Robust inference on effects attributable to mediators: A controlled-direct-effect-based approach for causal effect decomposition with multiple mediators

An-Shun Tai[1], Yi-Juan Du[1], Sheng-Hsuan Lin[1]*

1. Institute of Statistics, National Chiao Tung University, Hsin-Chu, Taiwan. 1001 University Road, Hsinchu, Taiwan 300

***Corresponding author**
Sheng-Hsuan Lin, MD, ScD
Institute of Statistics, National Chiao Tung University, Hsin-Chu, Taiwan
1001 University Road,
Hsinchu, Taiwan 300
Cell: +886 (3) 5712121 ext.56822
E-mail: shenglin@stat.nctu.edu.tw

Summary

Effect decomposition is a critical technique for mechanism investigation in settings with multiple causally ordered mediators. Causal mediation analysis is a standard method for effect decomposition, but the assumptions required for the identification process are extremely strong. By extending the framework of controlled direct effects, this study proposes the effect attributable to mediators (EAM) as a novel measure for effect decomposition. For policy making, EAM represents how much an effect can be eliminated by setting mediators to certain values. From the perspective of mechanism investigation, EAM contains information about how much a particular mediator or set of mediators is involved in the causal mechanism through mediation, interaction, or both. The assumptions of EAM for identification are considerably weaker than the those of causal mediation analysis. We develop a semiparametric estimator of EAM with robustness to model misspecification. The asymptotic property is fully realized. We applied EAM to assess the magnitude of the effect of hepatitis C virus infection on mortality, which was eliminated by controlling alanine aminotransferase and treating hepatocellular carcinoma.

Keywords: Controlled direct effect; Doubly robust estimator; Effect attributable to mediators; Effect decomposition; Multiple mediators; Semiparametric inference

1. Introduction

1.1. Research question

Effect decomposition is critical for investigating the mechanism of a confirmed causal effect. In settings with a single mediator, causal mediation analysis is a typical technique for effect decomposition (Pearl 2001, Robins and Greenland 1992). For settings with multiple causally ordered mediators, several methods have been proposed based on the framework of causal mediation analysis (Daniel *et al.* 2015, Fasanelli *et al.* 2019, Huang and Yang 2017, Lin 2019, Lin and VanderWeele 2017, Steen *et al.* 2017, Tai and Lin 2020, VanderWeele and Vansteelandt 2014, VanderWeele *et al.* 2014). Although these methods have contributed considerably to investigations on mechanisms, three limitations have been noted. First, the measure of each path is defined based on a cross-world counterfactual model (Avin *et al.* 2005, VanderWeele and Vansteelandt 2014), which cannot be verified through randomized controlled trials (RCTs). Second, the assumptions required for identification are strong (Albert and Nelson 2011, Avin, Shpitser and Pearl 2005, Daniel, De Stavola, Cousens and Vansteelandt 2015). For example, one assumption is that the mediator–outcome confounders cannot be affected by the exposure or by previous mediators. This assumption, however, is easily violated in longitudinal studies. Another assumption is that all mediator–mediator confounders and exposure–mediator confounders are comprehensively adjusted for. This unconfoundedness assumption, however, is also difficult to fulfil in practice. Third, and finally, causal mediation analysis answers questions such as “how much of the causal effect passes through the mediation effect of a particular mediator or set of mediators?” This question, however, differs from the substantive ones that researchers may be more interested in, such as “how much of the causal effect is attributed to a particular mediator or set of mediators, including mechanisms such as interaction and mediation?” This particular substantive question is highly related to similar scientific

questions, such as “how much of the causal effect can be eliminated by setting a particular mediator or set of mediators to a certain value or set of values?” This question is critical for policy makers to determine resource allocation when intervening in risk factors to prevent disease, especial when the exposure is difficult to modify. However, existing methods based on causal mediation analysis cannot answer such questions comprehensively.

1.2. Contributions of this study

To fill this research gap, we propose the effect attributable to mediators (EAM) as an alternative approach for effect decomposition based on the controlled direct effect (CDE) framework. CDE is a well-developed technique for effect decomposition in settings with a single mediator, in which the direct effect is measured under the intervention on a mediator. This method plays an essential role in policy making (Pearl 2001, Robins and Greenland 1992, VanderWeele 2011, VanderWeele 2013) and mechanism investigation (VanderWeele 2014). The generalized CDE for multiple mediators is defined to assess the contrast between two counterfactuals if the exposure had a different value when a particular set of mediators is intervened on. By disentangling the representations of the generalized CDEs into mediator-set-specific expressions, EAM is defined as the difference between two CDEs; defined as such, the EAM interprets the effect eliminated when specific mediators are intervened on. Furthermore, by intervening on certain mediators, EAM can assess how the exposure-induced mechanism is attributable to the remaining mediators. Similar to CDE identification, for which the requisite assumptions are considerably weaker than those for natural direct and indirect effects (VanderWeele 2011), EAM only requires two assumptions: no unmeasured confounding between the exposure and outcome and no unmeasured confounding between mediators and the outcome. The cross-world exchangeability required for natural direct and indirect effects is a strong and untestable assumption; however, the unconfoundedness assumptions required for EAM are testable. Through the requirement of fewer and testable assumptions, EAM

estimation can not only be verified by clinical trials but also be applied to a broader range of circumstances than natural direct and indirect effects can.

This study makes three substantial contributions. First, as mentioned, the proposed EAM yields an entirely new effect decomposition of the total effect (TE) in the presence of multiple mediators. In general, the analysis of EAM characterizes the mechanism of exposure on the outcome attributable to a set of mediators. By contrast, causal mediation analysis of multiple mediators mainly focuses on mediation and aims to decompose the TE into several path-specific effects (PSEs), which quantify the causal effects mediated through a certain mediation pathway (Avin, Shpitser and Pearl 2005). Similar to CDE (VanderWeele 2013), if no interaction is assumed, the proposed EAM can be regarded as an alternative to a PSE. Previous methods are either restricted to performing partial decomposition, through which all PSEs either cannot be identified (Fasanelli, Giraudo, Ricceri, Valeri and Zugna 2019, Huang and Yang 2017, Steen, Loeys, Moerkerke and Vansteelandt 2017, VanderWeele and Vansteelandt 2014, VanderWeele, Vansteelandt and Robins 2014) or are limited by either strong assumptions or changing definitions (Daniel, De Stavola, Cousens and Vansteelandt 2015, Lin 2019, Lin and VanderWeele 2017). By contrast, EAMs are derived from the complete decomposition of the TE under weaker assumptions.

Second, the proposed EAM is defined as a general form of the controlled mediated effect (CME) (VanderWeele 2011). Although the CME is a causal interpretation for mediation, two limitations arise. First, the CMEs do not sum up to the TE. Second, the complete-mediation condition is required (i.e., the effect of exposure should be completely mediated by mediators). In this article, CMEs are shown to be a subset of EAMs when the complete-mediation condition is assumed. The gap between the TE and the sum of the CMEs is another type of EAM. Moreover, identification of the CMEs may be restricted to conditions in which the mediators are causally independent, whereas the EAM is still applicable when mediators are independent

or causally ordered.

Third, we derive both parametric and semiparametric estimators, including a doubly robust semiparametric estimator for the EAM. Empirically, many confounding factors are collected in nonexperimental studies to adjust for confounding bias. Due to the problems of dimensionality, nonparametric methods are usually not practical in such cases. Recently, powerful robust semiparametric methods have been proposed in the context of mediation analysis (Goetgeluk *et al.* 2008, Tchetgen Tchetgen and Shpitser 2012). A doubly robust (DR) estimator of the EAM is constructed based on two standard approaches, the regression-based estimator and the inverse probability weighting (IPW) estimator. The proposed DR estimator operates on the union of the two semiparametric model spaces and is therefore less sensitive to model misspecification. Furthermore, our DR estimator is consistent and asymptotically normal. As demonstrated by extensive simulation studies, the DR estimator results in less bias and a higher coverage rate than the regression-based and IPW estimators do in settings where the model of the outcome or exposure is misspecified.

The remainder of this article is organized as follows: Section 2 introduces the definitions, assumptions, and identifications of the EAM with two ordered mediators. Section 3 proposes the DR estimators of the EAM and demonstrates its asymptotic properties. Section 4 presents an extension to an arbitrary number of ordered mediators. Section 5 discusses the relation between the EAM and the CME. Section 6 presents the results of a simulation study conducted to evaluate the performance of the proposed estimators. Section 7 applies the EAM to the dataset of the Risk Evaluation of Viral Load Elevation and Associated Liver Disease/Cancer (REVEAL) from Taiwan. Finally, in Section 8, we conclude our study, discussing our contributions and the limitations of our method.

2. EAMs with two causally ordered mediators

2.1. Definitions of the EAM

For simplicity, we introduce the EAM for two ordered mediators in this section. Extension to an arbitrary number of multiple ordered mediators is discussed in Section 4.

Let A denote the exposure, Y the outcome of interest, and M_1 and M_2 the two ordered mediators—where M_1 affects M_2. All confounders among A, M_1, M_2, and Y are comprehensively collected in the baseline confounder C_0, and in the time-varying confounders C_1 and C_2. The causal relationship among A, M_1, M_2, Y, and all confounders is shown in Figure 1(a). The counterfactual model (also called the potential outcome model) is introduced as follows (Little and Rubin 2000). Let $Y(a)$ be the hypothetical value of Y given that A is set equal to a; $Y(a, m_1)$ be the hypothetical value of Y given that A and M_1 are set equal to a and m_1, respectively; $Y(a, m_2)$ be the hypothetical value of Y given that A and M_2 are set equal to a and m_2, respectively; and $Y(a, m_1, m_2)$ be the hypothetical value of Y given that A, M_1, and M_2 are set equal to a, m_1, and m_1, respectively.

Robins and Greenland (1992) and Pearl (2001) have defined the CDE for a single mediator. Herein, we introduce CDEs in the presence of two mediators as follows.

Definition 1. (CDE with two mediators)
Suppose that M_1 *and* M_2 *are mediators and* a, a^*, m_1, *and* m_2 *are given values. The CDEs are defined as follows:*
$CDE_1(m_1) \equiv \varphi_1(a, m_1) - \varphi_1(a^*, m_1)$,
$CDE_2(m_2) \equiv \varphi_2(a, m_2) - \varphi_2(a^*, m_2)$, *and*
$CDE_3(m_1, m_2) \equiv \varphi_3(a, m_1, m_2) - \varphi_3(a^*, m_1, m_2)$,
where $\varphi_1(a, m_1) \equiv E(Y(a, m_1))$, $\varphi_2(a, m_2) \equiv E\big(Y(a, m_2)\big)$, *and* $\varphi_3(a, m_1, m_2) \equiv E\big(Y(a, m_1, m_2)\big)$.

Hereafter, $\varphi_1(a, m_1)$, $\varphi_2(a, m_2)$, and $\varphi_3(a, m_1, m_2)$ are referred to as the multimediation parameters. $\mathrm{CDE}_1(m_1)$ represents the effect of A on Y that is not attributable to M_1. Likewise, $\mathrm{CDE}_2(m_2)$ measures the effect of A on Y that is not attributable to M_2. Finally, $\mathrm{CDE}_3(m_1, m_2)$ represents the effect of A on Y that is not attributable to M_1 or M_2.

Additionally, the TE is required for the establishment of the EAMs and is defined as $\text{TE} \equiv \varphi_{TE}(a) - \varphi_{TE}(a^*)$, where $\varphi_{TE}(a) \equiv E(Y(a))$ (Pearl 2001).

Subsequently, we simplify the representations of the CDEs and TE into mediator-set-specific expressions, and the EAMs are defined accordingly.

Definition 2. (EAM with two mediators)
Based on the CDEs in Definition 1 and TE, the EAMs are given as follows:
$EAM_{P_0}(m_1, m_2) \equiv CDE_3(m_1, m_2)$,
$EAM_{P_1}(m_1, m_2) \equiv CDE_2(m_2) - CDE_3(m_1, m_2)$,
$EAM_{P_2}(m_1, m_2) \equiv CDE_1(m_1) - CDE_3(m_1, m_2)$, *and*
$EAM_{P_3}(m_1, m_2) \equiv TE + CDE_3(m_1, m_2) - CDE_1(m_1) - CDE_2(m_2)$,
where P_0, P_1, P_2, *and* P_3 *are the null set,* $\{M_1\}$, $\{M_2\}$, *and* $\{M_1, M_2\}$, *respectively.*

In Definition 2, $EAM_{P_0}(m_1, m_2)$, which is defined as $\text{CDE}_3(m_1, m_2)$, measures the alteration of Y directly caused by A when M_1 and M_2 are set equal to m_1 and m_2, respectively. In general, $EAM_{P_0}(m_1, m_2)$ is the effect that is not due to either interaction or mediation of M_1 and M_2.

$EAM_{P_1}(m_1, m_2)$ measures the contrast between the effect of A on Y when M_1 is set equal to m_1 and the effect of A on Y when M_1 is not intervened on and when M_2 is set equal to m_2. That is, $EAM_{P_1}(m_1, m_2)$ is the effect eliminated by setting M_1 equal to m_1 when M_2 is set equal to m_2. Conversely, $EAM_{P_1}(m_1, m_2)$ can also be interpreted as the effect solely attributable to $P_1 = \{M_1\}$ through the mechanism, which is mediation, interaction, or both. Similarly, $EAM_{P_2}(m_1, m_2)$ measures the contrast between the effect of A on Y when M_2 is set equal to m_2 and the effect of A on Y when M_2 is not intervened on when M_1 is set equal to m_1. Thus, $EAM_{P_2}(m_1, m_2)$ is the effect eliminated by setting M_2 equal to m_2 when M_1 is set equal to m_1, which can be interpreted as the effect solely attributable to $P_2 = \{M_2\}$ through the mechanism.

Finally, we provide the interpretation of $EAM_{P_3}(m_1, m_2)$. To this end, we rewrite the

formulation of $EAM_{P_3}(m_1, m_2)$ to the following:

$$EAM_{P_3}(m_1, m_2) \equiv [TE - \mathrm{CDE}_1(m_1)] - [\mathrm{CDE}_2(m_2) - \mathrm{CDE}_3(m_1, m_2)].$$

$TE - \mathrm{CDE}_1(m_1)$ represents the difference between the effect when M_1 is intervened on and the effect when M_1 is not intervened on. $\mathrm{CDE}_2(m_2) - \mathrm{CDE}_3(m_1, m_2)$ represents the difference between the effect when M_1 is intervened on and the effect when M_1 is not intervened on under the case that M_2 is intervened on as m_2. When the effect mediated sequentially through M_1 and M_2 and the effect of the $A - (M_1, M_2)$ interaction on the outcome are both zero, the difference between different settings for M_2 is zero. Thus, $EAM_{P_3}(m_1, m_2)$ captures the effect through the mechanism attributable to $P_3 = \{M_1, M_2\}$.

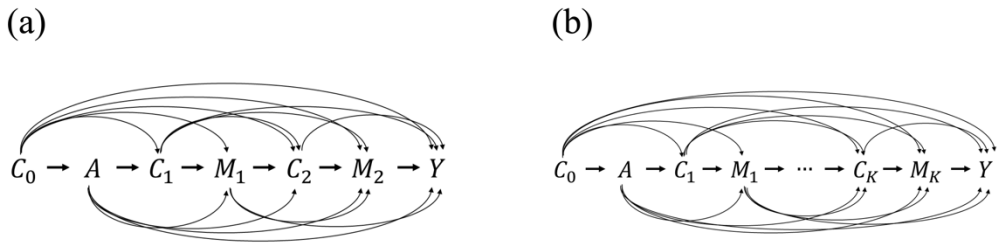

Figure 1. Causal relationships illustrated by direct acyclic graphs (DAGs). (a) Causal relationship among exposure, two causally ordered multiple mediators, outcome variable, baseline confounder, and time-varying confounders, which are denoted by A, (M_1, M_2), Y, C_0, and (C_1, C_2), respectively. (b) DAG illustrating the causal relationship for multiple mediators $(M_1, M_2, \ldots, M_K)$, in which $(C_1, C_2, C_3, \ldots, C_K)$ denotes the corresponding time-varying confounders.

2.2. Assumptions and identification

By definition, the identification of all EAMs can be achieved through the identification of $\mathrm{CDE}_1(m_1)$, $\mathrm{CDE}_2(m_2)$, $\mathrm{CDE}_3(m_1, m_2)$, and the TE. The following two sets of exchangeability assumptions must be satisfied:

(A1) No unmeasured confounding among the exposure and outcome,

i.e., (A1.1), $Y(a, m_1) \perp\!\!\!\perp A | C_0$; (A1.2), $Y(a, m_2) \perp\!\!\!\perp A | C_0$; (A1.3), $Y(a, m_1, m_2) \perp\!\!\!\perp A | C_0$; and

(A1.4) $Y(a) \perp\!\!\!\perp A | C_0$

(A2) No unmeasured confounding among the mediators and outcome,

i.e., (A2.1), $Y(a, m_1) \perp\!\!\!\perp M_1 | C_0, A, C_1$; (A2.2), $Y(a, m_2) \perp\!\!\!\perp M_2 | C_0, A, C_1, M_1, C_2$;

(A2.3), $Y(a, m_1, m_2) \perp\!\!\!\perp M_1 | C_0, A, C_1$; and (A2.4), $Y(a, m_1, m_2) \perp\!\!\!\perp M_2 | C_0, A, C_1, M_1, C_2$.

We further define the following consistency assumption:

$Y(a, m_1, m_2) = Y$, if $A = a$, $M_1 = m_1$ and $M_2 = m_2$;

$M_1(a) = M_1$, if $A = a$; and $M_2(a, m_1) = M_2$, if $A = a$ and $M_1 = m_1$.

Assuming (A1.4) and the consistency assumption, $\varphi_{TE}(a)$ can be immediately identified as $\int_{c_0} E(Y|a, c_0) dF(c_0)$. The identification of the multimediation parameters, $\varphi_1(a, m_1)$, $\varphi_2(a, m_2)$, and $\varphi_3(a, m_1, m_2)$, is detailed in Theorem 1.

Theorem 1. (Identification of multimediation parameters)

Suppose that the causal structure among A, M_1, M_2, Y, C_0, C_1, and C_2 follows the structure shown in Figure 1. Then, under (A1.1), (A2.1), and the consistency assumption, $\varphi_1(a, m_1)$ is identified as

$$\int_{c_0, c_1, c_2, m_2} E(Y|c_0, a, c_1, m_1, c_2, m_2) dF(m_2|c_0, a, c_1, m_1, c_2) \times dF(c_2|c_0, a, c_1, m_1) dF(c_1|c_0, a) dF(c_0);$$

under (A1.2), (A2.2), and the consistency assumption, $\varphi_2(a, m_2)$ is identified as

$$\int_{c_0, c_1, m_1, c_2} E(Y|c_0, a, c_1, m_1, c_2, m_2) dF(c_2|c_0, a, c_1, m_1) \times dF(m_1|c_0, a, c_1) dF(c_1|c_0, a) dF(c_0);$$

and under (A1.1), (A2.3), (A2.4), and the consistency assumption, $\varphi_3(a, m_1, m_2)$ is identified as

$$\int_{c_0, c_1, c_2} E(Y|c_0, a, c_1, m_1, c_2, m_2) dF(c_2|c_0, a, c_1, m_1)\, dF(c_1|c_0, a) dF(c_0),$$

where a, m_1, and m_2 are the prespecified values.

Theorem 1 follows from the *g* formula for causal effects (Pearl 2001, Pearl and Robins 1995, Robins 1987, Robins 1986). The proof for Theorem 1 is presented in Appendix A. The identifiability of the EAMs can be derived from the linear combinations of multimediation parameters based on Definition 2.

Similar to the identification of the CDEs (VanderWeele 2011), identifying the EAMs only requires the assumptions of no exposure–outcome confounding and no mediator–outcome

confounding. The assumptions of exposure–mediator and mediator–mediator unconfoundedness are not required. This constitutes a significant advantage for investigating causal mechanisms with multiple mediators because the exposure–mediator confounders and mediator–mediator confounders are sometimes difficult to collect comprehensively. In addition, all the EAMs—or, equivalently, φ_1, φ_2, and φ_3—can be verified through RCTs. For example, $\varphi_3(a, m_1, m_2)$ and $\varphi_3(a^*, m_1, m_2)$ can be estimated, respectively, by $E(Y|A = a, M_1 = m_1, M_2 = m_2)$ and $E(Y|A = a^*, M_1 = m_1, M_2 = m_2)$ in an RCT when A is randomly assigned to a and a^* and (M_1, M_2) is assigned to (m_1, m_2). By contrast, all causal estimands in causal mediation analysis are defined based on cross-world counterfactuals, which cannot be verified through RCTs.

3. Robust estimation

In this section, we present an estimation procedure for

$$\Phi = \{\varphi_1(a, m_1), \varphi_2(a, m_2), \varphi_3(a, m_1, m_2), \varphi_{TE}(a)\}.$$

The estimation of the EAMs can immediately be achieved through a linear transformation between the EAMs and Φ. Subsequently, we propose three approaches to develop the estimators of Φ in different model spaces. Given the corresponding model space, the estimators are consistent and asymptotically normal. Importantly, the DR estimators are less sensitive to the model misspecification that the regression-based and IPW estimators are.

Theorem 1 shows that estimating Φ requires estimating the conditional expectation of the outcome and the conditional densities for mediators and time-varying confounders. Moreover, in some of the following approaches, we must specify the density function of the exposure. Accordingly, we consider the following two model spaces:

(A) $\mathcal{M}_A$: the models for the outcome, mediators, and time-varying confounders are correctly and separately specified;

(B) $\mathcal{M}_B$: the models for the exposure and mediators are correctly and separately specified.

Let $f(m_2|c_0,a,c_1,m_1,c_2;\boldsymbol{\beta_2})$ denote the density function of $M_2|C_0,A,C_1,M_1,C_2$ evaluated at c_0, a, c_1, m_1, and c_2 with parameter $\boldsymbol{\beta_2}$. Similarly, let $f(c_2|c_0,a,c_1,m_1;\boldsymbol{\gamma_2})$, $f(m_1|c_0,a,c_1;\boldsymbol{\beta_1})$, $f(c_1|c_0,a;\boldsymbol{\gamma_1})$, and $f(a|c_0;\boldsymbol{\delta})$ be the conditional density functions of C_2 with parameter $\boldsymbol{\gamma_2}$, M_1 with parameter $\boldsymbol{\beta_1}$, C_1 with parameter $\boldsymbol{\gamma_1}$, and A with parameter $\boldsymbol{\delta}$, respectively. $E(Y|c_0,a,c_1,m_1,c_2,m_2;\boldsymbol{\alpha})$ is the expectation of Y evaluated at c_0, a, c_1, m_1, and c_2 with parameter $\boldsymbol{\alpha}$.

We propose parametric estimation as the first approach. This approach adopts the maximum likelihood estimator (MLE) when parametric models are specified for M_2, C_2, M_1, and C_1, and an empirical distribution is specified for C_0. In causal inference, the parametric estimation relies on the regression model. Thus, this approach is referred to as the regression-based estimation approach. By the plug-in principle (Casella and Berger 2002) and Theorem 1, the MLEs of $\varphi_{TE}(a)$, $\varphi_1(a,m_1)$, $\varphi_2(a,m_2)$, and $\varphi_3(a,m_1,m_2)$, are given by

$$\hat{\varphi}_{TE}(a) = \mathbb{P}_n\big(E(Y|a,C_0)\big)$$

$$\hat{\varphi}_1^{Reg}(a,m_1) = \mathbb{P}_n[\int_{c_1,c_2,m_2} E(Y|C_0,a,c_1,m_1,c_2,m_2;\hat{\boldsymbol{\alpha}})dF\big(m_2\big|C_0,a,c_1,m_1,c_2;\hat{\boldsymbol{\beta}}_{\boldsymbol{2}}\big) \times dF(c_2|C_0,a,c_1,m_1;\hat{\boldsymbol{\gamma}}_{\boldsymbol{2}})dF(c_1|C_0,a;\hat{\boldsymbol{\gamma}}_{\boldsymbol{1}})\,],$$

$$\hat{\varphi}_2^{Reg}(a,m_2) = \mathbb{P}_n[\int_{c_1,c_2,m_1} E(Y|C_0,a,c_1,m_1,c_2,m_2;\hat{\boldsymbol{\alpha}})\,dF(c_2|C_0,a,c_1,m_1;\hat{\boldsymbol{\gamma}}_{\boldsymbol{2}}) \times dF\big(m_1\big|C_0,a,c_1;\hat{\boldsymbol{\beta}}_{\boldsymbol{1}}\big)dF(c_1|C_0,a;\hat{\boldsymbol{\gamma}}_{\boldsymbol{1}})],\ \text{and}$$

$$\hat{\varphi}_3^{Reg}(a,m_1,m_2) = \mathbb{P}_n[\int_{c_1,c_2} E(Y|C_0,a,c_1,m_1,c_2,m_2;\hat{\boldsymbol{\alpha}})\,dF(c_2|C_0,a,c_1,m_1;\hat{\boldsymbol{\gamma}}_{\boldsymbol{2}}) \times dF(c_1|C_0,a;\hat{\boldsymbol{\gamma}}_{\boldsymbol{1}})\,],$$

where $\mathbb{P}_n[\cdot] = n^{-1}\sum_i[\cdot]_i$ is the empirical average operator, and $\hat{\boldsymbol{\alpha}}$, $\hat{\boldsymbol{\beta}}_{\boldsymbol{1}}$, $\hat{\boldsymbol{\beta}}_{\boldsymbol{2}}$, $\hat{\boldsymbol{\gamma}}_{\boldsymbol{1}}$, and $\hat{\boldsymbol{\gamma}}_{\boldsymbol{2}}$ are the MLEs of $\boldsymbol{\alpha}$, $\boldsymbol{\beta_1}$, $\boldsymbol{\beta_2}$, $\boldsymbol{\gamma_1}$, and $\boldsymbol{\gamma_2}$, respectively. The regression-based estimators are consistent only when the density functions of mediators, time-varying confounders, and the outcome are correctly specified (i.e., the model space is assumed to be $\mathcal{M}_A$).

Next, we propose the IPW estimators of the EAMs. Lemma 1 supports the construction of IPW estimators.

Lemma 1. (IPW)

Suppose that the model space is $\mathcal{M}_B$. *If the first moment of* $(Y, C_0, A, C_1, M_1, C_2, M_2)$ *is finite, then*

$$\mathbb{P}_n\left[\frac{I(A=a)}{f(A|C_0;\boldsymbol{\delta})}Y\right] \xrightarrow{p} \varphi_{TE}(a),$$

$$\mathbb{P}_n\left[\frac{I(M_1=m_1)I(A=a)}{f(M_1|C_0,A,C_1;\boldsymbol{\beta_1})f(A|C_0;\boldsymbol{\delta})}Y\right] \xrightarrow{p} \varphi_1(a,m_1),$$

$$\mathbb{P}_n\left[\frac{I(M_2=m_2)I(A=a)}{f(M_2|C_0,A,C_1,M_1,C_2;\boldsymbol{\beta_2})f(A|C_0;\boldsymbol{\delta})}Y\right] \xrightarrow{p} \varphi_2(a,m_2), and$$

$$\mathbb{P}_n\left[\frac{I(M_1=m_1,M_2=m_2)I(A=a)}{f(M_2|C_0,A,C_1,M_1,C_2;\boldsymbol{\beta_2})f(M_1|C_0,A,C_1;\boldsymbol{\beta_1})f(A|C_0;\boldsymbol{\delta})}Y\right] \xrightarrow{p} \varphi_3(a,m_1,m_2),$$

when $n \to \infty$. *In these expressions,* a, m_1, *and* m_1 *are prespecified,* $I(\cdot)$ *and* $f(\cdot)$ *represent the indicator function and density function, respectively, and* $\mathbb{P}_n[\cdot] = n^{-1}\sum_i[\cdot]_i$.

The four expressions in Lemma 1 provide an unbiased estimator of Φ. The poof is shown in Appendix B. Based on this unbiasedness, the asymptotic properties directly follow from the weak law of large numbers. Thus, the IPW estimators are defined as

$$\widehat{\varphi}_{TE}(a) = \mathbb{P}_n\left[\frac{I(A=a)}{f(A|C_0;\widehat{\boldsymbol{\delta}})}Y\right],$$

$$\widehat{\varphi}_1^{IPW}(a,m_1) = \mathbb{P}_n\left[\frac{I(M_1=m_1)I(A=a)}{f(M_1|C_0,A,C_1;\widehat{\boldsymbol{\beta}}_{\boldsymbol{1}})f(A|C_0;\widehat{\boldsymbol{\delta}})} \times Y\right],$$

$$\widehat{\varphi}_2^{IPW}(a,m_2) = \mathbb{P}_n\left[\frac{I(M_2=m_2)I(A=a)}{f(M_2|C_0,A,C_1,M_1,C_2;\widehat{\boldsymbol{\beta}}_{\boldsymbol{2}})f(A|C_0;\widehat{\boldsymbol{\delta}})}Y\right], \text{and}$$

$$\widehat{\varphi}_3^{IPW}(a,m_1,m_2) = \mathbb{P}_n\left[\frac{I(M_1=m_1,M_2=m_2)I(A=a)}{f(M_2|C_0,A,C_1,M_1,C_2;\widehat{\boldsymbol{\beta}}_{\boldsymbol{2}})f(M_1|C_0,A,C_1;\widehat{\boldsymbol{\beta}}_{\boldsymbol{1}})f(A|C_0;\widehat{\boldsymbol{\delta}})}Y\right],$$

where $\widehat{\boldsymbol{\beta}}_{\boldsymbol{1}}$, $\widehat{\boldsymbol{\beta}}_{\boldsymbol{2}}$, and $\widehat{\boldsymbol{\delta}}$ are the MLEs of $\boldsymbol{\beta_1}$, $\boldsymbol{\beta_2}$, and $\boldsymbol{\delta}$, respectively. Essentially, the IPW estimator is semiparametric. Given standard regularity conditions, IPW estimators are consistent under $\mathcal{M}_B$. The asymptotic variance of IPW estimators can be consistently estimated using the sandwich estimator.

Regression-based and IPW estimators may be severely biased if their corresponding models are misspecified. That is, IPW estimators generally fail to be consistent when their models of the exposure and mediators are misspecified, even if the model of the outcome is correct. Similarly, regression-based estimators must have correct density functions of the

outcome and mediators for estimation consistency. Therefore, a robust estimator is required that remains consistent when one, but not necessarily both, of $\mathcal{M}_A$ and $\mathcal{M}_B$ is correctly specified. We derive the DR estimators of Φ on the union model space $\mathcal{M}_U = \mathcal{M}_A \cup \mathcal{M}_B$. To avoid lengthy formulations, we focus on deriving the DR estimator of $\varphi_1(a, m_1)$ in the main context. The derivations for the remaining parts of Φ are provided in Appendix C. Consider the following estimating equation:

$$U_1(\mu, \boldsymbol{\theta}; m_1, a) = \frac{I(M_1 = m_1)I(A = a)}{f(M_1|C_0, A, C_1; \boldsymbol{\beta_1})f(A|C_0; \boldsymbol{\delta})}\{Y - Q_1(a, m_1; \boldsymbol{\alpha}, \boldsymbol{\gamma_1})\}$$

$$+\{Q_1(a, m_1; \boldsymbol{\alpha}, \boldsymbol{\gamma_1}) - \mu\},$$

where $\boldsymbol{\theta} = \{\boldsymbol{\alpha}, \boldsymbol{\beta_1}, \boldsymbol{\gamma_1}, \boldsymbol{\delta}\}$ and

$$Q_1(a, m_1; \boldsymbol{\alpha}, \boldsymbol{\gamma_1}) = \int_{c_1} E(Y|C_0, a, c_1, m_1; \boldsymbol{\alpha})\, dF(c_1|C_0, a; \boldsymbol{\gamma_1}).$$

We define $\widehat{U}_1\big(\mu, \widehat{\boldsymbol{\theta}}; m_1, a\big)$ as $U_1(\mu, \boldsymbol{\theta}; m_1, a)$ evaluated at $\widehat{\boldsymbol{\alpha}}$, $\widehat{\boldsymbol{\beta_1}}$, and $\widehat{\boldsymbol{\delta}}$. By solving $\mathbb{P}_n\big[\widehat{U}_1\big(\mu, \widehat{\boldsymbol{\theta}}; m_1, a\big)\big] = 0$, the proposed DR estimator of $\varphi_1(a, m_1)$ is defined as

$$\hat{\varphi}_1^{DR}(a, m_1) = \mathbb{P}_n[\frac{I(M_1 = m_1)I(A = a)}{f\big(M_1\big|C_0, A, C_1; \widehat{\boldsymbol{\beta_1}}\big)f\big(A\big|C_0; \widehat{\boldsymbol{\delta}}\big)}\{Y - Q_1(a, m_1; \widehat{\boldsymbol{\alpha}}, \widehat{\boldsymbol{\gamma}}_1)\}$$

$$+Q_1(a, m_1; \widehat{\boldsymbol{\alpha}}, \widehat{\boldsymbol{\gamma}}_1)\ ].$$

According to following theorem, $\hat{\varphi}_1^{DR}(a, m_1)$ is consistent and asymptotically normal (CAN) under the union model space of $\mathcal{M}_A$ and $\mathcal{M}_B$.

Theorem 2. (Asymptotic property)

Suppose that assumptions (A1), (A2), and consistency hold and that the regularity conditions of Theorem A.1 in Robins et al. (1992) hold. Then, $\hat{\varphi}_1^{DR}(a, m_1)$ is regular and asymptotically linear under $\mathcal{M}_U = \mathcal{M}_A \cup \mathcal{M}_B$ with the influence function

$$\zeta_1(\mu^*, \boldsymbol{\theta}^*; m_1, a)$$

$$= U_1(\mu^*, \boldsymbol{\theta}^*; m_1, a) + E(\frac{\partial U_1(\mu^*, \boldsymbol{\theta}; m_1, a)}{\partial \boldsymbol{\theta}^{\boldsymbol{T}}})E(\frac{\partial \Lambda_{\boldsymbol{1}}(\boldsymbol{\theta}; m_1, a)}{\partial \boldsymbol{\theta}^{\boldsymbol{T}}})^{-1}\Lambda_{\boldsymbol{1}}(\boldsymbol{\theta}; m_1, a)\Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}^*},$$

where $\mu^* = \varphi_1(a, m_1)$, $\boldsymbol{\theta}^*$ is the probability limit of $\widehat{\boldsymbol{\theta}}$, and $\Lambda_{\boldsymbol{1}}(\boldsymbol{\theta}; m_1, a)$ is the collection of score functions for the MLEs of $f(M_1|C_0, A, C_1; \boldsymbol{\beta_1})$, $f(A|C_0; \boldsymbol{\delta})$, $f(C_1|C_0, a; \boldsymbol{\gamma_1})$, and $E(Y|C_0, A, C_1, M_1, C_2, M_2; \boldsymbol{\alpha})$. Consequently, by the central limit theorem and Slutsky's theorem,

$\hat{\varphi}_1^{DR}(a, m_1)$ *is a CAN estimator of* $\varphi_1(a, m_1)$ *under the union model space* $\mathcal{M}_U$.

According to Theorem 2,

$$\sqrt{n}\left(\hat{\varphi}_1^{DR}(a, m_1) - \varphi_1(a, m_1)\right) = \sqrt{n}\left(\sum\nolimits_i \zeta_{1i}(\mu^*, \boldsymbol{\theta}^*; m_1, a)\right) + o_p(1),$$

where $o_p(1)$ converges to zero probability and ζ_{1i} is the influence function ζ_1 evaluated at the *i*th observation. Consequently, we have $\sqrt{n}\left(\hat{\varphi}_1^{DR}(a, m_1) - \varphi_1(a, m_1)\right) \xrightarrow{d} N(0, \sigma_{\mu^*}^2)$, where $\sigma_{\mu^*}^2 = E(\zeta_1(\mu^*, \boldsymbol{\theta}^*; m_1, a)^2)$. The proof of Theorem 2 is provided in Appendix C. Theorem 2 and Appendix C offer the DR estimators of $\varphi_1(a, m_1)$, $\varphi_2(a, m_2)$, $\varphi_3(a, m_1, m_2)$, and $\varphi_{TE}(a)$, which are referred to as $\hat{\varphi}_1^{DR}(a, m_1)$, $\hat{\varphi}_2^{DR}(a, m_2)$, $\hat{\varphi}_3^{DR}(a, m_1, m_2)$, and $\hat{\varphi}_{TE}^{DR}(a)$, respectively.

4. EAM with an arbitrary number of multiple causally ordered mediators

4.1. General formulation of the EAMs

In this section, we provide the general formulations of EAMs when the number of mediators is arbitrary. The exposure A, outcome Y, and baseline confounder C_0 are still assumed. Additionally, we assume K causally ordered multiple mediators $\mathbf{M} = (M_1, M_2, \dots, M_K)$ and K time-varying confounders $\mathbf{C} = (C_1, C_2, C_3, \dots, C_K)$, where C_k represents the confounders among M_k and Y for $k \in \{1, 2, \dots, K\}$. The causal relationships between these variables are illustrated in the DAG in Figure 1(b).

For K mediators, we consider $2^K - 1$ distinct CDEs and 2^K EAMs. To clearly define each CDE and EAM, we introduce a definitional system for a generalized setting. For the *d*th EAM, the vector of binary variables $P_d = \left(p_d(M_1), \dots, p_d(M_K)\right)$ represents the correspondence to a subset of $\{M_1, \dots, M_K\}$ that is attributed to the *d*th EAM. In P_d, $p_d(M_k)$ is an indicator function where $p_d(M_k) = 1$ if the *d*th EAM is attributable to M_k and $p_d(M_k) = 0$ otherwise. Because $\left(p_d(M_1), \dots, p_d(M_K)\right)$ has a one-to-one relation with the

subscript d, we set $d = \sum_{k=1}^{K} p_d(M_k) \times 2^{k-1}$. Hence, we have $d \in \{0,1,2,\dots,2^K - 1\}$. We subsequently define the following set:

$$\mathbb{Q} = \{P_d | P_d = (p_d(M_1), \dots, p_d(M_K)) \text{ and } d = \sum_{k=1}^{K} p_d(M_k) \times 2^{k-1}\}.$$

Based on $\mathbb{Q}$, we further define $g_d(m_{(1,K)})$ as a regime of $M_1, \dots, M_K$ corresponding to P_d under a specified value $m_{(1,K)} = (m_1, \dots, m_K)$. In the regime $g_d(m_{(1,K)})$, we set M_k equal to m_k when $p_d(M_k) = 1$, and we do not intervene in M_k when $p_d(M_k) = 0$. Eventually, each regime other than $g_0(m_{(1,K)})$ corresponds to a CDE, and $g_0(m_{(1,K)})$ represents the TE. Based on $\mathbb{Q}$, we provide the definition for the generalized CDE as follows:

Definition 3. (Generalized CDE)
Suppose that $(M_1, \dots, M_K)$ *are* K *ordered mediators and* a, a^*, *and* $m_{(1,K)} = (m_1, m_2, \dots, m_K)$ *are specified values. Given a regime* $g_d(m_{(1,K)})$, *the corresponding multimediation parameter* $\varphi_d(a, m_{(1,K)})$ *is defined as*

$$\varphi_d(a, m_{(1,K)}) \equiv E(Y(a, g_d(m_{(1,K)}))).$$

Accordingly, the corresponding generalized CDE is defined as

$$CDE_d(m_{(1,K)}) = \varphi_d(a, m_{(1,K)}) - \varphi_d(a^*, m_{(1,K)}).$$

To provide the intuition underlying these notations, we rewrite them in the case with two mediators, as detailed in Section 2. P_0 corresponds to the vector $(p_0(M_1), p_0(M_2)) = (0,0)$, and $g_0(m_1, m_2)$ thus represents the situation in which neither mediator is intervened on. For the corresponding counterfactuals in Definition 3, $Y(a, g_0(m_{(1,2)}))$ is identical to $Y(a)$. Likewise, P_1 and P_2 correspond to the vectors $(p_1(M_1), p_1(M_2)) = (1,0)$ and $(p_2(M_1), p_2(M_2)) = (0,1)$, respectively. In the regimes $g_1(m_1, m_2)$ and $g_2(m_1, m_2)$, one mediator is intervened on, and the other mediator is not intervened on. Thus, $Y(a, g_1(m_{(1,2)})) = Y(a, m_1)$ and $Y(a, g_2(m_{(1,2)})) = Y(a, m_2)$. P_3 corresponds to the vector $(p_3(M_1), p_3(M_2)) = (1,1)$. In regime $g_0(m_1, m_2)$, all mediators are intervened on. The corresponding counterfactual $Y(a, g_3(m_{(1,2)}))$ is identical to $Y(a, m_1, m_2)$.

The general formulations of the EAMs are given in Definition 4.

Definition 4. (Generalized EAM)

Based on $\mathbb{Q}$*, we have*

$$\begin{pmatrix} CDE_0(m_{(1,K)}) \\ \vdots \\ CDE_{2^K-1}(m_{(1,K)},) \end{pmatrix} = H \begin{pmatrix} EAM_{\boldsymbol{P_0}}(m_{(1,K)}) \\ \vdots \\ EAM_{\boldsymbol{P_{2^K-1}}}(m_{(1,K)}) \end{pmatrix}$$

where H *is a transformation matrix with* $H_{j_1 j_2} = I(< P_{j_1-1}, P_{j_2-1} >= 0)$. $I(\cdot)$ *represents the indicator function, and* $<\cdot,\cdot>$ *is the inner product.*

Thus, the generalized EAMs are defined as

$$\begin{pmatrix} EAM_{P_0}(m_{(1,K)}) \\ \vdots \\ EAM_{P_{2^K-1}}(m_{(1,K)}) \end{pmatrix} = H^{-1} \begin{pmatrix} CDE_0(m_{(1,K)}) \\ \vdots \\ CDE_{2^K-1}(m_{(1,K)}) \end{pmatrix}.$$

Because H is invertible, the generalized EAMs in Definition 4 are well-defined. In Definition 4, $EAM_{P_d}(m_{(1,K)})$ is interpreted as the effect of A on Y through the mechanism attributable to the P_d-defined subset of $\{M_1, \dots, M_K\}$.

4.2. Identification

Similar to the identification process presented in Section 2, we must specify the assumptions for identifying generalized EAMs. The following two assumptions are required:

(B1) No unmeasured confounding among the exposure and outcome,

$$Y\left(a, \mathcal{g}_d(m_{(1,K)})\right) \perp\!\!\!\perp A | C_0.$$

(B2) No unmeasured confounding among the mediators and outcome

$$Y\left(a, \mathcal{g}_d(m_{(1,K)})\right) \perp\!\!\!\perp M_k | C_{(0,k)}, A, M_{(1,k-1)} \text{ if } p_d(M_k) = 1 \text{ for } k \in \{1,2,\dots,K\},$$

where $X_{(p,q)}$ represents the vector $(X_p, \dots, X_q)$ if $p \le q$, and $X_{(p,q)}$ is a null set if $p > q$. (B1) and (B2) are the generalized versions of (A1) and (A2), respectively. We further adopt the following consistency assumption: (1) for the outcome, $Y\left(a, m_{(1,K)}\right) = Y$ if $A = a$ and $M_{(1,K)} = m_{(1,K)}$; (2) for the mediators, $M_k\left(a, m_{(1,k-1)}\right) = M_k$ if $A = a$ and $M_{(1,k-1)} = m_{(1,k-1)}$ for $k \in \{1,2,\dots,K\}$. The identification of multimediation parameters is given in Theorem 3.

Theorem 3. (Identification of $\varphi_d\left(a, m_{(1,K)}\right)$**)**

Under (B1), (B2), and the consistency assumption, the multimediation parameter $\varphi_d\left(a, m_{(1,K)}\right)$ *in Definition 1 can be nonparametrically identified as follows:*

$$\varphi_d(a, m_{(1,K)}) = \int_{c_{(0,K)}} \int_{\text{Å}_d} E(Y|c_0, a, c_{(1,K)}, m_{(1,K)}) \times \prod_{k=1}^{K}(1 - p_d(M_k))dF(m_k|c_0, a, c_{(1,k)}, m_{(1,k-1)}) \times \prod_{k=1}^{K} dF(c_k|c_0, a, c_{(1,k-1)}, m_{(1,k-1)})\, dF_{C_0},$$

where $p_d(M_k)$ *represents the k*th *element of* $\boldsymbol{P_d}$ *and* $\text{Å}_d = \{(1 - p_d(M_k)) \times m_k | k = 1, \dots, K\}$.

The proof of Theorem 3 is presented in Appendix D.

5. Relation to CME

Although the CDE, natural direct effect, and natural indirect effect (Pearl 2001) are well-defined, no generally applicable definition of the controlled indirect effect is available. VanderWeele (2011) first proposed the complete-mediation condition, under which CME is well-defined. In the case of two mediators, the CME of A on Y through M_1 controlling for M_2 at m_2 is defined as $\text{CME}_{M_1}(m_2) = E(Y(a, m_2)) - E(Y(a^*, m_2))$. Likewise, $\text{CME}_{M_2}(m_1)$ is defined as $E(Y(a, m_1)) - E(Y(a^*, m_1))$, which represents the CME of A on Y through M_2 controlling for M_1 at m_1. $\text{CME}_{M_2}(m_1)$ and $\text{CME}_{M_1}(m_2)$ are mathematically identical to $\text{CDE}_1(m_1)$ and $\text{CDE}_2(m_2)$, respectively. In the framework of our proposed method, the complete-mediation condition indicates that $EAM_{P_0}(m_1, m_2) = 0$. Thus, we can derive $\text{CME}_{M_2}(m_1) = EAM_{P_1}(m_1, m_2)$ and $\text{CME}_{M_1}(m_2) = EAM_{P_2}(m_1, m_2)$. Based on Definition 2, $\text{TE} - \text{CME}_{M_2}(m_1) - \text{CME}_{M_1}(m_2)$ is equivalent to $EAM_{P_3}(m_1, m_2)$. That is, the CME cannot capture $EAM_{P_3}(m_1, m_2)$, which is the effect of the exposure on the outcome jointly attributable to M_1 and M_2. Thus, $EAM_{P_3}(m_1, m_2)$ fills the gap between the TE and CME.

Empirically, the complete-mediation condition is an extremely strong assumption for clinical studies or biological experiments in which a few mediators are collected. As shown in the application to liver disease in Section 7, it is unrealistic to assume that the effect of hepatitis C virus infection on mortality is completely mediated through the level of alanine

aminotransferase and through liver cancer status. Thus, the CME is not applicable in such cases, whereas the proposed EAM is unrestricted to the complete-mediation condition. If the mediators are collected comprehensively, the complete-mediation condition can be approximately satisfied. In this case, a general formulation of the CME with an arbitrary number of mediators is required. However, the original paper lacks a general formulation of the CME. Fortunately, as explained above in this section, the CMEs are equivalent to some of the EAMs if the complete-mediation condition is satisfied. The generalized EAM can improve the utility of the CME. Thus, the proposed EAM generalizes the CME by relaxing the complete-mediation condition, allowing for an arbitrary number of mediators, and filling the gap between the TE and CME.

Furthermore, for identification, VanderWeele (2011) adopted two assumptions, namely $Y(a, m_2) \perp\!\!\!\perp A|C_0$ and $Y(a, m_2) \perp\!\!\!\perp M_2|C_0, A, C_2$, to identify $\mathrm{CME}_{M_1}(m_2)$ as

$$\int_{c_0,c_2} E(Y|c_0, a, c_2, m_2)\, dF(c_2|c_0, a) dF(c_0) - \int_{c_0,c_2} E(Y|c_0, a^*, c_2, m_2)\, dF(c_2|c_0, a^*) dF(c_0).$$

Under $Y(a, m_1) \perp\!\!\!\perp A|C_0$ and $Y(a, m_1) \perp\!\!\!\perp M_1|C_0, A, C_1$, $\mathrm{CME}_{M_2}(m_1)$ is identified as

$$\int_{c_0,c_1} E(Y|c_0, a, c_1, m_1)\, dF(c_1|c_0, a) dF(c_0) - \int_{c_0,c_1} E(Y|c_0, a^*, c_1, m_1)\, dF(c_1|c_0, a^*) dF(c_0).$$

These identifications are detailed in (VanderWeele 2011). These identification results are restricted to the parallel mediation structure, in which mediators are causally independent. If M_1 is the cause of M_2, then the assumption $Y(a, m_2) \perp\!\!\!\perp M_2|C_0, A, C_2$ is not valid. This follows from the nonparametric structural equation model (Appendix E). Therefore, with respect to identification and assumptions, the CME is a special case of the EAM because the EAM allows for dependent mediators.

6. Simulation

We conducted three simulation studies with different model settings to compare the performance of three proposed estimators. The data generation for the simulations proceeded

as follows:

$C_0 \sim Ber(p = expit(0.5))$,
$A|C_0 \sim Ber(p = expit(0.5C_0))$,
$C_1|C_0, A \sim Ber(p = expit(0.1 - 0.5C_0 - 0.5A))$,
$M_1|C_0, A, C_1 \sim Ber(p = expit(0.1 - 0.5C_0 + A - 0.5C_1))$,
$C_2|C_0, A, C_1, M_1 \sim Ber(p = expit(0.1 - 0.5C_0 + 0.5A - 0.5C_1 - 0.5M_1))$,
$M_2|C_0, A, C_1, M_1, C_2 \sim Ber(p = expit(0.1 - 0.5C_0 - A - 0.1C_1 - 0.1M_1 - 0.1C_2))$,
$Y|C_0, A, C_1, M_1, C_2, M_2 \sim Ber(p = expit(0.1 - 0.5C_0 - A - 0.5C_1 + M_1 - 0.5C_2 - 0.5M_2))$,

where Ber denotes the Bernoulli distribution function and $expit$ denotes the expit function. Following the DAG in Figure 1(a), we generated, as the covariates of the outcome Y, a baseline confounder C_0, two time-varying confounders C_1 and C_2, and two mediators M_1 and M_2. All variables were set as binary variables to adhere to the conditions of the motivating example. Simulations were performed 1,000 times with the sample sizes of 500 and 1,000. The performance of each of the three proposed estimators was evaluated separately under the following scenarios:

Scenario (1): the model of the outcome is correct; the model of the exposure is correct;
Scenario (2): the model of the outcome is incorrect; the model of the exposure is correct;
Scenario (3): the model of the outcome is correct; the model of the exposure is incorrect.

Specifically, Scenario (1) considers the case in which all models are correctly specified according to the simulation setting; Scenario (2) considers the case in which first, the model of Y is fitted by the Binomial distribution with the probit link function rather than with the expit link function and second, the remaining models are correctly specified; and Scenario (3) considers the case in which first, the model of A is fitted by the Binomial distribution with the probit link function rather than with the expit link function and second, the remaining models are correctly specified. Simulations under Scenarios (2) and (3) enable assessment of the robustness of the three proposed estimators when models are misspecified. The results are summarized in Table 1 for the sample size of 1,000 and in Appendix F for the sample size of

500. In the simulations, the interventions of mediators for the EAM are set equal to 0.

The results confirmed that the three estimators were consistent under Scenario (1) for the sample sizes of both 500 and 1,000. By contrast, the regression-based and IPW estimators were biased in Scenarios (2) and (3), respectively. In particular, the regression-based estimator exhibited dramatic model misspecification. For example, in Scenario (2), the coverage of the 95% confidence interval of the regression-based estimator was considerably lower than 0.95, whereas, in Scenario (3), the coverage of the 95% confidence interval of the IPW estimator was close to 0.95. Therefore, the IPW estimator is less biased than the regression-based estimator is when the model specification of some variables is uncertain. Furthermore, the results of our simulation studies demonstrated that the DR estimators of the EAMs and TE are guaranteed to be robust to model misspecification. In terms of bias and coverage rate, the DR estimator outperformed the IPW and regression-based estimators across all three scenarios.

Table 1. Simulation results (sample size = 1000)

	Doubly robust estimator			Inverse probability weighting estimator			Regression-based estimator		
	Bias	**SE**	**COV**	**Bias**	**SE**	**COV**	**Bias**	**SE**	**COV**
Scenario (1): the model of outcome is **correct**; the model of exposure is **correct.**									
$\mathbf{EAM_{P_0}}$	<0.001	0.010	0.945	<0.001	0.031	0.946	<0.001	0.029	0.951
$\mathbf{EAM_{P_1}}$	0.001	0.013	0.947	0.001	0.022	0.952	<0.001	0.012	0.955
$\mathbf{EAM_{P_2}}$	-0.001	0.011	0.955	-0.001	0.018	0.963	0.001	0.012	0.948
$\mathbf{EAM_{P_3}}$	<0.001	0.016	0.953	<0.001	0.011	0.957	<0.001	0.010	0.955
TE	0.001	0.007	0.954	0.001	0.023	0.957	0.001	0.031	0.946
Scenario (2): the model of outcome is **incorrect**; the model of exposure is **correct.**									
$\mathbf{EAM_{P_0}}$	0.001	0.012	0.962	<0.001	0.031	0.953	0.051	0.021	0.340
$\mathbf{EAM_{P_1}}$	0.002	0.012	0.953	0.001	0.022	0.947	0.002	0.008	0.936
$\mathbf{EAM_{P_2}}$	0.001	0.009	0.953	0.002	0.018	0.949	-0.014	0.007	0.404
$\mathbf{EAM_{P_3}}$	-0.002	0.012	0.949	-0.002	0.010	0.958	0.002	0.007	0.229
TE	0.001	0.006	0.951	0.001	0.023	0.951	0.041	0.021	0.527
Scenario (3): the model of outcome is **correct**; the model of exposure is **incorrect.**									
$\mathbf{EAM_{P_0}}$	0.003	0.010	0.944	0.031	0.027	0.918	0.002	0.029	0.944
$\mathbf{EAM_{P_1}}$	-0.001	0.013	0.949	0.012	0.021	0.949	0.001	0.012	0.951

$\mathbf{EAM_{P_2}}$	<0.001	0.010	0.951	-0.003	0.017	0.938	-0.001	0.011	0.964
$\mathbf{EAM_{P_3}}$	<0.001	0.015	0.944	-0.002	0.009	0.947	<0.001	0.010	0.944
TE	0.002	0.007	0.951	0.038	0.014	0.751	0.002	0.031	0.944

P_0 is the null set; P_1: $\{M_1\}$; P_2: $\{M_2\}$; P_3, $\{M_1, M_2\}$;
SE, standard error; CI, confidence interval; TE, total effect; EAM, effect attributable to mediators; COV, coverage of the 95% confidence interval.

7. Application

A study of hepatitis C virus (HCV)-induced liver disease motivated the present work. The development of an HCV vaccine faces challenges, and a vaccine capable of protecting against hepatitis C is not available. Thus, an intervention to prevent HCV infection cannot be implemented, and researchers have focused on intervention on mediators. Multiple risk factors of liver disease have been confirmed by association studies (Chen *et al.* 2008, Harman *et al.* 2015). Some risk factors, such as abnormal alanine aminotransferase (ALT) level and hepatocellular carcinoma (HCC), are potential mediators in the effect of HCV infection on mortality. Because treatments have been developed for abnormal liver function and early diagnosed HCC, intervention in these mediators is not only applicable but also critical for patients with HCV-positive status.

We applied the proposed EAM to the REVEAL nationwide population-based cohort study in Taiwan, and we investigated the effect of HCV infection on mortality (Chen, Yang, Yang, Liu, Chen, You, Wang, Sun, Lu and Chen 2008) by intervening in ALT and HCC. In REVEAL, 23,820 participants aged 30 to 65 years were recruited from seven townships of Taiwan during 1991 to 1992. After removing missing values, 23,724 samples were retained for the analysis. HCV infection status and ALT were measured at baseline. HCC status and death were confirmed through computerized data linkages with the national cancer registry and death certification system, respectively, in 2008. HCV, ALT, and HCC were coded as binary variables according to the following rules: HCV = 1 indicates HCV infection and HCV = 0 indicates no HCV infection; ALT = 1 indicates abnormal ALT level (>40 U/L) and ALT = 0

indicates normal ALT level; HCC = 1 indicates HCC diagnosis and HCC = 0 indicates no HCC diagnosis. For individual outcomes, Y = 1 indicates death and Y = 0 indicates survival. Additionally, age, gender, smoking status, and alcohol status were included as baseline confounders. We did not consider any time-varying confounders in this application. The causal diagram is shown in Figure 2.

We applied EAMs to this application by intervening in two mediators and setting them to normal status (i.e., ALT = 0 and HCC = 0). Four subsets of $\{\mathrm{ALT}, \mathrm{HCC}\}$ are defined as follows: P_0 is the null set; $\mathrm{P}_1 = \{\mathrm{ALT}\}$; $\mathrm{P}_2 = \{\mathrm{HCC}\}$; and $\mathrm{P}_3 = \{\mathrm{ALT}, \mathrm{HCC}\}$. All variables were fitted by logistic regression models according to the causal relationship shown in Figure 2. Subsequently, four EAMs and the TE were estimated by using the DR, IPW, and regression-based estimators. The estimates, 95% confidence intervals (CIs), and *p* values are listed in Table 2. First, the three approaches yielded similar estimates of TE, and the results of the DR, IPW, and regression-based methods were 0.096 (95% CI: 0.087–0.106), 0.097 (95% CI: 0.083–0.110), and 0.092 (95% CI: 0.071–0.114), respectively. For the estimation of EAMs, Table 2 indicates that the results of the DR, IPW, and regression-based approaches were slightly different. As discussed in Section 3, the DR estimator is theoretically more robust than the IPW and regression-based estimators are when an incorrect model specification of the outcome or exposure occurs. In addition, the result of the DR estimation is relatively plausible for explaining the mechanism of HCV-induced liver injury. The host immune response to HCV infection usually leads to active hepatitis, along with abnormal ALT level, thus contributing to the development of cancer.

In the results of the DR estimation, the three estimates of the EAMs differed significantly from zero. $\widehat{\mathrm{EAM}}_{\mathrm{P}_0}$ = 0.048 (95% CI: 0.045–0.051) corresponds to the effect of HCV infection on mortality without any mechanisms attributable to mediators. $\widehat{\mathrm{EAM}}_{\mathrm{P}_1}$ = 0.015 (95% CI: 0.012–0.017) corresponds to the effect of HCV infection on mortality attributable to abnormal

ALT level. $\widehat{EAM}_{P_3}$ = 0.023 (95% CI: 0.009–0.037) corresponds to the effect of HCV infection on mortality jointly attributable to abnormal ALT level and HCC. Therefore, HCV-induced mortality was 15.6% (=1.5/9.6) for that attributable to ALT solely and 24% (=2.3/9.6) for that attributable to ALT level and HCC jointly, indicating that intervention in both risk factors is potentially more powerful and cost-effective compared with just intervening in one risk factor.

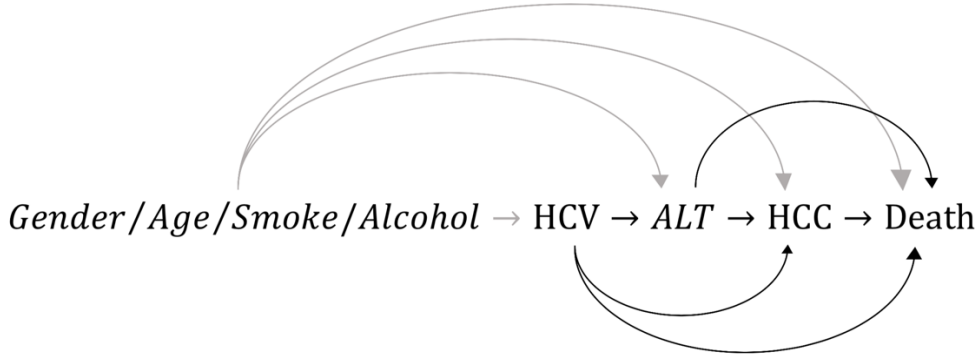

Figure 2. Causal diagram of HCC cohort. Black arcs represent the paths of interest.
Abbreviations: HCV, hepatitis C virus; ALT, abnormal alanine aminotransferase; HCC, hepatocellular carcinoma.

Table 2. Estimation of TE and all EAMs in HCC cohort.

	Doubly robust estimator			Inverse probability weighting estimator			Regression-based estimator		
	Estimate (SE)	**95% CI**	**P-value**	**Estimate (SE)**	**95% CI**	**P-value**	**Estimate (SE)**	**95% CI**	**P-value**
$\mathbf{EAM_{P_0}}$	0.048 (0.002)	(0.045, 0.051)	<0.001	0.058 (0.008)	(0.042, 0.074)	<0.001	0.041 (0.010)	(0.022, 0.060)	<0.001
$\mathbf{EAM_{P_1}}$	0.015 (0.001)	(0.012, 0.017)	<0.001	0.005 (0.006)	(-0.007, 0.018)	0.394	0.006 (0.002)	(0.002, 0.010)	0.002
$\mathbf{EAM_{P_2}}$	0.011 (0.006)	(-0.001, 0.022)	0.069	0.011 (0.001)	(0.008, 0.014)	<0.001	0.021 (0.003)	(0.015, 0.027)	<0.001
$\mathbf{EAM_{P_3}}$	0.023 (0.007)	(0.009, 0.037)	0.002	0.023 (0.002)	(0.019, 0.027)	<0.001	0.024 (0.003)	(0.017, 0.030)	<0.001
TE	0.096 (0.005)	(0.087, 0.106)	<0.001	0.097 (0.007)	(0.083, 0.110)	<0.001	0.092 (0.011)	(0.071, 0.114)	<0.001

Set: P_0 is the null set; P_1:{ALT}; P_2: {HCC}; P_3: {ALT, HCC}.
SE, standard error; CI: confidence interval; TE, total effect; EAM, effect attributable to mediators; HCV, hepatitis C virus; ALT, abnormal alanine aminotransferase; HCC, hepatocellular carcinoma.

8. Discussion

In this article, we propose the EAM as an alternative approach for effect decomposition in addition to causal mediation analysis in settings with multiple causally ordered mediators. The EAM relaxes the assumptions, required in previous methods, that are related to cross-world counterfactuals and identification (Daniel, De Stavola, Cousens and Vansteelandt 2015, Lin 2019, Lin and VanderWeele 2017). Under the assumption of no interaction, the EAMs are identical to the causal effects mediated through mediation paths. Because the EAM enables intervention in multiple mediators affecting the whole population at the statuses of interest, the development of the EAM contributes to both policy making and mechanism investigation. In the analysis of the REVEAL dataset, we estimated the EAMs of the effects of HCV on mortality attributable to ALT level and HCC; this can facilitate the implementation of health promotion policies.

Although the proposed EAM focuses on the risk difference scale, extending the EAM to the risk ratio scale or the odds ratio scale is trivial. The major limitation of this work is that the proposed method does not apply to longitudinal studies. In a longitudinal study, time-varying exposures and death-truncation are two primary issues that should be addressed. Moreover, the extension to survival outcomes is critical in medical and biological research. More generalized methods that fit all circumstances will be developed in future studies. Finally, the EAM methodology was implemented using the statistical software R, and the algorithm is available for download.

Acknowledgments

We thank Professor Hwai-I Yang (Genomics Research Center, Academia Sinica, Taipei, Taiwan) for suggesting revisions to refine the model and for providing the example dataset. We also thank Ying-Wen Liang for formulating the initial version of EAM. This study was supported by a grant from the Ministry of Science and Technology, Taiwan (No. 109-2636-B-009 -001). This manuscript was edited by Wallace Academic Editing.

References

J. M. Albert and S. Nelson (2011) Generalized causal mediation analysis. *Biometrics,* 1028-1038.

C. Avin, I. Shpitser and J. Pearl (2005) Identifiability of path-specific effects In *Proceedings of the 19th international joint conference on Artificial intelligence*, pp. 357-363: Morgan Kaufmann Publishers Inc.

G. Casella and R. L. Berger (2002) *Statistical inference*: Duxbury Pacific Grove, CA.

C. L. Chen, H. I. Yang, W. S. Yang, C. J. Liu, P. J. Chen, S. L. You, L. Y. Wang, C. A. Sun, S. N. Lu and D. S. Chen (2008) Metabolic factors and risk of hepatocellular carcinoma by chronic hepatitis B/C infection: a follow-up study in Taiwan. *Gastroenterology,* 111-121.

R. M. Daniel, B. L. De Stavola, S. N. Cousens and S. Vansteelandt (2015) Causal mediation analysis with multiple mediators. *Biometrics,* 1-14.

F. Fasanelli, M. T. Giraudo, F. Ricceri, L. Valeri and D. Zugna (2019) Marginal Time-Dependent Causal Effects in Mediation Analysis With Survival Data. *American journal of epidemiology,* 967-974.

S. Goetgeluk, S. Vansteelandt and E. Goetghebeur (2008) Estimation of controlled direct effects. *Journal of the Royal Statistical Society Series B,* 1049-1066.

D. J. Harman, S. D. Ryder, M. W. James, M. Jelpke, D. S. Ottey, E. A. Wilkes, T. R. Card, G. P. Aithal and I. N. Guha (2015) Direct targeting of risk factors significantly increases the detection of liver cirrhosis in primary care: a cross-sectional diagnostic study utilising transient elastography. *BMJ open,* e007516.

Y.-T. Huang and H.-I. Yang (2017) Causal mediation analysis of survival outcome with multiple mediators. *Epidemiology,* 370-378.

S.-H. Lin (2019) Generalized interventional approach for causal mediation analysis with causally ordered multiple mediators. *Harvard University Biostatistics Working Paper Series*.

S.-H. Lin and T. VanderWeele (2017) Interventional approach for path-specific effects. *Journal of Causal Inference*.

R. J. Little and D. B. Rubin (2000) Causal effects in clinical and epidemiological studies via potential outcomes: concepts and analytical approaches. *Annual review of public health,* 121-145.

J. Pearl (2001) Direct and indirect effects In *Proceedings of the Seventeenth conference on Uncertainty in artificial intelligence*, pp. 411-420, San Francisco, CA, USA: Morgan kaufmann publishers Inc.

J. Pearl and J. M. Robins (1995) Probabilistic evaluation of sequential plans from causal models with hidden variables In *Proceedings of the Eleventh conference on Uncertainty in artificial intelligence*, pp. 444–453, Montréal, Qué, Canada: Morgan Kaufmann Publishers Inc.

J. M. Robins (1987) Addendum to “a new approach to causal inference in mortality studies with a sustained exposure period—application to control of the healthy worker survivor effect”. *Computers & Mathematics with Applications,* 923-945.

J. M. Robins (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,* 1393-1512.

J. M. Robins and S. Greenland (1992) Identifiability and exchangeability for direct and indirect effects. *Epidemiology* 143-155.

J. M. Robins, S. D. Mark and W. K. Newey (1992) Estimating exposure effects by modelling the expectation of exposure conditional on confounders. *Biometrics,* 479-495.

J. Steen, T. Loeys, B. Moerkerke and S. Vansteelandt (2017) Flexible mediation analysis with multiple mediators. *American journal of epidemiology,* 184-193.

A.-S. Tai and S.-H. Lin (2020) Integrated multiple mediation analysis: A robustness–specificity trade-off in causal structure. *Harvard University Biostatistics Working Paper Series*.

E. J. Tchetgen Tchetgen and I. Shpitser (2012) Semiparametric theory for causal mediation analysis: efficiency bounds, multiple robustness and sensitivity analysis. *The Annals of Statistics,* 1816-1845.

T. J. VanderWeele (2011) Controlled direct and mediated effects: definition, identification and bounds. *Scandinavian Journal of Statistics,* 551-563.

T. J. VanderWeele (2013) Policy-relevant proportions for direct effects. *Epidemiology,* 175-176.

T. J. VanderWeele (2014) A unification of mediation and interaction: a four-way decomposition. *Epidemiology,* 749-761.

T. J. VanderWeele and S. Vansteelandt (2014) Mediation Analysis with Multiple Mediators. *Epidemiologic methods,* 95-115.

T. J. VanderWeele, S. Vansteelandt and J. M. Robins (2014) Effect decomposition in the presence of an exposure-induced mediator-outcome confounder. *Epidemiology,* 300-306.