Current Epidemiology Reports (2025) 12:20
https://doi.org/10.1007/s40471-025-00369-3

Natural Effects and Separable Effects: Insights into Mediation Analysis

Etsuji Suzuki[1] · **Tomohiro Shinozaki**[2,3] · **Eiji Yamamoto**[4]

Accepted: 30 July 2025

Abstract

Purpose of Review We compare natural effects and separable effects under nonparametric structural equation models with independent errors, highlighting their similarities and differences. By examining their required properties and sufficient conditions for identification, we aim to provide deeper insights into mediation analysis.

Recent Findings If certain assumptions about confounding, positivity, and consistency are met, we can identify natural direct and indirect effects under nonparametric structural equation models with independent errors. However, these effects have been criticized because they rely on a specific cross-world quantity, and the so-called cross-world independence assumption cannot be empirically verified. Furthermore, interventions on the mediator may sometimes be challenging to even conceive. As an alternative approach, separable effects have recently been proposed and applied in mediation analysis, often under finest fully randomized causally interpretable structured tree graph models. These effects are defined without relying on any cross-world quantities and are claimed to be identifiable under assumptions that are testable in principle, thereby addressing some of the challenges associated with natural direct and indirect effects.

Summary To conduct meaningful mediation analysis, it is crucial to clearly define the research question of interest, and the choice of methods should align with the nature of the question and the assumptions researchers are willing to make. Examining the underlying philosophical perspectives on causation and manipulation can provide valuable insights.

Keywords Causality · Counterfactuals · Cross-world independence assumption · Directed acyclic graphs · Mediation analysis · Nonparametric structural equation models with independent errors

Introduction

The assessment of mediation provides a valuable approach to gaining a deeper understanding of cause–effect relationships by examining whether and how a mediator transmits the effect of an exposure or intervention to an outcome [1–7]. By defining and identifying direct and indirect effects, mediation analysis enables us to disentangle complex causal mechanisms, providing insights into underlying biological, behavioral, or social processes. To this end, causal mediation analysis within the counterfactual framework has gained increasing attention across various disciplines in recent years [8–10]. Additionally, the AGReMA statement (A Guideline for Reporting Mediation Analyses) was developed to provide consolidated recommendations for reporting mediation analyses [11].

As is well appreciated in the literature on causal mediation, the total effect of the exposure on the outcome can be decomposed into natural direct and indirect effects [12, 13]. If certain assumptions about confounding, positivity, and consistency are met, the so-called mediation formula can be used to identify these effects in nonparametric structural equation models with independent errors (NPSEM-IE) [13, 14]. However, natural direct and indirect effects have been criticized because, as explained below, these rely on a specific cross-world quantity, and the so-called cross-world independence assumption—part of a set of sufficient, but not necessary, assumptions—is not

✉ Etsuji Suzuki
etsuji-s@cc.okayama-u.ac.jp

[1] Department of Epidemiology, Graduate School of Medicine, Dentistry and Pharmaceutical Sciences, Okayama University, 2-5-1 Shikata-cho, Kita-ku, Okayama 700-8558, Japan

[2] Interfaculty Initiative in Information Studies, the University of Tokyo, Tokyo, Japan

[3] Department of Biostatistics, School of Public Health, Graduate School of Medicine, the University of Tokyo, Tokyo, Japan

[4] Okayama University of Science, Okayama, Japan

Published online: 21 October 2025

empirically verifiable [15–18]. Furthermore, interventions on the mediator may sometimes be challenging to even conceive.

As an alternative approach, separable effects have recently been proposed and applied in mediation analysis [15–20], often in finest fully randomized causally interpretable structured tree graph (FFRCISTG) models [21]. Under this approach, the exposure is assumed to be separated into two (or more) components, one having a direct effect only on the mediator and the other having a direct effect only on the outcome. Furthermore, each separable component can be intervened separately in principle, and the total effect can be decomposed into separable direct and indirect effects. These effects are defined without relying on any cross-world quantities and are claimed to be identifiable under assumptions that are testable in principle [15], thereby addressing some of the challenges associated with natural direct and indirect effects [22].

In this article, we compare natural effects and separable effects under NPSEM-IE, highlighting their similarities and differences. Additionally, we illustrate these two approaches graphically using causal directed acyclic graphs (DAGs) [23, 24], incorporating potential outcomes determined by NPSEM-IE. By examining their required properties and sufficient conditions for identification, we aim to provide deeper insights into mediation analysis. Furthermore, to compare the two approaches, we examine their underlying philosophical perspectives on causation and manipulation. We then briefly review the controlled direct effect and interventional effects before concluding the article.

Natural Direct and Indirect Effects

Notations and Definitions

We let A denote an exposure of interest, Y an outcome of interest, and M a potential mediator of interest, as depicted in the causal DAG in Fig. 1a. For example, in the context of mediation analysis, Hernán and Robins [18] considered a randomized trial among cigarette smokers, letting A denote smoking cessation, M the presence of hypertension at 6 months, and Y the incidence of myocardial infarction within 1 year, assuming that no individuals experienced the outcome Y during the first 6 months. Similar examples were used in related literature [15, 16]. Throughout the present article, we assume that the set of baseline covariates not affected by the exposure, denoted as $\boldsymbol{C}$, is empty unless stated otherwise. However, a similar discussion applies, conditional on $\boldsymbol{C}=\boldsymbol{c}$, followed by marginalizing

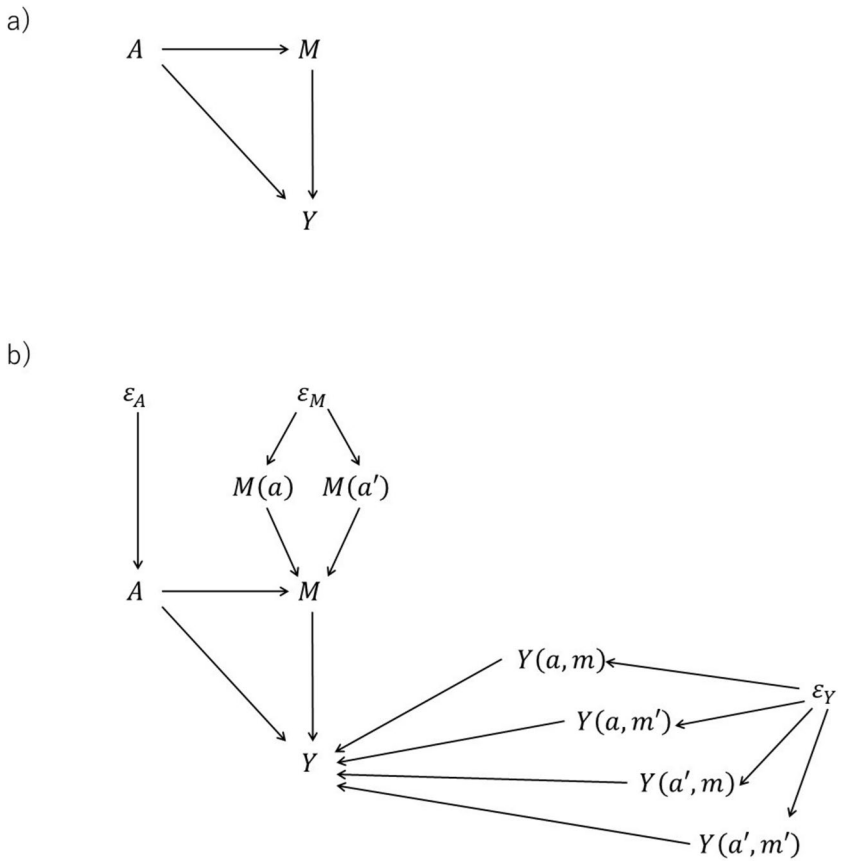

Fig. 1 **a** A causal directed acyclic graph (DAG) with exposure A, mediator M, and outcome Y. **b** A causal DAG incorporating the potential outcomes as well as error terms

over the possible values of $\boldsymbol{C}$. We presuppose that at least hypothetical interventions on A and M are conceivable.

In the counterfactual framework, we let $Y(a)$ and $M(a)$ denote the potential outcomes of Y and M, respectively, if, possibly contrary to fact, there had been interventions to set A to a. Additionally, we let $Y(a,m)$ denote the potential outcomes of Y if, possibly contrary to fact, there had been interventions to set A to a and to set M to m. Throughout this article, we assume that positivity and consistency hold [25–28]; see Nguyen et al. [5] for an in-depth discussion about these assumptions in the context of mediation analysis. Furthermore, we make a generalized consistency or composition assumption, $Y(a) = Y(a, M(a))$ [29–31], where the nested counterfactual on the right-hand side is sometimes referred to as a compound potential outcome [32–34]. Note that the composition assumption is needed not for identification but for interpretation of the natural effects [5].

Suppose that a and a' are two values of the exposure we wish to compare, the latter of which is a reference condition; for example, for binary exposure, we may have $a = 1$ and $a' = 0$. Similarly, m and m' are two values of the mediator. Then, the total effect on Y of setting the exposure to $A = a$ versus $A = a'$ in the population of interest is defined as $E[Y(a)] - E\left[Y\left(a'\right)\right]$, or equivalently $E[Y(a, M(a))] - E\left[Y\left(a', M\left(a'\right)\right)\right]$ under the composition assumption. As is well appreciated in the literature on causal mediation, even when there are interactions and nonlinearities, the total effect of A on Y can be decomposed into the pure direct effect (PDE) and the total indirect effect (TIE), as follows [12, 13]:

Definition 1

$$\text{PDE} \triangleq E\left[Y\left(a, M\left(a'\right)\right)\right] - E\left[Y\left(a', M\left(a'\right)\right)\right],$$
$$\text{TIE} \triangleq E[Y(a, M(a))] - E\left[Y\left(a, M\left(a'\right)\right)\right].$$

Alternatively, the total effect can be decomposed into the total direct effect (TDE) and the pure indirect effect (PIE), as follows:

Definition 2

$$\text{TDE} \triangleq E[Y(a, M(a))] - E\left[Y\left(a', M(a)\right)\right],$$
$$\text{PIE} \triangleq E\left[Y\left(a', M(a)\right)\right] - E\left[Y\left(a', M\left(a'\right)\right)\right].$$

Note that Definitions 1 and 2 are based on the counterfactual framework, which is completely general in terms of the models that it can accommodate. These two different decompositions essentially arise from different ways of accounting for an interaction between the exposure and the mediator; these become equivalent if there is no interaction. In this article, we use Definition 1, referring to the PDE and the TIE as the natural direct effect and natural indirect effect, respectively. Note that the counterfactual $Y\left(a, M\left(a'\right)\right)$ where $a \neq a'$ is referred to as a "cross-world" counterfactual [16] because two different levels of A are nested within the counterfactual for Y. To assess the extent to which the total effect operates through the mediator, the "proportion mediated" is sometimes used, which is defined on the difference scale as the ratio of the natural indirect effect to the total effect $\left(E[Y(a, M(a))] - E\left[Y\left(a, M\left(a'\right)\right)\right]\right) / \left(E[Y(a)] - E\left[Y\left(a'\right)\right]\right)$.

On a related issue, in the sufficient cause framework [35], Suzuki et al. [34] demonstrated that, under the assumption of sufficient cause positive monotonicity of the exposure and the mediator, although the PIE implies the presence of mediating pathways, it does not necessarily imply their operation because a non-M-mediating path may operate to induce Y. However, this is not the case for TIE, and a non-zero TIE implies the operation—not simply the presence—of mediation. This also supports the use of Definition 1. For details, see the related literature [34, 36].

Nonparametric Structural Equations for Natural Effects

In this article, we assume that a causal DAG represents an NPSEM-IE [14], which means that (i) each variable is some arbitrary general function of the other variables with arrows to that variable and a random error term and that (ii) the random error terms are independent of one another. Thus, Fig. 1a implies the following nonparametric structural equations for the observable (or factual) variables, A, M, and Y [14]:

$$\begin{cases} A = f_A\left(\varepsilon_A\right) = g_A\left(\varepsilon_A\right), \\ M = f_M\left(A, \varepsilon_M\right) = g_M\left(\varepsilon_A, \varepsilon_M\right), \\ Y = f_Y\left(A, M, \varepsilon_Y\right) = g_Y\left(\varepsilon_A, \varepsilon_M, \varepsilon_Y\right), \end{cases}$$

where ε_A, ε_M, and ε_Y are mutually independent. Note that $f_V(\cdot)$ and $g_V(\cdot)$ are arbitrary functions for generating a variable V, and the latter is used when all causal variables are error terms ε. Because the error terms are exogeneous variables, $g_V(\cdot)$ may be regarded as a "reduced form" in the econometrics literature. Each equation shows how an individual response variable changes as its direct (parent) causal variables change and can thus be interpreted from a perspective of the potential-outcome model for that response. Therefore, Fig. 1a implies the following nonparametric structural equations for the potential outcomes:

$$\begin{cases} M(a) = f_M\left(a, \varepsilon_M\right) = g_M^*\left(a, \varepsilon_M\right) \ (\forall a), \\ Y(a, m) = f_Y\left(a, m, \varepsilon_Y\right) = g_Y^*(a, m, \varepsilon_Y) \ (\forall a, m). \end{cases}$$

Note that we use $g_M^*(\cdot)$ and $g_Y^*(\cdot)$ because their functional forms may differ from $g_M(\cdot)$ and $g_Y(\cdot)$, respectively. Figure 1b shows a causal DAG incorporating the potential outcomes

of M and Y, as well as the error terms for A, M, and Y. Note that an arrow exerts from each potential outcome to the corresponding observed variable, such that each observed variable has its direct causal variable(s) and the corresponding potential outcomes as parents. For example, the parents of Y are A, M, $Y(a,m)$, $Y(a,m')$, $Y(a',m)$, and $Y(a',m')$. Note that ε_M and ε_Y are common causes of the potential outcomes of M and Y, respectively. The observed variables and their potential outcomes are endogenous variables, whereas the error terms are exogeneous variables in the system of structural equations.

Identification of Natural Effects

In NPSEM-IE, the natural direct and indirect effects can be identified if the following four assumptions hold [13, 37, 38]:

$$Y(a,m) \perp\!\!\!\perp A \ (\forall a,m), \tag{1}$$

$$Y(a,m) \perp\!\!\!\perp M(a) | A = a \ (\forall a,m), \tag{2}$$

$$M(a) \perp\!\!\!\perp A \ (\forall a), \tag{3}$$

$$Y(a,m) \perp\!\!\!\perp M(a') \ (\forall a,a',m). \tag{4}$$

Note that these are a set of sufficient independence conditions, although weaker assumptions relevant to the natural direct and indirect effects are often sufficient [5]. See the Appendix for further discussion on assumption (2).

It is worth noting that, unlike assumptions (1) to (3), assumption (4) is the so-called cross-world independence assumption [37] because it involves counterfactuals referring to two different "worlds" or scenarios. Specifically, assumption (4) states that the counterfactual values of the outcome if A were set to a are independent of those of the mediator if A were set to a'. The cross-world independence assumption is assumed under an NPSEM-IE, which is sometimes referred to as a "multiple-worlds model" [37]. By contrast, the cross-world independence assumptions are not assumed under an FFRCISTG model [16], which is sometimes referred to as a "single-world model" [37]. Although the natural direct and indirect effects are ontologically defined under an FFRCISTG model, they are not point-identified; however, their sharp bounds can be obtained [15]. These differences reflect important epistemological distinctions between NPSEM-IE and FFRCISTG models [16]. To summarize, as noted by Shpitser et al. [39] (p. 826), FFRCISTG models are "ontologically liberal, but epistemologically conservative."

Incorporating potential outcomes and error terms into the causal DAG has the advantage of visually illustrating that assumptions (1) to (4) hold (Fig. 1b). For example, regarding assumption (4), there are no open paths between $Y(a,m)$ and $M(a')$ and they are d-separated, which implies their independence under the rule of compatibility [40] or connectivity [41]. Note that the same argument applies between $Y(a,m')$ and $M(a')$, among others. Throughout the text, we focus on a specific set of nodes without loss of generality when discussing d-separation. If assumptions (1) to (4) hold, $E[Y(a,M(a'))]$ is identified by the following formula [14, 42]:

$$E[Y(a,M(a'))] = \sum_m E[Y|A=a,M=m]P(M=m|A=a'),$$

which is a special case of the "mediational g-formula" for time-fixed exposure and mediator [43]. Consequently, the natural direct and indirect effects are identified and given by the empirical expressions (see Online Appendix A). Note that we use $Y(a,M(a')) = Y(a,m)$ if $M(a') = m$ to identify these effects, which is specifically referred to as the "consistency of the cross-world potential outcome" by Nguyen et al. [5]. See Online Appendix B for further discussion on identification of the total effect.

Next, let us consider a scenario in which there is a mediator–outcome confounder H that is not affected by the exposure A (Fig. 2a). This implies the following nonparametric structural equations for the observable variables:

$$\begin{cases} A = f_A(\varepsilon_A) = g_A(\varepsilon_A), \\ H = f_H(\varepsilon_H) = g_H(\varepsilon_H), \\ M = f_M(A,H,\varepsilon_M) = g_M(\varepsilon_A,\varepsilon_H,\varepsilon_M), \\ Y = f_Y(A,M,H,\varepsilon_Y) = g_Y(\varepsilon_A,\varepsilon_M,\varepsilon_H,\varepsilon_Y). \end{cases}$$

Accordingly, we can obtain the following nonparametric structural equations for the potential outcomes:

$$\begin{cases} M(a) = f_M(a,H,\varepsilon_M) = g_M^*(a,\varepsilon_H,\varepsilon_M) \ (\forall a), \\ Y(a,m) = f_Y(a,m,H,\varepsilon_Y) = g_Y^*(a,m,\varepsilon_H,\varepsilon_Y) \ (\forall a,m). \end{cases}$$

In Fig. 2b, we visually show the relationship by incorporating the potential outcomes and the error terms. As indicated in the nonparametric structural equations above, H is depicted as a common cause of M and its two potential outcomes, as well as Y and its four potential outcomes (Fig. 2b). In this case, unlike Fig. 1b, although assumptions (1) and (3) hold, assumptions (2) and (4) *generally* do not. This point is visually illustrated in Fig. 2b. First, assumption (2) does not generally hold because, among those with $A = a$, there is an open path between $Y(a,m)$ and $M(a)$: $Y(a,m) \leftarrow H \rightarrow M(a)$. Note that this is based on the rule of weak faithfulness [40]; under the assumption of faithfulness, which is the converse property of compatibility [24], assumption (2) does not hold. However, there are no open paths between $Y(a,m)$ and $M(a)$ conditional on H among those with $A = a$; $Y(a,m) \perp\!\!\!\perp M(a) | (A = a, H)$

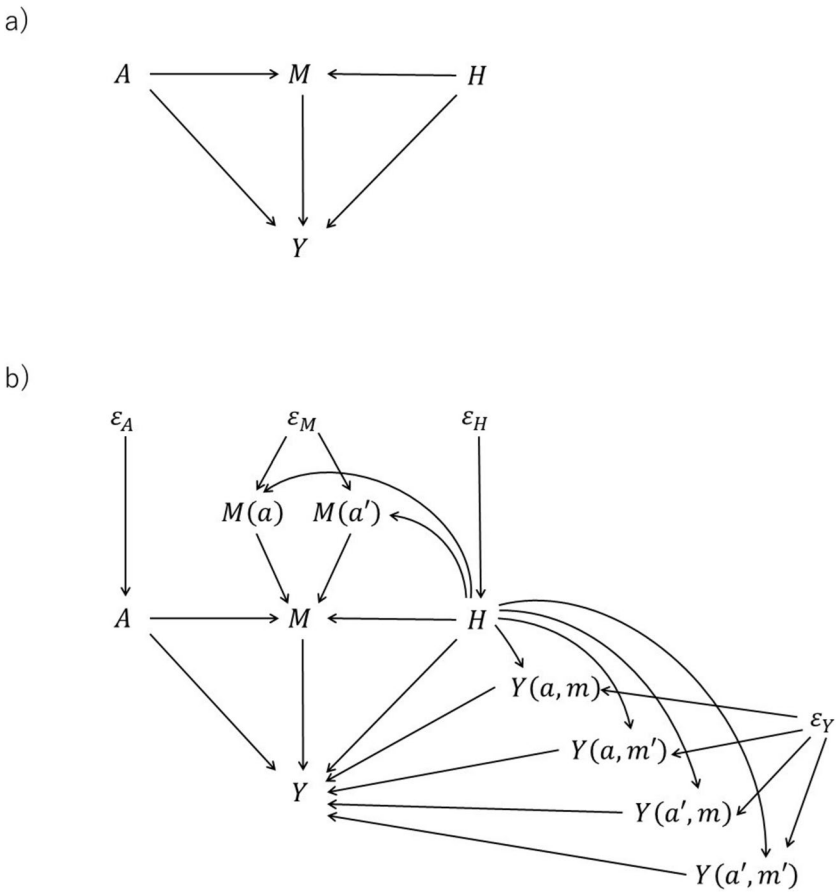

Fig. 2 **a** A causal directed acyclic graph (DAG) with exposure A, mediator M, and outcome Y when there is a mediator–outcome confounder H. **b** A causal DAG incorporating the potential outcomes as well as error terms

holds. Similarly, regarding assumption (4), there is an open path between $Y(a,m)$ and $M(a')$: $Y(a,m) \leftarrow H \rightarrow M(a')$, and assumption (4) does not generally hold. However, conditional on H, there are no open paths between $Y(a,m)$ and $M(a')$; $Y(a,m) \perp\!\!\!\perp M(a')|H$ holds. This is the so-called conditional cross-world independence assumption. Additionally, because H is not a collider, conditioning on H does not open any path, neither between $Y(a,m)$ and A, nor between $M(a)$ and A. Indeed, the following four assumptions hold in Fig. 2b:

$$Y(a,m) \perp\!\!\!\perp A|H \ (\forall a,m), \tag{5}$$

$$Y(a,m) \perp\!\!\!\perp M(a)|(A=a,H) \ (\forall a,m), \tag{6}$$

$$M(a) \perp\!\!\!\perp A|H \ (\forall a), \tag{7}$$

$$Y(a,m) \perp\!\!\!\perp M(a')|H \ (\forall a,a',m). \tag{8}$$

Under assumptions (5) to (8), $E\left[Y\left(a,M\left(a'\right)\right)\right]$ is identified as

$$E\left[Y\left(a,M\left(a'\right)\right)\right] = \sum_h \sum_m E[Y|A=a,M=m,H=h]P\left(M=m|A=a',H=h\right)P(H=h),$$

and the natural direct and indirect effects are identified and given by the empirical expressions (see Online Appendix C). See the Appendix for further discussion on assumption (6).

Finally, let us consider a situation in which there is an exposure-induced mediator–outcome confounder L (Fig. 3a). Sometimes, L is referred to as a "recanting witness" for A [44]. Figure 3a implies the following nonparametric structural equations for the observable variables:

$$\begin{cases} A = f_A\left(\varepsilon_A\right) = g_A\left(\varepsilon_A\right), \\ L = f_L\left(A,\varepsilon_L\right) = g_L\left(\varepsilon_A,\varepsilon_L\right), \\ M = f_M\left(A,L,\varepsilon_M\right) = g_M\left(\varepsilon_A,\varepsilon_L,\varepsilon_M\right), \\ Y = f_Y\left(A,M,L,\varepsilon_Y\right) = g_Y\left(\varepsilon_A,\varepsilon_M,\varepsilon_L,\varepsilon_Y\right). \end{cases}$$

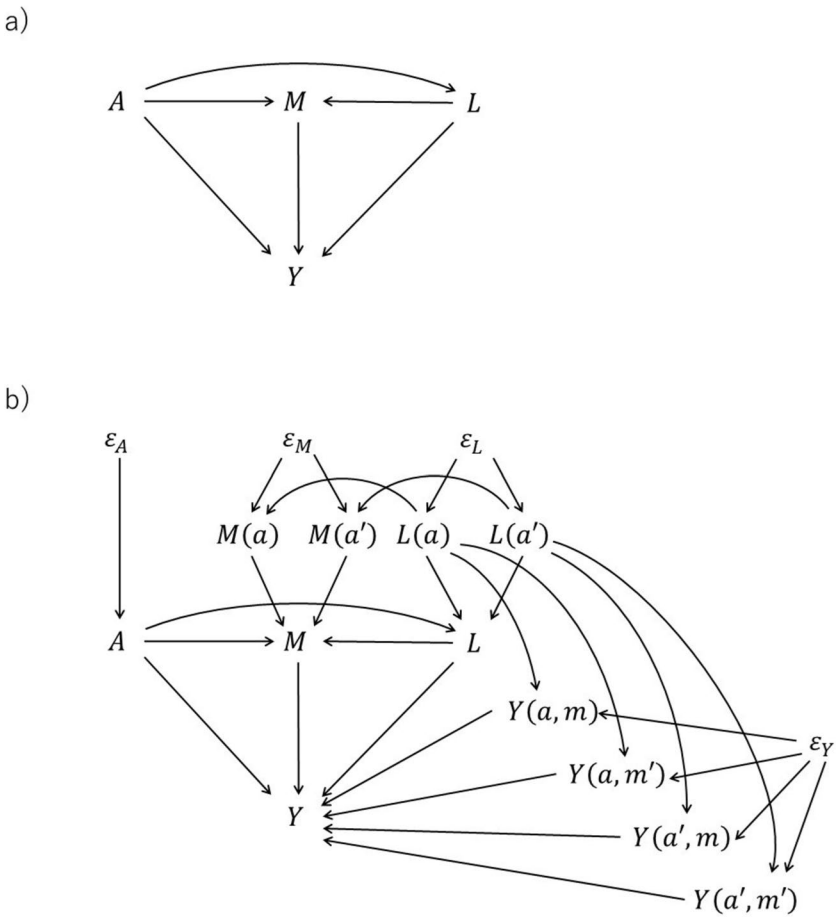

Fig. 3 **a** A causal directed acyclic graph (DAG) with exposure A, mediator M, and outcome Y when there is an exposure-induced mediator–outcome confounder L. **b** A causal DAG incorporating the potential outcomes as well as error terms

Accordingly, the following nonparametric structural equations for the potential outcomes can be obtained:

$$\begin{cases} L(a) = f_L(a, \varepsilon_L) = g_L^*(a, \varepsilon_L) \ (\forall a), \\ M(a) = f_M(a, L(a), \varepsilon_M) = g_M^*(a, \varepsilon_L, \varepsilon_M) \ (\forall a), \\ Y(a,m) = f_Y(a, m, L(a), \varepsilon_Y) = g_Y^*(a, m, \varepsilon_L, \varepsilon_Y) \ (\forall a, m), \end{cases}$$

where $L(a)$ denotes the potential outcome of L if, possibly contrary to fact, there had been interventions to set A to a. In Fig. 3b, we visually show the relationship by incorporating the potential outcomes and the error terms. As indicated in the nonparametric structural equations for the potential outcomes, arrows exert from $L(a)$ to $M(a)$, $Y(a,m)$, and $Y(a,m')$. Similarly, arrows exert from $L(a')$ to $M(a')$, $Y(a',m)$, and $Y(a',m')$. As in Fig. 2b, although assumptions (1) and (3) hold in Fig. 3b, assumptions (2) and (4) *generally* do not. Assumption (2) does not generally hold because, among those with $A = a$, there is an open path between $Y(a,m)$ and $M(a)$: $Y(a,m) \leftarrow L(a) \rightarrow M(a)$. Similarly, assumption (4) does not generally hold because there is an open path between $Y(a,m)$ and $M(a')$: $Y(a,m) \leftarrow L(a) \leftarrow \varepsilon_L \rightarrow L(a') \rightarrow M(a')$. Next, given that assumptions (5) to (8) hold in Fig. 2b, let us examine whether the following assumptions hold in Fig. 3b:

$$Y(a,m) \perp\!\!\!\perp A | L \ (\forall a, m), \tag{9}$$

$$Y(a,m) \perp\!\!\!\perp M(a) | (A = a, L(a)) \ (\forall a, m), \tag{10}$$

$$M(a) \perp\!\!\!\perp A | L \ (\forall a), \tag{11}$$

$$Y(a,m) \perp\!\!\!\perp M(a') | L \ (\forall a, a', m). \tag{12}$$

To state the conclusion first, only assumption (10) holds because there are no open paths between $Y(a,m)$ and $M(a)$ conditional on $L(a)$ among those with $A = a$; see the Appendix for further discussion on assumption (10). However, assumptions (9), (11), and (12) do not *generally* hold in Fig. 3b. Specifically, assumption (9) does not generally hold because, conditional on L, there is an open path between $Y(a,m)$ and A: $Y(a,m) \leftarrow L(a) \rightarrow L \leftarrow A$. Similarly, assumption (11) does not generally hold because, conditional on L, there is an open path between

$M(a)$ and A: $M(a) \leftarrow L(a) \rightarrow L \leftarrow A$. Finally, assumption (12) does not generally hold because, conditional on L, there are two open paths between $Y(a,m)$ and $M(a')$: $Y(a,m) \leftarrow L(a) \leftarrow \varepsilon_L \rightarrow L(a') \rightarrow M(a')$ and $Y(a,m) \leftarrow L(a) \rightarrow L \leftarrow L(a') \rightarrow M(a')$. Thus, if there is an exposure-induced mediator–outcome confounder L, the cross-world independence assumption does not generally hold, with or without conditioning on L [37].

To summarize, if there is an effect of the exposure that confounds the mediator–outcome relationship, as in Fig. 3a, the natural direct and indirect effects are not *generally* identified irrespective of whether data are available on L, except under strong assumptions [44], such as no interaction between the exposure and mediator at the individual level [45]. In other words, the absence of the exposure-induced mediator–outcome confounder L is a sufficient but not a necessary condition for identification of the natural direct and indirect effects. Even when an exposure-induced mediator–outcome confounder is present, the separable direct and indirect effects can still be identified from the data, provided certain assumptions hold. In the next section, we discuss these effects.

Separable Direct and Indirect Effects

Notations and Definitions

The basic idea underlying separable effects is that the exposure A can be decomposed into two separable components N and O, where the separable component N directly affects M but not Y; by contrast, the separable component O directly affects Y but not M. In their example of a randomized trial examining the effect of smoking cessation on myocardial infarction, Hernán and Robins [18] considered that N represents nicotine exposure and O represents exposure to the other non-nicotine components of a cigarette. Similar examples have been used in the related literature [15, 16]. Figure 4a shows a causal DAG including N and O. The absence of an arrow from N to Y encodes an assumption that N does not have a direct effect on Y. Similarly, the absence of an arrow from O to M encodes an assumption that O does not have an effect on M. Note that the bold arrows from A to N and O indicate deterministic relationships [15]. The two separable components N and O are not observed, and we observe only the value of A; in observed data, $A \equiv N \equiv O$ holds. However, we assume

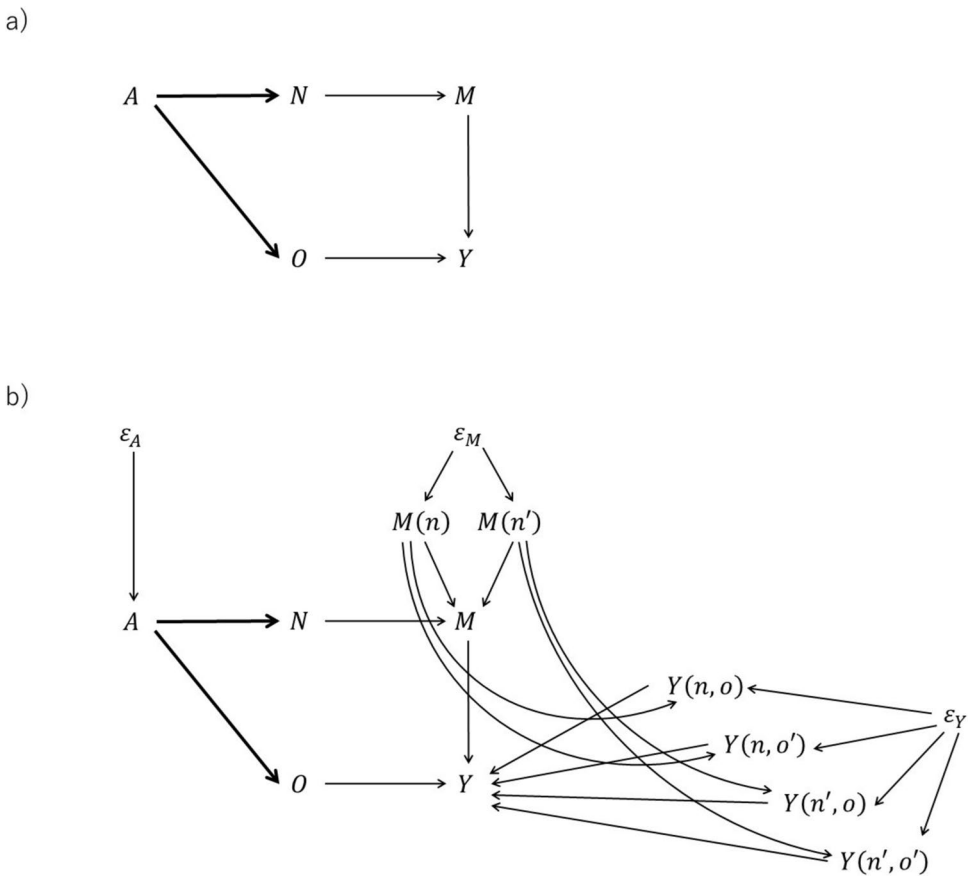

Fig. 4 **a** A causal directed acyclic graph (DAG) with exposure A, mediator M, and outcome Y, where A is assumed to be decomposed into two separable components N and O. The bold arrows from A to N and O indicate deterministic relationships. **b** A causal DAG incorporating the potential outcomes as well as error terms

that a future trial could be designed in which interventions are applied separately to separable components N and O. The relationships between N and M and between O and Y are not deterministic.

When considering the separable effects in the context of mediation, we do not consider interventions on the mediator M itself; rather, we consider separate interventions on the separable components N and O. Like the exposure A, suppose that n and n' are two values of the separable component N we wish to compare, and similarly, o and o' are two values of the separable component O. Then, let us consider a four-arm randomized controlled trial on the separable components N and O, comparing their two values: $(N,O)=(n,o),\left(n,o'\right),\left(n',o\right),\left(n',o'\right)$. In the counterfactual framework, we let $Y(n,o)$ denote the potential outcomes of Y if, possibly contrary to fact, there had been interventions to set N to n and to set O to o, where $Y(n,o)=Y(a)$ and $Y\left(n',o'\right)=Y\left(a'\right)$ hold based on the deterministic relationships between A, N, and O. We also let $M(n,o)$ denote the potential outcome of M if, possibly contrary to fact, there had been interventions to set N to n and to set O to o.

Accordingly, the total effect of A on Y of setting the exposure to $A=a$ versus $A=a'$ in the population of interest can be expressed as $E[Y(n,o)]-E\left[Y\left(n',o'\right)\right]$ in the four-arm randomized controlled trial. Using $E\left[Y\left(n',o\right)\right]$, the total effect can be decomposed into the separable direct effect (SDE) and the separable indirect effect (SIE). as follows:

Definition 3

$$\mathrm{SDE}\left(n'\right)\triangleq E\left[Y\left(n',o\right)\right]-E\left[Y\left(n',o'\right)\right],$$
$$\mathrm{SIE}(o)\triangleq E[Y(n,o)]-E\left[Y\left(n',o\right)\right].$$

Note that $E\left[Y\left(n',o\right)\right]$ is a hypothetical quantity because it cannot be observed even partly in the current data. Alternatively, the total effect can be also decomposed using $E\left[Y\left(n,o'\right)\right]$, as below:

Definition 4

$$\mathrm{SDE}(n)\triangleq E[Y(n,o)]-E\left[Y\left(n,o'\right)\right],$$
$$\mathrm{SIE}\left(o'\right)\triangleq E\left[Y\left(n,o'\right)\right]-E\left[Y\left(n',o'\right)\right].$$

Like Definitions 1 and 2, Definitions 3 and 4 are based on the counterfactual framework, which is completely general in terms of the models that it can accommodate. In this article, we use Definition 3 for the separable direct and indirect effects, which corresponds to Definition 1 for the natural direct and indirect effects. Note that $Y\left(n',o\right)$ and $Y\left(n,o'\right)$ are (single-world) counterfactuals involving only N and O, not the mediator M; Robins et al. [16] (p. 747) refer to them as "non-cross-world" counterfactuals. Unlike the natural direct and indirect effects and the controlled direct effect (defined later), the separable direct and indirect effects do not require reference to counterfactuals indexed by m (e.g., $Y(a,m)$). Recent studies have also discussed separable effects in other contexts, such as competing events [46–49] and the generalized frontdoor formula [50].

Nonparametric Structural Equations for Separable Effects

Although separable effects are often explained using single-world intervention graphs under the FFRCISTG models [39, 51], we use NPSEM-IE and causal DAGs to highlight the differences and similarities between natural effects and separable effects. Using the same reasoning as for natural direct and indirect effects, Fig. 4a implies the following nonparametric structural equations for the observable (or factual) variables A, N, O, M, and Y:

$$\begin{cases} A=f_A\left(\varepsilon_A\right)=g_A\left(\varepsilon_A\right),\\ N=f_N(A)=g_N\left(\varepsilon_A\right),\\ O=f_O(A)=g_O\left(\varepsilon_A\right),\\ M=f_M\left(N,\varepsilon_M\right)=g_M\left(\varepsilon_A,\varepsilon_M\right),\\ Y=f_Y\left(O,M,\varepsilon_Y\right)=g_Y\left(\varepsilon_A,\varepsilon_M,\varepsilon_Y\right). \end{cases}$$

Note that, because of the deterministic relationships between A, N, and O, we do not consider error terms ε_N or ε_O; rather, the separable components N and O are governed by the error term ε_A via the exposure A, such that $g_N\left(\varepsilon_A\right)=g_O\left(\varepsilon_A\right)=g_A\left(\varepsilon_A\right)$, and hence, $f_N(A)=f_O(A)=A$ in the observed data.

Following the same logic, the following nonparametric structural equations for the potential outcomes can be obtained:

$$\begin{cases} N(a)=O(a)=a\ (\forall a),\\ M(n,o)=M(n)=f_M\left(n,\varepsilon_M\right)=g_M^*\left(n,\varepsilon_M\right)\ (\forall n,o),\\ Y(n,o)=f_Y\left(o,M(n,o),\varepsilon_Y\right)=g_Y^*\left(n,o,\varepsilon_M,\varepsilon_Y\right)\ (\forall n,o), \end{cases}$$

where $N(a)$ and $O(a)$ denote the potential outcomes of N and O, respectively, if, possibly contrary to fact, there had been interventions to set A to a. Note that the first equation, $N(a)=O(a)=a$, indicates the deterministic relationships between A, N, and O [16]. Note also that, because we assume that O is not a cause of M for every individual, we can write $M(n,o)$ as $M(n)$. This is an individual-level assumption, sometimes referred to as the "isolation assumption" (see Online Appendix D for further discussion) [46]. Figure 4b presents a causal diagram incorporating the potential outcomes of M and Y, as well as the error terms for A, M, and Y. As mentioned above, there are no error terms ε_N or ε_O. As indicated in the nonparametric structural equations for the

potential outcomes, arrows exert from $M(n)$ to $Y(n,o)$ and $Y(n,o')$. Similarly, arrows exert from $M(n')$ to $Y(n',o)$ and $Y(n',o')$. These are the primary differences between Figs. 1b and 4b. However, there are also some similarities between them; a total of three arrows go to M and a total of six arrows go to Y. This point is related to the fact that, although we consider interventions on A and M in the natural direct and indirect effects, we consider interventions on N and O in the separable direct and indirect effects.

Finally, note that, if we implement interventions on N and O in a future (actual) trial, the nonparametric structural equations for the observable variables will become

$$\begin{cases} N = f_N^*(\varepsilon_N) = g_N^*(\varepsilon_N), \\ O = f_O^*(\varepsilon_O) = g_O^*(\varepsilon_O), \\ M = f_M(N, \varepsilon_M) = g_M^{**}(\varepsilon_N, \varepsilon_M), \\ Y = f_Y(O, M, \varepsilon_Y) = g_Y^{**}(\varepsilon_N, \varepsilon_O, \varepsilon_M, \varepsilon_Y), \end{cases}$$

which are completely different from the previous ones. However, the nonparametric structural equations for the potential outcomes become

$$\begin{cases} M(n,o) = M(n) = f_M(n, \varepsilon_M) = g_M^*(n, \varepsilon_M)\ (\forall n, o), \\ Y(n,o) = f_Y(o, M(n,o), \varepsilon_Y) = g_Y^*(n, o, \varepsilon_M, \varepsilon_Y)\ (\forall n, o), \end{cases}$$

which are identical to the nonparametric structural equations for the potential outcomes of M and Y in the current trial; this shows that a similar discussion applies in a future trial, where, unlike the current trial, $(N,O) = (a,a')$ can be implemented $(\forall a, a')$ [52]. In the following discussion, although we assume that a future trial could be designed to apply interventions separately to the separable components N and O, we only observe the value of A.

Identification of Separable Effects

Unlike natural direct and indirect effects, because separable effects are defined without relying on any cross-world quantities, they are claimed to be identifiable under assumptions that are testable in principle [15]. Regarding the identification condition for the separable direct and indirect effects, note that the following three assumptions hold in Fig. 4b:

$$(Y(n,o), M(n,o)) \perp\!\!\!\perp (N,O)\ (\forall n, o), \tag{13}$$

$$M \perp\!\!\!\perp O | N, \tag{14}$$

$$Y \perp\!\!\!\perp N | (O, M). \tag{15}$$

Assumptions (14) and (15) trivially hold in the observed data because $N = O$. In a future trial where N and O are separately intervened, assumptions (14) and (15) are testable [52]. The fact that these assumptions hold in both the observed data and future trial in the same population ensures that $E[Y(n',o)]$ in the future trial is identifiable from the observed data. This is consistent with the discussion in the last paragraph of the previous section. Under assumptions (13) to (15), $E[Y(n',o)]$ is identified as

$$E[Y(n',o)] = \sum_m E[Y|A=a, M=m] P(M=m|A=a'),$$

which is identical to the identification formula for $E[Y(a, M(a'))]$, and the separable direct and indirect effects are identified and given by the empirical expressions (see Online Appendix E). As explained in Online Appendix F, these effects are identified under weaker assumptions. The positivity assumption is addressed in footnote b of Table 1. When there is a common cause H of the mediator M and the outcome Y, the separable direct and indirect effects in the subgroup with $H=h$ are similarly identified and given by the empirical expressions, and the separable direct and indirect effects in the total population are obtained by marginalizing them over $H=h$.

Next, we consider situations in which there is an exposure-induced mediator–outcome confounder L. Following Robins et al. [16], we consider three scenarios. First, we consider a scenario in which N is a parent of L, but O is not, as described in Fig. 5a. In this case, we obtain the following nonparametric structural equations for the observable variables:

$$\begin{cases} A = f_A(\varepsilon_A) = g_A(\varepsilon_A), \\ N = f_N(A) = g_N(\varepsilon_A), \\ O = f_O(A) = g_O(\varepsilon_A), \\ M = f_M(N, L, \varepsilon_M) = g_M(\varepsilon_A, \varepsilon_L, \varepsilon_M), \\ L = f_L(N, \varepsilon_L) = g_L(\varepsilon_A, \varepsilon_L), \\ Y = f_Y(O, M, L, \varepsilon_Y) = g_Y(\varepsilon_A, \varepsilon_M, \varepsilon_L, \varepsilon_Y). \end{cases}$$

Accordingly, the following nonparametric structural equations for the potential outcomes can be obtained:

$$\begin{cases} N(a) = O(a) = a\ (\forall a), \\ M(n,o) = M(n) = f_M(n, L(n), \varepsilon_M) = g_M^*(n, \varepsilon_L, \varepsilon_M)\ (\forall n, o), \\ L(n,o) = L(n) = f_L(n, \varepsilon_L) = g_L^*(n, \varepsilon_L)\ (\forall n, o), \\ Y(n,o) = f_Y(o, M(n,o), L(n,o), \varepsilon_Y) = g_Y^*(n, o, \varepsilon_M, \varepsilon_L, \varepsilon_Y)\ (\forall n, o), \end{cases}$$

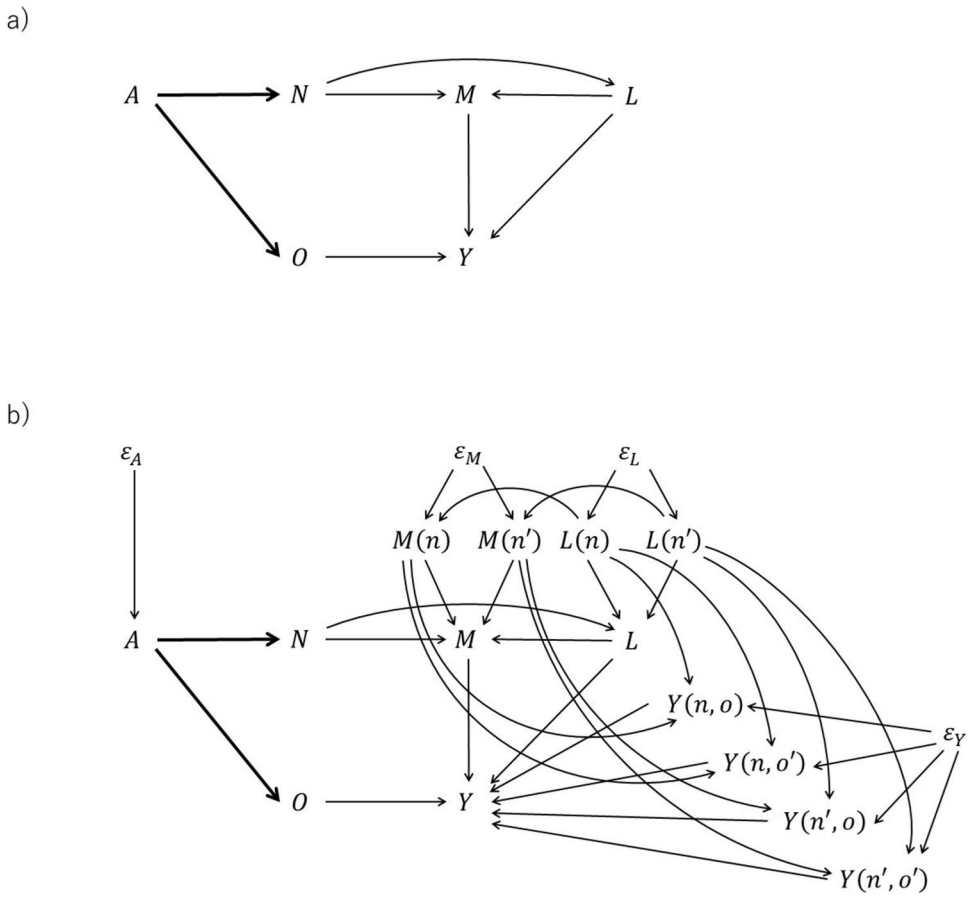

Fig. 5 **a** A causal directed acyclic graph (DAG) with exposure A, mediator M, and outcome Y, where A is assumed to be decomposed into two separable components N and O, and an exposure-induced mediator–outcome confounder L is present. We consider a scenario in which N is a parent of L, but O is not. The bold arrows from A to N and O indicate deterministic relationships. **b** A causal DAG incorporating the potential outcomes as well as error terms

where $L(n, o)$ denotes the potential outcome of L if, possibly contrary to fact, there had been interventions to set N to n and to set O to o. Because we assume that O is not a cause of L for every individual in this setting, we can write $L(n, o)$ as $L(n)$. Figure 5b shows these relationships visually, in which the following assumptions hold:

$$\big(Y(n,o), L(n,o), M(n,o)\big) \perp\!\!\!\perp (N, O)\ (\forall n, o), \tag{16}$$

$$M \perp\!\!\!\perp O | (L, N), \tag{17}$$

$$Y \perp\!\!\!\perp N | (L, M, O), \tag{18}$$

$$L \perp\!\!\!\perp O | N. \tag{19}$$

Under assumptions (16) to (19), $E\big[Y\big(n', o\big)\big]$ is identified as

$$E\big[Y\big(n', o\big)\big] = \sum_{m,l} E[Y|M = m, L = l, A = a] P\big(M = m|L = l, A = a'\big) P\big(L = l|A = a'\big),$$

and the separable direct and indirect effects are identified and given by the empirical expressions (see Online Appendix G). As explained in Online Appendix H, these effects are identified under weaker assumptions.

Next, we consider a scenario in which O is a parent of L, but N is not, as described in Fig. 6a. In this case, we obtain the following nonparametric structural equations for the observable variables:

$$\begin{cases} A = f_A\big(\varepsilon_A\big) = g_A\big(\varepsilon_A\big), \\ N = f_N(A) = g_N\big(\varepsilon_A\big), \\ O = f_O(A) = g_O\big(\varepsilon_A\big), \\ M = f_M\big(N, L, \varepsilon_M\big) = g_M\big(\varepsilon_A, \varepsilon_L, \varepsilon_M\big), \\ L = f_L\big(O, \varepsilon_L\big) = g_L\big(\varepsilon_A, \varepsilon_L\big), \\ Y = f_Y\big(O, M, L, \varepsilon_Y\big) = g_Y\big(\varepsilon_A, \varepsilon_M, \varepsilon_L, \varepsilon_Y\big). \end{cases}$$

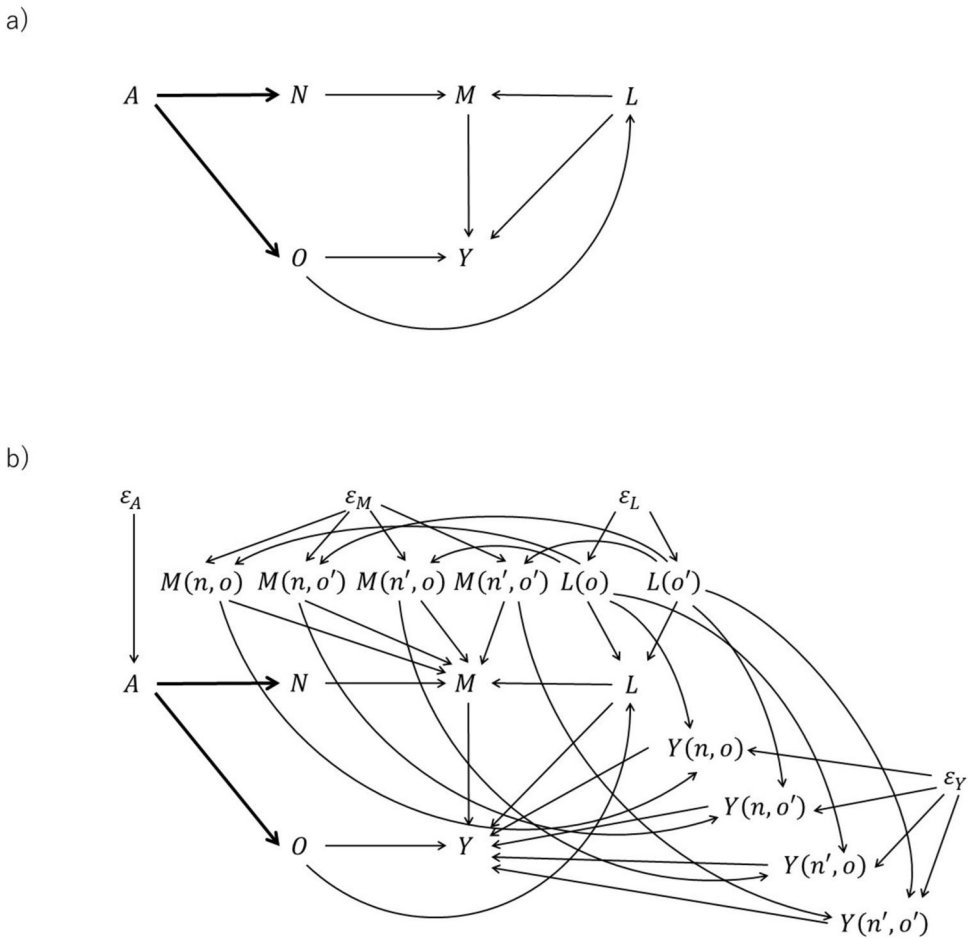

Fig. 6 **a** A causal directed acyclic graph (DAG) with exposure A, mediator M, and outcome Y, where A is assumed to be decomposed into two separable components N and O, and an exposure-induced mediator–outcome confounder L is present. We consider a scenario in which O is a parent of L, but N is not. The bold arrows from A to N and O indicate deterministic relationships. **b** A causal DAG incorporating the potential outcomes as well as error terms

Accordingly, the following nonparametric structural equations for the potential outcomes can be obtained:

$$\begin{cases} N(a) = O(a) = a\ (\forall a), \\ M(n,o) = f_M\big(n, L(n,o), \varepsilon_M\big) = g_M^*\big(n, o, \varepsilon_L, \varepsilon_M\big)\ (\forall n, o), \\ L(n,o) = L(o) = f_L\big(o, \varepsilon_L\big) = g_L^*\big(o, \varepsilon_L\big)\ (\forall n, o), \\ Y(n,o) = f_Y\big(o, M(n,o), L(n,o), \varepsilon_Y\big) = g_Y^*\big(n, o, \varepsilon_M, \varepsilon_L, \varepsilon_Y\big)\ (\forall n, o). \end{cases}$$

Because we assume that N is not a cause of L for every individual in this setting, we can write $L(n,o)$ as $L(o)$. Figure 6b shows these relationships visually; this figure is slightly more complicated than Fig. 5b because M is influenced by both N and O (via L), and we draw four potential outcomes of M. In Fig. 6b, the following assumptions hold:

$$\big(Y(n,o), L(n,o), M(n,o)\big) \perp\!\!\!\perp (N, O)\ (\forall n, o), \tag{16}$$

$$M \perp\!\!\!\perp O \,|\, (L, N), \tag{17}$$

$$Y \perp\!\!\!\perp N \,|\, (L, M, O), \tag{18}$$

$$L \perp\!\!\!\perp N \,|\, O. \tag{20}$$

Under assumptions (16), (17), (18), and (20), $E\big[Y\big(n', o\big)\big]$ is identified as

$$E\big[Y\big(n', o\big)\big] = \sum_{m,l} E[Y|M = m, L = l, A = a] P\big(M = m|L = l, A = a'\big) P(L = l|A = a),$$

and the separable direct and indirect effects are identified and given by the empirical expressions (see Online Appendix I). As explained in Online Appendix J, these effects are identified under weaker assumptions.

Finally, we consider a scenario in which both N and O are parents of L, as described in Fig. 7a. In this case, we obtain the following nonparametric structural equations for the observable variables:

$$\begin{cases} A = f_A(\varepsilon_A) = g_A(\varepsilon_A), \\ N = f_N(A) = g_N(\varepsilon_A), \\ O = f_O(A) = g_O(\varepsilon_A), \\ M = f_M(N, L, \varepsilon_M) = g_M(\varepsilon_A, \varepsilon_L, \varepsilon_M), \\ L = f_L(N, O, \varepsilon_L) = g_L(\varepsilon_A, \varepsilon_L), \\ Y = f_Y(O, M, L, \varepsilon_Y) = g_Y(\varepsilon_A, \varepsilon_M, \varepsilon_L, \varepsilon_Y). \end{cases}$$

Accordingly, the following nonparametric structural equations for the potential outcomes can be obtained:

$$\begin{cases} N(a) = O(a) = a \ (\forall a), \\ M(n,o) = f_M(n, L(n,o), \varepsilon_M) = g_M^*(n, o, \varepsilon_L, \varepsilon_M) \ (\forall n, o), \\ L(n,o) = f_L(n, o, \varepsilon_L) = g_L^*(n, o, \varepsilon_L) \ (\forall n, o), \\ Y(n,o) = f_Y(o, M(n,o), L(n,o), \varepsilon_Y) = g_Y^*(n, o, \varepsilon_M, \varepsilon_L, \varepsilon_Y) \ (\forall n, o). \end{cases}$$

Figure 7b shows these relationships visually; this figure is even more complicated than Fig. 6b because L is influenced by both N and O, and we draw four potential outcomes of L. In Fig. 7b, the following three assumptions hold:

$$\big(Y(n,o), L(n,o), M(n,o)\big) \perp\!\!\!\perp (N, O) \ (\forall n, o), \tag{16}$$

$$M \perp\!\!\!\perp O | (L, N), \tag{17}$$

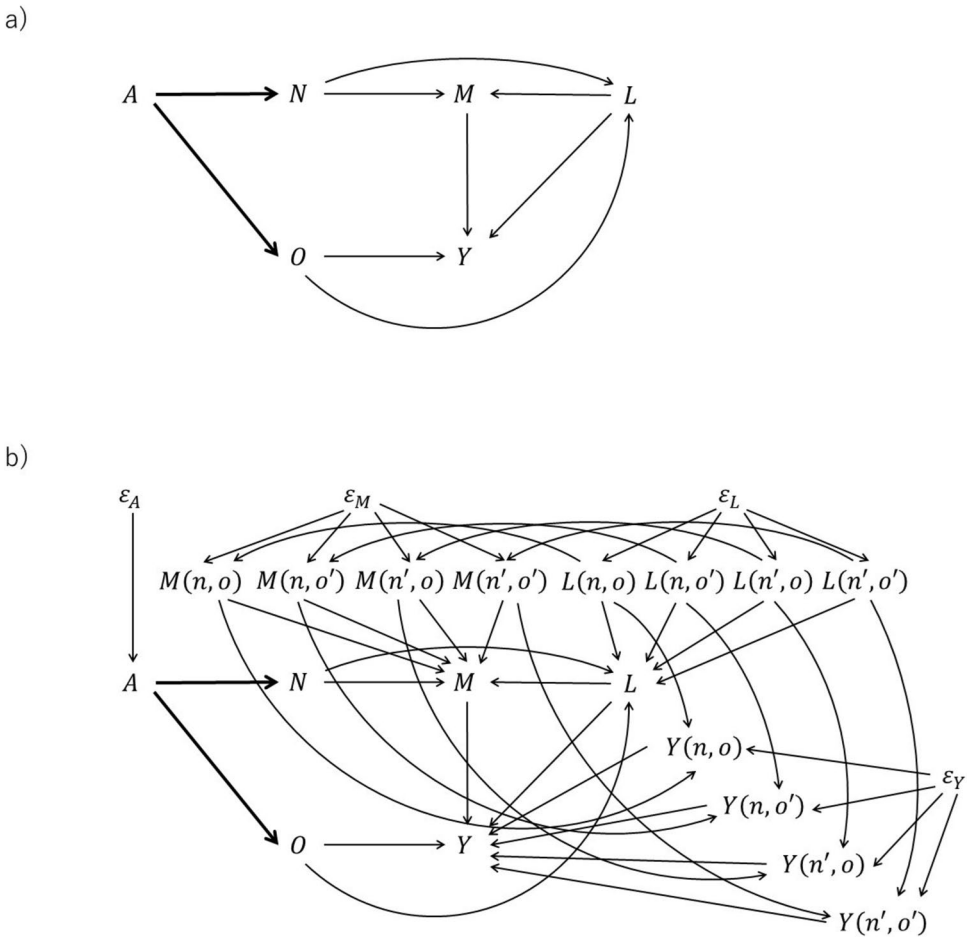

Fig. 7 **a** A causal directed acyclic graph (DAG) with exposure A, mediator M, and outcome Y, where A is assumed to be decomposed into two separable components N and O, and an exposure-induced mediator–outcome confounder L is present. We consider a scenario in which both N and O are parents of L. The bold arrows from A to N and O indicate deterministic relationships. **b** A causal DAG incorporating the potential outcomes as well as error terms

$$Y \perp\!\!\!\perp N | (L, M, O). \quad (18)$$

However, because there are direct paths $N \rightarrow L$ and $O \rightarrow L$ in Fig. 7b, the following assumptions do not *generally* hold:

$$L \perp\!\!\!\perp O | N, \quad (19)$$

$$L \perp\!\!\!\perp N | O. \quad (20)$$

Therefore, when L has both N and O as parents, $E\left[Y\left(n', o\right)\right]$ is not generally identified from the observed data, and the separable direct and indirect effects cannot generally be identified.

Philosophical Contrasts: Causation and Manipulation

We have contrasted natural effects and separable effects in the context of mediation, each of which has its own advantages and important considerations (Table 1). Despite some similarities, the two approaches differ fundamentally, with these differences likely rooted in contrasting philosophical

Table 1 Comparison of natural effects and separable effects

	Natural direct and indirect effects	Separable direct and indirect effects
Definitions[a]	NDE $\triangleq E\left[Y\left(a, M(a')\right)\right] - E\left[Y\left(a', M(a')\right)\right]$	SDE $\triangleq E[Y(n', o)] - E[Y(n', o')]$
	NIE $\triangleq E\left[Y\left(a, M(a)\right)\right] - E\left[Y\left(a, M(a')\right)\right]$	SIE $\triangleq E[Y(n, o)] - E[Y(n', o)]$
Counterfactuals	"Cross-world" counterfactual (e.g., $Y\left(a, M(a')\right)$)	"Non-cross-world" counterfactual (e.g., $Y(n', o)$)
Required properties	Positivity	Positivity[b]
	Consistency[c]	Consistency[c]
	Composition[d]	$A \equiv N \equiv O$ in the observed data
	Well-defined intervention on M	Each separable component can be intervened separately.
Sufficient independence conditions for identification[e] (when $\boldsymbol{C} = \phi$)	$Y(a, m) \perp\!\!\!\perp A\ (\forall a, m)$ (1)	$\left(Y(n, o), M(n, o)\right) \perp\!\!\!\perp (N, O)\ (\forall n, o)$ (13)
	$Y(a, m) \perp\!\!\!\perp M(a) \mid A = a\ (\forall a, m)$ (2)	$M \perp\!\!\!\perp O \mid N$ (14)
	$M(a) \perp\!\!\!\perp A\ (\forall a)$ (3)	$Y \perp\!\!\!\perp N \mid (O, M)$ (15)
	$Y(a, m) \perp\!\!\!\perp M(a')\ (\forall a, a', m)$ (4)	
Identifying formulae	$E\left[Y\left(a, M(a')\right)\right] = \sum_m E[Y \mid A = a, M = m] P(M = m \mid A = a')$	$E[Y(n', o)] = \sum_m E[Y \mid A = a, M = m] P(M = m \mid A = a')$
Advantages	Causal reasoning based on DAGs and d-separation of the observed variables suffices.	Makes questions of mediation empirically testable in future randomized controlled trials.
	Does not require considering the separable components, each of which can be intervened separately.	Does not require reference to counterfactuals indexed by m (e.g., $Y(a, m)$).
Commonly-used causal models	NPSEM-IE	FFRCISTG models
Philosophical dicta	"Causation first, manipulation second"	"No causation without manipulation"

DAG: directed acyclic graph, FFRCISTG: finest fully randomized causally interpretable structured tree graph, NDE: natural direct effect, NIE: natural indirect effect, NPSEM-IE: nonparametric structural equation models with independent errors, SDE: separable direct effect, SIE: separable indirect effect

[a]We consider an exposure A, a mediator M, and an outcome Y. We also consider that the exposure A can be decomposed into two separable components N and O, where the separable component N directly affects M but not Y; by contrast, the separable component O directly affects Y but not M. See Definitions 1 and 3 and the main text for details. Note that NDE and NIE here correspond to PDE and TIE, respectively, in the main text. SDE and SIE here correspond to SDE(n') and SIE(o), respectively, in the main text

[b]In the observed data where $A \equiv N \equiv O$, no individual has data $(N, O) = \left(n', o\right)$, and positivity for $(N, O) = \left(n', o\right)$ does not hold. However, the mediational g-formula is a function of the observed data distribution only because $N = n'$ if and only if $A = a'$ and $O = o$ if and only if $A = a$. See Technical Point 23.2 of Hernán and Robins[18]

[c]Pearl [70] argues that consistency is a theorem in the logic of counterfactuals, whereas Robins and Richardson [15] explain that the FFRCISTG model satisfies the consistency assumption

[d]Composition is needed not for identification but for interpretation

[e]Assumption (2) is equivalent to $Y(a, m) \perp\!\!\!\perp M \mid A = a\ (\forall a, m)$ under the consistency assumption. Regarding assumption (13), the following equivalence relationship holds by the weak union and decomposition graphoid axioms ($\Rightarrow$) and the contraction graphoid axiom ($\Leftarrow$) [14]: $\left(Y(n, o), M(n, o)\right) \perp\!\!\!\perp (N, O)\ (\forall n, o) \Leftrightarrow \left(Y(n, o) \perp\!\!\!\perp (N, O) \mid M(n, o)\right)\ (\forall n, o) \wedge \left(M(n, o) \perp\!\!\!\perp (N, O)\right)\ (\forall n, o)$

perspectives on causation and manipulation. One key distinction lies in the role of experimental intervention in establishing causal relationships. Recall that experiment is one of the nine Bradford Hill viewpoints [53].

Some researchers may argue that genuine causal inferences are only possible if certain variables can be carefully manipulated, asserting an ontological primacy of manipulation relative to causation. This perspective is often encapsulated by the dictum "no causation without manipulation" [54], emphasizing that causation should be understood through experimental control and intervention. From this standpoint, the separable effects are particularly valuable because they align with the principle that mediation should be empirically testable in future trials. In this context, the target trial framework has been proposed as a way to formalize causal inference by structuring observational analyses around the design principles of a hypothetical randomized experiment [55]. This approach reinforces the idea that causal questions should be framed in a way that facilitates empirical validation through intervention-based studies. When the counterfactual outcomes are not sufficiently well defined, observational data may still be quite useful for non-causal prediction [18, 56].

Others may argue that whereas manipulation and experiments play a crucial role in scientific inquiry, cause–effect relationships are ultimately governed by natural laws, many—if not most—of which extend beyond human control, asserting an ontological primacy of causation relative to manipulation. From this perspective, scientific progress has often been driven by uncovering these natural laws through careful observation, logical reasoning, and statistical inference, even in the absence of direct experimental manipulation. This perspective may be summarized by the dictum "causation first, manipulation second" [14] (p. 43), suggesting that the existence of causal relationships does not necessarily depend on our ability to manipulate variables experimentally. Within this framework, natural effects may provide valuable insights into causal mechanisms, even if we must consider the cross-world independence assumptions. In this context, the "ladder of causation" was introduced as a conceptual framework that categorizes causal reasoning into three hierarchical levels: associational, interventional, and counterfactual [57]. This hierarchy underscores the idea that causality extends beyond experimental manipulation, supporting the use of natural effects as a means of exploring mediation and causal pathways within complex systems.

Additional Remarks on Other Effects

Before concluding this review article, it is worth briefly highlighting two types of effect, which can be identified even in the presence of an exposure-induced mediator–outcome confounder. The first is the controlled direct effect. Although natural direct and indirect effects may be of greater interest in evaluating the action of various mechanisms and the importance of different pathways as well as for effect decomposition, the controlled direct effect is often of greater interest in policy evaluation. The controlled direct effect (CDE) of A on Y, setting the mediator M at a specific level m, is defined as follows [12, 13]:

Definition 5

$$\mathrm{CDE}(m) \triangleq E[Y(a,m)] - E\left[Y\left(a',m\right)\right],$$

which may vary across different levels of m. Note that indirect effects cannot be defined in a similar manner, and the difference between the total effect and the controlled direct effect cannot generally be interpreted as an indirect effect [29, 58]. The controlled direct effect is identifiable in the settings depicted in Figs. 1, 2, and 3. Specifically, this effect can be identified under assumptions (1) and (2) in Fig. 1b, assumptions (1) and (5) in Fig. 2b, and assumptions (1) and (10) in Fig. 3b (see Online Appendix K). Note that the natural direct effect and the controlled direct effect coincide when there is no interaction between the exposure and mediator. On a related issue, the proportion eliminated has also been proposed as a policy-relevant proportion for direct effects [59], which is defined on the difference scale as $\left(\left(E[Y(a)] - E\left[Y\left(a'\right)\right]\right) - \left(E[Y(a,m)] - E\left[Y\left(a',m\right)\right]\right)\right) / \left(E[Y(a)] - E\left[Y\left(a'\right)\right]\right)$. Further discussion on the controlled direct effect and the proportion eliminated can be found in the related literature [59–61].

As for the second effect, when there is an exposure-induced mediator–outcome confounder, an alternative approach is to consider randomized interventional analogs of the natural direct and indirect effects [38, 62, 63]. Apparently, these effects first appeared in Didelez et al. [64], and they are often denoted interventional direct and indirect effects, although some authors also use the label stochastic effects, which is short for stochastic interventional effects [65]. The interventional direct and indirect effects (IDE and IIE, respectively) are defined as follows:

Definition 6

$$\mathrm{IDE}\left(a'\right) \triangleq E\left[Y\left(a,G\left(a'\right)\right)\right] - E\left[Y\left(a',G\left(a'\right)\right)\right],$$
$$\mathrm{IIE}(a) \triangleq E[Y(a,G(a))] - E\left[Y\left(a,G\left(a'\right)\right)\right],$$

where $G(a)$ denotes a random draw from the distribution of the mediator among those with exposure status a. Alternatively, the interventional direct and indirect effects can be also defined as follows:

Definition 7

$$\mathrm{IDE}(a) \triangleq E[Y(a, G(a))] - E\left[Y\left(a', G(a)\right)\right],$$
$$\mathrm{IIE}\left(a'\right) \triangleq E\left[Y\left(a', G(a)\right)\right] - E\left[Y\left(a', G\left(a'\right)\right)\right].$$

Note that the interventional direct and indirect effects decompose not the total effect but the overall effect (i.e., $E[Y(a, G(a))] - E\left[Y\left(a', G\left(a'\right)\right)\right]$) [62]. The interventional direct and indirect effects are analogs arising not from fixing the mediator for each individual to the level it would have been under a particular exposure but rather from fixing it to a level that is randomly chosen from the distribution of the mediator among all those with a particular exposure [62]. In other words, unlike natural direct and indirect effects, interventional direct and indirect effects are contrasts of interventions that set the exposure to a specific value and the mediator distribution to a specific distribution; thus, it is not meaningful to talk about interventional direct and indirect effects for the individual [4]. Even in the presence of an exposure-induced mediator–outcome confounder L, interventional direct and indirect effects are identifiable from the data if—in addition to assumptions (1) and (10)—assumption (3) holds, as in Fig. 3b [38] (see Online Appendix L). Furthermore, even in the presence of an exposure-induced mediator–outcome confounder L, if assumption (4) holds in addition to assumptions (1), (3), and (10), the interventional direct and indirect effects are identified and become identical to the natural direct and indirect effects, respectively; the overall effect also becomes identical to the total effect. However, recall that assumption (4) does not *generally* hold in Fig. 3b, and when an exposure-induced mediator–outcome confounder L is present, it becomes challenging to consider specific causal structures that satisfy assumptions (1), (3), (4), and (10); see Online Appendix L for further discussion. Nguyen et al. [4] emphasized the importance of defining interventional effects more broadly to better align with the scientific research question, noting that the controlled direct effect belongs to the broader class of interventional effects; for further details, see Nguyen et al. [4].

Conclusions

To conduct meaningful mediation analysis, it is crucial to clearly define the research question of interest, and the choice of methods should align with the nature of the question and the assumptions researchers are willing to make [66, 67]. For example, although separable effects are defined without relying on any cross-world quantities and are claimed to be identifiable under assumptions that are testable in principle [15], researchers may still find it challenging to envision such a future trial in certain research settings. Indeed, as admitted by Robins et al. [16] and Hernán and Robins [18], although rare, cross-world counterfactuals can at least be conceptually observed in situations where a valid crossover trial is feasible. Regardless of whether one considers natural or separable effects, the estimated causal effect may well be interpreted as the effect of an intervention that has not actually been implemented in the observed data. In this regard, examining the underlying philosophical perspectives on causation and manipulation can provide valuable insights.

Ultimately, mediation is inherently a causal concept, and its validity relies on a well-defined causal structure and the justification of assumptions for identification. Comparing and contrasting natural effects and separable effects provides valuable insights into the foundations of mediation analysis, deepening our understanding of both its theoretical basis and practical implications.

Key References

- VanderWeele TJ. Explanation in Causal Inference: Methods for Mediation and Interaction. New York, NY: Oxford University Press; 2015.
  - This comprehensive textbook on mediation and interaction provides readers with valuable explanations and insights into the mechanisms in causal inference.
- Nguyen TQ, Schmid I, Stuart EA. Clarifying causal mediation analysis for the applied researcher: Defining effects based on what we want to learn. Psychol Methods. 2021;26:255–71. https://doi.org/10.1037/met0000299.
  - This article provides a detailed discussion about the meaning and relevance of causal estimands in mediation analysis, distinguishing between two general perspectives: the explanatory perspective and the interventional perspective.
- Nguyen TQ, Schmid I, Ogburn EL, et al. Clarifying causal mediation analysis: Effect identification via three assumptions and five potential outcomes. J Causal Inference. 2022;10:246–79. https://doi.org/10.1515/jci-2021–0049.
  - This article systematically explains the identifying assumptions in causal mediation analysis and is recommended for use alongside the companion paper on causal estimands [4].

- Robins JM, Richardson TS. Alternative graphical causal models and the identification of direct effects. In: Shrout PE, Keyes KM, Ornstein K, editors. Causality and Psychopathology: Finding the Determinants of Disorders and Their Cures. New York, NY: Oxford University Press; 2011. p. 103–58.
  - This book chapter introduces the concept of separable effects in mediation analysis, using an expanded graph in which treatment is decomposed into multiple separable components.
- Robins JM, Richardson TS, Shpitser I. An interventionist approach to mediation analysis. In: Geffner H, Dechter R, Halpern JY, editors. Probabilistic and Causal Inference: The Works of Judea Pearl. Association for Computing Machinery; 2022. p. 713–64. https://doi.org/10.1145/3501714.3501754.
  - This book chapter provides a detailed discussion of the interventionist approach to mediation analysis.
- Hernán MA, Robins JM. Causal Inference: What If. Boca Raton, FL: Chapman & Hall/CRC; 2020.
  - The chapter on causal mediation describes its theoretical framework using hypothetical interventions that can be mapped to a target trial.
- Andrews RM, Didelez V. Insights into the cross-world independence assumption of causal mediation analysis. Epidemiology. 2021;32:209–19. https://doi.org/10.1097/EDE.0000000000001313.
  - This article provides an overview of the cross-world independence assumption, discussing the relationship between assumptions for causal mediation analyses, causal models, and nonparametric identification of natural direct and indirect effects.
- Shpitser I, Richardson TS, Robins JM. Multivariate counterfactual systems and causal graphical models. In: Geffner H, Dechter R, Halpern JY, editors. Probabilistic and Causal Inference: The Works of Judea Pearl. Association for Computing Machinery; 2022. p. 813–52. https://doi.org/10.1145/3501714.3501757.
  - This book chapter describes graphical counterfactual models corresponding to the FFRCISTG models and serves as a companion chapter to that of Robins et al. [16].
- VanderWeele TJ, Vansteelandt S, Robins JM. Effect decomposition in the presence of an exposure-induced mediator-outcome confounder. Epidemiology. 2014;25:300–6. https://doi.org/10.1097/EDE.0000000000000034.
  - This article describes three approaches to effect decomposition that yield quantities interpretable as direct and indirect effects, which can be identified from data even in the presence of an exposure-induced mediator–outcome confounder.

Appendix

Regarding assumption (2), some researchers have provided the following alternative assumption [1, 29, 62, 68, 69]:

$$Y(a,m) \perp\!\!\!\perp M|A\ (\forall a,m), \quad (2^*)$$

which is stronger than assumption (2). To understand this, note that the following equivalence relationship holds:

$$Y(a,m) \perp\!\!\!\perp M|A\ (\forall a,m) \quad (2^*)$$

$$\Leftrightarrow (Y(a,m) \perp\!\!\!\perp M(a)|A=a)\ (\forall a,m) \quad (2)$$

$$\wedge\ (Y(a,m) \perp\!\!\!\perp M(a')|A=a')\ (\forall a,a',m).$$

On the right-hand side, the former condition, assumption (2), is based on "single-world" independence whereas the latter represents "cross-world" independence; like assumption (2), this holds in Fig. 1b because we use NPSEM-IE. This discussion demonstrates that assumption (2*) entails "cross-world" independence.

In the main text, we use assumption (2) to emphasize that assumptions (1) to (3) pertain to "single-world" independence whereas assumption (4) represents "cross-world" independence. A similar discussion applies to assumptions (6) and (10).

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s40471-025-00369-3.

Acknowledgements The authors are grateful to Mats J. Stensrud, Ryan M. Andrews, and Chiara Di Maria for their valuable comments on an earlier version of this manuscript. We also thank Analisa Avila, MPH, ELS, of Edanz (https://jp.edanz.com/ac) for editing a draft of the manuscript.

Author Contributions E.S., T.S., and E.Y. contributed to the conceptualization of this manuscript. E.S. conducted the literature review, drafted the manuscript, including the web appendix, and prepared all tables and

figures. E.Y. provided substantial input during the drafting process, while T.S. and E.Y. critically revised the manuscript for important intellectual content. All authors approved the final version for publication.

Funding Open Access funding provided by Okayama University. Dr. Suzuki received funding from the Japan Society for the Promotion of Science (KAKENHI grants JP19KK0418 and JP23K09740). Dr. Shinozaki received funding from the Japan Society for the Promotion of Science (KAKENHI grant JP24K14864).

Data Availability No datasets were generated or analysed during the current study.

Declarations

Competing Interests The authors declare no competing interests.

Human and Animal Rights and Informed Consent This article does not contain any studies with human or animal subjects performed by any of the authors.

References

1. VanderWeele TJ. Explanation in Causal Inference: Methods for Mediation and Interaction. New York, NY: Oxford University Press; 2015.
2. VanderWeele TJ. Mediation analysis: a practitioner's guide. Annu Rev Public Health. 2016;37:17–32. https://doi.org/10.1146/annurev-publhealth-032315-021402.
3. VanderWeele TJ. Mediation analysis. In: Lash TL, VanderWeele TJ, Haneuse S, et al., editors. Modern Epidemiology. 4th ed. Philadelphia, PA: Wolters Kluwer; 2021. p. 655–75.
4. Nguyen TQ, Schmid I, Stuart EA. Clarifying causal mediation analysis for the applied researcher: defining effects based on what we want to learn. Psychol Methods. 2021;26:255–71. https://doi.org/10.1037/met0000299.
5. Nguyen TQ, Schmid I, Ogburn EL, et al. Clarifying causal mediation analysis: effect identification via three assumptions and five potential outcomes. J Causal Inference. 2022;10:246–79. https://doi.org/10.1515/jci-2021-0049.
6. Nguyen TQ, Ogburn EL, Schmid I, et al. Causal mediation analysis: from simple to more robust strategies for estimation of marginal natural (in)direct effects. Stat Surveys. 2023;17:1–41. https://doi.org/10.1214/22-SS140.
7. Imai K, Keele L, Tingley D. A general approach to causal mediation analysis. Psychol Methods. 2010;15:309–34. https://doi.org/10.1037/a0020761.
8. Valeri L. Causal mediation analysis in pregnancy studies: the case of environmental epigenetics. Curr Epidemiol Rep. 2017;4:117–23. https://doi.org/10.1007/s40471-017-0112-1.
9. Qin X. An introduction to causal mediation analysis. Asia Pac Educ Rev. 2024;25:703–17. https://doi.org/10.1007/s12564-024-09962-5.
10. Celli V. Causal mediation analysis in economics: objectives, assumptions, models. J Econ Surv. 2022;36:214–34. https://doi.org/10.1111/joes.12452.
11. Lee H, Cashin AG, Lamb SE, et al. A guideline for reporting mediation analyses of randomized trials and observational studies: the AGReMA statement. JAMA. 2021;326:1045–56. https://doi.org/10.1001/jama.2021.14075.
12. Robins JM, Greenland S. Identifiability and exchangeability for direct and indirect effects. Epidemiology. 1992;3:143–55. https://doi.org/10.1097/00001648-199203000-00013.
13. Pearl J. Direct and indirect effects. Proceedings of the Seventeenth Conference on Uncertainty in Artificial Intelligenc. 2001;411–20. https://doi.org/10.48550/arXiv.1301.2300
14. Pearl J. Causality: Models, Reasoning, and Inference. 2nd ed. New York, NY: Cambridge University Press; 2009.
15. Robins JM, Richardson TS. Alternative graphical causal models and the identification of direct effects. In: Shrout PE, Keyes KM, Ornstein K, editors. Causality and Psychopathology: Finding the Determinants of Disorders and Their Cures. New York, NY: Oxford University Press; 2011. p. 103–58.
16. Robins JM, Richardson TS, Shpitser I. An interventionist approach to mediation analysis. In: Geffner H, Dechter R, Halpern JY, editors. Probabilistic and Causal Inference: The Works of Judea Pearl. Association for Computing Machinery; 2022. p. 713–64. https://doi.org/10.1145/3501714.3501754.
17. Didelez V. Defining causal mediation with a longitudinal mediator and a survival outcome. Lifetime Data Anal. 2019;25:593–610. https://doi.org/10.1007/s10985-018-9449-0.
18. Hernán MA, Robins JM. Causal Inference: What If. Boca Raton, FL: Chapman & Hall/CRC; 2020.
19. Di Maria C, Didelez V. Longitudinal mediation analysis with multilevel and latent growth models: a separable effects causal approach. BMC Med Res Methodol. 2024;24: 248. https://doi.org/10.1186/s12874-024-02358-4.
20. Andrews RM, Shpitser I, Didelez V, et al. Examining the causal mediating role of cardiovascular disease on the effect of subclinical cardiovascular disease on cognitive impairment via separable effects. J Gerontol A Biol Sci Med Sci. 2023;78:1172–8. https://doi.org/10.1093/gerona/glad077.
21. Robins JM. 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. 1986;7:1393–512. https://doi.org/10.1016/0270-0255(86)90088-6.
22. Schuler MS, Coffman DL, Stuart EA, et al. Practical challenges in mediation analysis: a guide for applied researchers. Health Serv Outcomes Res Methodol. 2025;25:57–84. https://doi.org/10.1007/s10742-024-00327-4.
23. Greenland S, Pearl J, Robins JM. Causal diagrams for epidemiologic research. Epidemiology. 1999;10:37–48. https://doi.org/10.1097/00001648-199901000-00008.
24. Suzuki E, Shinozaki T, Yamamoto E. Causal diagrams: pitfalls and tips. J Epidemiol. 2020;30:153–62. https://doi.org/10.2188/jea.JE20190192.
25. Westreich D, Cole SR. Invited commentary: positivity in practice. Am J Epidemiol. 2010;171:674–7. https://doi.org/10.1093/aje/kwp436.
26. Petersen ML, Porter KE, Gruber S, et al. Diagnosing and responding to violations in the positivity assumption. Stat Methods Med Res. 2012;21:31–54. https://doi.org/10.1177/0962280210386207.
27. Cole SR, Frangakis CE. The consistency statement in causal inference: a definition or an assumption? Epidemiology. 2009;20:3–5. https://doi.org/10.1097/EDE.0b013e31818ef366.

28. VanderWeele TJ. Concerning the consistency assumption in causal inference. Epidemiology. 2009;20:880–3. https://doi.org/10.1097/EDE.0b013e3181bd5638.
29. VanderWeele TJ, Vansteelandt S. Conceptual issues concerning mediation, interventions and composition. Statist Interf. 2009;2:457–68. https://doi.org/10.4310/SII.2009.v2.n4.a7.
30. Shpitser I, VanderWeele TJ. A complete graphical criterion for the adjustment formula in mediation analysis. Int J Biostat. 2011;7: 16. https://doi.org/10.2202/1557-4679.1297.
31. Vansteelandt S. Estimation of direct and indirect effects. In: Berzuini C, Dawid P, Bernardinelli L, editors. Causality: Statistical Perspectives and Applications. Hoboken, NJ: Wiley; 2012. p. 126–50.
32. Hafeman DM. A sufficient cause based approach to the assessment of mediation. Eur J Epidemiol. 2008;23:711–21. https://doi.org/10.1007/s10654-008-9286-7.
33. Hafeman DM, Schwartz S. Opening the Black Box: a motivation for the assessment of mediation. Int J Epidemiol. 2009;38:838–45. https://doi.org/10.1093/ije/dyn372.
34. Suzuki E, Yamamoto E, Tsuda T. Identification of operating mediation and mechanism in the sufficient-component cause framework. Eur J Epidemiol. 2011;26:347–57. https://doi.org/10.1007/s10654-011-9568-3.
35. Rothman KJ. Causes. Am J Epidemiol. 1976;104:587–92. https://doi.org/10.1093/oxfordjournals.aje.a112335.
36. Suzuki E. Unraveling causality: Innovations in epidemiologic methods. JMA J. 2025;8:323–37. https://doi.org/10.31662/jmaj.2024-0246.
37. Andrews RM, Didelez V. Insights into the cross-world independence assumption of causal mediation analysis. Epidemiology. 2021;32:209–19. https://doi.org/10.1097/EDE.0000000000001313.
38. Vansteelandt S, Daniel RM. Interventional effects for mediation analysis with multiple mediators. Epidemiology. 2017;28:258–65. https://doi.org/10.1097/EDE.0000000000000596.
39. Shpitser I, Richardson TS, Robins JM. Multivariate counterfactual systems and causal graphical models. In: Geffner H, Dechter R, Halpern JY, editors. Probabilistic and Causal Inference: The Works of Judea Pearl. Association for Computing Machinery; 2022. p. 813–52. https://doi.org/10.1145/3501714.3501757
40. Glymour MM, Greenland S. Causal diagrams. In: Rothman KJ, Greenland S, Lash TL, editors. Modern Epidemiology. 3rd ed. Philadelphia, PA: Lippincott Williams & Wilkins; 2008. p. 183–209.
41. VanderWeele TJ, Rothman KJ. Formal causal models. In: Lash TL, VanderWeele TJ, Haneuse S, et al., editors. Modern Epidemiology. 4th ed. Philadelphia, PA: Wolters Kluwer; 2021. p. 33–51.
42. Pearl J. The causal mediation formula: a guide to the assessment of pathways and mechanisms. Prev Sci. 2012;13:426–36. https://doi.org/10.1007/s11121-011-0270-1.
43. VanderWeele TJ, Tchetgen Tchetgen EJ. Mediation analysis with time varying exposures and mediators. J R Stat Soc Series B Stat Methodol. 2017;79:917–38. https://doi.org/10.1111/rssb.12194.
44. Tchetgen Tchetgen EJ, VanderWeele TJ. Identification of natural direct effects when a confounder of the mediator is directly affected by exposure. Epidemiology. 2014;25:282–91. https://doi.org/10.1097/EDE.0000000000000054.
45. Robins JM. Semantics of causal DAG models and the identification of direct and indirect effects. In: Green PJ, Hjort NL, Richardson S, editors. Highly Structured Stochastic Systems. New York, NY: Oxford University Press; 2003. p. 70–81.
46. Stensrud MJ, Hernán MA, Tchetgen Tchetgen EJ, et al. A generalized theory of separable effects in competing event settings. Lifetime Data Anal. 2021;27:588–631. https://doi.org/10.1007/s10985-021-09530-8.
47. Stensrud MJ, Young JG, Didelez V, et al. Separable effects for causal inference in the presence of competing events. J Am Stat Assoc. 2022;117:175–83. https://doi.org/10.1080/01621459.2020.1765783.
48. Stensrud MJ, Robins JM, Sarvet A, et al. Conditional separable effects. J Am Stat Assoc. 2023;118:2671–83. https://doi.org/10.1080/01621459.2022.2071276.
49. Janvin M, Young JG, Ryalen PC, et al. Causal inference with recurrent and competing events. Lifetime Data Anal. 2024;30:59–118. https://doi.org/10.1007/s10985-023-09594-8.
50. Wen L, Sarvet AL, Stensrud MJ. Causal effects of intervening variables in settings with unmeasured confounding. J Mach Learn Res. 2024;25: Article 345.
51. Richardson TS, Robins JM. Single world intervention graphs (SWIGs): a unification of the counterfactual and graphical approaches to causality. Center for Statistics and the Social Sciences, University of Washington, Working Paper. 2013;128
52. Richardson TS, Robins JM. Thomas S. Richardson and James M. Robins' contribution to the Discussion of 'Parameterizing and simulating from causal models' by Evans and Didelez. J R Stat Soc Series B Stat Methodol. 2024;86:578–80. https://doi.org/10.1093/jrsssb/qkae020.
53. Hill AB. The environment and disease: association or causation? Proc R Soc Med. 1965;58:295–300. https://doi.org/10.1177/003591576505800503.
54. Holland PW. Statistics and causal inference. J Am Stat Assoc. 1986;81:945–60. https://doi.org/10.1080/01621459.1986.10478354.
55. Hernán MA, Robins JM. Using big data to emulate a target trial when a randomized trial is not available. Am J Epidemiol. 2016;183:758–64. https://doi.org/10.1093/aje/kwv254.
56. Hernán MA, Hsu J, Healy B. A second chance to get causal inference right: a classification of data science tasks. Chance. 2019;32:42–9. https://doi.org/10.1080/09332480.2019.1579578.
57. Pearl J, Mackenzie D. The Book of Why: The New Science of Cause and Effect. 1st ed. New York, NY: Basic Books; 2018.
58. Kaufman JS, MacLehose RF, Kaufman S. A further critique of the analytic strategy of adjusting for covariates to identify biologic mediation. Epidemiol Perspect Innov. 2004;1: 4. https://doi.org/10.1186/1742-5573-1-4.
59. VanderWeele TJ. Policy-relevant proportions for direct effects. Epidemiology. 2013;24:175–6. https://doi.org/10.1097/EDE.0b013e3182781410.
60. Suzuki E, Mitsuhashi T, Tsuda T, et al. Alternative definitions of "proportion eliminated." Epidemiology. 2014;25:308–9. https://doi.org/10.1097/EDE.0000000000000050.
61. Suzuki E, Evans D, Chaix B, et al. On the "proportion eliminated" for risk differences versus excess relative risks. Epidemiology. 2014;25:309–10. https://doi.org/10.1097/EDE.0000000000000060.
62. VanderWeele TJ, Vansteelandt S, Robins JM. Effect decomposition in the presence of an exposure-induced mediator-outcome confounder. Epidemiology. 2014;25:300–6. https://doi.org/10.1097/EDE.0000000000000034.
63. Moreno-Betancur M, Carlin JB. Understanding interventional effects: a more natural approach to mediation analysis? Epidemiology. 2018;29:614–7. https://doi.org/10.1097/EDE.0000000000000866.
64. Didelez V, Dawid AP, Geneletti S. Direct and indirect effects of sequential treatments. Proceedings of the Twenty-Second Conference on Uncertainty in Artificial Intelligence. 2006;138–46. https://doi.org/10.48550/arXiv.1206.6840.
65. Rudolph KE, Sofrygin O, Zheng W, et al. Robust and flexible estimation of stochastic mediation effects: a proposed method

and example in a randomized trial setting. Epidemiol Methods. 2018;7. https://doi.org/10.1515/em-2017-0007.
66. Shrier I, Suzuki E. The primary importance of the research question: implications for understanding natural versus controlled direct effects. Int J Epidemiol. 2022;51:1041–6. https://doi.org/10.1093/ije/dyac090.
67. Suzuki E. L or M_1—Critical challenges in mediation analysis. Epidemiology. 2025;36:686–9. https://doi.org/10.1097/EDE.0000000000001888.
68. VanderWeele TJ. Marginal structural models for the estimation of direct and indirect effects. Epidemiology. 2009;20:18–26. https://doi.org/10.1097/EDE.0b013e31818f69ce.
69. Vansteelandt S, VanderWeele TJ. Natural direct and indirect effects on the exposed: effect decomposition under weaker assumptions. Biometrics. 2012;68:1019–27. https://doi.org/10.1111/j.1541-0420.2012.01777.x.
70. Pearl J. On the consistency rule in causal inference: axiom, definition, assumption, or theorem? Epidemiology. 2010;21:872–5. https://doi.org/10.1097/EDE.0b013e3181f5d3fd.