

# Causal Mediation Analysis with Multiple Mediators

R. M. Daniel,<sup>1,\*</sup> B. L. De Stavola,<sup>1</sup> S. N. Cousens,<sup>1</sup> and S. Vansteelandt<sup>2</sup>

<sup>1</sup>Centre for Statistical Methodology, London School of Hygiene and Tropical Medicine, Keppel Street, London WC1E 7HT, U.K.

<sup>2</sup>Department of Applied Mathematics, Computer Science and Statistics, Ghent University, Belgium

\**email*: Rhian.Daniel@LSHTM.ac.uk

**SUMMARY.** In diverse fields of empirical research—including many in the biological sciences—attempts are made to decompose the effect of an exposure on an outcome into its effects via a number of different pathways. For example, we may wish to separate the effect of heavy alcohol consumption on systolic blood pressure (SBP) into effects via body mass index (BMI), via gamma-glutamyl transpeptidase (GGT), and via other pathways. Much progress has been made, mainly due to contributions from the field of causal inference, in understanding the precise nature of statistical estimands that capture such intuitive effects, the assumptions under which they can be identified, and statistical methods for doing so. These contributions have focused almost entirely on settings with a single mediator, or a set of mediators considered *en bloc*; in many applications, however, researchers attempt a much more ambitious decomposition into numerous path-specific effects through many mediators. In this article, we give counterfactual definitions of such path-specific estimands in settings with multiple mediators, when earlier mediators may affect later ones, showing that there are many ways in which decomposition can be done. We discuss the strong assumptions under which the effects are identified, suggesting a sensitivity analysis approach when a particular subset of the assumptions cannot be justified. These ideas are illustrated using data on alcohol consumption, SBP, BMI, and GGT from the Izhevsk Family Study. We aim to bridge the gap from “single mediator theory” to “multiple mediator practice,” highlighting the ambitious nature of this endeavor and giving practical suggestions on how to proceed.

**KEY WORDS:** Causal pathways; Decomposition; Multiple mediation; Natural path-specific effects.

## 1. Introduction

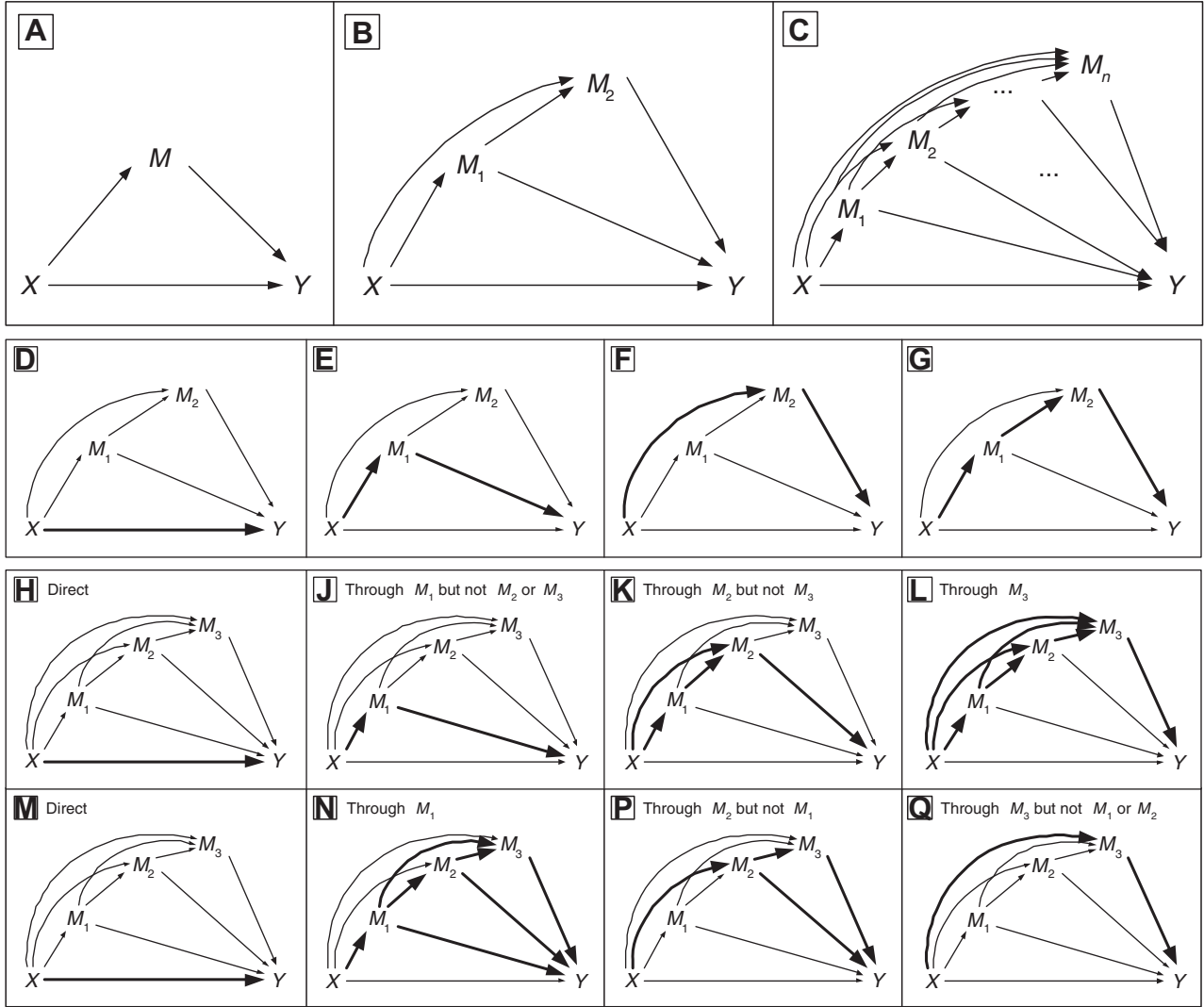
Exploring the relative strength of different pathways from an exposure to an outcome is a topic that has interested scientists across diverse fields for many decades. Early literature (Wright, 1921) through to the 1980s (Bentler, 1980; Baron and Kenny, 1986) focused on path analytic approaches, based on linear regression and structural equation models (SEMs). Under stringent parametric constraints, particular combinations of parameters from these models were taken to represent path-specific effects.

Starting with Robins and Greenland (1992), then Pearl (2001), followed by an explosion of recent contributions (see Ten Have and Joffe, 2012, and references therein, and more recent articles by VanderWeele and coauthors), the formal language and estimation methods from the field of causal inference have shone light on this problem and widened the scope of such analyses, under more explicit assumptions.

Robins and Greenland (1992) and Pearl (2001) used potential outcomes (Neyman, 1923; Rubin, 1978) to give model-free definitions of direct and indirect effect estimands. Informally, a direct effect acts around a mediating variable of interest, whereas the indirect effect acts through this mediator; “direct” thus refers to all other pathways other than through the mediator being considered. The mediator could be multivariate, but if so its constituent variables are considered *en bloc*: the direct effect acts around them all, and the indirect effect is through at least one of them without being further disentangled (Figure 1A).

In a setting with two mediators,  $M_1$  and  $M_2$  (see Figure 1B), there are four possible pathways from exposure ( $X$ ) to outcome ( $Y$ ): through  $M_1$  alone, through  $M_2$  alone, through both and through neither. In this article, our primary aim is to express the total causal effect of  $X$  on  $Y$  as the sum of separate effects along each of the possible pathways: the *finest possible decomposition*. The existing literature on multiple ( $>2$ ) pathways from exposure to outcome can be characterized as follows; either (1)  $M_2$  is the mediator of interest, and  $M_1$  is treated as a mediator–outcome confounder affected by exposure, leading to a coarser two-way decomposition into an effect (indirect) through  $M_2$  and an effect (direct) not through  $M_2$  (Tchetgen Tchetgen and Shpitser, 2012; Vansteelandt and VanderWeele, 2013; VanderWeele and Chiba, 2014; VanderWeele, Vansteelandt, and Robins, 2014), (2) path-specific effects are estimated, but not in such a way that their sum equals the total causal effect (Avin, Shpitser, and Pearl, 2005; Albert and Nelson, 2011), and (3) the multiple mediators do not causally affect one another (MacKinnon, 2000; Preacher and Hayes, 2008), that is, the arrow from  $M_1$  to  $M_2$  in Figure 1B is assumed absent, reducing the number of path-specific effects to three. Imai and Yamamoto (2013) fall into all three categories in different sections of their article, but at no point discuss the finest possible decomposition of the total causal effect in the presence of the arrow from  $M_1$  to  $M_2$ .

The outline for the remainder of the article is as follows. In Section 2 we briefly review mediation estimands in the single mediator setting. In Section 3 we give our proposed



**Figure 1.** Top line: representations of mediation with (A) one, (B) two, and (C)  $n$  mediators, causally ordered. Second line: a depiction of mediation through two causally ordered mediators, with each of the four paths from  $X$  to  $Y$  highlighted; (D) shows the direct path (through neither  $M_1$  nor  $M_2$ ), (E) the indirect path through  $M_1$  alone, (F) the indirect path through  $M_2$  alone, and (G) the indirect path through both  $M_1$  and  $M_2$ . Lines 3 and 4: an illustration of the two possible ways of defining mediator-specific natural effects through three mediators. (H)–(L) show the first way and (M)–(Q) the second.

classification of estimands when there are two causally ordered mediators, showing how decomposition can be achieved, and suggesting strategies for reducing complexity. Section 4 gives sufficient assumptions under which the estimands introduced in Section 3 can be identified, including details of a sensitivity analysis, and estimation methods are discussed briefly in Section 5. The approach is illustrated in Section 6 using data from the Izhevsk Family Study, and we conclude with some discursive remarks in Section 7. Extensions to  $n$  causally ordered mediators (Figure 1C) are given in the Web Appendix.

## 2. A Brief Review of Causal Mediation Estimands for One Mediator

We briefly review mediation estimands for a single mediator. A more detailed account is given in Daniel et al. (2014).

Consider an exposure  $X$ , mediator  $M$  and outcome  $Y$  (Figure 1A). The total, direct and indirect effects defined by Robins, Mark, and Newey (1992) and Pearl (2001) involve the counterfactual variables  $M(x)$ ,  $Y(x)$ ,  $Y(x, m)$ , and  $Y(x, M(x'))$ . These are, respectively, the value  $M$  would take were  $X$  set to  $x$ , the value  $Y$  would take were  $X$  set to  $x$ , the value  $Y$  would take were  $X$  set to  $x$  and  $M$  to  $m$ , and the value  $Y$  would take were  $X$  set to  $x$  and  $M$  to  $M(x')$ .

For simplicity, we take  $X$  to be binary. The *controlled direct effect* (CDE) at level  $m$  of  $M$  is  $E\{Y(1, m) - Y(0, m)\}$ , the *pure natural direct effect* (PNDE) is  $E\{Y(1, M(0)) - Y(0, M(0))\}$ , and the *total natural direct effect* (TNDE) is  $E\{Y(1, M(1)) - Y(0, M(1))\}$ . In each definition,  $M$  takes the same value in both halves of the contrast, corresponding to a “direct” effect. For the CDE, this value of  $M$  is the same for all individuals, whereas for the natural

direct effects, it differs by individual, according to the value that  $M$  would naturally take were  $X$  set to 0 (pure) or 1 (total).

The *pure natural indirect effect* (PNIE) is  $E\{Y(0, X(1)) - Y(0, M(0))\}$  and the *total natural indirect effect* (TNIE) is  $E\{Y(1, M(1)) - Y(1, M(0))\}$ . Note that these correspond to the idea of an indirect (mediated) effect, since they capture the effect on  $Y$  of changing  $X$ , but only via its effect on  $M$ . The first argument of the counterfactual is the same in both halves of each contrast, but this fixed value can be either 0 (pure) or 1 (total).

Note that the sum of the PNDE and TNIE and the sum of the TNDE and PNIE are the same, and that this quantity is the *total causal effect* (TCE) of  $X$  on  $Y$ :  $\text{PNDE} + \text{TNIE} = \text{TNDE} + \text{PNIE} = E\{Y(1, M(1)) - Y(0, M(0))\} = E\{Y(1) - Y(0)\} = \text{TCE}$ . That is, there are two definitions (pure and total) of natural direct and indirect effects, and two ways of decomposing the TCE into a sum of a natural direct and indirect effect. VanderWeele (2013) shows that the difference  $\text{TNDE} - \text{PNDE} = \text{TNIE} - \text{PNIE}$  corresponds to a “mediated interaction,” non-zero if and only if there is an effect of  $X$  on  $M$  and an interaction between  $X$  and  $M$  in their effect on  $Y$ . Thus the choice between the definitions/decompositions, which (in many contexts) is somewhat arbitrary, amounts to assigning the mediated interaction either to the direct or indirect effect.

### 3. Causal Mediation Estimands with Two Causally Ordered Mediators

Turning to the setting with two mediators (Figure 1B) we first note that  $M_1$  can affect  $M_2$  but not vice versa; in some applications, there may be doubt as to the direction of the arrow between  $M_1$  and  $M_2$ , which would introduce further difficulties beyond the scope of this article. We define four *path-specific effects*—one not mediated by either  $M_1$  or  $M_2$  (Figure 1D), one through  $M_1$  alone (Figure 1E), one through  $M_2$  alone (Figure 1F), and one through both  $M_1$  and  $M_2$  (Figure 1G)—such that these sum to the TCE.

#### 3.1. Potential Values of Mediators and Outcome

Let  $M_1(x)$ ,  $M_2(x, m_1)$ ,  $Y(x, m_1, m_2)$ ,  $M_2(x, M_1(x'))$ , and  $Y(x, M_1(x'), M_2(x'', M_1(x''')))$  be defined according to the obvious extensions of the definitions given in Section 2.

#### 3.2. Natural Direct Effects

Let the *natural-000 direct effect* through neither  $M_1$  nor  $M_2$  be  $\text{NDE-000} = E\{Y(1, M_1(0), M_2(0, M_1(0))) - Y(0, M_1(0), M_2(0, M_1(0)))\}$ . This is the obvious extension of the PNDE to two mediators and is the direct effect defined by Avin et al. (2005) and Albert and Nelson (2011). The first argument is the only one that changes, from 1 to 0, making it a direct effect. The other three arguments are fixed at 0; this is why we label it “000.” Rather than two types of effect (pure and total), there are now 8 types of effect—000, 100, 010, 001, 110, 101, 011, and 111—corresponding to each of the ways in which the other three arguments could be set. See Table 1 for all 8 definitions.

#### 3.3. Indirect Effects that Allow Decomposition

We now define indirect effects through  $M_1$  alone,  $M_2$  alone, and through both  $M_1$  and  $M_2$  such that their sum, together with the natural-000 direct effect, is equal to the TCE.

The *natural-100 indirect effect through  $M_1$  alone* is  $\text{NIE}_{1-100} = E\{Y(1, M_1(1), M_2(0, M_1(0))) - Y(1, M_1(0), M_2(0, M_1(0)))\}$ . Intuitively, this corresponds to an indirect effect of  $X$  on  $Y$  via  $M_1$  alone since it captures the effect of  $X$  on  $Y$  only through its effect on  $M_1$ , with the effect of  $M_1$  on  $M_2$  removed. The argument that differs between the two potential outcomes is the second one, the  $x$  shown here:  $Y(\cdot, M_1(x), M_2(\cdot, M_1(\cdot)))$ . The first argument is set to 1 in both potential outcomes, whereas the arguments that follow  $x$  are set to 0; this is why we label it “100.” Similarly, the *natural-110 indirect effect through  $M_2$  alone* is  $\text{NIE}_{2-110} = E\{Y(1, M_1(1), M_2(1, M_1(0))) - Y(1, M_1(1), M_2(0, M_1(0)))\}$ , and the *natural-111 indirect effect through both  $M_1$  and  $M_2$*  is  $\text{NIE}_{12-111} = E\{Y(1, M_1(1), M_2(1, M_1(1))) - Y(1, M_1(1), M_2(1, M_1(0)))\}$ , with each of the seven other types given in Table 1. Note that only the 000 effects have been defined in previous literature (Avin et al., 2005; Albert and Nelson, 2011).

For each effect type, we define its *level* to be the sum of the three fixed  $x$ -arguments. Thus  $\text{NDE-000}$  is a level-0 effect,  $\text{NIE}_{1-100}$  is a level-1 effect, etc.

Using the effects chosen above, it is easily verified that the total causal effect decomposes:

$$\text{TCE} = \text{NDE-000} + \text{NIE}_{1-100} + \text{NIE}_{2-110} + \text{NIE}_{12-111}. \quad (1)$$

Note that Albert and Nelson (2011) study  $\text{NDE-000} + \text{NIE}_{1-000} + \text{NIE}_{2-000} + \text{NIE}_{12-000}$ , and calculate each path-specific 000 effect as a proportion of this sum. Since this sum is *not* in general equal to the total causal effect, these proportions are not analogous to the “proportion mediated” typically calculated in settings with one mediator (Pearl, 2001).

#### 3.4. Alternative Decompositions

The decomposition given in (1) is not the only such decomposition. With one mediator there are two types (pure and total) of two path-specific effects (direct and indirect); with two mediators, there are eight types of four path-specific effects. Forming sums by choosing one type of each effect, with one mediator, we found that two out of the four possible sums equate to the TCE ( $\text{PNDE} + \text{TNIE} = \text{TNDE} + \text{PNIE} = \text{TCE}$ , but  $\text{PNDE} + \text{PNIE} \neq \text{TCE}$  and  $\text{TNDE} + \text{TNIE} \neq \text{TCE}$ ). With two mediators, there are  $8^4 = 4096$  possible sums, and 24 of them equate to the TCE. That is, there are exactly 24 ways of decomposing the TCE into a sum of its path-specific components through and around two mediators: the decomposition shown in (1) and 23 others (see Table 2). That these 24 are unique and represent all possible decompositions follows from the more general argument (for  $n$  mediators) given in Web Appendix A, where we show that there are  $(2^n)!$  ways of decomposing a TCE into a sum of path-specific effects through  $n$  mediators.

With  $n = 2$ , each decomposition includes one level-0, one level-1, one level-2, and one level-3 effect. In short, there are  $4! = 24$  ways of allocating these four levels to the four paths, and this gives rise to the 24 possible decompositions.

**Table 1**

The top half of this table gives the definitions of all natural path-specific effects when there are two causally ordered mediators. There are eight versions (one level-0, three level-1, three level-2, and one level-3) of each of the four effects (direct, indirect through  $M_1$  alone, indirect through  $M_2$  alone, and indirect through both  $M_1$  and  $M_2$ ). The ones shown in bold are the ones defined in Sections 3.2 and 3.3. Note that the level-0 effects are those studied by Avin et al. (2005) and Albert and Nelson (2011). The bottom half of the table gives the definitions of the mediator-specific effects introduced in Section 3.6.3.

Path	Level	Effect	Definition
Direct (through $\emptyset$ )	0	<b>NDE-000</b>	$E \{ Y(1, M_1(0), M_2(0, M_1(0))) - Y(0, M_1(0), M_2(0, M_1(0))) \}$
	1	NDE-100	$E \{ Y(1, M_1(1), M_2(0, M_1(0))) - Y(0, M_1(1), M_2(0, M_1(0))) \}$
	1	NDE-010	$E \{ Y(1, M_1(0), M_2(1, M_1(0))) - Y(0, M_1(0), M_2(1, M_1(0))) \}$
	1	NDE-001	$E \{ Y(1, M_1(0), M_2(0, M_1(1))) - Y(0, M_1(0), M_2(0, M_1(1))) \}$
	2	NDE-110	$E \{ Y(1, M_1(1), M_2(1, M_1(0))) - Y(0, M_1(1), M_2(1, M_1(0))) \}$
	2	NDE-101	$E \{ Y(1, M_1(1), M_2(0, M_1(1))) - Y(0, M_1(1), M_2(0, M_1(1))) \}$
	2	NDE-011	$E \{ Y(1, M_1(0), M_2(1, M_1(1))) - Y(0, M_1(0), M_2(1, M_1(1))) \}$
	3	NDE-111	$E \{ Y(1, M_1(1), M_2(1, M_1(1))) - Y(0, M_1(1), M_2(1, M_1(1))) \}$
Indirect through $M_1$ only	0	NIE <sub>1</sub> -000	$E \{ Y(0, M_1(1), M_2(0, M_1(0))) - Y(0, M_1(0), M_2(0, M_1(0))) \}$
	1	<b>NIE<sub>1</sub>-100</b>	$E \{ Y(1, M_1(1), M_2(0, M_1(0))) - Y(1, M_1(0), M_2(0, M_1(0))) \}$
	1	NIE <sub>1</sub> -010	$E \{ Y(0, M_1(1), M_2(1, M_1(0))) - Y(0, M_1(0), M_2(1, M_1(0))) \}$
	1	NIE <sub>1</sub> -001	$E \{ Y(0, M_1(1), M_2(0, M_1(1))) - Y(0, M_1(0), M_2(0, M_1(1))) \}$
	2	NIE <sub>1</sub> -110	$E \{ Y(1, M_1(1), M_2(1, M_1(0))) - Y(1, M_1(0), M_2(1, M_1(0))) \}$
	2	NIE <sub>1</sub> -101	$E \{ Y(1, M_1(1), M_2(0, M_1(1))) - Y(1, M_1(0), M_2(0, M_1(1))) \}$
	2	NIE <sub>1</sub> -011	$E \{ Y(0, M_1(1), M_2(1, M_1(1))) - Y(0, M_1(0), M_2(1, M_1(1))) \}$
	3	NIE <sub>1</sub> -111	$E \{ Y(1, M_1(1), M_2(1, M_1(1))) - Y(1, M_1(0), M_2(1, M_1(1))) \}$
Indirect through $M_2$ only	0	NIE <sub>2</sub> -000	$E \{ Y(0, M_1(0), M_2(1, M_1(0))) - Y(0, M_1(0), M_2(0, M_1(0))) \}$
	1	NIE <sub>2</sub> -100	$E \{ Y(1, M_1(0), M_2(1, M_1(0))) - Y(1, M_1(0), M_2(0, M_1(0))) \}$
	1	NIE <sub>2</sub> -010	$E \{ Y(0, M_1(1), M_2(1, M_1(0))) - Y(0, M_1(1), M_2(0, M_1(0))) \}$
	1	NIE <sub>2</sub> -001	$E \{ Y(0, M_1(0), M_2(1, M_1(1))) - Y(0, M_1(0), M_2(0, M_1(1))) \}$
	2	<b>NIE<sub>2</sub>-110</b>	$E \{ Y(1, M_1(1), M_2(1, M_1(0))) - Y(1, M_1(1), M_2(0, M_1(0))) \}$
	2	NIE <sub>2</sub> -101	$E \{ Y(1, M_1(0), M_2(1, M_1(1))) - Y(1, M_1(0), M_2(0, M_1(1))) \}$
	2	NIE <sub>2</sub> -011	$E \{ Y(0, M_1(1), M_2(1, M_1(1))) - Y(0, M_1(1), M_2(0, M_1(1))) \}$
	3	NIE <sub>2</sub> -111	$E \{ Y(1, M_1(1), M_2(1, M_1(1))) - Y(1, M_1(1), M_2(0, M_1(1))) \}$
Indirect through both $M_1$ and $M_2$	0	NIE <sub>12</sub> -000	$E \{ Y(0, M_1(0), M_2(0, M_1(1))) - Y(0, M_1(0), M_2(0, M_1(0))) \}$
	1	NIE <sub>12</sub> -100	$E \{ Y(1, M_1(0), M_2(0, M_1(1))) - Y(1, M_1(0), M_2(0, M_1(0))) \}$
	1	NIE <sub>12</sub> -010	$E \{ Y(0, M_1(1), M_2(0, M_1(1))) - Y(0, M_1(1), M_2(0, M_1(0))) \}$
	1	NIE <sub>12</sub> -001	$E \{ Y(0, M_1(0), M_2(1, M_1(1))) - Y(0, M_1(0), M_2(1, M_1(0))) \}$
	2	NIE <sub>12</sub> -110	$E \{ Y(1, M_1(1), M_2(0, M_1(1))) - Y(1, M_1(1), M_2(0, M_1(0))) \}$
	2	NIE <sub>12</sub> -101	$E \{ Y(1, M_1(0), M_2(1, M_1(1))) - Y(1, M_1(0), M_2(1, M_1(0))) \}$
	2	NIE <sub>12</sub> -011	$E \{ Y(0, M_1(1), M_2(1, M_1(1))) - Y(0, M_1(1), M_2(1, M_1(0))) \}$
	3	<b>NIE<sub>12</sub>-111</b>	$E \{ Y(1, M_1(1), M_2(1, M_1(1))) - Y(1, M_1(1), M_2(1, M_1(0))) \}$
		MS <sup>1</sup> -NDE-00	$E \{ Y(1, M_1(0), M_2(0, M_1(0))) - Y(0, M_1(0), M_2(0, M_1(0))) \}$
		MS <sup>1</sup> -NDE-01	$E \{ Y(1, M_1(0), M_2(1, M_1(1))) - Y(0, M_1(0), M_2(1, M_1(1))) \}$
		MS <sup>1</sup> -NDE-10	$E \{ Y(1, M_1(1), M_2(0, M_1(0))) - Y(0, M_1(1), M_2(0, M_1(0))) \}$
		MS <sup>1</sup> -NDE-11	$E \{ Y(1, M_1(1), M_2(1, M_1(1))) - Y(0, M_1(1), M_2(1, M_1(1))) \}$
		MS <sup>1</sup> -NIE <sub>1</sub> -00	$E \{ Y(0, M_1(1), M_2(0, M_1(0))) - Y(0, M_1(0), M_2(0, M_1(0))) \}$
		MS <sup>1</sup> -NIE <sub>1</sub> -01	$E \{ Y(0, M_1(1), M_2(1, M_1(1))) - Y(0, M_1(0), M_2(1, M_1(1))) \}$
		MS <sup>1</sup> -NIE <sub>1</sub> -10	$E \{ Y(1, M_1(1), M_2(0, M_1(0))) - Y(1, M_1(0), M_2(0, M_1(0))) \}$
		MS <sup>1</sup> -NIE <sub>1</sub> -11	$E \{ Y(1, M_1(1), M_2(1, M_1(1))) - Y(1, M_1(0), M_2(1, M_1(1))) \}$

(Continued)

**Table 1**  
Continued

Path	Level	Effect	Definition
		MS <sup>1</sup> -NIE <sub>2</sub> -00	$E \{Y(0, M_1(0), M_2(1, M_1(1))) - Y(0, M_1(0), M_2(0, M_1(0)))\}$
		MS <sup>1</sup> -NIE <sub>2</sub> -01	$E \{Y(0, M_1(1), M_2(1, M_1(1))) - Y(0, M_1(1), M_2(0, M_1(0)))\}$
		MS <sup>1</sup> -NIE <sub>2</sub> -10	$E \{Y(1, M_1(0), M_2(1, M_1(1))) - Y(1, M_1(0), M_2(0, M_1(0)))\}$
		MS <sup>1</sup> -NIE <sub>2</sub> -11	$E \{Y(1, M_1(1), M_2(1, M_1(1))) - Y(1, M_1(1), M_2(0, M_1(0)))\}$
		MS <sup>2</sup> -NDE-00	$E \{Y(1, M_1(0), M_2(0, M_1(0))) - Y(0, M_1(0), M_2(0, M_1(0)))\}$
		MS <sup>2</sup> -NDE-01	$E \{Y(1, M_1(0), M_2(1, M_1(0))) - Y(0, M_1(0), M_2(1, M_1(0)))\}$
		MS <sup>2</sup> -NDE-10	$E \{Y(1, M_1(1), M_2(0, M_1(1))) - Y(0, M_1(1), M_2(0, M_1(1)))\}$
		MS <sup>2</sup> -NDE-11	$E \{Y(1, M_1(1), M_2(1, M_1(1))) - Y(0, M_1(1), M_2(1, M_1(1)))\}$
		MS <sup>2</sup> -NIE <sub>1</sub> -00	$E \{Y(0, M_1(1), M_2(0, M_1(1))) - Y(0, M_1(0), M_2(0, M_1(0)))\}$
		MS <sup>2</sup> -NIE <sub>1</sub> -01	$E \{Y(0, M_1(1), M_2(1, M_1(1))) - Y(0, M_1(0), M_2(1, M_1(0)))\}$
		MS <sup>2</sup> -NIE <sub>1</sub> -10	$E \{Y(1, M_1(1), M_2(0, M_1(1))) - Y(1, M_1(0), M_2(0, M_1(0)))\}$
		MS <sup>2</sup> -NIE <sub>1</sub> -11	$E \{Y(1, M_1(1), M_2(1, M_1(1))) - Y(1, M_1(0), M_2(1, M_1(0)))\}$
		MS <sup>2</sup> -NIE <sub>2</sub> -00	$E \{Y(0, M_1(0), M_2(1, M_1(0))) - Y(0, M_1(0), M_2(0, M_1(0)))\}$
		MS <sup>2</sup> -NIE <sub>2</sub> -01	$E \{Y(0, M_1(1), M_2(1, M_1(1))) - Y(0, M_1(1), M_2(0, M_1(1)))\}$
		MS <sup>2</sup> -NIE <sub>2</sub> -10	$E \{Y(1, M_1(0), M_2(1, M_1(0))) - Y(1, M_1(0), M_2(0, M_1(0)))\}$
		MS <sup>2</sup> -NIE <sub>2</sub> -11	$E \{Y(1, M_1(1), M_2(1, M_1(1))) - Y(1, M_1(1), M_2(0, M_1(1)))\}$

**Table 2**

All 24 possible decompositions of the total causal effect (TCE) into a direct effect (NDE), an indirect effect via  $M_1$  alone (NIE<sub>1</sub>), an indirect effect via  $M_2$  alone (NIE<sub>2</sub>), and an indirect effect via both  $M_1$  and  $M_2$  (NIE<sub>12</sub>). In each decomposition, there is one level-0 effect, one level-1 effect, one level-2 effect, and one level-3 effect. The definitions of each of these effects is given in Table 1. In columns 2-5, the effect types are labeled: 1=000, 2=100, 3=010, 4=001, 5=110, 6=101, 7=011, and 8=111.

Decomposition	Effect and type				TCE =
	NDE	NIE <sub>1</sub>	NIE <sub>2</sub>	NIE <sub>12</sub>	
1	1	2	5	8	NDE-000 + NIE <sub>1</sub> -100 + NIE <sub>2</sub> -110 + NIE <sub>12</sub> -111
2	1	2	8	5	NDE-000 + NIE <sub>1</sub> -100 + NIE <sub>2</sub> -111 + NIE <sub>12</sub> -110
3	1	5	2	8	NDE-000 + NIE <sub>1</sub> -110 + NIE <sub>2</sub> -100 + NIE <sub>12</sub> -111
4	1	6	8	2	NDE-000 + NIE <sub>1</sub> -101 + NIE <sub>2</sub> -111 + NIE <sub>12</sub> -100
5	1	8	2	6	NDE-000 + NIE <sub>1</sub> -111 + NIE <sub>2</sub> -100 + NIE <sub>12</sub> -101
6	1	8	6	2	NDE-000 + NIE <sub>1</sub> -111 + NIE <sub>2</sub> -101 + NIE <sub>12</sub> -100
7	2	1	5	8	NDE-100 + NIE <sub>1</sub> -000 + NIE <sub>2</sub> -110 + NIE <sub>12</sub> -111
8	2	1	8	5	NDE-100 + NIE <sub>1</sub> -000 + NIE <sub>2</sub> -111 + NIE <sub>12</sub> -110
9	3	5	1	8	NDE-010 + NIE <sub>1</sub> -110 + NIE <sub>2</sub> -000 + NIE <sub>12</sub> -111
10	3	8	1	6	NDE-010 + NIE <sub>1</sub> -111 + NIE <sub>2</sub> -000 + NIE <sub>12</sub> -101
11	4	6	8	1	NDE-001 + NIE <sub>1</sub> -101 + NIE <sub>2</sub> -111 + NIE <sub>12</sub> -000
12	4	8	6	1	NDE-001 + NIE <sub>1</sub> -111 + NIE <sub>2</sub> -101 + NIE <sub>12</sub> -000
13	5	1	3	8	NDE-110 + NIE <sub>1</sub> -000 + NIE <sub>2</sub> -010 + NIE <sub>12</sub> -111
14	5	3	1	8	NDE-110 + NIE <sub>1</sub> -010 + NIE <sub>2</sub> -000 + NIE <sub>12</sub> -111
15	6	1	8	3	NDE-101 + NIE <sub>1</sub> -000 + NIE <sub>2</sub> -111 + NIE <sub>12</sub> -010
16	6	4	8	1	NDE-101 + NIE <sub>1</sub> -001 + NIE <sub>2</sub> -111 + NIE <sub>12</sub> -000
17	7	8	1	4	NDE-011 + NIE <sub>1</sub> -111 + NIE <sub>2</sub> -000 + NIE <sub>12</sub> -001
18	7	8	4	1	NDE-011 + NIE <sub>1</sub> -111 + NIE <sub>2</sub> -001 + NIE <sub>12</sub> -000
19	8	1	3	7	NDE-111 + NIE <sub>1</sub> -000 + NIE <sub>2</sub> -010 + NIE <sub>12</sub> -011
20	8	1	7	3	NDE-111 + NIE <sub>1</sub> -000 + NIE <sub>2</sub> -011 + NIE <sub>12</sub> -010
21	8	3	1	7	NDE-111 + NIE <sub>1</sub> -010 + NIE <sub>2</sub> -000 + NIE <sub>12</sub> -011
22	8	4	7	1	NDE-111 + NIE <sub>1</sub> -001 + NIE <sub>2</sub> -011 + NIE <sub>12</sub> -000
23	8	7	1	4	NDE-111 + NIE <sub>1</sub> -011 + NIE <sub>2</sub> -000 + NIE <sub>12</sub> -001
24	8	7	4	1	NDE-111 + NIE <sub>1</sub> -011 + NIE <sub>2</sub> -001 + NIE <sub>12</sub> -000

### 3.5. Example: Linear Structural Equation Model with Interactions

For illustration, we suppose that the data were generated from a linear structural equation model with interactions (and, for simplicity, no confounders), that is, a model implying the following conditional expectations:  $E(M_1|X) = \alpha_0 + \alpha_x X$ ,  $E(M_2|X, M_1) = \beta_0 + \beta_x X + \beta_{m_1} M_1 + \beta_{xm_1} X M_1$  and  $E(Y|X, M_1, M_2) = \gamma_0 + \gamma_x X + \gamma_{m_1} M_1 + \gamma_{m_2} M_2 + \gamma_{xm_1} X M_1 + \gamma_{xm_2} X M_2 + \gamma_{m_1 m_2} M_1 M_2 + \gamma_{xm_1 m_2} X M_1 M_2$ . Note that once interaction terms (or other nonlinearities) are included in the SEM, the simple method of multiplying path coefficients to calculate path-specific effects cannot be applied (VanderWeele and Vansteelandt, 2009).

In Web Appendix B we derive each of the 32 path-specific estimands in this special case in terms of the parameters above, together with certain conditional variance/covariance terms. For example, we have that

$$\begin{aligned} \text{NDE-000} = & \gamma_x + \underbrace{\gamma_{xm_1} \alpha_0 + (\gamma_{xm_2} + \gamma_{xm_1 m_2} \alpha_0) (\beta_0 + \beta_{m_1} \alpha_0)} \\ & + \gamma_{xm_1 m_2} \beta_{m_1} \sigma_{m_1}^2 \end{aligned}$$

where  $\sigma_{m_1}^2 = \text{Var}(M_1|X)$ , and

$$\begin{aligned} \text{NIE}_2\text{-000} = & \gamma_{m_2} \beta_x + \underbrace{\gamma_{m_1 m_2} \beta_x \alpha_0 + \beta_{xm_1} \alpha_0 (\gamma_{m_2} + \gamma_{m_1 m_2} \alpha_0)} \\ & + \gamma_{m_1 m_2} \beta_{xm_1} \sigma_{m_1}^2, \end{aligned}$$

where the terms denoted by the underbraces could be set to zero by adding appropriate constants to  $M_1$  and  $M_2$  (so that  $\alpha_0 = \beta_0 = 0$ ); although in the presence of interactions these terms differ for different effect types (see Web Appendix B). Note that  $\text{NIE}_2\text{-000}$ , for example, contains  $\gamma_{m_2} \beta_x$ , the term that would result from applying the ‘‘product of coefficients’’ methods to a linear model without interactions (Wright, 1921). It also has a further term involving  $\sigma_{m_1}^2$  if there are two interactions present. A similar expression is seen for  $\text{NDE-000}$ , where the ‘‘standard’’ direct effect ( $\gamma_x$ ) appears along with a variance term. The formulæ for some of the other effects involves the covariance of  $M_1(0)$  and  $M_1(1)$ ; we return to this point later. Note that the natural effects derived here would coincide with those used in the LSEM approach in the absence of all interactions.

### 3.6. Practical Suggestions for Reducing Complexity

With two mediators, it can be feasible to estimate all 32 path-specific effects, and hence all 24 decompositions, and compare them. However, with more mediators, the complexity grows at such a rate that this becomes impractical, even for three mediators (see Web Appendix A). In this section, we give three suggestions for reducing this complexity.

**3.6.1. Focusing on effects of greatest substantive interest.** Depending on the exposure, it can often be argued that the 000 effects are substantively most interesting, and easiest to interpret. For example, if  $X = 1$  denotes a new experimental medical treatment, with  $X = 0$  for the standard treatment, then the 000 effects are most naturally interpreted, since they entail setting the free arguments in the effect to what they

would be under the standard treatment. If, in addition, one particular mediator is of greater interest than the others, then the number of decompositions one needs to consider could be partially reduced by focusing only on decompositions that include level 000 of the indirect effect through the mediator of interest (e.g., for  $M_2$ , decompositions 9, 10, 14, 17, 21, and 23 in Table 2). With two mediators, therefore, this strategy reduces the number of decompositions from 24 to 6.

**3.6.2. Summary natural path-specific effects.** We define the *summary natural path-specific effects* SNDE (direct),  $\text{SNIE}_1$  (through  $M_1$  only),  $\text{SNIE}_2$  (through  $M_2$  only) and  $\text{SNIE}_{12}$  (through both  $M_1$  and  $M_2$ ) as follows:

$$\begin{aligned} \text{SNDE} &= \frac{1}{4} (\text{NDE-000} + \text{NDE-111}) + \frac{1}{12} \sum_{0 < i+j+k < 3} \text{NDE-ijk}, \\ \text{SNIE}_1 &= \frac{1}{4} (\text{NIE}_1\text{-000} + \text{NIE}_1\text{-111}) + \frac{1}{12} \sum_{0 < i+j+k < 3} \text{NIE}_1\text{-ijk}, \end{aligned}$$

and similarly for  $\text{SNIE}_2$  and  $\text{SNIE}_{12}$ .

The weights ( $\frac{1}{4}$  and  $\frac{1}{12}$ ) follow from how the path-specific types contribute to each of the 24 decompositions: in columns 2–5 of Table 2, each type-1 and type-8 (000 and 111) effect appears 6 times, and each of the other effect types appears twice. It follows therefore that

$$\text{SNDE} + \text{SNIE}_1 + \text{SNIE}_2 + \text{SNIE}_{12} = \text{TCE} \quad (2)$$

and (2) represents a summary of the 24 decompositions, which itself is a decomposition of the TCE into four (summary) path-specific effects. Whereas with one mediator, the summary direct and indirect effects can be interpreted as the direct and indirect effects that would be seen in a particular randomized experiment (see Web Appendix C), we are not aware of a similar intuitive interpretation of the summary effects for two or more mediators.

When summarizing the effects in this fashion, it would be useful also to consider the variability of the component effects, so that this information is not entirely lost. For example, for the direct effects, we define:

$$\begin{aligned} \text{var-NDE} &= \frac{1}{4} \{ (\text{NDE-000} - \text{SNDE})^2 + (\text{NDE-111} - \text{SNDE})^2 \} \\ &+ \frac{1}{12} \sum_{0 < i+j+k < 3} (\text{NDE-ijk} - \text{SNDE})^2, \end{aligned}$$

weighted to reflect that the SNDE will be closer to  $\text{NDE-000}$  and  $\text{NDE-111}$  than to the other effects. Similar expressions for  $\text{var-NIE}_1$ ,  $\text{var-NIE}_2$  and  $\text{var-NIE}_{12}$  are omitted.

**3.6.3. Mediator-specific natural effects.** Another option is to focus on a coarser decomposition. Indeed, as the number of mediators increases, we are unlikely to be interested in each of the  $2^n$  path-specific effects. For example, with two mediators, we could combine the effect through both  $M_1$  and  $M_2$  with either the effect through  $M_1$  alone, or with the effect through  $M_2$  alone, leaving us with a decomposition into only

three effects: the direct effect, and two *mediator-specific* effects. Graphically, the path shown in Figure 1G could either be combined with that of Figure 1F or with that of Figure 1E. Both lead to natural nested interpretations as follows. In the former (combining G and F, which we will denote as MS<sup>1</sup>, mediator-specific type 1) the mediator-specific direct effect is the effect through neither  $M_1$  nor  $M_2$ , the mediator-specific effect through  $M_1$  is the effect through  $M_1$  but not through  $M_2$ , and the mediator-specific effect through  $M_2$  is all of the effect through  $M_2$ . Similar definitions would apply to the latter (combining G and E, which we will denote as MS<sup>2</sup>). It is perhaps easier to understand this “nesting” argument, by generalizing to three mediators, as shown in Figure 1H–Q.

The algebraic definitions are given in the bottom half of Table 1. Note that such a sequential treatment of multiple mediators is also discussed in VanderWeele and Vansteelandt (2014).

These summaries do not assume no exposure–mediator or no mediator–mediator interactions, as would be required in linear structural equation modeling (see Web Appendix D). Discrepancies between these and estimates obtained under a no-interactions assumption would prompt more closely studying the original contributing path-specific effects.

## 4. Assumptions That Permit Identification

### 4.1. Identification Assumptions

Sufficient assumptions for the identification of the TCE are:

- (T.1) *Consistency of  $X$  on  $Y$* :  $Y(x) = Y$  if  $X = x$ . For those with exposure  $x$ , outcome  $Y$  and potential outcome  $Y(x)$  coincide (Rubin, 1978; Cole and Frangakis, 2009).
- (T.2) *No unmeasured confounding of the  $X$ – $Y$  relationship*: Formally,  $Y(x) \perp\!\!\!\perp X | \mathbf{C} = \mathbf{c}$  for all  $(\mathbf{c}, x)$ , where  $\mathbf{C}$  is a set of measured background confounders, not affected by  $X$ .

Assumption (T.1) is required for the TCE to be interpretable as the effect that would be seen in a hypothetical experiment in which we intervene on  $X$  in a well-defined fashion. The consistency assumption then states that the results are relevant for any kind of intervention which is such that it would have produced the data we have for those for whom  $X = x$  is naturally observed.

Assumption (T.2) states that, after taking into account observed background confounders  $\mathbf{C}$ , any remaining association between  $X$  and  $Y$  can be given a causal interpretation.

This intuition carries through to the extensions of these assumptions in the remainder of this section.

For the CDE, a sufficient set of assumptions is:

- (C.1) *Consistency of  $(X, M)$  on  $Y$* :  $Y(x, m) = Y$  if  $X = x$  and  $M = m$ .
- (C.2) *No unmeasured confounding of the  $(X, M)$ – $Y$  relationship*:  $Y(x, m) \perp\!\!\!\perp X | \mathbf{C} = \mathbf{c}$  for all  $\mathbf{c}$  and  $Y(x, m) \perp\!\!\!\perp M | \mathbf{C} = \mathbf{c}, X = x, \mathbf{L} = \mathbf{l}$  for all  $(\mathbf{c}, x, \mathbf{l}, m)$ , where  $\mathbf{L}$  is a set of measured intermediate confounders, where “intermediate” is used to denote that  $\mathbf{L}$  may be affected by  $X$  (but not by  $M$ ).

If we assume that the data are generated from a non-parametric structural equation model (NPSEM, see Pearl, 2009; Daniel et al., 2014) then, for the identification of the PNDE, TNDE, PNIE and TNIE, a sufficient set of assumptions is (C.1), (C.2), and, in addition:

- (N.3) *Consistency of  $X$  on  $M$* :  $M(x) = M$  if  $X = x$ .
- (N.4) *No unmeasured confounding of the  $X$ – $M$  relationship*:  $M(x) \perp\!\!\!\perp X | \mathbf{C} = \mathbf{c}$  for all  $\mathbf{c}$ .
- (N.5) *No mediator–outcome confounders affected by  $X$* , that is, no intermediate confounders  $\mathbf{L}$ .

Without the NPSEM assumption, (N.5) is replaced by  $Y(x, m) \perp\!\!\!\perp M(x') | \mathbf{C} = \mathbf{c}, \forall \mathbf{c}, m, x = 0, 1, x' = 0, 1$ , which is more difficult to interpret. Either version of assumption (N.5) can be relaxed, but only under strong parametric restrictions. For further details of all aspects of this subsection, see Daniel et al. (2014).

### 4.2. Assumptions for Identifying Path-Specific Effects with Two Causally Ordered Mediators

4.2.1. *Non-parametric identification*. For the CDE with two mediators ( $E\{Y(1, m_1, m_2) - Y(0, m_1, m_2)\}$ ), (C.1) and (C.2) generalize to:

- (MC.1) *Consistency of  $(X, M_1, M_2)$  on  $Y$* .
- (MC.2) *No unmeasured confounding of the  $(X, M_1, M_2)$ – $Y$  relationship*:  $Y(x, m_1, m_2) \perp\!\!\!\perp X | \mathbf{C} = \mathbf{c}$  for all  $(\mathbf{c}, x, m_1, m_2)$ ,  $Y(x, m_1, m_2) \perp\!\!\!\perp M_1 | \mathbf{C} = \mathbf{c}, X = x, \mathbf{L}_1 = \mathbf{l}_1$  for all  $(\mathbf{c}, x, \mathbf{l}_1, m_1, m_2)$  and  $Y(x, m_1, m_2) \perp\!\!\!\perp M_2 | \mathbf{C} = \mathbf{c}, X = x, \mathbf{L}_1 = \mathbf{l}_1, M_1 = m_1, \mathbf{L}_2 = \mathbf{l}_2$  for all  $(\mathbf{c}, x, \mathbf{l}_1, m_1, \mathbf{l}_2, m_2)$ , where  $\mathbf{C}$  are measured background confounders (unaffected by  $X, M_1$  or  $M_2$ ),  $\mathbf{L}_1$  is a set of measured intermediate confounders, which may be affected by  $X$ , but not by  $M_1$ , and  $\mathbf{L}_2$  is a second set of measured intermediate confounders, which may be affected by  $X$  and/or  $M_1$ , but not by  $M_2$ . See Web Figure 5A.

Under (MC.1) and (MC.2), the CDE is then identified using the  $g$ -computation formula (Robins, 1986); see Web Appendix E.

The generalizations of (N.3)–(N.5) (for the natural effects) to two mediators, under the assumption that the data are generated from a NPSEM, are as follows:

- (MN.3) *Consistency of  $X$  on  $M_1$  and of  $(X, M_1)$  on  $M_2$* .
- (MN.4) *No unmeasured confounding of the  $X$ – $M_1$  or  $(X, M_1)$ – $M_2$  relationships*:  $M_1(x) \perp\!\!\!\perp X | \mathbf{C} = \mathbf{c}$  for all  $(\mathbf{c}, x)$ ,  $M_2(x, m_1) \perp\!\!\!\perp M_1 | \mathbf{C} = \mathbf{c}, X = x, \mathbf{L}_1 = \mathbf{l}_1$  for all  $(\mathbf{c}, x, \mathbf{l}_1, m_1)$  and  $M_2(x, m_1) \perp\!\!\!\perp X | \mathbf{C} = \mathbf{c}$  for all  $(\mathbf{c}, x)$ .
- (MN.5) *No mediator–outcome confounder affected by  $X$* , that is, no  $(\mathbf{L}_1, \mathbf{L}_2)$  (Web Appendix F).

Each half of each of the natural path-specific effects in Table 1 is of the form

$$E\{Y(x, M_1(x'), M_2(x'', M_1(x''')))\} \quad (3)$$

and thus if we could identify (3) under assumptions (MC.1), (MC.2) and (MN.3)–(MN.5), all effects in Table 1 would be identified. To this end, we have the following result:

**THEOREM 1.** *Under assumptions (MC.1), (MC.2) and (MN.3)–(MN.5), we have that:*

$$\begin{aligned} & E \{ Y(x, M_1(x'), M_2(x''), M_1(x''')) \} \\ &= \int_{\mathcal{C}} \int_{\mathcal{M}_1} \int_{\mathcal{M}_1} \int_{\mathcal{M}_2} E \{ Y | \mathbf{C} = \mathbf{c}, X = x, M_1 = m_1, M_2 = m_2 \} \\ & \quad \times f_{M_2 | \mathbf{C}, X, M_1} (m_2 | \mathbf{c}, x'', m'_1) \boxed{f_{M_1(x''') | \mathbf{C}, M_1(x')} (m'_1 | \mathbf{c}, m_1)} \\ & \quad \times f_{M_1 | \mathbf{C}, X} (m_1 | \mathbf{c}, x') \\ & \quad \times f_{\mathbf{C}} (\mathbf{c}) d\mu_{M_2} (m_2) d\mu_{M_1} (m'_1) d\mu_{M_1} (m_1) d\mu_{\mathbf{C}} (\mathbf{c}). \quad (4) \end{aligned}$$

For the proof, see Web Appendix H.

Note that (4) involves one density (shown in a box) not written as a function of the distribution of the observed data. A sensitivity analysis when this is unknown is discussed in the next section. There are two special cases in which the boxed quantity in (4) is not required, or is trivially known.

**Special case 1:**  $x' = x''$

If  $x' = x''$ , then  $f_{M_1(x') | \mathbf{C}, M_1(x')} (m'_1 | \mathbf{c}, m_1) = I(m_1 = m'_1)$ . Thus all path-specific estimands in which  $x' = x''$  in both halves of the expression are nonparametrically identified under assumptions (MC.1), (MC.2), and (MN.3)–(MN.5). These are: NDE-000, NDE-010, NDE-101, NDE-111, NIE<sub>2</sub>-000, NIE<sub>2</sub>-100, NIE<sub>2</sub>-011, and NIE<sub>2</sub>-111. Also, note that MS<sup>1</sup>-NDE-00 and MS<sup>1</sup>-NDE-11, together with all of the MS<sup>2</sup>-NDE, MS<sup>2</sup>-NIE<sub>1</sub> and MS<sup>2</sup>-NIE<sub>2</sub> effects, are made up of effects in which  $x' = x''$ , and thus are also identified under assumptions (MC.1), (MC.2) and (MN.3)–(MN.5).

**Special case 2:** No effect of  $M_1$  on  $M_2$

If there is no effect of  $M_1$  on  $M_2$ , the calculation above simplifies as follows

**COROLLARY 1.** *Under assumptions (MC.1), (MC.2), and (MN.3)–(MN.5), if there is no effect of  $M_1$  on  $M_2$ :*

$$\begin{aligned} & E \{ Y(x, M_1(x'), M_2(x'')) \} \\ &= \int_{\mathcal{C}} \int_{\mathcal{M}_1} \int_{\mathcal{M}_2} E \{ Y | \mathbf{C} = \mathbf{c}, X = x, M_1 = m_1, M_2 = m_2 \} \\ & \quad \times f_{M_2 | \mathbf{C}, X} (m_2 | \mathbf{c}, x'') f_{M_1 | \mathbf{C}, X} (m_1 | \mathbf{c}, x') f_{\mathbf{C}} (\mathbf{c}) \\ & \quad \times d\mu_{M_2} (m_2) d\mu_{M_1} (m_1) d\mu_{\mathbf{C}} (\mathbf{c}). \end{aligned}$$

All effects (when  $M_1$  does not affect  $M_2$ ) are thus nonparametrically identified under assumptions (MC.1), (MC.2), and (MN.3)–(MN.5).

In the absence of an effect of  $M_1$  on  $M_2$ , the definitions and decompositions given in Section 3 simplify. There is no longer a path through both  $M_1$  and  $M_2$ , and thus the fourth argument in each half of each effect disappears. This leaves 12 effects, and 6 decompositions; the effects are

listed in Table 3, with the decompositions given in Web Table 1. Some of these effects and decompositions correspond to those given by Imai and Yamamoto (2013); in particular, Imai and Yamamoto define the NDE-00, NDE-01, NDE-10, NDE-11, NIE<sub>1</sub>-00, NIE<sub>1</sub>-11, NIE<sub>2</sub>-00, and NIE<sub>2</sub>-11, but not the remaining 4 effects (see Table 3) and they point out that TCE = NDE-01 + NIE<sub>1</sub>-00 + NIE<sub>2</sub>-11 = NDE-10 + NIE<sub>1</sub>-11 + NIE<sub>2</sub>-00 but do not give the other four possible decompositions (see Web Table 1). A summary of the comparison between the estimands defined and identified in the current manuscript versus those defined and identified in the previous literature is given in Web Table 2.

Note that the decompositions given in Web Table 1 apply also to the mediator-specific natural effects defined in Section 3.6.3.

As already noted, Avin et al. (2005) define only 000 effects, but, by symmetry, their identification result applies also to the 111 effects. Insofar as they can be compared, our result agrees with that of Avin et al. since they conclude that the effect along the direct pathway ( $X \rightarrow Y$ ) and the effect along the indirect pathway through  $M_2$  alone ( $X \rightarrow M_2 \rightarrow Y$ ) are identifiable, but that the effects along the other two pathways ( $X \rightarrow M_1 \rightarrow Y$  and  $X \rightarrow M_1 \rightarrow M_2 \rightarrow Y$ ) are not. This corresponds to what we find, since NDE-000, NDE-111, NIE<sub>2</sub>-000, and NIE<sub>2</sub>-111 are all included in our list of effects which can be estimated without the sensitivity parameter, whereas none of the NIE<sub>1</sub> or NIE<sub>12</sub> effects is included in this list.

**4.2.2. Identification and sensitivity analysis under a particular parametric model.** When there is an effect of  $M_1$  on  $M_2$ , the effects not listed under “special case 1” above require knowledge of the boxed quantity in (4) when  $x' \neq x''$ . Under most estimation strategies (see Section 5), we would assume a parametric model for the distribution of  $M_1$  given  $X$  and  $\mathbf{C}$ , for example that  $M_1 | \mathbf{C}, X \sim N(f(\mathbf{C}, X; \boldsymbol{\beta}), \sigma^2)$ , and we would estimate the parameters  $\boldsymbol{\beta}$  and  $\sigma^2$  from data on  $\mathbf{C}$ ,  $X$  and  $M_1$ . Under assumptions (MN.3) and (MN.4) and if our model for  $M_1 | \mathbf{C}, X$  is correctly specified, this gives us  $M_1(x) | \mathbf{C} \sim N(f(\mathbf{C}, x; \boldsymbol{\beta}), \sigma^2)$  for  $x = 0, 1$ . In this case, in order to know the boxed quantity in (4), we would need, in addition to this model, the conditional correlation between  $M_1(0)$  and  $M_1(1)$  given  $\mathbf{C}$ . There is no information in the data on the value of this correlation; a sensible approach would thus be to vary this parameter in a sensitivity analysis.

For example, consider the following form for the SEM for  $M_1$ :  $M_1 = h(\mathbf{C}, X) + U_{M_1,0}(1 - X) + U_{M_1,1}X + U_{M_1,2}$ , for some function  $h(\mathbf{C}, X)$ , where  $U_{M_1} = (U_{M_1,0}, U_{M_1,1}, U_{M_1,2})$  and

$$\begin{pmatrix} U_{M_1,0} \\ U_{M_1,1} \\ U_{M_1,2} \end{pmatrix} \sim N \left( \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \sigma^2 \begin{pmatrix} 1 - \kappa^2 & 0 & 0 \\ 0 & 1 - \kappa^2 & 0 \\ 0 & 0 & \kappa^2 \end{pmatrix} \right).$$

Then  $M_1(1) | M_1(0), \mathbf{C} \sim N(h(\mathbf{c}, 1) + \kappa^2(M_1(0) - h(\mathbf{c}, 0)), (1 - \kappa^4)\sigma^2)$ . Note that  $\sigma^2 = \text{Var}(M_1 | \mathbf{C}, X)$  can be estimated from the data. However, the data contain no information on  $\kappa^2$ , the proportion of residual variance shared across worlds; this becomes the sensitivity parameter, to be varied from 0 to 1. For more details, see Web Appendix J. An example of this sort of sensitivity analysis is given in Section 6.



Table 3

The definitions of all natural path-specific effects when there are two mediators that are not causally ordered. There are four versions (one level-0, two level-1, and one level-2) of each of the three effects (direct, indirect through  $M_1$ , and indirect through  $M_2$ ; note that there is no effect through both  $M_1$  and  $M_2$  when the mediators are not causally ordered).

Path	Level	Effect	Definition
Direct (through $\emptyset$ )	0	NDE-00	$E \{Y(1, M_1(0), M_2(0)) - Y(0, M_1(0), M_2(0))\}$
	1	NDE-10	$E \{Y(1, M_1(1), M_2(0)) - Y(0, M_1(1), M_2(0))\}$
	1	NDE-01	$E \{Y(1, M_1(0), M_2(1)) - Y(0, M_1(0), M_2(1))\}$
	2	NDE-11	$E \{Y(1, M_1(1), M_2(1)) - Y(0, M_1(1), M_2(1))\}$
Indirect through $M_1$	0	NIE <sub>1</sub> -00	$E \{Y(0, M_1(1), M_2(0)) - Y(0, M_1(0), M_2(0))\}$
	1	NIE <sub>1</sub> -10	$E \{Y(1, M_1(1), M_2(0)) - Y(1, M_1(0), M_2(0))\}$
	1	NIE <sub>1</sub> -01	$E \{Y(0, M_1(1), M_2(1)) - Y(0, M_1(0), M_2(1))\}$
	2	NIE <sub>1</sub> -11	$E \{Y(1, M_1(1), M_2(1)) - Y(1, M_1(0), M_2(1))\}$
Indirect through $M_2$	0	NIE <sub>2</sub> -00	$E \{Y(0, M_1(0), M_2(1)) - Y(0, M_1(0), M_2(0))\}$
	1	NIE <sub>2</sub> -10	$E \{Y(1, M_1(0), M_2(1)) - Y(1, M_1(0), M_2(0))\}$
	1	NIE <sub>2</sub> -01	$E \{Y(0, M_1(1), M_2(1)) - Y(0, M_1(1), M_2(0))\}$
	2	NIE <sub>2</sub> -11	$E \{Y(1, M_1(1), M_2(1)) - Y(1, M_1(1), M_2(0))\}$

A similar approach was taken by Daniels et al. (2012), for discrete mediators by Albert and Nelson (2011), and in the context of treatment noncompliance by Roy, Hogan, and Marcus (2008). Note that this sensitivity analysis solely assesses sensitivity to the arbitrary choice of conditional distribution of  $M_1(1)$  given  $M_1(0)$  and  $\mathbf{C}$ ; it does not explore sensitivity to departures from the other assumptions, namely (MC.1), (MC.2), and (MN.3)–(MN.5). An extensive literature on sensitivity analyses with respect to the single mediator versions of these assumptions exists, including in the presence of mediator–outcome confounders affected by the exposure (see, e.g., Imai, Keele, and Yamamoto, 2010; Tchetgen Tchetgen and Shpitser, 2012; VanderWeele and Chiba, 2014). In future work, we will extend these sorts of sensitivity analyses to the current setting.

An alternative route to parametric identification and sensitivity analysis would be to extend the “no interaction” assumption made by Robins and Greenland (1992) and relaxed by Imai and Yamamoto (2013). Given, however, that the 24 possible decompositions differ precisely when interactions are present, assuming them away may not be as attractive.

In Web Appendix K, we show what our identification results imply for the special case of the linear model with interactions introduced in Section 3.5, and in Web Appendix L, we show how identification is achieved, up to a set of sensitivity parameters, in the presence of a restricted pattern of intermediate confounding.

## 5. A Note on Estimation Methods

The most obvious estimation approach is to posit parametric (regression) models for each density/expectation in the identifying equations above, to estimate their parameters from the observed data (e.g., by maximum likelihood), and then to evaluate the integrals analytically. Pearl (2009) calls this approach the *mediation formula*. Closely related to the *g-computation formula* (Robins, 1986), which can be used to

estimate controlled direct effects in the presence of intermediate confounding, the mediation formula makes the additional step of integrating over the (conditional counterfactual) mediator distribution, in order to obtain natural effects (VanderWeele and Vansteelandt, 2009, 2010). When the integration is too cumbersome to be done analytically, it can instead be done by Monte Carlo simulation (Robins, 1986; Imai, Keele, and Tingley, 2010; Daniel, De Stavola, and Cousens, 2011).

The advantage of relying heavily on parametric models is that this approach is efficient when all models are correct; however, as pointed out by Robins and Wasserman (1997) and further discussed by Vansteelandt, Bekaert, and Lange (2012), the disadvantage is that it can be essentially impossible to specify these models such that they imply a sensible parsimonious model for the direct effect of interest. For this reason, and, more generally, to reduce reliance on parametric modelling assumptions, alternative semiparametric estimation approaches have been suggested (van der Laan and Petersen, 2008; VanderWeele, 2009; Tchetgen Tchetgen and Shpitser, 2012; Vansteelandt et al., 2012; Zheng and van der Laan, 2012). G-computation has nevertheless turned out to be rather successful in recent empirical applications (Young et al., 2011; Westreich et al., 2012).

We therefore adopt the fully parametric approach, implemented by Monte Carlo simulation, extending it to handle multiple mediators and incorporating the sensitivity analysis of Section 4.2.2. In future work, semiparametric estimation methods will be explored.

## 6. An Illustrative Data Example: The Izhevsk Family Study

### 6.1. Data and Question of Interest

The population-based controls from a case-control study conducted in Izhevsk, Russia (Leon et al., 2007) are used to study

the effect of heavy drinking during the previous year (defined as the consumption of >10 L ethanol) on systolic blood pressure (SBP), measured in mmHg. We decompose this into an effects via body mass index (BMI), via gamma-glutamyl transpeptidase (GGT), via both BMI and GGT, and a direct effect, that is, via other pathways. BMI is known to affect GGT (and not vice versa), and thus the set-up is as we have discussed, with  $M_1 = \text{BMI}$  and  $M_2 = \text{GGT}$ . We estimate the path-specific effects using data on 1275 men with complete information on yearly ethanol consumption (from which “heavy drinking” is derived) and all baseline confounders: age (treated as a continuous variable), socio-economic status (SES) score (the first principal component from an asset score analysis), smoking status (current/ex/never), and cigarettes per day ( $\leq 10, 10\text{--}20, >20$ ): together we label these confounders  $\mathbf{C}$  (Leon et al., 2007). Note that in this setting there are no (measured) intermediate confounders. Subjects with missing values of BMI, GGT and/or SBP are not excluded, since these partially observed records can be incorporated, under the missing at random assumption (Rubin, 1976). Some descriptive statistics are shown in Web Table 3.

### 6.2. Estimation by Parametric G-computation via Monte Carlo Simulation

Flexible parametric models for  $M_1 | \mathbf{C}, X$ ,  $M_2 | \mathbf{C}, X, M_1$ , and  $Y | \mathbf{C}, X, M_1, M_2$  were explored. To render the normality assumption for the errors more tenable,  $M_1$  and  $M_2$  (i.e., BMI and GGT) were log-transformed. All models included all possible two- and three-way interactions between exposure and mediators, so that the path-specific effects of different types differ as much as the data dictate. In addition, quadratic terms for the continuous variables (age, SES, BMI, and GGT) were considered where relevant, as well as interactions between exposure and confounders; these were included only if they improved the AIC (see Web Appendix M).

Write  $E(M_1 | \mathbf{C}, X) = v_1(\mathbf{C}, X; \beta_1)$ ,  $E(M_2 | \mathbf{C}, X, M_1) = v_2(\mathbf{C}, X, M_1; \beta_2)$ , and  $E(Y | \mathbf{C}, X, M_1, M_2) = v_3(\mathbf{C}, X, M_1, M_2; \beta_3)$  for the conditional expectations implied by this model, and let the error variances be  $\sigma_1^2$ ,  $\sigma_2^2$  and  $\sigma_3^2$ , respectively.

The estimation of path-specific effects is carried out as follows.

- (1) Estimate the parameters  $(\beta, \sigma)$  by OLS/ML.
- (2) For each subject  $i$ , draw  $V_i$  from  $N(0, \kappa^2 \hat{\sigma}_1^2)$ .  $\kappa$  is the sensitivity parameter (see Web Appendix J), to be varied between 0 (no cross-world correlation conditional on  $\mathbf{C}$ ) and 1 (perfect cross-world correlation conditional on  $\mathbf{C}$ ).
- (3) For  $x = 0, 1$ , draw  $M_{1,i}(x)$  for each  $i$  from  $N\left(v_1(\mathbf{C}_i, x; \hat{\beta}_1) + V_i, (1 - \kappa^2) \hat{\sigma}_1^2\right)$ .
- (4) For  $x = 0, 1$  and  $x' = 0, 1$ , draw  $M_{2,i}(x, M_{1,i}(x'))$  from  $N\left(v_2(\mathbf{C}_i, x, M_{1,i}(x'); \hat{\beta}_2), \hat{\sigma}_2^2\right)$ .
- (5) For  $x = 0, 1$ ,  $x' = 0, 1$ ,  $x'' = 0, 1$  and  $x''' = 0, 1$ , draw  $Y_i(x, M_{1,i}(x'), M_{2,i}(x''), M_{1,i}(x'''))$  from  $N\left(v_3(\mathbf{C}_i, x, M_{1,i}(x'), M_{2,i}(x''), M_{1,i}(x''')); \hat{\beta}_3\right), \hat{\sigma}_3^2$ .
- (6) To estimate each of the 32 effects  $E\{Y(x, M_1(x'), M_2(x''), M_1(x''')) - Y(z, M_1(z'), M_2(z''), M_1(z'''))\}$ , the empirical

average across all subjects of  $Y_i(x, M_{1,i}(x'), M_{2,i}(x''), M_{1,i}(x''')) - Y_i(z, M_{1,i}(z'), M_{2,i}(z''), M_{1,i}(z'''))$  is found.

To decrease Monte Carlo error, the simulation is done on a dataset 1000 times the size of the original (with the values of  $\mathbf{C}$  copied 1000 times), although the estimation of the parameters  $\sigma_1^2, \sigma_2^2, \sigma_3^2$  is based on the original sample. Standard errors are computed using the nonparametric bootstrap. For comparison, a LSEM (with no interactions) is also fitted.

### 6.3. Results

The results are shown in Tables 4 and 5 and Web Figures 1 and 2. There is evidence of a total effect of heavy drinking on SBP, but the associated confidence interval is wide (mean difference 7.63 mmHg, 95% CI 3.89–11.37). Only a small proportion (1.7%) of the large variation in SBP across this sample of men is explained by the dichotomous heavy drinking variable. It is not surprising therefore that the estimates of the various path-specific effects are also imprecise. Examination of the residual distribution for each contributing associational model shows good agreement with the assumption of normality while evidence for the interaction terms was weak (see Web Table 5). There is evidence of a small indirect effect through GGT alone (mean difference ranging from 2.85 to 3.10 mmHg, lower 95% confidence limit ranging from 1.05 to 1.43, upper 95% confidence limit ranging from 4.31 to 5.06), little evidence of path-specific effects through either BMI alone or both BMI and GGT, with the remaining part of the total effect attributed to a direct effect via other pathways (mean difference ranging from 5.07 to 5.25 mmHg, lower 95% confidence limit ranging from 1.35 to 1.48, upper 95% confidence limit ranging from 8.76 to 9.03). There is little variation between the eight versions of each effect. As a consequence, when we depict the 24 possible decompositions in Figure 2, they are all similar, which suggests—in this example—that conclusions about the comparative strengths of different pathways could be drawn from just one particular decomposition.

Due to the lack of important interactions, the summary effects included in Table 4 and Web Figure 1 are similar to each of the 8 effects in each instance. They are also similar to the results obtained when assuming no exposure–mediator interactions as implicitly done when fitting a traditional LSEM and multiplying path coefficients (note however the narrower CIs in the latter, due to the assumption of no interactions). The mediator-specific effects (Table 5 and Web Figure 2) also show a similar picture, with little difference between the two ways of defining the mediator-specific effects, due to the small magnitude of the path-specific effect through both BMI and GGT.

The results appear to be insensitive to variations in  $\kappa$  (Tables 4 and 5), and confirm that some effects do not depend on  $\kappa$  as theory suggests (see Web Figures 16–19).

### 6.4. Limitations

The exposure (heavy drinking) is likely subject to misclassification. This is of particular concern in mediation analyses, if either of the mediators (in this case GGT) is a good proxy

Table 4

Estimates, SEs, and 95% confidence intervals for the total causal effect (TCE), followed by each of the path-specific effects we have defined. All estimates are for mean differences in SBP measured in mmHg. The results are given for three values of the sensitivity parameter  $\kappa$ : 1, 0.5 and 0.

Effect	$\kappa = 1$		$\kappa = 0.5$		$\kappa = 0$	
	Estimate	95% CI	Estimate	95% CI	Estimate	95% CI
TCE	7.63	(3.89,11.37)	7.62	(4.04,11.20)	7.64	(4.05,11.24)
NDE-000	5.08	(1.38,8.78)	5.08	(1.35,8.81)	5.09	(1.53,8.66)
NDE-100	5.07	(1.35,8.79)	4.92	(1.13,8.70)	4.88	(1.25,8.51)
NDE-010	5.25	(1.47,9.03)	5.25	(1.52,8.98)	5.26	(1.62,8.90)
NDE-001	5.24	(1.48,9.01)	5.17	(1.45,8.90)	5.16	(1.52,8.80)
NDE-110	5.11	(1.46,8.76)	4.96	(1.28,8.65)	4.91	(1.39,8.44)
NDE-101	5.10	(1.44,8.77)	5.11	(1.43,8.79)	5.11	(1.59,8.62)
NDE-011	5.21	(1.45,8.97)	5.14	(1.44,8.83)	5.12	(1.50,8.73)
NDE-111	5.21	(1.46,8.95)	5.21	(1.53,8.89)	5.21	(1.62,8.80)
SNDE	5.15	(1.57,8.73)	5.12	(1.56,8.68)	5.11	(1.68,8.54)
$\sqrt{\text{var-NDE}}$	0.07		0.10		0.11	
DE <sup>nointer</sup>	5.24	(1.72,8.76)	5.24	(1.73,8.75)	5.24	(1.86,8.62)
NIE <sub>1</sub> -000	-0.36	(-1.01,0.29)	-0.23	(-1.02,0.56)	-0.18	(-1.05,0.69)
NIE <sub>1</sub> -100	-0.36	(-1.01,0.28)	-0.50	(-1.13,0.13)	-0.54	(-1.22,0.14)
NIE <sub>1</sub> -010	-0.39	(-1.00,0.21)	-0.39	(-1.06,0.28)	-0.36	(-1.09,0.37)
NIE <sub>1</sub> -001	-0.39	(-1.00,0.21)	-0.42	(-1.07,0.22)	-0.41	(-1.08,0.27)
NIE <sub>1</sub> -110	-0.36	(-1.01,0.29)	-0.23	(-1.02,0.56)	-0.18	(-1.05,0.69)
NIE <sub>1</sub> -101	-0.43	(-1.09,0.24)	-0.50	(-1.21,0.21)	-0.51	(-1.23,0.21)
NIE <sub>1</sub> -011	-0.39	(-1.00,0.23)	-0.41	(-1.03,0.22)	-0.39	(-1.03,0.25)
NIE <sub>1</sub> -111	-0.42	(-1.09,0.24)	-0.37	(-1.10,0.36)	-0.34	(-1.12,0.45)
SNIE <sub>1</sub>	-0.40	(-0.98,0.18)	-0.39	(-0.99,0.20)	-0.37	(-0.99,0.24)
$\sqrt{\text{var-NIE}_1}$	0.02		0.07		0.09	
IE <sub>1</sub> <sup>nointer</sup>	-0.39	(-0.94,0.16)	-0.40	(-0.96,0.15)	-0.38	(-0.94,0.18)
NIE <sub>2</sub> -000	2.85	(1.39,4.32)	2.86	(1.45,4.27)	2.85	(1.41,4.30)
NIE <sub>2</sub> -100	3.04	(1.09,4.98)	3.04	(1.04,5.04)	3.03	(1.04,5.03)
NIE <sub>2</sub> -010	2.96	(1.05,4.87)	3.03	(1.09,4.97)	3.06	(1.10,5.01)
NIE <sub>2</sub> -001	2.85	(1.39,4.32)	2.86	(1.45,4.27)	2.85	(1.41,4.30)
NIE <sub>2</sub> -110	2.96	(1.05,4.87)	3.03	(1.09,4.97)	3.06	(1.10,5.01)
NIE <sub>2</sub> -101	3.10	(1.14,5.06)	3.18	(1.14,5.22)	3.19	(1.13,5.25)
NIE <sub>2</sub> -011	2.93	(1.43,4.44)	2.94	(1.46,4.41)	2.93	(1.46,4.40)
NIE <sub>2</sub> -111	3.04	(1.09,4.98)	3.04	(1.04,5.04)	3.03	(1.04,5.03)
SNIE <sub>2</sub>	2.96	(1.56,4.36)	2.97	(1.56,4.38)	2.97	(1.55,4.39)
$\sqrt{\text{var-NIE}_2}$	0.08		0.10		0.10	
IE <sub>2</sub> <sup>nointer</sup>	2.34	(1.27,3.41)	2.34	(1.24,3.45)	2.34	(1.26,3.42)
NIE <sub>12</sub> -000	-0.05	(-0.17,0.08)	-0.04	(-0.25,0.17)	-0.03	(-0.32,0.25)
NIE <sub>12</sub> -100	-0.05	(-0.20,0.10)	0.01	(-0.20,0.23)	0.04	(-0.24,0.32)
NIE <sub>12</sub> -010	-0.12	(-0.29,0.06)	-0.14	(-0.44,0.17)	-0.13	(-0.51,0.24)
NIE <sub>12</sub> -001	-0.05	(-0.20,0.10)	0.01	(-0.20,0.23)	0.04	(-0.24,0.32)
NIE <sub>12</sub> -110	-0.12	(-0.34,0.09)	0000	(-0.35,0.36)	0.06	(-0.36,0.48)
NIE <sub>12</sub> -101	-0.05	(-0.21,0.10)	-0.12	(-0.44,0.20)	-0.14	(-0.53,0.26)
NIE <sub>12</sub> -011	-0.05	(-0.17,0.08)	-0.06	(-0.26,0.15)	-0.05	(-0.32,0.21)
NIE <sub>12</sub> -111	-0.05	(-0.20,0.10)	0.01	(-0.20,0.23)	0.04	(-0.24,0.32)
SNIE <sub>12</sub>	-0.08	(-0.22,0.05)	-0.07	(-0.22,0.07)	-0.06	(-0.22,0.09)
$\sqrt{\text{var-NIE}_{12}}$	0.04		0.08		0.10	
IE <sub>12</sub> <sup>nointer</sup>	-0.07	(-0.19,0.04)	-0.08	(-0.19,0.03)	-0.07	(-0.19,0.04)

**Table 5**

Estimates, SEs, and 95% confidence intervals for the total causal effect (TCE), followed by each of the mediator-specific effects we have defined. All estimates are for mean differences in SBP measured in mmHg. The results are given for three values of the sensitivity parameter  $\kappa$ : 1, 0.5 and 0.

Effect	$\kappa = 1$		$\kappa = 0.5$		$\kappa = 0$	
	Estimate	95% CI	Estimate	95% CI	Estimate	95% CI
TCE	7.63	(3.89,11.37)	7.62	(4.04,11.20)	7.64	(4.05,11.24)
MS <sup>1</sup> -NDE-00	5.08	(1.38,8.78)	5.08	(1.35,8.81)	5.09	(1.53,8.66)
MS <sup>1</sup> -NDE-10	5.11	(1.46,8.76)	4.96	(1.28,8.65)	4.91	(1.39,8.44)
MS <sup>1</sup> -NDE-01	5.24	(1.48,9.01)	5.17	(1.45,8.90)	5.16	(1.52,8.80)
MS <sup>1</sup> -NDE-11	5.21	(1.46,8.95)	5.21	(1.53,8.89)	5.21	(1.62,8.80)
MS <sup>1</sup> -NIE <sub>1</sub> -00	-0.39	(-100,0.21)	-0.39	(-1.06,0.28)	-0.36	(-1.09,0.37)
MS <sup>1</sup> -NIE <sub>1</sub> -10	-0.36	(-1.01,0.28)	-0.50	(-1.13,0.13)	-0.54	(-1.22,0.14)
MS <sup>1</sup> -NIE <sub>1</sub> -01	-0.39	(-100,0.23)	-0.41	(-1.03,0.22)	-0.39	(-1.03,0.25)
MS <sup>1</sup> -NIE <sub>1</sub> -11	-0.42	(-1.09,0.24)	-0.37	(-1.10,0.36)	-0.34	(-1.12,0.45)
MS <sup>1</sup> -NIE <sub>2</sub> -00	2.81	(1.35,4.26)	2.82	(1.41,4.23)	2.82	(1.38,4.26)
MS <sup>1</sup> -NIE <sub>2</sub> -10	2.97	(1.09,4.86)	2.91	(0.98,4.84)	2.89	(0.98,4.8)
MS <sup>1</sup> -NIE <sub>2</sub> -01	2.82	(1.35,4.28)	2.80	(1.39,4.21)	2.80	(1.36,4.24)
MS <sup>1</sup> -NIE <sub>2</sub> -11	2.91	(1.03,4.79)	3.04	(1.10,4.99)	3.09	(1.11,5.08)
MS <sup>2</sup> -NDE-00	5.08	(1.38,8.78)	5.08	(1.35,8.81)	5.09	(1.53,8.66)
MS <sup>2</sup> -NDE-10	5.10	(1.44,8.77)	5.11	(1.43,8.79)	5.11	(1.59,8.62)
MS <sup>2</sup> -NDE-01	5.25	(1.47,9.03)	5.25	(1.52,8.98)	5.26	(1.62,8.90)
MS <sup>2</sup> -NDE-11	5.21	(1.46,8.95)	5.21	(1.53,8.89)	5.21	(1.62,8.80)
MS <sup>2</sup> -NIE <sub>1</sub> -00	-0.51	(-1.25,0.23)	-0.53	(-1.27,0.22)	-0.49	(-1.25,0.26)
MS <sup>2</sup> -NIE <sub>1</sub> -10	-0.49	(-1.27,0.29)	-0.50	(-1.29,0.29)	-0.48	(-1.30,0.34)
MS <sup>2</sup> -NIE <sub>1</sub> -01	-0.43	(-1.10,0.24)	-0.45	(-1.12,0.23)	-0.42	(-1.11,0.27)
MS <sup>2</sup> -NIE <sub>1</sub> -11	-0.48	(-1.2,0.24)	-0.49	(-1.23,0.25)	-0.47	(-1.21,0.27)
MS <sup>2</sup> -NIE <sub>2</sub> -00	2.85	(1.39,4.32)	2.86	(1.45,4.27)	2.85	(1.41,4.30)
MS <sup>2</sup> -NIE <sub>2</sub> -10	3.02	(1.11,4.94)	3.03	(1.06,4.99)	3.02	(1.06,4.99)
MS <sup>2</sup> -NIE <sub>2</sub> -01	2.93	(1.43,4.44)	2.94	(1.46,4.41)	2.93	(1.46,4.40)
MS <sup>2</sup> -NIE <sub>2</sub> -11	3.04	(1.09,4.98)	3.04	(1.04,5.04)	3.03	(1.04,5.03)

for the true exposure, leading to an inflation of the estimated indirect effect. A feature of the Izhevsk Family Study, not exploited here, is that extremely rich information was collected (from both the subjects and a proxy) on the types, quantities and patterns of alcohol consumption. In these analyses we used only the information on estimated total ethanol consumption in one year and simplified it into a binary variable (heavy/not heavy). Concerns that the indirect effect through GGT could be inflated due to GGT's role as a good proxy for true alcohol exposure could potentially be reduced by incorporating more of the collected alcohol information.

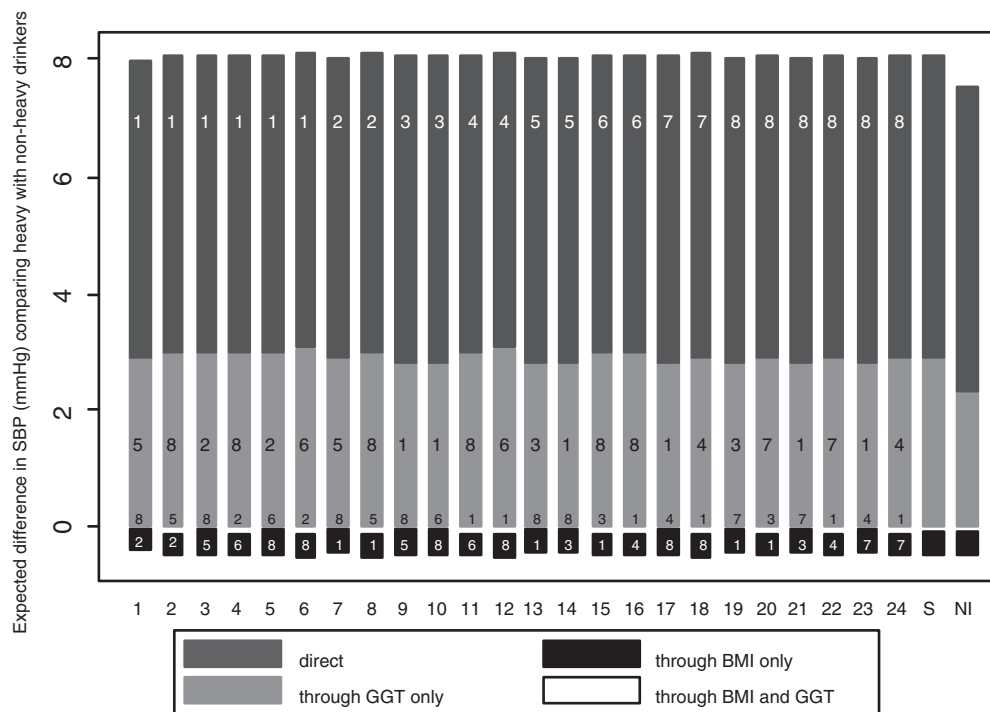
In this setting, assumptions (T.2), (MC.2), and (MN.4) imply that age, SES and smoking are sufficient to control for confounding of the alcohol-BMI, alcohol-GGT and alcohol-SBP relationships, and that BMI and alcohol, in addition to these baseline confounders are sufficient to control for confounding of the GGT-SBP relationship. In addition, we assume that all the specified parametric models are correctly specified, and that the assumptions made regarding the missing data mechanisms justified.

## 7. Concluding Remarks

Researchers are often interested in a decomposition into multiple path-specific effects through many mediators, but due to the focus in the causal inference literature primarily on one mediator, multiple mediator analyses are typically performed using LSEM, ignoring interactive and nonlinear effects, and often ignoring the effect of one mediator on another. We have shown that extending the mediation framework to multiple mediators gives rise to complexities (in terms of multiplicity of definitions) and challenges (for identification) beyond what might have been anticipated. As well as outlining these, we have provided suggestions on how to proceed in practice, via coarser decompositions and summary effects. Important future developments include extending semiparametric estimation approaches to estimate the effects defined here.

## 8. Supplementary Materials

Web Appendices, Tables and Figures, referenced in Sections 1, 3.4–3.6, 4.2, and 6.1–6.3 are available with this paper at the *Biometrics* website on Wiley Online Library.



**Figure 2.** With  $\kappa = 1$  (perfect correlation between  $M_1(0)$  and  $M_1(1)$  given **C**), all 24 possible decompositions of the total causal effect of heavy drinking on SBP into four path-specific components: a direct effect unmediated by BMI or GGT, an indirect effect via BMI alone, an indirect effect via GGT alone, and an indirect effect via both BMI and GGT. The numbers superimposed on the bars represent the code for that effect type (as defined in the caption of Table 2). The numbers along the  $x$ -axis represent the decomposition number, also defined in Table 2.

#### ACKNOWLEDGEMENTS

We thank the Izhevsk Family Study, funded by the Wellcome Trust, for providing the data for the illustrative example, and Professor David Leon for his guidance on this aspect of the work. RhMD is supported by the UK Medical Research Council (Career Development Award in Biostatistics, G1002283). BLDS is supported in part by the UK Economic and Social Research Council (NCRM Pathways node, ES/I025561/1). SV would like to thank the Flemish Research Council (FWO) for financial support (Grant G.0111.12). The article has benefited greatly from suggestions and comments made by two anonymous referees, an Associate Editor and one of the Editors of the journal; we thank all four for their input.

#### REFERENCES

- 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 19th Joint Conference on Artificial Intelligence*, 357–363. San Francisco, CA: Morgan Kaufmann Publishers Inc.
- 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.
- Bentler, P. M. (1980). Multivariate analysis with latent variables: Causal modeling. *Annual Review of Psychology* **31**, 419–456.
- Cole, S. R. and Frangakis, C. E. (2009). The consistency statement in causal inference: A definition or an assumption? *Epidemiology* **20**, 3–5.
- Daniel, R. M., De Stavola, B. L., and Cousens, S. N. (2011). g-formula: Estimating causal effects in the presence of time-varying confounding or mediation using the g-computation formula. *The Stata Journal* **11**, 479–517.
- Daniel, R. M., De Stavola, B. L., Cousens, S. N., and Vansteelandt, S. (2014). A review of causal mediation analysis with one mediator. Technical Report, Department of Medical Statistics, London School of Hygiene and Tropical Medicine.
- Daniels, M. J., Roy, J. A., Kim, C., Hogan, J. W., and Perri, M. G. (2012). Bayesian inference for the causal effect of mediation. *Biometrics* **68**, 1028–1036.
- Imai, K., Keele, L., and Tingley, D. (2010). A general approach to causal mediation analysis. *Psychological Methods* **15**, 309–334.
- Imai, K., Keele, L., and Yamamoto, T. (2010). Identification, inference, and sensitivity analysis for causal mediation effects. *Statistical Science* **25**, 51–71.
- Imai, K. and Yamamoto, T. (2013). Identification and sensitivity analysis for multiple causal mechanisms: Revisiting evidence from framing experiments. *Political Analysis* **21**, 141.
- Leon, D. A., Saburova, L., Tomkins, S., Andreev, E., Kiryanov, N., McKee, M., and Shkolnikov, V. M. (2007). Hazardous alcohol drinking and premature mortality in Russia: A population based case-control study. *Lancet* **16**, 2001–2009.
- MacKinnon, D. P. (2000). Contrasts in multiple mediator models. In: *Multivariate Applications in Substance Use Research*, Rose, J. S., et al. (eds), 141–160, Mahwah, NJ: Lawrence Erlbaum Associates Publishers.

- Neyman, J. (1923). On the application of probability theory to agricultural experiments. Essay on principles. Section 9. *Statistical Science* **5**, 465–472.
- Pearl, J. (2001). Direct and indirect effects. In *Proceedings of the 17th Conference in Uncertainty in Artificial Intelligence*, 411–420. San Francisco, CA: Morgan Kaufmann Publishers Inc.
- Pearl, J. (2009). *Causality*, 2nd edition. New York: Cambridge University Press.
- Preacher, K. J. and Hayes, A. F. (2008). Asymptotic and resampling strategies for assessing and comparing indirect effects in multiple mediator models. *Behavior Research Methods* **40**, 879–891.
- Robins, J. M. (1986). A new approach to causal inference in mortality studies with a sustained exposure period—Application to control of the healthy worker survivor effect. *Mathematical Modelling* **7**, 1393–1512.
- Robins, J. M. and Greenland, S. (1992). Identifiability and exchangeability for direct and indirect effects. *Epidemiology* **3**, 143–155.
- Robins, J. M., Mark, S. D., and Newey, W. K. (1992). Estimating exposure effects by modelling the expectation of exposure conditional on confounders. *Biometrics* **48**, 479.
- Robins, J. M. and Wasserman, L. (1997). Estimation of effects of sequential treatments by reparameterizing directed acyclic graphs. In *Proceedings of the 13th Joint Conference on Artificial Intelligence*, 409–420. San Francisco, CA: Morgan Kaufmann Publishers Inc.
- Roy, J., Hogan, J. W., and Marcus, B. H. (2008). Principal stratification with predictors of compliance for randomized trials with 2 active treatments. *Biostatistics* **9**, 277–289.
- Rubin, D. B. (1976). Inference and missing data. *Biometrika* **63**, 581–592.
- Rubin, D. B. (1978). Bayesian inference for causal effects: The role of randomization. *The Annals of Statistics* **6**, 34–58.
- 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**, 1816–1845.
- Ten Have, T. R. and Joffe, M. (2012). A review of causal estimation of effects in mediation analyses. *Statistical Methods in Medical Research* **21**, 77–107.
- van der Laan, M. J. and Petersen, M. L. (2008). Direct effect models. *International Journal of Biostatistics* **4**, Article 23.
- VanderWeele, T. (2009). Marginal structural models for the estimation of direct and indirect effects. *Epidemiology* **20**, 18–26.
- VanderWeele, T. J. (2013). A three-way decomposition of a total effect into direct, indirect, and interactive effects. *Epidemiology* **24**, 224–232.
- VanderWeele, T. J. and Chiba, Y. (2014). Sensitivity analysis for direct and indirect effects in the presence of a mediator-outcome confounder that may be affected by the exposure. *Epidemiology Biostatistics and Public Health* **11**.
- VanderWeele, T. 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. and Vansteelandt, S. (2014). Mediation analysis with multiple mediators. *Epidemiologic Methods* **2**, 95–115.
- Vansteelandt, S., Bekaert, M., and Lange, T. (2012). Imputation strategies for the estimation of natural direct and indirect effects. *Epidemiologic Methods* **1**, 131–158.
- Vansteelandt, S. and VanderWeele, T. J. (2013). Natural direct and indirect effects on the exposed: Effect decomposition under weaker assumptions. *Biometrics* **68**, 1019–1027.
- VanderWeele, T. J., Vansteelandt, S., and Robins, J. M. (2014). Methods for effect decomposition in the presence of an exposure-induced mediator-outcome confounder. *Epidemiology* **25**, 300–306.
- Westreich, D., Cole, S. R., Young, J. G., Palella, F., Tien, P. C., Kingsley, L., Gange, S. J., and Hernán, M. A. (2012). The parametric g-formula to estimate the effect of highly active antiretroviral therapy on incident AIDS or death. *Statistics in Medicine* **31**, 2000.
- Wright, S. (1921). Correlation and Causation: Part I—Method of path coefficients. *Journal of Agriculture Research* **20**, 557–575.
- Young, J. G., Cain, L. E., Robins, J. M., O’Reilly, E. J., and Hernán, M. A. (2011). Comparative effectiveness of dynamic treatment regimes: An application of the parametric g-formula. *Statistics in Biosciences* **3**, 119–143.
- Zheng, W. and van der Laan, M. J. (2012). Targeted maximum likelihood estimation of natural direct effects. *International Journal of Biostatistics* **8**, Article 1.

Received June 2013. Revised August 2014.

Accepted September 2014.