American Journal of Epidemiology

Vol. 186, No. 2
DOI: 10.1093/aje/kwx051
Advance Access publication:
June 20, 2017

Practice of Epidemiology

Flexible Mediation Analysis With Multiple Mediators

Johan Steen*, Tom Loeys, Beatrijs Moerkerke, and Stijn Vansteelandt

* Correspondence to Dr. Johan Steen, Department of Intensive Care, Ghent University Hospital, De Pintelaan 185, 9000 Ghent, Belgium (e-mail: johan.steen@ugent.be).

Initially submitted March 9, 2016; accepted for publication August 4, 2016.

The advent of counterfactual-based mediation analysis has triggered enormous progress on how, and under what assumptions, one may disentangle path-specific effects upon combining arbitrary (possibly nonlinear) models for mediator and outcome. However, current developments have largely focused on single mediators because required identification assumptions prohibit simple extensions to settings with multiple mediators that may depend on one another. In this article, we propose a procedure for obtaining fine-grained decompositions that may still be recovered from observed data in such complex settings. We first show that existing analytical approaches target specific instances of a more general set of decompositions and may therefore fail to provide a comprehensive assessment of the processes that underpin cause-effect relationships between exposure and outcome. We then outline conditions for obtaining the remaining set of decompositions. Because the number of targeted decompositions increases rapidly with the number of mediators, we introduce natural effects models along with estimation methods that allow for flexible and parsimonious modeling. Our procedure can easily be implemented using off-the-shelf software and is illustrated using a reanalysis of the World Health Organization's Large Analysis and Review of European Housing and Health Status (WHO-LARES) study on the effect of mold exposure on mental health (2002–2003).

causal inference; decomposition; dichotomous outcome; epidemiologic methods; flexible modeling; mediation analysis; multiple mediators

Abbreviations: DAG, directed acyclic graph; SEM, structural equation model; WHO-LARES, World Health Organization's Large Analysis and Review of European Housing and Health Status.

Mediation analysis is widely conducted to deepen understanding of the mechanisms behind established cause-effect relationships. It does so by separating the indirect effect that operates through a given intermediate (or mediator) from the remaining direct effect and by quantifying their respective contributions to the overall exposure effect. Epidemiologists often focus on multiple mediators, either because interest lies in multiple mechanisms or because the association between the mediator of interest and the outcome is confounded by an earlier intermediate. However, as the number of definable causal pathways from exposure to outcome grows exponentially with an increasing number of mediators being considered, so does the complexity related to their identification and estimation (1).

Although analyses with multiple mediators have a long tradition in the structural equation model (SEM) literature, complications related to effect decomposition have long been obscured because SEM-based definitions of path-specific effects rely on stringent parametric constraints (2). Recent contributions building on the counterfactual framework have helped to reveal intricacies related to nonparametric identification of path-specific effects (3). Accordingly, counterfactual-based approaches to effect decomposition in the presence of causally ordered mediators have been put forward. These approaches have mainly illustrated that progress can be made either by incorporating sensitivity analyses to obtain the finest possible decomposition (1, 4) or by focusing on coarser decompositions that require weaker assumptions (5, 6).

In the current article, we extend this second line of research by proposing a simple estimation procedure for effect decomposition in the presence of causally ordered mediators. Such settings give rise to a large number of possible decompositions (1). For instance, applications with only 3 sequential mediators

already yield 24 possible ways of partitioning the total causal effect into path-specific effects that can be identified, under certain conditions, without imposing parametric restrictions. Existing approaches (5) are limited because they recover only a subset of all such targeted decompositions. They may therefore give an incomplete assessment of the processes that underlie cause-effect relationships, especially in the presence of interaction. The multitude of possible decompositions, however, calls for parsimonious modeling strategies. We therefore extend so-called natural effects models (7, 8), a class of marginal structural models for mediation analysis, along with accompanying fitting strategies. Besides parsimony, our procedure offers greater modeling flexibility than prevailing Monte Carlo approaches (1, 4). For didactic purposes, we present our approach for 2 sequential mediators, although it easily extends to more mediators (see Web Appendix 1, available at https://academic.oup.com/aje).

EFFECT DECOMPOSITION INTO PATH-SPECIFIC EFFECTS

Decomposition in a single-mediator setting

Notation, definitions, and identification. Within the counterfactual framework, causal effects are defined by comparing counterfactual outcomes under different exposure regimes. The total effect of a binary exposure ($A = 1$ for exposed, $A = 0$ for unexposed) on an outcome Y is obtained by contrasting $Y(1)$ and $Y(0)$, with $Y(a)$ the counterfactual outcome that would be observed if A were set, possibly contrary to the fact, to a. The population-average exposure effect (i.e., the exposure effect for the target population; also referred to as the average treatment effect) can then be expressed in terms of mean differences $E\{Y(1) - Y(0)\}$, relative risks $P\{Y(1) = 1\}/P\{Y(0) = 1\}$, etc.

Expressions for direct and mediated effects can similarly be obtained by invoking nested counterfactuals $Y(a,M(a'))$. For instance, one can isolate part of the effect that is transmitted by M by leaving the exposure unchanged at $A = 1$ but changing the mediator from $M(1)$—the natural value it would have taken under exposure—to $M(0)$, the value it would have taken under no exposure. Comparison of nested counterfactuals $Y(1,M(1))$ and $Y(1,M(0))$ is central to the definition of natural indirect effects (9, 10). Definitions of natural direct effects can similarly be obtained by comparing $Y(1,M(0))$ and $Y(0,M(0))$. This contrast captures the intuitive notion of blocking the exposure's effect on the mediator by keeping the latter fixed at the level it would have taken in the absence of exposure.

Natural effects combine to produce the total effect, irrespective of the scale of interest or the presence of interactions or nonlinearities. For instance, on the additive scale, the total causal effect decomposes into the sum of the natural direct and indirect effect

$$E\{Y(1) - Y(0)\} = E\{Y(1,M(0)) - Y(0,M(0))\} + E\{Y(1,M(1)) - Y(1,M(0))\},$$

given the composition assumption that $Y(a,M(a)) = Y(a)$.

Nonparametric identification of natural effects can be obtained under a set of sufficient conditions (11), which state that for any value of a, a', and m

$$Y(a,m) \perp\!\!\!\perp A \mid C \quad (1)$$

$$Y(a,m) \perp\!\!\!\perp M \mid A = a, C \quad (2)$$

$$M(a) \perp\!\!\!\perp A \mid C \quad (3)$$

$$Y(a,m) \perp\!\!\!\perp M(a') \mid C, \quad (4)$$

where $U \perp\!\!\!\perp V \mid W$ denotes that U and V are independent conditional on W.

These conditions require a set of measured baseline covariates C that suffices to deconfound not only (i) the effect of exposure A on outcome Y and (ii) the effect of mediator M on outcome Y conditional on exposure A, as dictated in the SEM literature (12), but also (iii) the effect of exposure A on mediator M. Assumption 4 is a strong assumption, commonly referred to as Pearl's (9) "cross-world" independence assumption. If the data are assumed to be generated from a nonparametric SEM with independent errors (13), assumptions 1–4 can be shown to hold if, in addition to conditions i–iii, (iv) none of the mediator-outcome confounders are affected by exposure. In this article, we will further discuss identification conditions, such as i–iv, as represented in causal directed acyclic graphs (DAGs) (such as Figure 1) interpreted as nonparametric SEMs with independent errors.

Natural effects models. Natural direct and indirect effects can be parameterized by so-called natural effects models (7, 8, 14). These express the mean of nested counterfactuals in terms of hypothetical exposure levels a and a' and therefore naturally extend marginal structural models to allow for effect decomposition. For instance, in the following saturated model for a binary exposure A

$$E\{Y(a,M(a'))\} = \beta_0 + \beta_1 a + \beta_2 a' + \beta_3 a a', \quad (5)$$

for a, a' equal to 0 or 1, β_1 and $\beta_2 + \beta_3$ respectively capture the natural direct and indirect effect as expressed above, that is,

$$E\{Y(1,M(0)) - Y(0,M(0))\} = \beta_1,$$

$$E\{Y(1,M(1)) - Y(1,M(0))\} = \beta_2 + \beta_3.$$

This 2-way decomposition of the total effect ($\beta_1 + \beta_2 + \beta_3$) into the so-called pure direct and total indirect effect is not unique (10). A different decomposition into the so-called total direct and pure indirect effect arises from differently apportioning the interaction term β_3 as follows:

$$E\{Y(1,M(1)) - Y(0,M(1))\} = \beta_1 + \beta_3,$$

$$E\{Y(0,M(1)) - Y(0,M(0))\} = \beta_2.$$

Model 5 is a special case of the wider class of generalized linear natural effects models

$$E\{Y(a,M(a'))|C^*\} = g^{-1}\{\beta^\top W(a,a',C^*)\},$$

with $W(a,a',C^*)$ a known vector with components that may depend on a, a', and (possibly) a set of baseline covariates C^* (with $C^* \subseteq C$), β an unknown parameter vector and link function $g(\cdot)$. In model 5, which encodes population-average rather than stratum-specific natural effects (i.e. conditional average natural effects within strata defined by specific levels of baseline covariates), C^* is the empty set, $\beta = (\beta_0,\beta_1,\beta_2,\beta_3)^\top$, $W(a,a',C^*) = (1,a,a',aa')^\top$, and $g(\cdot)$ is the identity link. The inclusion of a nonempty set C^* additionally enables parameterizing effect modification by baseline covariates.

Decomposition in a setting with 2 sequential mediators

In most mediation analyses, even when interest lies in a single mediator, one cannot ignore the possible presence of multiple mediators, as illustrated in the following motivating example.

Motivating example. For illustrative purposes, we revisit previous analyses (8, 15) on survey data from 5,882 adult respondents from a World Health Organization project, the Large Analysis and Review of European Housing and Health Status (WHO-LARES) project (16). These analyses focused on the effect of living in damp and moldy conditions (binary exposure A) on the risk of depression (binary outcome Y) and put forward perceived control over one's home as a putative mediating mechanism (M). Corresponding natural direct and indirect effects (via perceived control) were estimated under the assumption that available individual and housing characteristics (C) were sufficient to control for confounding so that conditions i–iii were met (as reflected by the DAG in Figure 1). Kaufman (17), however, indicated that mold exposure is likely to also cause physical illness, which may, in turn, compromise both one's sense of control and mental health. This hypothetical scenario (as reflected by the DAG in Figure 2) therefore violates assumption iv and thus hinders identification of the targeted natural effects discussed earlier. It moreover implies that both physical illness (M_1) and perceived control (M_2) act as sequential mediators, giving rise to a finest possible decomposition that involves 4 distinct pathways from exposure to outcome (i.e., pathways $A \to Y$, $A \to M_1 \to Y$, $A \to M_2 \to Y$ and $A \to M_1 \to M_2 \to Y$).

In the remainder of this section, we first outline a sequential approach that bears close resemblance to that of VanderWeele and Vansteelandt (5), starting from a coarse 2-way decomposition that is next refined into a 3-way decomposition. We then demonstrate how natural effects models can be extended to parameterize component effects of the resulting and alternative decompositions, and we articulate required identification conditions.

A sequential approach. Let $Y(a,M_1(a'),M_2(a',M_1(a')))$ be the counterfactual outcome that would be observed if A were set to a and M_1 and M_2 were set to the natural value they would have taken if A had been a'. The first stage then corresponds to a 2-way decomposition with respect to the joint mediator $\{M_1,M_2\}$, separating pathway $A \to Y$ from the remaining pathways as follows:

$$E\{Y(1) - Y(0)\}$$
$$= E\{Y(1,M_1(1),M_2(1,M_1(1))) - Y(1,M_1(0),M_2(0,M_1(0)))\} \quad (6)$$
$$+ E\{Y(1,M_1(0),M_2(0,M_1(0))) - Y(0,M_1(0),M_2(0,M_1(0)))\}. \quad (7)$$

That is, the effect transmitted along either one or both mediators, or so-called joint natural indirect effect (expression 6), is separated from the remaining effect through neither of the mediators, or the joint natural direct effect (expression 7), denoted $E_{A\to Y}(0,0)$ (see Appendix Table 1).

In a second stage, a more fine-grained, 3-way decomposition can be obtained by further partitioning expression 6 into the entire effect transmitted along M_1 and the effect transmitted along M_2 only, respectively denoted $E_{A\to M_1 Y}(1,1)$ and $E_{A\to M_2\to Y}(1,0)$ (see Appendix Table 1):

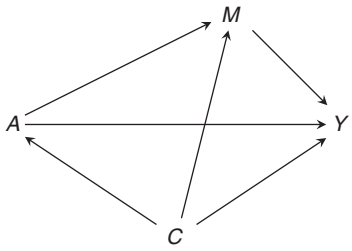

Figure 1. Causal directed acyclic graph (DAG) with exposure A, outcome Y, a single mediator M, and baseline confounders C. This DAG visually encodes causal assumptions that allow identification of natural direct and indirect effects (mediated through M) from observed data. These assumptions include no unmeasured confounding of exposure-outcome, mediator-outcome, and exposure-mediator relationships (encoded by the absence of unmeasured confounders U), and no exposure-induced confounding of the mediator-outcome relationship.

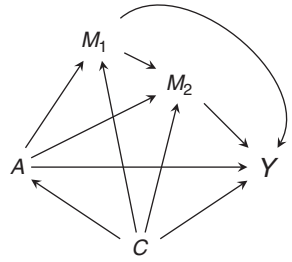

Figure 2. Causal directed acyclic graph (DAG) with exposure A, outcome Y, 2 sequential mediators M_1 and M_2, and baseline confounders C. Considering M_2 to be the mediator of interest, this DAG encodes a weaker set of causal assumptions than those encoded in the DAG in Figure 1. Because the assumption of no exposure-induced confounding of the mediator-outcome relationship is relaxed (allowing M_2 and Y to be confounded by an earlier mediator M_1), natural direct and indirect effects with respect to M_2 are no longer identifiable from observed data. In contrast, because M_1 (or the joint mediator $\{M_1, M_2\}$) is assumed not to be subject to exposure-induced confounding, natural direct and indirect effects with respect to M_1 ($\{M_1, M_2\}$) are still identifiable.

$$E\{Y(1,M_1(1),M_2(1,M_1(1))) - Y(1,M_1(0),M_2(0,M_1(0)))\}$$
$$= E\{Y(1,M_1(1),M_2(1,M_1(1))) - Y(1,M_1(0),M_2(1,M_1(0)))\} \quad (8)$$

$$+ E\{Y(1,M_1(0),M_2(1,M_1(0))) - Y(1,M_1(0),M_2(0,M_1(0)))\} \quad (9)$$

The first contrast (expression 8) captures the notion of activating all paths along M_1 that feed into Y, either directly or indirectly via M_2, while blocking all other pathways. It corresponds to the natural indirect effect as defined with respect to M_1 (i.e., along the combined pathways $A \rightarrow M_1 \rightarrow Y$ and $A \rightarrow M_1 \rightarrow M_2 \rightarrow Y$), under the composition assumption that $Y(a,M_1(a'),M_2(a,M_1(a'))) = Y(a,M_1(a'))$. The second contrast (expression 9) expresses the so-called seminatural indirect effect (18) or partial indirect effect (19) with respect to M_2 (i.e., $A \rightarrow M_2 \rightarrow Y$), in that it captures only part of the effect mediated by M_2 that bypasses M_1.

Further decomposition will generally fail without imposing strong parametric constraints, as in the linear SEM framework (3) (although see Daniel et al. (1) for a sensitivity analysis approach). Likewise, alternative decompositions of expression 6 that involve the natural indirect effect with respect to M_2 (instead of M_1—i.e., along the combined pathways $A \rightarrow M_2 \rightarrow Y$ and $A \rightarrow M_1 \rightarrow M_2 \rightarrow Y$) cannot be recovered without making certain no-interaction assumptions (19–23). These decompositions are beyond the scope of this paper (see Web Appendix 1 for a detailed overview and comparison of targeted decompositions).

Natural effects models. Natural effects models can be extended to characterize the 3-way decomposition of the previous section. For instance, in the following saturated natural effects model for a binary exposure A,

$$E\{Y(a,M_1(a'),M_2(a'',M_1(a')))\}$$
$$= \theta_0 + \theta_1 a + \theta_2 a' + \theta_3 a'' + \theta_4 aa' + \theta_5 aa''$$
$$+ \theta_6 a'a'' + \theta_7 aa'a'', \quad (10)$$

for a, a', and a'' equal to 0 or 1, the total effect, $\sum_{i=1}^{7} \theta_i$, can be partitioned into the joint natural direct effect

$$E_{A\rightarrow Y}(0,0) = \theta_1,$$

and the joint natural indirect effect

$$E_{A\rightarrow M_1Y}(1,1) + E_{A\rightarrow M_2\rightarrow Y}(1,0) = \sum_{i=2}^{7} \theta_i.$$

The latter can be further partitioned into the natural indirect effect with respect to M_1

$$E_{A\rightarrow M_1Y}(1,1) = \theta_2 + \theta_4 + \theta_6 + \theta_7,$$

and the partial indirect effect with respect to M_2 (see Appendix Table 1)

$$E_{A\rightarrow M_2\rightarrow Y}(1,0) = \theta_3 + \theta_5.$$

Model 10 is a special case of the wider class of generalized linear natural effects models for 3-way decomposition

$$E\{Y(a,M_1(a'),M_2(a'',M_1(a')))|C^*\}$$
$$= g^{-1}\{\theta^{\top} W(a,a',a'',C^*)\},$$

with $W(a,a',a'',C^*)$ a known vector with components that may depend on a, a', a'', and (possibly) covariates C^*.

Different ways of accounting for the interaction terms θ_4 to θ_7 yield another 5 possible decompositions, listed in Appendix Table 2. For instance, θ_4 can be apportioned to either $E_{A\rightarrow Y}$ or $E_{A\rightarrow M_1Y}$. Similarly, θ_5 can be apportioned to $E_{A\rightarrow Y}$ or $E_{A\rightarrow M_2\rightarrow Y}$, θ_6 to $E_{A\rightarrow M_1Y}$, or $E_{A\rightarrow M_2\rightarrow Y}$ and θ_7 to any of the 3 components. VanderWeele and Vansteelandt (5) focus only on the first 2 decompositions shown in Appendix Table 2 as their sequential approach builds on identification of $E_{A\rightarrow Y}(0,0)$ and $E_{A\rightarrow M_1Y}(1,1)$, as outlined in the previous section. The remaining 4 decompositions involve instances of $E_{A\rightarrow Y}(a',a'')$ with $a' \neq a''$ and instances of $E_{A\rightarrow M_1Y}(a, a'')$ with $a \neq a''$, which require slightly stronger identification assumptions, as discussed next.

Identification. Two-way decomposition into joint natural direct and indirect effects can be obtained if assumptions (1)–(4) hold with respect to the joint mediator $\{M_1,M_2\}$. We refer to the corresponding conditions in nonparametric SEMs with independent errors as i′–iv′.

Such first-stage decomposition can be obtained for the DAG in Figure 2, but also for the DAGs in Figure 3A and 3B. This may come as a surprise because the effect of M_1 on M_2 is confounded either by an unmeasured confounder U (Figure 3A) or (measured) intermediate confounder L (Figure 3B). However, this does not hinder identification of the joint natural direct and indirect effect because conditions i′–iv′ do not impose restrictions on the structural relationship between the mediators. The other DAGs, however, do not enable such 2-way decomposition. In Figures 3C and 3D, ii′ and iv′ are violated because of unmeasured confounding by U and intermediate confounding by L, respectively.

All 6 of the 3-way decompositions in Appendix Table 2 can be recovered under nonparametric SEMs with independent errors if, in addition to i′–iv′, (v′) the effect of M_1 on M_2 is unconfounded within strata of $\{A,C\}$, and (vi′) none of the M_1–M_2 confounders are affected by exposure. In contrast to assumptions i′–iv′, v′ and vi′ do not allow for unmeasured or intermediate confounding of the effect of M_1 on M_2. Consequently, these assumptions are violated in all DAGs in Figure 3 but not the one in Figure 2. However, decomposition with respect to the 3 sequential mediators L, M_1, and M_2 becomes possible under more general identification conditions for multiple mediators (see Web Appendix 1).

Finally, consistent with VanderWeele and Vansteelandt (5), we show in Web Appendix 1 that the first 2 decompositions in Appendix Table 2 necessitate slightly weaker assumptions than i′–vi′. In Web Appendix 1, we also provide a more detailed and formal discussion of identification assumptions, as well as extensions to more than 2 mediators.

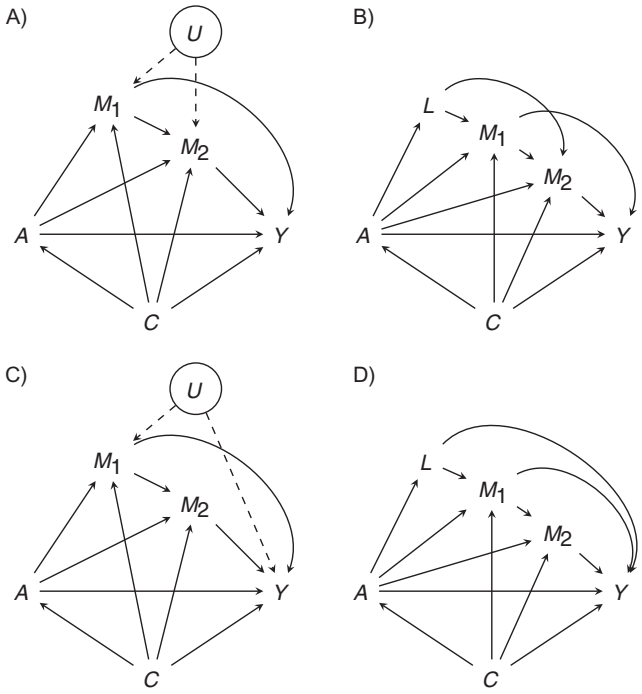

Figure 3. Variations of the causal directed acyclic graph (DAG) in Figure 2 that encode further weakening of causal assumptions (with respect to those encoded in the DAG in Figure 2). As assumptions are weakened, certain component path-specific effects may or may not be identifiable from observed data. Unobserved (A) or exposure-induced (B) confounding of the relationship between the mediators renders natural direct and indirect effect with respect to M_1 nonidentifiable but does not hinder identification of the natural direct and indirect effect with respect to the joint mediator $\{M_1, M_2\}$. Unobserved (C) or exposure-induced (D) confounding of the relationship between M_1 and Y, on the other hand, also renders natural direct and indirect effect with respect to the joint mediator $\{M_1, M_2\}$ nonidentifiable.

ESTIMATION APPROACH

Vansteelandt et al. (8) proposed an imputation procedure for fitting natural effects models for single mediators (see also Steen et al. (14) and Loeys et al. (24)). Below we describe how this procedure can be extended to recover all possible 3-way decompositions in Appendix Table 2 in settings with a binary exposure (coded 0/1) and 2 sequential mediators. We first focus on estimation of component effects as defined within strata of C, a covariate set assumed to be sufficient for conditions i′–vi′ to be met, and next describe how population-average analogs can be obtained. In Web Appendix 1 we provide some intuition as to why this procedure works and how it relates to Monte Carlo procedures based on generalizations of Pearl's (9, 25) mediation formula (1, 4).

1. Fit a suitable model for the probability (density) of either
   a. the first mediator conditional on exposure and covariate set C; for instance, a logistic regression model for binary M_1

$$\text{logit}\,P(M_1 = 1|A,C) = \beta_0 + \beta_1 A + \beta_2^\top C \qquad (11)$$

   b. or the second mediator conditional on exposure, the first mediator and covariate set C; for instance, a linear regression model for normally distributed M_2 with constant variance σ^2

$$f(M_2 \mid A,M_1,C) = N(\gamma_0 + \gamma_1 A + \gamma_2 M_1 + \gamma_3 A M_1 + \gamma_4^\top C, \sigma^2). \qquad (12)$$

2. Fit a suitable model for the outcome mean conditional on exposure, both mediators and covariate set C; for instance, a logistic regression model for binary outcome Y

$$\text{logit}\,P(Y = 1|A,M_1,M_2,C) = \delta_0 + \delta_1 A + \delta_2 M_1 + \delta_3 M_2 + \delta_4 A M_1 + \delta_5 A M_2 + \delta_6 M_1 M_2 + \delta_7 A M_1 M_2 + \delta_8^\top C. \qquad (13)$$

3. Construct an extended data set by replicating the observed data set 4 times. A similar step has previously been described by Lange et al. (26) and is best understood in terms of sequential duplication. For the first duplication, add 3 auxiliary variables a, a', and a''. Let a take on the value of the observed exposure A_i for the first replication and of the counterfactual exposure $1 - A_i$ for the second replication (for each individual i). Let both a' and a'' take on the observed exposure level for both replications. Next, duplicate the resulting extended data once again, now letting a' (a'') take on counterfactual exposure level $1 - A_i$ if model 11 (or 12) is selected as the working model (as illustrated in Web Appendix 2).
4. If model 11 is selected, compute weights

$$W_{1i,a'} = \frac{\hat{P}(M_1 = M_{1i} \mid A = a', C_i)}{\hat{P}(M_1 = M_{1i} \mid A = a'', C_i)} = \frac{\hat{P}(M_1 = M_{1i} \mid A = a', C_i)}{\hat{P}(M_1 = M_{1i} \mid A = A_i, C_i)},$$

   or, if model 12 is selected, compute weights

$$W_{2i,a''} = \frac{\hat{f}(M_2 = M_{2i} \mid A = a'', M_{1i}, C_i)}{\hat{f}(M_2 = M_{2i} \mid A = a', M_{1i}, C_i)} = \frac{\hat{f}(M_2 = M_{2i} \mid A = a'', M_{1i}, C_i)}{\hat{f}(M_2 = M_{2i} \mid A = A_i, M_{1i}, C_i)}$$

   for each row in the extended data set.
5. Impute nested counterfactuals $Y_i(a,M_{1i}(a'),M_{2i}(a'',M_{1i}(a')))$ as fitted values $\hat{E}(Y_i \mid A = a, M_{1i}, M_{2i}, C_i)$ from outcome model 13 in step 2, for each row in the extended data set.
6. Fit a natural effects model of interest for $E\{Y(a,M_{1i}(a'), M_{2i}(a'',M_{1i}(a')))|C\}$ to the extended data by regressing the imputed outcomes on a, a', a'', and C, weighting by the weights obtained in step 4.

This weighted imputation procedure is schematically displayed in Web Appendix 2. In contrast to direct application of

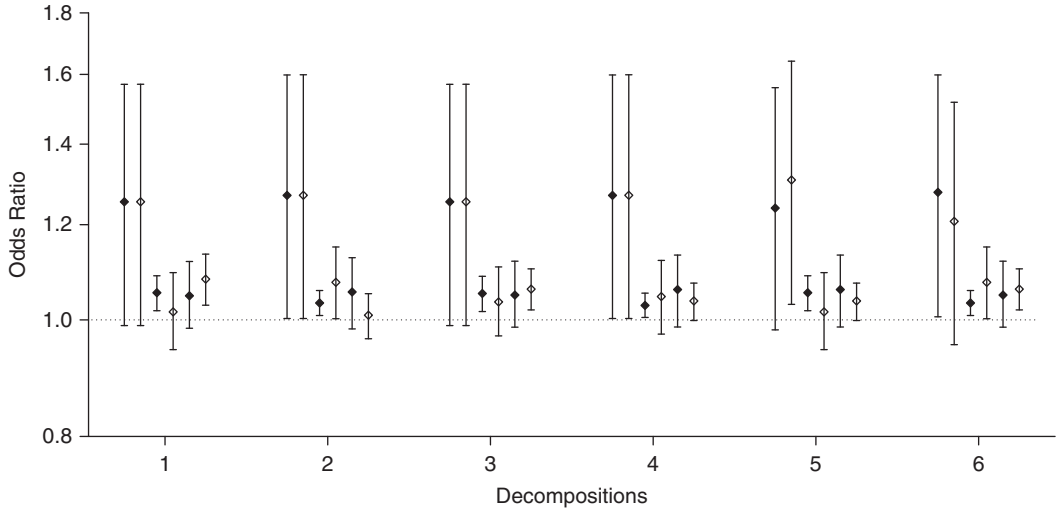

Figure 4. Odds ratio estimates and corresponding 95% confidence intervals for each of the conditional analogs of the component effects displayed in Appendix Table 2 (on the log odds ratio scale), using data from the World Health Organization's Large Analysis and Review of European Housing and Health Status, 2002–2003. Components are grouped per decomposition and displayed in the same order as in Appendix Table 2. Estimates are based on natural effects model 14, fitted upon weighting by $W_{1i,a'}$ (black diamond) or $W_{2i,a''}$ (white diamond).

the generalized mediation formula (1, 4), which relies on a model for the distribution of each of the mediators, our procedure requires only one of these models. This allows investigators to weigh by the ratio of densities of the mediator whose corresponding model they believe is less prone to misspecification. If, for instance, M_1 is binary and M_2 continuous, as in the examples given for models 11 and 12, weighting for M_1 would be most appropriate, because it allows analysts to refrain from modeling the (conditional) relationship between the mediators and making distributional assumptions.

The natural effects model from step 6 can be fitted to the weighted imputations to obtain estimates for stratum-specific component effects. If both exposure A and confounders C are discrete, saturated models can be fitted as long as C is not high-dimensional. In all other cases, our approach demands model restrictions. This improves interpretability of the results but also increases the risk of misspecification of the natural effects model, which may, in turn, lead to biased estimation of the component effects. However, as long as the structure of the imputation model chosen is sufficiently rich to minimize the risk of it being misspecified, results from an overly restrictive natural effects model may still be viewed as a useful summary (8).

Component effects within strata of C^*, a subset of C, can be obtained by fitting a natural effects model for $E\{Y(a, M_1(a'),M_2(a'',M_1(a')))|C^*\}$ conditional on a, a', a'', and C^* upon multiplying the weights from step 4 by $\hat{P}(A = A_i|C_i^*)/\hat{P}(A = A_i|C_i)$. If C^* is empty, the corresponding natural effects model encodes population-average rather than stratum-specific effects, and the numerator can possibly be replaced by 1. Inverse weighting then enables transporting results to the entire target population as it accounts for the possibly selective nature of subjects with observed exposure $A = A_i$.

Finally, standard errors and confidence intervals for this imputation estimator can be obtained using a bootstrap procedure (including steps 1–6). Bootstrapping is preferred over use of default standard errors for parameter estimates of natural effects models returned by statistical software, because the latter fail to account for uncertainty due to estimation of the working models.

Web Appendix 1 provides a detailed description of how to adapt the above procedure to continuous exposures (building on (8)), and to settings without interactions between component effects (building on an estimation procedure similar to the one described in (6)). It also explains how to implement our procedure and obtain bootstrap-based standard errors and confidence intervals in R.

In the next section, we reassess the mediating mechanisms from the empirical example introduced earlier by applying our suggested procedure to obtain a 3-way decomposition of the total effect of dampness or mold exposure (A) on the presence of depressive symptoms (Y).

MOTIVATING EXAMPLE REVISITED

Following Kaufman (17), we allow for the possibility that mold-related illness ($M_1 = 1$ in the presence of at least 1 physical condition known to be related to mold exposure or 0 otherwise) affects perceived control (M_2), as measured on a 5-point Likert scale (reverse coded), but not vice versa. The available set of covariates (C) was assumed sufficient for conditions i′–vi′ to be met. A logistic natural effects model

$$\begin{aligned}\operatorname{logit} P\{Y(a,M_1(a'),M_2(a'',M_1(a'))) = 1|C\} \\ = \eta_0 + \eta_1 a + \eta_2 a' + \eta_3 a'' + \eta_4 aa' + \eta_5 aa'' \\ + \eta_6 a'a'' + \eta_7 aa'a'' + \eta_8^{\top} C\end{aligned} \quad (14)$$

was fitted to decompose the total effect odds ratio of dampness or mold exposure A on the presence of depressive symptoms Y (conditional on baseline covariates C), which was estimated to be 1.38 (95% confidence interval: 1.09, 1.73). This was done following steps 1–6 of the previous section. First, mediator models 11 (for the probability of mold-related illness M_1) and 12 (for the density of perceived control M_2) and an extended version of outcome model 13 were fitted to the original data. The latter was used to impute nested counterfactuals in the data set that was extended according to whether model 11 or 12 was chosen to calculate regression weights for natural effects model 14. Each of the working models was specified to include all possible 2- and 3-way interactions between exposure and mediators to ensure that different decompositions resulting from model 14 appropriately reflected differences dictated by the data. For simplicity of exposition, we excluded interaction or polynomial terms involving baseline covariates. A more elaborate model focusing on effect modification by covariates, as well as a marginal natural effects model, is described in further detail in Web Appendix 1, which provides a more detailed report of the analyses of this section. A total of 1,000 bootstrap samples were drawn to calculate 95% (standard normal) bootstrap confidence intervals.

Results for all possible 3-way decompositions are displayed in Figure 4. Because different choices of working models yielded similar estimates, we report only estimates obtained upon weighting by the ratio of probabilities of M_1. The odds ratio for the joint natural direct effect, $\exp(E_{A\to Y}(0,0|C))$, was 1.25 (95% confidence interval: 0.99, 1.57). The odds of depression within a population (with specific individual and housing characteristics as defined within strata of C) would thus increase by 25% if all individuals were to be moved from a dry dwelling to a damp and moldy residence with neither their physical condition nor their sense of control over their living environment being affected by it. Its complement, the joint natural indirect effect odds ratio, was 1.10 (95% confidence interval: 1.03, 1.19). That is, if all individuals were exposed to residential dampness and mold, then the effect of changing both their physical condition and perceived control to what it would be if they were not living under such poor housing conditions would be to reduce the odds of depression by 9%. A reduction of 5% would be attributed to changing their physical condition; $\exp(\hat{E}_{A\to M_1 Y}(1,1|C)) = 1.05$ (95% CI: 1.02, 1.09). Another reduction of about 4% would be attributed to additionally changing their perceived control, in as far as earlier changes in their physical condition would not yet have done so; $\exp(\hat{E}_{A\to M_2\to Y}(1,0|C)) = 1.05$ (95% CI: 0.98, 1.12).

Natural effects model 14 not only permits estimation of the component effects but also enables probing potential interactions between causal mechanisms. For instance, a multivariate Wald test based on the bootstrap normal approximation indicated the mediating mechanisms captured by $E_{A\to M_1 Y}$ and $E_{A\to M_2\to Y}$ did not interact in their effect on the outcome—that is, the null that $\eta_6 = \eta_7 = 0$ could not be rejected at the 5% level ($\chi^2 = 1.35$; $P = 0.51$). In addition, there were no substantial differences between decompositions in Figure 4—that is, the null that $\eta_4 = \eta_5 = \eta_6 = \eta_7 = 0$ could not be rejected at the 5% level ($\chi^2 = 3.43$; $P = 0.49$).

Table 1. Estimates and of the Component Effects Odds Ratios[a], Using Data From the World Health Organization's Large Analysis and Review of European Housing and Health Status, 2002–2003

Component	Weighted by $W_{1i,a'}$		Weighted by $W_{2i,a''}$	
	Estimate	95% CI	Estimate	95% CI
$\exp(E_{A\to Y})$[b]	1.260	1.000, 1.573	1.259	1.000, 1.571
$\exp(E_{A\to M_1 Y})$[c]	1.042	1.015, 1.069	1.041	0.995, 1.089
$\exp(E_{A\to M_2\to Y})$[c]	1.052	1.008, 1.098	1.048	1.016, 1.079

Abbreviation: CI, confidence interval.

[a] Component effects as parameterized in the following natural effects model: $\text{logit}P\{Y(a,M_1(a'),M_2(a'',M_1(a'))) = 1|C\} = \zeta_0 + \zeta_1 a + \zeta_2 a' + \zeta_3 a'' + \zeta_4^{\top} C$.

[b] The natural direct effect odds ratio of exposure to damp and moldy conditions (A) on risk of depression (Y) through neither physical condition (M_1) nor perceived control (M_2).

[c] The natural indirect effect odds ratio mediated by exposure-induced changes in physical condition (M_1).

[d] The partial indirect effect odds ratio mediated solely by exposure-induced changes in perceived control over one's home (M_2).

The absence of such interactions not only facilitates interpretation of the component effects, it may also lead to more precise estimates when fitting a natural effects model that excludes these interaction terms. However, as the estimates and their 95% confidence intervals in Table 1 suggest, this did not result in the anticipated efficiency gain. Interestingly, in the absence of interactions, one may refrain from modeling mediator densities altogether by adopting a fully imputation-based estimation procedure (see Web Appendix 1).

Finally, note that this illustrative analysis is likely oversimplistic because the assumptions encoded in the DAG in Figure 2 may well be violated. For instance, possible attempts to control mold growth, such as cleaning or ventilating the house, are possibly affected by the level of mold exposure and may, in turn, influence both mold-related illness and perceived control over one's home. The level of exposure may therefore be inherently time-varying, adding another level of complexity.

DISCUSSION

In this paper, we focused on the finest decomposition that can be obtained in settings with multiple, causally ordered mediators without introducing sensitivity parameters (1, 4, 20) or parametric assumptions, as in the SEM tradition (see De Stavola et al. (27) for a review). We pointed out that previous approaches with a similar focus yield only a subset of all possible decompositions (5). Moreover, we proposed a flexible approach for estimating component effects and derived sufficient conditions for their identification.

Our estimation approach combines imputation- and weighting-based methods to fit a novel class of natural effects models (7, 8, 14, 24) for multiple mediators. As opposed to Monte Carlo approaches (1, 4), which dictate modeling the joint density of the mediators, our approach necessitates modeling the density of only one of the mediators, enabling practitioners to

select the mediator they feel most confident about modeling correctly. In the absence of interactions between component effects, one may even avoid modeling mediator densities altogether, at the expense of an additional model for the outcome, as discussed in Web Appendix 1. This may be particularly attractive in settings with large numbers of mediators because it dramatically reduces modeling demands. Nonetheless, when the joint density is correctly specified, fully parametric Monte Carlo approaches yield more efficient estimators for the component effects. Alternatively, one could refrain from modeling the outcome and, instead, opt for an approach that exclusively relies on weighting. However, this then requires correct specification of the joint density of the mediators, as in Lange et al. (26) and Taguri et al. (28) for settings with multiple causally unrelated mediators (see also VanderWeele et al. (6) for a similar approach in settings with intermediate confounding). Unless there are major concerns for model extrapolation due to inadequate modeling of the outcome (8), we generally discourage such approaches, especially when dealing with continuous mediators, because typical issues of instability, characteristic for weighting methods, tend to be exacerbated when combining density weights for each of the mediators.

In addition to added flexibility in choice of working models, natural effects modeling owes much of its attractiveness to its parsimonious parameterization. It enables testing certain hypotheses of interest (especially those concerning effect modification by baseline covariates) that, in particular settings, cannot be tested by direct application of the mediation formula (8, 24). In our illustration, we have demonstrated that differences between decompositions listed in Appendix Table 2, captured by the interaction terms of the natural effects models, can be formally tested in a straightforward manner.

Although we have restricted our presentation to applications with only 2 sequential mediators, results can straightforwardly be extended to settings with more mediators. In Web Appendix 1, we illustrate that, in such complex settings, our set of assumptions leads to a manageable and piecemeal identification procedure. Moreover, in settings where the structural dependence between certain subsequent mediators is unclear, these groups of mediators can simply be treated as joint mediators in order to render identification assumptions of the corresponding component effects more plausible (5, 28).

ACKNOWLEDGMENTS

Author affiliations: Department of Applied Mathematics, Computer Science and Statistics, Faculty of Sciences, Ghent University, Ghent, Belgium (Johan Steen, Stijn Vansteelandt); and Department of Data Analysis, Faculty of Psychology and Educational Sciences, Ghent University, Ghent, Belgium (Tom Loeys, Beatrijs Moerkerke).

This work was supported by Research Foundation Flanders (FWO Grant G.0111.12).

We thank Vanessa Didelez, Karel Vermeulen, and 2 anonymous referees for helpful suggestions and corrections on an earlier draft of this paper (and appendices). We also thank the World Health Organization's European Centre for Environment and Health, Bonn office, for providing the Large Analysis and Review of European Housing and Health Status (LARES) data set used in this paper to illustrate the methods.

The paper reflects the authors' opinion and not necessarily the position of the World Health Organization.

Conflict of interest: none declared.

REFERENCES

1. Daniel RM, De Stavola BL, Cousens SN, et al. Causal mediation analysis with multiple mediators. *Biometrics*. 2015; 71(1):1–14.
2. Taylor AB, MacKinnon DP, Tein JY. Tests of the three-path mediated effect. *Organ Res Methods*. 2008;11(2):241–269.
3. Avin C, Shpitser I, Pearl J. Identifiability of path-specific effects. In: *Proceedings of the 19th International Joint Conference on Artificial Intelligence, IJCAI'05*. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc.; 2005:357–363.
4. Albert JM, Nelson S. Generalized causal mediation analysis. *Biometrics*. 2011;67(3):1028–1038.
5. VanderWeele TJ, Vansteelandt S. Mediation analysis with multiple mediators. *Epidemiol Methods*. 2014;2(1):95–115.
6. VanderWeele TJ, Vansteelandt S, Robins JM. Effect decomposition in the presence of an exposure-induced mediator-outcome confounder. *Epidemiology*. 2014;25(2): 300–306.
7. Lange T, Vansteelandt S, Bekaert M. A simple unified approach for estimating natural direct and indirect effects. *Am J Epidemiol*. 2012;176(3):190–195.
8. Vansteelandt S, Bekaert M, Lange T. Imputation strategies for the estimation of natural direct and indirect effects. *Epidemiol Methods*. 2012;1(1):131–158.
9. Pearl J. Direct and indirect effects. In: Breese J, Koller D, eds. *Proceedings of the Seventeenth Conference on Uncertainty in Artificial Intelligence (UAI-01)*. San Francisco, CA: Morgan Kaufmann; 2001:411–420.
10. Robins JM, Greenland S. Identifiability and exchangeability for direct and indirect effects. *Epidemiology*. 1992;3(2): 143–155.
11. VanderWeele TJ, Vansteelandt S. Conceptual issues concerning mediation, interventions and composition. *Stat Interface*. 2009;2(4):457–468.
12. Judd CM, Kenny DA. Process analysis: estimating mediation in treatment evaluations. *Eval Rev*. 1981;5(5):602–619.
13. Robins JM, Richardson TS. Alternative graphical causal models and the identification of direct effects. In: Shrout P, Keyes KM, Ornstein K, eds. *Causality and Psychopathology: Finding the Determinants of Disorders and Their Cures*. Oxford, England: Oxford University Press; 2010:103–158.
14. Steen J, Loeys T, Moerkerke B, et al. Medflex: an R package for flexible mediation analysis using natural effect models. *J Stat Softw*. 2017;76(11):doi:10.18637/jss.v076.i11.
15. VanderWeele TJ, Vansteelandt S. Odds ratios for mediation analysis for a dichotomous outcome. *Am J Epidemiol*. 2010; 172(12):1339–1348.
16. Shenassa ED, Daskalakis C, Liebhaber A, et al. Dampness and mold in the home and depression: an examination of mold-related illness and perceived control of one's home as possible depression pathways. *Am J Public Health*. 2007;97(10): 1893–1899.

17. Kaufman JS. Invited commentary: decomposing with a lot of supposing. *Am J Epidemiol*. 2010;172(12):1349–1351.
18. Pearl J. Interpretation and identification of causal mediation. *Psychol Methods*. 2014;19(4):459–481.
19. Huber M. Identifying causal mechanisms (primarily) based on inverse probability weighting. *J Appl Econ*. 2014;29(6):920–943.
20. Imai K, Yamamoto T. Identification and sensitivity analysis for multiple causal mechanisms: revisiting evidence from framing experiments. *Pol Anal*. 2013;21(2):141–171.
21. Petersen ML, Sinisi SE, van der Laan MJ. Estimation of direct causal effects. *Epidemiology*. 2006;17(3):276–284.
22. Robins JM. Semantics of causal DAG models and the identification of direct and indirect effects. In: Green PJ, Hjort NL, Richardson S, eds. *Highly Structured Stochastic Systems*. New York, NY: Oxford University Press; 2003:70–81.
23. 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.
24. Loeys T, Moerkerke B, De Smet O, et al. Flexible mediation analysis in the presence of nonlinear relations: beyond the mediation formula. *Multivariate Behav Res*. 2013;48(6): 871–894.
25. Pearl J. The mediation formula: a guide to the assessment of causal pathways in nonlinear models. In: Berzuini C, Dawid P, Bernardinelli L, eds. *Causality: Statistical Perspectives and Applications*. Chichester, UK: John Wiley & Sons; 2012: 151–179.
26. Lange T, Rasmussen M, Thygesen LC. Assessing natural direct and indirect effects through multiple pathways. *Am J Epidemiol*. 2014;179(4):513–518.
27. De Stavola BL, Daniel RM, Ploubidis GB, et al. Mediation analysis with intermediate confounding: structural equation modeling viewed through the causal inference lens. *Am J Epidemiol*. 2015;181(1):64–80.
28. Taguri M, Featherstone J, Cheng J. Causal mediation analysis with multiple causally non-ordered mediators. *Stat Methods Med Res*. 2015;doi:10.1177/0962280215615899.

(Appendix follows)

APPENDIX

Appendix Table 1. Shorthand Notation for the Component Effects From a 3-Way Decomposition in the Presence of 2 Causally Ordered Mediators M_1 and M_2 and Their Parameterization in Model 10[a,b]

$E_{A\to Y}(a',a'') = g(E\{Y(1,M_1(a'), M_2(a'',M_1(a')))\}) - g(E\{Y(0,M_1(a'), M_2(a'',M_1(a')))\})$	$= \theta_1 + \theta_4 a' + \theta_5 a'' + \theta_7 a'a''$
$E_{A\to M_1Y}(a,a'') = g(E\{Y(a,M_1(1), M_2(a'',M_1(1)))\}) - g(E\{Y(a,M_1(0), M_2(a'',M_1(0)))\})$	$= \theta_2 + \theta_4 a + \theta_6 a'' + \theta_7 aa''$
$E_{A\to M_2\to Y}(a,a') = g(E\{Y(a,M_1(a'), M_2(1,M_1(a')))\}) - g(E\{Y(a,M_1(a'), M_2(0,M_1(a')))\})$	$= \theta_3 + \theta_5 a + \theta_6 a' + \theta_7 aa'$

[a] Model 10: $E\{Y(a,M_1(a'), M_2(a'',M_1(a')))\} = \theta_0 + \theta_1 a + \theta_2 a' + \theta_3 a'' + \theta_4 aa' + \theta_5 aa'' + \theta_6 a'a'' + \theta_7 aa'a''$. In this model the link function $g(\cdot)$ is the identity link.

[b] In the motivating example, we wish to decompose the effect of living in damp and moldy conditions (binary exposure A) on the risk of depression (binary outcome Y) into $E_{A\to M_1Y}$, the part of the effect that is mediated by exposure-induced changes in physical condition (M_1), $E_{A\to M_2\to Y}$, the part that is mediated solely by exposure-induced changes in perceived control over one's home (M_2), and $E_{A\to Y}$, the remaining direct effect through neither of the putative mediators. For a binary exposure A, each of these components has 4 different instances, depending on hypothetical exposure levels a, a', and a'' (see also Appendix Table 2).

Appendix Table 2. All 6 Possible 3-Way Decompositions and Their Parameterization in Model 10[a–c]

1. $E_{A\to Y}(0,0) + E_{A\to M_1Y}(1,1) + E_{A\to M_2\to Y}(1,0)$	$= (\theta_1) + (\theta_2 + \theta_4 + \theta_6 + \theta_7) + (\theta_3 + \theta_5)$
2. $E_{A\to Y}(1,1) + E_{A\to M_1Y}(0,0) + E_{A\to M_2\to Y}(0,1)$	$= (\theta_1 + \theta_4 + \theta_5 + \theta_7) + (\theta_2) + (\theta_3 + \theta_6)$
3. $E_{A\to Y}(0,0) + E_{A\to M_1Y}(1,0) + E_{A\to M_2\to Y}(1,1)$	$= (\theta_1) + (\theta_2 + \theta_4) + (\theta_3 + \theta_5 + \theta_6 + \theta_7)$
4. $E_{A\to Y}(1,1) + E_{A\to M_1Y}(0,1) + E_{A\to M_2\to Y}(0,0)$	$= (\theta_1 + \theta_4 + \theta_5 + \theta_7) + (\theta_2 + \theta_6) + (\theta_3)$
5. $E_{A\to Y}(0,1) + E_{A\to M_1Y}(1,1) + E_{A\to M_2\to Y}(0,0)$	$= (\theta_1 + \theta_5) + (\theta_2 + \theta_4 + \theta_6 + \theta_7) + (\theta_3)$
6. $E_{A\to Y}(1,0) + E_{A\to M_1Y}(0,0) + E_{A\to M_2\to Y}(1,1)$	$= (\theta_1 + \theta_4) + (\theta_2) + (\theta_3 + \theta_5 + \theta_6 + \theta_7)$

[a] Each component on the left side of the equation is represented by a linear combination of parameters on the right side (grouped in parentheses).

[b] Model 10: $E\{Y(a,M_1(a'), M_2(a'',M_1(a')))\} = \theta_0 + \theta_1 a + \theta_2 a' + \theta_3 a'' + \theta_4 aa' + \theta_5 aa'' + \theta_6 a'a'' + \theta_7 aa'a''$.

[c] In the motivating example, we wish to decompose the effect of living in damp and moldy conditions (binary exposure A) on the risk of depression (binary outcome Y) into $E_{A\to M_1Y}$, the part of the effect that is mediated by exposure-induced changes in physical condition (M_1), $E_{A\to M_2\to Y}$, the part that is mediated solely by exposure-induced changes in perceived control over one's home (M_2), and $E_{A\to Y}$, the remaining direct effect through neither of the putative mediators.