The generalized front-door criterion for estimation of indirect causal effects of a confounded treatment

Isabel Fulcher, Ilya Shpitser, and Eric Tchetgen Tchetgen

Department of Biostatistics, Harvard University

March 26, 2018

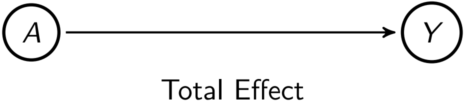
Overview

- Background and motivation
- Nonparametric identification
- Robust estimation
- Simulation study

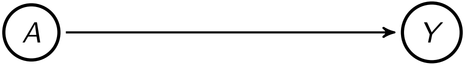
Background: the main goal

The goal of this work is to develop methodology to identify and estimate a novel causal effect, the population intervention indirect effect, in the presence of exposure-outcome unmeasured confounding.

Background: mediation analysis



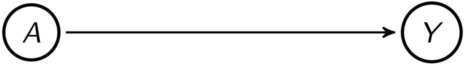
Background: mediation analysis



Total Effect

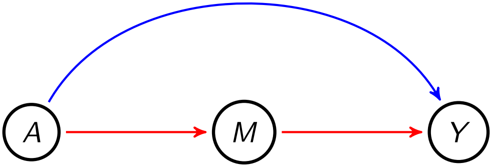
i.e. average causal effect

Background: mediation analysis

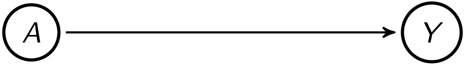


Total Effect

i.e. average causal effect

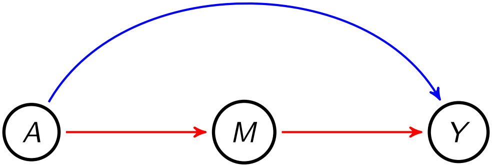


Background: mediation analysis



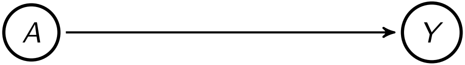
Total Effect

i.e. average causal effect



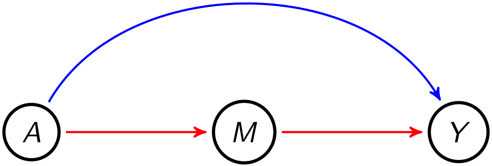
Direct and Indirect Effects

Background: mediation analysis



Total Effect

i.e. average causal effect

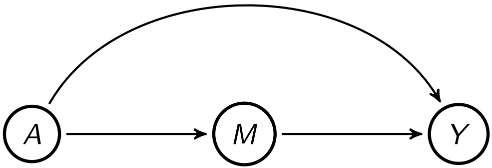


Direct and Indirect Effects

i.e. natural direct and indirect effects

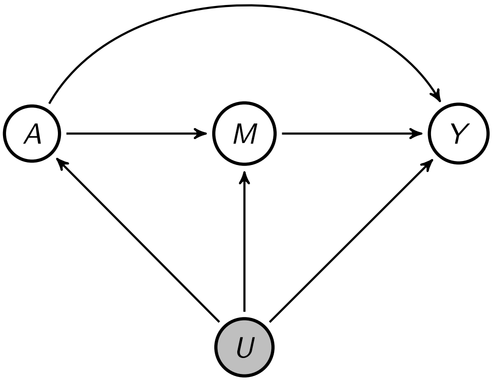
Background: classic mediation setting

Before discussing estimation of these effects, we need *identification*...



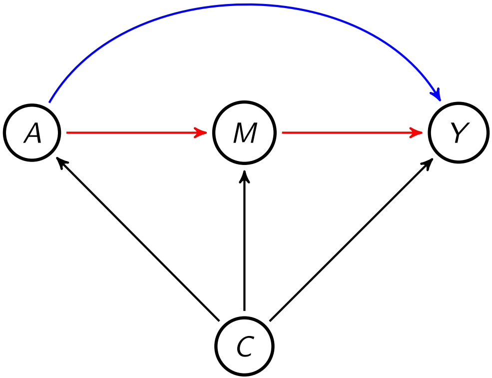
Background: classic mediation setting

Before discussing estimation of these effects, we need *identification*...



Background: classic mediation setting

Before discussing estimation of these effects, we need *identification*...



Background: classic mediation setting

Recent work has been done to relax these assumptions,

- Sensitivity analyses that produce a range of possible values for the effect estimate

Background: classic mediation setting

Recent work has been done to relax these assumptions,

- Sensitivity analyses that produce a range of possible values for the effect estimate
- Most work has focused on randomized exposure

Background: classic mediation setting

Recent work has been done to relax these assumptions,

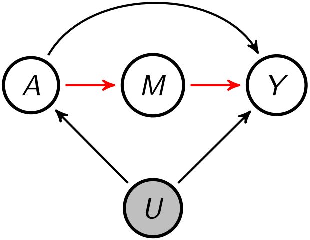
- Sensitivity analyses that produce a range of possible values for the effect estimate
- Most work has focused on randomized exposure
- What if the exposure is not randomized?

Background: classic mediation setting

Recent work has been done to relax these assumptions,

- Sensitivity analyses that produce a range of possible values for the effect estimate
- Most work has focused on randomized exposure
- What if the exposure is not randomized?

We provide nonparametric identification for an indirect effect in the presence of unmeasured confounding of the exposure-outcome relation.



Motivating example: birth surveillance study

- Interest lies in quantifying the extent to which the effect of HIV infection on adverse birth outcomes (AE) operates through maternal medication use during pregnancy (ART)

Motivating example: birth surveillance study

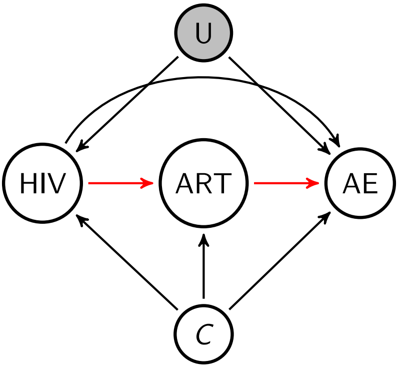
- Interest lies in quantifying the extent to which the effect of HIV infection on adverse birth outcomes (AE) operates through maternal medication use during pregnancy (ART)
- The exposure (HIV) cannot conceivably be randomized

Motivating example: birth surveillance study

- Interest lies in quantifying the extent to which the effect of HIV infection on adverse birth outcomes (AE) operates through maternal medication use during pregnancy (ART)
- The exposure (HIV) cannot conceivably be randomized
- Key behavioral risk factors for HIV-infection and adverse birth outcomes are not collected (U)

Motivating example: birth surveillance study

- Interest lies in quantifying the extent to which the effect of HIV infection on adverse birth outcomes (AE) operates through maternal medication use during pregnancy (ART)
- The exposure (HIV) cannot conceivably be randomized
- Key behavioral risk factors for HIV-infection and adverse birth outcomes are not collected (U)



Population intervention effect as total effect

As the exposure is not randomized, we choose to decompose a total effect known as the **Population Intervention Effect** (PIE) into direct and indirect components (Hubbard and Van Der Laan, 2008).

Population intervention effect as total effect

As the exposure is not randomized, we choose to decompose a total effect known as the **Population Intervention Effect** (PIE) into direct and indirect components (Hubbard and Van Der Laan, 2008).

The PIE measures the expected change of an outcome from its observed value in a population to that in the same observed population had contrary to fact no one been exposed

Notation and definitions

Define the following counterfactual variables,

- $M(0)$ mediator in the absence of exposure

Notation and definitions

Define the following counterfactual variables,

- $M(0)$ mediator in the absence of exposure
- $Y(0, M(0))$ outcome in the absence of exposure

Notation and definitions

Define the following counterfactual variables,

- $M(0)$ mediator in the absence of exposure
- $Y(0, M(0))$ outcome in the absence of exposure
- $Y(A, M(0))$ outcome if the exposure had taken its natural level and mediator value in the absence of exposure

Notation and definitions

Define the following counterfactual variables,

- $M(0)$ mediator in the absence of exposure
- $Y(0, M(0))$ outcome in the absence of exposure
- $Y(A, M(0))$ outcome if the exposure had taken its natural level and mediator value in the absence of exposure

Decomposition of the PIE interest:

$$
\begin{aligned}
PIE &= E[Y] - E[Y(0, M(0))] \\
&= \underbrace{E[Y] - E[Y(A, M(0))]}_{\text{Population Intervention Indirect Effect}} + \underbrace{E[Y(A, M(0))] - E[Y(0, M(0))]}_{\text{Population Intervention Direct Effect}}
\end{aligned}
$$

Notation and definitions

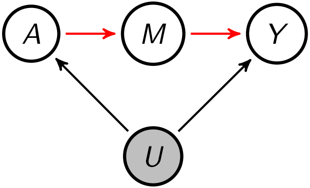
Define the following counterfactual variables,

- $M(0)$ mediator in the absence of exposure
- $Y(0, M(0))$ outcome in the absence of exposure
- $Y(A, M(0))$ outcome if the exposure had taken its natural level and mediator value in the absence of exposure

Decomposition of the PIE interest:

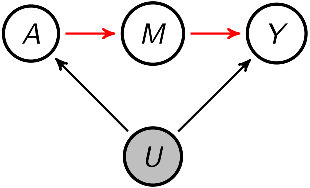
$$\begin{aligned} PIE &= E[Y] - E[Y(0, M(0))] \\ &= \underbrace{E[Y] - \mathbf{E[Y(A,M(0))]}}_{\text{Population Intervention Indirect Effect}} + \underbrace{\mathbf{E[Y(A,M(0))]} - E[Y(0, M(0))]}_{\text{Population Intervention Direct Effect}} \end{aligned}$$

Connection to Pearl's front-door criterion



1. A1. $M(0) \perp A \mid C = c \quad \forall\ c$
2. A2. $Y(a, m) \perp M \mid A = a, C = c \quad \forall\ m, a, c$
3. F1. $Y(a, m) = Y(0, m) = Y(m) \quad \forall\ a, m$

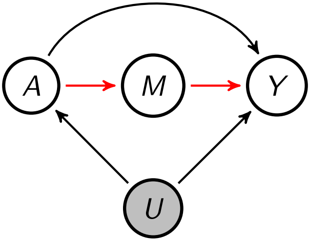
Connection to Pearl's front-door criterion



1. A1. $M(0) \perp A \mid C = c \quad \forall\ c$
2. A2. $Y(a, m) \perp M \mid A = a, C = c \quad \forall\ m, a, c$
3. F1. $Y(a, m) = Y(0, m) = Y(m) \quad \forall\ a, m$

$$PIIE = E(Y) - E(Y(0, M(0))) = PIE$$

Connection to Pearl's front-door criterion



1. A1. $M(0) \perp A \mid C = c \quad \forall\ c$
2. A2. $Y(a, m) \perp M \mid A = a, C = c \quad \forall\ m, a, c$
3. A3. $Y(a, m) \perp M(0) \mid A = a, C = c \quad \forall\ m, a, c$

$$PIIE = E(Y) - E(Y(A, M(0)))$$

Nonparametric identification of the PIIE

Theorem (1)

If assumptions A1-A3 are satisfied, the population intervention indirect effect (PIIE) is nonparametrically identified. That is,

$$PIIE = E[Y] - E[Y(A, M(0))] = E[Y] - \Psi$$

$$\begin{aligned}\Psi = \sum_{m,c} & Pr(M = m|A = 0, C = c) \\ & \times \sum_{a} E(Y|A = a, M = m, C = c) \\ & \times Pr(A = a|C = c)Pr(C = c)\end{aligned}$$

Estimation of the PIIE

Type	Model	Estimator
Parametric (*mle*)	$\mathcal{M}_{y,m,a}$	$\hat{\Psi}_{mle} = \frac{1}{n}\sum_{i=1}^{n}\left\{\sum_{m}\hat{P}r(M_i = m \mid A_i = a^*, C_i = c) \times \sum_{a}\hat{E}(Y \mid A_i = a, M_i = m, C_i = c)\hat{P}r(A_i = a \mid C_i = c)\right\}$
Semiparametric (*sp1*)	$\mathcal{M}_{m}$	$\hat{\Psi}_{sp1} = \frac{1}{n}\sum_{i=1}^{n} Y_i \frac{\hat{f}(M_i \mid 0, C_i)}{\hat{f}(M_i \mid A_i, C_i)}$
Semiparametric (*sp1*)	$\mathcal{M}_{y,a}$	$\hat{\Psi}_{sp2} = \frac{1}{n}\sum_{i=1}^{n}\frac{I(A_i = 0)}{\hat{P}r(A_i = 0 \mid C_i)} E\big(E\{Y \mid A_i, M_i, C_i\} \mid C_i\big)$
Doubly-robust semiparametric (*dr*)	$\mathcal{M}_{m \bigcup y,a}$	$\hat{\Psi}_{dr} = \frac{1}{n}\sum_{i=1}^{n}[Y - \hat{E}(Y \mid A_i, M_i, C_i)]\frac{\hat{f}(M_i \mid 0, C_i)}{\hat{f}(M_i \mid A_i, C_i)} + \frac{I(A_i = 0)}{\hat{P}r(A_i = 0 \mid C_i)}\Big(\sum_{a}\hat{E}(Y \mid a, M_i, C_i]\hat{f}(a \mid C_i) - \sum_{a,\bar{m}}\hat{E}(Y \mid a, \bar{m}, C_i)\hat{f}(\bar{m} \mid A_i, C_i)\hat{f}(a \mid C_i)\Big) + \sum_{m}\hat{E}(Y \mid A_i, m, C_i]\hat{f}(m \mid 0, C_i)$

Simulation study setup

Goal of simulation study is two-fold:

1. Verify robustness to exposure-outcome confounding for all confounders
2. Assess performance under various forms of model misspecification

Simulation study setup

- Set of binary confounders C

$$C_1 \sim Ber(.6), \quad C_2 \mid C_1 \sim Ber(\text{expit}(1 + .5c_1)), C_3 \sim Ber(.3)$$

- Binary exposure A

$$A \mid C_1, C_2 \sim Ber(\text{expit}(.5 + .2c_1 + .4c_2 + .5c_1c_2 + .2c_3))$$

- Continuous mediator M

$$M \mid A, C_1, C_2 \sim N(1 + a - 2c_1 + 2c_2 + 8c_1c_2,\ 4)$$

- Continuous outcome Y

$$Y \mid A, M, C_1, C_2 \sim N(1 + 2a + 2m - 8am + 3c_1 + c_2 + c_1c_2 + c_3,\ 1)$$

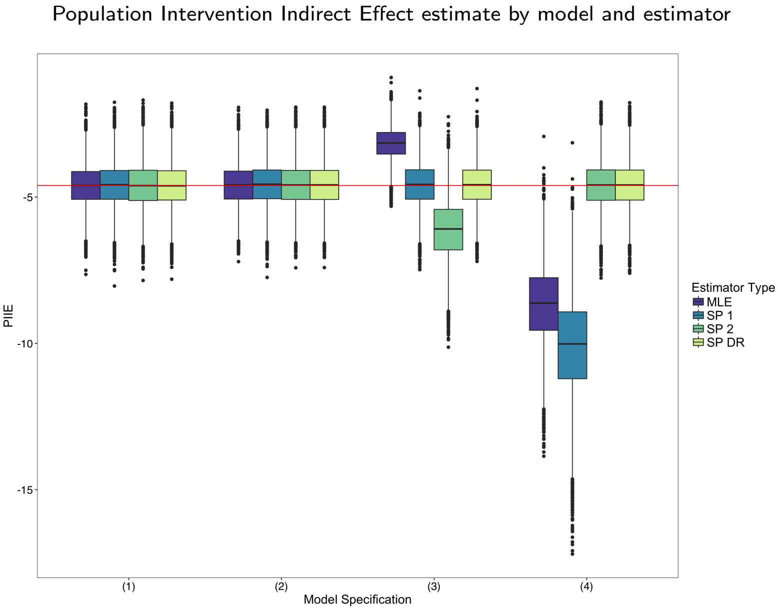
Simulation study setup

Simulations were run 10,000 times on a sample size of 1,000 under the following scenarios:

- (1) $\mathcal{M}_{y,m,a}$: models for A, M, Y correct
- (2) $\mathcal{M}'_{y,m,a}$: same as above with confounder of $A - Y$ relationship excluded from the models
- (3) $\mathcal{M}_m$: models for M correct[a]
- (4) $\mathcal{M}_{y,a}$: models for A, Y correct[b]

[a] Y misspecified by leaving out AM interaction
[b] M misspecified by leaving out interaction term between confounders

Simulation study results



http://bit.ly/enar-frontdoor

Key references

[1] Hubbard, A. E., & Van Der Laan, M. J. (2008). Population intervention models in causal inference. Biometrika, 95(1), 35-47.

[2] Pearl, J. (2009). Causality. Cambridge University Press.

[3] VanderWeele, T. (2015). Explanation in causal inference: methods for mediation and interaction. Oxford University Press.