# Targeted Machine Learning for Average Causal Effect Estimation Using the Front-Door Functional and its Extensions

Anna Guo

Department of Biostatistics and Bioinformatics
Rollins School of Public Health
Emory University

anna.guo@emory.edu

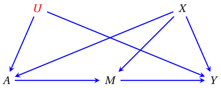
In collaborations with: David Benkeser (Emory), Razieh Nabi (Emory)

# Key takeaways

**R packages**

- `fdtmle`
- `ADMGtmle`

# Key takeaways

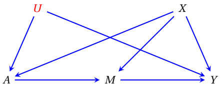



**(Front-door Model)**

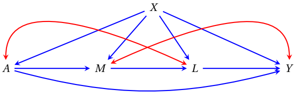
## R packages

- `fdtmle`
- `ADMGtmle`

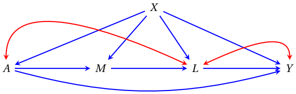
# Key takeaways




**(Front-door Model)**




**(Extension of Front-door model)**

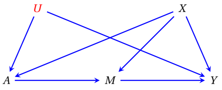



**(Extension of Front-door model)**

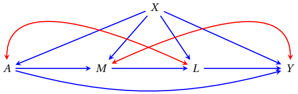
## R packages

- `fdtmle`
- `ADMGtmle`

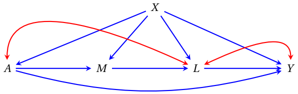
# Key takeaways




(Front-door Model)




(Extension of Front-door model)




(Extension of Front-door model)

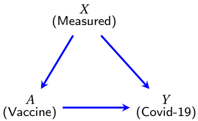
## R packages

- `fdtmle`
- `ADMGtmle`

## Functionalities

- Identification
- Estimation: doubly robust onestep & TMLE estimators
- Efficiency & Asymptotic Behavior

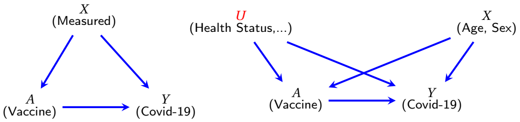
# Vaccine study




1. **Objective:** evaluate the causal effect of vaccine on Covid-19 incidence.

   Average Causal Effect: $\mathbb{E}(Y^{a=1}) - \mathbb{E}(Y^{a=0})$, Target Parameter: $\psi := \mathbb{E}\left[Y^{a_0}\right], a_0 \in \{0,1\}$

# Vaccine study

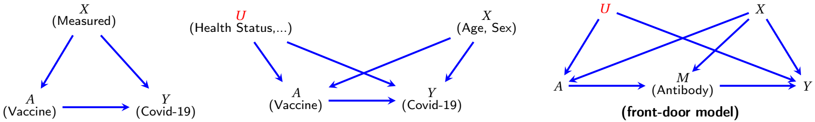



1. **Objective:** evaluate the causal effect of vaccine on Covid-19 incidence.

   Average Causal Effect: $\mathbb{E}(Y^{a=1}) - \mathbb{E}(Y^{a=0})$, Target Parameter: $\psi := \mathbb{E}\left[Y^{a_0}\right], a_0 \in \{0,1\}$

2. **Challenge:** presence of unmeasured confounders $U$.

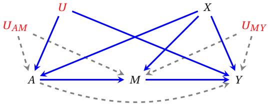
# Vaccine study




1. **Objective:** evaluate the causal effect of vaccine on Covid-19 incidence.

   Average Causal Effect: $\mathbb{E}(Y^{a=1}) - \mathbb{E}(Y^{a=0})$, Target Parameter: $\psi := \mathbb{E}\left[Y^{a_0}\right], a_0 \in \{0, 1\}$

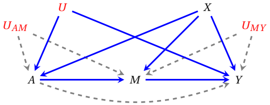
2. **Challenge:** presence of unmeasured confounders $U$.
3. **Solution:** presence of mediator(s) $M \Longrightarrow$ front-door model (Pearl, 1995)




(Encoded assumptions)

**Identification Assumptions:**

1. No direct effect: $Y^{a,m} = Y^{m}, \forall a, m$;
2. Conditional ignorability: $M^{a} \perp A \mid X$ & $Y^{m} \perp M \mid A, X$;
3. Consistency: $M^{a} = M$ when $A = a$ & $Y^{m} = Y$ when $M = m$;
4. Positivity: $P(A = 1 \mid X = x) > 0, P(M = m \mid A = a, X = x) > 0, \forall x$ with $P(X = x) > 0$




(Encoded assumptions)

**Identification Assumptions:**

1. No direct effect: $Y^{a,m} = Y^m, \forall a, m$;
2. Conditional ignorability: $M^a \perp A \mid X$ & $Y^m \perp M \mid A, X$;
3. Consistency: $M^a = M$ when $A = a$ & $Y^m = Y$ when $M = m$;
4. Positivity: $P(A = 1 \mid X = x) > 0, P(M = m \mid A = a, X = x) > 0, \forall x$ with $P(X = x) > 0$

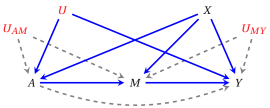
**Identification functional:**

$$\psi(P) = \iint \sum_{a=0}^{1} \underbrace{\mathbb{E}(Y \mid m, a, x)}_{\mu(m,a,x)} \underbrace{p(a \mid x)}_{\pi(a|x)} \underbrace{p(m \mid a_0, x)}_{f_M(m|a_0,x)} \underbrace{p(x)}_{p_X(x)} \, dm \, dx \quad \textbf{(target estimand).}$$

The above functional depends on the following **nuisance parameters**:

- outcome regression: $\mathbb{E}[Y \mid m, a, x]$, denoted by $\mu(m, a, x)$

# Identification




(Encoded assumptions)

**Identification Assumptions:**

1. No direct effect: $Y^{a,m} = Y^m, \forall a, m$;
2. Conditional ignorability: $M^a \perp A \mid X$ & $Y^m \perp M \mid A, X$;
3. Consistency: $M^a = M$ when $A = a$ & $Y^m = Y$ when $M = m$;
4. Positivity: $P(A = 1 \mid X = x) > 0, P(M = m \mid A = a, X = x) > 0, \forall x$ with $P(X = x) > 0$

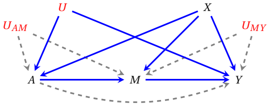
**Identification functional:**

$$\psi(P) = \iint \sum_{a=0}^{1} \underbrace{\mathbb{E}(Y \mid m, a, x)}_{\mu(m,a,x)} \underbrace{p(a \mid x)}_{\pi(a|x)} \underbrace{p(m \mid a_0, x)}_{f_M(m|a_0,x)} \underbrace{p(x)}_{p_X(x)} \, dm \, dx \quad \textbf{(target estimand)}.$$

The above functional depends on the following **nuisance parameters**:

- outcome regression: $\mathbb{E}[Y \mid m, a, x]$, denoted by $\mu(m, a, x)$
- propensity score: $p(a \mid x)$, denoted by $\pi(a|x)$

# Identification




(Encoded assumptions)

**Identification Assumptions:**

1. No direct effect: $Y^{a,m} = Y^m, \forall a, m$;
2. Conditional ignorability: $M^a \perp A \mid X$ & $Y^m \perp M \mid A, X$;
3. Consistency: $M^a = M$ when $A = a$ & $Y^m = Y$ when $M = m$;
4. Positivity: $P(A=1 \mid X=x) > 0, P(M=m \mid A=a, X=x) > 0, \forall x$ with $P(X=x) > 0$

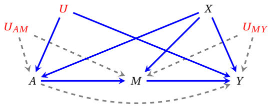
**Identification functional:**

$$\psi(P) = \iint \sum_{a=0}^{1} \underbrace{\mathbb{E}(Y \mid m, a, x)}_{\mu(m,a,x)} \underbrace{p(a \mid x)}_{\pi(a|x)} \underbrace{p(m \mid a_0, x)}_{f_M(m|a_0,x)} \underbrace{p(x)}_{p_X(x)} \, dm \, dx \quad \textbf{(target estimand).}$$

The above functional depends on the following **nuisance parameters**:

- outcome regression: $\mathbb{E}[Y \mid m, a, x]$, denoted by $\mu(m, a, x)$
- propensity score: $p(a \mid x)$, denoted by $\pi(a|x)$
- mediator density: $p(m \mid a_0, x)$, denoted by $f_M(m \mid a_0, x)$

# Identification




(Encoded assumptions)

**Identification Assumptions:**

1. No direct effect: $Y^{a,m} = Y^m, \forall a, m$;
2. Conditional ignorability: $M^a \perp A \mid X$ & $Y^m \perp M \mid A, X$;
3. Consistency: $M^a = M$ when $A = a$ & $Y^m = Y$ when $M = m$;
4. Positivity: $P(A=1 \mid X=x) > 0, P(M=m \mid A=a, X=x) > 0, \forall x$ with $P(X=x) > 0$

**Identification functional:**

$$\psi(P) = \iint \sum_{a=0}^{1} \underbrace{\mathbb{E}(Y \mid m, a, x)}_{\mu(m,a,x)} \underbrace{p(a \mid x)}_{\pi(a|x)} \underbrace{p(m \mid a_0, x)}_{f_M(m|a_0,x)} \underbrace{p(x)}_{p_X(x)} \, dm \, dx \quad \textbf{(target estimand).}$$

The above functional depends on the following **nuisance parameters**:

- outcome regression: $\mathbb{E}[Y \mid m, a, x]$, denoted by $\mu(m, a, x)$
- propensity score: $p(a \mid x)$, denoted by $\pi(a|x)$
- mediator density: $p(m \mid a_0, x)$, denoted by $f_M(m \mid a_0, x)$
- covariates density: $p(x)$, denoted by $p_X(x)$.

Let $Q = \{\mu, \pi, f_M, p_X\}$, we can write $\psi(P)$ as $\psi(Q)$.

# Estimation - Plugin Estimator

$$\psi(Q) = \iint \sum_{a=0}^{1} \mu(m,a,x)\ \pi(a \mid x)\ f_M(m \mid a_0, x)\ p_X(x)\ dm\ dx \quad \textbf{(ID functional)}$$

$$\psi(\widehat{Q}) = \frac{1}{n} \sum_{i=1}^{n} \sum_{m} \sum_{a=0}^{1} \widehat{\mu}(m,a,X_i)\ \widehat{\pi}(a \mid X_i)\ \widehat{f}_M(m \mid a_0, X_i) \quad \textbf{(plugin estimator)}.$$

# Estimation - Plugin Estimator

$$\psi(Q) = \iint \sum_{a=0}^{1} \mu(m,a,x)\ \pi(a \mid x)\ f_M(m \mid a_0, x)\ p_X(x)\ dm\ dx \quad \textbf{(ID functional)}$$

$$\psi(\widehat{Q}) = \frac{1}{n} \sum_{i=1}^{n} \sum_{m} \sum_{a=0}^{1} \widehat{\mu}(m,a,X_i)\ \widehat{\pi}(a \mid X_i)\ \color{red}{\widehat{f}_M(m \mid a_0, X_i)} \quad \textbf{(plugin estimator)}.$$

**Limitations:**

- density estimation and numeric integration

# Estimation - Plugin Estimator

$$\psi(Q) = \iint \sum_{a=0}^{1} \mu(m,a,x)\ \pi(a \mid x)\ f_M(m \mid a_0, x)\ p_X(x)\ dm\ dx \quad \textbf{(ID functional)}$$

$$\psi(\widehat{Q}) = \frac{1}{n} \sum_{i=1}^{n} \sum_{m} \sum_{a=0}^{1} \widehat{\mu}(m,a,X_i)\ \widehat{\pi}(a \mid X_i)\ \widehat{f}_M(m \mid a_0, X_i) \quad \textbf{(plugin estimator).}$$

**Limitations:**

- density estimation and numeric integration
- first-order bias: $\psi(\widehat{Q}) = \psi(Q) - \underbrace{P\Phi(\widehat{Q})}_{\text{first-order bias}} + \underbrace{R_2(\widehat{Q}, Q)}_{\text{remainder term}}$ (von Mises Expansion).

where $\Phi(Q)$ is the **efficient influence function** of $\psi(Q)$, which is unique under nonparametric models, such as the front-door model.

- Correction for the first-order bias yields the **onestep estimator**

$$\psi(\widehat{Q}) = \psi(Q) - P\Phi(\widehat{Q}) + R_2(\widehat{Q}, Q) \quad \Longrightarrow \quad \psi^+(\widehat{Q}) = \psi(\widehat{Q}) + P_n\Phi(\widehat{Q}).$$

- Fulcher et al. (2020) first derived $\Phi(Q)$, and suggested the following estimator:

$$\psi^+(\widehat{Q}) = \frac{1}{n}\sum_{i=1}^{n} \frac{\widehat{f}_M(M_i \mid a_0, X_i)}{\widehat{f}_M(M_i \mid A_i, X_i)}\{Y_i - \widehat{\mu}(M_i, A_i, X_i)\}$$
$$+ \frac{\mathbb{I}(A_i = a_0)}{\widehat{\pi}(a_0 \mid X_i)}\Big\{\sum_a \widehat{\mu}(M_i, a, X_i)\widehat{\pi}(a \mid X_i) - \int \sum_a \widehat{\mu}(M_i, a, X_i)\widehat{\pi}(a \mid X_i)\widehat{f}_M(m \mid a_0, X_i)\,dm\Big\}$$
$$+ \int \widehat{\mu}(m, A_i, X_i)\widehat{f}_M(m \mid a_0, X_i)\,dm$$

- Nuisance estimates: $\widehat{Q} = \{\widehat{\mu}, \widehat{\pi}, \widehat{f}_M\}$, while $p_X$ is emipirically evaluated
- Double robustness: $\psi^+(\widehat{Q})$ is a consistent estimator if either $\widehat{f}_M$ or $\{\widehat{\mu}, \widehat{\pi}\}$ is correctly specified in parametric models.

**Limitations**

- The proposed onestep estimator requires estimating the mediator density $f_M(M|A,X)$. A daunting task under large collection of mediators of different types.
- Onestep estimator may yield estimates that is outside of the range of the target parameter, especially for binary outcome.
- The asymptotic behavior of this one-step estimator, along with its second-order remainder term, requires additional investigation.

**Limitations**

- The proposed onestep estimator requires estimating the mediator density $f_M(M|A,X)$. A daunting task under large collection of mediators of different types.
- Onestep estimator may yield estimates that is outside of the range of the target parameter, especially for binary outcome.
- The asymptotic behavior of this one-step estimator, along with its second-order remainder term, requires additional investigation.

**Solutions**

- Re-parameterize $\psi(Q)$ to avoid direct density estimation
- Adopt Targeted Minimum Loss Based Estimation (TMLE) (Van der Laan et al., 2011)

Nuisances that involving the **conditional density** $f_M$:

$$\psi^+(\widehat{Q}) = \frac{1}{n}\sum_{i=1}^{n} \frac{\widehat{f}_M(M_i \mid a_0, X_i)}{\widehat{f}_M(M_i \mid A_i, X_i)} \{Y_i - \widehat{\mu}(M_i, A_i, X_i)\}$$
$$+ \frac{\mathbb{I}(A_i = a_0)}{\widehat{\pi}(a_0 \mid X_i)} \Big\{ \sum_a \widehat{\mu}(M_i, a, X_i)\widehat{\pi}(a \mid X_i) - \int \sum_a \widehat{\mu}(M_i, a, X_i)\widehat{\pi}(a \mid X_i)\widehat{f}_M(m \mid a_0, X_i)\, dm \Big\}$$
$$+ \int \widehat{\mu}(m, A_i, X_i)\widehat{f}_M(m \mid a_0, X_i)\, dm$$

- Mediator density ratio estimation via **Bayes rule**

$$f_M^r(m, a, x) = \frac{f_M(m \mid a_0, x)}{f_M(m \mid a, x)} = \frac{\lambda(a_0 \mid x, m)}{\lambda(a \mid x, m)} \times \frac{\pi(a \mid x)}{\pi(a_0 \mid x)},$$

where $\lambda(a \mid x, m) = p(A = a \mid X = x, M = m)$.

# Estimation - Onestep Estimator

Nuisances that involving the **conditional density** $f_M$:

$$\psi^{+}(\widehat{Q}) = \frac{1}{n}\sum_{i=1}^{n} \frac{\widehat{f}_M(M_i \mid a_0, X_i)}{\widehat{f}_M(M_i \mid A_i, X_i)} \{Y_i - \widehat{\mu}(M_i, A_i, X_i)\}$$

$$+ \frac{\mathbb{I}(A_i = a_0)}{\widehat{\pi}(a_0 \mid X_i)} \Big\{ \sum_a \widehat{\mu}(M_i, a, X_i)\widehat{\pi}(a \mid X_i) - \underbrace{\int \overbrace{\sum_a \widehat{\mu}(M_i, a, X_i)\widehat{\pi}(a \mid X_i)}^{\xi(M_i, X_i)} \widehat{f}_M(m \mid a_0, X_i) dm}_{\mathbb{E}(\xi(M_i, X_i) \mid a_0, X_i)} \Big\}$$

$$+ \int \widehat{\mu}(m, A_i, X_i)\widehat{f}_M(m \mid a_0, X_i)\, dm$$

- Mediator density ratio estimation via **Bayes rule**

$$f_M^r(m, a, x) = \frac{f_M(m \mid a_0, x)}{f_M(m \mid a, x)} = \frac{\lambda(a_0 \mid x, m)}{\lambda(a \mid x, m)} \times \frac{\pi(a \mid x)}{\pi(a_0 \mid x)},$$

where $\lambda(a \mid x, m) = p(A = a \mid X = x, M = m)$.

- **Sequential regression** for estimating $\gamma(x)$: $\gamma(x) = \mathbb{E}(\xi(m, x) \mid a_0, x)$

## Estimation - Onestep Estimator

Nuisances that involving the **conditional density** $f_M$:

$$\psi^+(\widehat{Q}) = \frac{1}{n}\sum_{i=1}^{n} \frac{\widehat{f}_M(M_i \mid a_0, X_i)}{\widehat{f}_M(M_i \mid A_i, X_i)}\{Y_i - \widehat{\mu}(M_i, A_i, X_i)\}$$

$$+ \frac{\mathbb{I}(A_i = a_0)}{\widehat{\pi}(a_0 \mid X_i)}\Big\{\sum_a \widehat{\mu}(M_i, a, X_i)\widehat{\pi}(a \mid X_i) - \int \sum_a \widehat{\mu}(M_i, a, X_i)\widehat{\pi}(a \mid X_i)\widehat{f}_M(m \mid a_0, X_i)\, dm\Big\}$$

$$+ \underbrace{\int \widehat{\mu}(m, A_i, X_i)\,\widehat{f}_M(m \mid a_0, X_i) dm}_{\kappa_a = \mathbb{E}[\mu(M_i, a, X_i) \mid A = a_0, X],\ \text{for } a \in \{0,1\}}$$

- Mediator density ratio estimation via **Bayes rule**

$$f_M^r(m, a, x) = \frac{f_M(m \mid a_0, x)}{f_M(m \mid a, x)} = \frac{\lambda(a_0 \mid x, m)}{\lambda(a \mid x, m)} \times \frac{\pi(a \mid x)}{\pi(a_0 \mid x)},$$

where $\lambda(a \mid x, m) = p(A = a \mid X = x, M = m)$.

- **Sequential regression** for estimating $\gamma(x)$: $\gamma(x) = \mathbb{E}(\xi(m, x) \mid a_0, x)$
- **Sequential regression** for estimating $\eta(x)$

$$\eta(a, x) = \int \mu(m, a, x) f_{M|A,X}(m \mid a_0, x)\, dm = A\kappa_1(X) + (1 - A)\kappa_0(X).$$

- Reparameterized onestep estimator

$$
\begin{aligned}
\psi_2^+(\widehat{Q}) = \frac{1}{n}\sum_{i=1}^{n}\Big[\widehat{f}_M^r\left(M_i, A_i, X_i\right)\left\{Y_i - \widehat{\mu}\left(M_i, A_i, X_i\right)\right\} \\
+ \frac{\mathbb{I}\left(A_i = a_0\right)}{\widehat{\pi}\left(a_0 \mid X_i\right)}\left\{\widehat{\xi}\left(M_i, X_i\right) - \widehat{\gamma}\left(X_i\right)\right\} \\
+ \left\{\widehat{\kappa}_1\left(X_i\right) - \widehat{\kappa}_0\left(X_i\right)\right\}\left\{A_i - \widehat{\pi}\left(1 \mid X_i\right)\right\} \\
+\widehat{\gamma}\left(X_i\right)\Big]. \qquad \text{(second one-step estimator)}
\end{aligned}
$$

new set of nuisance parameters: $Q = \{\mu, \pi, p_X, \underbrace{f_M^r, \gamma, \kappa}_{f_M}\}$ or $\{\mu, \pi, p_X, \underbrace{\lambda, \gamma, \kappa}_{f_M}\}$.

# Estimation - Onestep Estimator (Reparameterized)

- Reparameterized onestep estimator

$$\psi_2^+(\widehat{Q}) = \frac{1}{n}\sum_{i=1}^{n}\Big[\widehat{f}_M^r(M_i, A_i, X_i)\{Y_i - \widehat{\mu}(M_i, A_i, X_i)\} \\ + \frac{\mathbb{I}(A_i = a_0)}{\widehat{\pi}(a_0 \mid X_i)}\{\widehat{\xi}(M_i, X_i) - \widehat{\gamma}(X_i)\} \\ + \{\widehat{\kappa}_1(X_i) - \widehat{\kappa}_0(X_i)\}\{A_i - \widehat{\pi}(1 \mid X_i)\} \\ + \widehat{\gamma}(X_i)\Big]. \qquad \text{(second one-step estimator)}$$

new set of nuisance parameters: $Q = \{\mu, \pi, p_X, \underbrace{f_M^r, \gamma, \kappa}_{f_M}\}$ or $\{\mu, \pi, p_X, \underbrace{\lambda, \gamma, \kappa}_{f_M}\}$.

- The onestep estimator resolves the first-order bias by adding $P_n\Phi(\widehat{Q})$ on top of the plugin estimator.

- Reparameterized onestep estimator

$$\begin{aligned}\psi_2^{+}(\widehat{Q}) = \frac{1}{n}\sum_{i=1}^{n}\Big[&\widehat{f}_M^r\left(M_i, A_i, X_i\right)\left\{Y_i - \widehat{\mu}\left(M_i, A_i, X_i\right)\right\} \\ &+ \frac{\mathbb{I}\left(A_i = a_0\right)}{\widehat{\pi}\left(a_0 \mid X_i\right)}\left\{\widehat{\xi}\left(M_i, X_i\right) - \widehat{\gamma}\left(X_i\right)\right\} \\ &+ \left\{\widehat{\kappa}_1\left(X_i\right) - \widehat{\kappa}_0\left(X_i\right)\right\}\left\{A_i - \widehat{\pi}\left(1 \mid X_i\right)\right\} \\ &+ \widehat{\gamma}\left(X_i\right)\Big]. \qquad \text{(second one-step estimator)}\end{aligned}$$

new set of nuisance parameters: $Q = \{\mu, \pi, p_X, \underbrace{f_M^r, \gamma, \kappa}_{f_M}\}$ or $\{\mu, \pi, p_X, \underbrace{\lambda, \gamma, \kappa}_{f_M}\}$.

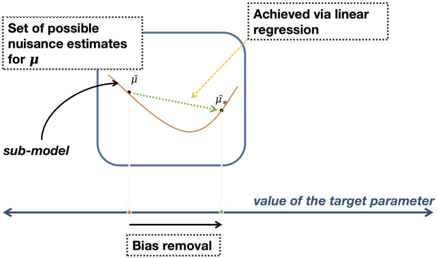
- The onestep estimator resolves the first-order bias by adding $P_n\Phi(\widehat{Q})$ on top of the plugin estimator.
- **Unaddressed**: estimates that fall out of the parameter space.

- Update $\widehat{Q} \Longrightarrow \widehat{Q}^*$ s.t $P_n\Phi(\widehat{Q}^*) \approx 0$. **TMLE estimator** is defined as $\psi_2(\widehat{Q}^*)$

### The TMLE procedure

1. Obtain an initial estimate of the nuisances $\hat{Q} = \{\hat{\mu}, \hat{\pi}, \cdots\}$
2. **loss functions** $L : Q \to R$
   **parametric submodels** $\widehat{Q}_\epsilon$ s.t

   (C1) $\widehat{Q} = \widehat{Q}_{\epsilon=0}$
   (C2) $Q = \text{argmin}_{\widetilde{Q}\in\mathcal{M}_Q} \int L(\widetilde{Q})dQ$
   (C3) $\frac{\partial}{\partial\epsilon} L\left(\widehat{Q}_\epsilon\right)\Big|_{\epsilon=0} = \Phi(\widehat{Q})$



**Theorem (Asymptotic linearity of $\psi_2(\hat{Q}^*)$)**

*Assume the nuisance estimates $\hat{Q}^\star = (\hat{\mu}^*, \hat{\pi}^*, \hat{\gamma}^*, \hat{\kappa}, \hat{\lambda})$ have the following $L^2(P)$ rates of convergence:*

$$||\hat{\pi}^* - \pi|| = o_P(n^{-\frac{1}{k}}), \quad ||\hat{\mu}^* - \mu|| = o_P(n^{-\frac{1}{q}})$$
$$||\hat{\gamma}^* - \gamma|| = o_P(n^{-\frac{1}{j}}), \quad ||\hat{\kappa}_a - \kappa_a|| = o_P(n^{-\frac{1}{\ell}}), \quad ||\hat{\lambda} - \lambda|| = o_P(n^{-\frac{1}{d}}).$$

*The TMLE $\psi_2(\hat{Q}^*)$ is asymptotically linear if the following condition as well as the Donsker condition are satisfied.*

$$\frac{1}{q} + \frac{1}{k} \geq \frac{1}{2}, \quad \frac{1}{d} + \frac{1}{q} \geq \frac{1}{2}, \quad \frac{1}{k} + \frac{1}{j} \geq \frac{1}{2}, \quad \frac{1}{k} + \frac{1}{\ell} \geq \frac{1}{2}.$$

### Theorem (Asymptotic linearity of $\psi_2(\hat{Q}^*)$)

*Assume the nuisance estimates $\hat{Q}^\star = (\hat{\mu}^*, \hat{\pi}^*, \hat{\gamma}^*, \hat{\kappa}, \hat{\lambda})$ have the following $L^2(P)$ rates of convergence:*

$$||\hat{\pi}^* - \pi|| = o_P(n^{-\frac{1}{k}}), \quad ||\hat{\mu}^* - \mu|| = o_P(n^{-\frac{1}{q}})$$
$$||\hat{\gamma}^* - \gamma|| = o_P(n^{-\frac{1}{j}}), \quad ||\hat{\kappa}_a - \kappa_a|| = o_P(n^{-\frac{1}{\ell}}), \quad ||\hat{\lambda} - \lambda|| = o_P(n^{-\frac{1}{d}}).$$

*The TMLE $\psi_2(\hat{Q}^*)$ is asymptotically linear if the following condition as well as the Donsker condition are satisfied.*

$$\frac{1}{q} + \frac{1}{k} \geq \frac{1}{2}, \quad \frac{1}{d} + \frac{1}{q} \geq \frac{1}{2}, \quad \frac{1}{k} + \frac{1}{j} \geq \frac{1}{2}, \quad \frac{1}{k} + \frac{1}{\ell} \geq \frac{1}{2}.$$

- Embrace a larger set of machine learning & statistical models

## Theorem (Asymptotic linearity of $\psi_2(\hat{Q}^*)$)

*Assume the nuisance estimates* $\hat{Q}^\star = (\hat{\mu}^*, \hat{\pi}^*, \hat{\gamma}^*, \hat{\kappa}, \hat{\lambda})$ *have the following* $L^2(P)$ *rates of convergence:*

$$||\hat{\pi}^* - \pi|| = o_P(n^{-\frac{1}{k}}), \quad ||\hat{\mu}^* - \mu|| = o_P(n^{-\frac{1}{q}})$$

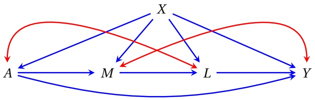
$$||\hat{\gamma}^* - \gamma|| = o_P(n^{-\frac{1}{j}}), \quad ||\hat{\kappa}_a - \kappa_a|| = o_P(n^{-\frac{1}{\ell}}), \quad ||\hat{\lambda} - \lambda|| = o_P(n^{-\frac{1}{d}}).$$

*The TMLE* $\psi_2(\hat{Q}^*)$ *is asymptotically linear if the following condition as well as the Donsker condition are satisfied.*

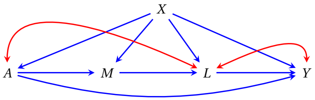
$$\frac{1}{q} + \frac{1}{k} \geq \frac{1}{2}, \quad \frac{1}{d} + \frac{1}{q} \geq \frac{1}{2}, \quad \frac{1}{k} + \frac{1}{j} \geq \frac{1}{2}, \quad \frac{1}{k} + \frac{1}{\ell} \geq \frac{1}{2}.$$

- Embrace a larger set of machine learning & statistical models
- Cross-fitting as an alternative of Donsker condition.

Generalize front-door model to settings with multiple mediators that have different confounding behavior

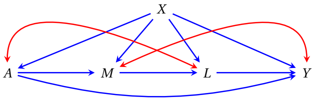



(a)

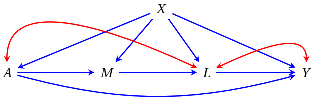



(b)

Generalize front-door model to settings with multiple mediators that have different confounding behavior




(a)

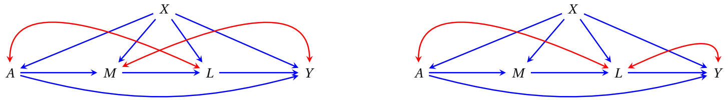



(b)

- primal fixability of $A$: no bidrected $\leftrightarrow$ path from $A$ to its children.

# Extension of the Front-door Model - Identification

Generalize front-door model to settings with multiple mediators that have different confounding behavior

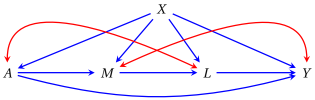



(a) (b)

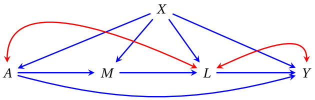
- primal fixability of $A$: no bidrected $\leftrightarrow$ path from $A$ to its children.
- primal fixability of $A \Longleftrightarrow$ identifiability of the causal effect of $A$ on $V \setminus A$ (Tian and Pearl, 2002)

# Extension of the Front-door Model - Identification

Generalize front-door model to settings with multiple mediators that have different confounding behavior




(a)



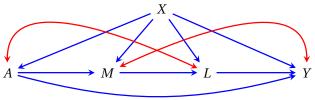

(b)

- primal fixability of $A$: no bidrected $\leftrightarrow$ path from $A$ to its children.
- primal fixability of $A \Longleftrightarrow$ identifiability of the causal effect of $A$ on $V \setminus A$ (Tian and Pearl, 2002)
- identification of $\mathbb{E}(Y^{a_0})$

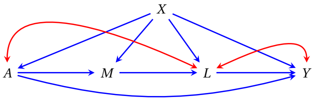
$$\psi(t) = \sum_{V \setminus A} Y \times \prod_{M_i \in \mathbb{M}} p\left(M_i \mid \text{mp}_{\mathcal{G}}\left(M_i\right)\right)\Big|_{A=a_0} \times \sum_{A} \prod_{L_i \in \mathbb{L}} p\left(L_i \mid \text{mp}_{\mathcal{G}}\left(L_i\right)\right) \times p(\mathbb{C}),$$

# Extension of the Front-door Model - Identification

Generalize front-door model to settings with multiple mediators that have different confounding behavior




(a)



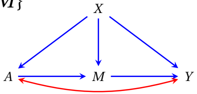

(b)

- primal fixability of $A$: no bidrected $\leftrightarrow$ path from $A$ to its children.
- primal fixability of $A \iff$ identifiability of the causal effect of $A$ on $V \setminus A$ (Tian and Pearl, 2002)
- identification of $\mathbb{E}(Y^{a_0})$

$$\psi(t) = \sum_{V \setminus A} Y \times \prod_{M_i \in \mathbb{M}} p\left(M_i \mid \text{mp}_{\mathcal{G}}\left(M_i\right)\right)\Big|_{A=a_0} \times \sum_{A} \prod_{L_i \in \mathbb{L}} p\left(L_i \mid \text{mp}_{\mathcal{G}}\left(L_i\right)\right) \times p(\mathbb{C}),$$

- front-door model as an example: $\mathbb{C} = \{X\}, \quad \mathbb{L} = \{A, Y\}, \quad \mathbb{M} = \{M\}$

$$\psi(t) = \sum_{X,M,Y} Y \times p(M \mid a_0, X) \times \sum_{A} p(Y \mid M, A, X) p(A \mid X) p(X)$$



- the EIF of $\mathbb{E}(Y^{a_0})$, which is unique under nonparametrically saturated model (Bhattacharya et al., 2022)

$$
\begin{aligned}
U_{\psi_t} = & \sum_{M_i \in \mathbb{M}} \Big\{ \frac{\mathbb{I}(A=a_0)}{\prod_{L_i \prec M_i} p(L_i \mid \mathrm{mp}_{\mathcal{G}}(L_i))} \times \Big( \sum_{A \cup \{\succ M_i\}} Y \times \prod_{\substack{V_i \in \mathbb{L} \\ \{\succ M_i\}}} p(V_i \mid \mathrm{mp}_{\mathcal{G}}(V_i))\big|_{A=a_0 \text{ if } V_i \in \mathbb{M}} \\
& \qquad - \sum_{A \cup \{\succeq M_i\}} Y \times \prod_{\substack{V_i \in \mathbb{L} \\ \{\succeq M_i\}}} p(V_i \mid \mathrm{mp}_{\mathcal{G}}(V_i))\big|_{A=a_0 \text{ if } V_i \in \mathbb{M}} \Big) \Big\} \\
& + \sum_{L_i \in \mathbb{L} \setminus A} \Big\{ \frac{\prod_{M_i \prec L_i} p(M_i \mid \mathrm{mp}_{\mathcal{G}}(M_i))\big|_{A=a_0}}{\prod_{M_i \prec L_i} p(M_i \mid \mathrm{mp}_{\mathcal{G}}(M_i))} \times \Big( \sum_{\{\succ L_i\}} Y \times \prod_{V_i \succ L_i} p(V_i \mid \mathrm{mp}_{\mathcal{G}}(V_i))\big|_{A=a_0 \text{ if } V_i \in \mathbb{M}} \\
& \qquad - \sum_{\{\succeq L_i\}} Y \times \prod_{V_i \succeq L_i} p(V_i \mid \mathrm{mp}_{\mathcal{G}}(V_i))\big|_{A=a_0 \text{ if } V_i \in \mathbb{M}} \Big) \Big\} \\
& + \sum_{V \setminus \{A, \mathbb{C}\}} Y \times \prod_{M_i \in \mathbb{M}} p\left(M_i \mid \mathrm{mp}_{\mathcal{G}}\left(M_i\right)\right)\big|_{A=a_0} \times \prod_{L_i \in \mathbb{L} \setminus A} p\left(L_i \mid \mathrm{mp}_{\mathcal{G}}\left(L_i\right)\right) - \psi(t),
\end{aligned}
$$

- **plugin estimator:** $\tau : A \to Z_1 \to \cdots \to Z_k \to Y$

$$\psi(t) = E\,(\mathbb{I}(A = a_0)\,Y) + \mathbb{E}\Bigg\{\underbrace{\mathbb{E}\Big[\cdots \underbrace{\mathbb{E}\Big[\mu \mid \mathrm{mp}_{\mathcal{G}}^{-A}(Z_k), a_{Z_k}\Big]}_{B_k} \cdots \mid \mathrm{mp}_{\mathcal{G}}^{-A}(Z_1), a_{Z_1}\Big]}_{B_1}\Bigg\},$$

where $a_{Z_k} = a_0$ if $Z_k \in \mathbb{M}$ and $= 1 - a_0$ if $Z_k \in \mathbb{L}$.

- **onestep estimator**

$$Q = \{\mu, \pi, B_1, \cdots, B_k, f^r_{Z_1}, \cdots, f^r_{Z_k}\}, \quad f^r_{Z_k} = \frac{p(Z_k \mid \mathrm{mp}_{\mathcal{G}}^{-A}(Z_k), a_{Z_k})}{p(Z_k \mid \mathrm{mp}_{\mathcal{G}}^{-A}(Z_k), 1 - a_{Z_k})}$$

- **TMLE estimator**
  order of updating nuisances: $\pi^{(t)}$, $\mu^{(t)}, B_k^{(t)}$, $\cdots$, $B_1^{(t)}, \cdots$

# Simulations

We conducted extensive simulation studies, under various types of mediators (binary, continuous, multivariate):

- Confirming theoretical properties of our proposed estimators.
- Comparing TMLE vs. one-step in settings with weak overlap (near positivity violations).
- Performances under misspecified parametric models vs. flexible estimation (using a super learner).
- Impact of cross-fitting on proposed estimators (using random forests).

# Open-source package and software

- `fdtmle` package in R
  - Conducting causal inference using the front-door criterion
  - https://github.com/annaguo-bios/fdtmle
- `ADMGtmle` package in R
  - Conducting causal inference in graphical models with unmeasured variables via the extension of the front-door criterion
  - https://github.com/annaguo-bios/ADMGtmle

# Concluding Remarks

**Summary:** We proposed an estimation scheme for ACE using onestep and TMLE estimator that avoids density estimation based on the front-door model and its extension.

# Concluding Remarks

**Summary:** We proposed an estimation scheme for ACE using onestep and TMLE estimator that avoids density estimation based on the front-door model and its extension.

**Future work**

- Assumption violation and sensitivity analysis

**Summary:** We proposed an estimation scheme for ACE using onestep and TMLE estimator that avoids density estimation based on the front-door model and its extension.

**Future work**

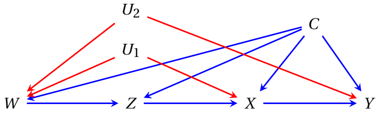
- Assumption violation and sensitivity analysis
- Extension to **semiparametric** graphical models that encodes regular independence constrain and/or interventional constrain (Verma constrain)

# Concluding Remarks

**Summary:** We proposed an estimation scheme for ACE using onestep and TMLE estimator that avoids density estimation based on the front-door model and its extension.

**Future work**

- Assumption violation and sensitivity analysis
- Extension to **semiparametric** graphical models that encodes regular independence constrain and/or interventional constrain (Verma constrain)
- Extension to graphical models that can **not be** identified via either backdoor criterion or front-door criterion




"Napkin graph"

# Targeted Machine Learning for Average Causal Effect Estimation Using the Front-Door Functional

**Anna Guo,[1] David Benkeser[1] and Razieh Nabi[1,*]**

[1]Department of Biostatistics and Bioinformatics, Rollins School of Public Health, Emory University, Atlanta, GA, USA

*Corresponding author. razieh.nabi@emory.edu

SCAN ME

(arXiv:2312.10234)

# Q&A

Anna Guo, Ph.D. candidate (she/her/hers)
Department of Biostatistics and Bioinformatics
Emory University
@AnnaGuo617
anna.guo@emory.edu
https://guoanna.com

# References I


Bhattacharya, R., Nabi, R., and Shpitser, I. (2022). Semiparametric inference for causal effects in graphical models with hidden variables. *Journal of Machine Learning Research*, 23(295):1–76.

Fulcher, I. R., Shpitser, I., Marealle, S., and Tchetgen Tchetgen, E. J. (2020). Robust inference on population indirect causal effects: the generalized front door criterion. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 82(1):199–214.

Pearl, J. (1995). Causal diagrams for empirical research. *Biometrika*, 82(4):669–688.

Tian, J. and Pearl, J. (2002). A general identification condition for causal effects. In *Aaai/iaai*, pages 567–573.

Van der Laan, M. J., Rose, S., et al. (2011). *Targeted learning: causal inference for observational and experimental data*, volume 4. Springer.

For a TMLE $\psi_2(\hat{Q}^*)$ of $\psi(Q)$,

$$\psi(\hat{Q}^*) - \psi(Q) = P_n\Phi(Q) - P_n\Phi(\hat{Q}^*) + (P_n - P)\left\{\Phi(\hat{Q}^*) - \Phi(Q)\right\} + R_2(\hat{Q}^*, Q) \ .$$

In order to establish asymptotic linearity of the TMLE, we will require

(A1) *Donsker estimates*: $\Phi(\hat{Q}^*) - \Phi(Q)$ falls in a $P$-Donsker class with probability tending to 1 ;

(A2) $L^2(P)$*-consistent influence function estimates*: $P\{\Phi(\hat{Q}^*) - \Phi(Q)\}^2 = o_P(1)$ ;

(A3) *Successful targeting of nuisance parameters*: $P_n\Phi(\hat{Q}^*) = o_P(n^{-1/2})$ .

(A4) *Bounded nuisance estimates*: for all $a, m, x$, $\hat{\pi}^*(a \mid x) > \delta_1$ for some $\delta_1 > 0$ and $\hat{\lambda}(a \mid m, x) > \delta_2$ for some $\delta_2 > 0$.

# Appendix - Simulation1

Front-door model with bivariate mediator $M$

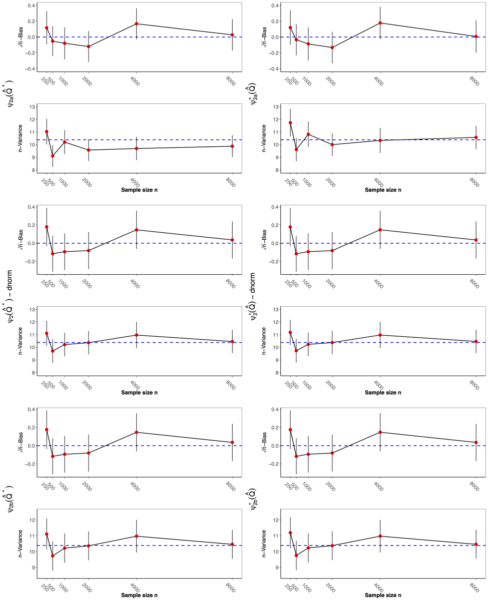


**Table 1.** Comparative analysis of TMLEs and one-step estimators under violation of the positivity assumption.

| | Univariate Binary | | Univariate Continuous | | | | | | Bivariate Continuous | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | $\psi_1(\hat{Q}^\star)$ | $\psi_1^+(\hat{Q})$ | $\psi_1(\hat{Q}^\star)$ | $\psi_1^+(\hat{Q})$ | $\psi_{2a}(\hat{Q}^\star)$ | $\psi_{2a}^+(\hat{Q})$ | $\psi_{2b}(\hat{Q}^\star)$ | $\psi_{2b}^+(\hat{Q})$ | $\psi_{2a}(\hat{Q}^\star)$ | $\psi_{2a}^+(\hat{Q})$ | $\psi_{2b}(\hat{Q}^\star)$ | $\psi_{2b}^+(\hat{Q})$ |
| **n=500** | | | | | | | | | | | | |
| Bias | -0.004 | -0.010 | -0.022 | -0.004 | -0.002 | 0.000 | -0.002 | -0.012 | -0.012 | 0.153 | -0.031 | -0.065 |
| SD | 0.078 | 0.418 | 0.135 | 0.799 | 0.432 | 2.524 | 0.405 | 1.191 | 0.610 | 5.096 | 0.495 | 1.447 |
| MSE | 0.006 | 0.174 | 0.019 | 0.638 | 0.187 | 6.363 | 0.164 | 1.418 | 0.372 | 25.965 | 0.245 | 2.097 |
| CI coverage | 91.2% | 95.4% | 96.6% | 95.2% | 98.4% | 97.1% | 98.3% | 97.3% | 99.4% | 98.2% | 98.5% | 97.7% |
| CI width | 0.317 | 0.854 | 1.533 | 1.531 | 4.764 | 5.705 | 2.720 | 3.447 | 10.115 | 12.100 | 2.854 | 3.834 |
| **n=1000** | | | | | | | | | | | | |
| Bias | 0.000 | -0.002 | -0.012 | -0.018 | -0.004 | 0.041 | -0.003 | 0.020 | -0.015 | -0.078 | -0.003 | -0.001 |
| SD | 0.056 | 0.207 | 0.101 | 0.470 | 0.342 | 1.394 | 0.338 | 0.787 | 0.389 | 1.841 | 0.333 | 0.716 |
| MSE | 0.003 | 0.043 | 0.010 | 0.221 | 0.117 | 1.942 | 0.114 | 0.619 | 0.152 | 3.391 | 0.111 | 0.513 |
| CI coverage | 92.1% | 95.4% | 96% | 94.3% | 98.5% | 96.3% | 98% | 97.1% | 99.4% | 97.1% | 99% | 96.4% |
| CI width | 0.240 | 0.492 | 0.931 | 0.930 | 3.071 | 3.460 | 1.861 | 2.178 | 4.809 | 5.365 | 1.852 | 2.136 |
| **n=2000** | | | | | | | | | | | | |
| Bias | 0.000 | -0.002 | -0.005 | 0.010 | 0.009 | 0.010 | 0.009 | 0.014 | 0.003 | -0.006 | 0.008 | 0.022 |
| SD | 0.039 | 0.114 | 0.068 | 0.239 | 0.238 | 0.699 | 0.243 | 0.481 | 0.319 | 0.980 | 0.276 | 0.489 |
| MSE | 0.001 | 0.013 | 0.005 | 0.057 | 0.057 | 0.488 | 0.059 | 0.231 | 0.102 | 0.959 | 0.076 | 0.240 |
| CI coverage | 94.1% | 96.2% | 97.4% | 96% | 99.2% | 96.9% | 98.7% | 96% | 99.2% | 96.9% | 98.6% | 97.4% |
| CI width | 0.175 | 0.318 | 0.602 | 0.602 | 1.960 | 2.092 | 1.321 | 1.454 | 2.989 | 3.209 | 1.351 | 1.504 |

**Table 2.** Comparative analysis of TMLEs and one-step estimators under model misspecifications.

| | **TMLEs** | | | | | | | | | **One-step estimators** | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | *Univariate Binary* | | | *Univariate Continuous* | | | | | | *Univariate Binary* | | | *Univariate Continuous* | | | | | |
| | $\psi_1(\hat{Q}^\star)$ | | | $\psi_{2a}(\hat{Q}^\star)$ | | | $\psi_{2b}(\hat{Q}^\star)$ | | | $\psi_1^+(\hat{Q})$ | | | $\psi_{2a}^+(\hat{Q})$ | | | $\psi_{2b}^+(\hat{Q})$ | | |
| | Linear | SL | CF | Linear | SL | CF | Linear | SL | CF | Linear | SL | CF | Linear | SL | CF | Linear | SL | CF |
| **n=500** | | | | | | | | | | | | | | | | | | |
| Bias | -0.016 | -0.001 | -0.010 | -0.081 | -0.020 | -0.037 | -0.081 | -0.016 | -0.038 | -0.017 | -0.008 | -0.005 | -0.081 | -0.021 | -0.039 | -0.081 | -0.016 | -0.037 |
| SD | 0.043 | 0.050 | 0.071 | 0.099 | 0.123 | 0.128 | 0.099 | 0.116 | 0.123 | 0.043 | 0.048 | 0.183 | 0.099 | 0.128 | 0.133 | 0.099 | 0.115 | 0.126 |
| MSE | 0.002 | 0.003 | 0.005 | 0.016 | 0.016 | 0.018 | 0.016 | 0.014 | 0.016 | 0.002 | 0.002 | 0.033 | 0.016 | 0.017 | 0.019 | 0.016 | 0.014 | 0.017 |
| CI coverage | 84.2% | 83.2% | 82.8% | 85.5% | 97% | 96.8% | 85.5% | 91.5% | 91.8% | 83.1% | 80% | 81.5% | 85.5% | 96.8% | 96.5% | 85.5% | 91.4% | 91.4% |
| CI width | 0.161 | 0.154 | 0.172 | 0.398 | 0.567 | 0.596 | 0.399 | 0.398 | 0.444 | 0.158 | 0.143 | 0.176 | 0.399 | 0.560 | 0.589 | 0.399 | 0.397 | 0.444 |
| **n=1000** | | | | | | | | | | | | | | | | | | |
| Bias | -0.018 | -0.003 | -0.008 | -0.081 | -0.012 | -0.027 | -0.081 | -0.009 | -0.023 | -0.018 | -0.006 | -0.008 | -0.081 | -0.013 | -0.029 | -0.081 | -0.009 | -0.023 |
| SD | 0.030 | 0.035 | 0.035 | 0.074 | 0.088 | 0.089 | 0.074 | 0.088 | 0.089 | 0.030 | 0.034 | 0.035 | 0.074 | 0.092 | 0.092 | 0.074 | 0.087 | 0.089 |
| MSE | 0.001 | 0.001 | 0.001 | 0.012 | 0.008 | 0.009 | 0.012 | 0.008 | 0.008 | 0.001 | 0.001 | 0.001 | 0.012 | 0.009 | 0.009 | 0.012 | 0.008 | 0.008 |
| CI coverage | 81.5% | 87.3% | 85.3% | 74.6% | 98.2% | 97.2% | 74.6% | 90.1% | 89.9% | 80.8% | 83.6% | 84.2% | 74.6% | 96.8% | 96.6% | 74.6% | 90.3% | 89.8% |
| CI width | 0.111 | 0.113 | 0.117 | 0.282 | 0.403 | 0.416 | 0.282 | 0.293 | 0.311 | 0.109 | 0.106 | 0.110 | 0.282 | 0.400 | 0.412 | 0.282 | 0.292 | 0.310 |
| **n=2000** | | | | | | | | | | | | | | | | | | |
| Bias | -0.018 | -0.002 | -0.005 | -0.084 | -0.008 | -0.019 | -0.084 | -0.005 | -0.016 | -0.018 | -0.004 | -0.005 | -0.084 | -0.008 | -0.018 | -0.084 | -0.005 | -0.016 |
| SD | 0.020 | 0.023 | 0.024 | 0.050 | 0.060 | 0.059 | 0.050 | 0.060 | 0.059 | 0.020 | 0.023 | 0.023 | 0.050 | 0.062 | 0.061 | 0.050 | 0.060 | 0.059 |
| MSE | 0.001 | 0.001 | 0.001 | 0.010 | 0.004 | 0.004 | 0.010 | 0.004 | 0.004 | 0.001 | 0.001 | 0.001 | 0.010 | 0.004 | 0.004 | 0.010 | 0.004 | 0.004 |
| CI coverage | 76.9% | 89.7% | 88.4% | 60.5% | 97.9% | 98% | 60.4% | 92.2% | 92.5% | 75.4% | 87.2% | 87.4% | 60.5% | 97.3% | 97.6% | 60.4% | 92.1% | 92.3% |
| CI width | 0.077 | 0.083 | 0.084 | 0.198 | 0.288 | 0.293 | 0.198 | 0.214 | 0.222 | 0.076 | 0.079 | 0.081 | 0.198 | 0.286 | 0.291 | 0.198 | 0.213 | 0.221 |

**Table 3.** Comparative analysis for the impact of cross-fitting on TMLEs and one-step estimators in conjunction with the use of random forests. RF refers to random forest with 500 trees and a minimum node size of 5 for a continuous variable and 1 for binary, and CF denotes random forest with cross fitting using 5 folds.

| | **TMLEs** | | | | | | **One-step estimators** | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | *Univariate Binary* | | *Univariate Continuous* | | | | *Univariate Binary* | | *Univariate Continuous* | | | |
| | $\psi_1(\hat{Q}^\star)$ | | $\psi_{2a}(\hat{Q}^\star)$ | | $\psi_{2b}(\hat{Q}^\star)$ | | $\psi_1^+(\hat{Q})$ | | $\psi_{2a}^+(\hat{Q})$ | | $\psi_{2b}^+(\hat{Q})$ | |
| | RF | CF | RF | CF | RF | CF | RF | CF | RF | CF | RF | CF |
| **n=500** | | | | | | | | | | | | |
| Bias | -0.162 | -0.020 | -0.312 | 0.055 | -0.486 | 0.017 | -0.103 | -0.028 | 0.009 | 0.066 | -0.492 | 0.014 |
| SD | 0.166 | 0.140 | 0.372 | 0.331 | 0.369 | 0.285 | 0.051 | 0.128 | 0.432 | 0.318 | 0.373 | 0.286 |
| MSE | 0.054 | 0.020 | 0.235 | 0.113 | 0.373 | 0.081 | 0.013 | 0.017 | 0.186 | 0.105 | 0.381 | 0.082 |
| CI coverage | 17.4% | 82.8% | 48.8% | 86.9% | 36.1% | 87.3% | 18.8% | 86.3% | 56.7% | 87.6% | 35.5% | 87% |
| CI width | 0.128 | 0.389 | 0.681 | 0.980 | 0.717 | 0.862 | 0.119 | 0.388 | 0.682 | 0.977 | 0.718 | 0.861 |
| **n=1000** | | | | | | | | | | | | |
| Bias | -0.162 | -0.016 | -0.329 | 0.054 | -0.490 | 0.008 | -0.100 | -0.021 | -0.017 | 0.059 | -0.497 | 0.005 |
| SD | 0.114 | 0.096 | 0.252 | 0.212 | 0.267 | 0.221 | 0.040 | 0.091 | 0.286 | 0.215 | 0.271 | 0.221 |
| MSE | 0.039 | 0.009 | 0.172 | 0.048 | 0.312 | 0.049 | 0.012 | 0.009 | 0.082 | 0.049 | 0.320 | 0.049 |
| CI coverage | 13.3% | 88.5% | 30.1% | 88.6% | 19.5% | 86.6% | 12.4% | 89.7% | 52.4% | 88.3% | 18.3% | 87.1% |
| CI width | 0.101 | 0.315 | 0.417 | 0.690 | 0.520 | 0.656 | 0.098 | 0.315 | 0.420 | 0.689 | 0.520 | 0.655 |
| **n=2000** | | | | | | | | | | | | |
| Bias | -0.161 | -0.010 | -0.326 | 0.063 | -0.473 | 0.019 | -0.096 | -0.013 | -0.041 | 0.065 | -0.479 | 0.016 |
| SD | 0.083 | 0.074 | 0.176 | 0.148 | 0.186 | 0.164 | 0.034 | 0.072 | 0.197 | 0.150 | 0.189 | 0.164 |
| MSE | 0.033 | 0.006 | 0.137 | 0.026 | 0.259 | 0.027 | 0.010 | 0.005 | 0.041 | 0.027 | 0.265 | 0.027 |
| CI coverage | 7.8% | 90.4% | 14.4% | 89.8% | 6.4% | 86.5% | 8.9% | 90.7% | 56.6% | 88.9% | 6.3% | 86.5% |
| CI width | 0.081 | 0.246 | 0.292 | 0.520 | 0.376 | 0.499 | 0.080 | 0.246 | 0.294 | 0.519 | 0.376 | 0.499 |