Harvard University
Harvard University Biostatistics Working Paper Series

Year 2012 *Paper* 148

On Identification of Natural Direct Effects when a Confounder of the Mediator is Directly Affected by Exposure

Eric J. Tchetgen Tchetgen* Tyler J. VanderWeele†

*Harvard University, etchetge@hsph.harvard.edu

†Harvard University, tvanderw@hsph.harvard.edu

Original Article

On identification of natural direct effects when a confounder of the mediator is directly affected by exposure

Eric J. Tchetgen Tchetgen and Tyler J. VanderWeele

Departments of Epidemiology and Biostatistics,

Harvard University

Abstract

Natural direct and indirect effects formalize traditional notions of mediation analysis into a rigorous causal framework and have recently received considerable attention in epidemiology and in the social sciences. Sufficient conditions for identification of natural direct effects were formulated by Judea Pearl under a nonparametric structural equations model, which assumes certain independencies between potential outcomes. A common situation in epidemiology is that a confounder of the mediator is affected by the exposure, in which case, natural direct effects fail to be nonparametrically identified without additional assumptions, even under Pearl's nonparametric structural equations model. In this paper, the authors show that when a single binary confounder of the mediator is affected by the exposure; the natural direct effect is nonparametrically identified under a monotonicity assumption about the effect of the exposure on the confounder. A similar result is shown to hold for a vector of binary confounders of the mediator under a certain independence assumption about the confounders. Finally, the authors show that natural direct effects are more generally identified if there is no-additive mean interaction between the mediator and confounders of the mediator affected by exposure. When correct, this latter assumption is particularly appealing because it does not require monotonicity of effects of the exposure, additionally,

it places no restriction on the nature of the confounders of the mediator which can be continuous or polytomous.

KEY WORDS: Natural direct effect,potential outcomes, identification.

There has recently developed a literature in causal inference concerned with the definition, identification and estimation of direct and indirect effects in fully non-parametric models[1−14] primarily based on ideas developed by Robins and Greenland[1], and Pearl.[2] This recent literature uses the language of potential outcomes to give a non-parametric definition of effects involved in mediation analysis known as controlled direct effects, natural direct and indirect effects, and path-specific effects. These effects, despite being defined in a fully non-parametric way, can nevertheless be sometimes identified and estimated from observational data.[2]

The current paper concerns natural direct effects, also known as pure direct effects, which capture the effects of an exposure when one intervenes to set a mediator to the (random) level it would have had in the absence of exposure.[1,2] Such effects generally differ from controlled direct effects which refer to exposure effects that arise upon intervening to set the mediator to a fixed level that may differ from its actual observed value.[1,2,4] Natural direct and indirect effects combine to produce an exposure total effect, and as Pearl previously noted, are more useful than controlled direct effects for understanding the underlying mechanism by which the exposure operates.[2]

Sufficient conditions for nonparametric identification of natural direct and indirect effects were given by Pearl[2], under a nonparametric structural equations model (NPSEM), which assumes certain independencies between potential outcomes. A common situation in epidemiology is that a confounder of the mediator is affected by the exposure, in which case, Avin et al[4] establish that natural direct effects fail to be nonparametrically identified without additional assumptions, even under Pearl's nonparametric structural equations model. In this paper, the authors show that when

a single binary confounder of the mediator is affected by the exposure, natural direct effects are nonparametrically identified under a monotonicity assumption about the effect of the exposure on the confounder. A similar result is shown to hold for a vector of binary confounders of the mediator under a certain independence assumption of the confounders. Finally, the authors show that natural direct effects are more generally identified if there is no-additive interaction between the mediator and the confounders of the mediator affected by the exposure in the outcome regression. When correct, this latter assumption is appealing because it does not require monotonicity of effects of exposure, additionally, it places no restriction on the nature of the confounders which can be continuous or polytomous, however, the approach is no longer nonparametric.

Notation and definitions

We introduce the notation and definitions we will be using throughout. Let E denote the exposure or treatment received by an individual, let Y denote a post-treatment outcome, and let M denote the value of a post-treatment intermediate variable that may serve as a mediator for the treatment-outcome relationship. Let C denote the value of a set of pre-exposure confounding variables of the effects of E and M. Throughout, we will assume independent and identically distributed sampling of C, E, M and Y. If there is no confounder of the mediator effect on the outcome that is itself affected by the exposure then the relationships between these variable may be depicted in the causal diagram in Figure 1.

Insert Figure 1

We now consider counterfactuals or potential outcomes, under possible interventions on the variables.[15,16] Let $Y(e)$ denote a subject's outcome if treatment E were set, possibly contrary to

fact, to e. In the context of mediation there will also be potential outcomes for the intermediate variable. Let $M(e)$ denote a subject's counterfactual value of the intermediate M if treatment E were set to the value e. Finally, let $Y(e,m)$ denote a subject's counterfactual value for Y if E were set to e and M were set to m. Similar definitions hold for $Y(e,m,c)$ and $M(e,c)$.

Nonparametric structural equations models and natural direct effects

The exposition is framed around a nonparametric structural equation theory of causal inference, described by Judea Pearl.[17] Structural equations provide a nonparametric algebraic interpretation of the diagram of Figure 1 corresponding to four functions, one for each variable on the causal graph:

$$C = g_C\left(\varepsilon_C\right) \tag{1}$$

$$E = g_E\left(C, \varepsilon_E\right) \tag{2}$$

$$M = g_M(C, E, \varepsilon_M) \tag{3}$$

$$Y = g_Y(C, E, M, \varepsilon_Y) \tag{4}$$

Each of the nonparametric functions $\{g_C, g_E, g_M, g_Y\}$ represents a causal mechanism that determines the value of the left-hand-side variable, known as the output, from variables on the right, known as the inputs. The errors $(\varepsilon_C, \varepsilon_E, \varepsilon_M, \varepsilon_Y)$ stand for all factors not included on the graph that could possibly affect their corresponding outputs when all other inputs are held constant. To be consistent with the causal graph presented in Figure 1, we require that these errors be mutually independent, but we allow their distribution to remain arbitrary. If they were not independent we would include an additional unmeasured variable U on the diagram with arrows into the relevant

variables to induce independence. Lack of a causal effect of a given variable on an output is encoded by an absence of the variable from the right-hand side. For example, consider a modification of Figure 1 obtained upon deleting the arrow $E \rightarrow Y$ indicating the absence of a direct effect of E on Y. This no direct effect is encoded by replacing equation (4) with $Y = g_Y(C, M, \varepsilon_Y)$ in the NPSEM. The absence of E from the arguments of g_Y encodes the assumption that variations in E will leave Y unchanged, as long as variables C, M and ε_Y remain constant, which is also consistent with the assumption that there are no unmeasured common causes of Y and E.

As stated by Pearl[17], the invariance of structural equations permits their use as a basis for modeling causal effects and potential outcomes. In fact, to emulate the intervention in which one sets $\{E = e\}$ for all individuals simply amounts to replacing the equation for E with $E = e$, producing the following set of modified equations:

$$\begin{aligned} C &= g_C\left(\varepsilon_C\right) \\ E &= e \\ M\left(e\right) &= g_M(C, e, \varepsilon_M) \\ Y\left(e\right) &= g_Y(C, e, M\left(e\right), \varepsilon_Y) \end{aligned}$$

with $\{M\left(e\right), Y\left(e\right) = Y\left(e, M\left(e\right)\right)\}$ denoting the potential outcomes had the exposure been set to e.

Under the above NPSEM, the independence of errors $\varepsilon_M \perp\!\!\!\perp \varepsilon_Y$ implies independence of potential outcomes for different exposure values:

$$Y\left(e, m, c\right) \perp\!\!\!\perp M\left(e^*, c\right) \tag{5}$$

where $M(e^*, c) = g_M(c, e^*, \varepsilon_M)$ and $Y(e, m, c) = g_Y(c, e, m, \varepsilon_Y)$ are obtained upon intervening on (E, C) and (E, M, C) respectively, and e, e^* take values in $\{0, 1\}$.

Robins and Greenland[1] and Pearl[2] considered the following decomposition of individual total effect of exposure:

$$\begin{aligned} Y(e) - Y(e^*) &= Y(e, M(e)) - Y(e^*, M(e^*)) \\ &= \underbrace{Y(e, M(e^*)) - Y(e^*, M(e^*))}_{\text{Natural direct effect}} + \underbrace{Y(e, M(e)) - Y(e, M(e^*))}_{\text{Natural indirect effect}} \end{aligned}$$

where e^* indicates a reference or baseline value of E; for instance it is common to chose $e^* = 0$ for binary E, and e represents an active value of treatment. The first contrast on the right hand side of the second line displayed above defines individual natural direct effect of treatment E on outcome Y. The potential outcome $Y(e^*, M(e^*))$ captures the behavior of Y under the baseline treatment value, while $Y(e, M(e^*))$ describes the behavior of Y under the active treatment value, in a hypothetical situation where the mediator behaves as if treatment were set to baseline. The second contrast on the right hand side of the expression in the display above corresponds to the natural indirect effect of treatment E on outcome Y. The potential outcome $Y(e, M(e))$ describes the behavior of Y under the active treatment value, while the second "subtracts off" the behavior of Y under the active treatment value, in a hypothetical situation where the mediator behaves as if treatment were set to baseline. In graphical terms, the individual natural indirect effect quantifies for the individual, the effect of E on Y along the indirect causal pathway $E \rightarrow M \rightarrow Y$, but not along the direct arrow from E to Y. Because potential outcomes under conflicting exposure status are never jointly observed, individual causal effects are generally not identified. However, one can hope that under certain assumptions, population average causal effects would become identified. It is well known that the average total effect of E on Y is identified given data on (C, E, Y) in the

causal diagram of Figure 1, and is given by the g-formula of Robins[18] :

$$TE\left(e,e^{*}\right)\equiv\mathbb{E}\left\{Y\left(e\right)-Y\left(e^{*}\right)\right\}=\sum_{c}\left[\mathbb{E}\left\{Y|e,c\right\}-\mathbb{E}\left\{Y|e^{*},c\right\}\right]\Pr(C=c) \tag{6}$$

where E stands for expectation. Pearl[2] proved that under the NPSEM for the causal graph of Figure 1, the average natural direct effect is identified by

$$\begin{aligned}NDE\left(e,e^{*}\right)&\equiv\mathbb{E}\left\{Y\left(e,M\left(e^{*}\right)\right)-Y\left(e^{*},M\left(e^{*}\right)\right)\right\}\\&=\sum_{c}\left[\mathbb{E}\left\{Y|e,m,c\right\}-\mathbb{E}\left\{Y|e^{*},m,c\right\}\right]\Pr(M=m|E=e^{*},C=c)\Pr(C=c)\end{aligned}$$

Therefore the average natural indirect effect is obtained under the NPSEM by $NIE\left(e,e^{*}\right)=TE\left(e,e^{*}\right)-NDE\left(e,e^{*}\right)$. A variety of statistical methods for estimating $NDE\left(e,e^{*}\right)$ and $NIE\left(e,e^{*}\right)$ have been proposed in recent literature.[6,9–14] Tchetgen Tchetgen and Shpitser compares several of these methods and develop a semiparametric approach with attractive robustness and efficiency properties.[14]

Next, we consider a common setting in epidemiology displayed in Figure 2. In this causal diagram, a common cause N of the mediator and outcome is directly affected by exposure. In other words, N is simultaneously a confounder of the effects of the mediator M on Y, and on the causal pathway from exposure to outcome.

Insert Figure 2.

Similar to figure 1, we suppose the NPSEM for this causal diagram is given by:

$$C = g_C(\varepsilon_C) \tag{7}$$

$$E = g_E(C, \varepsilon_E) \tag{8}$$

$$N = g_N(C, E, \varepsilon_N) \tag{9}$$

$$M = g_M(C, E, N, \varepsilon_M) \tag{10}$$

$$Y = g_Y(C, E, M, N, \varepsilon_Y) \tag{11}$$

where as before $\{g_C, g_E, g_N, g_M, g_Y\}$ are nonparametric functions, and the errors $(\varepsilon_C, \varepsilon_E, \varepsilon_N, \varepsilon_M, \varepsilon_Y)$ are mutually independent. The total effect of E on Y remains identified by equation (6) using data (C, E, Y) only, so that the presence of N brings no new difficulty. The situation is quite different if identification of natural direct effect $NDE(e, e^*) = \mathbb{E}\{Y(e, M(e^*)) - Y(e^*, M(e^*))\}$ is in view. In particular, according to a result by Avin et al, a causal effect along a specific path is not identifiable in a fully observable NPSEM if and only if there exists a so called "recanting witness," namely a random variable that mediates the causal pathway of interest from E to Y, while at the same time mediating another causal pathway from E to Y which is not of interest. Note that the direct effects of E on outcome Y in Figure 2 consists of the two pathways $E \to Y$ and $E \to N \to Y$. But the variable N also mediates the indirect effect $E \to N \to M \to Y$ which is not of interest when direct effects are in view, and therefore N is a recanting witness for the direct effect path $E \to N \to Y$. This in turn implies that $NDE(e, e^*)$ is not nonparametrically identified without an additional assumption under NPSEM (7) – (11) for the causal diagram in Figure 2. To understand why identification fails, it is useful to consider the following expression for the average of the potential outcome $\mathbb{E}\{Y(e, M(e^*))\}$, under the NPSEM given in Robins and

Richardson[19]:

$$\sum_{m,n,n',c} \mathbb{E}\left(Y|e,m,n,c\right) \Pr(M=m|\ e^*,\ n^{'},c) \Pr\left(N\left(e\right)=n, N(e^*)=n'|c\right) \Pr(C=c) \tag{12}$$

therefore, although $\mathbb{E}\left(Y|e,m,n,c\right), \Pr(M=m|\ e^*,\ n^{'},c)$ and $\Pr(C=c)$ are identified from data (Y,E,M,N,C), identification of $NDE\left(e,e^*\right)$ fails because it requires the conditional joint density $\Pr(N\left(e\right)=n,\ N(e^*)=n'|c)$ which involves the potential outcomes of the recanting witness N for conflicting exposure values and therefore is not identified even under the NPSEM.

Robins and Richardson[19] show that equation (12) becomes identified under either of the following two assumptions:

(i) if $N(e) \perp\!\!\!\perp N(e^*)$, then $\mathbb{E}\left\{Y\left(e,M\left(e^*\right)\right)\right\}$ is identified by the following formula

$$\sum_{m,n,n',c} \mathbb{E}\left(Y|e,m,n,c\right) \Pr(M=m|\ e^*,\ n^{'},c) \Pr\left(N=n|e,c\right) \Pr\left(N=n'|e^*,c\right) \Pr(C=c)$$

(ii) or if $N(e)$ is a deterministic function of $N(e^*)$, say $N(e)=h(N(e^*))$ in which case, $\mathbb{E}\left\{Y\left(e,M\left(e^*\right)\right)\right\}$ is given by

$$\sum_{m,n,n',c} \mathbb{E}\left(Y|e,m,n,c\right) \Pr(M=m|\ e^*,\ n^{'},c) \Pr\left(N=n'|e^*,c\right) I(n=h\left(n'\right)) \Pr(C=c)$$

For scalar N taking values in a continuous state space, Robins and Richardson showed that under a condition of rank preservation, there will always exist a function h such that (ii) holds. However, rank preservation is often not biologically plausible.[18] Likewise, assumption (i) seems unrealistic for applications in the health sciences, because in such applications, it is usual that potential outcomes under various exposure values are correlated because of unknown behavioral or environmental risk factors, or unknown genetic risk factors for the outcome. In the next section,

we explore less stringent assumptions for identification of $NDE\,(e, e^*)$ under an NPSEM.

Identification with a binary recanting witness under effect monotonicity

Suppose that N is binary. Then consider the following monotonicity assumption.

$E - N$ Monotonicity Assumption: If $e^* < e$ then $N(e^*) \leq N(e)$ for all individuals.

This type of monotonicity assumption is often used in recent epidemiologic literature, particularly in the context of causal inference.[20–23] The monotonicity assumption is particularly easy to interpret for binary exposure and counfounder E and N. Then it simply states that if a person experiences the confounder when unexposed, that is $N(0) = 1$, then it must be that he or she would also experience the confounder when exposed, that is $N(1) = 1$. However, for a person without the confounder when unexposed, that is $N(0) = 0$, the potential outcome $N(1)$ can either be 0 or 1. In the appendix, we use this assumption to show the following result.

Result 1: Assuming the NPSEM $(7) - (11)$, *suppose that N is binary, and $E - N$ Monotonicity Assumption holds, then $E\{Y\,(e, M\,(e^*))\}$ is nonparametrically identified by the following formula* $z\,(e, e^*) =$

$$\sum_{m,n,n',c} \mathbb{E}\,(Y|e, m, n, c)\Pr(M = m|\ e^*,\ n^{'}, c) f\left(n, n^{'}, e, e^*, c\right)\Pr(C = c)$$

where

$$f\left(n, n^{'}, e, e^{*}, c\right) = \begin{cases} \Pr\left\{N = 1 | e^{*}, c\right\} \text{ if } n' = n = 1 \\ \Pr\left\{N = 1 | e, c\right\} - \Pr\left\{N = 1 | e^{*}, c\right\} \text{ if } n' = 0 \text{ and } n = 1 \\ 0 \text{ if } n' = 1 \text{ and } n = 0 \\ \Pr\left\{N = 0 | e, c\right\} \text{ if } n' = n = 0 \end{cases}$$

The theorem states that under the NPSEM $(7) - (11)$ for which $E - N$ monotonicity holds, the joint conditional density $\Pr\left(N\left(e\right) = n, N(e^{*}) = n' | c\right)$ of potential outcomes for conflicting values $E = e, e^{*}$ is identified by $f\left(n, n^{'}, e, e^{*}, c\right)$, and therefore, by equation (12), $\mathbb{E}\left\{Y\left(e, M\left(e^{*}\right)\right)\right\}$ is nonparametrically identified by $z\left(e, e^{*}\right)$. Note that because of monotonicity $\Pr\{N = 1 | e, c\} \geq \Pr\{N = 1 | e^{*}, c\}$ and therefore $\Pr\left(N\left(e\right) = 1, N(e^{*}) = 0 | c\right) = f\left(1, 0, e, e^{*}, c\right) = \Pr\{N = 1 | e, c\} - \Pr\{N = 1 | e^{*}, c\} \geq 0$. In addition, since $E\left\{Y\left(e^{*}, M\left(e^{*}\right)\right)\right\}$ is nonparametrically identified and equal to the second term on the right hand-side of equation (6), we conclude that $NDE\left(e, e^{*}\right)$ is nonparametrically identified. Likewise, it follows that $NIE\left(e, e^{*}\right)$ is also nonparametrically identified by $TE\left(e, e^{*}\right) - NDE\left(e, e^{*}\right)$.

Result 1 generalizes somewhat beyond the simple case of a single binary recanting witness; specifically, suppose that N consists of multiple binary variables $N = (N_1, ..., N_k)$, and suppose that the NPSEM $(7) - (11)$ holds upon replacing equation (9) with the k equations:

$$N_j = g_{N_j}\left(C, E, \varepsilon_{N_j}\right), \ j = 1, ..., k$$

such that $\left\{\varepsilon_{N_j} : j = 1, ..., k\right\}$ are mutually independent and are jointly independent of $\left\{\varepsilon_C, \varepsilon_E, \varepsilon_M, \varepsilon_Y\right\}$.

Result 2: Assuming the NPSEM $(7) - (11)$, *suppose that the* $E - N$ *Monotonicity Assumption*

holds for each of $(N_1, ..., N_k) = N$*, then* $E\{Y(e, M(e^*))\}$ *is nonparametrically identified by the following formula:*

$$\sum_{m,n,n',c} \mathbb{E}(Y|e, m, n, c) \Pr(M = m|\ e^*,\ n^{'}, c) \prod_{j=1}^{k} f_j\left(n_j, n_j^{'}, e, e^*, c\right) \Pr(C = c)$$

where

$$f_j\left(n_j, n_j^{'}, e, e^*, c\right) = \begin{cases} \Pr\{N_j = 1|e^*, c\} \text{ if } n_j' = n_j = 1 \\ \Pr\{N_j = 1|e, c\} - \Pr\{N_j = 1|e^*, c\} \text{ if } n_j' = 0 \text{ and } n_j = 1 \\ 0 \text{ if } n_j' = 1 \text{ and } n_j = 0 \\ \Pr\{N_j = 0|e, c\} \text{ if } n_j' = n_j = 0 \end{cases}$$

Crucially, we note that the above Results 1 and 2 place no restriction on the functional form of the regression function $\mathbb{E}(Y|e, m, n, c)$, and therefore interactions are easy to accommodate as well as nonlinearities corresponding to log or logit link, e.g. for a nonnegative outcome or a binary response respectively, but also nonlinearities in dose response relations between (E, M, N, C) and Y.

Additionally, the above results extend to the survival context. For instance, consider the hazard function for the survival outcome Y given (E, M, N, C) on the additive scale:

$$\lambda(y|e, m, n, c) = \lambda_0(y) + \gamma(y, e, m, n, c)$$

where $\gamma(y, e, m, n, c)$ is unrestricted except for $\gamma(y, e^*, m^*, n^*, c^*) = 0$ for all $y > 0$, and $\lambda_0(y)$ is

the baseline hazard of Y given $(E = e^*, M = m^*, N = n^*, C = c^*)$. In the appendix, we derive a general expression for the hazard function of $Y(e, M(e^*))$ evaluated at y under the monotonicity assumption.

It is also straightforward to extend the above results when conditional effects are in view, say $NDE(e, e^*, c) = E\{Y(e, M(e^*)) - Y(e^*, M(e^*))|c\}$. Similar to marginal effects the challenge in identification of such effects lies in the need to identify $E\{Y(e, M(e^*))|c\}$ which can be shown under the NPSEM to be equal to:

$$\sum_{m,n,n',c} \mathbb{E}(Y|e, m, n, c) \Pr(M = m|\ e^*,\ n', c) \Pr(N(e) = n, N(e^*) = n'|c)$$

Under monotonicity, it essentially follows from the proof of Result 1 for binary N that the above formula is identified by

$$\sum_{m,n,n',c} \mathbb{E}(Y|e, m, n, c) \Pr(M = m|\ e^*,\ n', c) f\left(n, n', e, e^*, c\right)$$

In the following section, we consider a further decomposition of $NDE(e, e^*)$ to account for the mediating role of the recanting witness N.

Decomposition of $NDE(e, e^*)$

Consider again the setting of a binary recanting witness N. Recall that $NDE(e, e^*)$ captures the effects along the following two pathways: $E \to Y$ and $E \to N \to Y$. We note that

$$\begin{aligned}\mathbb{E}\{Y(e, M(e^*))\} - \mathbb{E}\{Y(e^*, M(e^*))\} = {} & \mathbb{E}\{Y(e, M(e^*), N(e)) - Y(e, M(e^*), N(e^*))\} \\ & + \mathbb{E}\{Y(e, M(e^*), N(e^*)) - Y(e^*, M(e^*), N(e^*))\}\end{aligned}$$

$\mathbb{E}\{Y(e, M(e^*), N(e^*)) - Y(e^*, M(e^*), N(e^*))\}$ captures the pathway $E \rightarrow Y$, the portion of the direct effect not mediated by N, while $\mathbb{E}\{Y(e, M(e^*), N(e)) - Y(e, M(e^*), N(e^*))\}$ captures the pathway $E \rightarrow N \rightarrow Y$, the portion of the direct effect mediated by N. Under monotonicity of the effects of exposure on the recanting witness, we show next that $\mathbb{E}\{Y(e, M(e^*), N(e^*))\}$ is identified and therefore, both of these effects are nonparametrically identified.

Corollary 1: Assuming the NPSEM $(7)-(11)$, *suppose that N is binary, and $E-N$ Monotonicity Assumption holds, then*

$$\mathbb{E}\{Y(e, M(e^*), N(e^*))\} = \sum_{m,n,n',c} \mathbb{E}(Y|e, m, n', c) \Pr(M = m|\ e^*,\ n', c) f\left(n, n', e, e^*, c\right) \Pr(C = c)$$

with $f\left(n, n', e, e^*, c\right)$ given in Result 1.

Corollary 1 extends to the context of multivariate binary confounder N under the assumptions listed in Result 2. Details are omitted but are easily deduced from the presentation. Despite these important generalizations of Result 1, identification under monotonicity is still somewhat limited in that each N_j is restricted to be binary $j = 1, ..., k$. In epidemiologic applications, confounders are often measured as polytomous factors, or as continuous factors, in which case Results 1 and 2 are not particularly useful. In the following section, we give a simple condition that identifies $NDE(e, e^*)$ with a (possibly multivariate) recanting witness of a polytomous or continuous nature.

Identification assuming no-mediator-recanting witness interaction

To state the identification result, we require additional notation. Let (c^*, m^*, n^*) denote a reference

value of (c, m, n) and define

$$\begin{aligned}
\beta_m(e, m, c) &= \mathbb{E}(Y|e, m, n^*, c) - \mathbb{E}(Y|e, m^*, n^*, c) \\
\beta_n(e, \ n, c) &= \mathbb{E}(Y|e, m^*, n, c) - \mathbb{E}(Y|e, m^*, n^*, c) \\
\beta_{m,n}(e, m, n, c) &= \mathbb{E}(Y|e, m, n, c) - \mathbb{E}(Y|e, m^*, n, c) - \mathbb{E}(Y|e, m^*, n, c) + \mathbb{E}(Y|e, m^*, n^*, c) \\
\overline{\beta}_{e,c}(e, c) &= \mathbb{E}(Y|e, m^*, n^*, c)
\end{aligned}$$

$\beta_m(e, m, c)$ and $\beta_n(e, \ n, c)$ encode on the additive scale, the average main effects of M and N on Y within levels of E and C, when $n = n^*$ and $m = m^*$ respectively. $\beta_{m,n}(e, m, n, c)$ encodes the interaction between m and n on the additive scale, within levels of E and C, and $\beta_{e,c}(e, c)$ is the average outcome within levels of E and C. The average outcome can be decomposed on the additive scale in terms of $\beta_m, \beta_n, \beta_{m,n}, \overline{\beta}_{e,c}$:

$$\mathbb{E}(Y|e, m, n, c) = \beta_m(e, m, c) + \beta_n(e, n, c) + \beta_{m,n}(e, m, n, c) + \overline{\beta}_{e,c}(e, c)$$

Crucially, note that this decomposition is fully nonparametric. Consider the following No-$M - N$ Average Interaction Assumption on the additive scale.

No M-N average interaction assumption: The average additive interaction between M and N is zero; that is

$$\beta_{m,n}(e, m, n, c) = 0$$

Result 4: Assuming the NPSEM (7)−(11), *suppose that no* $M - N$ *average interaction assumption*

holds, then $\mathbb{E}\{Y(e, M(e^*))\}$ *is identified by the following formula*

$$\sum_{m,c} \beta_m(e, m, c) \Pr(M = m|\ e^*, c) \Pr(C = c) + \sum_{n,c} \beta_n(e,\ n, c) \Pr(N = n|e, c) \Pr(C = c)$$
$$+ \sum_{c} \overline{\beta}_{e,c}(e, c) \Pr(C = c)$$

The no- $M - N$ average interaction assumption is testable, since the assumption places a restriction on the observed data distribution. In principle, a nonparametric test of interaction could be performed to assess this restriction as long as the observed data is not too high dimensional. In practice, a simple parametric test of interaction could be used, or alternatively, a semiparametric multiply robust test of additive interaction could be used to accommodate high dimensional data also using simple parametric models, while minimizing the risk for bias due to modeling error.[24] The above result states that when the assumption of no interaction holds, $\mathbb{E}\{Y(e, M(e^*))\}$ is identified under the NPSEM and therefore $NDE(e, e^*)$ is identified by the following simple expression

$$NDE(e, e^*) = \textstyle\sum_{m,c} \{\beta_m(e, m, c) - \beta_m(e^*, m, c)\} \Pr(M = m|\ e^*, c) \Pr(C = c)$$
$$+ \textstyle\sum_{n,c} \{\beta_n(e,\ n, c) \Pr(N = n|e, c) - \beta_n(e^*,\ n, c) \Pr(N = n|e^*, c)\} \Pr(C = c)$$
$$+ \textstyle\sum_{c} \left\{\overline{\beta}_{e,c}(e, c) - \overline{\beta}_{e,c}(e^*, c)\right\} \Pr(C = c).$$

Estimation of $NDE(e, e^*)$

Inference under the $E - N$ Monotonicity Assumption is relatively straightforward using standard parametric models. To ground the discussion, consider the case of continuous Y and M, and

binary E, N. Then, consider the following regression models:

$$\mathbb{E}\left(Y|e,m,n,c\right) = \alpha_0 + \alpha_c'c + \alpha_e e + \alpha_m m + \alpha_n n + \alpha_{mn} mn \tag{13}$$

$$\mathbb{E}\left(M|e,n,c\right) = \theta_0 + \theta_c'c + \theta_e e + \theta_n n \tag{14}$$

$$\text{logit}\Pr\{N=1|e,c\} = \eta_0 + \eta_e e + \eta_c c \tag{15}$$

where for simplicity we allow for a potential interaction between M and N in the model for Y, and otherwise, all covariate effects are assumed to be linear in this model, as well as in the model for M. These assumptions could of course be relaxed to incorporate additional interaction and possible nonlinearity in continuous factors. Closed-form expressions for $NDE(e,e^*)$ and $NDE(e,e^*,c)$, with $e=1$ and $e^*=0$ under models (13)-(15) are given by:

$$NDE(1,0) = \alpha_e + \mathbb{E}\left\{\left(\alpha_{mn}\theta_0 + \alpha_{mn}\theta_c'C + \alpha_n\right)\omega\left(C\right)\right\}$$

$$NDE(1,0,c) = \alpha_e + \left(\alpha_{mn}\theta_0 + \alpha_{mn}\theta_c'c + \alpha_n\right)\omega\left(c\right)$$

where

$$\begin{aligned}\omega\left(c\right) &= \Pr\{N=1|e=1,c\} - \Pr\{N=1|e=0,c\} \\ &= \left[\left\{1+\exp\left(-\eta_0-\eta_e-\eta_c C\right)\right\}^{-1} - \left\{1+\exp\left(-\eta_0-\eta_c C\right)\right\}^{-1}\right]\end{aligned}$$

Above, α_e captures the effect of E along the direct path $E \to Y$ in Figure 2, whereas

$$\left(\alpha_{mn}\theta_0 + \alpha_{mn}\theta_c'c + \alpha_n\right)\omega\left(c\right)$$

captures the effect of E along the path $E \to N \to Y$ accounting for interaction between M and N.

Regression parameters in (13) and (14) can be estimated via ordinary least-squares using standard statistical software, and the parameters in (15) can similarly be obtained via maximum likelihood estimation of logistic regression. The estimator of $NDE(1,0)$ is then obtained upon replacing unknown parameters by their estimates. A standard application of the delta method can be used to compute standard errors, or alternatively, one could apply the nonparametric bootstrap. In the next section, we provide a simple data illustration of the methods described in this section using Proc NLMIXED in SAS which also delivers valid standard error estimates and 95% confidence intervals.

Inference about $NDE(1,0)$ and $NDE(1,0,c)$ can likewise be obtained even when the monotonicity assumption does not apply, provided that the no-interaction assumption of Result 4 holds. Specifically, suppose that N is now continuous, assume:

$$\mathbb{E}(Y|e,m,n,c) = \alpha_0 + \alpha_c'c + \alpha_e e + \alpha_m m + \alpha_n n + \alpha_{em} em + \alpha_{en} e \tag{16}$$

$$\mathbb{E}(N|e,c) = \eta_0 + \eta_e e + \eta_c c \tag{17}$$

$$\mathbb{E}(M|e,n,c) = \theta_0 + \theta_c'c + \theta_e e + \theta_n n \tag{18}$$

The regression model for Y in the above display differs from the previous model in that now it incorporates possible interactions between E and M, and E and N, and by assumption, no interaction between M and N. Additionally, suppose that M is modeled as in Equation (14),

which implies that

$$\mathbb{E}\left(M|e,c\right) = \theta_0^\dagger + \theta_c^{\dagger\prime} c + \theta_e^\dagger e$$

$$\text{where } \theta_0^\dagger = \theta_0 + \theta_n \eta_0$$

$$\theta_c^\dagger = \theta_c + \theta_n \eta_c$$

$$\theta_e^\dagger = \theta_e + \theta_n \eta_e$$

Applying Result 4, we obtain the following simple expressions:

$$NDE(1,0) = \alpha_e + \alpha_{em}\left(\theta_0^\dagger + \theta_c^{\dagger\prime}\mathbb{E}\left(C\right)\right) + \alpha_n\eta_e + \alpha_{en}\left(\eta_0 + \eta_e + \eta_c\mathbb{E}\left(C\right)\right)$$

$$NDE(1,0,c) = \alpha_e + \alpha_{em}\left(\ \theta_0^\dagger + \theta_c^{\dagger\prime} c\right) + \alpha_n\eta_e + \alpha_{en}\left(\eta_0 + \eta_e + \eta_c c\right)$$

As was the case under monotonicity, the coefficients in the above regression models for Y,M and N can be obtained using ordinary least-squares. $NDE(1,0)$ and $NDE(1,0,c)$ are estimated using the expression in the above display evaluated at the estimated parameter values. In principle, the delta method could be used to derive analytical estimates of standard errors, alternatively, the nonparametric bootstrap could also be used, and may be more convenient in practice. Next, we provide an example illustrating how these estimators and respective standard errors can be obtained using Proc NLMIXED in SAS.

A data illustration using Proc NLMIXED in SAS

We illustrate the methodology developed above in the context of simulated data. We first generate data as would be observed in a randomized study of sample size 500 where E is randomized with

probability $1/2$, N is dichotomous with event probability

$$\Pr(N = 1|E) = \{1 + \exp(-0.5 - 0.75E)\}^{-1},$$

and M and Y are continuous:

$$Y = 60 + 2E + 3M + 1.5N + MN + N(0, 1.5)$$

$$M = 30 + 3E + 4N + N(0, 1)$$

There is no pre-exposure confounder C in these simulated data. Then, under monotonicity, $NDE(1, 0) = 6.877$ in these data. We illustrate how this effect estimate can be obtained in Proc NLMIXED in SAS by providing sample code in the appendix. Using this sample code, we obtained an estimate of $NDE(1, 0)$ equal to 6.383 (95% confidence interval=$(3.8328 - 8.9347)$).

To further illustrate the methods developed in this paper, consider an alternative data generating mechanism mimicking an observational study with a single binary confounder

$$C \sim Bernoulli(1/3);$$

$$E|C \sim Bernoulli((1 + \exp(-0.5 - 0.6C))^{-1});$$

$$N = 50 - 0.75E + 0.5C + N(0, 1);$$

$$M = 30 + 3E + 3C + 2N + N(0, 1);$$

$$u \sim N(0, 2)$$

$$Y = 60 + 2E + 4C + 3M + 1.5N + EN + 2EM + N(0, 1.5)$$

We aim to estimate $NDE(1,0,c)$, which under the assumption of no $M-N$ interaction, is equal to

$$NDE(1,0,1) = 318.63,$$

$$NDE(1,0,0) = 310.13.$$

We further illustrate how this effect estimate can be obtained in Proc NLMIXED in SAS by providing sample code in the appendix. Using this sample code, we obtained the following estimates

$$NDE(1,0,1) = 317.52: \quad 95\% \text{ confidence interval} = (316.62 - 318.43)$$

$$NDE(1,0,0) = 309.60: \quad 95\% \text{ confidence interval} = (308.86 - 310.35)$$

Final Remarks

Natural direct effects have previously been shown not to be nonparametrically identified under an NPSEM when a variable affected by exposure confounds the mediator effects on the outcome. The primary contribution of this paper has been to show that natural direct effects are nonparametrically identified in the simple case of a binary recanting witness, but also in the case where the recanting witness is a vector of binary variables that satisfy a certain independence condition, provided that exposure is known to have a monotonic effect on the recanting witness. For more general settings outside of these special cases, parametric assumptions are required for identification of natural direct effects. For instance, in this paper, we have show that when interaction between the mediator and the recanting witness is absent under an NPSEM, natural direct effects are identified from the observed data. Sample SAS code is provided to facilitate the use of

proposed methods, together with the analysis of a simulated data set which confirms good finite sample performance of the proposed methods.

In addition to the relevance to questions of mediation in observational epidemiology in the context of an exposure-induced mediator-outcome confounder, our results are also of relevance to a broad class of intervention trials. Particularly within the context of experiments in psychology and the behavioral sciences, an intervention (e.g. cognitive behavioral therapy) is randomized to evaluate whether it improves some outcome (e.g. depressive symptoms). Often the theory behind the intervention is motivated by a belief that the intervention will principally operate through some intermediate (e.g. attitudes toward negative life experience) and the intervention is often designed so as to target this intermediate. In the context of such trials it is then of interest to assess mediation. Typically these trials involve some degree of non-compliance. Compliance to treatment is itself of course affected by the treatment, but compliance will likely affect both the intermediate and the outcome. Compliance in these trials serves as a mediator-outcome confounder affected by treatment. Our results are relevant to assessing mediation in such trials whenever compliance is all-or-nothing (i.e. a binary mediator-outcome confounder) or when compliance status and the mediator do not interact in their effects on the outcome (though the treatment and the mediator and the treatment and compliance status would still be allowed to interact). The methods we have developed in this paper provide useful tools for mediation analysis in such settings.

APPENDIX

Proof of Result 1: Assuming the NPSEM $(7)-(11)$,

$$
\begin{aligned}
E\left\{Y\left(e, M\left(e^{*}\right)\right)\right\} &= \sum_{y,m,n,n',c} y f\left(Y\left(e,m,n,c\right)=y, M\left(e^{*},n',c\right)=m, N\left(e,c\right)=n, N(e^{*},c)=n', C=c\right) \\
&= \sum_{y,m,n,n',c} y f\left(Y\left(e,m,n,c\right)=y | M\left(e^{*},n',c\right)=m, N\left(e,c\right)=n, N(e^{*},c)=n', C=c\right) \\
&\times f\left(M\left(e^{*},n',c\right)=m | N\left(e,c\right)=n, N(e^{*},c)=n', C=c\right) \\
&\times f\left(N\left(e,c\right)=n, N(e^{*},c)=n' | C=c\right) \\
&\times f\left(C=c\right) \\
&= \sum_{y,m,n,n',c} y f\left(Y\left(e,m,n,c\right)=y | E=e, M\left(e,n,c\right)=m, N\left(e,c\right)=n, C=c\right) \\
&\times f\left(M\left(e^{*},n',c\right)=m | E=e^{*}, N(e^{*},c)=n', C=c\right) \\
&\times f\left(N\left(e,c\right)=n, N(e^{*},c)=n' | C=c\right) \\
&\times f\left(C=c\right) \qquad \text{(by NPSEM independence)} \\
&= \sum_{y,m,n,n',c} y f\left(Y=y | E=e, M=m, N=n, C=c\right) \\
&\times f\left(M=m | E=e^{*}, N=n', C=c\right) f\left(N(e,c)=n, N(e^{*},c)=n' | C=c\right) f\left(C=c\right)
\end{aligned}
$$

(by consistency)

then, under $E-N$ *Monotonicity Assumption,*

$$\Pr\left(N\left(e\right)=0, N(e^{*})=1 | c\right)=0.$$

and

$$\Pr\left(N\left(e\right)=1, N(e^{*})=1 | c\right)=\Pr\left(\ N(e^{*})=1 | c\right)=\Pr\{N=1 | e^{*}, c\}$$

Similarly,

$$\Pr\left(N\left(e\right)=0,N(e^{*})=0|c\right)=\Pr\left(\ N(e)=0|c\right)=\Pr\{N=0|e,c\}.$$

This implies

$$\begin{aligned}
&\Pr\left(N\left(e\right)=1,N(e^{*})=0|c\right)\\
&=1-\Pr\left(N\left(e\right)=1,N(e^{*})=1|c\right)-\Pr\left(N\left(e\right)=0,N(e^{*})=0|c\right)\\
&=\Pr\{N=1|e,c\}-\Pr\{N=1|e^{*},c\}
\end{aligned}$$

proving the result.

Proof of Result 2:

$$\begin{aligned}
E\left\{Y\left(e,M\left(e^{*}\right)\right)\right\}=&\sum_{y,m,n,n',c}yf\left(Y=y|E=e,M=m,N=n,C=c\right)\\
&\times f\left(M=m|E=e^{*},N=n^{*},C=c\right)\\
&\times f\left(N(e,c)=n,N(e^{*},c)=n^{*}|C=c\right)\\
&\times f\left(C=c\right)\\
=&\sum_{y,m,n,n',c}yf\left(Y=y|E=e,M=m,N=n,C=c\right)\\
&\times f\left(M=m|E=e^{*},N=n^{*},C=c\right)\\
&\times\prod_{j=1}^{k}f\left(N_{j}(e,c)=n_{j},N_{j}(e^{*},c)=n_{j}^{'}|C=c\right)f\left(C=c\right)
\end{aligned}$$

by independence of $(N_1(e,c),N_2(e^*,c))$,...$(N_j(e,c),N_j(e^*,c))$. *Next by monotonicity of the effect*

of E on each N_j, one obtains as in the proof of Result 1

$$f\left(N_j(e,c)=n_j, N_j(e^*,c)=n_j^{'}|C=c\right)=f_j\left(n_j,n_j^{'},e,e^*,c\right)$$

proving the result.

Result 3: Assuming the NPSEM $(7)-(11)$, *suppose that the* $E-N$ *Monotonicity Assumption holds for each of* $(N_1,...,N_k)=N$, *then the hazard function of* $Y\left(e,M\left(e^*\right)\right)$ *evaluated at* y *is nonparametrically identified and satisfies an additive hazards model of the form:*

$$\lambda_0(y)$$
$$+\frac{\sum_{m,n,n',c}\gamma\left(y,e,m,n,c\right)\Gamma\left(y,e,m,n,c\right)\Pr(M=m|\ e^*,\ n^{'},c)\prod_{j=1}^{k}f_j\left(n_j,n_j^{'},e,e^*,c\right)\Pr(C=c)}{\sum_{m,n,n',c}\ \Gamma\left(y,e,m,n,c\right)\Pr(M=m|\ e^*,\ n^{'},c)\prod_{j=1}^{k}f_j\left(n_j,n_j^{'},e,e^*,c\right)\Pr(C=c)}$$

where $\Gamma\left(y,e,m,n,c\right)=\exp\left\{-\int_0^y\gamma\left(u,e,m,n,c\right)du\right\}$ *and* $f_j\left(n_j,n_j^{'},e,e^*,c\right)$ *defined in Result 2.*

Proof of Result 3: By Result 1, the log-survival curve of $Y\left(e,M\left(e^*\right)\right)$ at y is given by

$$\log\sum_{m,n,n',c}\exp\left\{-\int_0^y[\lambda_0(y)+\gamma\left(u,e,m,n,c\right)]du\right\}$$
$$\times f\left(M=m|E=e^*,N=n',C=c\right)\prod_{j=1}^{k}f_j\left(n_j,n_j^{'},e,e^*,c\right)f\left(C=c\right)$$

The result follows upon differentiation of this function with respect to y *and multiplication by* (-1).

Proof of Result 4: Recall that $\mathbb{E}\{Y(e, M(e^*))\}$

$$= \sum_{m,n,n',c} \mathbb{E}(Y|e,m,n,c)\, f(M=m|E=e^*, N=n', C=c)$$

$$\times f(N(e,c)=n, N(e^*,c)=n'|C=c)\, f(C=c)$$

$$= \sum_{m,n,n',c} \left(\beta_m(e,m,c) + \beta_n(e,n,c) + \overbrace{\beta_{m,n}(e,m,n,c)}^{=0 \text{ by assumption}} + \overline{\beta}_{e,c}(e,c) \right)$$

$$\times f(M=m|E=e^*, N=n', C=c)\, f(N(e,c)=n, N(e^*,c)=n'|C=c)\, f(C=c)$$

$$= \sum_{m,n,n',c} \left(\beta_m(e,m,c) + \beta_n(e,n,c) + \overline{\beta}_{e,c}(e,c)\right)$$

$$\times f(M=m|E=e^*, N=n', C=c)\, f(N(e,c)=n, N(e^*,c)=n'|C=c)\, f(C=c)$$

$$= \sum_{m,n',c} \beta_m(e,m,c)\, f(M=m|E=e^*, N=n', C=c)\, f(N(e^*,c)=n'|C=c)\, f(C=c)$$

$$+ \sum_{n,c} \beta_n(e,n,c)\, f(N(e,c)=n|C=c)\, f(C=c) + \sum_{c} \overline{\beta}_{e,c}(e,c)\, f(C=c)$$

$$= \sum_{m,c} \beta_m(e,m,c)\, f(M=m|E=e^*, C=c)\, f(C=c)$$

$$+ \sum_{n,c} \beta_n(e,n,c)\, f(N=n|E=e, C=c)\, f(C=c) + \sum_{c} \overline{\beta}_{e,c}(e,c)\, f(C=c)$$

proving the result.

SAS CODE

The data set 'data_example' contains variables E, N, M, Y from the simulated randomized study example.

The first sample code produced the maximum likelihood estimate of $NDE(1,0)$ reported in the text.


```
proc nlmixed data=sample_example;
```

```
parms alpha_0=1 alpha_e=2 alpha_m=3 alpha_n=4 alpha_mn=2
theta_0=2 theta_e=3 theta_n=1
eta_0=-1 eta_e=0.4 sigma_y=0.5 sigma_m=2;
MuY= alpha_0+alpha_e*E+alpha_m*M+alpha_n*N
+alpha_mn*M*N;
ll_y=-((Y-MuY)**2)/(2*sigma_y)-0.5*log(sigma_y);
MUm=theta_0+theta_e*E+theta_n*N;
ll_m=-((M-MuM)**2)/(2*sigma_M)-0.5*log(sigma_M);
p_n=(1+exp(-(eta_0+eta_e*E)))**-1;
ll_n= N*log (p_n)+(1-N)*log(1-p_n);
ll_o=ll_y+ll_m+ll_n;
omega= (1+exp(-(eta_0+eta_e)))**-1-(1+exp(-(eta_0)))**-1;
model Y ~general(ll_o);
estimate 'nde' alpha_e+(alpha_mn*theta_0+alpha_n)*omega;
run;
```

The data set 'data_example_2' contains variables C, E, N, M, Y from the simulated observation study example.The following sample code produces the maximum likelihood estimates of $NDE(1, 0, c), c = 0, 1$, reported in the text.

```
proc nlmixed data=sample_example;
parms alpha_0=1 alpha_e=2 alpha_c=1 alpha_m=3 alpha_n=4
alpha_me=2 alpha_ne=1 theta_0=2 theta_e=3 theta_c=1 theta_n=0.5
eta_0=-1 eta_e=0.5 eta_c=1 sigma_y=0.5 sigma_m=2 sigma_n=1 ;
MuY= alpha_0+alpha_c*c+alpha_e*E+alpha_m*M+alpha_n*N+alpha_ne*E*N
```

```
+alpha_me*E*M;
ll_y=-((Y-MuY)**2)/(2*sigma_y)-0.5*log(sigma_y);
MUm=theta_0+theta_c*C+theta_e*E+theta_n*N;
ll_m=-((M-MuM)**2)/(2*sigma_M)-0.5*log(sigma_M);
MUn=eta_0+eta_c*C+eta_e*E;
ll_n=      -((N-MuN)**2)/(2*sigma_N)-0.5*log(sigma_N);
ll_o=ll_y+ll_m+ll_n;
theta_00 = theta_0+theta_n*eta_0;
theta_cc = theta_c+theta_n*eta_c;
theta_ee = theta_e+theta_n*eta_e;
model N~general(ll_o);
estimate 'nde(1,0,1)' alpha_e+alpha_me*(theta_00+theta_cc)
+alpha_n*eta_e+alpha_ne*(eta_0+eta_e+eta_c);
estimate 'nde(1,0,0)' alpha_e+alpha_me*(theta_00)
+alpha_n*eta_e+alpha_ne*(eta_0+eta_e);
run;
```

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] Robins, J. (2003). Semantics of causal DAG models and the identification of direct and indirect effects. In P. Green, N. Hjort, and S. Richardson (Eds.), Highly Structured Stochastic Systems, pp. 70–81. Oxford, UK: Oxford University Press.

[4] Avin, C., I. Shpitser, and J. Pearl (2005). Identifiability of path-specific effects. In IJCAI-05, Proceedings of the Nineteenth International Joint Conference on Artificial Intelligence, Edinburgh, Scotland, UK, July 30-August 5, 2005, pp. 357–363.

[5] Petersen M, Sinisi, S; van der Laan, M.(2006) Estimation of Direct Causal Effects.Epidemiology. Volume 17, 3, 276-284.

[6] van der Laan, M, Petersen, M. (2005) Direct Effect Models. U.C. Berkeley Division of Biostatistics Working Paper Series. Working Paper 187. http://www.bepress.com/ucbbiostat/paper187

[7] Goetgeluk, S., Vansteelandt, S. and Goetghebeur, E. (2008). Estimation of controlled direct effects. Journal of the Royal Statistical Soceity, Series B, 70:1049-1066.

[8] Vansteelandt, S. (2009). Estimating direct effects in cohort and case-control studies. Epidemiology, 20:851-860.

[9] VanderWeele, T.J. (2009). Marginal structural models for the estimation of direct and indirect effects. Epidemiology, 20:18-26.

[10] VanderWeele, T.J. and Vansteelandt, S. (2009). Conceptual issues concerning mediation, interventions and composition. Statistics and its Interface, 2, 457-468.

[11] 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.

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

[13] Tchetgen Tchetgen, E J (2011) On Causal Mediation Analysis with a Survival Outcome.International Journal of Biostatistics: Vol. 7: Iss. 1, Article 33.

[14] Tchetgen Tchetgen EJ and Shpitser I (2012) Semiparametric Theory for Causal Mediation Analysis: efficiency bounds, multiple robustness, and sensitivity analysis. Annals of Statistics. In press

[15] Rubin DB. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. Journal Educational Psychology, 66:688-701.

[16] Rubin DB. (1978). Bayesian inference for causal effects: The role of randomization. Annals of Statistics, 6:34-58.

[17] Pearl, J. *Causality: Models, Reasoning, and Inference*, 2nd ed. New York: Cambridge University Press; 2009.

[18] Robins JM. (1997). Causal Inference from Complex Longitudinal Data. Latent Variable Modeling and Applications to Causality. Lecture Notes in Statistics (120), M. Berkane, Editor. NY: Springer Verlag, pp. 69-117.

[19] Robins JM, Richardson, TS. (2010). Alternative graphical causal models and the identification of direct effects. To appear in Causality and Psychopathology: Finding the Determinants of Disorders and Their Cures. P. Shrout, Editor. Oxford University Press.

[20] Angrist J, Imbens GW, Rubin DB. Identification of causal effects using instrumental variables. J Am Stat Assoc. 1996;91:444–455

[21] VanderWeele, T.J. and Robins, J.M. (2009). The properties of monotonic effects on directed acyclic graphs. Journal of Machine Learning Research - Special Topic on Causality, 10:699-718.

[22] VanderWeele, T.J. and Tchetgen Tchetgen, E.J. (2011). Bounding the infectiousness effect in vaccine trials. Epidemiology, 22:686-693.

[23] Suzuki E, Yamamoto E and Tsuda T.On the Relations Between Excess Fraction, Attributable Fraction, and Etiologic Fraction Am. J. Epidemiol. first published online February 16, 2012 doi:10.1093/aje/kwr333.

[24] Tein, J.-Y., MacKinnon, D.P. (2003). Estimating mediated effects with survival data. In H. Yanai, A. O. Rikkyo, K. Shigemasu, Y. Kano, & J. J. Meulman (Eds.): New Developments on Psychometrics. p. 405-412. Tokyo, Japan: Springer-Verlag Tokyo Inc.

[25] Lange, T. and Hansen, J.V. (2011). Direct and indirect effects in a survival context. Epidemiology, 22:575-581.

[26] VanderWeele, T.J. (2011b). Causal mediation analysis with survival data. Epidemiology, 22:582-585.

[27] Tchetgen Tchetgen, E.J. (2011). On causal mediation analysis with a survival outcome. The International Journal of Biostatistics: Vol. 7: Iss. 1, Article 33.

[28] Vansteelandt S., VanderWeele T.J., Tchetgen E, Robins J.M., Multiply robust inference for statistical interactions. Journal of the American Statistical Association, December 2008, Vol. 103, No. 484, Theory and Methods

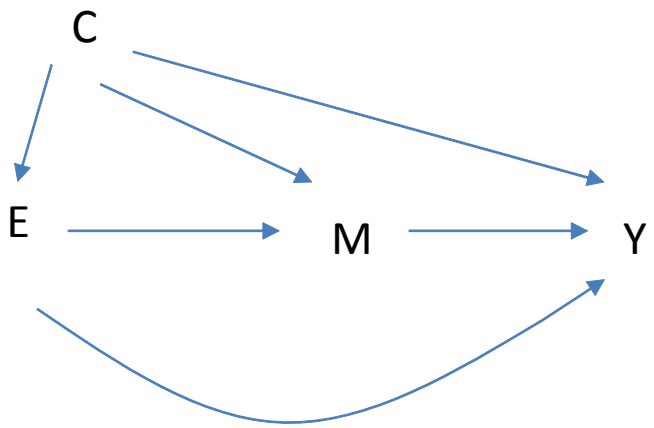

Figure 1. No confounder of M-Y relation is affected by E

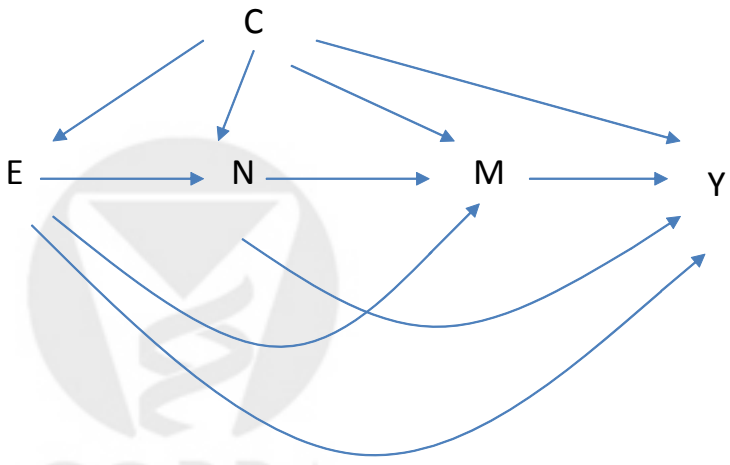

Figure 2. N is a confounder of M-Y relation that is affected by E