Tyler VanderWeele* and Stijn Vansteelandt

Mediation Analysis with Multiple Mediators

Abstract: Recent advances in the causal inference literature on mediation have extended traditional approaches to direct and indirect effects to settings that allow for interactions and non-linearities. In this article, these approaches from causal inference are further extended to settings in which multiple mediators may be of interest. Two analytic approaches, one based on regression and one based on weighting are proposed to estimate the effect mediated through multiple mediators and the effects through other pathways. The approaches proposed here accommodate exposure–mediator interactions and, to a certain extent, mediator–mediator interactions as well. The methods handle binary or continuous mediators and binary, continuous or count outcomes. When the mediators affect one another, the strategy of trying to assess direct and indirect effects one mediator at a time will in general fail; the approach given in this article can still be used. A characterization is moreover given as to when the sum of the mediated effects for multiple mediators considered separately will be equal to the mediated effect of all of the mediators considered jointly. The approach proposed in this article is robust to unmeasured common causes of two or more mediators.

Keywords: direct and indirect effects, joint effects mediation, regression, weighting

***Corresponding author: Tyler VanderWeele,** Departments of Epidemiology and Biostatistics, Harvard University, 677 Huntington Avenue, Boston, MA 02138, USA, E-mail: tvanderw@hsph.harvard.edu
Stijn Vansteelandt, Applied Mathematics and Computer Science, Ghent University, Ghent, Belgium, E-mail: stijn.vansteelandt@ugent.be

1 Introduction

The causal inference literature has provided important advances in mediation analysis in clarifying the assumptions needed for a causal interpretation of direct and indirect effect estimates and in developing approaches to mediation analysis allowing for models with interactions and non-linearities (Robins and Greenland, 1992; Pearl, 2001; van der Laan, Petersen, 2008; VanderWeele and Vansteelandt, 2009, 2010; Imai et al., 2010; Tchetgen Tchetgen and Shpitser, 2012; Valeri and VanderWeele, 2013). With a few notable exceptions (Avin et al., 2005; Albert and Nelson, 2011; Imai and Yamamoto, 2012; Zheng and van der Laan, 2012) most of this literature has been restricted to the setting of a single mediator.

When multiple mediators are of interest one approach would be to consider the mediators one at a time. As described below, however, this will in general require that the mediators do not affect one another. In this article we develop an approach that allows an investigator to assess mediation with multiple mediators simultaneously, and which can also accommodate cases in which the mediators affect one another. We allow for potential exposure–mediator interactions as well as, to a certain extent, mediator–mediator interactions. When the ordering of the mediators is known we also discuss how the magnitude of certain path-specific effects can be inferred by applying the approach sequentially. We give two different statistical techniques – one based on regression and one based on weighting – to estimate direct and indirect effects in these cases. We show how our approach is robust to unmeasured common causes of two or more mediators, whereas handling the mediators one-by-one will fail in these cases. When mediators are handled one-by-one, the sum of the proportion mediated on the additive scale for the

mediators can sometimes total more than 100%, even if the direction of mediation is the same for all mediators and pathways. We show that when this arises, it is because the mediators in fact affect one another or because of mediator–mediator interactions. The methods we propose in this article can handle these situations. In a related article (VanderWeele et al., 2013) we discuss effect decomposition in the presence of exposure-induced mediator outcome confounding in which there are multiple mediators but only one mediator is of principal interest. In this article we focus on assessing direct and indirect effects when multiple mediators are of interest simultaneously. Tchetgen Tchetgen and Shpitser (2012), Zheng and van der Laan (2012) and Vansteelandt et al. (2012) develop semi-parametric approaches that can also potentially be used for multiple mediators. Here we develop a simple parametric approach that is easy to implement extending the work of VanderWeele and Vansteelandt (2009, 2010) and Valeri and VanderWeele (2013) to handle multiple mediators.

2 Direct and indirect effects for a single mediator: a review

Let A denote the exposure for an individual, let Y denote some outcome, and let M denote the value of a single potential mediator that may be on the pathway from exposure to outcome. Let C denote some set of confounding variables that may affect the exposure, mediator and/or outcome. The relationships between A, M, Y and C are given in Figure 1.

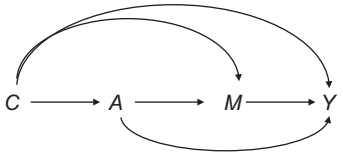

Figure 1 Mediation with a single mediator M, exposure A, outcome Y, and confounders C

There may be other mediators as well but when focusing on only one mediator these would be represented by the direct path from A to Y not through M.

Let Y_a denote a subject's potential or counterfactual outcome if exposure A were set, possibly contrary to fact, to a. Let M_a denote a subject's counterfactual value of the intermediate M if exposure A were set to the value a. Let Y_{am} denote a subject's counterfactual value for Y if A were set to a and M were set to m. For a single mediator, Robins and Greenland (1992) and Pearl (2001) provided the following definitions. The controlled direct effect of exposure A on outcome Y comparing $A = a$ with $A = a^*$ and setting M to m is defined by $Y_{am} - Y_{a^*m}$ and measures the effect of A on Y not mediated through M, i.e. the effect of A on Y after intervening to fix the mediator to some value m. The natural direct effect of exposure A on outcome Y comparing $A = a$ with $A = a^*$ intervening to set M to what it would have been if exposure had been $A = a^*$ is defined as $Y_{aM_{a^*}} - Y_{a^*M_{a^*}}$. Essentially, the natural direct effect assumes that the intermediate M is set to M_{a^*}, the level it would have been for each individual had exposure been a^*, and then compares the direct effect of exposure. The natural indirect effect is defined as $Y_{aM_a} - Y_{aM_{a^*}}$. The natural indirect effect assumes that exposure is set to some level $A = a$ and then compares what would have happened if the mediator were set to what it would have been if exposure had been a versus what would have happened if the mediator were set to what it would have been if exposure had been a^*. A total effect can be decomposed into a natural direct and indirect effect. The total effect $Y_a - Y_{a^*}$ can be written as $Y_a - Y_{a^*} = Y_{aM_a} - Y_{a^*M_{a^*}} = (Y_{aM_a} - Y_{aM_{a^*}}) + (Y_{aM_{a^*}} - Y_{a^*M_{a^*}})$, where the first expression in the sum is the indirect or mediated effect and the second expression is the natural direct effect. We can likewise define average controlled direct effect and natural direct and indirect effects conditional on covariates $C = c$ by $E[Y_{am} - Y_{a^*m}|c]$, $E[Y_{aM_{a^*}} - Y_{a^*M_{a^*}}|c]$, and $E[Y_{aM_a} - Y_{aM_{a^*}}|c]$ respectively.

With a binary outcome, we can define direct and indirect effects on a risk ratio or odds ratio scale (VanderWeele and Vansteelandt, 2010). On the odds ratio scale, the total effect conditional on $C = c$ is given by $OR^{TE}_{a,a^*|c} = \frac{P(Y_a=1|c)/\{1-P(Y_a=1|c)\}}{P(Y_{a^*}=1|c)/\{1-P(Y_{a^*}=1|c)\}}$. The controlled direct effect on the odds ratio scale conditional on $C = c$ is given by $OR^{CDE}_{a,a^*|c}(m) = \frac{P(Y_{am}=1|c)/\{1-P(Y_{am}=1|c)\}}{P(Y_{a^*m}=1|c)/\{1-P(Y_{a^*m}=1|c)\}}$. The natural direct effect on the odds ratio scale conditional on $C = c$ is given by $OR^{NDE}_{a,a^*|c}(a^*) = \frac{P(Y_{aM_{a^*}}=1|c)/\{1-P(Y_{aM_{a^*}}=1|c)\}}{P(Y_{a^*M_{a^*}}=1|c)/\{1-P(Y_{a^*M_{a^*}}=1|c)\}}$. The natural indirect effect on the odds ratio scale conditional on $C = c$ is given by $OR^{NIE}_{a,a^*|c}(a) = \frac{P(Y_{aM_a}=1|c)/\{1-P(Y_{aM_a}=1|c)\}}{P(Y_{aM_{a^*}}=1|c)/\{1-P(Y_{aM_{a^*}}=1|c)\}}$. On a risk ratio scale conditional on $C = c$, the total effect is given by $RR^{TE}_{a,a^*|c} = \frac{P(Y_a=1|c)}{P(Y_{a^*}=1|c)}$, the controlled direct effect is given by $RR^{CDE}_{a,a^*|c}(m) = \frac{P(Y_{am}=1|c)}{P(Y_{a^*m}=1|c)}$, and the natural direct effect is given by $RR^{NDE}_{a,a^*|c}(a^*) = \frac{P(Y_{aM_{a^*}}=1|c)}{P(Y_{a^*M_{a^*}}=1|c)}$. The natural indirect effect on the risk ratio scale conditional on $C = c$ is given by $RR^{NIE}_{a,a^*|c}(a) = \frac{P(Y_{aM_a}=1|c)}{P(Y_{aM_{a^*}}=1|c)}$. The total effect then decomposes into the product of the natural direct and indirect effects on the odds ratio or risk ratio scale: $OR^{TE}_{a,a^*|c} = OR^{NIE}_{a,a^*|c}(a) \times OR^{NDE}_{a,a^*|c}(a^*)$ and $RR^{TE}_{a,a^*|c} = RR^{NIE}_{a,a^*|c}(a) \times RR^{NDE}_{a,a^*|c}(a^*)$.

The identification of direct and indirect effects requires various no unmeasured confounding assumptions. We will use the notation $A \perp\!\!\!\perp B|C$ to denote that A is independent of B conditional on C. Controlled direct effects are identified if control is made for a covariate set C that includes all confounders of not only the exposure–outcome relationship but also the mediator–outcome relationship. In counterfactual notation, we require that for all a and m,

$$Y_{am} \perp\!\!\!\perp A|C \quad [1]$$

$$Y_{am} \perp\!\!\!\perp M|\{A, C\}. \quad [2]$$

Assumption [1] can be interpreted as: conditional on C, there is no unmeasured confounding for the exposure–outcome relationship. Assumption [2] can be interpreted as there is no unmeasured confounding for the mediator–outcome relationship conditional on (A,C). Natural direct and indirect effects will be identified if four no unmeasured confounding assumptions hold. Natural direct and indirect effects will be identified if, in addition to assumptions [1] and [2], the following two assumptions hold, that for all a, a^* and m,

$$M_a \perp\!\!\!\perp A|C \quad [3]$$

$$Y_{am} \perp\!\!\!\perp M_{a^*}|C. \quad [4]$$

Assumption [3] can be interpreted as: conditional on C, there is no unmeasured confounding of the exposure–mediator relationship. On a causal diagram interpreted as a set of non-parametric structural equations (Pearl, 2009), if assumption [2] holds, then assumption [4] will hold if there is no effect L of exposure A that itself affects both M and Y, i.e. no mediator–outcome confounder that is itself affected by the exposure A. If, however, there is an effect of the exposure that confounds the mediator–outcome relationship, then natural direct and indirect effects will not in general be identified irrespective of whether data is available on that variable or not (Avin et al., 2005). Exceptions arise under strong assumptions about no interaction at the individual level (Robins, 2003).

We will now present a simple regression-based approach for the estimation of direct and indirect effects, which we will extend below to allow for multiple mediators. Suppose assumptions [1]–[4] hold, that Y and M are continuous and that the following regression models for Y and M are correctly specified:

$$E[Y|a, m, c] = \theta_0 + \theta_1 a + \theta_2 m + \theta_3 am + \theta_4' c$$

$$E[M|a, c] = \beta_0 + \beta_1 a + \beta_2' c.$$

Then it can be shown (VanderWeele and Vansteelandt, 2009) that the average controlled direct effect and the average natural direct and indirect effects are given by

$$E[Y_{am} - Y_{a^*m}|c] = (\theta_1 + \theta_3 m)(a - a^*)$$

$$E\left[Y_{aM_{a^*}} - Y_{a^*M_{a^*}}|c\right] = \{\theta_1 + \theta_3(\beta_0 + \beta_1 a^* + \beta_2' c)\}(a - a^*)$$

$$E\left[Y_{aM_a} - Y_{aM_{a^*}}|c\right] = (\theta_2\beta_1 + \theta_3\beta_1 a)(a - a^*).$$

VanderWeele and Vansteelandt (2009) also derived standard errors for these effects using the delta method; alternatively bootstrapping can also be used. If there is no interaction between A and M so that $\theta_3 = 0$, then these expressions reduce to the expressions of Baron and Kenny (1986) employed in the psychology literature. The controlled direct effect and the natural direct effect are then both equal to $\theta_1(a - a^*)$ and the natural indirect effect is $\theta_2\beta_1(a - a^*)$.

Likewise for a binary outcome, suppose that the following models fit the observed data:

$$\text{logit}\{P(Y = 1|a, m, c)\} = \theta_0 + \theta_1 a + \theta_2 m + \theta_3 am + \theta_4' c$$

$$E[M|a, c] = \beta_0 + \beta_1 a + \beta_2' c.$$

and that the error term in the regression model for M is normally distributed with mean 0 and conditional variance σ^2. If the covariates C sufficed to control for confounding, satisfying assumptions [1]–[4] above, and the outcome were rare, then the conditional controlled direct effect and the average natural direct and indirect effects on the odds ratio scale would be given by (VanderWeele and Vansteelandt, 2010):

$$\log\left\{OR^{CDE}_{a,a^*|c}(m)\right\} = (\theta_1 + \theta_3 m)(a - a^*)$$

$$\log\left\{OR^{NDE}_{a,a^*|c}(a^*)\right\} \approx \left\{\theta_1 + \theta_3(\beta_0 + \beta_1 a^* + \beta_2' c + \theta_2\sigma^2)\right\}(a - a^*)$$

$$+ 0.5\theta_3^2\sigma^2(a^2 - a^{*2})$$

$$\log\left\{OR^{NIE}_{a,a^*|c}(a)\right\} \approx (\theta_2\beta_1 + \theta_3\beta_1 a)(a - a^*).$$

The approximations hold to the extent that the outcome is rare. These expressions would also hold exactly for a rare or common binary outcome if the logistic model was replaced by a log-linear model and natural direct and indirect effects on the risk ratio scale were used. Valeri and VanderWeele (2013) derived similar expressions for either binary or continuous outcomes when the mediator is binary. Similar expressions also hold if the binary outcome is replaced by a count outcome with a Poisson or negative binomial model and rate ratios are used (Valeri and VanderWeele, 2013).

3 Direct and indirect effects for multiple mediators

3.1 Notation, definitions and assumptions with multiple mediators

Suppose now that there are multiple mediators of interest, $\mathbf{M} = (M^{(1)}, \ldots, M^{(K)})$ and that we are interested in the effects mediated through $(M^{(1)}, \ldots, M^{(K)})$ jointly and the effects independent of $(M^{(1)}, \ldots, M^{(K)})$. We can define controlled direct effects and natural direct and indirect effects in a similar way as before simply replacing our single mediator M with the entire vector of mediators $\mathbf{M} = (M^{(1)}, \ldots, M^{(K)})$. Thus, let $\mathbf{M}_a$ be the counterfactual value of $\mathbf{M}$ if exposure A were set to the value a and let Y_{am} denote the counterfactual value for Y if A were set to a and $\mathbf{M}$ were set to $\mathbf{m}$. The controlled direct effect is defined by $Y_{a\mathbf{m}} - Y_{a^*\mathbf{m}}$; the natural direct effect is defined as $Y_{a\mathbf{M}_{a^*}} - Y_{a^*\mathbf{M}_{a^*}}$; the natural indirect effect is defined as $Y_{a\mathbf{M}_a} - Y_{a\mathbf{M}_{a^*}}$; and

once again the total effect can be decomposed into a natural direct and indirect effect: $Y_a - Y_{a^*} = Y_{a\mathbf{M}_a} - Y_{a^*\mathbf{M}_{a^*}} = (Y_{a\mathbf{M}_a} - Y_{a\mathbf{M}_{a^*}}) + (Y_{a\mathbf{M}_{a^*}} - Y_{a^*\mathbf{M}_{a^*}})$.

Suppose again that the four assumptions about confounding hold but now with respect to the whole set of mediators $\mathbf{M} = (M^{(1)}, \ldots, M^{(K)})$. In other words suppose we have (1) $Y_{a\mathbf{m}} \perp\!\!\!\perp A|C$, (2) $Y_{a\mathbf{m}} \perp\!\!\!\perp \mathbf{M}|\{A, C\}$, (3) $\mathbf{M}_a \perp\!\!\!\perp A|C$, and (4) $Y_{a\mathbf{m}} \perp\!\!\!\perp \mathbf{M}_{a^*}|C$. We once again need to control for all exposure–outcome, mediator–outcome, and exposure–mediator confounders, but note that now for assumptions [2] and [3] control must be made for the mediator–outcome confounders for all of the mediators, not just one and likewise control must be made for the exposure–outcome confounders for all of the mediators, not just one. Assumption [4] again requires that there be no effect of the exposure that confounds the mediator–outcome relationship for any of the mediators. If there were such a variable then to proceed it would have to be included in the mediator vector $\mathbf{M}$ if assumption [4] were not to be violated.

3.2 Regression-based approach for multiple mediators with a continuous outcome

Under these assumptions the natural direct and indirect effects can once again be estimated using a parametric regression-based approach. We will use one regression for the outcome Y and a separate regression for each of the mediators. For simplicity and ease of implementation we will use parametric rather than non-parametric regression here. We will begin with the case of a continuous outcome with continuous mediators and no interactions and we will consider extensions allowing for exposure–mediator interactions, mediator–mediator interactions as well as binary mediators and outcomes below.

Suppose then that assumptions [1]–[4] held for the vector of mediators $\mathbf{M} = (M^{(1)}, \ldots, M^{(K)})$ and that the following regressions are correctly specified and fit to the data:

$$E[Y|a, \mathbf{m}, c] = \theta_0 + \theta_1 a + \theta_2^{(1)} m^{(1)} + \theta_2^{(2)} m^{(2)} + \cdots + \theta_2^{(K)} m^{(K)} + \theta_4' c$$

$$E[M^{(i)}|a, c] = \beta_0^{(i)} + \beta_1^{(i)} a + \beta_2^{(i)'} c \text{ for } i = 1, \ldots, K.$$

We show in the Appendix that the controlled direct effect and natural direct and indirect effects are then given by

$$E[Y_{a\mathbf{m}} - Y_{a^*\mathbf{m}}|c] = \theta_1(a - a^*)$$

$$E\left[Y_{a\mathbf{M}_{a^*}} - Y_{a^*\mathbf{M}_{a^*}}|c\right] = \theta_1(a - a^*) \quad [5]$$

$$E\left[Y_{a\mathbf{M}_a} - Y_{a\mathbf{M}_{a^*}}|c\right] = \left[\beta_1^{(1)}\theta_2^{(1)} + \cdots + \beta_1^{(K)}\theta_2^{(K)}\right](a - a^*).$$

The direct effects are perhaps exactly what one might expect, simply the coefficient for the exposure, θ_1, in the model that contains all of the mediators. The natural indirect effect is equal to the sum over the various mediators, $M^{(1)}, \ldots, M^{(K)}$, of the product of the coefficient for the exposure ($\beta_1^{(k)}$ for the kth mediator) in the model for the mediator and the coefficient for the mediator ($\theta_2^{(k)}$ for the kth mediator) in the model for the outcome that has all the mediators. The indirect effect has this fairly intuitive form. Note, however, that this is different from applying the approach to mediation for a single mediator described in Section 2 one mediator at a time and then summing up the indirect effects. This is because if the mediators were handled one at a time then a different regression for Y would be fit for each mediator and only one mediator would be included in each of these regressions. The approach described in this section fits only a single regression for Y which includes all of the mediators under consideration.

In the Appendix we show that these two approaches will coincide if the mediators do not affect one another (or more precisely, if the mediators are independent of one another conditional on A and C) but they will diverge otherwise. They will diverge if they affect one another because certain pathways will be

counted twice if the mediation analysis is done one at a time. For example if there are two mediators where $M^{(1)}$ affects $M^{(2)}$ as in Figure 2 and if the analysis were done one mediator at a time then the path $A - M^{(1)} - M^{(2)} - Y$ would be included in the indirect effect both for the analysis for $M^{(1)}$ and for the analysis for $M^{(2)}$. If the two "indirect effects" were summed, the path would essentially be counted twice.

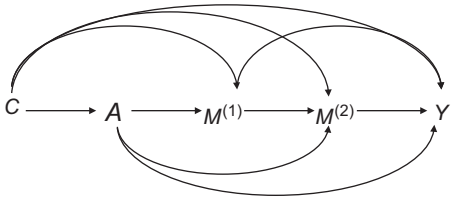

Figure 2 Mediation with two mediators of interest

The approach described in this section circumvents this difficulty by fitting only one regression for Y. As will be seen in the next subsection, however, the approach described here will also be able to be used even in the presence of interaction. Moreover, later in this article we will see that even if the mediators are in fact dependent on one another, the approach described here will be robust to unmeasured common causes of two or more mediators, whereas the approach considering the mediators one at a time will not be robust to such unmeasured variables.

When the mediators affect one another the approach of handling one mediator at a time also suffers from another difficulty. This other difficulty is that for the second (and potentially each subsequent) mediator, assumption [2] will not hold if the mediators are considered one at a time. This is because $M^{(1)}$ may affect both $M^{(2)}$ and Y and thus be a mediator–outcome confounder. Including $M^{(1)}$ in the covariate set C would not remedy this when $M^{(1)}$ is affected by A, for then assumption [4] is still violated. Assumption [4] may nonetheless hold with regard to a whole collection of mediators $\mathbf{M} = (M^{(1)}, \ldots, M^{(K)})$ without holding for each mediator individually. When mediators are considered one mediator at a time, natural direct and indirect effects will thus often not be identified except under strong assumptions about the absence of interaction (cf. Robins, 2003). See VanderWeele et al. (in press) for methods relevant to such exposure–induced mediator–outcome confounding.

3.3 Exposure–mediator interactions, binary mediators, and mediator–mediator interactions

In this subsection we will discuss how the simple approach above can be easily adapted to allow for exposure–mediator interactions, binary mediators, and, to a certain extent, mediator–mediator interactions. We will describe how to go about estimating causal effects for each of these variations; justification for all statements is given in the Appendix. Although it would not be difficult to give analytic standard errors for each of the variations below using the delta method, there are, as will be seen, numerous possible variations and new formulae would have to be derived in each case. We therefore recommend bootstrapping in the estimation of standard errors for the purposes of simplicity.

Suppose we wished to allow for an interaction between the exposure A and a mediator $M^{(i)}$ in the model for Y for example so that the outcome model became:

$$E[Y|a, \mathbf{m}, c] = \theta_0 + \theta_1 a + \theta_2^{(1)} m^{(1)} + \theta_2^{(2)} m^{(2)} + \cdots + \theta_2^{(K)} m^{(K)} + \theta_3^{(i)} a m^{(i)} + \theta_4' c.$$

Provided the models are correctly specified, the expressions for the controlled direct effect and natural direct and indirect effects in eq. [5] are then modified by adding $\theta_3^{(i)} m^{(i)}(a - a^*)$ to the controlled direct

effect, $\theta_3^{(i)}\{\beta_0^{(i)} + \beta_1^{(i)}a^* + \beta_2^{(i)'}c\}(a - a^*)$ to the natural direct effect and $\theta_3^{(i)}\beta_1^{(i)}a(a - a^*)$ to the natural indirect effect so that the effects become:

$$E[Y_{a\mathbf{m}} - Y_{a^*\mathbf{m}}|c] = \left\{\theta_1 + \theta_3^{(i)}m^{(i)}\right\}(a - a^*)$$

$$E\left[Y_{a\mathbf{M}_{a^*}} - Y_{a^*\mathbf{M}_{a^*}}|c\right] = \left[\theta_1 + \theta_3^{(i)}\left\{\beta_0^{(i)} + \beta_1^{(i)}a^* + \beta_2^{(i)'}c\right\}\right](a - a^*)$$

$$E\left[Y_{a\mathbf{M}_a} - Y_{a\mathbf{M}_{a^*}}|c\right] = \left[\beta_1^{(1)}\theta_2^{(1)} + \cdots + \beta_1^{(K)}\theta_2^{(K)} + \theta_3^{(i)}\beta_1^{(i)}a\right](a - a^*).$$

If a further interaction is thought to be present between the exposure and another of the mediators, for example mediator j say, then the same terms could once again be added to these expressions: $\theta_3^{(j)}m^{(j)}(a - a^*)$ to the controlled direct effect, $\theta_3^{(j)}\{\beta_0^{(j)} + \beta_1^{(j)}a^* + \beta_2^{(j)'}c\}(a - a^*)$ to the natural direct effect, and $\theta_3^{(j)}\beta_1^{(j)}a(a - a^*)$ to the natural indirect effect. And similarly for other exposure–mediator interactions; any number of exposure–mediator interactions could be accommodated in this manner.

Thus far we have assumed all mediators are continuous. Suppose that one or more of the mediators is binary, say mediator j, and that we fit a logistic regression model for $M^{(j)}$ (instead of a linear regression model):

$$\text{logit}P\left[M^{(j)} = 1|a, c\right] = \beta_0^{(j)} + \beta_1^{(j)}a + \beta_2^{(j)'}c.$$

Provided the models are correctly specified, the expressions for the controlled direct effect and natural direct effect then remain the same, but the expression for the natural indirect effect is modified. Instead of the term $\beta_1^{(j)}\theta_2^{(j)}(a - a^*)$ for the jth mediator we would include in the natural indirect effect the term: $\left(\frac{\exp\{\beta_0^{(j)}+\beta_1^{(j)}a+\beta_2^{(j)'}c\}}{1+\exp\{\beta_0^{(j)}+\beta_1^{(j)}a+\beta_2^{(j)'}c\}} - \frac{\exp\{\beta_0^{(j)}+\beta_1^{(j)}a^*+\beta_2^{(j)'}c\}}{1+\exp\{\beta_0^{(j)}+\beta_1^{(j)}a^*+\beta_2^{(j)'}c\}}\right)\theta_2^{(j)}$; i.e. we would replace $\beta_1^{(j)}(a - a^*)$ in the expression in eq. [5] with $\left(\frac{\exp\{\beta_0^{(j)}+\beta_1^{(j)}a+\beta_2^{(j)'}c\}}{1+\exp\{\beta_0^{(j)}+\beta_1^{(j)}a+\beta_2^{(j)'}c\}} - \frac{\exp\{\beta_0^{(j)}+\beta_1^{(j)}a^*+\beta_2^{(j)'}c\}}{1+\exp\{\beta_0^{(j)}+\beta_1^{(j)}a^*+\beta_2^{(j)'}c\}}\right)$ and we would likewise do this for each mediator that were binary. If we also wanted to allow for exposure–mediator interaction with a mediator that was binary, we would further add $\theta_3^{(j)}m^{(j)}(a - a^*)$ to the controlled direct effect, $\theta_3^{(j)}\frac{\exp\{\beta_0^{(j)}+\beta_1^{(j)}a^*+\beta_2^{(j)'}c\}}{1+\exp\{\beta_0^{(j)}+\beta_1^{(j)}a^*+\beta_2^{(j)'}c\}}(a - a^*)$ to the natural direct effect, and $\left(\frac{\exp\{\beta_0^{(j)}+\beta_1^{(j)}a+\beta_2^{(j)'}c\}}{1+\exp\{\beta_0^{(j)}+\beta_1^{(j)}a+\beta_2^{(j)'}c\}} - \frac{\exp\{\beta_0^{(j)}+\beta_1^{(j)}a^*+\beta_2^{(j)'}c\}}{1+\exp\{\beta_0^{(j)}+\beta_1^{(j)}a^*+\beta_2^{(j)'}c\}}\right)\theta_3^{(j)}a$ to the natural indirect effect. We could once again do this for each mediator that were binary and for which we wanted to include an exposure–mediator interaction.

Finally, suppose, that a binary exposure variable A were randomized and that no covariates were needed for assumptions [1]–[4] to hold. Suppose that we wanted to allow for mediator–mediator interaction between two mediators i and j as for example in the regression model:

$$E[Y|a, \mathbf{m}] = \theta_0 + \theta_1 a + \theta_2^{(1)}m^{(1)} + \theta_2^{(2)}m^{(2)} + \cdots + \theta_2^{(K)}m^{(K)} + \tau^{(ij)}m^{(j)}m^{(i)}.$$

Provided the models are correctly specified, the controlled direct effect and natural direct effect will both be exactly the same as described above. However, the natural indirect effect needs to be modified further. In particular, we could fit a linear regression model for the product

$$E[M^{(i)}M^{(j)}|a] = \beta_0^{(ij)} + \beta_1^{(ij)}a.$$

We then would add the term $\tau^{(ij)}\beta_1^{(ij)}(a - a^*)$ to the natural indirect effect. Unfortunately, as discussed in the appendix, if covariates C are included in the model then this can lead to issues of model compatibility between the models for $M^{(i)}$ and $M^{(j)}$ and that for the product $M^{(i)}M^{(j)}$. In section 4 we present an alternative

weighting approach that circumvents this issue and is applicable to settings with mediator–mediator interactions.

3.4 A regression-based approach for multiple mediators with a binary outcome

When the outcome is binary, rather than continuous, a similar approach to that described above, can also be employed but it is subject to the restriction that it will only work when all of the mediators are continuous. In the next section, however, we will also describe a weighting-based approach which can be used if the mediators are binary (or if some are binary and some are continuous) and that can also accommodate potential mediator–mediator interactions.

Suppose then the outcome is binary and all mediators are continuous and the following two models are correctly specified and fit to the data:

$$\text{logit}[P\{Y=1|a,\mathbf{m},c\}] = \theta_0 + \theta_1 a + \theta_2^{(1)} m^{(1)} + \theta_2^{(2)} m^{(2)} + \cdots + \theta_2^{(K)} m^{(K)} + \theta_4' c$$

$$E[M^{(i)} = 1|a,c] = \beta_0^{(i)} + \beta_1^{(i)} a + \beta_2^{(i)'} c \text{ for } i = 1,\ldots,K.$$

Suppose also now that the mediators $M^{(1)},\ldots,M^{(K)}$ follow a multivariate normal distribution conditional on A and C. We show in the Appendix that when the models are correctly specified and when assumptions [1]–[4] hold and when the outcome is rare (or the logistic regression model is replaced by a log-linear model for a common outcome with the effect measures interpreted as relative risks rather than odds ratios) then the log of the controlled direct effect and natural direct and indirect effect odds ratios are given by:

$$\log\left\{OR^{CDE}_{a,a^*|c}(m)\right\} = \theta_1(a-a^*)$$

$$\log\left\{OR^{NDE}_{a,a^*|c}(a^*)\right\} \approx \theta_1(a-a^*)$$

$$\log\left\{OR^{NIE}_{a,a^*|c}(a)\right\} \approx \left[\beta_1^{(1)}\theta_2^{(1)} + \cdots + \beta_1^{(K)}\theta_2^{(K)}\right](a-a^*).$$

These were the same expressions we had obtained in eq. [5] for a continuous outcome. If we wish to allow for exposure–mediator interactions we can simply add additional terms. Suppose we wished to allow for an interaction $\theta_3^{(i)} a m^{(i)}$ between the exposure A and a mediator $M^{(i)}$ in the model for Y. The expressions for the controlled direct effect and natural indirect effects are then modified by adding $\theta_3^{(i)} m^{(i)}(a-a^*)$ to the controlled direct effect, and $\theta_3^{(i)}\beta_1^{(i)} a(a-a^*)$ to the natural indirect effect; however, the expression for the natural direct effect is more complicated as it involves the correlation between the mediators; it is given in the Appendix.

If the data come from a case-control study with a rare outcome, then this same approach and the same expressions can be used but the regression models for the mediators are then fit only among the controls; see VanderWeele and Vansteelandt (2010) for further details. The same expressions can likewise be used for the ratio of expectations if the outcome is a count outcome and the logistic regression model is replaced by a log-linear model.

The approach described here for a binary or count outcome will, as noted above, only apply if all of the mediators are continuous. If one or more of the mediators are binary then an alternative approach will need to be used. Moreover, as we have seen, even if all of the mediators are continuous, the expressions for the natural direct effect become more complicated if there are exposure–mediator interactions. All of this motivates the alternative weighting approach, described below, which can much more flexibly accommodate binary outcomes.

3.5 Assessing mediators sequentially

Before moving on a few additional comments merit attention. First, the approach we have described so far does not necessarily require knowing the ordering of the mediators $M^{(1)}, \ldots, M^{(K)}$, though it does again require that there is no further variable that is affected by the exposure and that goes on to affect one of the mediators $M^{(1)}, \ldots, M^{(K)}$ and also the outcome. If there is such a variable it needs to be included in the set $M^{(1)}, \ldots, M^{(K)}$. If the ordering of $M^{(1)}, \ldots, M^{(K)}$ is known then some further progress can also be made. One could for example, begin with the first mediator $M^{(1)}$ and use the approach described here to examine the portion of the effect mediated through $M^{(1)}$. One could then consider $M^{(1)}$ and $M^{(2)}$ jointly and use the approach described here to examine what proportion of the effect is mediated through both $M^{(1)}$ and $M^{(2)}$ together. Doing so would allow one to assess the additional contribution of $M^{(2)}$ beyond $M^{(1)}$ alone. Note that the difference between the two will potentially be different than simply the effect mediated through $M^{(2)}$ itself because for example, $M^{(1)}$ and $M^{(2)}$ may share common pathways (if for example $M^{(1)}$ affected $M^{(2)}$ or if, as discussed later in this article, $M^{(1)}$ and $M^{(2)}$ interact in their effects). One could further then consider $(M^{(1)}, M^{(2)}, M^{(3)})$ and examine the proportion mediated by all three jointly along with the additional contribution of $M^{(3)}$ beyond $(M^{(1)}, M^{(2)})$. One could carry on this process, adding sequentially one mediator at a time, until all K mediators are included.

Undertaking this sequential approach does, however, place additional restrictions on the models being used. This is because for each group of mediators a different model is being fit for Y. In particular, as is discussed at greater length in the Appendix, for the various models for Y to be compatible with one another when the exposure is binary, it is necessary that either (i) there are no exposure–mediator or mediator–mediator interactions or (ii) the models must be extended to allow for exposure–covariate interaction. See the appendix for further details.

4 A weighting approach

Because of the aforementioned concerns about model incompatibility that arise because of mediator–mediator interactions and because the addition of mediators increases the need for modeling, we also present a simple alternative approach based on inverse probability weighting. This alternative weighting approach does not require models for the mediators. Instead a model for the exposure is used and this then essentially overcomes the issue of model incompatibility. This weighting approach can be used for essentially any type of outcome, including non-rare binary outcomes, and it can be used regardless of whether there are exposure–mediator or mediator–mediator interactions. However, as with other weighting approaches, its performance is best when the exposure is binary or discrete with only a few levels.

The weighting approach estimates the marginal natural direct effect, $E[Y_{a\mathbf{M}_{a^*}}] - E[Y_{a^*\mathbf{M}_{a^*}}]$, and the marginal natural indirect effect, $E[Y_{a\mathbf{M}_a}] - E[Y_{a\mathbf{M}_{a^*}}]$. Doing so, requires the estimation of three counterfactuals: $E[Y_{a\mathbf{M}_{a^*}}]$, $E[Y_{a^*\mathbf{M}_{a^*}}]$, and $E[Y_{a\mathbf{M}_a}]$. For the counterfactual $E[Y_{a^*\mathbf{M}_{a^*}}]$ we can obtain this by taking a weighted average of the subjects with $A = a^*$ where each subject i is given a weight

$$\frac{P(A = a^*)}{P(A = a^*|c_i)}$$

where c_i denotes the actual covariate value for subject i. For a binary exposure with $a = 1$ and $a^* = 0$, the probabilities $P(A = 0|c_i)$ could be fit, for example, using logistic regression and obtaining the predicted probabilities for $P(A = 0|c_i)$ for each subject i. The approach requires that such models are correctly specified. Likewise we can obtain the counterfactual $E[Y_{a\mathbf{M}_a}]$ by taking a weighted average of the subjects with $A = a$ where each subject i is given a weight

$$\frac{P(A = a)}{P(A = a|c_i)}$$

where again c_i denotes the actual covariate value for subject i. And again for a binary exposure with $a = 1$ and $a^* = 0$, the probabilities $P(A = 1|c_i)$ could be fit, for example, using logistic regression and obtaining the predicted probabilities for $P(A = 1|c_i)$ for each subject i.

Finally, for the counterfactual $E[Y_{a\mathbf{M}_{a^*}}]$, for each subject i with $A_i = a^*$, one uses an outcome model $E(Y|a, \mathbf{m}_i, c_i)$ (which can include exposure–mediator or mediator–mediator interactions) to obtain a predicted estimate of the outcome if the individual had had exposure $A_i = a$ rather than $A_i = a^*$, but using the individual's own values of the mediator, $\mathbf{M} = \mathbf{m}_i$, and covariates, $C = c_i$. The weighting approach requires that the outcome model is correctly specified. Once these predicted values are calculated one can obtain an estimate for the counterfactual $E[Y_{a\mathbf{M}_{a^*}}]$ by taking a weighted average of these predicted values for subjects with $A_i = a^*$ where each subject i is given the weight

$$\frac{P(A = a^*)}{P(A = a^*|c_i)}.$$

Once the various counterfactuals are obtained, we can estimate the natural direct effect by taking the difference $E[Y_{a\mathbf{M}_{a^*}}] - E[Y_{a^*\mathbf{M}_{a^*}}]$, and the natural indirect effect by taking the difference $E[Y_{a\mathbf{M}_a}] - E[Y_{a\mathbf{M}_{a^*}}]$. Alternatively, on a ratio scale we could obtain risk ratios for the natural direct effect by taking the ratio $E[Y_{a\mathbf{M}_{a^*}}]/E[Y_{a^*\mathbf{M}_{a^*}}]$, and for the natural indirect effect by taking the ratio $E[Y_{a\mathbf{M}_a}]/E[Y_{a\mathbf{M}_{a^*}}]$ and likewise for effects on an odds ratio scale. We recommend bootstrapping for confidence intervals. This approach essentially constitutes a straightforward generalization of the result in Albert (2012) to the case of a vector of mediators and is also closely related to the imputation approach of Vansteelandt et al. (2012). SAS code to implement this weighting approach is given in the Appendix. As with other weighting approaches, the approach here can be unstable if some of the probabilities in the denominator are very small so that some of the weights are very large; it is also good to check overlap in the distribution of the weights among the exposed and unexposed.

Both the weighting approach in this section and the regression approach in the previous section assume that the models are correctly specified. However, these different approaches make different modeling assumptions in that they require different models to be correctly specified. The regression approach in the previous section requires that the model for the outcome and the models for each of the mediators are correctly specified; no model for the exposure is needed in the regression approach. The approach will be biased if the models for the either the outcome or the mediators are mis-specified. In contrast, the weighting approach requires that the model for the outcome and the model for the exposure are correctly specified; no models for the mediators are needed. The weighting approach will be biased if the models for the either the outcome or the exposure is mis-specified.

5 Illustration

To illustrate the weighting approach we will analyze 2003 US birth certificate data and will consider whether the exposure, A, of adequate or inadequate prenatal care ($n = 2,629,247$; those with intermediate or superadequate care are excluded from the analysis for the purposes of this illustration) on preterm birth (Y) is mediated by maternal smoking and/or drinking ($M^{(1)}$) or pre-eclampsia ($M^{(2)}$). Adequacy of prenatal care categories are determined from data on the month prenatal care was initiated, on the number of visits, and on gestational age, according to the American College of Gynecologists recommendation as encoded in a modification of the APNCU index (Kotelchuck, 1994; VanderWeele et al., 2009). In this analysis we will take age category (below 20 years, between 20 and 35 years, or above 35 years), ethnicity (black, Hispanic, native American, white), education and marital status as baseline confounders (C). Our analysis is certainly a simplification of a more complex reality as prenatal care and maternal smoking are both ultimately time-varying and pre-eclampsia and preterm birth could be conceived of as processes whereas we will treat them as dichotomous.

Inverse probability weights were constructed on the basis of logistic regression models for adequate care. In view of the large sample size and the resulting computational burden, standard errors and confidence intervals were constructed using the subsampling bootstrap (Politis and Romano, 1994). This is similar to the bootstrap, but involved repeating the analysis for 1,000 subsamples of size $n = 13,146$ (0.5% of the total sample size); on the basis of the empirical standard deviation of the 1,000 estimates, the standard error of the estimates that were obtained from the analysis of the full data set can be inferred (accounting for correlation resulting from the fact that some data points may be shared between subsamples).

In the sequential approach, we first considered maternal drinking and smoking as mediators. This shows that the direct effect of adequate care, through pathways other maternal smoking or drinking ($M^{(1)}$), is a 5.6% (95% CI 5.5% to 5.7%) reduction in the risk of preterm birth and that the mediated effect via maternal smoking and/or drinking is a 0.09% (95% CI 0.08% to 0.10%) reduction in the risk of preterm birth. When pre-eclampsia is considered as an additional mediator ($M^{(2)}$), we found essentially the same direct and indirect effects of 5.6% (95% CI 5.5% to 5.7%) and 0.09% (95% CI 0.08% to 0.10%). The effect of adequate prenatal care on preterm birth by pathways through pre-eclampsia, but not through maternal smoking and drinking thus seem minimal.

6 Some further properties: robustness to mediator confounding and joint versus summed proportion mediated

As noted above, when multiple mediators are of interest, the approach of considering mediators one at a time will only be appropriate if the mediators do not affect one another. If one of the mediators of interest affects another then assumption [4] will be violated for one or more mediators. The approach we have described in this article can, however, still be used. The approach described here has other advantages, even if the mediators do not affect one another. Suppose for example the mediators do not affect each other but there is an unmeasured common cause U of two or more mediators as in Figure 3.

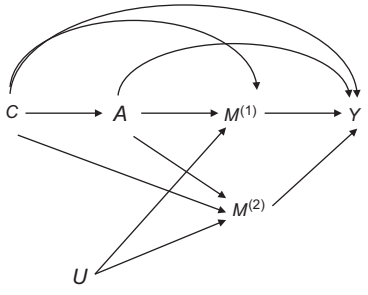

Figure 3 Two mediators with an unmeasured common cause

In this case, the approach of considering the mediators $M^{(1)}$ and $M^{(2)}$ one-by-one will be biased because when $M^{(1)}$ alone is considered, U will be an unmeasured confounder for the effect of $M^{(1)}$ on Y (it affects Y through $M^{(2)}$) and when $M^{(2)}$ alone is considered, U will be an unmeasured confounder of the effect of $M^{(2)}$ on Y (it affects Y through $M^{(1)}$). However, when $M^{(1)}$ and $M^{(2)}$ are considered jointly as in this article, U no longer serves as a confounder for the joint effect of $(M^{(1)}, M^{(2)})$ on Y because U no longer affects Y, except through $(M^{(1)}, M^{(2)})$.

When the mediators affect one another then, as discussed above, we generally cannot estimate the natural direct and indirect effects for one or more mediators. Even if we could, the sum of the proportion mediated can be more than 100%, even if all pathways affect the outcome in the same direction. This is because certain paths may be counted twice. In Figure 2, if the analysis were done one mediator at a time then the path $A - M^{(1)} - M^{(2)} - Y$ would be included in the indirect effect both for the analysis of $M^{(1)}$ and for the analysis

of $M^{(2)}$. The approach described in this article circumvents this difficulty. However, we might then think that if the mediators do not affect one another – if for example, the mediators are independent of one another conditional on A and C – then the sum of two mediated effects considered separately should equal the joint mediated effect when both mediators are considered together. In fact, even if the mediators are independent and do not affect each other this need not hold. The sum of two mediated effects considered separately may diverge from the joint mediated effect when there are interactions between the effects of the two mediators on the outcome. Note that such interaction can arise even if the mediators do not affect each other. We show in the Appendix that if there is no additive interaction in the effects of the two mediators at the individual counterfactual level, and if the two mediators do not affect each other then the sum of two mediated effects on the additive scale, considered separately, will equal the joint mediated effect when both are considered together. If the two diverge then either the mediators must affect one another or there must be an additive interaction at the individual level between the effects of the two mediators.

In some applications, mediators are considered one at a time and the proportion mediated is calculated for each of these. Sometimes, when doing this, the sum of the proportion mediated can exceed 100%. One possible explanation for this is that there are other pathways (that operate through other mediators) that affect the outcomes in the opposite direction from those under consideration. The sum of the proportion mediated may exceed 100% if there are other mediators with a "negative" proportion mediated. If this is thought not to be the case, i.e. if all pathways are thought to operate on the outcome in the same direction, then the true proportion mediated for all mediators (known and unknown) considered jointly must be 100%. If the sum of the proportion mediated when each measured mediator is considered separately exceeds 100% then the sum and the joint proportion mediated would be different and thus it must be the case that either the mediators affect one another or that there are interactions between the effects of the mediators on the outcome. The approach described in this article could accommodate these complications by considering all mediators jointly in contrast to the approach of assessing mediators one at a time which cannot. In summary, if the sum of the proportion mediated exceeds 100% then one of the following must be true: (i) there are other mediators with a negative proportion mediated; (ii) the mediators affect one another; (iii) there are interactions between the effects of the mediators on the outcome.

Comparing the sum of the mediated effects to a joint mediated effect (or examining if the sum of the proportion mediated exceeds 100% if all mediators are thought to operate in the same direction) would thus constitute one strategy whereby an investigator could assess whether the approach of examining one mediator at a time might fail. An alternative approach might consist of examining the independence of the mediators more directly. For example, in the case of two mediators, $M^{(1)}$ and $M^{(2)}$, a regression of $M^{(2)}$ on $M^{(1)}, A, C$ should have $M^{(1)}$ independent of $M^{(2)}$ in the regression. Statistical dependence between the two conditional on A and C would indicate that the approach of examining mediators one at a time cannot be used.

Statistical dependence between $M^{(1)}$ and $M^{(2)}$ conditional on A and C cannot distinguish between Figures 2 and 3 (or Figure 4 in which one mediator affects the other and they share an unmeasured common cause), but in either case, the approach of examining the mediators one at a time fails because assumptions required for such an approach are then violated.

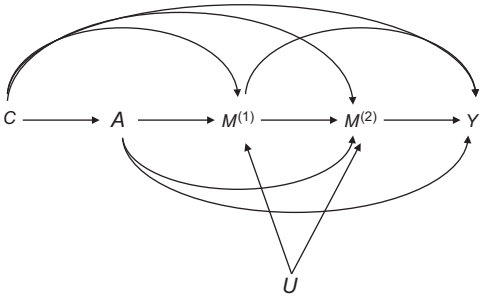

Figure 4 Two mediators in which one affects the other and they share an unmeasured common cause U

7 Discussion

In this article we have considered methods for assessing the effects of an exposure on an outcome through several mediators considered jointly. We have described a regression-based approach and a weighting-based approach, which are simple extensions of methods for a single mediator (VanderWeele and Vansteelandt, 2009, 2010; Valeri and VanderWeele, 2013; Albert, 2012). The weighting-based approach allowed for somewhat more flexibility in settings with a binary outcome, in allowing for mediator–mediator interaction and mediators of different types (e.g., continuous and categorical). Otherwise, however, the regression-based approach is perhaps to be preferred on the grounds that it will in general yield more efficient estimates.

The approaches considered here are parametric and require correct model specification. Both approaches demand correct specification of an outcome model, but differ in that the regression-based approach additionally requires correct specification of models for each of the mediators, whereas the weighting approach additionally requires correct specification of a model for the exposure. When the number of mediators is large, the weighting approach is thus considerably less demanding in terms of the number of models it requires. It is moreover less demanding in settings where the exposure is highly predictive of (one of) the mediators, since correct specification of a model for the mediators is then more difficult. The weighting approach is particularly desirable when the exposure is randomly assigned so that the model for the exposure is known by design: it then merely requires correct specification of a model for the outcome.

Recent work has proposed multiply robust estimators that specify models for the exposure, mediator and outcome and yield valid inferences if at least two of these models are correctly specified (Tchetgen Tchetgen and Shiptser, 2012; Vansteelandt et al., 2012; Zheng and van der Laan, 2012). These semiparametric approaches have been studied in settings with a single mediator, but can in principle deal with multiple mediators as well. They have greater robustness to model mis-specification but are more difficult to implement in practice. Robustness against model mis-specification can alternatively be improved using a slight modification of the weighting approach, which uses outcome predictions obtained via more general statistical learning methods rather than parametric methods. The practical performance of these different variations and approaches deserves further study, particularly in settings where the number of confounders is large and model mis-specification thus more likely.

The assumptions required for the estimation for the approaches proposed here, like all work on mediation, are quite strong. Sensitivity analysis for direct and indirect effects are now available for a single mediator (VanderWeele, 2010; Imai et al., 2010) and these approaches could perhaps be applied and extended to settings with multiple mediators. However, by being able to handle multiple mediators, at least one of the assumptions for the estimation of direct and indirect effects is in some sense made more plausible: namely the assumption that there is no mediator-outcome confounder that is itself affected by the exposure. The approach described in this article renders this assumption more plausible in that if there were such a variable then it itself could be included in the mediator vector and the methods could still be employed. The methods we have provided here are relatively general and also fairly straightforward to use in practice and we hope will be of use in settings in which multiple mediators might be of interest.

Appendix

Regression-based approach

Consider the regression models:

$$E[Y|a,\mathbf{m},c] = \theta_0 + \theta_1 a + \sum_{i=1}^{K} \theta_2^{(i)} m^{(i)} + \sum_{i=1}^{K} \theta_3^{(i)} a m^{(i)} + \sum_{i=1,j=1}^{K,K} \tau^{(ij)} m^{(i)} m^{(j)} + \theta_4' c$$

$$E[M^{(i)}|a,c] = \beta_0^{(i)} + \beta_1^{(i)} a + \beta_2^{(i)'} c \text{ for } M^{(i)} \text{continuous}$$

$$\text{logit}\left\{P[M^{(i)} = 1|a, c]\right\} = \beta_0^{(i)} + \beta_1^{(i)}a + \beta_2^{(i)'}c \text{ for } M^{(i)}\text{binary}$$

Under assumptions [1] and [2], we have for the controlled direct effect:

$$\begin{aligned} E[Y_{am} - Y_{a^*m}|c] &= E[Y|a, \mathbf{m}, c] - E[Y|a^*, \mathbf{m}, c] \\ &= \theta_1 a + \sum_{i=1}^{K} \theta_3^{(i)} a m^{(i)}. \end{aligned}$$

Under assumptions [1]–[4] we have by Pearl's mediation formula

$$\begin{aligned} &E\left[Y_{a\mathbf{M}_{a^*}}|c\right] \\ &= \int_{\mathbf{m}} E[Y|c, a, \mathbf{m}]dp(\mathbf{m}|c, a*) \\ &= \int_m \left\{\theta_0 + \theta_1 a + \sum_{i=1}^{K} \theta_2^{(i)} m^{(i)} + \sum_{i=1}^{K} \theta_3^{(i)} a m^{(i)} + \sum_{i=1,j=1}^{K,K} \tau^{(ij)} m^{(i)} m^{(j)} + \theta_4' c\right\} dP(\mathbf{m}|c, a^*) \\ &= \theta_0 + \theta_1 a + \sum_{i=1}^{K} \theta_2^{(i)} E\left[M^{(i)}|c, a^*\right] + \sum_{i=1}^{K} \theta_3^{(i)} a E\left[M^{(i)}|c, a^*\right] \\ &\quad + \sum_{i=1,j=1}^{K,K} \tau^{(ij)} E\left[M^{(i)} M^{(j)}|c, a^*\right] + \theta_4' c. \end{aligned}$$

Thus the natural direct effect is given by:

$$\begin{aligned} &E\left[Y_{a\mathbf{M}_{a^*}} - Y_{a^*\mathbf{M}_{a^*}}|c\right] \\ &= \theta_0 + \theta_1 a + \sum_{i=1}^{K} \theta_2^{(i)} E\left[M^{(i)}|c, a^*\right] + \sum_{i=1}^{K} \theta_3^{(i)} a E\left[M^{(i)}|c, a^*\right] \\ &\quad + \sum_{i=1,j=1}^{K,K} \tau^{(ij)} E\left[M^{(i)} M^{(j)}|c, a^*\right] + \theta\prime_4 c \\ &\quad - \left[\theta_0 + \theta_1 a^* + \sum_{i=1}^{K} \theta_2^{(i)} E\left[M^{(i)}|c, a^*\right] + \sum_{i=1}^{K} \theta_3^{(i)} a^* E\left[M^{(i)}|c, a^*\right]\right. \\ &\quad \left. + \sum_{i=1,j=1}^{K,K} \tau^{(ij)} E\left[M^{(i)} M^{(j)}|c, a^*\right] + \theta_4' c\right] \\ &= \left\{\theta_1 + \sum_{i=1}^{K} \theta_3^{(i)} E\left[M^{(i)}|c, a^*\right]\right\}(a - a^*). \end{aligned}$$

If $M^{(i)}$ is continuous, $E[M^{(i)}|c, a^*] = \beta_0^{(i)} + \beta_1^{(i)}a + \beta_2^{(i)'}c$. If $M^{(i)}$ is binary, $E[M^{(i)}|c, a^*] = \frac{\exp\{\beta_0^{(i)} + \beta_1^{(i)}a^* + \beta_2^{(i)'}c\}}{1+\exp\{\beta_0^{(i)} + \beta_1^{(i)}a^* + \beta_2^{(i)'}c\}}$. The natural indirect effect is given by

$$\begin{aligned} &E[Y_{a\mathbf{M}_a} - Y_{a\mathbf{M}_{a^*}}|c] \\ &= \theta_0 + \theta_1 a + \sum_{i=1}^{K} \theta_2^{(i)} E\left[M^{(i)}|c, a\right] + \sum_{i=1}^{K} \theta_3^{(i)} a E\left[M^{(i)}|c, a\right] \\ &\quad + \sum_{i=1,j=1}^{K,K} \tau^{(ij)} E\left[M^{(i)} M^{(j)}|c, a\right] + \theta_4' c \\ &\quad - \left\{\theta_0 + \theta_1 a + \sum_{i=1}^{K} \theta_2^{(i)} E\left[M^{(i)}|c, a^*\right] + \sum_{i=1}^{K} \theta_3^{(i)} a E\left[M^{(i)}|c, a^*\right]\right. \end{aligned}$$

$$+\sum_{i=1,j=1}^{K,K}\tau^{(ij)}E\left[M^{(i)}M^{(j)}|c,a^*\right]+\theta\prime_4 c\Big\}$$

$$=\sum_{i=1}^{K}\left\{\theta_2^{(i)}+\theta_3^{(i)}a\right\}\left\{E\left[M^{(i)}|c,a\right]-E\left[M^{(i)}|c,a^*\right]\right\}$$

$$+\sum_{i=1,j=1}^{K,K}\tau^{(ij)}\left\{E\left[M^{(i)}M^{(j)}|c,a\right]-E\left[M^{(i)}M^{(j)}|c,a^*\right]\right\}.$$

If $M^{(i)}$ is continuous, $E[M^{(i)}|c,a]-E[M^{(i)}|c,a^*]=\beta_1^{(i)}(a-a^*)$. If $M^{(i)}$ is binary, $E[M^{(i)}|c,a]-E[M^{(i)}|c,a^*]=\frac{\exp\{\beta_0^{(i)}+\beta_1^{(i)}a+\beta_2^{(i)'}c\}}{1+\exp\{\beta_0^{(i)}+\beta_1^{(i)}a+\beta_2^{(i)'}c\}}-\frac{\exp\{\beta_0^{(i)}+\beta_1^{(i)}a^*+\beta_2^{(i)'}c\}}{1+\exp\{\beta_0^{(i)}+\beta_1^{(i)}a^*+\beta_2^{(i)'}c\}}$.

For the mediator–mediator interaction terms, we could consider the following models: $E[M^{(i)}M^{(j)}|a,c]=\beta_0^{(ij)}+\beta_1^{(ij)}a+\beta_2^{(ij)'}c$ for at least one of $M^{(i)},M^{(j)}$ continuous $\log it\{P[M^{(i)}M^{(j)}=1|a,c]\}=\beta_0^{(ij)}+\beta_1^{(ij)}a+\beta_2^{(ij)'}c$ for $M^{(i)},M^{(j)}$ both binary.

Under these models, if at least one of $M^{(i)}$ or $M^{(j)}$ is continuous, $E[M^{(i)}M^{(j)}|c,a]-E[M^{(i)}M^{(j)}|c,a^*]=\beta_1^{(ij)}(a-a^*)$. If both $M^{(i)}$ and $M^{(j)}$ are binary $E[M^{(i)}M^{(j)}|c,a]-E[M^{(i)}M^{(j)}|c,a^*]=\frac{\exp\{\beta_0^{(ij)}+\beta_1^{(ij)}a+\beta_2^{(ij)'}c\}}{1+\exp\{\beta_0^{(ij)}+\beta_1^{(ij)}a+\beta_2^{(ij)'}c\}}-\frac{\exp\{\beta_0^{(ij)}+\beta_1^{(ij)}a^*+\beta_2^{(ij)'}c\}}{1+\exp\{\beta_0^{(ij)}+\beta_1^{(ij)}a^*+\beta_2^{(ij)'}c\}}$. Note, however, if covariates C are included in the model and a mediator–mediator interaction term, $\tau^{(ij)}m^{(i)}m^{(j)}$ is also included, then this can lead to issues of model compatibility between the models for $M^{(i)}$ and $M^{(j)}$ and that for the product $M^{(i)}M^{(j)}$. For continuous mediators, a possible solution would be to assume a constant covariance matrix, in which case the average product follows from knowledge of the covariance and means. For instance, we would have that

$$E[M^{(i)}M^{(j)}|c,a]=(\beta_0^{(i)}+\beta_1^{(i)}a+\beta_2^{(i)'}c)(\beta_0^{(j)}+\beta_1^{(j)}a+\beta_2^{(j)'}c)+\mathrm{Cov}(\varepsilon^{(i)},\varepsilon^{(j)})$$

where $\varepsilon^{(i)},\varepsilon^{(j)}$ are the residuals in the models for both mediators. For dichotomous mediators, the Plackett copula could be used; that is, on top of the models for each mediator separately, one could postulate the model $\mathrm{odds}(M^{(i)}=1|M^{(j)}=1,a,c)/\mathrm{odds}(M^{(i)}=1|M^{(j)}=0,a,c)=\alpha$ with α unknown; α could be estimated using standard software for alternating logistic regression, which is, for instance, available via the option "logor = exch" in proc genmod. With $p^{(i)}\equiv E[M^{(i)}|c,a]$ and $p^{(j)}\equiv E[M^{(j)}|c,a]$, we then have that

$$E\left[M^{(i)}M^{(j)}|c,a\right]=\frac{1}{2(\alpha-1)}\Big[1-(p^{(i)}+p^{(j)})(1-\alpha)$$

$$-\left\{(1-(p^{(i)}+p^{(j)})(1-\alpha)^2-4p^{(i)}p^{(j)}\alpha(\alpha-1)\right\}^{1/2}\Big].$$

Binary outcome

Now suppose that the outcome is binary and rare and the following regression models are fit to the data:

$$\mathrm{logit}[P\{Y=1|a,\mathbf{m},c\}]=\theta_0+\theta_1 a+\sum_{i=1}^{K}\theta_2^{(i)}m^{(i)}+\sum_{i=1}^{K}\theta_3^{(i)}am^{(i)}+\theta_4'c$$

$$E\left[M^{(i)}=1|a,c\right]=\beta_0^{(i)}+\beta_1^{(i)}a+\beta_2^{(i)'}c \text{ for } i=1,\ldots,K.$$

with the vector of mediators M following a multivariate normal distribution conditional on A and C with conditional covariance matrix Σ. Under assumptions [1]–[4] we then have $\mathrm{logit}\{P(Y_{a\mathbf{M}_{a^*}}=1|c)\}$

$$\approx\log\left\{P\left(Y_{a\mathbf{M}_{a^*}}=1|c\right)\right\}$$

$$= \log\left\{\int P(Y=1|a,\mathbf{m},c)dP(\mathbf{m}|a^*,c)\right\} \text{ by } [1]-[4]$$
$$\approx \log\left\{\int \exp(\theta_0+\theta_1 a+\sum_{i=1}^K \theta_2^{(i)}m^{(i)}+\sum_{i=1}^K \theta_3^{(i)}am^{(i)}+\theta_4'c)dP(\mathbf{m}|a^*,c)\right\}$$
$$= \log\left\{\exp(\theta_0+\theta_1 a+\theta_4'c)\int \exp\left\{\sum_{i=1}^K \theta_2^{(i)}m^{(i)}+\theta_3^{(i)}am^{(i)}\right\}dP(\mathbf{m}|a^*,c)\right\}$$
$$= \log\left\{\exp(\theta_0+\theta_1 a+\theta_4'c)E\left[e^{\sum_{i=1}^K(\theta_2^{(i)}+\theta_3^{(i)}a)M^{(i)}}|a^*,c\right]\right\}$$

Note that conditional on a^* and c, $\sum_{i=1}^K(\theta_2^{(i)}+\theta_3^{(i)}a)M^{(i)}$ follows a normal distribution with mean $\sum_{i=1}^K(\theta_2^{(i)}+\theta_3^{(i)}a)(\beta_0^{(i)}+\beta_1^{(i)}a+\beta_2^{(i)'}c)$ and variance $(\theta_2+\theta_3a)'\Sigma(\theta_2+\theta_3a)$, where $\theta_2=(\theta_2^{(1)},\ldots,\theta_2^{(K)})'$ and $\theta_3=(\theta_3^{(1)},\ldots,\theta_3^{(K)})'$. It thus follows that $\text{logit}\{P(Y_{a\mathbf{M}_{a^*}}=1|c)\}$

$$= \log\left\{\exp(\theta_0+\theta_1 a+\theta_4'c)\exp\left(\sum_{i=1}^K(\theta_2^{(i)}+\theta_3^{(i)}a)(\beta_0^{(i)}+\beta_1^{(i)}a^*+\beta_2^{(i)'}c)\right.\right.$$
$$\left.\left.+\frac{1}{2}(\theta_2+\theta_3a)'\Sigma(\theta_2+\theta_3a)\right)\right\}$$
$$= \theta_0+\theta_1 a+\theta_4'c+\sum_{i=1}^K(\theta_2^{(i)}+\theta_3^{(i)}a)\left(\beta_0^{(i)}+\beta_1^{(i)}a^*+\beta_2^{(i)'}c\right)+\frac{1}{2}(\theta_2+\theta_3a)'\Sigma(\theta_2+\theta_3a).$$

The log of natural direct effect odds ratio is then given by:

$$\log\left\{OR^{NDE}_{a,a^*|c}(a^*)\right\}$$
$$= \text{logit}\left\{P\left(Y_{aM_{a^*}}=1|c\right)\right\}-\text{logit}\left\{P\left(Y_{a^*M_{a^*}}=1|c\right)\right\}$$
$$\approx \theta_0+\theta_1 a+\theta_4'c+\sum_{i=1}^K(\theta_2^{(i)}+\theta_3^{(i)}a)\left(\beta_0^{(i)}+\beta_1^{(i)}a^*+\beta_2^{(i)'}c\right)+\frac{1}{2}(\theta_2+\theta_3a)'\Sigma(\theta_2+\theta_3a).$$
$$-\left\{\theta_0+\theta_1 a^*+\theta_4'c+\sum_{i=1}^K(\theta_2^{(i)}+\theta_3^{(i)}a^*)(\beta_0^{(i)}+\beta_1^{(i)}a^*+\beta_2^{(i)'}c)+\frac{1}{2}(\theta_2+\theta_3a^*)'\Sigma(\theta_2+\theta_3a^*)\right\}$$
$$= \left\{\theta_1+\sum_{i=1}^K\theta_3^{(i)}\left(\beta_0^{(i)}+\beta_1^{(i)}a^*+\beta_2^{(i)'}c+\theta_2^{(i)}\sigma^{(i)2}\right)\right\}(a-a^*)$$
$$+0.5(\theta_2+\theta_3a)'\Sigma(\theta_2+\theta_3a)-0.5(\theta_2+\theta_3a^*)'\Sigma(\theta_2+\theta_3a^*).$$

The log of natural indirect effect odds ratio is then given by:

$$\log\left\{OR^{NIE}_{a,a^*|c}(a)\right\}$$
$$= \text{logit}\{P(Y_{a\mathbf{M}_a}=1|c)\}-\text{logit}\left\{P\left(Y_{a\mathbf{M}_{a^*}}=1|c\right)\right\}$$
$$\approx \theta_0+\theta_1 a+\theta_4'c+\sum_{i=1}^K\left(\theta_2^{(i)}+\theta_3^{(i)}a\right)\left(\beta_0^{(i)}+\beta_1^{(i)}a+\beta_2^{(i)'}c\right)+\frac{1}{2}(\theta_2+\theta_3a)'\Sigma(\theta_2+\theta_3a)$$
$$-\left[\theta_0+\theta_1 a+\theta_4'c+\sum_{i=1}^K\left(\theta_2^{(i)}+\theta_3^{(i)}a\right)\left(\beta_0^{(i)}+\beta_1^{(i)}a^*+\beta_2^{(i)'}c\right)+\frac{1}{2}(\theta_2+\theta_3a)'\Sigma(\theta_2+\theta_3a)\right]$$
$$= \sum_{i=1}^K\left(\theta_2^{(i)}+\theta_3^{(i)}a\right)\beta_1^{(i)}(a-a^*).$$

Sequential approach

Let $\overline{\mathbf{M}}^{(k)}$ be the subset of mediators $(M^{(1)}, \ldots, M^{(k)})$. Consider the regression models:

$$E\left[Y|a, \bar{\mathbf{m}}^{(k)}, c\right] = \theta_0 + \theta_1 a + \sum_{i=1}^{k} \theta_2^{(i)} m^{(i)} + \sum_{i=1}^{k} \theta_3^{(i)} a m^{(i)} + \theta_4' c$$

$$E\left[M^{(i)}|a, c\right] = \beta_0^{(i)} + \beta_1^{(i)} a + \beta_2^{(i)'} c$$

for $i = 1, \ldots, k$. Under assumptions [1]–[4], we then have by Pearl's mediation formula that the exposure effect that is mediated by the first k mediators equals:

$$E\left[Y_{a\overline{\mathbf{M}}_a^{(k)}} - Y_{a\overline{\mathbf{M}}_{a^*}^{(k)}}|c\right] = \sum_{i=1}^{k} \left\{\theta_2^{(i)} + \theta_3^{(i)} a\right\} \left\{E[M^{(i)}|c, a] - E[M^{(i)}|c, a^*]\right\}$$

$$\sum_{i=1}^{k} \left\{\theta_2^{(i)} + \theta_3^{(i)} a\right\} \beta_1^{(i)} (a - a^*)$$

While this approach is valid for each fixed k, a concern is that the models for $E[Y|a, \bar{\mathbf{m}}^{(k)}, c]$ and $E[M^{(k)}|a, c]$ may not be compatible across k. To illustrate the difficulty of correct specification of these models across k, suppose that

$$E\left[M^{(i)}|\bar{\mathbf{m}}^{(i-1)}, a, c\right] = \gamma_0^{(i)} + \gamma_1^{(i)} a + \gamma_2^{(i)'} c + \gamma_3^{(i)'} \bar{\mathbf{m}}^{(i-1)},$$

for $i = 1, \ldots, k$, which is compatible with the aforementioned models for $E[M^{(i)}|a, c]$. Then the model for $E[Y|a, \bar{\mathbf{m}}^{(k)}, c]$ implies that

$$E\left[Y|a, \bar{\mathbf{m}}^{(k-1)}, c\right] = \theta_0 + \theta_1 a + \sum_{i=1}^{k-1} \theta_2^{(i)} m^{(i)} + \sum_{i=1}^{k-1} \theta_3^{(i)} a m^{(i)} + \theta_4' c$$

$$+ \left(\theta_2^{(k)} + \theta_3^{(k)} a\right) \left(\gamma_0^{(i)} + \gamma_1^{(i)} a + \gamma_2^{(i)'} c + \gamma_3^{(i)'} \bar{\mathbf{m}}^{(i-1)}\right)$$

This model is no longer of the same form as it includes interactions between a and c, as well as squared terms a^2 that were not previously allowed for. This can be remedied by extending the outcome regression model to include such terms:

$$E\left[Y|a, \bar{\mathbf{m}}^{(k)}, c\right] = \theta_0 + \theta_1 a + \sum_{i=1}^{k} \theta_2^{(i)} m^{(i)} + \sum_{i=1}^{k} \theta_3^{(i)} a m^{(i)} + \theta_4' c + \theta_5' ac + \theta_6' a^2.$$

With this extension one still has

$$E\left[Y_{a\overline{\mathbf{M}}_a^{(k)}} - Y_{a\overline{\mathbf{M}}_{a^*}^{(k)}}|c\right] = \sum_{i=1}^{k} \left\{\theta_2^{(i)} + \theta_3^{(i)} a\right\} \beta_1^{(i)} (a - a^*).$$

If the exposure is binary and we allow for exposure–covariate interaction in the outcome model then this particular problem of correct model specification is thus remedied. Alternatively, if there are no exposure–mediator interactions in the outcome model then the models will remain compatible with each other.

Weighting approach

The weighting approach is based on the following identity:

$$E\left[Y_{a\mathbf{M}_{a^*}}\right] = \int E(Y|a, \mathbf{m}, c) f(\mathbf{M} = \mathbf{m}|a^*, c) f(c) d\mathbf{m} dc$$

$$= \int I(A = a^*)E(Y|a, \mathbf{M}, c)f(\mathbf{M}|A, c)f(c)d\mathbf{M}dAdc$$

$$= \int \frac{I(A = a^*)}{P(A = a^*|c)}E(Y|a, \mathbf{M}, c)f(\mathbf{M}, A, c)d\mathbf{M}dAdc$$

$$= E\left[\frac{I(A = a^*)}{P(A = a^*|c)}E(Y|a, \mathbf{M}, c)\right]$$

$$= E\left[\frac{P(A = a^*)}{P(A = a^*|c)}E(Y|a, \mathbf{M}, c)|A = a^*\right].$$

For the sequential approach, one likewise obtains that

$$E\left[Y_{a\overline{\mathbf{M}}_{a^*}^{(k)}}\right] = E\left[\frac{I(A = a^*)}{P(A = a^*|c)}E(Y|a, \overline{\mathbf{M}}^{(k)}, c)\right].$$

Applying the sequential approach thus demands models for the conditional expectation $E(Y|a, \overline{\mathbf{M}}^{(k)}, c)$; a possible concern is then that these models for different k may not be compatible with each other.

Sum of individual mediated effects versus joint mediated effects

Consider two mediators, $M^{(1)}$ and $M^{(2)}$, and suppose that neither affects the other. For simplicity assume a binary exposure. The natural indirect effect through $M^{(1)}$ is by definition $Y_{1M_1^{(1)}} - Y_{1M_0^{(1)}}$. The natural indirect effect through $M^{(2)}$ is by definition $Y_{1M_1^{(2)}} - Y_{1M_0^{(2)}}$. The natural indirect effect through $(M^{(1)}M^{(2)})$, is by definition $Y_{1M_1^{(1)}M_1^{(2)}} - Y_{1M_0^{(1)}M_0^{(2)}}$. If the mediators do not affect each other then the natural indirect effect through $M^{(1)}$ is equal to $Y_{1M_1^{(1)}} - Y_{1M_0^{(1)}} = Y_{1M_1^{(1)}M_1^{(2)}} - Y_{1M_0^{(1)}M_1^{(2)}}$ and the natural indirect effect through $M^{(2)}$ is equal to $Y_{1M_1^{(1)}M_1^{(2)}} - Y_{1M_1^{(1)}M_0^{(2)}}$. The sum of the two natural indirect effects for $M^{(1)}$ and $M^{(2)}$ considered separately is thus $\left\{Y_{1M_1^{(1)}M_1^{(2)}} - Y_{1M_0^{(1)}M_1^{(2)}}\right\} + \left\{Y_{1M_1^{(1)}M_1^{(2)}} - Y_{1M_1^{(1)}M_0^{(2)}}\right\}$. The difference between the sum of the two natural indirect effects for $M^{(1)}$ and $M^{(2)}$ considered separately and the natural indirect effect through $(M^{(1)}M^{(2)})$ jointly is then given by:

$$\left\{Y_{1M_1^{(1)}M_1^{(2)}} - Y_{1M_0^{(1)}M_1^{(2)}}\right\} + \left\{Y_{1M_1^{(1)}M_1^{(2)}} - Y_{1M_1^{(1)}M_0^{(2)}}\right\} - \left\{Y_{1M_1^{(1)}M_1^{(2)}} - Y_{1M_0^{(1)}M_0^{(2)}}\right\}$$

$$= Y_{1M_1^{(1)}M_1^{(2)}} - Y_{1M_0^{(1)}M_1^{(2)}} - Y_{1M_1^{(1)}M_0^{(2)}} + Y_{1M_0^{(1)}M_0^{(2)}}.$$

This difference in some sense captures the effect mediated by the interaction between $M^{(1)}$ and $M^{(2)}$.

We now show that if the mediators do not affect each other and if there is no interaction between $M^{(1)}$ and $M^{(2)}$ at the individual counterfactual level then the sum of the two natural indirect effects for $M^{(1)}$ and $M^{(2)}$ considered separately and the natural indirect effect through $(M^{(1)}M^{(2)})$ jointly must be equal. We will say that there is no interaction between $M^{(1)}$ and $M^{(2)}$ at the individual counterfactual level if for any a and any two values $m^{(1)}, m^{(1)^\dagger}$ of $M^{(1)}$, $Y_{am^{(1)}m^{(2)}} - Y_{am^{(1)^\dagger}m^{(2)}}$ is constant across $m^{(2)}$ or, equivalently, for any a and any two values $m^{(2)}, m^{(2)^\dagger}$ of $M^{(2)}$, $Y_{am^{(1)}m^{(2)}} - Y_{am^{(1)}m^{(2)^\dagger}}$ is constant across $m^{(1)}$. If this is the case then $Y_{1M_1^{(1)}m^{(2)}} - Y_{1M_0^{(1)}m^{(2)}}$ must be constant across $m^{(2)}$ and thus $Y_{1M_1^{(1)}M_1^{(2)}} - Y_{1M_0^{(1)}M_1^{(2)}}$ must be equal to $Y_{1M_1^{(1)}M_0^{(2)}} - Y_{1M_0^{(1)}M_0^{(2)}}$ and so $Y_{1M_1^{(1)}M_1^{(2)}} - Y_{1M_0^{(1)}M_1^{(2)}} - Y_{1M_1^{(1)}M_0^{(2)}} + Y_{1M_0^{(1)}M_0^{(2)}}$ must be equal to 0.

We now also show a similar result with linear models on the additive scale with interaction on the average level of population rather than the individual level. The approach that uses one mediator at a time, uses regression models for $E[Y|a, m^{(i)}, c]$, which are of the form

$$E[Y|a,m^{(i)},c] = \theta_0 + \theta_1 a + \theta_2^{(i)} m^{(i)} + \theta_3^{(i)} a m^{(i)} + \tau^{(ii)} m^{(i)} m^{(i)}$$
$$+ \sum_{j\neq i} \tau^{(ij)} m^{(i)} E\left(M^{(j)}|a,m^{(i)},c\right) + \sum_{j\neq i} \theta_2^{(j)} E\left(M^{(j)}|a,m^{(i)},c\right)$$
$$+ \sum_{j\neq i} \theta_3^{(j)} a E\left(M^{(j)}|a,m^{(i)},c\right) + \sum_{j\neq i,k\neq i} \tau^{(jk)} E\left(M^{(j)}M^{(k)}|a,m^{(i)},c\right) + \theta_4' c.$$

Under this model, the natural indirect effect is given by:

$$E\left[Y_{aM_a^{(i)}} - Y_{aM_{a^*}^{(i)}}|c\right]$$
$$= \left\{\theta_2^{(i)} + \theta_3^{(i)} a\right\}\left\{E\left[M^{(i)}|c,a\right] - E\left[M^{(i)}|c,a^*\right]\right\} + \tau^{(ii)}\left\{E\left[M^{(i)}M^{(i)}|c,a\right] - E\left[M^{(i)}M^{(i)}|c,a^*\right]\right\}$$
$$+ \sum_{j\neq i} \tau^{(ij)} E\left(M^{(i)}M^{(j)}|a,c\right) - E\left\{m^{(i)} E\left(M^{(j)}|a,m^{(i)},c\right)|a^*,c\right\}$$
$$+ \sum_{j\neq i}\left\{\theta_2^{(j)} + \theta_3^{(j)} a\right\}\left[E\left(M^{(j)}|a,c\right) - E\left\{E\left(M^{(j)}|a,m^{(i)},c\right)|a,c\right\}\right]$$
$$+ \sum_{j\neq i,k\neq i} \tau^{(jk)} E\left(M^{(j)}M^{(k)}|a,c\right) - E\left\{E\left(M^{(j)}M^{(k)}|a,m^{(i)},c\right)|a^*,c\right\}.$$

When $M^{(i)}$ is independent of all other mediators (conditional on a and c), then this reduces to

$$\left\{\theta_2^{(i)} + \theta_3^{(i)} a\right\}\left\{E\left[M^{(i)}|c,a\right] - E\left[M^{(i)}|c,a^*\right]\right\} + \tau^{(ii)}\left\{E\left[M^{(i)}M^{(i)}|c,a\right] - E\left[M^{(i)}M^{(i)}|c,a^*\right]\right\}$$
$$+ \sum_{j\neq i} \tau^{(ij)} E\left[M^{(j)}|c,a\right]\left\{E\left[M^{(i)}|c,a\right] - E\left[M^{(i)}|c,a^*\right]\right\}.$$

It follows from this that the natural indirect effect $E[Y_{a\mathbf{M}_a} - Y_{a\mathbf{M}_{a^*}}|c]$ equals the sum of the individual natural indirect effects when either none of the mediators is affected by the exposure, or none of the mediators affects the outcome, or the mediators are mutually independent (conditional on a and c) and there are no mediator–mediator interactions, but not generally otherwise.

SAS implementation

We describe how the proposed weighting approach given above can be implemented in SAS statistical software (SAS Institute, Inc., Cary, North Carolina). Below we let `c`, `a`, `m` and `y` correspond to the observed confounders C, exposure A, mediator M and outcome Y, and assume, for the illustration, that A and Y are dichotomous.

```
proc logistic data = mydata;
model a = c;
score data = mydata out = preda;
run;
data preda;
set preda;
pa1 = P_1;
run;
data mydata0;
set mydata;
```

```
a = 0; output;
run;
data mydata1;
set mydata;
a = 1; output;
run;
proc logistic data = mydata;
model y = a m c;
score data = mydata0 out = predy0;
score data = mydata1 out = predy1;
run;
data predy0;
set predy0;
py0 = P_1;
run;
data predy1;
set predy1;
py1 = P_1;
run;
data mydataw;
merge preda predy0 predy1 mydata;
run;
data mydataw;
set mydataw;
w = a/pa1 + (1-a)/(1-pa1);
run;
```

The mean $E[Y_{1\mathbf{M}_0}]$ (except for standard errors) can now be estimated using:

```
proc reg data = mydataw;
where a = 0;
model py1 = ;
weight w;
run;
```

and the mean $E[Y_{0\mathbf{M}_1}]$ (except for standard errors) using:

```
proc reg data = mydataw;
where a = 1;
model py0 = ;
weight w;
run;
```

References

Albert, J. M. (2012). Distribution-free mediation analysis for nonlinear models with confounding. Epidemiology, 23:879–888.

Albert, J. M. and Nelson, S. (2011). Generalized causal mediation analysis. Biometrics, 67:1028–1038.

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

Baron, R. M. and Kenny, D. A. (1986). The moderator-mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. Journal of Personality and Social Psychology, 51:1173–1182.

Imai, K., Keele, L., and Tingley, D. (2010). A general approach to causal mediation analysis. Psychological Methods, 15:309–334.

Imai, K. and Yamamoto, T. (2012). Identification and sensitivity analysis for multiple causal mechanisms: revisiting evidence from framing experiments. Political Analysis, 21:141–171.

Kotelchuck, M. (1994). An evaluation of the Kessner adequacy of prenatal care index and a proposed adequacy of prenatal care utilization index. American Journal of Public Health, 84:1414–1420.

Pearl, J. (2001). Direct and indirect effects. In: Proceedings of the Seventeenth Conference on Uncertainty and Artificial Intelligence. San Francisco: Morgan Kaufmann, 411–420.

Pearl, J. (2009). Causality: Models, Reasoning, and Inference. 2nd Edition. Cambridge: Cambridge University Press.

Politis, D. N. and Romano, J. P. (1994). Large sample confidence regions based on subsamples under minimal assumptions. Annals of Statistics, 22:2031–2050.

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

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

Tchetgen Tchetgen, E. J. and Shpitser, I. (2012). Semiparametric theory for causal mediation analysis: Efficiency bounds, multiple robustness, and sensitivity analysis. Annals of Statistics, 40(3):1816–1845.

Valeri, L. and VanderWeele, T. J. (2013). Mediation analysis allowing for exposure-mediator interactions and causal interpretation: Theoretical assumptions and implementation with SAS and SPSS macros. Psychological Methods, 18:137–150.

van der Laan, M. J. and Petersen, M. L. (2008). Direct effect models. International Journal of Biostatistics, 4:Article 23.

VanderWeele, T. J. (2010). Bias formulas for sensitivity analysis for direct and indirect effects. Epidemiology, 21:540–551.

VanderWeele, T. J., Lantos, J. D., Siddique, J., and Lauderdale, D. S. (2009). A comparison of four prenatal care indices in birth outcome models: Comparable results for predicting small-for-gestational-age outcome but different results for preterm birth or infant mortality. Journal of Clinical Epidemiology, 62:438–445.

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

VanderWeele, T. J. and Vansteelandt, S. (2010). Odds ratios for mediation analysis for a dichotomous outcome. American Journal of Epidemiology, 172:1339–1348.

VanderWeele T. J., Vansteelandt, S., and Robins, J. M. (2013). Effect decomposition in the presence of an exposure-induced mediator-outcome confounder. Epidemiology, in press.

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

Zheng, W. and van der Laan, J. (2012). Targeted maximum likelihood estimation of natural direct effects. International Journal of Biostatistics, 8:Article 1.