HHS Public Access
Author manuscript
J R Stat Soc Series B Stat Methodol. Author manuscript; available in PMC 2018 June 01.

Published in final edited form as:
J R Stat Soc Series B Stat Methodol. 2017 June ; 79(3): 917–938. doi:10.1111/rssb.12194.

Mediation analysis with time varying exposures and mediators

Tyler J. VanderWeele and **Eric J. Tchetgen Tchetgen**
Harvard T.H. Chan School of Public Health, Departments of Biostatistics and Epidemiology, 677 Huntington Avenue, Boston MA 02115, USA

Summary

In this paper we consider causal mediation analysis when exposures and mediators vary over time. We give non-parametric identification results, discuss parametric implementation, and also provide a weighting approach to direct and indirect effects based on combining the results of two marginal structural models. We also discuss how our results give rise to a causal interpretation of the effect estimates produced from longitudinal structural equation models. When there are time-varying confounders affected by prior exposure and mediator, natural direct and indirect effects are not identified. However, we define a randomized interventional analogue of natural direct and indirect effects that are identified in this setting. The formula that identifies these effects we refer to as the "mediational g-formula." When there is no mediation, the mediational g-formula reduces to Robins' regular g-formula for longitudinal data. When there are no time-varying confounders affected by prior exposure and mediator values, then the mediational g-formula reduces to a longitudinal version of Pearl's mediation formula. However, the mediational g-formula itself can accommodate both mediation and time-varying confounders and constitutes a general approach to mediation analysis with time-varying exposures and mediators.

Keywords

counterfactual; direct and indirect effect; longitudinal data; mediation; pathway analysis; time-varying confounding

1. Introduction

There has recently been considerable methodologic development on approaches to mediation and pathway analysis from within the causal inference literature (Robins and Greenland, 1992; Pearl, 2001; van der Laan and Petersen, 2008; Goetgeluk et al., 2008; VanderWeele and Vansteelandt, 2009; Imai et al., 2010; Tchetgen Tchetgen and Shpitser, 2012, 2014; Lange and Hansen, 2011; Martinusen et al., 2011; Vansteelandt et al., 2012; VanderWeele, 2015). This work has extended traditional approaches for mediation analysis to settings with interactions and non-linearities and has clarified the no-unmeasured confounding assumptions that suffice for a causal interpretation of direct and indirect effects. Almost all of this literature has considered a single exposure at one point in time, a single mediator, and a single outcome. There is also now a literature on a single exposure, mediator and outcome but with a time-dependent confounder that is affected by the exposure and which itself affects both the mediator and the outcome (Albert and Nelson, 2011; Imai and Yamamoto, 2013; VanderWeele et al., 2014; Tchetgen Tchetgen and VanderWeele, 2014;

Author Manuscript
Author Manuscript
Author Manuscript
Author Manuscript

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Daniel et al., 2015); however this literature also does not allow the exposures and the mediators themselves to vary over time. In practice, often longitudinal data are available and both the exposure and the mediator vary over time. There is currently very little work in the causal inference literature with exposures and mediators that vary over time. Only a few papers in the causal inference briefly touch on such settings with longitudinal data (van der Laan and Petersen, 2008; VanderWeele, 2009; Shpitser, 2013) and an approach that fully accommodates time-varying exposures and mediators and time-varying confounding is yet to be developed. Although some work has been done in psychology on mediation analysis with longitudinal data (cf. MacKinnon, 2008), this does not fall within a formal causal framework and issues of time-varying confounding are not addressed.

Some of the difficulty is that the concepts of natural direct and indirect effects (Robins and Greenland, 1992; Pearl, 2001) that have been employed in the causal inference literature on mediation are not identified from the data in many settings involving time-varying exposures and mediators. In particular whenever there is a mediator-outcome confounder affected by the exposure, these natural direct and indirect effects are not non-parametrically identified irrespective of whether data is available on the exposure-induced confounder or not (Avin et al., 2005). In longitudinal settings such exposure-induced confounding may be very common. In this paper we propose an approach to pathway analysis that can be used in settings with time-varying exposures and mediators. To do so, instead of using the natural direct and indirect effects commonly employed in the literature we use a randomized interventional analogue of natural direct and indirect effects (cf. Didelez et al., 2006; VanderWeele et al., 2014) that can be identified from longitudinal data under weaker assumptions than the natural direct and indirect effects. The approach we develop draws upon both Robins' g-formula (Robins, 1986) and Pearl's mediation formula (Pearl, 2001) but unites these in a single framework that allows one to assess mediation with time-varying exposures and mediators in the presence of time-varying confounders. We will refer to the resulting empirical expression as the mediational g-formula. In the absence of time-varying confounders it reduces to a time-varying analogue of Pearl's mediational formula. In the absence of mediation it reduces to Robins' g-formula. However, the approach with the mediational g-formula can handle both mediation and time-varying confounding; it unites the g-formula and the mediation formula together in a single framework. It is applicable to assess questions of mediation over a broad range of contexts. We illustrate how the approach can be implemented by fitting two marginal structural models (with code in the Online Supplement); we also show how it can be implemented using sets of linear structural equation models common in the social sciences (MacKinnon, 2008) and how it clarifies the interpretation of effects in this context. However, these are only two settings in which the approach can be used. It is much more general and can be applied to numerous settings. We believe the mediational g-formula will lie at the foundation of numerous future developments concerning the assessment of mediation with time-varying exposures and mediators.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

2. Natural Direct and Indirect Effects Versus Randomized Interventional Analogues

In this section we will review the definitions and identification assumptions for the natural direct and indirect effects defined in the causal inference literature on mediation. We will moreover contrast this to randomized interventional analogues of natural direct and indirect effects which can be identified under weaker assumptions and which will, in the following section, be extended to settings with time-varying exposures and mediators.

Let A denote the exposure of interest; Y, the outcome and M, the potential mediator, and V a set of baseline covariates not affected by the exposure. For now we will assume that the exposure and mediator only occur at one point in time. We will let Y_a and M_a denote, respectively, the values of the outcome and mediator that would have been observed had exposure A been set to level a. We will let Y_{am} denote the value of the outcome that would have been observed had exposure A been set to level a, and mediator M been set to level m. These counterfactual or potential outcome variables, Y_a, M_a and Y_{am} all presuppose that at least hypothetical interventions on A and M are conceivable. A further assumption is generally made, sometimes referred to as the “consistency assumption”, that when the observed exposure $A = a$, the counterfactual outcomes Y_a and M_a are, respectively, equal to the observed outcomes Y and M, and likewise when observed $A = a$ and $M = m$, the counterfactual outcome Y_{am} is equal to Y, along with a “composition” assumption that $Y_a = Y_{aM_a}$.

Using these counterfactuals, Robins and Greenland (1992) and Pearl (2001) defined what have since come to be called controlled direct effects and natural direct and indirect effects. The average controlled direct effect, conditional on covariates $V = v$, comparing exposure level $A = a$ with $A = a^*$ (for a binary exposure $a = 1$, $a^* = 0$) and fixing the mediator to level m, is defined by $E[Y_{am} - Y_{a^*m}|v]$ and captures the effect of exposure A on outcome Y, intervening to fix M to m; it may be different for different levels of m. The natural direct effect, conditional on covariates $V = v$, is defined as $E[Y_{aM_{a^*}} - Y_{a^*M_{a^*}}|v]$ and differs from controlled direct effects in that the intermediate M is set to the level M_{a^*}, the level that it would have naturally been if the exposure had taken value $A = a^*$. Similarly, the average natural indirect effect, conditional on $V = v$, can be defined as $E[Y_{aM_a} - Y_{aM_{a^*}}|v]$, which compares the effect of the mediator at levels M_a and M_{a^*}on the outcome when exposure is set to $A = a$. Natural direct and indirect effects have the property that a total effect, $E[Y_a - Y_{a^*}|v]$, decomposes into a natural direct and indirect effect: $E[Y_a - Y_{a^*}|v] = E[Y_{aM_a} - Y_{a^*M_{a^*}}|v] = E[Y_{aM_a} - Y_{aM_{a^*}}|v] + E[Y_{aM_{a^*}} - Y_{a^*M_{a^*}}|v]$; the decomposition holds even when there are interactions and non-linearities.

In general, stronger no-unmeasured-confounding assumptions are required to identify direct and indirect effects than total effects. On a causal diagram interpreted as a set of non-parametric structural equations (Pearl, 1995, 2009), the following four assumptions suffice to identify natural direct and indirect effects from data (Pearl, 2001; Shpitser and VanderWeele, 2011): (i) the effect the exposure A on the outcome Y is unconfounded conditional on V; (ii) the effect the mediator M on the outcome Y is unconfounded conditional on (V, A); (iii) the effect the exposure A on the mediator M is unconfounded

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

conditional on V; and (iv) none of the mediator-outcome confounders are affected by the exposure. The assumptions would hold in a nonparametric structural equation model given by Figure 1.

Only assumptions (i) and (ii) are required to estimate controlled direct effects. Assumptions (i)–(iv) in the text, stated formally in terms of counterfactual independence, are: (i) $Y_{am} \perp\!\!\!\perp A|V$, (ii) $Y_{am} \perp\!\!\!\perp M|\{A, V\}$, (iii) $M_a \perp\!\!\!\perp A|V$, (iv) $Y_{am} \perp\!\!\!\perp M_{a^*}|V$. Under these assumptions natural direct and indirect effects are identified (Pearl, 2001) and given by the following expressions:

$$\begin{aligned} E[Y_{aM_{a^*}} - Y_{a^*M_{a^*}}|v] &= \textstyle\sum_m \{E[Y|a,m,v] - E[Y|a^*,m,v]\}P(m|a^*,v). \\ E[Y_{aM_a} - Y_{aM_{a^*}}|v] &= \textstyle\sum_m E[Y|a,m,v]\{P(m|a,v) - P(m|a^*,v)\}. \end{aligned}$$

Importantly, however, if there is a mediator-outcome confounder L affected by exposure then assumption (iv) will fail and natural direct and indirect effects will not be identified from the data (Avin et al., 2005). Assumption (iv) would thus be violated in Figure 2. The counterfactual independence assumption (iv) that $Y_{am} \perp\!\!\!\perp M_{a^*}|V$ is also somewhat controversial for other reasons. Although it will hold in the causal diagram in Figure 1 if this diagram is interpreted as a non-parametric structural equation model as in Pearl (2009), there are other interpretations of causal diagrams wherein assumption (iv) may fail even in Figure 1 (Robins, 2003; Robins and Richardson, 2010) because these alternative interpretations impose fewer conditional counterfactual independencies than are implied by a structural equation model. Further discussion is provided elsewhere (Robins and Richardson, 2010; VanderWeele, 2015).

Even if this assumption, that $Y_{am} \perp\!\!\!\perp M_{a^*}|V$, fails, an analogue of natural direct and indirect effects, based on randomized interventions, can be identified from the data under assumptions (i)-(iii) alone. We will conclude this section with a discussion of these randomized interventional analogues of natural direct and indirect effects and in the following section we will consider longitudinal extensions of these effects. These randomized interventional analogues are essentially equivalent to those proposed by Didelez et al. (2006) and Geneletti (2007), but here we employ and extend these concepts to a longitudinal context for mediation.

Let $G_{a|v}$ denote a random draw from the distribution of the mediator with exposure status fixed to a conditional on $V = v$. The effect $E(Y_{aG_{a|v}}) - E(Y_{aG_{a^*|v}})$ is then the effect on the outcome of randomly assigning an individual who is given the exposure to a value of the mediator from the distribution of the mediator amongst those given exposure versus not given exposure (conditional on the covariates); this is an effect through the mediator. Next consider the effect $E(Y_{aG_{a^*|v}}) - E(Y_{a^*G_{a^*|v}})$; this is a direct effect comparing exposure versus no exposure with the mediator in both cases randomly drawn from the distribution of the population when given no exposure (conditional on the covariates). Finally, the overall effect $E(Y_{aG_{a|v}}) - E(Y_{a^*G_{a^*|v}})$ compares the expected outcome when (conditional on the covariates) having the exposure with the mediator randomly drawn from the distribution of the population when given the exposure (conditional on covariates) to the expected outcome

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

when not having the exposure with the mediator randomly drawn from the distribution of the population when not exposed. With effects thus defined we have the decomposition: $E(Y_{aG_{a|v}}) - E(Y_{a^*G_{a^*|v}}) = \{E(Y_{aG_{a|v}}) - E(Y_{aG_{a^*|v}})\} + \{E(Y_{aG_{a^*|v}}) - E(Y_{a^*G_{a^*|v}})\}$so that the overall effect decomposes into the sum of the effect through the mediator and the direct effect. We will refer to these effects as interventional direct and indirect effects since, unlike the natural direct and indirect effects, they are effects that could in principle be brought about in practice by interventions on the exposure and the mediator. These are not the natural direct and indirect effects considered earlier but are instead analogues arising from fixing the mediator for each individual, not to the level it would have been that for individual under a particular exposure, but rather, to a level that is randomly chosen from the distribution of the mediator amongst all of those with a particular exposure, conditional on the covariates. These effects are identified under assumptions (i)–(iii) alone (Vander-Weele et al., 2014). Under these assumptions (i)–(iii) the interventional direct and indirect, $\{E(Y_{aG_{a^*|v}}) - E(Y_{a^*G_{a^*|v}})\}$and $\{E(Y_{aG_{a|v}}) - E(Y_{aG_{a^*|v}})\}$, are in fact identified by the same empirical expression as those given above for natural direct and indirect effects, and the interventional total effect equals the regular total effect, points we will return to again below in the longitudinal setting. Note that assumption (iv) is not necessary for the identification of these interventional effects; it is not necessary because the mediator is being fixed to a level that is randomly chosen from the distribution of the mediator amongst all of those with a particular exposure, rather than fixed to the level it would have been for that individual under a different exposure status. Because assumptions (iv) is not necessary the interventional direct and indirect effects are also identified in interpretations of causal diagrams (Robins and Richardson, 2010) other than Pearl's non-parametric structural equations (cf. Vander-Weele et al., 2014). Moreover, even if there is a mediator-outcome confounder affected by the exposure as in Figure 2, the interventional direct and indirect effects may still be identified from the data but the empirical expressions equal to these effects no longer coincide with that given above for natural direct and indirect effects. They are instead, if Figure 2 is a causal diagram, given by (VanderWeele, et al., 2014):

$$E(Y_{aG_{a^*|v}}) - E(Y_{a^*G_{a^*|v}}) = \sum_{l,m}\{E[Y|a,l,m,v]P(l|a,v) - E[Y|a^*,l,m,v]P(l|a^*,v)\}P(m|a^*,v)$$

$$E(Y_{aG_{a|v}}) - E(Y_{aG_{a^*|v}}) = \sum_{l,m}E[Y|a,l,m,v]P(l|a,v)\{P(m|a,v) - P(m|a^*,v)\}.$$

3. Time-Varying Exposures and Mediators and the Mediational G-Formula

Suppose now that the exposure, mediators and possibly confounding variables vary over time. Let $(A(1), \ldots, A(T))$, $(M(1), \ldots, M(T))$, and $(L(1), \ldots, L(T))$ denote values of the exposures, mediator, and time-varying confounders at periods $0, \ldots, T$, with initial baseline covariates V, and subsequent temporal ordering $A(t)$, $M(t)$, $L(t)$. We will revisit this question of temporal ordering again later in the paper. The relationships among the variables are given in Figure 3. In what follows it is in principle possible to allow the mediator at each time $M(t)$ to denote a vector of mediators to allow for assessing mediation over time through a set of time-varying mediators.

For any variable W, let $\bar{W}(t) = (W(1), \ldots, W(t))$ and let $\bar{W} = \bar{W}(T) = (W(1), \ldots, W(T))$. Let $\underline{W}(t) = (W(t), \ldots, W(T))$. By convention, we let $W(t)$ denote the empty set for $t \leq 0$. Let $Y_{\overline{am}}$

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

be the counterfactual outcome if $\bar{A}$ were set to $\bar{a}$ and if $\bar{M}$ were set to $\bar{m}$. Let $M_{\bar{a}}(t)$ be the counterfactual value of $M(t)$ if $\bar{A}$ were set to $\bar{a}$. We assume consistency that when observed $\bar{A} = \bar{a}$ we have $M_{\bar{a}}(t) = M(t)$ and $Y_{\bar{a}}(t) = Y(t)$ and when observed $\bar{A} = \bar{a}$ and $\bar{M} = \bar{m}$ we have $Y_{\overline{am}} = Y$.

Note that if the entire vector $A = (A(1), \ldots, A(T))$ is taken as the exposure and $M = (M(1), \ldots, M(T))$ is taken as the mediator then the variable $L(1)$ is itself affected by the exposure (namely, by $A(1)$) and in turn confounds the mediator-outcome relationship between $M(2)$ and Y. From this it follows that natural direct and indirect effects are not identified in this setting (Avin et al., 2005). However, identification of interventional direct and indirect effects may once again be possible.

Let $\bar{G}_{\bar{a}|v}(t)$ denote a random draw from the distribution of the mediator $\bar{M}(t)$ that would have been observed in the population with baseline covariates $V = v$ if exposure status $\bar{A}$ had been fixed to $\bar{a}$. Note that at time t, $\bar{G}_{\bar{a}|v}(t)$ will only depend on $\bar{a}$ through time t. Let $\bar{a}$ and $\bar{a}^*$ be two distinct exposure histories. We once again have a decomposition, even with time-varying exposures and mediators: $E(Y_{\bar{a}\bar{G}_{\bar{a}|v}(t)}|v) - E(Y_{\bar{a}^*\bar{G}_{\bar{a}^*|v}}|v) = \{E(Y_{\bar{a}\bar{G}_{\bar{a}|v}}|v) - E(Y_{\bar{a}\bar{G}_{\bar{a}^*|v}}|v)\} + \{E(Y_{\bar{a}\bar{G}_{\bar{a}^*|v}}|v) - E(Y_{\bar{a}^*\bar{G}_{\bar{a}^*|v}}|v)\}$ with $\{E(Y_{\bar{a}\bar{G}_{\bar{a}|v}}|v) - E(Y_{\bar{a}\bar{G}_{\bar{a}^*|v}}|v)\}$ being the interventional indirect effect and $\{E(Y_{\bar{a}\bar{G}_{\bar{a}^*|v}}|v) - E(Y_{\bar{a}^*\bar{G}_{\bar{a}^*|v}}|v)\}$ the interventional direct effect. These effects will of course vary according to the exposure trajectories $\bar{a}$ and $\bar{a}^*$ being compared. For a binary exposure a common choice would be comparing $\bar{1}$ and $\bar{0}$. For a continuous exposure one possible choice would be taking $\bar{a}$ and $\bar{a}^*$ each as a constant separated by one standard deviation difference in the exposure distribution centered at the mean. But any two trajectories can in fact be compared.

The decomposition above is a decomposition of the interventional overall effect, $E(Y_{\bar{a}\bar{G}_{\bar{a}|v}(t)}|v) - E(Y_{\bar{a}^*G_{\bar{a}^*|v}}|v)$, into interventional direct and indirect effects. We can, however, also decompose an average treatment effect (just setting the exposure itself to different levels) into analogous components. In this setting, the average treatment effect, conditional on baseline covariates $V = v$, comparing exposure trajectories $\bar{a}$ and $\bar{a}^*$, is simply $E(Y_{\bar{a}}|v) - E(Y_{\bar{a}^*}|v)$. We can decompose this effect as follows $E(Y_{\bar{a}}|v) - E(Y_{\bar{a}^*}|v) = \{E(Y_{\bar{a}}|v) - E(Y_{\bar{a}\bar{G}_{\bar{a}^*|v}}|v)\}+\{E(Y_{\bar{a}\bar{G}_{\bar{a}^*|v}}|v) - E(Y_{\bar{a}^*}|v)\}$, where the first component, $E(Y_{\bar{a}}|v) - E(Y_{\bar{a}\bar{G}_{\bar{a}^*|v}}|v)$, is an interventional analogue of the natural indirect effect and examines how the outcome under exposure $\bar{a}$ would change if the mediator were fixed for each individual to a random draw from the distribution of the mediator under exposure $\bar{a}^*$ i.e. from $\bar{M}_{\bar{a}^*}$; and the second component, $E(Y_{\bar{a}\bar{G}_{\bar{a}^*|v}}|v) - E(Y_{\bar{a}^*}|v)$, is similarly an interventional analogue of the natural direct effect. The downside of the decomposition of the average treatment effect is that the direct effect here captures both the effect of changing the exposure from $\bar{a}^*$ to $\bar{a}$ but also the effect of having the mediator set to its natural level under $\bar{a}^*$ versus a random draw from the mediator under $\bar{a}^*$ since $E(Y_{\bar{a}\bar{G}_{\bar{a}^*|v}}|v) - E(Y_{\bar{a}^*}|v) = \{E(Y_{\bar{a}\bar{G}_{\bar{a}^*|v}}|v) - E(Y_{\bar{a}\bar{M}_{\bar{a}^*|v}}|v)\}+ \{E(Y_{\bar{a}\bar{M}_{\bar{a}^*|v}}|v) - E(Y_{\bar{a}^*\bar{M}_{\bar{a}^*|v}}|v)\}$, where the second term is the natural direct effect but the first term essentially captures the difference between having the mediator set to its natural level under $\bar{a}^*$ versus a random draw from the mediator under $\bar{a}^*$. This was not an issue with the direct effect in the decomposition of the interventional overall effect above. However, even with the decomposition of the average treatment effect, it is not an issue with the interventional indirect effect (which, within the context of mediation, will often be what is of

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

interest), since $E(Y_{\bar{a}}|v) - E(Y_{\bar{a}\bar{G}_{\bar{a}^*}|v}|v) = \{E(Y_{\bar{a}\bar{M}_{\bar{a}}|v}|v) - E(Y_{\bar{a}\bar{M}_{\bar{a}^*}|v}|v)\}+\{E(Y_{\bar{a}\bar{M}_{\bar{a}^*}|v}|v) - E(Y_{\bar{a}\bar{G}_{\bar{a}^*}|v}|v)\}$ where the first term is the natural indirect effect and the second still captures an effect through the mediator i.e. the effect of having the mediator set to its natural level under $\bar{a}^*$ versus a random draw from the mediator under $\bar{a}^*$. In any case, in the various decomposition above, the central task becomes identifying the expected counterfactual $E(Y_{\bar{a}\bar{G}_{\bar{a}^*}|v}|v)$ and we give a result below with an empirical expression to identify this quantity.

Although these interventional direct and indirect effects defined here are not identical with natural direct and indirect, they are in some sense the best we may be able to do as the natural direct and indirect effects themselves will not be identified when a mediator-outcome confounder is affected by the exposure; in such settings the interventional direct and indirect effects are then all that we can estimate. Moreover, several further comments merit attention. First, these interventional effects do in some sense capture mediated effects and pathways; the interventional indirect effect, $\{E(Y_{\bar{a}G_{\bar{a}}|v}|v) - E(Y_{\bar{a}G_{\bar{a}^*}|v}|v)\}$, will be non-zero only if the exposure changes the distribution of the mediator and that change in the distribution of the mediator changes the outcome. Second, when there are no mediator-outcome confounders affected by the exposure, it will be seen below that the interventional direct and indirect effects in fact do coincide with natural direct and indirect effects; thus when the latter effects are identified the interventional analogues in fact capture these effects. Third, when natural direct and indirect effects are not identified, it will only be in extremely pathological settings that the interventional analogue is non-zero, but there are in fact no natural indirect effect. For that to occur, it would be necessary that the exposure affects the mediator for a completely different set of individual than for whom the mediator affects the outcome i.e. there is no overlap in those for whom the exposure affects the mediator and for whom the mediator affects the outcome. Conversely for there to be a non-zero natural indirect effect with a zero interventional analogue of that effect would essentially require exact cancellations to occur. Nevertheless, the interventional analogues will not correspond exactly to the natural individual level variation because they ensure only that the distributions of the mediator are similar across exposure groups. In related work (Miles et al., 2015) we have considered bounds related to how far various empirical expressions can deviate from the true natural direct and indirect effects when these effects are not identified. Further extensions of that work to the time-varying setting would be of interest.

Finally, there are arguably some settings in which the interventional direct and indirect effects are in fact what is of principal substantive interest, rather than the natural direct and indirect effects. Suppose we were interested in whether a racial health disparity (race constituting the exposure, and health the outcome) was in some sense mediated by differences in socioeconomic distributions. The natural direct and indirect effects would entail hypothetical interventions on the mediator of fixing a black individual's socioeconomic status to what it would have been had they been white. Counterfactual queries of the form of what a black individual's socioeconomic status would have been had they been of a different race strike most people as strange or meaningless. Just as above we showed how we decomposed not simply the overall interventional effect, but also the average treatment effect using the interventional analogue ideas so also we can do something similar with disparities i.e. with actual observed health differences between black and white

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

individuals. If the exposure variable A for example indicates race (e.g. $A = 1$ for black and $A = 0$ for white), and M some marker of socioeconomic status (e.g. elementary school test scores), and Y some health outcome we could first consider the actual observed racial disparity, $E(Y|A = 1, v) - E(Y|A = 0, v)$. In this context let $Y_{G_{0|v}}$ denote the outcome that we would have observed if we fixed the socioeconomic variable M to a random draw from the underlying distribution of the mediator $\bar{M}(t)$ that would have been observed in the white population with $A = 0$ and baseline covariates $V = v$. We could then compare the actual racial disparity, $E(Y|A = 1, v) - E(Y|A = 0, v)$ with the disparity that would have remained if we had set the distribution of the socioeconomic distributions of the black individuals to be the same distribution as that of the white individuals i.e., $E(Y|A = 1, v) - E(Y_{\bar{G}_{0|v}}|A = 1, v)$. We can further decompose the actual racial disparity, $E(Y|A = 1, v) - E(Y|A = 0, v)$, as follows, $E(Y|A = 1, v) - E(Y|A = 0, v) = \{E(Y|A = 1, v) - E(Y_{\bar{G}_{0|v}}|A = 1, v)\} + \{E(Y_{\bar{G}_{0|v}}|A = 1, v) - E(Y|A = 0, v)\}$, where the first component, $\{E(Y|A = 1, v) - E(Y_{\bar{G}_{0|v}}|A = 1, v)\}$, is again the disparity that would have remained if we had set the distribution of the socioeconomic distribution of the black individuals to be the same distribution as that of the white individuals, and the second component, $\{E(Y_{\bar{G}_{0|v}}|A = 1, v) - E(Y|A = 0, v)\}$, is the portion of the disparity that would remain under such an intervention (VanderWeele and Robinson, 2014). These are once again interventional analogues of the natural direct and indirect effects, and the methods described below will be applicable to this setting as well. See VanderWeele and Robinson (2014) for further discussion of this race and health disparities context and corresponding assumptions. Note however, the interventional analogues arguably involve much less problematic comparisons. By randomly fixing the distributions to equal one another (and also decomposing the observed disparity itself, rather than "the effect of race"), we avoid peculiar counterfactuals of the form of what would have happened to an individual had they been of a different race. Thus, in some cases at least, the interventional analogues are not simply a second-best alternative to natural direct and indirect effects, but are themselves arguably the causal effects of interest.

Suppose now that at each time, conditional on the past, the exposure-outcome-, mediator-outcome-, and exposure-mediator-relationships are unconfounded. Formally, analogous to (i)–(iii): for all t, (i†) $Y_{\overline{am}} \perp\!\!\!\perp A(t)|\overline{A}(t-1)$, $\bar{M}(t-1)$, $\bar{L}(t-1)$, V and (ii†) $Y_{\overline{am}} \perp\!\!\!\perp M(t)|\overline{A}(t)$, $\bar{M}(t-1)$, $\bar{L}(t-1)$, V and (iii†) $M_{\bar{a}}(t) \perp\!\!\!\perp A(t)|\bar{A}(t-1)$, $\bar{M}(t-1)$, $\bar{L}(t-1)$, V. It is shown in the Online Supplement that although natural direct and indirect effects are not in general identified in this setting (Avin et al., 2005), the interventional direct and indirect effects, $\{E(Y_{\bar{a}\bar{G}_{\bar{a}|v}}|v) - E(Y_{\bar{a}\bar{G}_{\bar{a}}^{*}|v}|v)\}$and $\{E(Y_{\bar{a}\bar{G}_{\bar{a}}^{*}|v}|v) - E(Y_{\bar{a}^{*}\bar{G}_{\bar{a}}^{*}|v}|v)\}$, are identified and $E[Y_{\bar{a}\bar{G}_{\bar{a}}^{*}|v}|v]$ is given by:

$$E[Y_{\overline{a}\overline{G}_{\overline{a}^*|v}}|v]=\int_{\overline{m}}\int_{\overline{l}(T-1)}E[Y|\overline{a},\overline{m},\overline{l},v]\prod_{t=1}^{T-1}dP\{\overline{l}(t)|\overline{a}(t),\overline{m}(t),\overline{l}(t-1),v)\}$$
$$\times d[\int_{\overline{l}^{\dagger}(T-1)}\prod_{t=1}^{T}P\{m(t)|\overline{a}^*(t),\overline{m}(t-1),\overline{l}^{\dagger}(t-1),v\}dP\{\overline{l}^{\dagger}(t-1)|\overline{a}^*(t-1),\overline{m}(t-1),\overline{l}^{\dagger}(t-2),v)\}].$$

(1)

Author Manuscript

We refer to this expression in (1) as the mediational g-formula. We will denote this quantity by $Q(\bar{a}, \bar{a}^*)$. Our interventional direct and indirect effects are under assumptions (i†)–(iii†) then given by

$$\begin{aligned} E(Y_{\overline{a}\overline{G}_{\overline{a}|v}}|v)-E(Y_{\overline{a}\overline{G}_{\overline{a}^*|v}}|v) &= Q(\overline{a},\overline{a})-Q(\overline{a},\overline{a}^*) \\ E(Y_{\overline{a}\overline{G}_{\overline{a}^*|v}}|v)-E(Y_{\overline{a}^*\overline{G}_{\overline{a}^*|v}}|v) &= Q(\overline{a},\overline{a}^*)-Q(\overline{a}^*,\overline{a}^*) \end{aligned}$$

Note that if $\bar{L}$ is empty as in Figure 4 then the mediational g-formula reduces to

$$Q(\overline{a},\overline{a}^*)=\int_{\overline{m}}E[Y|\overline{a},\overline{m},v]\prod_{t=1}^{T}dP\{m(t)|\overline{a}^*(t),\overline{m}(t-1),v\}.$$

We show in the Online Supplement that if $\bar{L}$ is empty then, under a non-parametric structural equation model, natural direct effects are identified by the mediational g-formula and are equal to $Q(\bar{a}, \bar{a}^*) - Q(\bar{a}^*, \bar{a}^*)$ and natural indirect effects are identified by the mediational g-formula and are equal to $Q(\bar{a}, \bar{a}) - Q(\bar{a}, \bar{a}^*)$.

In other words if $\bar{L}$ is empty then the empirical expressions that suffice to identify the interventional direct and indirect effects under assumptions (i)–(iii) in fact also in this setting identify the natural direct and indirect effects as well by a time-varying analogue of Pearl's "mediation formula" (Pearl, 2012). However, even when $\bar{L}$ is not empty so we cannot identify the natural direct and indirect effects themselves, we still can, under assumptions (i†)–(iii†) identify the interventional direct and indirect effects. Note that we thus have that when the natural direct and indirect effects are identified from the data by the Pearl's mediation formula (cf. Shpitser and VanderWeele, 2011), they will coincide with the interventional direct and indirect effects. The only setting in which the natural and the interventional effects will diverge is when the natural direct and indirect effects are not empirically identified by Pearl's mediation formula (but it will of course not be possible in such settings to empirically compare the natural and interventional effects since the natural effects are then not identified from the data).

Note also that if M were empty then the expression in (1) simply reduces to:

$$\int_{\overline{m}}\int_{\bar{l}(T-1)}E[Y|\overline{a},\bar{l},v]\prod_{t=1}^{T-1}dP\{\bar{l}(t)|\overline{a}(t),\bar{l}(t-1),v\}$$

because, with M empty, $\int_{\bar{l}^\dagger(T-1)}\prod_{t=1}^{T}dP\{\bar{l}^\dagger(t-1)|\overline{a}^*(t-1),\bar{l}^\dagger(t-2),v\}=1$. Thus with M empty, the formula in (1) simply reduces to the regular g-formula of Robins (1986). We see then that, on the one hand, if there is no-time-varying confounding the "mediational g-formula" in (1) reduces to the time-varying analogue of the mediational formula. And if, on the other hand, there is no mediation, then the "mediational g-formula" reduces to the regular g-formula of Robins (1986).

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

We now consider some variations on this approach. First, suppose instead that after the initial baseline covariates V, the subsequent temporal ordering of the variables were $A(t)$, $L(t)$, $M(t)$, as in Figure 5, and that analogous to (i†)–(iii†) we have that: for all t, (i‡) $Y_{\overline{am}} \perp\!\!\!\perp A(t) | \overline{A}(t-1)$, $\bar{M}(t-1)$, $\bar{L}(t-1)$, V and (ii‡) $Y_{\overline{am}} \perp\!\!\!\perp M(t) | \overline{A}(t)$, $\bar{M}(t-1)$, $\bar{L}(t)$, V and (iii‡) $M_{\bar{a}}(t) \perp\!\!\!\perp A(t) | A(t-1), M(t-1), \bar{L}(t-1), V$

Under assumptions (i‡)–(iii‡) we would then have, using a similar derivation:

$$E[Y_{\overline{a}\overline{G}_{\overline{a}*|v}}|v]$$
$$=\int_{\overline{m}}\int_{\overline{l}(T-1)} E[Y|\overline{a},\overline{m},\overline{l},v]\prod_{t=1}^{T-1} dP\{\overline{l}(t)|\overline{a}(t),\overline{m}(t-1),\overline{l}(t-1),v)\}$$
$$\times d[\int_{\overline{l}^{\dagger}(T-1)} \prod_{t=1}^{T} P\{m(t)|\overline{a}^*(t),\overline{m}(t-1),\overline{l}^{\dagger}(t),v\}dP\{\overline{l}^{\dagger}(t)|\overline{a}^*(t),\overline{m}(t-1),\overline{l}^{\dagger}(t-1),v)\}].$$

As another variation instead of considering randomized interventions that fix the mediator $\bar{M}$ for each individual to a value randomly drawn from the distribution in the sub-population with baseline covariates $V = v$ if $\bar{A}$ had been fixed to $\bar{a}^*$, we could instead consider randomizing the mediator $\bar{M}$ for each individual to the value randomly drawn the distribution in the entire population if $\bar{A}$ had been fixed to $\bar{a}^*$. We then let $\bar{G}_{\bar{a}}(t)$ denote a random draw from the distribution of the mediator $\bar{M}(t)$ that would have been observed in the population if exposure $\bar{A}$ had been fixed to $\bar{a}$ and we have the decomposition: $E(Y_{\bar{a}\bar{G}_{\bar{a}}(t)}) - E(Y_{\bar{a}^*\bar{G}_{\bar{a}}^*}) = \{E(Y_{\bar{a}\bar{G}_{\bar{a}}}) - E(Y_{\bar{a}\bar{G}_{\bar{a}}^*})\} + \{E(Y_{\bar{a}\bar{G}_{\bar{a}}^*}) (- E(Y_{\bar{a}^*\bar{G}_{\bar{a}}^*})\}$. Under assumptions (i†)–(iii†) we have:

$$E[Y_{\overline{a}\overline{G}_{\overline{a}*}}]=$$
$$\int_{\overline{m}}\int_{\overline{l}(T-1)} E[Y|\overline{a},\overline{m},\overline{l},v][\prod_{t=1}^{T-1} dP\{\overline{l}(t)|\overline{a}(t),\overline{m}(t),\overline{l}(t-1)v)\}]dP(v)$$
$$\times d[\int_{\overline{l}^{\dagger}(T-1)} \prod_{t=1}^{T} P\{m(t)|\overline{a}^*(t),\overline{m}(t-1),\overline{l}^{\dagger}(t-1),v\}dP\{\overline{l}^{\dagger}(t-1)|\overline{a}^*(t-1),\overline{m}(t-1),\overline{l}^{\dagger}(t-2),v\}dP(v)]$$

and under assumptions assumptions (i‡)–(iii‡) we would then have:

$$E[Y_{\overline{a}\overline{G}_{\overline{a}*}}]$$
$$=\int_{\overline{m}}\int_{\overline{l}(T-1)} E[Y|\overline{a},\overline{m},\overline{l},v][\prod_{t=1}^{T-1} dP\{\overline{l}(t)|\overline{a}(t),\overline{m}(t-1),\overline{l}(t-1)v)\}]dP(v)$$
$$\times d[\int_{\overline{l}^{\dagger}(T-1)} \prod_{t=1}^{T} P\{m(t)|\overline{a}^*(t),\overline{m}(t-1),\overline{l}^{\dagger}(t),v\}dP\{\overline{l}^{\dagger}(t)|\overline{a}^*(t),\overline{m}(t-1),\overline{l}^{\dagger}(t-1),v)\}dP(v)].$$

Note that in all of the above variations, we have fixed the entire mediator $\bar{M}$ to a random draw from the mediator vector under a particular exposure history i.e. from the distribution of $\bar{M}_{\bar{a}}(T)$. While this could be viewed as a single draw from this multivariate distribution, it is the case that prior values of the mediator and time-varying confounders, e.g. $M(t-1)$ and $L(t-1)$, will affect the current mediator $M(t)$ and will also be affected by prior exposure $A(t-1)$. Thus, in practice, one would have to use methods for time-varying exposures to estimate the distribution of $\bar{M}_{\bar{a}}(T)$ and then draw from that distribution. The mediational g-formula expressions above take this dependence into account and in the following section we present an inverse probability weighting estimation using marginal structural models to likewise handle this dependence.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

As yet another alternative, though one we argue is not suitable for mediation analysis, we could have at each time t, fixed the mediator $M(t)$ for that time t, to a random draw from the mediator distribution under a particular exposure history up to that point in time t. Said another way, we could have fixed the mediator at each time t to a random draw from the mediator distribution under a specific exposure distribution marginally, rather than jointly, as in all the variations considered above. If we had proceeded in this manner the identifying expression would have differed. Doing so, however, we argue does not adequately allow for the analysis of pathways. To see this, consider the following example: suppose we were interested in assessing the extent to which the effect of marital status (which may be time-varying) on income is mediated by time-varying health status. Suppose that different individuals with different marital status histories have different health trajectories, and that at least some individuals have consistently poor health over time if and only if in the unmarried state, but that the vast majority are healthy over time in either marital state. Suppose that it is only a long-term poor health trajectory that substantially affects income. If we were to randomize the entire mediator vector jointly to a draw from the health trajectory distribution of those who were unmarried then some of these trajectories randomly drawn would be consistently low and would adversely affect income. Using the approaches described above we would see that some of the effect of marital status on income was mediated by preventing the consistently low health trajectories. However, if we were instead to randomize the mediator marginally at each time point to a random draw of the distribution of the unmarried population, the probability of obtaining a health trajectory that was consistently low over time would be very small (since at each time the majority are in the healthy state and thus to get a consistently low health trajectory would require low probability events at each of the individual time points). Consequently, if we were to randomize the mediator marginally at each time point, far fewer individuals with the mediator randomized marginally at each time according to the unmarried distribution would have a health trajectory which was consistently low at all time points than was actually the case with the actual unmarried population and thus there would be few individuals for whom income was substantially adversely affected by health and we would for the most part miss those pathways by which marital status affects income through consistently low health trajectories. To assess such pathways we need to randomize the mediator jointly at all time points to a random draw from the distribution of those with a particular exposure history, as in the approaches described above.

4. Estimation Using Marginal Structural Models

One possible estimation approach would be to use the identification formula in (1) and fit parametric models for each of $E[Y|\bar{a}, \bar{m}, \bar{l}, v]$, $P\{\bar{l}(t)|\bar{a}(t), \bar{m}(t), \bar{l}(t-1), v)\}$, and $P\{M(t)|\bar{a}(t), \bar{m}(t-1), \bar{l}(t-1), v\}$. This estimation approach is sometimes called a g-computation approach. It is described in the setting of time-varying exposures outside of the context of mediation elsewhere (Robins and Hernan, 2009). We will in fact consider one such approach in the context of MacKinnon's three wave longitudinal mediation model (MacKinnon, 2008) in the following section. However, in general such an approach requires fitting many parametric models and it can sometimes be difficult to specify these models so that they are compatible with one another and compatible with the null hypothesis of no effect; these

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

problems are discussed in the setting of time-varying exposures outside of the context of mediation elsewhere (Robins and Wasserman, 1997; Robins and Hernan, 2009). In this section we will instead develop a more parsimonious approach to estimating the interventional direct and indirect effects using marginal structural models and inverse probability of treatment weighting (Robins et al., 2000). In the context of mediation this will require fitting two marginal structural models.

For estimation, one reasonably straightforward approach entails positing a pair of marginal structural models (MSMs) for $E[Y_{\overline{am}}]$ and $P(\bar{M}_{\bar{a}}^* = \bar{m})$. These models can in turn be used to evaluate direct and indirect effects using the following expression:

$$E[Y_{\overline{a}G_{\overline{a}^*}}]=\int_{\overline{m}}E[Y_{\overline{am}}]dP(\overline{M}_{\overline{a}^*}=\overline{m})$$

Consider a scenario, in which Y is a continuous outcome. We assume the following simple marginal structural linear regression model for the outcome:

$$E[Y_{\overline{am}}]=\theta_0+\theta_1\text{cum}\,(\overline{a})+\theta_2\text{cum}\,(\overline{m}) \quad (2)$$

where $\theta_0 = \{\theta_0, \theta_1, \theta_2\}$, and cum $(\bar{a}) = \Sigma_{t \le T}\, a(t)$ and cum $(\bar{m}) = \Sigma_{t \le T}\, m(t)$ are the cumulative totals of $\bar{A}$ and $\bar{M}$ respectively. This MSM assumes that the joint effects of $\bar{M}$ and $\bar{A}$ is cumulative, with a single parameter θ_2 encoding the effect of the M process through cum $(\bar{m}) = \Sigma_{t \le T}\, m(t)$ and θ_1 encoding the effect of the A process through cum $(\bar{a}) = \Sigma_{t \le T}\, a(t)$. For continuous M or A, the model essentially states that the joint effects of $\bar{M}$ and $\bar{A}$ on Y operate strictly through their respective historical average levels, and that these two processes do not interact on the additive scale. A more flexible model, such as is given in detail in the Online Supplement, could also be specified to account for possibly more complex dose-response relationships between $(\bar{a},\bar{m})$ and $Y_{\overline{am}}$ and interactions between $\bar{m}$ and $\bar{a}$ could also be specified. Together with Model (2), suppose that the following MSM model holds for the mediator process

$$\text{g}^{-1}\left\{E(M_{\overline{a}}(t))\right\}=\beta_0\,(t)+\beta_1\,(t)\,\text{avg}\,(\overline{a}\,(t)) \quad (3)$$

where $g^{-1}(\cdot)$ is a link function, and $\beta = \{\beta_0(t), \beta_1(t) : t\}$ are potentially allowed to vary with time, and avg $(\bar{a}(t)) = \Sigma_{j \le t}\, a(j)/t$. It is easy to verify that models (2) and (3) gives the following

$$\begin{aligned} E[Y_{\overline{a}G_{\overline{a}^*}}]&=\int_{\overline{m}}E[Y_{\overline{am}}]dP(\overline{M}_{\overline{a}^*}=\overline{m}) \\ &=\int_{\overline{m}}\{\theta_0+\theta_1\text{cum}\,(\overline{a})+\theta_2\text{cum}\,(\overline{m})\}dP(\overline{M}_{\overline{a}^*}=\overline{m}) \\ &=\theta_0+\theta_1\text{cum}(\overline{a})+\theta_2\left(\sum_{t\le T}g\left(\beta_0\,(t)+\beta_1\,(t)\,\text{avg}\,(\overline{a}^*(t))\right)\right) \end{aligned}$$

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

In the special case where $M(t)$ is continuous, so that g^{-1} may be taken to be the identity link, one obtains the following expression for the interventional direct effect:

$$E(Y_{\overline{a}\overline{G}_{\overline{a}^*}}) - E(Y_{\overline{a}^*\overline{G}_{\overline{a}^*}}) = \theta_1\{\mathrm{cum}\,(\overline{a}) - \mathrm{cum}\,(\overline{a}^*)\},$$

and for the indirect effect:

$$E(Y_{\overline{a}\overline{G}_{\overline{a}}}) - E(Y_{\overline{a}\overline{G}_{\overline{a}^*}}) = \sum_{t\le T}\theta_2\beta_1\,(t)\,\{\mathrm{avg}\,(\overline{a}(t)) - \mathrm{avg}\,(\overline{a}^*\,(t))\}$$

The expressions above simplify further when $\beta_1\,(t) = \beta_1$ is assumed to be constant, and $a^*(t) = 0$ and $a\,(t) = 1$ for all t, giving $\beta_1\theta_2 T$ for the indirect effect and $\theta_1 T$ for the direct effect.

For estimation, standard inverse probability weighting may be used to estimate (β,θ), however, construction of the weights varies somewhat with the underlying identifying assumptions. Specifically, suppose that assumptions (i†)–(iii†) hold, then a consistent estimate of θ under model (2) can be obtained by weighted least squares regression of Y on (cum $\bar{M}$, cum) ($\bar{A}$) with estimated weight equal to

$$\prod_{t=1}^{T-1}\hat{P}\{A(t),M(t)|\overline{A}(t-1),\overline{M}(t-1),\overline{L}(t-1),V\}^{-1}$$

where

$$\hat{P}\{A(t),M(t)|\overline{A}(t-1),\overline{M}(t-1),\overline{L}(t-1),V\} = \hat{P}\{M(t)|\overline{A}(t),\overline{M}(t-1),\overline{L}(t-1),V\}\hat{P}\{A(t)|\overline{A}(t-1),\overline{M}(t-1),\overline{L}(t-1),V\}$$

is a maximum likelihood estimate of $P\{A(t), M(t)|\bar{A}(t-1), \bar{M}(t-1), \bar{L}(t-1), V)\}$ under a standard parametric model. The parameter $\beta_1(t)$ of the second MSM(3) is likewise estimated via inverse probability weighted regression with weight

$$\prod_{s=1}^{t}\hat{P}\{A(t)|\overline{A}(t-1),\overline{M}(t-1),\overline{L}(t-1),V\}^{-1}.$$

In the Online Supplement we give SAS code for fitting these marginal structural models using inverse probability weights, in the context of the example in Section 6. A similar

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

approach could also be developed for marginal structural models conditional on baseline covariates $V = v$.

The expressions given above for the interventional direct and indirect effects pertain to the simple linear marginal structural models in (2) and (3) which involve either the average or cumulative average of the exposure and/or mediator histories. It would not be difficult to modify these models, by adding higher-order terms, or incorporating lags, or separate effects on the outcome for each time point; one could derive alternative expressions for the interventional direct and indirect effects. In the Online Supplement we show, for example, how this can be done for a marginal structural model for the outcome that involves an interaction between the cumulative total of the exposure and the cumulative total of the mediator in their effects on the outcome. Below we also discuss the case of a binary outcome and log-linear or logistic marginal structural models. Various changes in the models will result in alternative empirical expressions, but the structure of the proof would follow a very similar development.

In settings in which certain values of the exposure or mediator histories are unlikely the exposure or mediator weights given above can become very large. To address this problem, it has become standard practice in fitting marginal structural models to truncate the weight, often at the 1st and 99th percentile of the weight distribution to help ensure more stable estimators (Cole and Hernán, 2008). To improve stability it may also be preferable to use so-called stabilized weights (Robins et al., 2000) replacing the weight for time t in model (2) by $\hat{P}\{M(t)|\bar{A}(t), \bar{M}(t-1)\}\ \hat{P}\{A(t)|\bar{A}(t-1), \bar{M}(t-1)\}/[\hat{P}\{M(t)|\bar{A}(t), \bar{M}(t-1), \bar{L}(t-1), V\}\ \hat{P}\{A(t)|\bar{A}(t-1), \bar{M}(t-1), \bar{L}(t-1), V\}]$ and the weight for time t in model (3) by $\hat{P}\{A(t)|\bar{A}(t-1)\}/\hat{P}\{A(t)|\bar{A}(t-1), \bar{M}(t-1), \bar{L}(t-1), V)\}$. This is especially important if the exposure or mediator are continuous (Robins et al., 2000).

If at each time point the mediator variable $M(t)$ in fact consists of a vector of mediators, it is still possible to proceed with the estimation in a similar manner, but one would require a separate weight for each component in the vector $M(t)$. Provided the ordering of the exposure, mediator and time-varying confounders was still $A(t), M(t), L(t)$ for the entire vector $M(t)$, one could choose any ordering of the components of $M(t)$ and the weight for a specific component for $M(t)$ would simply be conditional on $\bar{A}(t), \bar{M}(t-1), \bar{L}(t-1), V$ and the prior components in $M(t)$. The marginal structural model in (2) could be modified to include a separate cumulative average term for each component in $M(t)$.

It is also straightforward to modify the weights for estimation under the alternative identifying assumptions $(i^{\ddagger})$–$(iii^{\ddagger})$ if the ordering of the variables is $A(t)$, $L(t)$, $M(t)$, as in Figure 5. Specifically, estimation of θ under model (2) would instead use the following set of weights

$$\left[\prod_{t=1}^{T-1} \hat{P}\{M(t)|\overline{A}(t), \overline{M}(t-1), \overline{L}(t), V\}\hat{P}\{A(t)|\overline{A}(t-1), \overline{M}(t-1), \overline{L}(t-1), V)\}\right]^{-1}$$

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

while estimation of $\beta(t)$ in the second MSM (3) would use the same set of weights as above. In any of these cases, inference can proceed using bootstrapping, to appropriately account for variation due to estimation of the weights.

Suppose now that the outcome is binary and that we replace the linear marginal structural model for the outcome with a log-binomial marginal structural model:

$$\log\{E[Y_{\overline{am}}]\}=\theta_0+\theta_1\text{cum}\,(\overline{a})+\theta_2\text{cum}\,(\overline{m}).$$

Suppose further that $M_{\bar{a}}(t)$ are multivariate normally distributed with mean $\beta_0(t)+\ \beta_1\ (t)$ avg ($\bar{a}(t)$). We show in the Online Supplement that the interventional direct effect on a risk ratio scale is given by:

$$\log\{E(Y_{\overline{a}\overline{G}_{\overline{a}*}})/E(Y_{\overline{a}*\overline{G}_{\overline{a}*}})\}=\theta_1\,\{\text{cum}\,(\overline{a})-\text{cum}\,(\overline{a}^*)\}$$

and the interventional indirect effect on a risk ratio scale is once again given by:

$$\log\{E(Y_{\overline{a}\overline{G}_{\overline{a}}})/E(Y_{\overline{a}\overline{G}_{\overline{a}*}})\}=\theta_2\sum_{t\leq T}\beta_1\,(t)\,\{\text{avg}\,(\overline{a}\,(t))-\text{avg}\,(\overline{a}^*(t))\}.$$

As before, the expressions above further simplify when $\beta_0\ (t) = \beta_0$ and $\beta_1\ (t) = \beta_1$ are assumed to be constant and $a^*(t) = 0$ and $a\ (t) = 1$ for all t, giving $\theta_1 T$ for the direct effect and $\beta_1\theta_2 T$ for the indirect effect. If the binary outcome is rare, the same formulas hold approximately if a logistic marginal structural model is fit to the data. Similar expressions would also pertain to marginal structual Poisson or Negative Binomial models with log link. In the Online Supplement we also show how these formulas extend to the setting in which the marginal structural model for the outcome includes an interaction terms between the cumulative exposure total and the cumulative mediator total in their effects on the outcome.

5. A Counterfactual Analysis of MacKinnon's Three-Wave Mediation Model

MacKinnon (2008) considered a three-wave mediation model with linear structural equations as depicted in Figure 6. We relabel indices somewhat to correspond to the notation of this paper, and also add a set of baseline covariates C, which is allowed to have effects on all other variables on the diagram, but otherwise the model considered here is MacKinnon's model (MacKinnon, 2008, pp. 204–206, Autoregressive Model III). We let $A(0)$, $M(0)$ and $Y(0)$ denote baseline values of A, M and Y that could be included in the baseline covariates C but are given here to make clearer the relation with MacKinnon (2008). Consider then the following regression models:

$$\begin{aligned}
E[M(1)|m(0),y(0),\overline{a}(1),c] &= \beta_{10}+\beta_{11}a(0)+\beta_{12}a(1)+\beta_{13}m(0)+\beta_{14}y(0)+\beta_{15}^{'}c\\
E[M(2)|\overline{m}(1),\overline{y}(1),\overline{a}(2),c] &= \beta_{20}+\beta_{21}a(1)+\beta_{22}a(2)+\beta_{23}m(1)+\beta_{24}y(1)+\beta_{25}^{'}c\\
E[Y(1)|\overline{m}(1),y(0),\overline{a}(1),c] &= \theta_{10}+\theta_{11}a(0)+\theta_{12}a(1)+\theta_{13}m(0)+\theta_{14}m(1)+\theta_{15}y(0)+\theta_{16}^{'}c\\
E[Y(2)|\overline{m}(2),\overline{y}(1),\overline{a}(2),c] &= \theta_{20}+\theta_{21}a(1)+\theta_{22}a(2)+\theta_{23}m(1)+\theta_{24}m(2)+\theta_{25}y(1)+\theta_{26}^{'}c.
\end{aligned}$$

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Note that in these models, the mediator and the outcome depend only on the two most recent past exposure values. The mediator model depends only on the most recent past mediator value and the most recent past outcome value. The outcome model depends on the two most recent mediator values and the most recent outcome value.

We show that under assumptions (i†)–(iii†) with $V = (C, A(0), M(0), Y(0))$ and $L(1) = Y(1)$, with two intervention periods, $A(1)$ and $A(2)$, the interventional direct and indirect effects are given by:

$$\begin{aligned} E[Y_{\overline{a}G_{\overline{a}^*|v}}(2)|v]-E[Y_{\overline{a}^*G_{\overline{a}^*|v}}(2)|v] &= (\theta_{21}+\theta_{12}\theta_{25})[a(1)-a^*(1)]+\theta_{22}[a(2)-a^*(2)] \\ E[Y_{\overline{a}G_{\overline{a}|v}}(2)|v]-E[Y_{\overline{a}G_{\overline{a}^*|v}}(2)|v] &= \{\theta_{23}\beta_{12}+\theta_{25}\theta_{14}\beta_{12}+\beta_{21}\theta_{24}+\beta_{24}\theta_{12}\theta_{24}\}[a(1)-a^*(1)]+\beta_{22}\theta_{24}[a(2)-a^*(2)]. \end{aligned}$$

The first expression is the interventional direct effect with time-varying exposure and mediator and the second expression is the interventional indirect effect with time-varying exposure and mediator. A proof of this is given in the Online Supplement.

There is arguably a two-fold advantage of using data like that in Figure 5 and using a modeling approach like that described above, over simply applying the standard methods for mediation to one point in time e.g. using the variables $A(1)$, $M(1)$, $Y(1)$. First, by having multiple waves of data, we can control for baseline levels of the exposure, mediator and outcome, i.e. for $A(0)$, $M(0)$, $Y(0)$. This is potentially important because such baseline values of the exposure, mediator and outcome may serve as the most important confounders for the effects of subsequent values of exposure and mediator on the outcome. By including such baseline values of the exposure, mediator and outcome, in our covariate set, our confounding assumptions required for a causal interpretation of our estimates are rendered much more plausible. Second, by using multiple waves of subsequent exposure and mediator and outcome data (i.e. by using $A(1)$, $M(1)$, $Y(1)$, $A(2)$, $M(2)$, $Y(2)$ rather than just $A(1)$, $M(1)$, $Y(1)$) we may be able to more fully capture the dynamics of mediation over time. For example we can pick up, in our indirect effect estimates, mediated effects of $A(1)$ through $M(1)$ to $Y(2)$ directly and also those from $A(1)$ through $M(1)$ to $Y(1)$ to $Y(2)$ or from $A(1)$ to $M(2)$ to $Y(2)$, etc.

Here we have given a counterfactual analysis of one specific mediational model with three waves of data on the exposure, mediator and outcome (MacKinnon, 2008). A similar approach could in principle be used for other complex longitudinal models often used in the social sciences to provide counterfactual-based interpretations of direct and indirect effect estimates.

6. Illustration

We apply the marginal structural model approach to time-varying mediation to an example from psychology. Loneliness has been shown to be associated with subjective well-being, both cross-sectionally and longitudinally (Cacioppo et al., 2008). Likewise, loneliness predicts subsequent depression longitudinally, even after control is made for initial depression levels (Heikkinen and Kauppinen, 2004; Cacioppo et al., 2009). These

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

associations persist even after control is made for objective measures of social support (Cacioppo et al., 2009). Prior analyses considered these relationships using repeated measures marginal structural models (VanderWeele et al., 2011, 2012) and again found fairly strong evidence for effects of loneliness on both depression and subjective well-being. Although not unrelated, there is good empirical evidence that depression and subjective well-being are conceptually distinct. For example, recent empirical work (Gargiulo and Stokes, 2009) indicates that subjective well-being has relatively poor predictive ability for diagnosing clinical depression; moreover, many clinically depressed individuals having considerably higher levels of subjective well-being than might be anticipated (Cummins et al., 2007).

The relationships between loneliness, depression and subjective well-being are complicated by reciprocal relationships and feedback present among these various constructs. Loneliness may bring with it heightened depression potentially leading to social withdrawal and isolation and yet greater levels of loneliness. It may also be the case that individuals with a perception of high levels of subjective well-being may act more positively toward others prompting a positive response, closer social ties, less loneliness and subsequently a yet greater sense of subjective well-being (Hawkley et al., 2007). Depression itself of course likely contributes adversely to subjective well-being. In addition to these reciprocal relationships, the analysis of the effects of loneliness, depression and subjective well-being is further complicated by the fact that these effects may depend not simply on the level of a construct at a particular point in time but on the entire history of the psychological construct of interest.

A question that therefore arises - and one that can only be properly addressed with longitudinal data - is the extent to which the effects of loneliness on subjective well-being are mediated by depression and the extent to which it is through other pathways e.g. loneliness directly leading to a negative sense of well-being. Here we use five waves of data from the Chicago Health, Aging, and Social Relations Study (CHASRS), a population-based study of 229 individuals living in Cook County, Illinois, aged 50 to 67. Data are available at baseline and then subsequently collected once per year for four additional years. The data is structured in correspondence with Figure 5 and the notation above. The first wave was taken as baseline covariates V (including baseline measurements of loneliness, depression and subjective well-being); the next three waves were used for the time-varying exposure (loneliness) $\bar{A}$, the time-varying mediator (depressive symptoms) $\bar{M}$ and the time-varying confounders $\bar{L}$; the final wave was used only for the subjective well-being outcome Y. Subjective well-being in waves 2, 3 and 4 were taken as time-varying confounders. Loneliness was assessed using the UCLA-R, a 20-item questionnaire measuring general perception of social connection or isolation with scores ranging from 20 to 80 with higher scores indicating higher levels of loneliness. Depression was assessed using the Center for Epidemiologic Studies Depression Scale (CES-D), a 20-item measure with each scored 0–3, which after removing the one loneliness question gives a measure (CES-D-ML) with scores ranging from 0 to 57. Subjective well-being was assessed used the 5-item Satisfaction with Life Scale (Diener et al., 1985) in which respondents rate each item on a scale from 1 to 7 with scores ranging from 5 to 35. All measures were standardized.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

We apply the marginal structural model approach to the mediational g-formula and fit a linear marginal structural model for the effect of loneliness and depression on subjective well-being:

$$E[Y_{\overline{am}}]=\theta_0+\theta_1\text{cum}\,(\overline{a})+\theta_2\text{cum}\,(\overline{m}) \quad (2)$$

and a linear marginal structural model for the effect of loneliness on depression:

$$E[M_{\overline{a}}(t)]=\beta_0+\beta_1\text{avg}\,(\overline{a}\,(t)) \quad (3)$$

Estimation is carried out using inverse probability of treatment weights as described above with baseline adjustment for age, gender, ethnicity, marital status, education, income and baseline depressive symptoms, social support, loneliness, subjective well-being, psychiatric conditions and psychiatric medications as baseline confounders V, and with control for time-varying subjective well-being, social support, psychiatric conditions and psychiatric medications as time-varying confounders $L(t)$. Standard errors and confidence intervals for the direct and indirect effects were obtained by bootstrapping. Decisions about confounding control were made on substantive grounds and include most known confounders of the loneliness-depression, loneliness-subjective-well-being, and depression-subjective-well being relationships (Cacioppo et al., 2009, VanderWeele et al., 2011, 2012). The estimates obtained by fitting model (3) were $\theta_0 = 1.85(s.e. = 0.395)$; $\theta_1 = -0.091(s.e. = 0.039)$; $\theta_2 = -0.092(s.e. = 0.044)$. The estimates obtained by fitting model (2) were $\beta_0 = 0.14(s.e. = 0.35)$ and $\beta_1 = 0.36(s.e. = 0.061)$. As can be seen from the standard errors, all coefficients in both models were statistically significantly different from zero except the intercept β_{m0} in the mediator model. If we consider the direct and indirect effects for a one standard deviation change in loneliness across all time-points so that $a(t) = a^*(t)+1$ for all three exposure periods then we have from Section 4 that the interventional direct effect is given by: $\{E(Y_{\bar{a}G_{\bar{a}}^*}|v) - E(Y_{\bar{a}^*G_{\bar{a}}^*})\} = \beta_{ya}\{\text{cum}\,(\bar{a}) - \text{cum}\,(\bar{a}^*)\} = 3\theta_1 = -0.27$ (95%CI: −0.65, −0.07) and the interventional indirect effect is given by $\{E(Y_{\bar{a}G_{\bar{a}}}) - E(Y_{\bar{a}G_{\bar{a}}^*})\} = \beta_1\theta_2 T = 3\beta_1\theta_2 = -0.10$ (95%CI: −0.20, 0.00). We can sum the direct and indirect effects to obtain an overall effect of loneliness on subjective well-being for a one standard deviation change in loneliness across all time and this gives −0.37 (95%CI: −0.70, −0.22). The overall effect is the effect on subjective well-being of setting loneliness to the high trajectory and depression to a random draw from what it would have been on the high loneliness trajectory versus setting loneliness to the lower trajectory and depression to a random draw from what it would have been on the lower loneliness trajectory. The interventional direct effect is the effect on subjective well-being of setting loneliness to the high trajectory and depression to a random draw from what it would have been on the lower loneliness trajectory versus setting loneliness to the lower trajectory and depression to a random draw from what it would have been on the lower loneliness trajectory. The interventional indirect effect is the effect on subjective well-being of setting loneliness to the high trajectory and depression to a random draw from what it would have been on the high loneliness trajectory versus setting loneliness to the high trajectory and depression to a random draw from what it would have been on the lower loneliness trajectory. Thus of this overall effect of loneliness on subjective well-being,

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

−0.37 (95%CI: −0.70, −0.22), about one quarter of this, −0.10 (95%CI: −0.20, 0.00), seems to be mediated by affecting depressive symptoms. The depressive symptoms brought upon by loneliness thus does seem to be an important mechanism for the effect of loneliness on subjective well-being, but there appear to be other important pathways as well, independent of depression, related to subjective assessment of well-being or that quality of life is poorer when one is lonely.

7. Discussion

In this paper we have considered methods for time-varying exposures and mediators. One of the challenges here was the presence of mediator-outcome confounders affected by the exposure. This in general leads to lack of non-parametric identification of longitudinal analogues of natural direct and indirect effects. However we were able to show in this paper that it is still possible to identify interventional analogues of natural direct and indirect effects and these can in fact be used for effect decomposition. The empirical expression used for identification we referred to as the mediational g-formula. When identified, these interventional direct and indirect effects do reduce to the natural direct and indirect effects where there is no mediator-outcome confounder affected by exposure (e.g. when there are no time-varying confounders) but the interventional analogues can be estimated in a broader range of settings even when natural direct and indirect effects are not identified with the data. The methods in this paper thereby extend those in prior literature to settings with longitudinal data and exposures and mediators that vary over time. Such rich longitudinal data can potentially increase power in the analysis of direct and mediated effects and help better ensure that questions of temporality in thinking about causal effects are clearer. We have shown how the developments in this paper can be implemented by fitting two marginal structural models and how this constitutes a fairly general and flexible approach as discussed further in the Online Supplement. We have also shown how our mediational g-formula can be used to formalize and clarify the interpretation of effect estimates that come from longitudinal linear structural equation models common in the social sciences. But these two contexts are only two settings in which the approach we developed here using the mediational g-formula can be implemented. We believe that the mediational g-formula will lie at the foundation of the development of many subsequent methods for mediational analysis with longitudinal data.

Future research on this topic could develop a parametric approach to the mediational g-formula to try to increase power or could develop doubly robust estimators for the interventional direct and indirect effects identified by the mediational g-formula to both improve power and robustness. Recent work has also shown that in the context of both mediation and interaction, a total effect can be decomposed into four components: that due to just mediation, that due to just interaction, that due to both, and that due to neither (VanderWeele, 2014). In the Appendix and Online Supplement we show how this four-way decomposition can also be extended, using very similar ideas to those developed above with marginal structural models, to the context of four-way decompositions for time-varying exposures and mediators. The approach we have developed here provides a very general framework for mediation analysis with longitudinal data.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

The research was supported by the National Institutes of Health, U.S.A. grants ES017876 and AI104459.

References

Albert JM, Nelson S. Generalized causal mediation analysis. Biometrics. 2011; 67:1028–1038. [PubMed: 21306353]

Avin C, Shpitser I, Pearl J. Identifiability of path-specific effects. Proceedings of the International Joint Conferences on Artificial Intelligence. 2005:357–363.

Cacioppo, JT., Hawkley, LC., Kalil, A., et al. Happiness and the invisible threads of social connection: The Chicago Health, Aging, and Social Relations Study. In: Larsen, MER., editor. The Science of Well-being. New York: Guilford; 2008. p. 195-219.

Chen YH, Mukherjee B, Ferguson KK, Meeker JD, VanderWeele TJ. Mediation formula for binary outcome and time-varying exposure and mediator accounting for possible exposure-mediator interaction. American Journal of Epidemiology. in press.

Cole S, Hernán MA. Constructing inverse probability weights for marginal structural models. American Journal of Epidemiology. 2008; 168:656–664. [PubMed: 18682488]

Cummins, RA., Lau, ALD., Davern, M. Homeostatic mechanisms and subjective well-being. New York: Springer; 2007.

Daniel RM, De Stavola BL, Cousens SN, Vansteelandt S. Causal mediation analysis with multiple mediators. Biometrics. 2015; 71:1–14. [PubMed: 25351114]

Didelez V, Dawid AP, Geneletti S. Direct and indirect effects of sequential treatments. Proceedings of the Twenty-Second Conference on Uncertainty in Artificial Intelligence. 2006

Diener E, Emmons RA, Larsen RJ, et al. The Satisfaction With Life Scale. J Pers Assess. 1985; 49(1): 71–75. [PubMed: 16367493]

Gargiulo RA, Stokes MA. Subjective well-being as an indicator for clinical depression. Soc Indic Res. 2009; 92(3):517–527.

Geneletti S. Identifying direct and indirect effects in a non-counterfactual framework. Journal of the Royal Statistical Soceity, Series B. 2007; 69:199–216.

Goetgeluk S, Vansteelandt S, Goetghebeur E. Estimation of controlled direct effects. Journal of the Royal Statistical Society, Series B. 2008; 70:1049–1066.

Hawkley, LC., Preacher, K., Cacioppo, JT. Multilevel modeling of social interactions and mood in lonely and socially connected individuals. In: Dulmen, AOMv, editor. Oxford Handbook of Methods in Positive Psychology. New York: Oxford University Press; 2007. p. 559-575.

Heikkinen RL, Kauppinen M. Depressive symptoms in late life: a 10-year follow-up. Arch Gerontol Geriatr. 2004; 38(3):239–250. [PubMed: 15066310]

Imai K, Keele L, Tingley D. A general approach to causal mediation analysis. Psychological Methods. 2010; 15:309–334. [PubMed: 20954780]

Imai K, Yamamoto T. Identification and sensitivity analysis for multiple causal mechanisms: revisiting evidence from framing experiments. Political Analysis. 2013; 21:141–171.

Lange T, Hansen JV. Direct and indirect effects in a survival context. Epidemiology. 2011; 22:575–581. [PubMed: 21552129]

MacKinnon, DP. Introduction to Statistical Mediation Analysis. New York: Erlbaum; 2008.

Martinussen T, Vansteelandt S, Gerster M, von Bornemann Hjelmborg J. Estimation of direct effects for survival data by using the Aalen additive hazards model. Journal of the Royal Statistical Society, Series B. 2011; 73:773–788.

Pearl J. Causal diagrams for empirical research (with discussion). Biometrika. 1995; 82:669–710.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Pearl, J. Proceedings of the Seventeenth Conference on Uncertainty and Artificial Intelligence. San Francisco: Morgan Kaufmann; 2001. Direct and indirect effects; p. 411-420.

Pearl, J. Causality: Models, Reasoning, and Inference. 2. Cambridge: Cambridge University Press; 2009.

Pearl J. The causal mediation formula - a guide to the assessment of pathways and mechanisms. Prevention Science. 2012; 13:426–436. [PubMed: 22419385]

Robins, JM. Semantics of causal DAG models and the identification of direct and indirect effects. In: Green, P.Hjort, NL., Richardson, S., editors. Highly Structured Stochastic Systems. Oxford University Press; New York: 2003. p. 70-81.

Robins JM. A new approach to causal inference in mortality studies with sustained exposure period - application to control of the healthy worker survivor effect. Mathematical Modelling. 1986; 7:1393–1512.

Robins JM, Greenland S. Identifiability and exchangeability for direct and indirect effects. Epidemiology. 1992; 3:143–155. [PubMed: 1576220]

Robins, JM., Hernán, MA. Estimation of the causal effects of time-varying exposures. In: Fitzmaurice, G.Davidian, M.Verbeke, G., Molenberghs, G., editors. Longitudinal Data Analysis. New York: Chapman and Hall/CRC Press; 2009.

Robins, JM., Richardson, TS. Alternative graphical causal models and the identification of direct effects. In: Shrout, P., editor. Causality and Psychopathology: Finding the Determinants of Disorders and Their Cures. Oxford University Press; 2010.

Robins, JM., Wasserman, L. Estimation of effects of sequential treatments by reparameterizing directed acyclic graphs. In: Geiger, D., Shenoy, P., editors. Proceedings of the Thirteenth Conference on Uncertainty in Artificial Intelligence. Morgan Kaufmann; San Francisco: 1997. p. 409-420.

Shpitser I, Pearl J. Complete identification methods for the causal hierarchy. Journal of Machine Learning Research. 2008; 9:1941–1979.

Shpitser I, VanderWeele TJ. A complete graphical criterion for the adjustment formula in mediation analysis. International Journal of Biostatistics. 2011; 7:1–24. Article 16.

Shpitser I. Counterfactual graphical models for longitudinal mediation analysis with unobserved confounding. Cognitive Science. 2013; 37:1011–1035. [PubMed: 23899340]

Tchetgen Tchetgen EJ, Shpitser I. Semiparametric theory for causal mediation analysis: efficiency bounds, multiple robustness, and sensitivity analysis. Annals of Statistics. 2012; 40(3):1816–1845. [PubMed: 26770002]

Tchetgen Tchetgen EJ, Shpitser I. Estimation of a semiparametric natural direct effect model incorporating baseline covariates. Biometrika. 2014; 101:849–864. [PubMed: 25892739]

Tchetgen Tchetgen EJ, VanderWeele TJ. On identification of natural direct effects when a confounder of the mediator is directly affected by exposure. Epidemiology. 2014; 25:282–291. [PubMed: 24487211]

van der Laan MJ, Petersen ML. Direct effect models. International Journal of Biostatistics. 2008; 4 Article 23.

VanderWeele TJ. Marginal structural models for the estimation of direct and indirect effects. Epidemiology. 2009; 20:18–26. [PubMed: 19234398]

VanderWeele TJ. A unification of mediation and interaction: a four-way decomposition. Epidemiology. 2014; 25:749–761. [PubMed: 25000145]

VanderWeele, TJ. Explanation in Causal Inference: Methods for Mediation and Interaction. New York: Oxford University Press; 2015.

VanderWeele TJ, Hawkley LC, Thisted RA, Cacioppo JT. A marginal structural model for loneliness: implications for intervention trials and clinical practice. Journal of Consulting and Clinical Psychology. 2011; 79:225–235. [PubMed: 21443322]

VanderWeele TJ, Hawkley LC, Cacioppo JT. On the reciprocal effects between loneliness and subjective well-being. American Journal of Epidemiology. 2012; 176:777–784. [PubMed: 23077285]

VanderWeele TJ, Robinson W. On the causal interpretation of race in regressions adjusting for confounding and mediating variables. Epidemiology. 2014; 25:473–484. [PubMed: 24887159]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

VanderWeele TJ, Vansteelandt S. Conceptual issues concerning mediation, interventions and composition. Statistics and Its Interface. 2009; 2(4):457–468.

VanderWeele TJ, Vansteelandt S, Robins JM. Effect decomposition in the presence of an exposure-induced mediator-outcome confounder. Epidemiology. 2014; 25:300–306. [PubMed: 24487213]

Vansteelandt S, Bekaert M, Lange T. Imputation strategies for the estimation of natural direct and indirect effects. Epidemiologic Methods. 2012; 1:131–158.

Appendix. Four-Way Decomposition for Mediation and Interaction with Time-Varying Exposures and Mediators

VanderWeele (2014) showed that with a single binary exposure and mediator it was possible to decompose a total effect, $Y_1 - Y_0$, into four components: that due to just mediation, that due to just interaction, that due to both, and that due to neither. The decomposition was given as follows:

$$Y_1 - Y_0 = (Y_{10} - Y_{00}) + (Y_{11} - Y_{10} - Y_{01} + Y_{00})(M_0) + (Y_{11} - Y_{10} - Y_{01} + Y_{00})(M_1 - M_0) + (Y_{01} - Y_{00})(M_1 - M_0).$$

where the first component, $(Y_{10} - Y_{00})$, is the controlled direct effect, due to neither mediation nor interaction; the second component, $(Y_{11} - Y_{10} - Y_{01} + Y_{00})(M_0)$, was referred to as the reference interaction due to just interaction, not mediation; the third component, $(Y_{11} - Y_{10} - Y_{01} + Y_{00})(M_1 - M_0)$, was referred to as the mediated interaction, due to both mediation and interaction, and the fourth component $(Y_{01} - Y_{00})(M_1 - M_0) = Y_{0M_1} - Y_{0M_0}$ is the pure indirect effect, due to just mediation. See VanderWeele (2014) for further discussion and interpretation. These four components are identified under the same assumptions (i)–(iv) in the text for identifying natural direct and indirect effects. The controlled direct effect and the reference interaction sum to the natural direct effect, $Y_{1M_0} - Y_{0M_0}$. The mediated interaction and the pure indirect effect sum to the natural indirect effect, $Y_{1M_1} - Y_{1M_0}$ (VanderWeele, 2014).

The decomposition can also be generalized to an arbitrary exposure and mediator, with the controlled direct effect fixing the mediator to an arbitrary level m^*, not necessarily 0. We then have $Y_a - Y_{a^*}$

$$= (Y_{am^*} - Y_{a^*m^*}) + \sum_m (Y_{am} - Y_{a^*m} - Y_{am^*} + Y_{a^*m^*}) 1(M_{a^*} = m) + \sum_m (Y_{am} - Y_{a^*m}) \{1(M_a = m) - 1(M_{a^*} = m)\} + (Y_{a^*M_a} - Y_{a^*M_{a^*}}).$$

Here we show a similar four-way decomposition pertains to the interventional overall effect, $Y_{\bar{a}\bar{G}_{\bar{a}}} - Y_{\bar{a}^*\bar{G}_{\bar{a}^*}}$, given in the text. Specifically, we have that for any $\bar{a}$, $\bar{a}^*$, and $\bar{m}^*$:

$$\begin{aligned}Y_{\overline{a}\overline{G}_{\overline{a}}}-Y_{\overline{a}^*\overline{G}_{\overline{a}^*}}&=(Y_{\overline{a}\overline{G}_{\overline{a}}}-Y_{\overline{a}^*\overline{G}_{\overline{a}}})+(Y_{\overline{a}^*\overline{G}_{\overline{a}}}-Y_{\overline{a}^*\overline{G}_{\overline{a}^*}})\\&=(Y_{\overline{a}\overline{G}_{\overline{a}^*}}-Y_{\overline{a}^*\overline{G}_{\overline{a}^*}})+(Y_{\overline{a}^*\overline{G}_{\overline{a}}}-Y_{\overline{a}^*\overline{G}_{\overline{a}^*}})+(Y_{\overline{a}\overline{G}_{\overline{a}}}-Y_{\overline{a}^*\overline{G}_{\overline{a}}}-Y_{\overline{a}\overline{G}_{\overline{a}^*}}+Y_{\overline{a}^*\overline{G}_{\overline{a}^*}})\\&=(Y_{\overline{am}^*}-Y_{\overline{a}^*\overline{m}^*})+\{(Y_{\overline{a}\overline{G}_{\overline{a}^*}}-Y_{\overline{a}^*\overline{G}_{\overline{a}^*}})-(Y_{\overline{am}^*}-Y_{\overline{a}^*\overline{m}^*})\}+(Y_{\overline{a}\overline{G}_{\overline{a}}}-Y_{\overline{a}^*\overline{G}_{\overline{a}}}-Y_{\overline{a}\overline{G}_{\overline{a}^*}}+Y_{\overline{a}^*\overline{G}_{\overline{a}^*}})+(Y_{\overline{a}^*\overline{G}_{\overline{a}}}-Y_{\overline{a}^*\overline{G}_{\overline{a}^*}}).\end{aligned}$$

(A1)

This can be further rewritten as:

$$\begin{aligned}&(Y_{\overline{am}^*}-Y_{\overline{a}^*\overline{m}^*})+\textstyle\sum_{\overline{m}}\{(Y_{\overline{am}}-Y_{\overline{a}^*\overline{m}})-(Y_{\overline{am}^*}-Y_{\overline{a}^*\overline{m}^*})\}1(\overline{G}_{\overline{a}^*}=\overline{m})\\&+\textstyle\sum_{\overline{m}}\{(Y_{\overline{am}}-Y_{\overline{a}^*\overline{m}})1(\overline{G}_{\overline{a}|v}=\overline{m})-(Y_{\overline{am}}-Y_{\overline{a}^*\overline{m}})1(\overline{G}_{\overline{a}^*}=\overline{m})\}+(Y_{\overline{a}^*\overline{G}_{\overline{a}}}-Y_{\overline{a}^*\overline{G}_{\overline{a}^*}})\\&=(Y_{\overline{am}^*}-Y_{\overline{a}^*\overline{m}^*})+\textstyle\sum_{m}(Y_{\overline{am}}-Y_{\overline{a}^*\overline{m}}-Y_{\overline{am}^*}+Y_{\overline{a}^*\overline{m}^*}\}1(\overline{G}_{\overline{a}^*}=\overline{m})\\&+\textstyle\sum_{\overline{m}}(Y_{\overline{am}}-Y_{\overline{a}^*\overline{m}})\{1(\overline{G}_{\overline{a}|v}=\overline{m})-1(\overline{G}_{\overline{a}^*}=\overline{m})\}+(Y_{\overline{a}^*\overline{G}_{\overline{a}}}-Y_{\overline{a}^*\overline{G}_{\overline{a}^*}}).\end{aligned} \quad \text{(A2)}$$

This latter expression has the same form of the four-way decomposition for a single time-fixed exposure and mediator. The four terms in either (A1) or (A2) might thus once again be referred to as the controlled direct effect, the reference interaction, the mediated interaction, and the interventional pure direct effect. Similar decompositions hold when conditioning on $V = v$ in the random draw from $\bar{M}_{\bar{a}}$ i.e. for decomposing $Y_{\bar{a}\bar{G}_{\bar{a}}|V} - Y_{\bar{a}^* \bar{G}_{\bar{a}}^*|V}$.

From the former expression in (A1), it is clear that each of the four components of the four-way decomposition can be identified on average under the same confounding assumptions ($i^\dagger$)–($iii^\dagger$) in the text for the interventional direct and indirect effects since each term involves expectation of the form of either $E[Y_{\overline{am}^*}]$ or of $E[Y_{\bar{a}\bar{G}_{\bar{a}}^*}]$, which, as noted in the text, are identified from the data under assumptions ($i^\dagger$)–($iii^\dagger$).

Suppose now that the marginal structural models above, allowing for exposure-mediator interaction, were fit to the data under assumptions ($i^\dagger$)–($iii^\dagger$):

$$\begin{aligned}E[Y_{\overline{am}}]&=\theta_0+\theta_1\text{cum}(\overline{a})+\theta_2\text{cum}(\overline{m})+\theta_3\text{cum}(\overline{a})\text{ cum}(\overline{m})\\E[M_{\overline{a}}(t)]&=\beta_0(t)+\beta_1(t)\text{ avg}(\overline{a}(t)).\end{aligned}$$

We have directly $E[Y_{\overline{am}}]$ directly from the marginal structural model for the outcome and it is shown in the Online Supplement that $E[Y_{\bar{a}\bar{G}_{\bar{a}}^*}]$ =

$$\theta_0+\theta_1\text{cum}(\overline{a})+\theta_2\left(\sum_{t\le T}\{\beta_0(t)+\beta_1(t)\text{ avg}(\overline{a}^*(t))\}\right)+\theta_3\text{cum}(\overline{a})\left(\sum_{t\le T}\{\beta_0(t)+\beta_1(t)\text{ avg}(\overline{a}^*(t))\}\right)$$

and this gives the expectations of all counterfactual quantities needed for each of the four components. We thus have that the controlled direct effect is given by:

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

$$\begin{array}{lcl} CDE(\overline{m}^*) & : & =E[Y_{\overline{a}\overline{m}^*}-Y_{\overline{a}^*\overline{m}^*}] \\ & = & \{\theta_1+\theta_3\text{cum}\,(\overline{m}^*)\}\{\text{cum}\,(\overline{a})-\text{cum}\,(\overline{a}^*)\}. \end{array}$$

The reference interaction is given by:

$$\begin{array}{l} INT_{ref}(\overline{m}^*): \\ \quad =E[\sum_m(Y_{\overline{a}\overline{m}}-Y_{\overline{a}^*\overline{m}}-Y_{\overline{a}\overline{m}^*}+Y_{\overline{a}^*\overline{m}^*}\}1(\overline{G}_{\overline{a}^*}=\overline{m})] \\ =E[\{(Y_{\overline{a}\overline{G}_{\overline{a}^*}}-Y_{\overline{a}^*\overline{G}_{\overline{a}^*}})-(Y_{\overline{a}\overline{m}^*}-Y_{\overline{a}^*\overline{m}^*})\}]=\theta_1\,\{\text{cum}\,(\overline{a}) \\ \quad -\text{cum}\,(\overline{a}^*)\} \\ +\theta_3\,\{\text{cum}\,(\overline{a}) \\ -\text{cum}\,(\overline{a}^*)\}\,\left(\sum_{t\le T}\{\beta_0\,(t)+\beta_1\,(t)\,\text{avg}\,(\overline{a}^*(t))\}\right)-\{\theta_1+\theta_3\text{cum}\,(\overline{m}^*)\}\{\text{cum}\,(\overline{a}) \\ -\text{cum}\,(\overline{a}^*)\}=\theta_3\,\{\text{cum}\,(\overline{a}) \\ -\text{cum}\,(\overline{a}^*)\}\,\left(\sum_{t\le T}\{\beta_0\,(t)+\beta_1\,(t)\,\text{avg}\,(\overline{a}^*(t))\}\right)-\theta_3\text{cum}\,(\overline{m}^*)\,\{\text{cum}\,(\overline{a}) \\ \quad -\text{cum}\,(\overline{a}^*)\}. \end{array}$$

The mediated interaction is given by:

$$\begin{array}{lcl} INT_{med} & : & =E[\sum_m(Y_{\overline{a}\overline{m}}-Y_{\overline{a}^*\overline{m}})\{1(\overline{G}_{\overline{a}|v}=\overline{m})-1(\overline{G}_{\overline{a}^*}=\overline{m})\}] \\ & = & E[Y_{\overline{a}\overline{G}_{\overline{a}}}-Y_{\overline{a}\overline{G}_{\overline{a}^*}}-Y_{\overline{a}^*\overline{G}_{\overline{a}}}+Y_{\overline{a}^*\overline{G}_{\overline{a}^*}}] \\ & = & \{\theta_2+\theta_3\text{cum}\,(\overline{a})\}\sum_{t\le T}\beta_1\,(t)\,\{\text{avg}\,(\overline{a}\,(t))-\text{avg}\,(\overline{a}^*(t))\}-\{\theta_2+\theta_3\text{cum}\,(\overline{a}^*)\}\sum_{t\le T}\beta_1\,(t)\,\{\text{avg}\,(\overline{a}\,(t))-\text{avg}\,(\overline{a}^*(t))\} \\ & = & \theta_3\{\text{cum}\,(\overline{a})-\text{cum}\,(\overline{a}^*)\}\sum_{t\le T}\beta_1\,(t)\,\{\text{avg}\,(\overline{a}\,(t))-\text{avg}\,(\overline{a}^*(t))\} \end{array}$$

And the interventional pure indirect effect is given by:

$$\begin{array}{lcl} PIE & : & =E(Y_{\overline{a}^*\overline{G}_{\overline{a}}}-Y_{\overline{a}^*\overline{G}_{\overline{a}^*}}) \\ & = & \{\theta_2+\theta_3\text{cum}\,(\overline{a}^*)\}\sum_{t\le T}\beta_1\,(t)\,\{\text{avg}\,(\overline{a}\,(t))-\text{avg}\,(\overline{a}^*(t))\}. \end{array}$$

Note that if $\beta_0(t) = \beta_0$ and $\beta_1(t) = \beta_1$ are assumed to be constant, and $a^*(t) = 0$ and $a\,(t) = 1$ for all t, and the reference level for the mediator is selected as $\overline{m}^* = \overline{0}$ the expressions simplify considerably to

$$\begin{array}{rcl} CDE(m^*) & = & \theta_1 T \\ INT_{ref}(m^*) & = & \beta_0\theta_3 T^2 \\ INT_{med} & = & \beta_1\theta_3 T^2 \\ PIE & = & \beta_1\theta_2 T. \end{array}$$

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

These expressions are analogous to those given in VanderWeele (2014) for a single time-fixed exposure and mediator. Note here that, once again, the controlled direct effect and the reference interaction sum to the interventional direct effect given in the previous section of the online supplement, $\theta_1 T + \beta_0 \theta_3 T^2$. And the mediated interaction and the interventional pure indirect effect sum to the interventional indirect effect given in the previous section, $\beta_1 T\{\theta_2 + \theta_3 T\}$.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

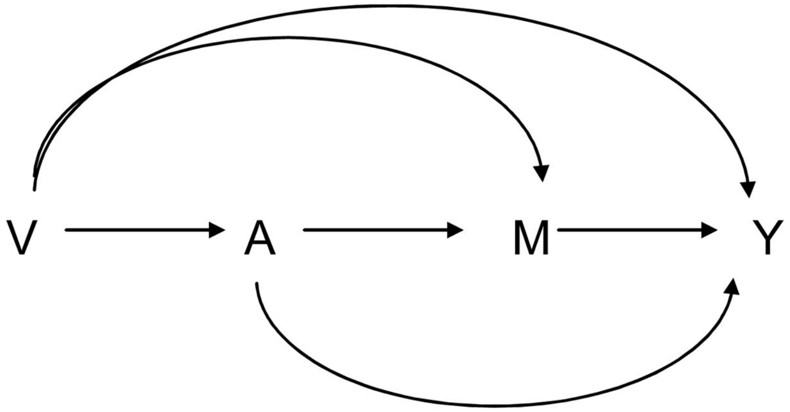


Figure 1.

Simple model for mediation.

Author Manuscript

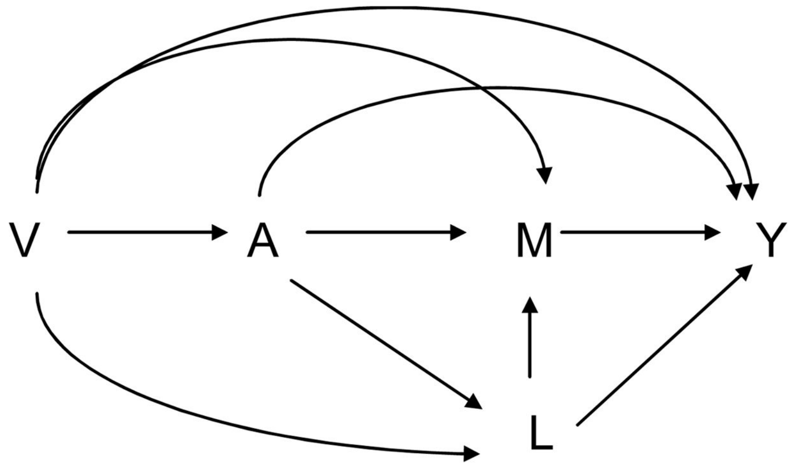


Figure 2.

Mediation with a mediator-outcome confounder L that is affected by exposure.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

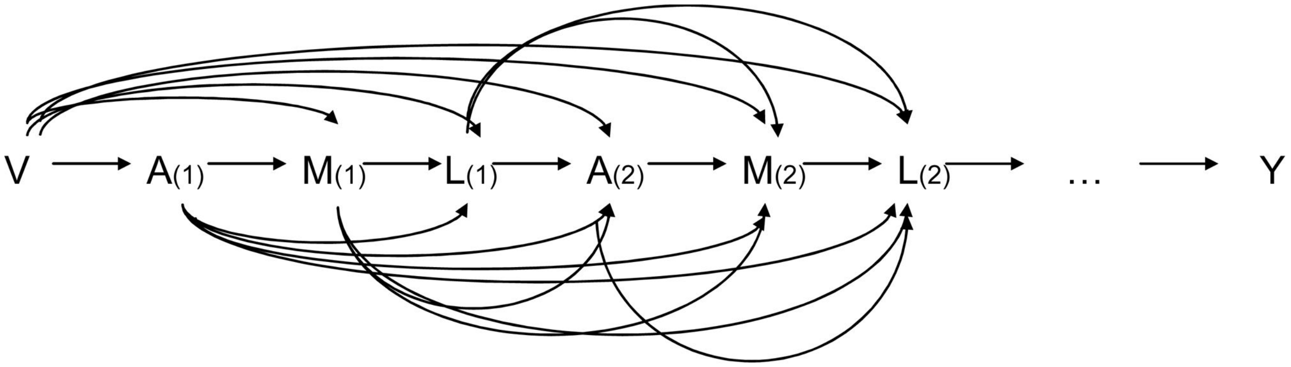


Figure 3.
Time-varying mediation with ordering of variables of A(t), M(t), L(t).

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

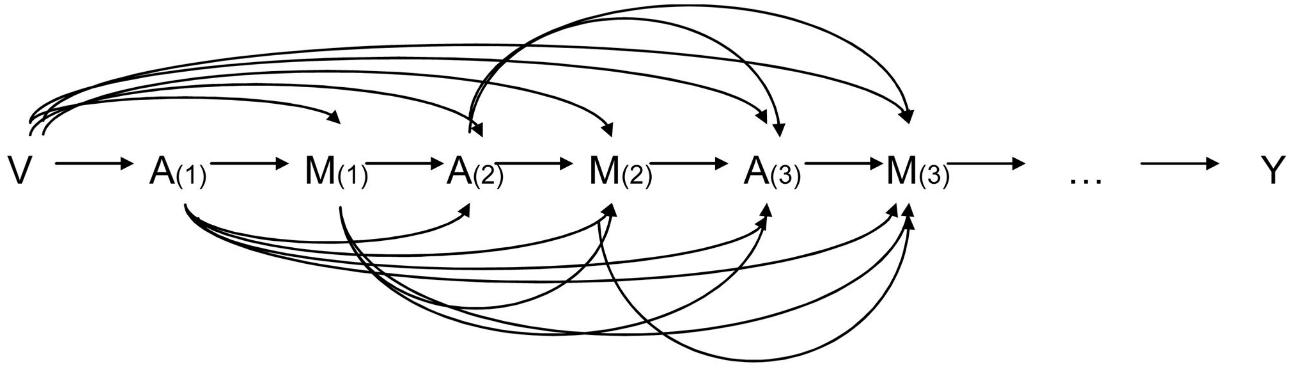


Figure 4.
Time-varying exposures and mediators, with no time-varying confounders.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

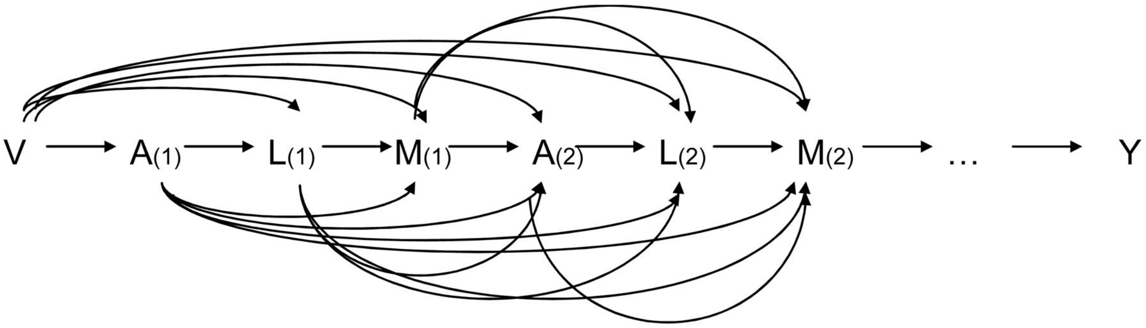


Figure 5.

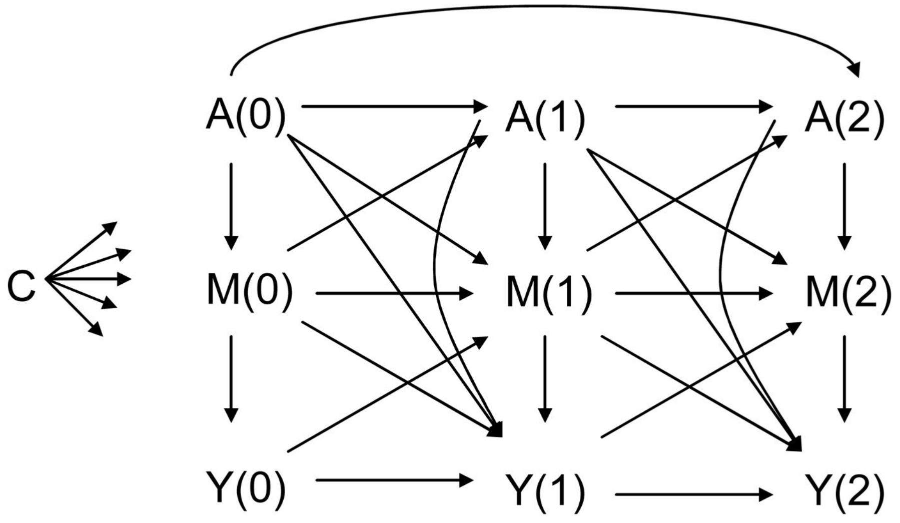
Time-varying mediation with variable ordering A(t), L(t), M(t).

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript


A(0)
A(1)
A(2)
C
M(0)
M(1)
M(2)
Y(0)
Y(1)
Y(2)