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

Anna Guo, David Benkeser, and Razieh Nabi

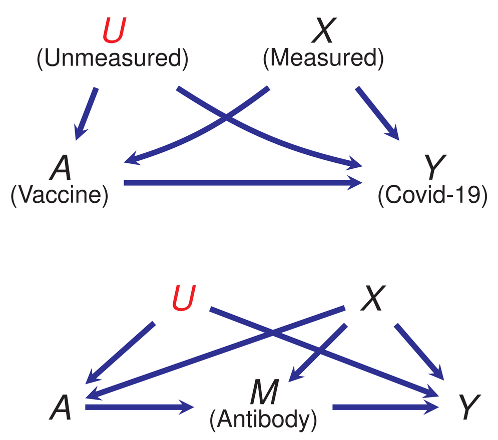
Dept. of Biostatistics & Bioinformatics, Emory University

EMORY UNIVERSITY

Causal Inference in the Presence of Unmeasured Confounders

Motivation.

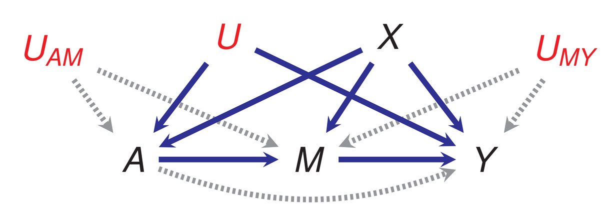
- Average Causal Effect (ACE) $:= \mathbb{E}[Y^1 - Y^0]$
- In the presence of unmeasured confounders, ACE is not identified.
- The front-door model [2] offers an alternative strategy to identify ACE.
- This work focuses on providing a flexible and robust estimation framework for ACE using the front-door functional.



(front-door model)

Target parameter. $\psi := \mathbb{E}[Y^{a_0}],\ a_0 \in \{0,1\}$

Identification assumptions.



(Encoded 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 $A = a$ and $Y^m = Y$ when $M = m$;
4. Positivity: $P(A=1 \mid X=x) > 0,\ P(M=m \mid A=a, X=x) > 0,\ \forall a, m, x$.

Identification functional for the target parameter.

- Let $P(O) = P(X, A, M, Y)$ denote observed data distribution.
- Let $\mu(m,a,x) = \mathbb{E}_P[Y \mid m, a, x]$, and $\pi(a \mid x) = P(a \mid x)$, and $p_{M|A,X}(m \mid a_0, x) = P(M = m \mid A = a_0, X = x)$ and $p_X(x) = P(X = x)$.

$$\psi(P) = \iint \sum_{a=0}^{1} \mu(m,a,x)\, \pi(a \mid x)\, p_{M|A,X}(m \mid a_0, x)\, p_X(x)\, dm\, dx \quad \textbf{(target estimand)}$$

Existing estimation strategies.

- Let $Q = \{\mu, \pi, p_{M|A,X}\}$ contain the nuisance functionals.
- Plug-in estimator: $\psi(\hat{Q}) = \frac{1}{n}\sum_{i=1}^{n}\sum_{m}\sum_{a=0}^{1} \hat{\mu}(m,a,X_i)\, \hat{\pi}(a \mid X_i)\, \hat{p}_{M|A,X}(m \mid a_0, X_i)$.
- **First order bias**: $\psi(\hat{Q}) = \psi(Q) - P\Phi(\hat{Q}) + R_2(\hat{Q}, Q)$.
- **Efficient influence function:**

$$\Phi(Q)(O_i) = \underbrace{\frac{p_{M|A,X}(a_0, X_i)}{p_{M|A,X}(A_i, X_i)}\{Y_i - \mu(M_i, A_i, X_i)\}}_{\Phi_Y(Q)(O_i)} + \underbrace{\frac{\mathbb{I}(A_i = a_0)}{\pi(a_0 \mid X_i)}\{\xi(M_i, X_i) - \theta(X_i)\}}_{\Phi_M(Q)(O_i)}$$
$$+ \underbrace{\{\eta(1, X_i) - \eta(0, X_i)\}\{A_i - \pi(1 \mid X_i)\}}_{\Phi_A(Q)(O_i)} + \underbrace{\theta(X_i) - \psi(Q)}_{\Phi_X(Q)(O_i)}. \qquad (1)$$

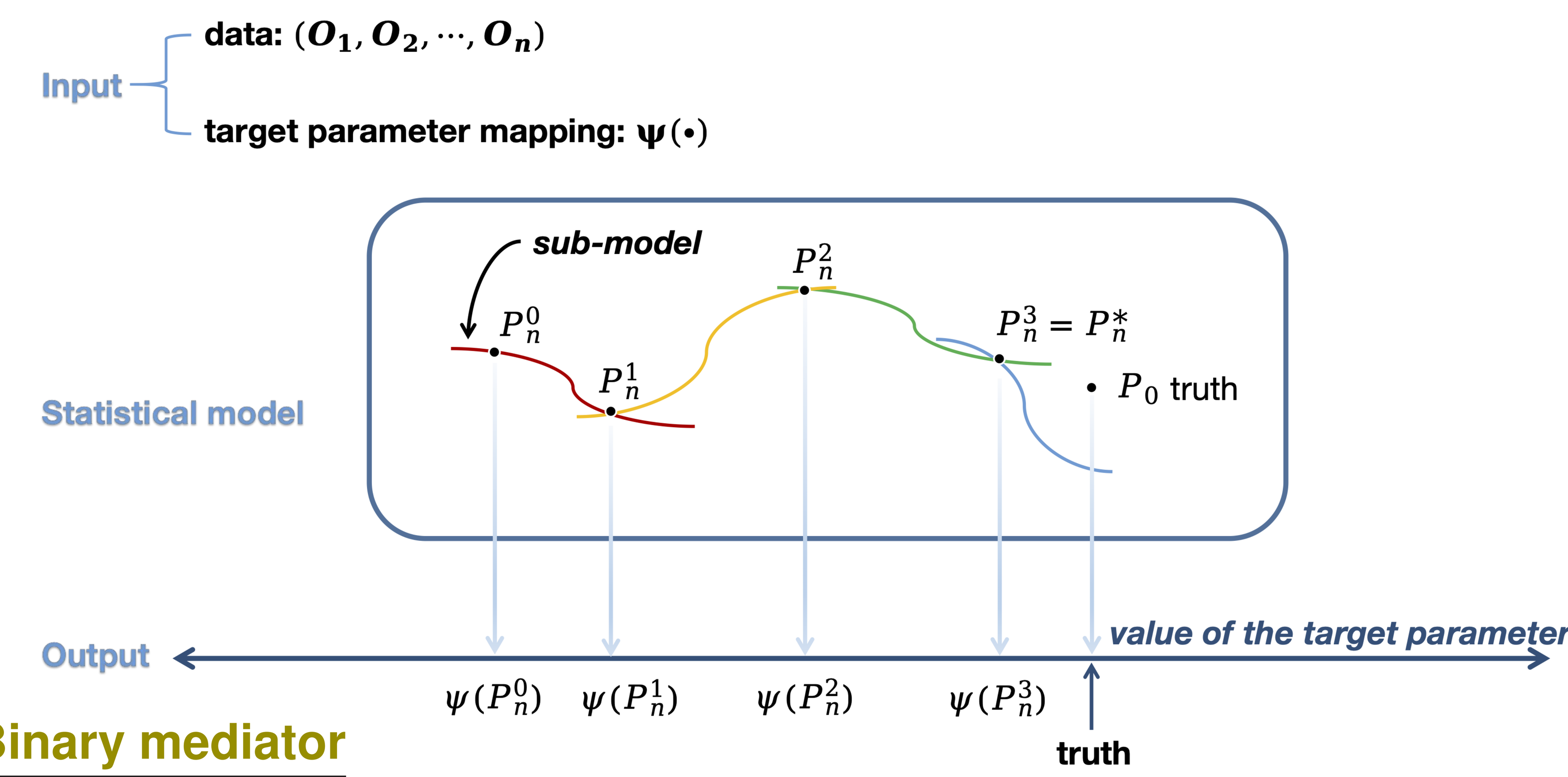
where $\xi(M,X) = \sum_{a=0}^{1} \mu(M,a,X)\pi(a \mid X)$, $\eta(A,X) = \int \mu(m,A,X) p_{M|A,X}(m \mid a_0, X)\, dm$, $\theta(X) = \int \xi(m,X) p_{M|A,X}(m \mid a_0, X)\, dm$.

- **Doubly robust one-step estimator** [1]: $\psi^{+}(\hat{Q}) = \psi(\hat{Q}) + P_n\Phi(\hat{Q})$.

Our approach: Targeted Minimum Loss Based Estimation (TMLE).

- Update $\hat{Q} \Longrightarrow Q_n^*$ such that $P_n\Phi(Q_n^*) \approx 0$.

Targeted Minimum Loss Based Estimation (TMLE) Procedure



Binary mediator

1. **Obtain initial nuisance estimates**: $\pi^{(0)}(A \mid X), p_{M|A,X}^{(0)}$ and $\mu^{(0)}(M,A,X)$.
2. **Define loss functions and parametric fluctuations**:
   - loss function $L(\widetilde{Q})(O)$: $P = \arg\min_{\widetilde{P}} PL(\widetilde{P})(O)$.
   - parametric submodel $\{P_{n,\epsilon}^0 : \epsilon \in \mathbb{R}\}$: $P_{n,\epsilon=0}^0 = P_n^0$, $\frac{d}{d\epsilon}L(P_{n,\epsilon}^0)\big|_{\epsilon=0} = \Phi(P_n^0)$.

$$L(p_{M|A,X})(A,X) = -\mathbb{I}(A = a_0)\{M \log p_{M|A,X}(1 \mid a_0, X) + (1-M)\log(1 - p_{M|A,X}(1 \mid a_0, X)\}$$
$$p_{M|A,X}(1 \mid a_0, X; \varepsilon_m) = \text{expit}\left\{\text{logit}\, p_{M|A,X}(1 \mid a_0, X) + \varepsilon_m \frac{1}{\pi(a_0 \mid X)}\big(\xi(1,X) - \xi(0,X)\big)\right\},\ \varepsilon_m \in \mathbb{R}$$

3. **Update nuisance estimates by solving optimization problem**:

$$\varepsilon_m^{(t)} = \arg\min_{\varepsilon_m} \sum_{i=1}^{n} L(p_{M|A,X})(A_i, X_i; \varepsilon_m).$$

   - Iterative update of $\pi(A \mid X)$ and $p_{M|A,X}$:
   – Define auxiliary covariates $H_m^{(t)}$:

$$H_m^{(t)} := \frac{1}{\pi^{(t)}(a_0 \mid X)}\big(\xi^{(t)}(1,X) - \xi^{(t)}(0,X)\big).$$

   – Then fit the following logistic regressions without an intercept:

$$M \sim \text{offset}\left(\text{logit}\, p_{M|A,X}^{(t)}(1 \mid a_0, X)\right) + H_m^{(t)}.$$

   – Repeat the optimization step until $\varepsilon_m^{(t)} = 0$.
   - Update $\mu(M,A,X)$ in one step.
4. **Return** $\psi^1(Q_n^*) = P_n[\theta(X; \mu^{(1)}, p_{M|A,X}^{(t)}, \pi^{(t)}]$ as the TMLE estimator.

Continuous mediator

- Density estimation for $p_{M|A,X}$ is needed.
- Optimization for $p_{M|A,X}$ can no longer be solved by regression.

Multivariate mediators

- Density estimation for $p_{M|A,X}$ is computational intensive.
- TMLE **targeting** $\theta(X)$ instead of $p_{M|A,X}$:
- Density ratio estimation: (i) nonparametric estimation; (ii) regression

$$p_{M|A,X}^r(M \mid A, X) := \frac{p_{M|A,X}(a_0, X)}{p_{M|A,X}(A, X)} = \frac{P(A = a_0 \mid X, M)}{P(A \mid X, M)} \times \frac{P(A \mid X)}{P(A = a_0 \mid X)}.$$

- Equally applicable to **continuous** mediators.

Asymptotic Behaviours and Robustness Properties

Binary & Continuous mediators

Assume nuisance estimates have the convergence rates as follows:

$$\left\{\int [\widehat{\pi}(1 \mid x) - \pi(1 \mid x)]^2\, dP(x)\right\}^{1/2} = o_P(n^{-\frac{1}{k}}),$$
$$\left\{\int [\widehat{p}_{M|A,X}(m \mid a, x) - p_{M|A,X}(m \mid a, x)]^2\, dP(x,a,m)\right\}^{1/2} = o_P(n^{-\frac{1}{b}}),$$
$$\left\{\int [\widehat{\mu}(m,a,x) - \mu(m,a,x)]^2\, dP(x,a,m)\right\}^{1/2} = o_P(n^{-\frac{1}{q}}).$$

Under standard regularity conditions, we have

$$R^2(P_n^*, P) \le o_P\left(n^{\max\left\{-\left(\frac{1}{b}+\frac{1}{q}\right), -\left(\frac{1}{b}+\frac{1}{k}\right)\right\}}\right).$$

- Asymptotically efficient if following nuisances combinations achieve $o_P(n^{-\frac{1}{2}})$ convergence:

$$(i)\ p_{M|A,X}(M \mid A, X), \qquad (ii)\ \{\mu(M,A,X),\ \pi(A \mid X)\}.$$

- All nuisances converge to the respective truth at a slower rate of $o_P(n^{-\frac{1}{4}})$.

Multivariate mediators

- Asymptotically efficient if following nuisances combinations achieve $o_P(n^{-\frac{1}{2}})$ convergence:

$(i)\ \{\pi(A \mid X),\ \mu(M,A,X)\}$, $\quad (ii)\ \{\theta(X; \widehat{\xi}),\ \eta(a^*, X; \widehat{\mu}),\ \mu(M,A,X)\}$
$(iii)\ \{\pi(A \mid X),\ p_{M|A,X}^r(M \mid A, X)\}$, $\quad (iv)\ \{\theta(X; \widehat{\xi}),\ \eta(a^*, X; \widehat{\mu}),\ p_{M|A,X}^r(M \mid A, X)\}$.

- All nuisances converge to the respective truth at a slower rate of $o_P(n^{-\frac{1}{4}})$.

Simulations

Table 1. Comparison between TMLE estimators and one-step EIF estimator under binary, continuous, and multivariate mediators.

	Proposed TMLE estimators			One-step EIF estimator		
	Binary M	*Continuous M*	*Multivariate M*	*Binary M*	*Continuous M*	*Multivariate M*
ATE bias (SD)	0 (0.034)	0 (0.106)	-0.004 (0.153)	0 (0.034)	0 (0.105)	0.213 (7.771)
E(Y^1) bias (SD)	-0.001 (0.054)	0.001 (0.092)	0 (0.114)	-0.001 (0.054)	0.001 (0.091)	0.076 (5.142)
E(Y^0) bias (SD)	-0.001 (0.058)	0.001 (0.091)	0.005 (0.124)	-0.001 (0.058)	0.001 (0.091)	-0.137 (5.917)

Note: ψ^1 is adopted under binary and continuous mediators, and ψ^2 is adopted under multivariate mediator.

Positivity violation

Table 2. Comparison between TMLE estimators and one-step EIF estimator under binary, continuous, and multivariate mediators in the presence of weak overlapping.

	Proposed TMLE estimators			One-step EIF estimator		
	Binary M	*Continuous M*	*Multivariate M*	*Binary M*	*Continuous M*	*Multivariate M*
ATE bias (SD)	0.006 (0.068)	-0.003 (0.719)	0.008 (0.85)	0.024 (1.024)	-0.089 (1.015)	-2.41 (49.788)
E(Y^1) bias (SD)	0 (0.05)	0.006 (0.241)	0.01 (0.248)	-0.002 (0.064)	0.002 (0.133)	0.116 (10.027)
E(Y^0) bias (SD)	-0.006 (0.078)	0.009 (0.682)	0.002 (0.811)	-0.026 (1.023)	0.091 (1.009)	2.526 (49.481)

Note: ψ^1 is adopted under binary mediator, and ψ^2 is adopted under continuous and multivariate mediators.

References

[1] Isabel R. Fulcher, Ilya Shpitser, Stella Marealle, and Eric J. Tchetgen Tchetgen. Robust inference on population indirect causal effects: The generalized front-door criterion. *Journal of the Royal Statistical Society, Series B*, 2019.

[2] Judea Pearl. Causal diagrams for empirical research. *Biometrika*, 82(4):669–688, 1995.