Estimation of natural indirect effects robust to unmeasured confounding and mediator measurement error

Isabel Rose Fulcher

Joint work with: Xu Shi & Eric Tchetgen Tchetgen

March 26, 2019

Harvard PEPFAR program in Nigeria

- Adult patients with HIV treated with antiretroviral therapy (ART) and followed for at least 1 year

Harvard PEPFAR program in Nigeria

- Adult patients with HIV treated with antiretroviral therapy (ART) and followed for at least 1 year
- Interest may lie in assessing differences in efficacy of ART regimens

$$\text{ART} \longrightarrow \text{Viral Load}$$

Harvard PEPFAR program in Nigeria

- Adult patients with HIV treated with antiretroviral therapy (ART) and followed for at least 1 year
- Interest may lie in assessing differences in efficacy of ART regimens

$$\text{ART} \longrightarrow \text{Viral Load}$$

- Or, the mechanisms responsible for differences in ART effectiveness

$$\begin{tikzcd}
\text{ART} \arrow[r] \arrow[rr, bend left] & \text{Adherence} \arrow[r] & \text{Viral load}
\end{tikzcd}$$

Harvard PEPFAR program in Nigeria

- Adult patients with HIV treated with antiretroviral therapy (ART) and followed for at least 1 year
- Interest may lie in assessing differences in efficacy of ART regimens

$$\text{ART} \longrightarrow \text{Viral Load}$$

- Or, the mechanisms responsible for differences in ART effectiveness

$$\begin{tikzcd} \text{ART} \arrow[r] \arrow[rr, bend left] & \text{Adherence} \arrow[r] & \text{Viral load} \end{tikzcd}$$

What is the mediating role of adherence on the relationship between antiretroviral therapy and viral load?

Mediation Analysis

- Mediation analysis seeks to understand the underlying relationship between an exposure and outcome through an intermediate variable

$$\begin{tikzcd} \text{ART} \arrow[r] \arrow[rr, bend left] & \text{Adherence} \arrow[r] & \text{Viral load} \end{tikzcd}$$

- Beyond evaluating the total effect of the exposure on outcome, one aims to evaluate:
  - **indirect effect** of the exposure on outcome through a given mediator
  - **direct effect** of the exposure on the outcome, not through the mediator

Causal Mediation Analysis

- Natural direct and indirect effects have emerged as the most popular forms of mediation causal effects
- Require stringent assumptions for identification
  - No unmeasured confounding for the exposure-outcome, exposure-mediator, and mediator-outcome relationships
  - No measurement error of the mediator
- Recent work has been devoted to addressing these challenges, but relies on either sensitivity analyses or additional data
- We establish conditions that obviate reliance on these analyses or data collection

Notation

Define counterfactual outcomes for adherence (M) and viral load (Y) under values of ART regimen (A) for each individual

$$
\begin{aligned}
M(a): &\quad M \text{ when intervening to set } A \text{ to } a \\
Y(a, M(a)): &\quad Y \text{ when intervening to set } A \text{ to } a \\
Y(a, M(a^*)): &\quad Y \text{ when intervening to set } A \text{ to } a \text{ and } M \text{ to the value it} \\
&\quad \text{would take when intervening to set } A \text{ to } a^*
\end{aligned}
$$

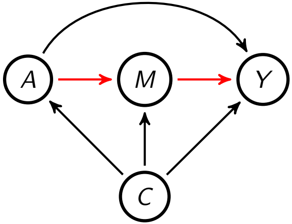
Notation

Define counterfactual outcomes for adherence (M) and viral load (Y) under values of ART regimen (A) for each individual

$$\begin{aligned} M(a): &\quad M \text{ when intervening to set } A \text{ to } a \\ Y(a, M(a)): &\quad Y \text{ when intervening to set } A \text{ to } a \\ Y(a, M(a^*)): &\quad Y \text{ when intervening to set } A \text{ to } a \text{ and } M \text{ to the value it} \\ &\quad \text{would take when intervening to set } A \text{ to } a^* \end{aligned}$$

$$\underbrace{Y(a, M(a)) - Y(a^*, M(a^*))}_{Total\ \ Effect} = \underbrace{Y(a, M(a)) - Y(a, M(a^*))}_{Natural\ \ Indirect\ \ Effect} + \underbrace{Y(a, M(a^*)) - Y(a^*, M(a^*))}_{Natural\ \ Direct\ \ Effect}$$

Identifying assumptions (standard)



M1. Standard consistency assumptions

M2. $M(a^*) \perp A \mid C = c \quad \forall\ a^*, c$

M3. $Y(a, m) \perp M(a^*) \mid A = a, C = c \quad \forall\ m, a, a^*, c$

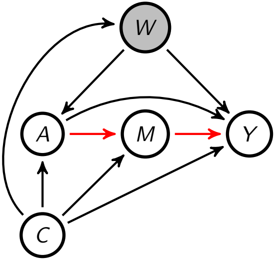
M4. $Y(a, m) \perp A \mid C = c \quad \forall\ m, a, c$

These assumptions allow us to nonparametrically identify the population average NIE,

$$NIE(a, a^*) = E[Y(a, M(a)) - Y(a, M(a^*))]$$

$$\overset{M1-4}{=} \int_c \int_m E[Y|M = m, A = a, C = c](f_M(m|a, c) - f_M(m|a^*, c))f_C(c)dm\ dc$$

Exposure-outcome unmeasured confounding (W)



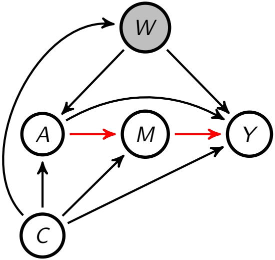
M1. Standard consistency assumptions

M2. $M(a^*) \perp A \mid C = c \quad \forall\ a^*, c$

M3. $Y(a, m) \perp M(a^*) \mid A = a, C = c \quad \forall\ m, a, a^*, c$

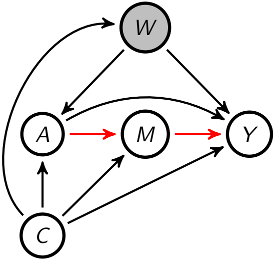
M4. $Y(a, m) \perp A \mid C = c \quad \forall\ m, a, c$

Exposure-outcome unmeasured confounding (W)



$$
\begin{aligned}
\textbf{M4}'. \ &(a)\ Y(a,m) \perp A \mid C = c, W = w \ \forall\ m, a, c, w \\
&(b)\ M \perp W \mid C = c, A = a \ \forall\ c, a \\
&(c)\ E[Y|M = m, a, w, c] \\
&\qquad - E[Y|M = 0, a, w, c] = \gamma_1(a, m, c)
\end{aligned}
$$

Exposure-outcome unmeasured confounding (W)



$$\begin{aligned} \mathbf{M4'}.\ (a)\ & Y(a,m) \perp A \mid C=c, W=w\ \forall\ m,a,c,w \\ (b)\ & M \perp W \mid C=c, A=a\ \forall\ c,a \\ (c)\ & E[Y|M=m,a,w,c] \\ & \quad - E[Y|M=0,a,w,c] = \gamma_1(a,m,c) \end{aligned}$$

Result 1

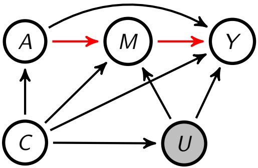
Under assumptions M1-M3 and M4$'$, the natural indirect effect is nonparametrically identified by,

$$NIE(a,a^*) = \int_c \int_m E[Y|M=m, A=a, C=c](f_M(m|a,c) - f_M(m|a^*,c))f_C(c)dm\ dc$$

Implications of Result 1

- The result implies that unmeasured confounding of $A - Y$ relationship can largely be ignored when targeting NIE
- For inference, existing parametric and semiparametric methods to target functional can be used without modification (Pearl, 2001; Vansteelandt & VanderWeele, 2012; Imai et al., 2010; Tchetgen Tchetgen & Shpitser, 2012)
- Importantly, this result only pertains to the NIE contrast as the mean counterfactuals are themselves not nonparameterically identified under these conditions

Mediator-outcome unmeasured confounding (U)



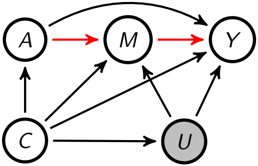
M1. Standard consistency assumptions

M2. $M(a^*) \perp A \mid C = c \quad \forall\ a^*, c$

M3. $Y(a, m) \perp M(a^*) \mid A = a, C = c \quad \forall\ m, a, a^*, c$

M4. $Y(a, m) \perp A \mid C = c \quad \forall\ m, a, c$

Mediator-outcome unmeasured confounding (U)



$\mathbf{M3'}$. (a) $Y(a,m) \perp M(a^*) \mid C = c, U = u \ \forall\ a, a^*, c, u$

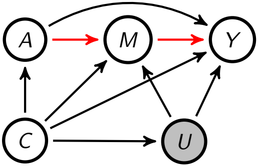
(b) $A \perp U \mid C = c \ \forall\ c$

(c) There is no additive $M - (U, A)$ or $A - U$ interaction in model for $E[Y|A, M, C, U]$

(d) There is no additive $A - U$ interaction in model for $E[M|A, U, C]$

(e) $\text{var}(M|A=1, C) - \text{var}(M|A=0, C) \neq 0$

Mediator-outcome unmeasured confounding (U)



$\mathbf{M3'}$. (a) $Y(a,m) \perp M(a^*) \mid C=c, U=u \;\forall\; a, a^*, c, u$

(b) $A \perp U \mid C=c \;\forall\; c$

(c) There is no additive $M-(U,A)$ or $A-U$ interaction in model for $E[Y|A,M,C,U]$

(d) There is no additive $A-U$ interaction in model for $E[M|A,U,C]$

(e) $\text{var}(M|A=1,C) - \text{var}(M|A=0,C) \neq 0$

M3′(c) and (d) imply the following modeling assumptions (excluding interactions with C for exposition):

$$E[Y|A,M=m,C,U] - E[Y|A,M=0,C,U] = \theta_m m \quad (1)$$

$$E[M|A=a,C,U] - E[M|A=0,C,U] = \beta_a a \quad (2)$$

Mediator-outcome unmeasured confounding

Under assumption M3′ parts (b) through (e), following Lewbel (2012), Tchetgen Tchetgen et al. (2017) established that the average causal effect of M on Y is θ_m identified by,

$$\theta_m = \frac{E\left[\{A - E[A|C]\}\{M - E[M|A, C]\}Y\right]}{E\left[\{A - E[A|C]\}\{M - E[M|A, C]\}M\right]} \tag{3}$$

They refer to equation (3) as the "G-Estimation under No Interaction with Unmeasured Selection" (GENIUS) estimator.

Mediator-outcome unmeasured confounding

Under assumption M3′ parts (b) through (e), following Lewbel (2012), Tchetgen Tchetgen et al. (2017) established that the average causal effect of M on Y is θ_m identified by,

$$\theta_m = \frac{E\left[\{A - E[A|C]\}\{M - E[M|A,C]\}Y\right]}{E\left[\{A - E[A|C]\}\{M - E[M|A,C]\}M\right]} \tag{3}$$

They refer to equation (3) as the "G-Estimation under No Interaction with Unmeasured Selection" (GENIUS) estimator.

Result 2

Under assumptions M1, M2, M3′, and M4 the natural indirect effect is uniquely identified by

$$NIE(a, a^*) = \theta_m \beta_a (a - a^*)$$

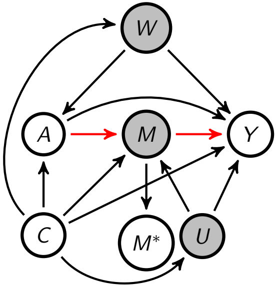
where θ_m is identified by equation (3) and β_a is identified by standard regression of M on A and C.

Combined and extended identification results

The identifying formula in Result 2 equally applies under the additional violations:

Combined and extended identification results

The identifying formula in Result 2 equally applies under the additional violations:



1. Unmeasured confounding of exposure outcome and mediator outcome (violation of M3 and M4)

2. Classical measurement error of the mediator
   - **E1.** $M^* = M + \epsilon$
   - **E2.** $(A, Y, M, C, W, U) \perp \epsilon$
   - **E3.** $E[\epsilon] = 0$

Estimation and inference

- The estimator of θ_m can be consistently estimated by solving the following equation,

$$0 = n^{-1} \sum_{i=1}^{n} \left[h(C_i)\{A_i - \hat{E}[A|C_i]\}\{M_i - \hat{E}[M|A_i, C_i]\}\{Y_i - \hat{\theta}_m M_i\} \right]$$

- The natural indirect effect estimator then follows from the product rule,

$$\widehat{NIE}(a, a^*) = \hat{\theta}_m \hat{\beta}_a (a - a^*)$$

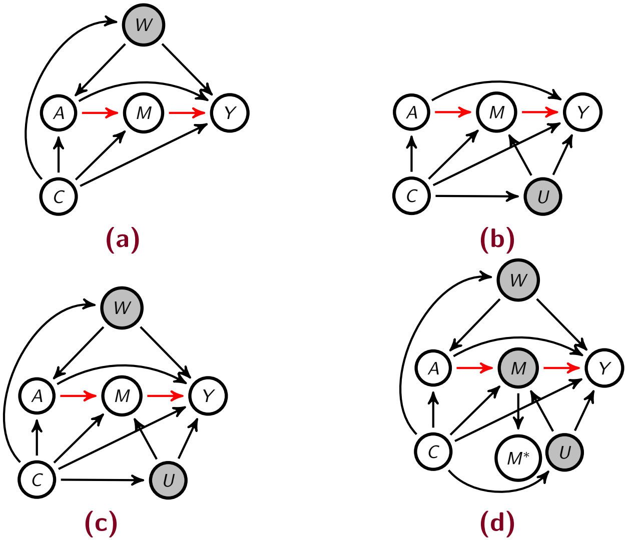
- For inference, the delta method or nonparametric bootstrap can be used to obtain variance estimates for the natural indirect effect

Simulation setup

We report simulation studies we performed under DAGs (a)-(d) in order to illustrate the following key properties:

1. The estimator of the NIE given by the standard mediation formula is unbiased in the presence of unmeasured confounding of the exposure-outcome relationship under M1-M3 and M4$'$ (**Result 1**)
2. The proposed GENIUS estimator of the NIE is robust to unmeasured confounding of the $M - Y$ and $A - Y$ association and measurement error of the mediator (**Result 2 & Extensions**)

Simulation setup



Simulation results

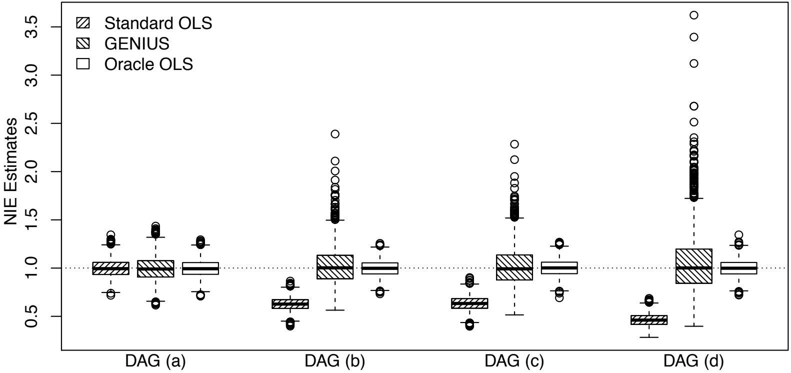


Figure: Boxplots of the estimates over 1000 samples using the standard OLS estimator, the GENIUS estimator, and the oracle OLS estimator under DAGs (a)-(d). True NIE is 1.

Conclusions

- Assumption M4′ is more general than the assumption M4 and yet the identifying formula remains the same
- However, the relaxed condition does not imply robustness against any type of unmeasured exposure-outcome confounding
- In contrast, Result 2 relies on more stringent assumptions, some of which can be empirically verified
- Full paper forthcoming in *Epidemiology* (November 2019) and currently available at:

http://bit.ly/fulcher_med