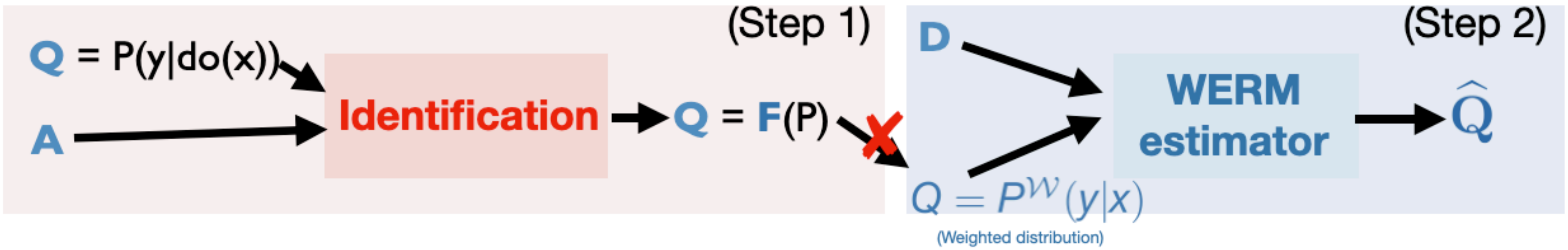
- A gap b/w causal effect *Identification* and *estimation.*

Summary

- A gap b/w causal effect *Identification* and *estimation.*

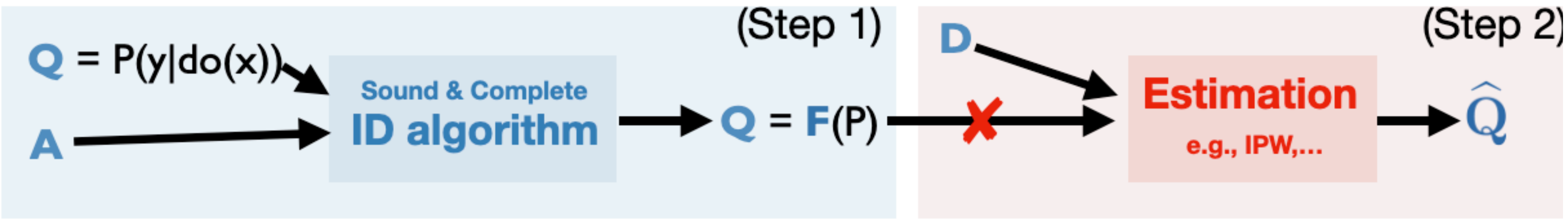


- A gap b/w ERM based estimators and the *causal inference.*

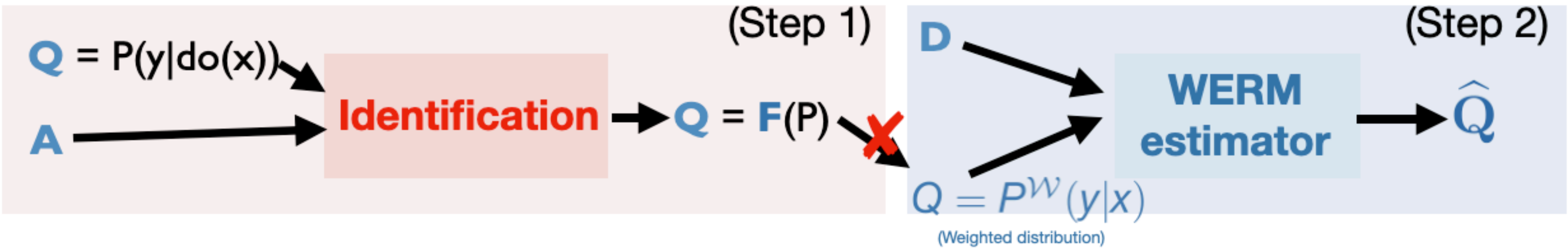


Summary

- A gap b/w causal effect *Identification* and *estimation.*



- A gap b/w ERM based estimators and the *causal inference.*



- We fill the gap between the causal inference and the ERM based estimation.