Effect decomposition through multiple causally non-ordered mediators in the presence of exposure-induced mediator-outcome confounding

Murthy N. Mittinty[1], John W. Lynch[1,2], Andrew B. Forbes[3] and Lyle C. Gurrin[4]

1 School of Public Health, University of Adelaide;
2 School of Social and Community Medicine, University of Bristol;
3 School of Population Health and Preventive Medicine Monash University;
4 School of Population and Global Health, University of Melbourne.

May 22, 2019

Abstract

Avin *et al.* (2005) showed that, in the presence of exposure-induced mediator-outcome confounding, decomposing the total causal effect (TCE) using standard conditional exchangeability assumptions is not possible even under a non-parametric structural equation model with all confounders observed. Subsequent research has investigated the assumptions required for such a decomposition to be identifiable and estimable from observed data. One approach was proposed by [1]. They decomposed the TCE under three different scenarios: (1) treating the mediator and the exposure-induced confounder as joint mediators; (2) generating path specific effects albeit without distinguishing between multiple distinct paths through the exposure-induced confounder; and (3) using so-called randomised interventional analogues where sampling values from the distribution of the mediator within the levels of the exposure effectively marginalises over the exposure-induced confounder. In this paper we extend their approach to the case where there are multiple mediators that do not influence each other directly but which are all

influenced by an exposure-induced mediator-outcome confounder. We provide a motivating example and results from a simulation study based on from our work in dental epidemiology featuring the 1982 Pelotas Birth Cohort in Brazil.

1 Introduction

Intervention researchers are often interested in decomposing the total causal effect of an exposure on an outcome into a series of path-specific effects that act through intermediate variables. A major advantage of the potential outcomes approach to mediation analysis is that it allows for the decomposition of the total causal effect of an exposure on an outcome into the direct effect of that exposure and the indirect effect that goes through an exposure-influenced mediator. Methods for such an effect decomposition analysis are also available for situations where the associations between exposure, mediator and outcome are non-linear or interactions are present. If, however, there is an intermediate variable that is affected by the exposure and that in turn confounds the mediator-outcome relationship then the direct and the indirect effects are not identifiable using standard conditional exchangeability assumptions and existing methods [2, 1, 3].

For such a situation methods have been developed under a number of different scenarios. One approach is to group variables by either (1) treating the mediator-outcome confounder and the exposure as joint exposures [4, 5, 6] (a scenario we do not consider further in this paper); or (2) treating the mediator-outcome confounder induced by the exposure and the mediator as joint mediators [1]. More generally, the problem of multiple mediators that cannot be assumed to be independent even given a set of covariates has been addressed using distinct definitions of the direct and indirect effects, differing sets of assumptions and for a range of study designs and outcome measures [7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18]. In [15] the mediation effect is defined as a rate of change so that the estimated effect is invariant to the scale and/or the difference in the magnitude of levels of the exposure. Tingley and colleagues [13] advocate the use of a parallel design requiring random sampling from the joint distribution of the exposure and the mediators.

The approaches to the decomposition of the total causal effect in the presence of exposure-induced multiple mediator-outcome confounding that we explore here and, in some instances, extend are those proposed by VanderWeele, Vansteelandt and Robins (2014) [1] (henceforth VVR14). Apart from their joint mediators and path-specific approaches, VVR14 introduced so-called randomized interventional analogue (in)direct effects. These can be identified under much weaker conditions than natural (in)direct effects, but sum to a total interventional causal effect, not the total causal effect. Recent work by Vansteelandt and Daniel

(2017) [11] (henceforth VD17) also uses randomised interventional analogues, although these are defined differently to those presented by VVR14. VD17 assume no exposure-induced mediator-outcome confounding which, along with the so-called "cross-world" assumption, allows them to decompose the total causal effect into randomised interventional analogues of direct and indirect effects. This approach identifies the components of the effect of an exposure on an outcome attributable each of a series of mediators even when the structural dependence between the multiple mediators is unknown. A detailed explanation of the distinction between the randomised interventional analogues of VVR14 and VD17 is provided by Moreno Betancur and Carlin in [19]. Taguri *et al*. (2018) [20] present three methods for multiple mediation analysis, the first essentially the same as our joint mediator approach, and the second and third similar but not identical to our approach to path-specific effects. They do not, however, extend (as we do) their methods to the use of randomised interventional analogues.

The setting for multiple mediators in which we are interested is different to that considered in VD17. We differentiate exposure-induced mediator-outcome confounders that are of not of interest from the mediators that we wish to target in a decomposition of the total causal effect. Provided the exposure-induced confounders can be handled appropriately in the effect decomposition we are not concerned that the mediated effects transferred through their descendants of cannot be individually identified. In addition, VD17 require the specification of regression models for the outcome at that capture the effects of covariates and any interactions between the exposure and mediators. In contrast, the propensity score-based weighting strategies proposed in this paper seek to deliver the same distribution of mediators across exposure (or treatment) groups. This allows us to compute the exposure effect on the outcome using the empirical relationship between the outcome and exposure, simplifying the specification of the outcome model; for further explanation, see [21]. We provide the details of implementing all three methods of VVR14 to multiple mediators including programming code for weighting-based regression approaches to estimation. The identification results that we present are nonparametric in the sense that they can be applied to any type of outcome (continuously-valued, binary, categorical or count) and the generalised linear models that are typically used to analyse them.

Recently [22] and [23] extended the analysis of single time point data structures with exposure-induced mediator-outcome confounding (represented by the causal diagram in Figure 3) to the scenario where the exposure, confounders, mediators and the outcome are measured repeatedly over multiple time points in a longitudinal setting. The approach that we propose can, in one sense, be seen as a special case of their scenario where the mediators are repeated measures of the same mediator and there is a single (i.e. not time-varying) exposure-induced mediator-outcome confounder. It differs from the approaches presented in [22]

and [23] in that we (1) are considering an arbitrary set of distinct mediators, not repeated measures ordered in time of a single univariate mediator; and (2) have spelt out a decomposition of the total causal effect in terms of natural indirect effects for each of the mediators [22] does not explore a decomposition of the total causal effect into time point-specific contributions of the mediator. In further recent work, Albert *et al*. (2018) [24] consider arbitrary groups of possibly inter-related mediators sequential in time, although they preference particular types of contrasts in estimation (*eg* transitions from low to high socio-economic status) to overcome to overcome problems with the identifiability of causal effects. Several other papers have sought to extend these methods for longitudinal mediation analysis to incorporate time-to-event in a survival setting see [25, 26, 27, 28, 29, 30, 31, 32, 33].

The structure of the paper is as follows. In Section 2 we provide a motivating example from our work in dental epidemiology. Section 3 introduces notation and reviews effect decomposition and the difficulty of providing such a decomposition in the presence of exposure-induced mediator-outcome confounding. Section 4 describes the three approaches of VVR14 and the extension we propose to the analysis of path-specific effects. Section 5 provides methods of estimation based on weighted regression. Section 6 is a discussion.

2 Motivating example: Socio-economic position and the risk of periodontitis

Here we present an example where the data structure suggests that causal mediation analysis will require us to accommodate exposure-induced confounding, and which motivates the subsequent methodological developments. This example application concerns the work of the first author in life-course epidemiology that seeks to explain an established disparity between childhood disadvantage and oral health in adult life: Low childhood socio-economic position (SEP) increases the risk of dental disease in adult life that appears to be independent of adult risk factors and low adult SEP, suggesting that socially patterned behaviours in childhood have a lasting influence on adult health [34, 35].

We explore this problem using information from the 1982 Pelotas Birth Cohort in Brazil. In 1997 a sample of 900 (from the total of 5,914) participants were randomly selected for an oral health study. During periodic oral examinations participants were examined for the presence of periodontal disease (bleeding on probing and assessment of probing depth and gingival margin level to identify clinical attachment loss, dental caries and other dental and oral conditions) which we can produce a severity index for periodontitis, an infectious inflammatory condition that affects the supporting tissue of the teeth. Two binary exposure variables mea-

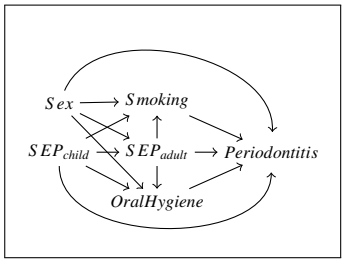

Figure 1: Causal Diagram: For mediation analysis of low SEP in childhood and the severity of periodontitis example. Low SEP in adult is an EIMOC of the association between smoking and poor oral hygiene and periodontitis

sured concurrently include the participants smoking status (current- or ex-smoker versus non-smoker), poor oral hygiene habits (frequency of brushing teeth: less than daily versus at least daily). Current, adult SEP is, along with low childhood SEP, available from information collected at baseline and early follow-up. Summary statistics for the cohort are presented in Table 1. We are interested in the extent to which the total causal effect of childhood SEP on the risk of periodontal disease in adult life is mediated through two important determinants of oral health, smoking and oral hygiene habits.

The directed acyclic graph (DAG) in Figure 1 was developed by the investigators to represent the causal structure of the problem. In this diagram low adult SEP is an exposure-induced confounder of the association between smoking and the risk of periodontal disease, and similarly for poor oral hygiene habits. We acknowledge that there may be unmeasured factors in adult life that continue to confound the associations between the mediators smoking and poor oral hygiene habits, and the outcome periodontal disease. We argue that a quality measure of adult SEP is likely to be closely correlated with these unmeasured factors (e.g. a mis-measured version of these factors) and that smoking and oral hygiene habits will be approximately conditional independent given adult SEP. Consequently, the common assumption that we rely on during statistical modelling, namely that there is no unmeasured confounding of the mediators and the outcome, will be approximately true.

We return to this example in Section 6 where we present the results of a simulation study using datasets that reflect the exposure patterns that were observed in the 1982 Pelotas Birth Cohort oral health study, and represent a substantive analysis of the full dataset from the same cohort study.

Characteristic	Sample (%)
Sex	
Male	3,037 (51.4)
Female	2,876 (48.6)
Socio-economic position at birth	
Middle or highest tertile	3,951 (66.8)
Poorest	1,963 (33.2)
Socio-economic position at age 23 years	
Middle or highest tertile	2,874 (66.9)
Poorest	1,422 (33.1)
Smoking status	
Current smoker	1,103 (25.7)
Non- or ex-smoker	3,193 (74.3)
Oral Hygiene (frequency of brushing)	
0-2 times a day	195 (27.4)
3+ times	517 (72.6)
Periodontitis	
Healthy	338 (62.7)
Any Periodontal disease	201 (37.3)
Periodontitis (Number of teeth with probing depth or gingival margin > 3mm)	
Sample size	535
Mean	3.36
Standard deviation	4.59
Maternal education	
0-8	4,413 (74.7)
9+years	1,494 (25.3)

Table 1: A statistical summary of demographic characteristics for the Pelotas 1982 Birth Cohort.

3 Effect decomposition

3.1 Notation, definitions and assumptions

We start with a simple directed acyclic graph (DAG) as shown in Figure 2, which acts as a causal diagram for the associations between the variables C, A, L, M and Y which we define below. We keep the notation as similar as possible to that used in [1] for the purpose of comparison and so that our extensions may be readily understood. Let A denote the exposure, Y the outcome, M the mediator, C a set of confounders and L, a potentially multivariate exposure-induced mediator-outcome confounder. Let Y_a and M_a denote the values of the potential outcome and potential mediator that would have been observed had the exposure A been set to level a. The total causal effect is expressed as the expected difference in the potential outcome Y_a where A is set to an observed level a and Y_{a^*} where A is set to counterfactual value a^*, denoted as $E[Y_a - Y_{a^*}]$. For subsequent mediation analysis, following Figure 2, let Y_{am} be the value of the potential outcome that would have been observed had A been set to a level a and M set to level m. The implied nesting structure is formalised by the composition assumption which is that $Y_a = Y_{aM_a} = Y(A = a, M(A = a))$ [36, 37].

For the simple mediation example derived from Figure 2, the total causal effect (TCE) $E[Y_a - Y_{a^*}]$ can be decomposed into the sum of natural direct and indirect effects as follows:

$$E[Y_a - Y_{a^*}] = E[Y_{aM_a} - Y_{a^*M_{a^*}}] = E[Y_{aM_a} - Y_{aM_{a^*}}] + E[Y_{aM_{a^*}} - Y_{a^*M_{a^*}}].$$

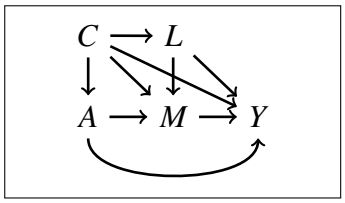

Figure 2: Directed acyclic graph of simple mediation with covariates.

This decomposition holds even when there are interaction terms between the exposure and mediator in the model for the outcome Y, and also when the relationships between the outcome and exposure or mediator and exposure are non-linear. For the direct and the indirect effects to be estimable from the data we require four assumptions: (i) the effect of the exposure A on the potential outcome Y_{am} is unconfounded conditional on C ($Y_{am} \perp\!\!\!\perp A \mid C$); (ii) the effect of the mediator M on the potential outcome Y_{am} is unconfounded conditional on C and A ($Y_{am} \perp\!\!\!\perp M \mid A = a, C$ for $a = 0, 1$); (iii) the effect of the exposure A on the mediator

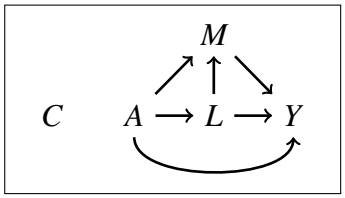

Figure 3: Mediation in the presence of exposure-induced mediator-outcome confounder. The global confounder C has hidden arrows to all nodes in the graph so once again all statements about independence are conditional on C.

M is unconfounded conditional on C ($M_a \perp\!\!\!\perp A \mid C$); and (iv) ["cross-world"] none of the mediator-outcome confounders are themselves affected by the exposure ($Y_{am} \perp\!\!\!\perp M_{a^*} \mid C$) where in each case $Y \perp\!\!\!\perp A \mid C$ is to be read as "Y is conditionally independent of A given C". We note that the conditional independence statement in (iv) only corresponds to the condition that no mediator-outcome confounders be affected by exposure under the additional assumption that the data (and counterfactual data) follow a Nonparametric Structural Equation Model (NPSEM). If assumptions (i)-(iv) hold, and the correct DAG representing the causal relationships between variables C, A, L, M and Y is Figure 2, then the empirical expressions for the direct and the indirect effects are:

$$E[Y_{aM_a^*} - Y_{a^*M_{a^*}}] = \sum_{c,m} \{E[Y \mid a,m,c] - E[Y \mid a^*,m,c]\} P(m \mid a^*,c) P(c) \qquad (1)$$

$$E[Y_{aM_a} - Y_{aM_{a^*}}] = \sum_{c,m} E[Y \mid a,m,c] \{P(m \mid a,c) - P(m \mid a^*,c)\} P(c) \qquad (2)$$

3.2 Exposure-induced mediator-outcome confounding with one mediator

We now consider the situation where there is an additional arrow in the causal diagram from A to L (see Figure 3) so that the exposure A induces the mediator-outcome confounder L. For such a situation, Avin *et al.* (2005) [2] showed that assumption (iv) is violated in the presence of exposure-induced mediator-outcome confounding even when using the nonparametric structural equation approach. This is sometimes referred to as the problem of the "recanting witness" who, when giving evidence on different occasions, tells apparently conflicting stories that are nevertheless consistent with the final conclusion. Before proceeding we need to extend the earlier notation to allow for three subscript indices, once again in a style that is consistent with VVR14. Let L_a be the value of the potential outcome that would have been observed had A been set to a level a, M_{al} be the value of the potential outcome that would have been observed had A been set to a level a

and L set to level l and Y_{alm} be the value of the potential outcome that would have been observed had A been set to a level a, L set to level l and M set to level m. The composition assumption implies that $M_a = M_{aL_a} = M(A = a, L(A = a))$ and, by iterative application, $Y_a = Y_{aM_a} = Y_{aL_aM_{aL_a}} = Y(A = a, L(A = a), M(A = a, L(A = a)))$. Thus multiple subscript indices imply, under the composition assumption, a potentially complex structure of nested functions that define L, M and, ultimately, Y.

If one is interested in estimating the total causal effect, $E[Y(A = a) - Y(A = a^*)]$, then this complex nesting structure does not pose a problem. Decomposing the total effect into the sum of direct and indirect effects in this situation is, however, more difficult since it relies on identification (and hence estimation using the observed data) of these effects, which requires additional assumptions. One possible way to proceed is to consider the following derivation for an expression of the average potential outcome $E[Y(A = a, M(A = a^*))]$ where all steps that assume independence are conditional on covariates C:

$$
\begin{aligned}
& E[Y(A=a, M(A=a^*))] \\
= & \sum_{l,m} E[Y(a,l,m) \mid L(A=a)=l, M(A=a^*)=m] \times f(L(A=a)=l, M(A=a^*)=m) \\
= & \sum_{l,l^*,m} E[Y(a,l,m) \mid L(A=a)=l, M(A=a^*, L(A=a^*)=l^*)=m, L(A=a^*)=l^*] \times \\
& f(L(A=a)=l, M(A=a^*, L(A=a^*)=l^*)=m, L(A=a^*)=l^*) \\
& \text{Law of Total Probability} \\
= & \sum_{l,l^*,m} E[Y(a,l,m) \mid \ldots] f(L(A=a)=l, L(A=a^*)=l^*) \times \\
& f(M(A=a^*, L(A=a^*)=l^*)=m \mid L(A=a)=l, L(A=a^*)=l^*) \\
= & \sum_{l,l^*,m} E[Y(a,l,m) \mid \ldots] f(L(A=a)=l, L(A=a^*)=l^*) \times \\
& f(M(A=a^*, L(A=a^*)=l^*)=m \mid L(A=a^*)=l^*) \\
& \text{Cross-Worlds Independence of } M_{a^*} = M_{a^* L_{a^*}} \text{ and } L_a \\
= & \sum_{l,l^*,m} E[Y(a,l,m) \mid \ldots] f(L(A=a)=l) f(L(A=a^*)=l^*) \times \\
& f(M(A=a^*, L(A=a^*)=l^*))=m \mid L(A=a^*)=l^*) \\
& \text{Independence of } L_a \text{ and } L_{a^*} \\
= & \sum_{l,l^*,m} E[Y(a,l,m) \mid L(A=a)=l, M(A=a^*, L(A=a^*)=l^*)=m, L(A=a^*)=l^*] \times \\
& f(L(A=a)=l) f(L(A=a^*)=l^*) \times \\
& f(M(A=a^*, L(A=a^*)=l^*)=m \mid L(A=a^*)=l^*) \\
= & \sum_{l^*,m} E[Y(a,m) \mid M(A=a^*, L(A=a^*)=l^*)=m, L(A=a^*)=l^*] \times \\
& f(L(A=a^*)=l^*) f(M(A=a^*, L(A=a^*)=l^*)=m \mid L(A=a^*)=l^*) \\
& \text{Law of Total Probability} \\
= & \sum_{l^*,m} E[Y(a,m)] f(L(A=a^*)=l^*) f(M(A=a^*, L(A=a^*)=l^*)=m \mid L(A=a^*)=l^*) \\
& \text{Exchangeability} \\
= & \sum_{l^*,m} E[Y \mid A=a, M=m] f(L=l^* \mid A=a^*) f(M=m \mid A=a^*, L=l^*) \\
& \text{Consistency}
\end{aligned}
$$

The fifth equality requires the additional assumption of the conditional independence of $L(A=a)$ and $L(A=a^*)$. Without this or another similar simplifying assumption, identification of the natural direct effects in Figure 3 that require simultaneously setting A to different values (say a and a^*) is not possible because

it requires knowledge of the joint density of $L(A = a)$ and $L(A = a^*)$. Put another way, we require the "dual behaviour" of the recanting witness L under $A = a$ and $A = a^*$ [2]. Tchetgen Tchetgen & VanderWeele (2014) [6] consider possible distributional relationships between either binary or continuously-valued $L(A = a)$ and $L(A = a^*)$ that Albert *et al.* (2018) [24] deal with explicitly by assuming a Gaussian bivariate copula and studying the sensitivity of results to the value of the correlation parameter. By assuming $L(A = a)$ and $L(A = a^*)$ to be independent de Stavola *et al.* (2015) [38] showed that the TCE can be decomposed using non-parametric structural equation models. Although the methods we subsequently propose make no assumptions about the joint distribution of L(a) and L(a*), decomposition of the TCE does, however, require additional assumptions about no unmeasured confounding of the association between potential outcomes that we state explicitly in the next section.

4 Effect decomposition for multiple mediators in the presence of exposure-induced mediator-outcome confounding

4.1 Approach 1: Joint mediators

In Figure 3 the effect of A on Y is mediated by M on one causal path that does not include L. In addition, it is mediated through two causal paths featuring L - one that contains M and one that does not. One way to proceed with a decomposition of the TCE is to consider (L, M) jointly as a single multivariate mediator. For this scenario of single exposure, single mediator and single exposure-induced mediator-outcome confounder, VVR14 described methods for estimating the direct and indirect effects by extending assumptions (i)-(iv) to include (v) $Y_{alm} \perp\!\!\!\perp A \mid C$, (vi) $Y_{alm} \perp\!\!\!\perp (L, M) \mid A = a, C$ for $a = 0, 1$, (vii) $(L_a, M_a) \perp\!\!\!\perp A \mid C$, (viii) $Y_{alm} \perp\!\!\!\perp (L_{a^*}, M_{a^*}) \mid C$. Assumption (vi) is achieved directly from the contraction property of conditional independence, namely $Y_{alm} \perp\!\!\!\perp L \mid A = a, C$ for $a = 0, 1$ and $Y_{alm} \perp\!\!\!\perp M \mid A = a, L = l, C$ for all combinations of a, l [36]. Assumption (viii) is satisfied if there is no effect of exposure A that confounds the relationship between the joint mediators (L, M) provided with additionally assume that assumption that the data (and counterfactual data) follow a Nonparametric Structural Equation Model (NPSEM). Following their derivations we describe this method explicitly when applied to the causal diagram in Figure 4 where there are multiple mediators $(M^i, i = 1, 2, \ldots, p)$ and an exposure-induced mediator-outcome confounder (L). The direct and indirect effects treating (L, M) as joint mediator are then defined as $E[Y_{aL_{a^*}M_{a^*}} - Y_{a^*L_{a^*}M_{a^*}}]$ and $E[Y_{aL_aM_a} - Y_{aL_{a^*}M_{a^*}}]$. The indirect effect here

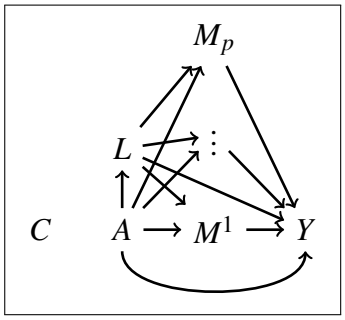

Figure 4: Mediation in the presence of an exposure-induced mediator-outcome confounder. The global confounder C has hidden arrows to all nodes in the graph so once again all statements about independence are conditional on C.

is the effect mediated through either M or L or both. We have the effect decomposition as

$$E[Y_a - Y_{a^*}] = E[Y_{aL_{a^*}M_{a^*}} - Y_{a^*L_{a^*}M_{a^*}}] + E[Y_{aL_aM_a} - Y_{aL_{a^*}M_{a^*}}]$$

Empirical expressions for the corresponding (pure natural) direct and (total natural) indirect effects can be found in VVR14. To decompose the TCE into direct and indirect effects where M^i for $i = 1, 2, \ldots, p$ is a vector of mediators we extend assumptions (v)-(viii) as follows: $(v^\dagger)$ $Y_{alm^1\ldots m^p} \perp\!\!\!\perp A \mid C$; $(vi^\dagger)$ $Y_{alm^1\ldots m^p} \perp\!\!\!\perp (L, M^1, \ldots, M^p) \mid \{A, C\}$; $(vii^\dagger)$ $(L_a, M^1_a, \ldots, M^p_a) \perp\!\!\!\perp A \mid C$, and $((viii)^\dagger)$ $Y_{alm^1\ldots m^p} \perp\!\!\!\perp (L_{a^*}, M^1_{a^*}, \ldots, M^p_{a^*}) \mid C$. In addition we assume that the mediators $(M^1, \ldots, M^p)$ are independent given C, A and L. Once again assumption $vi^\dagger$ is achieved directly from the contraction property of conditional independency. Assumption, $viii^\dagger$ is satisfied if there is no unmeasured exposure-induced mediator-outcome confounders between A and the joint mediators (L, M^i). With these assumptions one possible way of decomposing the TCE is:

$$E[Y_a - Y_{a^*}] = E[Y_{aL_{a^*}M^1_{a^*}\ldots M^p_{a^*}} - Y_{a^*L_{a^*}M^1_{a^*}\ldots M^p_{a^*}}] + E[Y_{aL_aM^1_a\ldots M^p_a} - Y_{aL_{a^*}M^1_{a^*}\ldots M^p_{a^*}}]$$

The empirical derivation of each of the direct and indirect effect is as follows

$$E[Y_{aL_{a^*}M^1_{a^*}\ldots M^p_{a^*}} - Y_{a^*L_{a^*}M^1_{a^*}\ldots M^p_{a^*}}] = \sum_{c,l,m^1,\ldots,m^p} \{E[Y \mid a, l, m^1, \ldots, m^p, c] - E[Y \mid a^*, l, m^1, \ldots, m^p, c]\}$$
$$P(l, m^1, \ldots, m^p \mid a^*, c)P(c) \quad (3)$$

and

$$E[Y_{aL_aM^1_a\ldots M^p_a} - Y_{aL_{a^*}M^1_{a^*}\ldots M^p_{a^*}}] = \sum_{c,l,m^1,\ldots,m^p} E[Y \mid a, l, m^1, \ldots, m^p, c]$$
$$\{P(l, m^1, \ldots, m^p \mid a, c) - P(l, m^1, \ldots, m^p \mid a^*, c)\}P(c) \quad (4)$$

Conditional analogues of these effects $E[Y_{aL_{a^*}M^1_{a^*}\ldots M^p_{a^*}} - Y_{a^*L_{a^*}M^1_{a^*}\ldots M^p_{a^*}} \mid c]$ and $E[Y_{aL_aM^1_a\ldots M^p_a} - Y_{aL_{a^*}M^1_{a^*}\ldots M^p_{a^*}} \mid c]$ are given by the same expression but without summing over the distribution of the covariates C. In practice these quantities are estimated using linear or logistic regression models for continuously-valued or binary outcomes respectively, examples of which we present in Sections 5 and 6. We are assuming that the joint distributions are known explicitly for the exposure-induced mediator-outcome confounder, L, and the mediators, M^i. If this is not the case then one may use (1) a copula to specify the joint distribution for L, M^1, M^2, …, M^p from the full set of marginal distributions; or (2) the inverse odds ratio weighting approach proposed in [12]. Here an exposure-mediator odds ratio is estimated by regressing the (binary) treatment status on the M^i's and L conditional on covariates C. The inverse of this odds ratio is then used in a weighted regression analysis to generate a natural direct effect. The total causal effect is obtained by regressing the outcome on the exposure and covariates but not the mediators. The sum of the natural indirect effects is computed as the difference between the total causal effect and the natural direct effect. This approach depends on the invariance of the odds ratio to which of two variables is declared to be the exposure (with the other as the outcome) which is alternatively interpreted as invariance to the time order of the variables. We offer, however, a note of caution since the odds ratio is not invariant to the choice of threshold for the dichotomisation of a continuously-valued underlying trait as implied by [12] —see [39].

4.2 Approach 2: Path-specific effects

The joint mediation approach is useful if the interest is in learning about the unmediated direct effect of an exposure on the outcome and the combined indirect effect of that exposure on the outcome through all pathways containing one or more mediators. The indirect effect through each of the individual mediators M^i for $i = 1, 2, \ldots, p$ will, however, be of primary interest in many social science and public health applications. Following [2] (see also [38]), VVR14 state that, for the case of single mediator M, the effects through pathways that involve only A (denoted $A \rightarrow Y$), through pathways that involve M but not L (denoted $A \rightarrow M \rightarrow Y$) or through pathways involving L with or without M (the combination of $A \rightarrow L \rightarrow M \rightarrow Y$ and $A \rightarrow L \rightarrow Y$, denoted $A \rightarrow LY$) can all be estimated. The effects through pathways involving M (the combination of $A \rightarrow L \rightarrow M \rightarrow Y$ and $A \rightarrow L \rightarrow Y$) cannot, however, be separately identified. For the case of single mediator (M) and single exposure-induced mediator-outcome confounder (L), VVR14 showed that it is possible to identify and estimate the direct effect $A \rightarrow Y$ and the two indirect effects $A \rightarrow M \rightarrow Y$ and $A \rightarrow LY$. They derived the following

expressions for these effects:

$$E_{A\to Y}(c) = E[Y_{aL_{a^*}M_{a^*}} - Y_{a^*L_{a^*}M_{a^*}} \mid c], \tag{5}$$

$$E_{A\to M\to Y}(c) = E[Y_{aL_{a^*}M_{a_{L_{a^*}}}} - Y_{aL_{a^*}M_{a^*}} \mid c], \tag{6}$$

and

$$E_{A\to LY}(c) = E[Y_{aL_aM_a} - Y_{aL_{a^*}M_{a_{L_{a^*}}}} \mid c]. \tag{7}$$

The total effect decomposes into the sum of these three effects: $E[Y_a - Y_{a^*}] = E_{A\to Y} + E_{A\to M\to Y} + E_{A\to LY}$.

We now extend the derivations for the case where there are two mediators (M^1, M^2). In this extension we assume that there is only one exposure-induced mediator-outcome confounder, L, that confounds the association between each of the mediators and the outcome but that the mediators are independent conditional on C, A and L (Figure 4). The path specific effects in case of two mediators are $A \to Y$, paths that involve M^1 but not L ($A \to M^1 \to Y$), pathways that involve M^2 but not L ($A \to M^2 \to Y$) and paths that involve only L (the combination of $A \to L \to M^1 \to Y$, $A \to L \to M^2 \to Y$ and $A \to L \to Y$). These effects are defined as:

$$\begin{aligned}
E_{A\to Y}(c) &= E[Y_{aL_{a^*}M^1_{a^*}M^2_{a^*}} - Y_{a^*L_{a^*}M^1_{a^*}M^2_{a^*}} \mid c] \\
E_{A\to M^1\to Y}(c) &= E[Y_{aL_{a^*}M^1_{a_{L_{a^*}}}M^2_{a^*}} - Y_{aL_{a^*}M^1_{a^*}M^2_{a^*}} \mid c] \\
E_{A\to M^2\to Y}(c) &= E[Y_{aL_{a^*}M^1_{a_{L_{a^*}}}M^2_{a_{L_{a^*}}}} - Y_{aL_{a^*}M^1_{a_{L_{a^*}}}M^2_{a^*}} \mid c] \\
&\text{and} \\
E_{A\to LY}(c) &= E[Y_{aL_aM^1_aM^2_a} - Y_{aL_{a^*}M^1_{a_{L_{a^*}}}M^2_{a_{L_{a^*}}}} \mid c]
\end{aligned}$$

We show in eAppendix A that the total effect decomposes into the sum of these four path-specific effects: $E[Y_a - Y_{a^*}] = E_{A\to Y} + E_{A\to M^1\to Y} + E_{A\to M^2\to Y} + E_{A\to LY}$. By using the arguments of [2], it follows that if the identifying assumptions $(v^\dagger)$ $Y_{alm^1\ldots m^p} \perp\!\!\!\perp A \mid C$; $(vi^\dagger)$ $Y_{alm^1\ldots m^p} \perp\!\!\!\perp (L, M^1, \ldots, M^p) \mid \{A, C\}$; $(vii^\dagger)$ $(L_a, M^1_a, \ldots, M^p_a) \perp\!\!\!\perp A \mid C$; and $((viii)^\dagger)$ $Y_{alm^1\ldots m^p} \perp\!\!\!\perp (L_{a^*}, M^1_{a^*}, \ldots, M^p_{a^*}) \mid C$ all hold, then these four

effects are identified and are given by the following empirical expressions:

$$E_{A\to Y} = \sum_{c,l,m^1,m^2} \{E[Y \mid a,c,l,m^1,m^2] - E[Y \mid a^*,c,l,m^1,m^2]\}P(l,m^1,m^2 \mid a^*,c)P(c)$$

$$E_{A\to M^1\to Y} = \sum_{c,l,m^1,m^2} E[Y \mid a,c,l,m^1,m^2]\{P(m^1 \mid a,c,l) - P(m^1 \mid a^*,c,l)\}P(l,m^2 \mid a^*,c)P(c)$$

$$E_{A\to M^2\to Y} = \sum_{c,l,m^1,m^2} E[Y \mid a,c,l,m^1,m^2]\{P(m^2 \mid a,c,l) - P(m^2 \mid a^*,c,l)\}P(l \mid a^*,c)$$
$$P(m^1 \mid a,c,l)P(c)$$

$$E_{A\to LY} = \sum_{c,l,m^1,m^2} E[Y \mid a,c,l,m^1,m^2]\{P(l \mid a,c) - P(l \mid a^*,c)\}P(m^1,m^2 \mid a,c,l)P(c)$$

We can now generalize to the case where there are more than two mediators. The proposed decomposition into path-specific effects when there are multiple mediators $M^i, i = 1,2...,p$ is given by

$$E[Y_a - Y_{a^*}] = E_{A\to Y} + E_{A\to M^1\to Y} + ... + E_{A\to M^p\to Y} + E_{A\to LY} \quad (8)$$

where

$$E_{A\to Y}(c) = E[Y_{aL_{a^*}M^1_{a^*}...M^p_{a^*}} - Y_{a^*L_{a^*}M^1_{a^*}...M^p_{a^*}} \mid c]$$
$$E_{A\to M^1\to Y}(c) = E[Y_{aL_{a^*}M^1_{aL_{a^*}}M^2_{a^*}...M^p_{a^*}} - Y_{aL_{a^*}M^1_{a^*}M^2_{a^*}...M^p_{a^*}} \mid c]$$
$$E_{A\to M^2\to Y}(c) = E[Y_{aL_{a^*}M^1_{aL_{a^*}}M^2_{aL_{a^*}}...M^p_{a^*}} - Y_{aL_{a^*}M^1_{aL_{a^*}}M^2_{a^*}...M^p_{a^*}} \mid c]$$
$$\vdots$$
$$E_{A\to M^p\to Y}(c) = E[Y_{aL_{a^*}M^1_{aL_{a^*}}...M^{p-1}_{aL_{a^*}}M^p_{aL_{a^*}}} - Y_{aL_{a^*}M^1_{aL_{a^*}}...M^{p-1}_{aL^*_a},M^p_{a^*}} \mid c]$$

and

$$E_{A\to LY}(c) = E[Y_{aL_aM^1_a...M^p_a} - Y_{aL_{a^*}M^1_{aL_{a^*}}...M^p_{aL_{a^*}}} \mid c].$$

We emphasise that in the above expressions for $A \to M^i \to Y$ for $i = 1,2,\ldots,p$ the subscripts in the contrast of the two potential outcomes for Y are the same with the exception of the subscript for M^i, which is aL_{a^*} in the first term and a^* or equivalently $a^*L_{a^*}$ in the second. These natural indirect effects are, in the language of [40], *total* natural indirect effects with respect to the exposure A and preceding mediators $M_1, M_2, \ldots, M_{i-1}$ since the value of the exposure itself and of the exposure in the first argument for the mediators is $A = a$, and *pure* natural indirect effects with respect to the confounder L and the subsequent mediators $M_{i+1}, M_{i+2}, \ldots, M_p$ since the value of the exposure in the first argument for L and

these mediators is $A = a^*$. The empirical expressions that employ probability distributions for exposures and outcomes are:

$$E_{A\to Y} = \sum_{c,l,m^1,\ldots,m^p} \{E[Y \mid a,c,l,m^1,\ldots,m^p] - E[Y \mid c,a^*,l,m^1,\ldots,m^p]\}$$
$$P(l,m^1,\ldots,m^p \mid a^*,c)P(c)$$
$$E_{A\to M^1\to Y} = \sum_{c,l,m^1,\ldots,m^p} E[Y \mid c,a,l,m^1,\ldots,m^p]\{P(m^1 \mid a,c,l) - P(m^1 \mid a^*,c,l)\}$$
$$P(l,m^2,\ldots,m^p \mid a^*,c)P(c)$$
$$\vdots$$

$$E_{A\to M^p\to Y} = \sum_{c,l,m^1,\ldots,m^p} E[Y \mid a,c,l,m^1,\ldots,m^p]\{P(m^p \mid a,c,l) - P(m^p \mid a^*,c,l)\}$$
$$P(l \mid a^*,c)P(m^1,\ldots,m^{p-1} \mid a,c,l)P(c)$$
$$E_{A\to LY} = \sum_{c,l,m^1,\ldots,m^p} E[Y \mid a,c,l,m^1,m^2]\{P(l \mid a,c) - P(l \mid a^*,c)\}$$
$$P(m^1,\ldots,m^p \mid a,c,l)P(c)$$

4.3 Approach 3: Interventional effects

Similar to the approach presented above, so-called interventional effects can also be defined for multiple paths. Once again we are interested in the indirect effect through each individual mediator rather than one or more of the M^i's and L jointly. Similar to the single mediator situation described in VVR14, we are not able to identify natural direct and indirect effects with M^i alone as the mediator of interest. We can, however, identify and estimate effects that set M^i to a value chosen randomly from its distribution given the value of the exposure A, mimicking the process of randomly allocating the value of the mediator stratified by the levels of the exposure A. Consistent with the notation of VVR14, let $G_{a|C}$ denote a random draw from the distribution of the mediator among those with exposure status a conditional on C. We continue to assume that the M^i's are conditionally independent given L although the interventional effects discussed below remain well-defined without this assumption.

Suppose that we wish to compare the outcome Y for two levels a and a^* of the exposure A. The contrast $E[Y_{aG_{a|C}}] - E[Y_{aG_{a^*|C}}]$ is then the effect on the expected outcome Y of randomly assigning to a person who has their exposure value set to $A = a$ a value of a mediator from the distribution of the mediator among those

whose exposure value is set to $A = a$ (given covariates) compared to randomly assigning to a person who is assigned the *same* exposure value $A = a$ a value of a mediator from the distribution of the mediator among those assigned a *different* exposure value $A = a^*$ (given covariates); this is an indirect effect through the mediator. Next, consider the effect $E[Y_{aG_{a^*|C}}] - E[Y_{a^*G_{a^*|C}}]$. This is the direct effect comparing those with exposure $A = a$ to those with exposure $A = a^*$ with the value of the mediator in both instances drawn randomly from the corresponding distribution of those with exposure $A = a^*$ (given covariates). Finally, the total interventional causal effect $E[Y_{aG_{a|C}}] - E[Y_{a^*G_{a^*|C}}]$ compares the expected outcome Y between those with exposure levels $A = a$ and $A = a^*$ where the value of the mediator is drawn randomly from its distribution in the populations with exposure levels $A = a$ and $A = a^*$ respectively. Extending the interventional effects to multiple paths is straightforward following the derivations in VVR14.

Let $G^1_{a|C}$, and $G^2_{a|C}$ denote the random draw from the distribution of mediators (M^1 and M^2) among those with exposure status a conditional on C. The total interventional causal effect is then defined as $E[Y_{aG^1_{a|C}G^2_{a|C}}] - E[Y_{a^*G^1_{a^*|C}G^2_{a^*|C}}]$. Once again the total interventional causal effect can be decomposed into direct and indirect effects:

$$E[Y_{aG^1_{a|C}G^2_{a|C}}] - E[Y_{a^*G^1_{a^*|C}G^2_{a^*|C}}] = \{E[Y_{aG^1_{a|C}G^2_{a|C}}] - E[Y_{aG^1_{a^*|C}G^2_{a^*|C}}]\} + \{E[Y_{aG^1_{a^*|C}G^2_{a^*|C}}] - E[Y_{a^*G^1_{a^*|C}G^2_{a^*|C}}]\}.$$

Identification of these effects is possible from the assumptions (i) $Y_{am^1m^2} \perp\!\!\!\perp A \mid C$ and (ii) $(M^1_a, M^2_a) \perp\!\!\!\perp A \mid C$ that conditional on C there is no unmeasured exposure-outcome or exposure-mediator confounding. We also need the assumption that (iii) $Y_{am^1m^2} \perp\!\!\!\perp (M^1, M^2) \mid \{A, C, L\}$, that is, conditional on (A, C, L), there is no unmeasured confounding of the mediator-outcome relationship. Figure 4 is consistent with these three assumptions even if the association between L and Y was confounded by unmeasured factors. If assumptions (i)-(iii) hold then the interventional effects are identified by

$$E[Y_{aG^1_{a^*|C}G^2_{a^*|C}}] - E[Y_{a^*G^1_{a^*|C}G^2_{a^*|C}}] = \sum_{c,l,m^1,m^2} \{E[Y \mid a, l, m^1, m^2, c]P(l \mid a, c) - E[Y \mid a^*, l, m^1, m^2, c]P(l \mid a^*, c)\}P(m^1m^2 \mid a^*, c)P(c)$$

$$E[Y_{aG^1_{a|C}G^2_{a^*|C}}] - E[Y_{aG^1_{a^*|C}G^2_{a^*|C}}] = \sum_{c,l,m^1,m^2} E[Y \mid a, l, m^1, m^2, c]P(l \mid a, c) \{P(m^1 \mid a, c) - P(m^1 \mid a^*, c)\} \times P(m^2 \mid a^*, c)P(c)$$

$$E[Y_{aG^1_{a|C}G^2_{a|C}}] - E[Y_{aG^1_{a|C}G^2_{a^*|C}}] = \sum_{c,l,m^1,m^2} E[Y \mid a,l,m^1,m^2,c]P(l \mid a,c)$$
$$\{P(m^2 \mid a,c) - P(m^2 \mid a^*,c)\} \times P(m^1 \mid a,c)P(c)$$

the first expression amounts to averaging controlled direct effects corresponding to different levels of m, using the distribution $P(m^1, m^2 \mid a^*, c)$. These expressions reduce to mediation formulae in (1) and (2) when L does not confound the association between M^1, M^2 and Y conditional on covariates C. In contrast to the effects derived in earlier sections (3.1) and (3.2), the effects derived in this section, based on the assumptions (i) - (iii), do not depend on the "cross-world" assumption (iv). The results from two mediators can be generalized to p mediators from the assumptions (i) $Y_{am^1\ldots m^p} \perp\!\!\!\perp A \mid C$; and (ii) $(M^1_a, ..., M^p_a) \perp\!\!\!\perp A \mid C$ that, conditional on C, there is no unmeasured exposure-outcome or exposure-mediator confounding, along with an additional assumption (iii) that $Y_{am^1\ldots m^p} \perp\!\!\!\perp (M^1, ..., M^p) \mid \{A, C, L\}$. That is, conditional on (A, C, L) there is no unmeasured confounding of the mediator-outcome relationship. Figure 3 is consistent with these three assumptions even if the association between L and Y is confounded by unmeasured factors. The generalized effects are then

$$E[Y_{aG^1_{a^*|C}\ldots G^p_{a^*|C}}] - E[Y_{a^*G^1_{a^*|C}\ldots G^p_{a^*|C}}] = \sum_{c,l,m^1,\ldots,m^p} \{E[Y \mid a,l,m^1,...,m^p,c]P(l \mid a,c)$$
$$- E[Y \mid a^*,l,m^1,m^2,c]P(l \mid a^*,c)\}P(m^1\ldots m^p \mid a^*,c)P(c)$$

$$E[Y_{aG^1_{a|C}\ldots G^p_{a^*|C}}] - E[Y_{aG^1_{a^*|C}\ldots G^2_{a^*|C}}] = \sum_{c,l,m^1,\ldots,m^p} E[Y \mid a,l,m^1,...,m^2,c]P(l \mid a,c)$$
$$\{P(m^1 \mid a,c) - P(m^1 \mid a^*,c)\}P(m^2\ldots m^p \mid a^*,c)P(c)$$
$$\vdots$$

$$E[Y_{aG^1_{a|C}\ldots G^p_{a|C}}] - E[Y_{aG^1_{a|C}\ldots G^{p-1}_{a|C}G^p_{a^*|C}}] = \sum_{c,l,m^1,m^2} E[Y \mid a,l,m^1,m^2,c]P(l \mid a,c)$$
$$\{P(m^p \mid a,c) - P(m^p \mid a^*,c)\}P(m^1\ldots m^{p-1} \mid a,c)P(c)$$

5 Estimators based on weighted regression

Here we develop weighting-based regression methods for estimating direct and indirect effects for each of the three approaches proposed above. Following VVR14,

we pursue an approach based on Marginal Structural Models that uses a model for the exposure ("treatment") to produce inverse-probability-of-treatment weights. The weighting approach can be used for any type of outcome (binary, categorical, count or continuously-valued) although it is best suited to scenarios where the exposure and mediators have a small number of categories. The regression approach can be used to derive both marginal and conditional expectations of potential outcomes.

For path-specific effects we take a sequential approach addressing each mediator in turn, which requires the additional constraint of specifying an order for the mediators. Assuming the M^i are independent given L, A and C, we write the joint density of $M^i = m^i, L = l$ given $A = a$ and $C = c$ in the formulae below as $P(m^i, l \mid a, c) = P(m^i \mid l, a, c)P(l \mid a, c)$. This presents an issue of model compatibility since it may be possible to provide a better fit to the data by allowing the form of the model for both M^i given L, A and C and L given A and C to depend on the index i despite there being only one "true" distribution for L given A and C.

5.1 Weighting under the joint approach

Under this approach a weighting-based estimator is obtained by duplicating the data set and adding an exposure variable A^*, that is 0 for the first replication and 1 for the next. For each person the weight is obtained as the ratio of the product of multiple logistic regression for L and $M^i, i = 1, 2, \ldots, p$ had the exposure been A^*, and the product of the logistic regressions had the exposure been as observed. The denominator will have an additional weight, the reciprocal of observed exposure to adjust for confounding of the exposure-outcome association:

$$w_1 = \frac{P(l \mid a^*, c) \prod_{i=1}^{p} P(m^i \mid l, a^*, c)}{P(a \mid c)P(l \mid a, c) \prod_{i=1}^{p} P(m^i \mid l, a, c)}$$

The natural direct effect $E[Y_{1L_0M_0^1 \ldots M_0^p} - Y_{0L_0M_0^1M_0^p}]$ is then obtained as the coefficient of A in a weighted regression of Y on A among subjects with $A^* = 0$; the natural indirect effect $E[Y_{1L_1M_1^1 \ldots M_1^p} - Y_{1L_0M_0^1 \ldots M_0^p}]$ is obtained as the coefficient of A^* among the subjects $A = 1$ in the duplicated data set.

5.2 Weighting for path-specific effects

Under approach 2, a weighting-based estimator can be obtained upon merging $p+2$ copies of the data set and adding exposure variables $A^{0*}, A^{*1}, A^{*2}, \ldots, A^{*p}$, where

A^{*i} equals the observed exposure A for the first $i+1$ replications and $1-A$ for the remaining $p-i+1$ replications for $i = 0, 1, 2, \ldots, p$. For each person, a weight is obtained by taking the product of the predicted probability (of the observed confounder value) from the logistic regression for L had the exposure been A^{*0} and the predicted probability (of the observed mediator values) from a series of logistic regressions for $M^1, M^2, \ldots, M^p$ had their exposures been $A^{*1}, A^{*2}, \ldots, A^{*p}$ respectively, divided by the product of the corresponding predicted probabilities from the logistic regressions had the exposure been as observed, with additional weighting by the reciprocal of the probability of the observed exposure:

$$w_2 = \frac{P(l \mid a^{*0}, c) \prod_{i=1}^{p} P(m^i \mid l, a^{*i}, c)}{P(a \mid c) P(l \mid a, c) \prod_{i=1}^{p} P(m^i \mid l, a, c)}$$

For a binary exposure, the natural direct effect $E_{A\to Y}$ is now obtained as the coefficient of A in a weighted regression of Y on A among persons with $A^{*i} = 0$ for all values of $i \geq 0$; the natural indirect effect $E_{A\to LY}$ is obtained as the coefficient of A^{*0} in a weighted regression of Y on A^{*0} among persons with $A = 1$ and $A^{*i} = 1$ for all values of $i \geq 1$; finally, the natural indirect effect $E_{A\to M^k\to Y}$ is obtained as the coefficient of A^{*k} in a weighted regression of Y on A^{*k} among those with $A = 1$, $A^{*0} = 0$, $A^{*i} = 1$ for $i < k$ and $A^{*i} = 0$ for $i > k$.

5.3 Weighting for interventional effects

This approach works in a similar way to the multiple pathways approach of Lange *et al.* (2014) [8] but the weights are slightly modified:

$$w_3 = \frac{\prod_{i=1}^{p} \sum_l P(l|a^{*i}, c) P(m^i \mid l, a^{*i}, c)}{P(a \mid c) \prod_{i=1}^{p} P(m^i \mid l, a, c)}$$

The weights w_3 differ from the weights w_2 in the previous sub-section in that the exposure-induced mediator-outcome confounder L is integrated out of its joint distributions with the mediators before we form the joint distribution of the mediators by multiplication. This apparent"marginal independence of the mediators is a deliberate feature of the interventional effects approach to causal mediation analysis. For details on the derivation on the weights refer to the eAppendix.

6 Data Simulation and Analysis

A motivating example for the methods implemented in this paper was presented in Section 2. Here we present the results of simulation study to demonstrate the estimation routines presented in Section 5. We generated 1000 single simulated datasets for each of three sample sizes (500, 1000 and 5000) based on the covariate structure observed in the 1982 Pelotas Birth Cohort. The component models used to generate the mediators and outcome data, that reflect the causal diagram in Figure 1, are as follows:

1. Low socio-economic position(SEP) in childhood ("SEP_{child}", the exposure A) is distributed with probability 0.5;

2. Sex (the covariate C) is distributed with probability of being male 0.5;

3. Conditional on A and C, Low SEP in adulthood ("SEP_{adult}", the exposure-induced mediator-outcome confounder L) is Bernoulli distributed with $logit[\ Pr(L=1)\ ] = -1.89 + 0.37$ Sex $+1.3\ \text{SEP}_{child}$;

4. Conditional on the A, C and L, Smoking (mediator M^1) is Bernoulli distributed with $logit[\ Pr(M^1=1)\ ] = -1.26 + 0.73\ \text{SEP}_{adult} + 0.95\ \text{SEP}_{child} + 0.26$ Sex;

5. Conditional on the A, C, and L, Oral-Hygiene (mediator M^2) is Bernoulli distributed with $logit[Pr(M^2=1)] = 0.19 + 0.67$ Sex $+0.95\ \text{SEP}_{child} + 0.95\ \text{SEP}_{adult}$;

6. Conditional on A, C, L, M^1 and M^2, two outcome variables were generated (i) a continuously-valued measure of the severity of periodontitis (the outcome Y) following a normal distribution mean$=0.122 + 0.55\ \text{SEP}_{child} + 0.65\text{SEP}_{adult} + 0.85$ Smoking $+0.75$ Oral Hygiene and standard deviation $= 0.5$; and (ii) a binary measure for the presence or absence of periodontitis, for which the continuously-valued outcome acted as the logit of the probability of the event "presence of periodontitis".

We compare five methods of causal mediation analysis with multiple mediators: joint mediators (VVR14), inverse odds ratio weighting (see Section 4.1 and [12]), path-specific (VVR14), randomised interventional analogues (VVR14) and multiple pathways (ignoring L and using the methods of [8] - see Section 5.3), which provides estimates of the maximum possible effect solely mediated by each of the mediators M_k for $k = 1, 2, \ldots, p$. We note that the corresponding estimates of the mediated effect of M_k using the methods of VVR14 including M_k and L will be lower in magnitude (closer to zero), although not as small as

Method	Relevant circumstance and study design
(a) Joint mediators (b) Inverse odds ratio weighting	Only the direct effect, total effect and combined indirect effect are of interest.
(c) Path-specific effects	Path-specific effects desired for all mediators including those influenced by other mediators. Requires the "cross-worlds" assumption.
(d) Randomised intervent. analogues (e) Multiple mediators	Intervention on mediators possible, *e.g.* a randomised controlled trial.

Table 2: Methods for the decomposition of the total causal effect and the circumstances under which each is appropriate.

the estimated effects from any of the methods that allow for all of the mediators. Table 2 suggests the circumstances and study designs under which the different decomposition methods are most relevant.

The mean of 1000 estimates of the total causal effect, direct and indirect effects for each of the five methods for each of three different samples sizes (500, 1,000 and 5,000) are presented in Table 3 (for a continuously-valued outcome) and Table 4 (for a binary outcome). 95% confidence intervals for these effect sizes were generated by taking the 2.5 and 97.5 percentiles of the observed distribution of estimates. The estimates of the total causal effect and the component direct and indirect effects show good agreement for all five methods, in many cases showing essentially identical results. The true values of all causal contrasts presented in Tables 3 and 4 could be determined by considering an array of the $2^6 = 64$ potential outcomes $Y_{a^1 L_{a^2} M^1_{a^3 L_{a^4}} M^2_{a^5 L_{a^6}}}$ where $a^i = 0$ or 1 for $i = 1, 2, \ldots, 6$, this substantially increases the computational burden of the simulations, so we make only relative comparisons here. The computing code for R and SAS used to simulate the data set and run the analyses is presented in eAppendix A.

The results of an analysis of data from the Pelotas 1982 Birth Cohort Study (see Section 2) using the same five methods as the simulation study above are presented in Table 5. The joint mediation inverse odds ratio weighting methods indicate that while childhood socio-economic position is a major contributor to inequalities in oral health even when excluding the pathways through adult socio-economic position and two of its consequences, smoking and poor oral hygiene habits, the effects of which appear to be quite small. That is not to say, of course, that adult socio-economic position, smoking and oral hygiene habits are not important determinants of the risk of periodontitis, but instead that only a small proportion of the strong association between childhood socio-economic status and the risk of periodontal disease in adult life acts through these intermediate factors.

Method	Joint Mediators	Inverse Odds Ratio Weighting	Path Specific	Randomised Interventional Analogues	Multiple Mediators
Effect					
n = 500					
Direct (SEP_{child})	0.50 (0.22, 0.78)	0.52 (0.26, 0.77)	0.48 (0.21, 0.74)	0.58 (0.37, 0.77)	0.59 (0.35, 0.82)
Indirect	0.50 (0.34, 0.67)	0.49 (0.33, 0.66)	0.52 (0.32, 0.74)		0.41 (0.29, 0.56)
SEP_{adult}			0.36 (0.04, 0.76)		
Smoking			0.26 (0.15, 0.38)	0.19 (0.12,0.29)	0.23 (0.14, 0.34)
Oral Hygiene			-0.12 (-0.37, 0.11)	0.15 (0.08, 0.24)	0.18 (0.11, 0.28)
Total	1.00 (0.75, 1.25)	1.00 (0.75, 1.25)	1.00 (0.75, 1.25)		1.00 (0.75, 1.25)
n = 1000					
Direct (SEP_{child})	0.48 (0.32, 0.66)	0.48 (0.32, 0.56)	0.47 (0.32, 0.63)	0.63 (0.46, 0.79)	0.59 (0.45, 0.74)
Indirect	0.52 (0.41, 0.63)	0.52 (0.32, 0.64)	0.53 (0.40, 0.65)		0.41 (0.33, 0.52)
SEP_{adult}			0.45 (0.23, 0.68)		
Smoking			0.22 (0.16, 0.30)	0.24 (0.17, 0.31)	0.24 (0.18, 0.31)
Oral Hygiene			-0.15 (-0.29, 0.01)	0.18 (0.12, 0.26)	0.18 (0.12, 0.24)
Total	1.00 (0.84, 1.16)	1.00 (0.84, 1.16)	1.00 (0.85, 1.15))		1.00 (0.84, 1.16)
n = 5000					
Direct (SEP_{child})	0.49 (0.42, 0.57)	0.50 (0.43, 0.57)	0.49 (0.43, 0.58)	0.67 (0.60, 0.75)	0.59 (0.53, 0.66)
Indirect	0.50 (0.45, 0.56)	0.50 (0.45, 0.56)	0.50 (0.45, 0.56)		0.41 (0.37, 0.45)
SEP_{adult}			0.49 (0.38, 0.60)		
Smoking			0.20 (0.17, 0.23)	0.21 (0.19, 0.25)	0.23 (0.20, 0.26)
Oral Hygiene			-0.18 (-0.26, 0.10)	0.16 (0.13, 0.19)	0.18 (0.15, 0.21)
Total	1.00 (0.93, 1.07)	1.00 (0.093, 1.07)	1.00 (0.93, 1.07)		1.00 (0.93, 1.07)

Table 3: The results of the analysis of 1,000 simulated datasets for each of three sample sizes using five different approaches to estimation for partitioning the total causal effect via multiple mediators using a linear regression model for a **continuously-valued outcome** variable, so causal effects are interpretable as differences in population means of potential outcomes. The target total causal effect was 1.00, with the estimated causal effects varying between 0.92 and 1.14. For ease of comparison, the mean estimates of the effect sizes and corresponding 95% confidence intervals have been standardised so that the total causal effect is 1.00. Entries left blank are quantities that are not estimated by the given method. The indirect effect is the sum of the effects for SEP_{adult}, smoking and oral hygiene for methods where these components are estimated. The total effect is the sum of the direct and indirect effect.

Method	Joint Mediators	Inverse Odds Ratio Weighting	Path Specific	Randomised Interventional Analogues	Multiple Mediators
Effect					
n = 500					
Direct (SEP_{child})	1.15 (0.64, 2.07)	1.20 (0.65, 2.09)	1.23 (0.67, 2.12)	1.33 (0.85, 2.10)	1.28 (0.77, 2.30)
Indirect	1.75 (1.29, 2.35)	1.66 (1.23, 2.38)	1.70 (1.22, 2.40)		1.53 (1.25, 1.96)
SEP_{adult}			1.06 (0.99, 1.45)		
Smoking			1.14 (0.89, 1.24)	1.11 (0.99, 1.28)	1.16 (1.03, 1.61)
Oral Hygiene			1.20 (1.06, 1.38)	1.20 (1.08, 1.43)	1.30 (1.13, 1.61)
Total	2.00 (1.21, 3.35)	2.00 (1.21, 3.36)	2.00 (1.21, 3.36)		2.00 (1.21, 3.34)
n = 1000					
Direct (SEP_{child})	1.35 (1.02, 1.81)	1.36 (1.03, 1.79)	1.46 (1.11, 1.97)	1.66 (1.22, 2.37)	1.51 (1.19, 1.97)
Indirect	1.48 (1.29, 1.70)	1.61 (1.29, 1.73)	1.44 (1.22, 1.68)		1.33 (1.21,1.50)
SEP_{adult}			1.12 (1.08, 1.22)		
Smoking			1.09 (1.01, 1.18)	1.10 (1.01, 1.21)	1.09 (1.02, 1.16)
Oral Hygiene			1.23 (1.12, 1.32)	1.27 (1.16, 1.44)	1.22 (1.14, 1.34)
Total	2.00 (1.57, 2.43)	2.00 (1.57, 2.43)	2.00 (1.57, 2.43)		2.00 (1.59,2.58)
n = 5000					
Direct (SEP_{child})	1.44 (1.28, 1.62)	1.44 (1.29, 1.63)	1.48 (1.26, 1.74)	1.80 (1.57, 2.11)	1.53 (1.38, 1.72)
Indirect	1.52 (1.30, 1.49)	1.39 (1.30, 1.49)	1.45 (1.26, 1.64)		1.30 (1.25, 1.37)
SEP_{adult}			1.10 (1.03, 1.19)		
Smoking			1.18 (1.11, 1.28)	1.20 (1.15, 1.25)	1.17 (1.13, 1.21)
Oral Hygiene			1.14 (1.11, 1.18)	1.12 (1.08, 1.18)	1.12 (1.08, 1.16)
Total	2.00 (1.80, 2.21)	2.00 (1.80, 2.07)	2.00 (1.80, 2.10)		2.00 (1.80, 2.20)

Table 4: The results of the analysis of 1,000 simulated datasets for each of three sample sizes using five different approaches to estimation for partitioning the total causal effect via multiple mediators using a logistic regression model for a **binary outcome** variable, so causal effects are interpretable as ratios of the odds of potential outcomes. The target total causal effect was an odds ratio of 2.00, with the estimated causal effects varying between 1.78 and 2.42. For ease of comparison, the mean estimates of the effect sizes and corresponding 95% confidence intervals have been standardised so that the total causal effect is 2.00.Entries left blank are quantities that are not estimated by the given method. The indirect effect is the product of the effects for SEP_{adult}, smoking and oral hygiene for methods where these components are estimated. The total effect is the product of the direct and indirect effect.

Method	Joint Mediators	Inverse Odds Ratio Weighting	Path Specific	Randomised Interventional Analogues	Multiple Mediators
Effect					
$n = 513$					
Direct (SEP_{child})	1.78 (0.46, 3.37)	1.80 (0.41, 3.47)	1.52 (0.29, 3.01)	1.94 (0.78, 3.21)	1.78 (0.72, 2.91)
Indirect	0.01 (-0.67, 0.67)	-0.05 (-0.86, 0.71)	0.06 (-0.66, 0.68)		0.01 (-0.13, 0.15)
SEP_{adult}			-0.03 (-1.05, 0.88)		
Smoking			0.07 (-0.55, 0.67)	0.02 (-0.08, 0.18)	0.01 (-0.07, 0.15)
Oral Hygiene			0.03 (-0.86, 0.93)	-0.02 (-0.14, 0.08)	-0.01 (-0.13, 0.08)
Total	1.79 (0.71, 2.95)	1.75 (0.71, 2.95)	1.58 (0.51, 2.74)		1.78 (0.76, 2.92)

Table 5: The results of an analysis of all data from the 1982 Pelotas Birth Cohort using five different approaches to analysis with multiple mediators. The outcome is the number of teeth with periodontitis (probing depth or gingival margin > 3mm) which we assumed to be continuously-valued. See Section 2, Section 6 and the caption of Table 3 for further explanation.

7 Discussion

We have implemented explicitly and, in one instance, extended the methods for the decomposition of the total causal effect proposed by VVR14 to the case where there are multiple mediators that do not influence each other directly but each of which has an association with the outcome that is subject to confounding by the same exposure-induced mediator-outcome confounder. For indirect effects that correspond to paths in the causal diagram that pass through the exposure-induced confounder, our multi-mediator decomposition does not distinguish between those that do and do not pass through the mediators. Any additional decomposition requires assumptions about the joint distribution of the exposure-induced confounder for different levels of the exposure.

We stated a series of conditional independence assumptions that are required for our proposed extension to multi-mediator problems to correspond to models with identifiable parameters that can be estimated from observed data. Many of these assumptions of conditional independence (sequential ignorability) are untestable.

To mitigate concerns about violations of these assumptions we propose an extension of the method of VanderWeele and Chiba (2014) [41] for assessing the sensitivity of the terms in any calculated decomposition of the total causal effect to the assumptions about no unmeasured confounding between the exposure A and any for the mediators, including the exposure-induced mediator-outcome confounder L. Details can be found in the eAppendix. In this method, a given value of the exposure A and each distinct set of values for the p mediators $M^1, M^2, \ldots, M^k, \ldots, M^{p-1}, M^p$ is allocated a sensitivity parameter that represents the true but unobservable difference between the observed exposure groups in the mean of the corresponding potential outcome for Y conditional on measured confounders C; if all of the conditional independence assumptions are true this will be zero. VanderWeele and Chiba (2014) show that the difference between the estimates and true values of the direct and indirect effects are functions of these sensitivity parameters and estimable probabilities representing the joint distribution of mediators given the exposure and the measured confounders. VanderWeele and Chiba (2014) suggest evaluating this bias for values of the sensitivity parameters between 10% and 25% of the standard deviation of the outcome scale, which, for binary outcomes, corresponds to prevalence differences of between 0.05 and 0.125. Implementing this method for realistically complex problem where there are a potentially large number of combinations of values for the mediators will be the subject of further work.

A question that arises after multiple potential mediators of an exposure-outcome relationship have been measured is whether and, if so, when it is acceptable to perform a series of single-mediator analyses? When, if at all, will the sum of each of

these single-mediator (pure or total) natural indirect effects sum to a well-defined (pure or total) natural, combined simultaneous indirect effect of all of the mediators? Under what conditions will such an equivalence hold? Does it depend on the definition of the combined indirect effect of multiple mediators? These questions have been studied in some detail recently by Daniel *et al.* (2015) [10]. For the scenarios we consider, a series of single-mediator analyses that include the exposure-induced mediator-outcome confounder will, if all assumptions are satisfied, estimate the same value for the indirect effect for each mediator as would have been obtained from either of our proposed multiple-mediator analyses, but only if there is no interaction between the exposure and any of the mediators in their effect on the outcome see eAppendix B. Moreover, the direct effect of the exposure and the indirect effect through the exposure-induced mediator-outcome confounder will vary according to which variable appears as the single mediator, since these effects will be required to absorb the indirect effects through any mediators not included in the analysis. If the assumption that the mediators are conditionally independent given the baseline covariates, the exposure, and the exposure-induced mediator-outcome confounder do not hold because there are two or more mediators that share an unmeasured common cause that is not also caused by the exposure (that is, it is not an exposure-induced mediator-outcome), then the single-mediator analysis will produce biased estimates of the indirect effects through the mediator, whereas the multiple mediator analysis will not. We feel that these drawbacks are sufficient reason to prefer the multiple mediator analyses proposed in this paper.

The proposed path-specific decomposition potentially depends on the order in which the mediators are considered. The indirect effects are generated by cycling through the mediators and, for a given mediator, comparing the expected value of the outcome where the preceding mediators are set to values they would take when the exposure is present and subsequent mediators are set to values they would take when the exposure is absent. The resulting component indirect effects are therefore total with respect to the exposure and some of the mediators and pure with respect to the other mediators (including the exposure-induced confounder), which is necessary if additivity of the natural direct and indirect effects to the total causal effect is to hold. We note that the decompositions based on interventional effects in the recent proposal of [11] also potentially depends on the order of the mediators for the same reasons described above. In the absence of exposure-mediator interactions, pure and total effects coincide and the effect decomposition will be invariant to the order chosen for the mediators. More generally, the consequences of altering the mediator order will need to be investigated empirically through a sensitivity analysis.

Our methods are still valid if the outcome models requires an interaction between two or more mediators. In this scenario, however, the sum of the indirect

effects through mediators M_1 and M_2 will not equal the joint mediated effect of (M_1, M_2). Taguri *et al*. (2018) [20] show that the joint mediated effect can be decomposed into a pure effect through M_1, a total effect through M_2 (or vice versa, total through M_1 and pure through M_2) plus a mediated interaction "MI. Without interaction then there is a unique decomposition of the joint mediated effect through M_1 and M_2 into an effect through M_1 and an effect through M_2 subject to the order restriction discussed above. We concur with Taguri *et al*. (2018) [20]. that further work is required to extend even the most recently proposed methods to consider multiple causally-ordered mediators that interact when influencing the outcome.

Acknowledgements

We thank Professor Aluisio J. D. Barros for his advice and for his permission to use data from the Pelotas 1982 Birth Cohort Study, and Dr Helena S. Schuch for assistance with data management.

References

[1] VanderWeele TJ, Vansteelandt S, Robins JM. Effect decomposition in the presence of an exposure-induced mediator-outcome confounder. *Epidemiology.* 2014;2:300–306.

[2] Avin C, Shpitser I, Pearl J. Identifiability of Path-specific Effects. In: IJCAI'05:357–363Morgan Kaufmann Publishers Inc; 2005; San Francisco, CA, USA.

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

[4] Robins JM, Richardson TS. Alternative graphical causal models and the identification of direct effects. In: Shrout P, ed. *Causality and Psychopathology: Finding the Determinants of Disorders and Their Cures*, Oxford, England: Oxford University Press 2010 (pp. 103–158).

[5] Richardson TS, Robins JM. *Single world intervention graphs (SWIGS): A unification of the counterfactual and graphical approaches to causality.* Working Paper 128: Center for Statistics and the Social Sciences, University of Washington; 2013.

[6] Tchetgen Tchetgen EJ, VanderWeele TJ. Identification of natural direct effects when a confounder of the mediator is directly affected by exposure. *Epidemiology.* 2014;25(2):282–291.

[7] Lange T, Vansteelandt S, Bekaert M. A simple unified approach for estimating natural direct and indirect effects. *American Journal of Epidemiology.* 2012;176(3):190–195.

[8] Lange T, Rasmussen M, Thygesen LC. Assessing natural direct and indirect effects through multiple pathways. *American Journal of Epidemiology.* 2014;179(4):513–518.

[9] VanderWeele TJ, Vansteelandt S. Mediation analysis with multiple mediators. *Epidemiologic Methods.* 2013;2(1):95–115.

[10] Daniel RM, De Stavola BL, Cousens SN, Vansteelandt S. Causal mediation analysis with multiple causally-ordered mediators. *Biometrics.* 2015;71:1–14.

[11] Vansteelandt S, Daniel RM. Interventional effects for mediation analysis with multiple mediators. *Epidemiology.* 2017;28:258–265.

[12] Nguyen QC, Osypuk TL, Schmidt NM, Glymour MM, Tchetgen Tchetgen EJ. Practical guidance for conducting mediation analysis with multiple mediators using inverse odds ratio weighting. *American Journal of Epidemiology.* 2015;181(5):349–356.

[13] Tingley D, Yamamoto T, Hirose K, Keele L, Imai K. mediation: R package for causal mediation analysis. *Journal of Statistical Software.* 2014;59(5):1–38.

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

[15] Yu Q, Fan Y, Wu X. General multiple mediation analysis with an application to explore racial disparities in breast cancer survival. *Journal of Biometrics and Biostatistics.* 2014;5(2):1–9.

[16] Albert JM, Nelson S. Generalized causal mediation analysis. *Biometrics.* 2011;67(3):1028–1038.

[17] Steen J, Loeys T, Moerkerke B, Vansteelandt S. medflex: An R Package for flexible mediation analysis using natural effect models. *Journal of Statistical Software.* 2017;76(11):1–46.

[18] Cheng J, Cheng NF, Guo Z, Gregorich S, Ismail AI, Gansky SA. Mediation analysis for count and zero-inflated count data. *Statistical Methods in Medical Research.* 2018;27(9):2756–2774.

[19] Moreno Betancur M, Carlin JB. Understanding interventional effects: a more natural approach to mediation analysis. *Epidemiology.* 2018;. In Press.

[20] Taguri M, Featherstone J, Cheng J. Causal mediation analysis with multiple causally non-ordered mediators. *Statistical Methods in Medical Research.* 2018;27(1):3–19.

[21] Vansteelandt S. Understanding counterfactual-based mediation analysis approaches and their differences. *Epidemiology.* 2012;23(6):889–891.

[22] VanderWeele TJ, Tchetgen Tchetgen EJ. Mediation analysis with time varying exposures and mediators. *Journal of the Royal Statistical Society: Series B (Statistical Methodology).* 2017;79(3):917–938.

[23] Chen Y-H, Mukherjee B, Ferguson KK, Meeker JD, VanderWeele TJ. Mediation formula for a binary outcome and a time-varying exposure and mediator, accounting for possible exposure-mediator Interaction. *American Journal of Epidemiology.* 2016;184(2):157.

[24] Albert JM, Cho JI, Liu Y, Nelson S. Generalized causal mediation and path analysis: Extensions and practical considerations. *Statistical Methods in Medical Research.* 2018;In press.

[25] Luo P, Geng Z. Causal mediation analysis for survival outcome with unobserved mediatoroutcome confounders. *Computational Statistics and Data Analysis.* 2016;93:336–347.

[26] Huang Y-T, Yang H-I. Causal Mediation Analysis of Survival Outcome with Multiple Mediators. *Epidemiology.* 2017;28(3):370–378.

[27] Wang W, Albert J. Causal mediation analysis for the Cox proportional hazards model with a smooth baseline hazard estimator. *Applied Statistics.* 2017;66(4):741–757.

[28] Fulcher IR, Tchetgen Tchetgen EJ, Williams PL. *Mediation Analysis for Censored Survival Data under an Accelerated Failure Time Model.* Working Paper 211: Department of Biostatistics, Harvard University; 2017.

[29] Lin S-H, Young J, Logan R, VanderWeele TJ. Mediation analysis for a survival outcome with timevarying exposures, mediators, and confounders. *Statistics in Medicine.* 2017;36:4153–4166.

[30] Lin S-H, Young J, Logan R, Tchetgen Tchetgen E, VanderWeele TJ. Parametric mediational g-formula approach to mediation analysis

with time-varying exposures, mediators, and confounders. *Epidemiology.* 2017;28:266–274.

[31] Zheng W, van der Laan M. Longitudinal mediation analysis with time-varying mediator and exposures, with application to survival outcomes. *Journal of Causal Inference.* 2017;5(2).

[32] Liu L, Zheng C, Kang J. Exploring causality mechanism in the joint analysis of longitudinal and survival data. *Statistics in Medicine.* 2018;37(26):3733–3744.

[33] Cho S-H, Huang Y-T. Mediation analysis with causally ordered mediators using Cox proportional hazards model. *Statistics in Medicine.* 2019;38:1566–1581.

[34] Nicolau B, Thomson WM, Steele JG, Allison PJ. Life-course epidemiology: concepts and theoretical models and its relevance to chronic oral conditions. *Community Dentistry and Oral Epidemiology.* 2007;35(4):241–249.

[35] Nascimento GG, Peres MA, Mittinty MN, et al. Diet-induced overweight and obesity and periodontitis risk: An application of the Parametric G-Formula in the 1982 Pelotas Birth Cohort. *American Journal of Epidemiology.* 2017;185(6):442.

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

[37] Shpitser I, Tchetgen Tchetgen EJ. Causal inference with a graphical hierarchy of interventions. *Ann Statist.* 2016;44(6):2433–2466.

[38] De Stavola BL, Daniel RM, Ploubidis GB, Micali N. Mediation analysis with intermediate confounding: Structural equation modelling viewed through the causal inference lens. *American Journal of Epidemiology.* 2015;181:64–80.

[39] Fleiss JL. On the asserted invariance of the odds ratio. *Journal of Epidemiology and Community Health.* 1970;24(1):45–46.

[40] Robins JM, Greenland S. Identifiability and exchangeability for direct and indirect effects. *Epidemiology.* 1992;3(2):143–155.

[41] VanderWeele TJ, Chiba Y. Sensitivity analysis for direct and indirect effects in the presence of exposure-induced mediator-outcome confounders. *Epidemiology and Biostatistics in Public Health.* 2014;11(2):pii: e9027.