Harvard University
Harvard University Biostatistics Working Paper Series

Year 2014 *Paper* 171

Bounds to Evaluate the Pure/natural Direct Effect without Cross-world Counterfactual Independence

Eric Tchetgen Tchetgen* Kelesitse Phiri†

*Harvard University, etchetge@hsph.harvard.edu
†Harvard University, kphiri@hsph.harvard.edu

Commentary

Bounds to evaluate the pure/natural direct effect without cross-world counterfactual independence

Eric J. Tchetgen Tchetgen[1,2] and Kelesitse Phiri[2]

Departments of Biostatistics[1] and Epidemiology[2], Harvard University

Corresponding author: Eric J. Tchetgen Tchetgen, Department of Epidemiology, Harvard School of Public Health 677 Huntington Avenue, Boston, MA 02115.

Abstract

Identification of the pure/natural direct effect relies on a cross-world counterfactual independence assumption which cannot be enforced experimentally and therefore is not strictly subject to scientific scrutiny. Here we extend the Robins-Richardson nonparametric bounds for binary exposure, mediator and outcome variables, to handle the situation of exposure-induced confounding. The new bounds are a useful alternative to the strong assumptions made by Tchetgen Tchetgen and VanderWeele (Epidemiology, 2014,Mar;25(2):282-91) who point identify the pure direct effect under a nonparametric structural equation model with independent errors which implies cross-world counterfactual independencies, and make an additional assumption about the exposure-dependent confounder.

The pure or natural direct effect quantifies the effect of an exposure on an outcome that is not mediated by an intermediate variable.[1-5] For exposure A, mediator M and outcome Y, let $M(a)$ and $Y(a) = Y(a, M(a))$ respectively define the counterfactual mediator and outcome had exposure taken value a. Likewise, let $Y(a, m)$ define the counterfactual outcome had exposure and mediator taken the value a and m, respectively. Finally let $Y(a, M(a^*))$ denote the counterfactual outcome had exposure taken value a and the mediator taken the value it would have under treatment a^*. The average pure direct effect on the additive scale is then defined for $a \neq a^*$:

$$PDE(a, a^*) = E\{Y(a, M(a^*)) - Y(a^*)\}$$

Throughout assume that as encoded in the diagram given in Figure 1.a, treatment is randomized,

$$A \perp\!\!\!\perp \{Y(a, m), M(a)\} \tag{1}$$

and also, M is randomized within levels of A,

$$Y(a, m) \perp\!\!\!\perp M(a)|A = a \tag{2}$$

Assumptions (1) and (2) are sufficient under the 3-node diagram to identify the total average causal effect $E\{Y(a) - Y(a^*)\}$ and the controlled direct effect $E\{Y(a, m) - Y(a^*, m)\}$. However, assumptions (1) and (2) taken to be the only counterfactual independencies for each value of a encoded in the 3-node causal diagram of Figure 1.a, do not suffice to identify counterfactual averages of the form $E\{Y(a, M(a^*))\}$, and therefore these assumptions alone cannot identify $PDE(a, a^*)$.

Identification of $PDE(a, a^*)$ has been controversial even in this simple setting, as it requires interpreting the diagram of Figure 1.a to encode additional counterfactual independencies, such that in addition to (1) and (2) , we also have for all a, a^* :[2]

$$Y(a, m) \perp\!\!\!\perp M(a^*)|A = a \tag{3}$$

This latter assumption is sometimes described as simply requiring that conditional on exposure, there is no unobserved confounding between the mediator and the outcome, as depicted in Figure 1.a.[2-4] However, the assumption is much stronger than suggested in Figure 1.a, since it entails assuming independencies about counterfactuals indexed by distinct treatment interventions, i.e. cross-world counterfactual independencies. Whereas, the traditional statement of no unobserved confounding (1) and (2) ensures that all counterfactuals involved in the independence statements are defined for a single treatment value a.[6,7] In principle, assumptions (1) and (2) can be made to hold experimentally, say by randomizing A and subsequently randomizing M conditional on A, while assumption (3) is more audacious and should be made with caution in practice, as it cannot be experimentally enforced and therefore is not strictly subject to scientific scrutiny.[6] The ambiguity as to which counterfactual independencies are naturally being assumed in Figure 1.a stems from the fact that the counterfactual variables in question (for the mediator and outcome variables) are themselves not directly represented on the causal diagram. As a remedy to this limiting feature of the traditional causal diagram, Richardson and Robins[6] have recently developed the single-world intervention graph (SWIG), a new causal diagram which allows one to represent counterfactuals indexed by a single world intervention directly on the graph. Figure 1.b depicts the SWIG corresponding to the causal diagram given in Figure 1.a., for an intervention which sets the treatment to a fixed value $\widetilde{a}$ for all persons in the population. One can then verify using d-separation, that unlike in Figure 1.a, in the SWIG of Figure 1.b., it is guaranteed that only assumptions (1) and

(2) (for intervention value $\widetilde{a} = a$), and no cross-world counterfactual independence is encoded in the graph. Note that to be correctly interpreted, d-separation statements for the SWIG of figure 1.b implicitly condition on variables for all split nodes to equal their value under the intervention, i.e. all independence statements are conditional on $A = a$ in Figure 1.b.

Upon ruling out cross-world independence assumptions, Robins and Richardson fail to point identify $PDE\,(a, a^*)$ and have proposed instead reporting nonparametric bounds for $PDE(a, a^*)$.[7] For binary Y, A and M, the Robins-Richardson bounds under assumptions (1) and (2) are given for $PDE\,(1, 0)$:

$$\begin{aligned}
&L \leq PDE\,(1,0) \leq U \\
&L = \max\{0, \Pr(M=0|A=0) + E[Y|A=1, M=0] - 1\} \\
&\quad + \max\{0, \Pr(M=1|A=0) + E[Y|A=1, M=1] - 1\} - E[Y|A=0] \\
&U = \min\{\Pr(M=0|A=0), E(Y|A=1, M=0)\} \\
&\quad + \min\{\Pr(M=1|A=0), E[Y|A=1, M=1] - E[Y|A=0]\}
\end{aligned} \tag{4}$$

Tchetgen Tchetgen and VanderWeele consider identification of $PDE(a, a^*)$ under the causal diagram depicted in Figure 2.a.[8] The mediator in this diagram is subject to exposure-induced confounding which by the recanting witness criterion, renders $PDE(a, a^*)$ non-identified from the observed data, even if instead of (3), one were to assume the nonparametric structural equations model with independent errors (NPSEM-IE) representation of the diagram in figure 2.a.[8] Under the NPSEM-IE, we have that conditioning on A and R recovers cross-world counterfactual independencies for the mediator and outcome,[7,8] i.e. for all a and a^*

$$Y(a, m) \perp\!\!\!\perp M(a^*)|A = a, R(a) = r \tag{5}$$

Tchetgen Tchetgen and VanderWeele[8] establish that $PDE(a, a^*)$ becomes identified in the NPSEM-IE for Figure 2.a, provided an additional assumption also holds, either

(i) the treatment and confounder R are binary and the effect of treatment on R is monotone at the individual level, i.e. $R(a^*) \leq R(a)$ for $a^* < a$, or,

(ii) there is no average additive interaction between R and M in their joint effects on Y, i.e.

$$E\,(Y|a, m, r) - E\,(Y|a, m^*, r) - E\,(Y|a, m^*, r) + E\,(Y|a, m^*, r^*) = 0$$

Alternative identifying assumptions under the NPSEM-IE were also considered by Robins and Richardson.[7] Note that Tchetgen Tchetgen and VanderWeele[8] thus continue to make a cross-world counterfactual independence assumption now given by (5). We extend the Robins-Richardson bounds for $PDE(a, a^*)$ so that they may be used in the presence of exposure-induced confounding as depicted in Figure 2.a, upon interpreting the causal diagram strictly as encoding assumptions (1) and

$$Y(a, m) \perp\!\!\!\perp M(a)|A = a, R = r \tag{6}$$

The SWIG corresponding to this causal diagram is given in Figure 2.b for the intervention $\widetilde{a}$. In the Appendix we establish that for binary A and M, and $a = 1$, $a^* = 0$, under the SWIG of Figure 2.b,

$$
\begin{aligned}
&L^* \leq PDE\,(1,0) \leq U^* \qquad (7)\\
&L^* = \max\{0, \Pr(M(0)=0) + E[Y(1,0)] - 1\}\\
&\quad + \max\{0, \Pr(M(0)=1) + E[Y(1,1)] - 1\} - E[Y(a=0)]\\
&U^* = \min\{\Pr(M(0)=0), E[Y(1,0)]\}\\
&\quad + \min\{\Pr(M(0)=1), E[Y(1,1)] - E[Y(a=0)]\}
\end{aligned}
$$

Therefore the bounds L^* and U^* are identified provided $E[Y(a,m)], E[Y(a)]$ and $E[M(a)]$ are themselves identified. Under the usual no unobserved confounding assumptions (1) and (6), i.e. under the SWIG in figure 2.b for $\widetilde{a} = a$, we indeed have

$$
\begin{aligned}
E[Y(a,m)] &= \sum_r E[Y|a,m,r]\, f(r|a),\\
E[Y(a)] &= E[Y|a],\\
E[M(a)] &= E[M|a].
\end{aligned}
$$

Appendix

SWIG bounds under (1) and (6): The derivation of L^* and U^* is analogous to that of L and U given in Robins and Richardson,[6] with a slight generalization to allow for the recanting witness R. Note that

$$\begin{aligned} E\{Y(1,M(0))\} &= E\{Y(1,1)|M(0)=1\}\Pr\{M(0)=1\} \\ &+ E\{Y(1,0)|M(0)=0\}\Pr\{M(0)=0\}, \end{aligned}$$

and

$$\begin{aligned} E\{Y(1,1)\} &= E\{Y(1,1)|M(0)=1\}\Pr\{M(0)=1\} \\ &+ E\{Y(1,1)|M(0)=0\}\Pr\{M(0)=0\}, \end{aligned}$$

which implies

$$\begin{aligned} \max\{0,(E\{Y(1,1)\}-\Pr\{M(0)=0\})/\Pr\{M(0)=1\}\} &\leq E\{Y(1,1)|M(0)=1\} \\ \min\{1,E\{Y(1,1)\}/\Pr\{M(0)=1\}\} &\geq E\{Y(1,1)|M(0)=1\} \end{aligned}$$

likewise

$$\begin{aligned} E\{Y(1,0)\} &= E\{Y(1,0)|M(0)=0\}\Pr\{M(0)=0\} \\ &+ E\{Y(1,0)|M(0)=1\}\Pr\{M(0)=1\} \end{aligned}$$

which implies

$$\begin{aligned} \max\{0,(E\{Y(1,0)\}-\Pr\{M(0)=1\})/\Pr\{M(0)=0\}\} &\leq E\{Y(1,0)|M(0)=0\} \\ \min\{1,E\{Y(1,0)\}/\Pr\{M(0)=0\}\} &\geq E\{Y(1,1)|M(0)=0\} \end{aligned}$$

It follows that

$$\begin{aligned} \max\{0,(E\{Y(1,1)\}-\Pr\{M(0)=0\})\}+\max\{0,(E\{Y(1,0)\}-\Pr\{M(0)=1\})\} &\leq E\{Y(1,M(0))\} \\ \min\{\Pr(M(0)=1),E\{Y(1,1)\}\}+\min\{\Pr(M(0)=0),E\{Y(1,0)\}\} &\geq E\{Y(1,M(0))\} \end{aligned}$$

proving the result.

Acknowledgments: The authors thank James Robins and Thomas Richardson for insightful comments and for providing the SWIG figures included in the paper.

References

[1] Robins, J. M. and S. Greenland (1992). Identifiability and exchangeability for direct and indirect effects. Epidemiology 3, 143–155.

[2] Pearl, J. (2001). Direct and indirect effects. In Proceedings of the 17th Annual Conference on Uncertainty in Artificial Intelligence (UAI-01), San Francisco, CA, pp. 411–42. Morgan Kaufmann.

[3] VanderWeele, T.J. and Vansteelandt, S. (2010). Odds ratios for mediation analysis for a dichotomous outcome - with discussion. American Journal of Epidemiology, 172, 1339-1348.

[4] Imai, K., Keele, L., and Yamamoto, T. (2010). Identification, inference and sensitivity analysis for causal mediation effects. Statistical Science 25, 51–71.

[5] Tchetgen Tchetgen EJ and Shpitser I (2012) Semiparametric Theory for Causal Mediation Analysis: efficiency bounds, multiple robustness, and sensitivity analysis).The Annals of Statistics. 40,3: 1816-1845.

[6] Richardson TS, and Robins JM (2013). Single World Intervention Graphs (SWIGs): A unification of the counterfactual and graphical approaches to causality. Technical Report 128, Center for Statistics and the Social Sciences, University of Washington.

[7] Robins, JM, and Richardson TS (2010). Alternative graphical causal models and the identification of direct effects. Causality and psychopathology: Finding the determinants of disorders and their cures (2010): 103-158.

[8] Tchetgen Tchetgen EJ and VanderWeele T. On identification of natural direct effects when a confounder of the mediator is directly affected by exposure. (2013) Epidemiology. Mar;25(2):282-91.

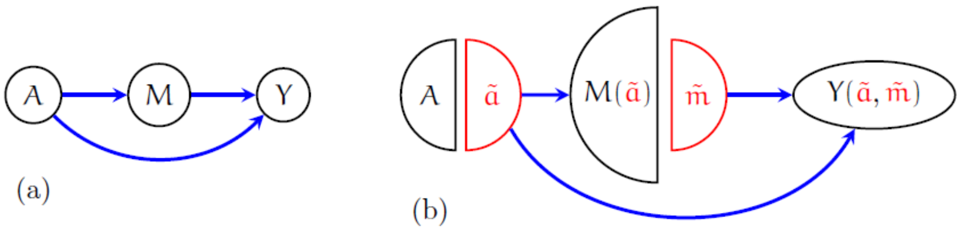

Figure 1: (a) Causal diagram $\mathcal{G}$ depicting the 3-node mediation graph with exposure A, mediator M, outcome Y in the absence of unmeasured confounding; (b) the corresponding SWIG $\mathcal{G}(\tilde{a}, \tilde{m})$ obtained from $\mathcal{G}$.

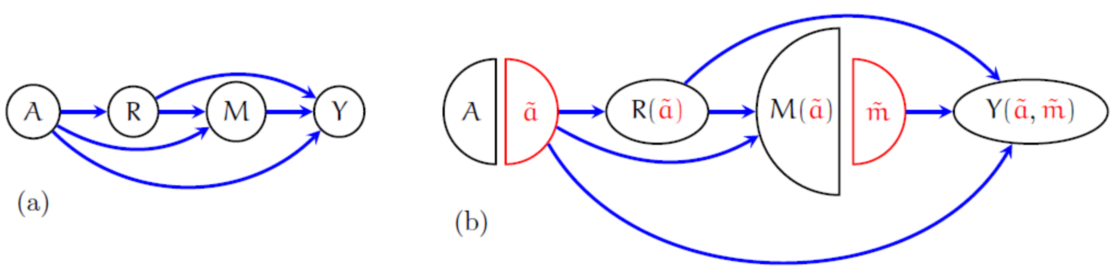

Figure 2: (a) Causal diagram $\mathcal{G}$ depicting the 4-node mediation graph with exposure A, mediator M, outcome Y and exposure-induced confounder R again unmeasured confounding is assumed absent; (b) the corresponding SWIG $\mathcal{G}(\tilde{a}, \tilde{m})$.