Online Appendix A: Identification of natural direct and indirect effects

Proposition 1

If assumptions (1) to (4) hold, the natural direct effect is identified and given by

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

and the natural indirect effect is identified and given by

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

Proof:

We have

$$\begin{aligned}
E\big[Y\big(a, M(a')\big)\big] &= \sum_m E\big[Y\big(a, M(a')\big)|M(a') = m\big]P(M(a') = m) \quad (\because) \text{ law of total expectation} \\
&= \sum_m E[Y(a, m)|M(a') = m]P(M(a') = m) \\
&= \sum_m E[Y(a, m)]P(M(a') = m|A = a') \quad (\because)\ Y(a, m) \perp\!\!\!\perp M(a'),\ \ M(a') \perp\!\!\!\perp A \\
&= \sum_m E[Y(a, m)|A = a]P(M = m|A = a') \quad (\because)\ Y(a, m) \perp\!\!\!\perp A, \text{consistency, positivity} \\
&= \sum_m E[Y(a, m)|A = a, M(a) = m]P(M = m|A = a') \quad (\because)\ Y(a, m) \perp\!\!\!\perp M(a)|A = a \\
&= \sum_m E[Y|A = a, M = m]P(M = m|A = a'). \quad (\because) \text{ consistency, positivity}
\end{aligned}$$

See Nguyen et al. [1] for an explanation of the positivity assumption. If we apply this result and replace a with a', we get

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

Thus, the natural direct effect is given by

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

If we apply the result and replace a' with a, we get

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

Thus, the natural indirect effect is given by

$$E\big[Y\big(a, M(a)\big)\big] - E\big[Y\big(a, M(a')\big)\big] = \sum_m E[Y|A = a, M = m]\big(P(M = m|A = a) - P(M = m|A = a')\big). \quad \blacksquare$$

The expression for the natural indirect effect is sometimes referred to as the “mediation formula” [2].

Online Appendix B: Identification of the total effect

Proposition 2

If $\left(Y(a,m),M(a)\right) \perp\!\!\!\perp I(A=a)\ (\forall a,m)$ holds, where $I(A=a)$ is the Bernoulli indicator random variable, the total effect is identified and given by

$$E[Y(a)]-E[Y(a')]=E[Y|A=a]-E[Y|A=a'].$$

Proof:

We have

$$\begin{aligned}
E[Y(a)] &= E\left[Y\left(a,M(a)\right)\right] \quad (\because)\ \text{composition}\\
&= \sum_m E\left[Y\left(a,M(a)\right)|M(a)=m\right]P(M(a)=m) \quad (\because)\ \text{law of total expectation}\\
&= \sum_m E[Y(a,m)|M(a)=m]P(M(a)=m)\\
&= \sum_m E[Y(a,m)|A=a,M(a)=m]P(M(a)=m|A=a) \quad (\because)\ \left(Y(a,m),M(a)\right) \perp\!\!\!\perp I(A=a)\\
&= \sum_m E[Y|A=a,M=m]P(M=m|A=a) \quad (\because)\ \text{consistency, positivity}\\
&= E[Y|A=a], \quad (\because)\ \text{law of total expectation}
\end{aligned}$$

where, in the fourth equation, the following equivalence relationship holds by the weak union and decomposition graphoid axioms (⇒) and the contraction graphoid axiom (⇐) [3]:

$$\left(Y(a,m),M(a)\right) \perp\!\!\!\perp I(A=a) \Leftrightarrow \left(Y(a,m) \perp\!\!\!\perp I(A=a)|M(a)\right) \wedge \left(M(a) \perp\!\!\!\perp I(A=a)\right).$$

See Nguyen et al. [1] for an explanation of the positivity assumption. If we apply this result and replace a with a', we get

$$E[Y(a')]=E[Y|A=a'].$$

Thus, the total effect is given by

$$E[Y(a)] - E[Y(a')] = E[Y|A = a] - E[Y|A = a']. \quad \blacksquare$$

Note that $\big(Y(a,m), M(a)\big) \perp\!\!\!\perp I(A = a)\ (\forall a, m)$ is weaker than assumptions (1) to (4), as below:

$$\begin{aligned} P(Y(a,m) = y, M(a) = m|A = a) &= P(Y(a,m) = y|A = a)P(M(a) = m|A = a) \quad (\because)\ Y(a,m) \perp\!\!\!\perp M(a)|A = a \\ &= P(Y(a,m) = y)P(M(a) = m) \quad (\because)\ Y(a,m) \perp\!\!\!\perp A,\ \ M(a) \perp\!\!\!\perp A \\ &= P(Y(a,m) = y, M(a) = m), \quad (\because)\ Y(a,m) \perp\!\!\!\perp M(a) \end{aligned}$$

where, in the third equation, $Y(a,m) \perp\!\!\!\perp M(a)$ is weaker than assumption (4). Thus, by definition, the expressions for $E\big[Y\big(a, M(a)\big)\big]$ and $E\big[Y\big(a', M(a')\big)\big]$ in Online Appendix A reduce to $E[Y|A = a]$ and $E[Y|A = a']$, respectively. In causal mediation analyses, it may be common to assume that the exposure *A* is (conditionally) randomized, such that $\big(Y(a,m), M(a)\big) \perp\!\!\!\perp A\ (\forall a, m)$ holds, which is not generally weaker than assumptions (1) to (4). However, when *A* is a binary variable, $\big(Y(a,m), M(a)\big) \perp\!\!\!\perp A\ (\forall a, m)$ is equivalent to $\big(Y(a,m), M(a)\big) \perp\!\!\!\perp I(A = a)\ (\forall a, m)$, which is (as shown above) weaker than assumptions (1) to (4).

By the same logic, $\big(Y(a,m), M(a)\big) \perp\!\!\!\perp I(A = a)|H\ (\forall a, m)$ is weaker than assumptions (5) to (8). If $\big(Y(a,m), M(a)\big) \perp\!\!\!\perp I(A = a)|H\ (\forall a, m)$ holds, as in Figure 2b, the total effect is identified and given by

$$E[Y(a)] - E[Y(a')] = \sum_h (E[Y|A = a, H = h] - E[Y|A = a', H = h])P(H = h).$$

However, note that $\big(Y(a,m), M(a)\big) \perp\!\!\!\perp I(A = a)\ (\forall a, m)$ also holds in Figure 2b; thus, the total effect is identified and given by $E[Y|A = a] - E[Y|A = a']$ (see Proposition 2). Indeed, the expressions for $E\big[Y\big(a, M(a)\big)\big]$ and $E\big[Y\big(a', M(a')\big)\big]$ in Online Appendix C become $\sum_h E[Y|A = a, H = h]P(H = h)$ and $\sum_h E[Y|A = a', H = h]P(H = h)$, respectively. Note that they reduce to $E[Y|A = a]$ and $E[Y|A = a']$, respectively, because *A* and *H* are d-separated in Figure 2b, which implies $P(H = h) = P(H = h|A = a) = P(H = h|A = a')$ holds.

Finally, In Figure 3b, $\big(Y(a,m), M(a)\big) \perp\!\!\!\perp I(A = a)|L\ (\forall a, m)$ does not generally hold. However, $\big(Y(a,m), M(a)\big) \perp\!\!\!\perp I(A = a)\ (\forall a, m)$ holds, and the total effect can be identified and given by $E[Y|A = a] - E[Y|A = a']$ (see Proposition 2).

Online Appendix C: Identification of natural direct and indirect effects when there is a mediator–outcome confounder *H*

Proposition 3

If assumptions (5) to (8) hold, the natural direct effect is identified and given by

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

and the natural indirect effect is identified and given by

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

Proof:

We have

$$\begin{aligned}
E\big[Y\big(a, M(a')\big)\big] &= \sum_h \sum_m E\big[Y\big(a, M(a')\big)|M(a') = m, H = h\big]P(M(a') = m|H = h)P(H = h) \quad (\because) \text{ law of total expectation} \\
&= \sum_h \sum_m E[Y(a, m)|M(a') = m, H = h]P(M(a') = m|H = h)P(H = h) \\
&= \sum_h \sum_m E[Y(a, m)|H = h]P(M(a') = m|A = a', H = h)P(H = h) \quad (\because)\ Y(a, m) \perp\!\!\!\perp M(a')|H,\ \ M(a') \perp\!\!\!\perp A|H \\
&= \sum_h \sum_m E[Y(a, m)|A = a, H = h]P(M = m|A = a', H = h)P(H = h) \quad (\because)\ Y(a, m) \perp\!\!\!\perp A|H, \text{consistency, positivity} \\
&= \sum_h \sum_m E[Y(a, m)|A = a, M(a) = m, H = h]P(M = m|A = a', H = h)P(H = h) \quad (\because)\ Y(a, m) \perp\!\!\!\perp M(a)|(A = a, H) \\
&= \sum_h \sum_m E[Y|A = a, M = m, H = h]P(M = m|A = a', H = h)P(H = h). \quad (\because) \text{ consistency, positivity}
\end{aligned}$$

See Nguyen et al. [1] for an explanation of the positivity assumption. If we apply this result and replace a with a', we get

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

Thus, the natural direct effect is given by

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

If we apply the result and replace a' with a, we get

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

Thus, the natural indirect effect is given by

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

The expression for the natural indirect effect is sometimes referred to as the “mediation formula” [2].

Proposition 4

If assumptions (1), (3), (4), and (6) hold, the natural direct effect is identified and given by

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

and the natural indirect effect is identified and given by

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

Proof:

We have

$$\begin{aligned}
E\big[Y\big(a,M(a')\big)\big]&=\sum_m E\big[Y\big(a,M(a')\big)|M(a')=m\big]P(M(a')=m)\quad (\because)\ \text{law of total expectation}\\
&=\sum_m E[Y(a,m)|M(a')=m]P(M(a')=m)\\
&=\sum_m E[Y(a,m)]P(M(a')=m)\quad (\because)\ Y(a,m)\perp\!\!\!\perp M(a')\\
&=\sum_m E[Y(a,m)|A=a]P(M(a')=m|A=a')\quad (\because)\ Y(a,m)\perp\!\!\!\perp A,\ \ M(a')\perp\!\!\!\perp A\\
&=\sum_m\left\{\sum_h E[Y(a,m)|A=a,H=h]P(H=h|A=a)\right\}P(M(a')=m|A=a')\quad (\because)\ \text{law of total expectation, positivity}\\
&=\sum_h\sum_m E[Y(a,m)|A=a,H=h]P(H=h|A=a)P(M(a')=m|A=a')\\
&=\sum_h\sum_m E[Y(a,m)|A=a,M(a)=m,H=h]P(H=h|A=a)P(M(a')=m|A=a')\quad (\because)\ Y(a,m)\perp\!\!\!\perp M(a)|(A=a,H)
\end{aligned}$$

$$= \sum_h \sum_m E[Y|A = a, M = m, H = h]P(H = h|A = a)P(M = m|A = a'). \quad (\because) \text{ consistency, positivity}$$

See Nguyen et al. [1] for an explanation of the positivity assumption. If we apply this result and replace a with a', we get

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

Thus, the natural direct effect is given by

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

If we apply the result and replace a' with a, we get

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

Thus, the natural indirect effect is given by

$$E\big[Y\big(a, M(a)\big)\big] - E\big[Y\big(a, M(a')\big)\big] = \sum_h \sum_m E[Y|A = a, M = m, H = h]P(H = h|A = a)\big(P(M = m|A = a) - P(M = m|A = a')\big). \quad \blacksquare$$

These are equivalent to the empirical formulae provided by VanderWeele [4]. In Figure 2b, although assumptions (1), (3), and (6) hold, assumption (4) does not *generally* hold. By contrast, assumptions (5) to (8) hold in Figure 2b. Therefore, in this context, the natural direct and indirect effects are readily identified using Proposition 3.

Online Appendix D: Isolation assumption and dismissible component condition

For simplicity, we consider a binary exposure *X* and a binary outcome *Y*, and we do not consider a setting of mediation. We let *Y*(*x*) denote the potential outcomes of *Y* if, possibly contrary to fact, there had been interventions to set *X* to *x*. In this case, there are four response types, as shown in Table S1. Note that the following equations hold:

$$P(Y(1) = 1) = r_1 + r_2,$$
$$P(Y(0) = 1) = r_1 + r_3,$$
$$P(Y = 1) = (p_1 + p_2)\pi + (q_1 + q_3)(1 - \pi) = r_1 + p_2\pi + q_3(1 - \pi),$$

where $\pi = P(X = 1)$ is the prevalence of the exposed group in the study population. These three quantities become identical if and only if $r_2 = r_3 = p_2\pi + q_3(1 - \pi)$ holds. Here, we describe the following four conditions:

1. Independence between *X* and *Y* (i.e., d-separation between *X* and *Y* in a causal DAG):

$$\begin{aligned} & Y \perp\!\!\!\perp X \\ & \quad \Leftrightarrow P(Y = 1|X = 1) = P(Y = 1|X = 0) \\ & \quad \Leftrightarrow p_1 + p_2 = q_1 + q_3. \end{aligned}$$

2. Exchangeability of *Y*(*x*) across *X* (i.e., no open backdoor paths from *X* to *Y* in a causal DAG):

$$\begin{aligned} & Y(x) \perp\!\!\!\perp X \ (x = 0,1) \\ & \quad \Leftrightarrow \big(P(Y(1) = 1|X = 1) = P(Y(1) = 1|X = 0)\big) \wedge \big(P(Y(0) = 1|X = 1) = P(Y(0) = 1|X = 0)\big) \\ & \quad \Leftrightarrow (p_1 + p_2 = q_1 + q_2) \wedge (p_1 + p_3 = q_1 + q_3). \end{aligned}$$

3. Dismissible component condition of *X* on *Y*(*x*):

$$\begin{aligned} & P(Y(1) = 1) = P(Y(0) = 1) \\ & \quad \Leftrightarrow r_2 = r_3 \\ & \quad \Leftrightarrow p_2\pi + q_2(1 - \pi) = p_3\pi + q_3(1 - \pi). \end{aligned}$$

4. Isolation assumption of X on $Y(x)$ (i.e., no direct arrow from X to Y in a causal DAG):

$$
\begin{aligned}
&Y(1) = Y(0) \text{ for all individuals} \\
&\qquad \Leftrightarrow r_2 = r_3 = 0 \\
&\qquad \Leftrightarrow P(Y(1) = 1) = P(Y(0) = 1) = P(Y = 1) = r_1 \\
&\qquad \Rightarrow r_2 = r_3 = p_2\pi + q_3(1 - \pi) \\
&\qquad \Leftrightarrow P(Y(1) = 1) = P(Y(0) = 1) = P(Y = 1).
\end{aligned}
$$

Three points are worth mentioning. First, although the isolation assumption is an individual-level assumption, the dismissible component condition is a population-level assumption [5, 6]. Thus, the isolation assumption implies the dismissible component condition. Second, if both the exchangeability condition and the isolation assumption hold, independence between X and Y holds. Graphically, this can be explained as follows: if there are no backdoor path(s) from X to Y and no direct arrows from X to Y, we can say that X and Y are d-separated. Third, under the exchangeability condition, the dismissible component condition is equivalent to the independence between X and Y. We show its proof below.

Proof:

$P(Y(x) = y) = P(Y(x') = y)$

$\Leftrightarrow P(Y(x) = y|X = x) = P(Y(x') = y|X = x') \quad (\because)\ Y(x) \perp\!\!\!\perp X$

$\Leftrightarrow P(Y = y|X = x) = P(Y = y|X = x') \quad (\because)$ consistency

$\Leftrightarrow Y \perp\!\!\!\perp X.$ ■

Table S1 Response types when considering a binary exposure *X* and a binary outcome *Y*

Response types	Potential outcomes $Y(x)$		Distributions in		
	$Y(1)$	$Y(0)$	Exposed group	Unexposed group	Total population[a]
Doomed	1	1	p_1	q_1	r_1
Causal	1	0	p_2	q_2	r_2
Preventive	0	1	p_3	q_3	r_3
Immune	0	0	p_4	q_4	r_4
Total			1	1	1

[a]Note that r_i can be calculated as $p_i \times P(X=1) + q_i \times P(X=0)$ $(i = 1,2,3,4)$, where $P(X=x)$ represents the prevalence of $X = x$ in the study population.

Online Appendix E: Identification of separable direct and indirect effects

Proposition 5

If assumptions (13) to (15) hold, the separable direct effect is identified and given by

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

and the separable indirect effect is identified and given by

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

Proof:

We have

$$\begin{aligned}
E[Y(n',o)] &= \sum_m E[Y(n',o)|M(n',o)=m]P(M(n',o)=m) \quad (\because)\text{ law of total expectation} \\
&= \sum_m E[Y(n',o)|M(n',o)=m, N=n', O=o]P(M(n',o)=m|N=n', O=o) \quad (\because)\,\big(Y(n',o), M(n',o)\big) \perp\!\!\!\perp (N,O) \\
&= \sum_m E[Y|M=m, N=n', O=o]P(M=m|N=n', O=o) \quad (\because)\text{ consistency} \\
&= \sum_m E[Y|M=m, N=n, O=o]P(M=m|N=n', O=o') \quad (\because)\, Y \perp\!\!\!\perp N|(O,M), M \perp\!\!\!\perp O|N \\
&= \sum_m E[Y|A=a, M=m]P(M=m|A=a'), \quad (\because)\text{ determinism, positivity}
\end{aligned}$$

where, in the second equation, the following equivalence relationship holds by the weak union and decomposition graphoid axioms (⇒) and the contraction graphoid axiom (⇐) [3]:

$$\big(Y(n',o), M(n',o)\big) \perp\!\!\!\perp (N,O) \Leftrightarrow \big(Y(n',o) \perp\!\!\!\perp (N,O) | M(n',o)\big) \wedge \big(M(n',o) \perp\!\!\!\perp (N,O)\big).$$

If we apply this result and replace o with o', we get

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

Thus, the separable direct effect is given by

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

If we apply the result and replace n' with n, we get

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

Thus, the separable indirect effect is given by

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

Online Appendix F: Identification of separable direct and indirect effects with weaker assumptions

Lemma 1

For some $x \in \{0,1\}$ and $x^* = 1 - x$, the following relationships hold:

$$\begin{pmatrix} \big(Y(a), M(a)\big) \perp\!\!\!\perp A\ (\forall a) & (13^*) \\ P(M(n = x, o = 0) = m) = P(M(n = x, o = 1) = m) & (14^*) \\ P(Y(n = 1, o = x^*) = y | M(n = 1, o = x^*) = m) = P(Y(n = 0, o = x^*) = y | M(n = 0, o = x^*) = m) & (15^*) \end{pmatrix}$$

$$\Leftarrow \begin{pmatrix} \big(Y(n, o), M(n, o)\big) \perp\!\!\!\perp (N, O)\ (\forall n, o) \\ P(M(n = x, o = 0) = m) = P(M(n = x, o = 1) = m) \\ P(Y(n = 1, o = x^*) = y | M(n = 1, o = x^*) = m) = P(Y(n = 0, o = x^*) = y | M(n = 0, o = x^*) = m) \end{pmatrix}$$

$$\Leftrightarrow \begin{pmatrix} \big(Y(n, o), M(n, o)\big) \perp\!\!\!\perp (N, O)\ (\forall n, o) \\ M \perp\!\!\!\perp O | N = x \\ Y \perp\!\!\!\perp N | (O = x^*, M) \end{pmatrix}$$

$$\Leftarrow \begin{pmatrix} \big(Y(n, o), M(n, o)\big) \perp\!\!\!\perp (N, O)\ (\forall n, o) & (13) \\ M \perp\!\!\!\perp O | N & (14) \\ Y \perp\!\!\!\perp N | (O, M) & (15) \end{pmatrix}.$$

Proof:

The relationship between the first and second sets of assumptions trivially holds from the deterministic relationship between *A*, *N*, and *O*. Additionally, the relationship between the third and fourth sets of assumptions also trivially holds. We here provide a proof for the equivalence relationship between the second and third sets of assumptions, using the following equivalence relationship, which holds by the weak union and decomposition graphoid axioms (⇒) and the contraction graphoid axiom (⇐) [3]:

$$\big(Y(n, o), M(n, o)\big) \perp\!\!\!\perp (N, O)\ (\forall n, o) \Leftrightarrow \big(Y(n, o) \perp\!\!\!\perp (N, O) | M(n, o)\big)\ (\forall n, o) \wedge \big(M(n, o) \perp\!\!\!\perp (N, O)\big)\ (\forall n, o).$$

$P(M(n = x, o = 0) = m) = P(M(n = x, o = 1) = m)$

$\Leftrightarrow P(M(n = x, o = 0) = m | N = x, O = 0) = P(M(n = x, o = 1) = m | N = x, O = 1) \quad (\because)\ M(n, o) \perp\!\!\!\perp (N, O)$

$\Leftrightarrow P(M = m | N = x, O = 0) = P(M = m | N = x, O = 1) \quad (\because)$ consistency

$\Leftrightarrow M \perp\!\!\!\perp O | N = x.$

$P(Y(n=1,o=x^*)=y|M(n=1,o=x^*)=m)=P(Y(n=0,o=x^*)=y|M(n=0,o=x^*)=m)$
$\Leftrightarrow P(Y(n=1,o=x^*)=y \mid M(n=1,o=x^*)=m, N=1, O=x^*)=P(Y(n=0,o=x^*)=y \mid M(n=0,o=x^*)=m, N=0, O=x^*)$
$(\because)\ Y(n,o) \perp\!\!\!\perp (N,O)|M(n,o)$
$\Leftrightarrow P(Y=y\ |M=m, N=1, O=x^*)=P(Y=y\ |M=m, N=0, O=x^*)$ $(\because)$ consistency
$\Leftrightarrow Y \perp\!\!\!\perp N|(O=x^*, M)$. ■

Assumptions (14*) and (15*) are equivalent to Equations (38.14) and (38.15) in Robins et al. [7], respectively, which are sometimes referred to as dismissible component conditions [5, 6]; see Online Appendix D for a related discussion.

Lemma 2

For some $x \in \{0,1\}$ and $x^* = 1 - x$, If assumptions (13*) to (15*) in Lemma 1 hold, the following equation holds:

$$P(Y(n = x, o = x^*) = y) = \sum_m P(Y = y|M = m, A = x^*)P(M = m|A = x).$$

Proof:

We have

$$\begin{aligned}
P(Y(n = x, o = x^*) = y) &= \sum_m P(Y(n = x, o = x^*) = y|M(n = x, o = x^*) = m)P(M(n = x, o = x^*) = m) \quad (\because) \text{ law of total probability} \\
&= \sum_m P(Y(n = x^*, o = x^*) = y|M(n = x^*, o = x^*) = m)P(M(n = x, o = x) = m) \quad (\because)\ (14^*), (15^*) \\
&= \sum_m P(Y(a = x^*) = y|M(a = x^*) = m)P(M(a = x) = m) \quad (\because) \text{ determinism} \\
&= \sum_m P(Y(a = x^*) = y|M(a = x^*) = m, A = x^*)P(M(a = x) = m|A = x) \quad (\because)\ (13^*) \\
&= \sum_m P(Y = y|M = m, A = x^*)P(M = m|A = x), \quad (\because) \text{ consistency, positivity}
\end{aligned}$$

where, in the fourth equation, the following equivalence relationship holds by the weak union and decomposition graphoid axioms (⇒) and the contraction graphoid axiom (⇐) [3]:

$$\big(Y(a), M(a)\big) \perp\!\!\!\perp A\ (\forall a) \Leftrightarrow \big(Y(a) \perp\!\!\!\perp A|M(a)\big)\ (\forall a) \wedge (M(a) \perp\!\!\!\perp A)\ (\forall a). \quad \blacksquare$$

For reference, see Proposition 38.1 in Robins et al. [7]. The second equation above is a summation over *m* of their Equation (38.16), which is shown below:

$$P(M(n = x, o = x^*) = m, Y(n = x, o = x^*) = y) = P(Y(n = x^*, o = x^*) = y|M(n = x^*, o = x^*) = m)P(M(n = x, o = x) = m).$$

Proposition 6

If assumptions (13*) to (15*) in Lemma 1 hold, the separable direct effect is identified and given by

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

and the separable indirect effect is identified and given by

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

Proof:

From Lemma 2, we have

$$P(Y(n',o)=y) = \sum_m P(Y=y|M=m, A=a)P(M=m|A=a').$$

Thus, we have

$$\begin{aligned}
E[Y(n',o)] &= \sum_y y \cdot P(Y(n',o)=y) \\
&= \sum_y y \cdot \sum_m P(Y=y|M=m, A=a)P(M=m|A=a') \\
&= \sum_m \left\{ \sum_y y \cdot P(Y=y|M=m, A=a) \right\} P(M=m|A=a') \\
&= \sum_m E[Y|A=a, M=m]P(M=m|A=a').
\end{aligned}$$

If we apply this result and replace o with o', we get

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

Thus, the separable direct effect is given by

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

If we apply the result and replace n' with n, we get

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

Thus, the separable indirect effect is given by

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

Finally, it is worth noting that the separable direct and indirect effects cannot be identified under the following set of assumptions:

$$\begin{pmatrix} \big(Y(a),M(a)\big) \perp\!\!\!\perp A \ (\forall a) & (13^*) \\ M \perp\!\!\!\perp O|N & (14) \\ Y \perp\!\!\!\perp N|(O,M) & (15) \end{pmatrix},$$

which, although weaker than the set of assumptions (13) to (15), does not imply the set of assumptions (13*) to (15*). In longitudinal settings, Di Maria and Didelez [8] provided A0, A1, and A2 in their article as sufficient assumptions to identify the separable direct and indirect effects. Their assumptions can be written in the current setting as follows:

$$\begin{pmatrix} E[Y(a)] = E[Y|A=a]\ (\forall a) & (13^{**}) \\ M \perp\!\!\!\perp O|N & (14) \\ Y \perp\!\!\!\perp N|(O,M) & (15) \end{pmatrix},$$

which is even weaker than the set of assumptions (13*), (14), and (15). These discussions highlight the importance of carefully considering the assumptions for identifying separable direct and indirect effects.

Online Appendix G: Identification of separable direct and indirect effects when only *N* is a parent of *L*

Proposition 7

If assumptions (16) to (19) hold, the separable direct effect is identified and given by

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

and the separable indirect effect is identified and given by

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

Proof:

We have

$$\begin{aligned}
E[Y(n',o)] &= \sum_{m,l} E[Y(n',o)|M(n',o)=m, L(n',o)=l]P(M(n',o)=m|L(n',o)=l)P(L(n',o)=l) \quad (\because) \text{ law of total expectation} \\
&= \sum_{m,l} E[Y(n',o)|M(n',o)=m, L(n',o)=l, N=n', O=o]P(M(n',o)=m|L(n',o)=l, N=n', O=o)P(L(n',o)=l|N=n', O=o) \quad (\because) \text{ (16)} \\
&= \sum_{m,l} E[Y|M=m, L=l, N=n', O=o]P(M=m|L=l, N=n', O=o)P(L=l|N=n', O=o) \quad (\because) \text{ consistency} \\
&= \sum_{m,l} E[Y|M=m, L=l, N=n, O=o]P(M=m|L=l, N=n', O=o')P(L=l|N=n', O=o') \quad (\because) \text{ (17), (18), (19)} \\
&= \sum_{m,l} E[Y|M=m, L=l, A=a]P(M=m|L=l, A=a')P(L=l|A=a'), \quad (\because) \text{ determinism, positivity}
\end{aligned}$$

where, in the second equation, the following equivalence relationship holds by the weak union and decomposition graphoid axioms (⇒) and the contraction graphoid axiom (⇐) [3]:

$$\big(Y(n',o),L(n',o),M(n',o)\big) \perp\!\!\!\perp (N,O) \Leftrightarrow \Big(Y(n',o) \perp\!\!\!\perp (N,O)\big|\big(M(n',o),L(n',o)\big)\Big) \wedge \big(M(n',o) \perp\!\!\!\perp (N,O)|L(n',o)\big) \wedge \big(L(n',o) \perp\!\!\!\perp (N,O)\big).$$

If we apply this result and replace o with o', we get

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

Thus, the separable direct effect is given by

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

If we apply the result and replace n' with n, we get

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

Thus, the separable indirect effect is given by

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

Online Appendix H: Identification of separable direct and indirect effects when only *N* is a parent of *L* with weaker assumptions

Lemma 3

For some $x \in \{0,1\}$ and $x^* = 1 - x$, the following relationships hold:

$$\begin{pmatrix} (Y(a), M(a), L(a)) \perp\!\!\!\perp A \ (\forall a) & (16^*) \\ P(M(n=x,o=0)=m|L(n=x,o=0)=l) = P(M(n=x,o=1)=m|L(n=x,o=1)=l) & (17^*) \\ P(Y(n=1,o=x^*)=y|M(n=1,o=x^*)=m, L(n=1,o=x^*)=l) = P(Y(n=0,o=x^*)=y|M(n=0,o=x^*)=m, L(n=0,o=x^*)=l) & (18^*) \\ P(L(n=x,o=0)=l) = P(L(n=x,o=1)=l) & (19^*) \end{pmatrix}$$

$$\Leftarrow \begin{pmatrix} \big(Y(n,o), L(n,o), M(n,o)\big) \perp\!\!\!\perp (N,O) \ (\forall n,o) \\ P(M(n=x,o=0)=m|L(n=x,o=0)=l) = P(M(n=x,o=1)=m|L(n=x,o=1)=l) \\ P(Y(n=1,o=x^*)=y|M(n=1,o=x^*)=m, L(n=1,o=x^*)=l) = P(Y(n=0,o=x^*)=y|M(n=0,o=x^*)=m, L(n=0,o=x^*)=l) \\ P(L(n=x,o=0)=l) = P(L(n=x,o=1)=l) \end{pmatrix}$$

$$\Leftrightarrow \begin{pmatrix} \big(Y(n,o), L(n,o), M(n,o)\big) \perp\!\!\!\perp (N,O) \ (\forall n,o) \\ M \perp\!\!\!\perp O|(L, N=x) \\ Y \perp\!\!\!\perp N|(L, M, O=x^*) \\ L \perp\!\!\!\perp O|N=x \end{pmatrix}$$

$$\Leftarrow \begin{pmatrix} \big(Y(n,o), L(n,o), M(n,o)\big) \perp\!\!\!\perp (N,O) \ (\forall n,o) & (16) \\ M \perp\!\!\!\perp O|(L,N) & (17) \\ Y \perp\!\!\!\perp N|(L,M,O) & (18) \\ L \perp\!\!\!\perp O|N & (19) \end{pmatrix}.$$

Proof:

The relationship between the first and second sets of assumptions trivially holds from the deterministic relationship between *A*, *N*, and *O*. Additionally, the relationship between the third and fourth sets of assumptions also trivially holds. Here, we provide a proof for the equivalence relationship between the second and third sets of assumptions, using the following equivalence relationship, which holds by the weak union and decomposition graphoid axioms (⇒) and the contraction graphoid axiom (⇐) [3]:

$$\big(Y(n,o), L(n,o), M(n,o)\big) \perp\!\!\!\perp (N,O) \ (\forall n,o) \Leftrightarrow \Big(Y(n,o) \perp\!\!\!\perp (N,O)|\big(M(n,o), L(n,o)\big)\Big) \ (\forall n,o) \wedge \big(M(n,o) \perp\!\!\!\perp (N,O)|L(n,o)\big) \ (\forall n,o) \wedge \big(L(n,o) \perp\!\!\!\perp (N,O)\big) \ (\forall n,o).$$

$P(M(n=x,o=0)=m|L(n=x,o=0)=l)=P(M(n=x,o=1)=m|L(n=x,o=1)=l)$
$\Leftrightarrow P(M(n=x,o=0)=m|L(n=x,o=0)=l,N=x,O=0)=P(M(n=x,o=1)=m|L(n=x,o=1)=l,N=x,O=1)$ $(\because)$ $M(n,o) \perp\!\!\!\perp (N,O)|L(n,o)$
$\Leftrightarrow P(M=m|L=l,N=x,O=0)=P(M=m|L=l,N=x,O=1)$ $(\because)$ consistency
$\Leftrightarrow M \perp\!\!\!\perp O|(L,N=x)$.

$P(Y(n=1,o=x^*)=y|M(n=1,o=x^*)=m,L(n=1,o=x^*)=l)=P(Y(n=0,o=x^*)=y|M(n=0,o=x^*)=m,L(n=0,o=x^*)=l)$
$\Leftrightarrow P(Y(n=1,o=x^*)=y|M(n=1,o=x^*)=m,L(n=1,o=x^*)=l,N=1,O=x^*)$
$=P(Y(n=0,o=x^*)=y|M(n=0,o=x^*)=m,L(n=0,o=x^*)=l,N=0,O=x^*)$ $(\because)$ $Y(n,o) \perp\!\!\!\perp (N,O)|\big(M(n,o),L(n,o)\big)$
$\Leftrightarrow P(Y=y\ |L=l,M=m,N=1,O=x^*)=P(Y=y\ |L=l,M=m,N=0,O=x^*)$ $(\because)$ consistency
$\Leftrightarrow Y \perp\!\!\!\perp N|(L,M,O=x^*)$.

$P(L(n=x,o=0)=l)=P(L(n=x,o=1)=l)$
$\Leftrightarrow P(L(n=x,o=0)=l|N=x,O=0)=P(L(n=x,o=1)=l|N=x,O=1)$ $(\because)$ $L(n,o) \perp\!\!\!\perp (N,O)$
$\Leftrightarrow P(L=l|N=x,O=0)=P(L=l|N=x,O=1)$ $(\because)$ consistency
$\Leftrightarrow L \perp\!\!\!\perp O|N=x$. ■

Lemma 4

For some $x \in \{0,1\}$ and $x^* = 1 - x$, if assumptions (16*) to (19*) in Lemma 3 hold, the following equation holds:

$$P(Y(n = x, o = x^*) = y) = \sum_{m,l} P(Y = y|M = m, L = l, A = x^*)P(M = m|L = l, A = x)P(L = l|A = x).$$

Proof:

We have

$$
\begin{aligned}
& P(Y(n = x, o = x^*) = y) \\
&= \sum_{m,l} P(Y(n = x, o = x^*) = y|M(n = x, o = x^*) = m, L(n = x, o = x^*) = l)P(M(n = x, o = x^*) = m|L(n = x, o = x^*) = l)P(L(n = x, o = x^*) = l) \\
& (\because) \text{ law of total probability} \\
&= \sum_{m,l} P(Y(n = x^*, o = x^*) = y|M(n = x^*, o = x^*) = m, L(n = x^*, o = x^*) = l)P(M(n = x, o = x) = m|L(n = x, o = x) = l)P(L(n = x, o = x) = l) \\
& (\because)\ (17^*), (18^*), (19^*) \\
&= \sum_{m,l} P(Y(a = x^*) = y|M(a = x^*) = m, L(a = x^*) = l)P(M(a = x) = m|L(a = x) = l)P(L(a = x) = l) \quad (\because) \text{ determinism} \\
&= \sum_{m,l} P(Y(a = x^*) = y|M(a = x^*) = m, L(a = x^*) = l, A = x^*)P(M(a = x) = m|L(a = x) = l, A = x)P(L(a = x) = l|A = x) \quad (\because)\ (16^*) \\
&= \sum_{m,l} P(Y = y|M = m, L = l, A = x^*)P(M = m|L = l, A = x)P(L = l|A = x), \quad (\because) \text{ consistency, positivity}
\end{aligned}
$$

where, in the fourth equation, the following equivalence relationship holds by the weak union and decomposition graphoid axioms (⇒) and the contraction graphoid axiom (⇐) [3]:

$$(Y(a), M(a), L(a)) \perp\!\!\!\perp A\ (\forall a) \Leftrightarrow \big(Y(a) \perp\!\!\!\perp A|(M(a), L(a))\big)\ (\forall a) \wedge \big(M(a) \perp\!\!\!\perp A|L(a)\big)\ (\forall a) \wedge (L(a) \perp\!\!\!\perp A)\ (\forall a). \quad \blacksquare$$

For reference, see Equation (38.17) in Robins et al. [7].

Proposition 8

If assumptions (16*) to (19*) in Lemma 3 hold, the separable direct effect is identified and given by

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

and the separable indirect effect is identified and given by

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

Proof:

From Lemma 4, we have

$$P(Y(n',o)=y) = \sum_{m,l} P(Y=y|M=m,L=l,A=a)P(M=m|L=l,A=a')P(L=l|A=a').$$

Thus, we have

$$\begin{aligned}
E[Y(n',o)] &= \sum_{y} y\cdot P(Y(n',o)=y) \\
&= \sum_{y} y\cdot \sum_{m,l} P(Y=y|M=m,L=l,A=a)P(M=m|L=l,A=a')P(L=l|A=a') \\
&= \sum_{m,l}\left\{\sum_{y} y\cdot P(Y=y|M=m,L=l,A=a)\right\}P(M=m|L=l,A=a')P(L=l|A=a') \\
&= \sum_{m,l} E[Y|M=m,L=l,A=a]P(M=m|L=l,A=a')P(L=l|A=a').
\end{aligned}$$

If we apply this result and replace o with o', we get

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

Thus, the separable direct effect is given by

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

If we apply the result and replace n' with n, we get

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

Thus, the separable indirect effect is identified and given by

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

Online Appendix I: Identification of separable direct and indirect effects when only *O* is a parent of *L*

Proposition 9

If assumptions (16) to (18) and (20) hold, the separable direct effect is identified and given by

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

and the separable indirect effect is identified and given by

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

Proof:

We have

$$\begin{aligned}
E[Y(n',o)] &= \sum_{m,l} E[Y(n',o)|M(n',o)=m, L(n',o)=l]P(M(n',o)=m|L(n',o)=l)P(L(n',o)=l) \quad (\because) \text{ law of total expectation} \\
&= \sum_{m,l} E[Y(n',o)|M(n',o)=m, L(n',o)=l, N=n', O=o]P(M(n',o)=m|L(n',o)=l, N=n', O=o)P(L(n',o)=l|N=n', O=o) \quad (\because)\ (16) \\
&= \sum_{m,l} E[Y|M=m, L=l, N=n', O=o]P(M=m|L=l, N=n', O=o)P(L=l|N=n', O=o) \quad (\because) \text{ consistency} \\
&= \sum_{m,l} E[Y|M=m, L=l, N=n, O=o]P(M=m|L=l, N=n', O=o')P(L=l|N=n, O=o) \quad (\because)\ (17), (18), (20) \\
&= \sum_{m,l} E[Y|M=m, L=l, A=a]P(M=m|L=l, A=a')P(L=l|A=a), \quad (\because) \text{ determinism, positivity}
\end{aligned}$$

where, in the second equation, the following equivalence relationship holds by the weak union and decomposition graphoid axioms (⇒) and the contraction graphoid axiom (⇐) [3]:

$$\big(Y(n',o), L(n',o), M(n',o)\big) \perp\!\!\!\perp (N,O) \Leftrightarrow \Big(Y(n',o) \perp\!\!\!\perp (N,O) | \big(M(n',o), L(n',o)\big)\Big) \wedge \big(M(n',o) \perp\!\!\!\perp (N,O) | L(n',o)\big) \wedge \big(L(n',o) \perp\!\!\!\perp (N,O)\big).$$

If we apply this result and replace o with o', we get

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

Thus, the separable direct effect is given by

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

If we apply the result and replace n' with n, we get

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

Thus, the separable indirect effect is given by

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

Online Appendix J: Identification of separable direct and indirect effects when only *O* is a parent of *L* with weaker assumptions

Lemma 5

For some $x \in \{0,1\}$ and $x^* = 1 - x$, the following relationships hold:

$$\begin{pmatrix} (Y(a), M(a), L(a)) \perp\!\!\!\perp A\ (\forall a) & (16^*) \\ P(M(n=x,o=0)=m|L(n=x,o=0)=l) = P(M(n=x,o=1)=m|L(n=x,o=1)=l) & (17^*) \\ P(Y(n=1,o=x^*)=y|M(n=1,o=x^*)=m, L(n=1,o=x^*)=l) = P(Y(n=0,o=x^*)=y|M(n=0,o=x^*)=m, L(n=0,o=x^*)=l) & (18^*) \\ P(L(n=1,o=x^*)=l) = P(L(n=0,o=x^*)=l) & (20^*) \end{pmatrix}$$

$$\Leftarrow \begin{pmatrix} \big(Y(n,o), L(n,o), M(n,o)\big) \perp\!\!\!\perp (N,O)\ (\forall n,o) \\ P(M(n=x,o=0)=m|L(n=x,o=0)=l) = P(M(n=x,o=1)=m|L(n=x,o=1)=l) \\ P(Y(n=1,o=x^*)=y|M(n=1,o=x^*)=m, L(n=1,o=x^*)=l) = P(Y(n=0,o=x^*)=y|M(n=0,o=x^*)=m, L(n=0,o=x^*)=l) \\ P(L(n=1,o=x^*)=l) = P(L(n=0,o=x^*)=l) \end{pmatrix}$$

$$\Leftrightarrow \begin{pmatrix} \big(Y(n,o), L(n,o), M(n,o)\big) \perp\!\!\!\perp (N,O)\ (\forall n,o) \\ M \perp\!\!\!\perp O|(L, N=x) \\ Y \perp\!\!\!\perp N|(L,M,O=x^*) \\ L \perp\!\!\!\perp N|O=x^* \end{pmatrix}$$

$$\Leftarrow \begin{pmatrix} \big(Y(n,o), L(n,o), M(n,o)\big) \perp\!\!\!\perp (N,O)\ (\forall n,o) & (16) \\ M \perp\!\!\!\perp O|(L,N) & (17) \\ Y \perp\!\!\!\perp N|(L,M,O) & (18) \\ L \perp\!\!\!\perp N|O & (20) \end{pmatrix}.$$

Proof:

See the proof for Lemma 3 and the proof below.

$P(L(n=1,o=x^*)=l) = P(L(n=0,o=x^*)=l)$

$\Leftrightarrow P(L(n=1,o=x^*)=l|N=1,O=x^*) = P(L(n=0,o=x^*)=l|N=0,O=x^*)$ $(\because)$ $L(n,o) \perp\!\!\!\perp (N,O)$

$\Leftrightarrow P(L=l|N=1,O=x^*) = P(L=l|N=0,O=x^*)$ $(\because)$ consistency

$\Leftrightarrow L \perp\!\!\!\perp N|O=x^*$. ■

Lemma 6

For some $x \in \{0,1\}$ and $x^* = 1 - x$, if assumptions (16*) to (18*) and (20*) in Lemma 5 hold, the following equation holds:

$$P(Y(n = x, o = x^*) = y) = \sum_{m,l} P(Y = y|M = m, L = l, A = x^*)P(M = m|L = l, A = x)P(L = l|A = x^*).$$

Proof:

We have

$$\begin{aligned}
&P(Y(n = x, o = x^*) = y) \\
&= \sum_{m,l} P(Y(n = x, o = x^*) = y|M(n = x, o = x^*) = m, L(n = x, o = x^*) = l)P(M(n = x, o = x^*) = m|L(n = x, o = x^*) = l)P(L(n = x, o = x^*) = l) \\
&(\because) \text{ law of total probability} \\
&= \sum_{m,l} P(Y(n = x^*, o = x^*) = y|M(n = x^*, o = x^*) = m, L(n = x^*, o = x^*) = l)P(M(n = x, o = x) = m|L(n = x, o = x) = l)P(L(n = x^*, o = x^*) = l) \\
&(\because)\ (17^*), (18^*), (20^*) \\
&= \sum_{m,l} P(Y(a = x^*) = y|M(a = x^*) = m, L(a = x^*) = l)P(M(a = x) = m|L(a = x) = l)P(L(a = x^*) = l) \quad (\because) \text{ determinism} \\
&= \sum_{m,l} P(Y(a = x^*) = y|M(a = x^*) = m, L(a = x^*) = l, A = x^*)P(M(a = x) = m|L(a = x) = l, A = x)P(L(a = x^*) = l|A = x^*) \quad (\because)\ (16^*) \\
&= \sum_{m,l} P(Y = y|M = m, L = l, A = x^*)P(M = m|L = l, A = x)P(L = l|A = x^*), \quad (\because) \text{ consistency, positivity}
\end{aligned}$$

where, in the fourth equation, the following equivalence relationship holds by the weak union and decomposition graphoid axioms (⇒) and the contraction graphoid axiom (⇐) [3]:

$$(Y(a), M(a), L(a)) \perp\!\!\!\perp A\ (\forall a) \Leftrightarrow \big(Y(a) \perp\!\!\!\perp A|(M(a), L(a))\big)\ (\forall a) \wedge \big(M(a) \perp\!\!\!\perp A|L(a)\big)\ (\forall a) \wedge (L(a) \perp\!\!\!\perp A)\ (\forall a). \quad \blacksquare$$

For reference, see Equation (38.18) in Robins et al. [7].

Proposition 10

If assumptions (16*) to (18*) and (20*) in Lemma 5 hold, the separable direct effect is identified and given by

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

and the separable indirect effect is identified and given by

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

Proof:

From Lemma 6, we have

$$P(Y(n',o)=y) = \sum_{m,l} P(Y=y|M=m,L=l,A=a)P(M=m|L=l,A=a')P(L=l|A=a).$$

Thus, we have

$$\begin{aligned}
E[Y(n',o)] &= \sum_{y} y \cdot P(Y(n',o)=y) \\
&= \sum_{y} y \cdot \sum_{m,l} P(Y=y|M=m,L=l,A=a)P(M=m|L=l,A=a')P(L=l|A=a) \\
&= \sum_{m,l} \left\{ \sum_{y} y \cdot P(Y=y|M=m,L=l,A=a) \right\} P(M=m|L=l,A=a')P(L=l|A=a) \\
&= \sum_{m,l} E[Y|M=m,L=l,A=a]P(M=m|L=l,A=a')P(L=l|A=a).
\end{aligned}$$

If we apply this result and replace o with o', we get

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

Thus, the separable direct effect is given by

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

If we apply the result and replace n' with n, we get

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

Thus, the separable indirect effect is identified and given by

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

Online Appendix K: Identification of controlled direct effect

Proposition 11

If assumptions (1) and (2) hold, the controlled direct effect (CDE) that sets the mediator *M* at *m*, or CDE(*m*) is identified and given by

$$E[Y(a,m)] - E[Y(a',m)] = E[Y|A=a, M=m] - E[Y|A=a', M=m].$$

Proof:

We have

$$\begin{aligned} E[Y(a,m)] &= E[Y(a,m)|A=a] \quad (\because)\ Y(a,m) \perp\!\!\!\perp A \\ &= E[Y(a,m)|A=a, M(a)=m] \quad (\because)\ Y(a,m) \perp\!\!\!\perp M(a)|A=a \\ &= E[Y|A=a, M=m]. \quad (\because)\ \text{consistency, positivity} \end{aligned}$$

See Nguyen et al. [1] for an explanation of the positivity assumption. If we apply this result and replace a with a', we get

$$E[Y(a',m)] = E[Y|A=a', M=m].$$

Thus, the CDE(*m*) is given by

$$E[Y(a,m)] - E[Y(a',m)] = E[Y|A=a, M=m] - E[Y|A=a', M=m]. \quad \blacksquare$$

Proposition 12

If assumptions (1) and (6) hold, the CDE that sets the mediator *M* at *m*, or CDE(*m*) is identified and given by

$$E[Y(a,m)] - E[Y(a',m)] = \sum_h \big(E[Y|A=a, M=m, H=h]P(H=h|A=a) - E[Y|A=a', M=m, H=h]P(H=h|A=a')\big).$$

Proof:

We have

$$\begin{aligned} E[Y(a,m)] &= E[Y(a,m)|A=a] \quad (\because)\ Y(a,m) \perp\!\!\!\perp A \\ &= \sum_h E[Y(a,m)|A=a, H=h]P(H=h|A=a) \quad (\because)\ \text{law of total expectation, positivity} \\ &= \sum_h E[Y(a,m)|A=a, M(a)=m, H=h]P(H=h|A=a) \quad (\because)\ Y(a,m) \perp\!\!\!\perp M(a)|(A=a,H) \\ &= \sum_h E[Y|A=a, M=m, H=h]P(H=h|A=a). \quad (\because)\ \text{consistency, positivity} \end{aligned}$$

See Nguyen et al. [1] for an explanation of the positivity assumption. If we apply this result and replace a with a', we get

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

Thus, the CDE(*m*) is given by

$$E[Y(a,m)] - E[Y(a',m)] = \sum_h \big(E[Y|A=a, M=m, H=h]P(H=h|A=a) - E[Y|A=a', M=m, H=h]P(H=h|A=a')\big). \quad \blacksquare$$

Proposition 13

If assumptions (5) and (6) hold, the CDE that sets the mediator *M* at *m*, or CDE(*m*), is identified and given by

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

Proof:

We have

$$\begin{aligned}
E[Y(a,m)] &= \sum_h E[Y(a,m)|H=h]P(H=h) \quad (\because) \text{ law of total expectation} \\
&= \sum_h E[Y(a,m)|A=a, H=h]P(H=h) \quad (\because)\ Y(a,m) \perp\!\!\!\perp A|H \\
&= \sum_h E[Y(a,m)|A=a, M(a)=m, H=h]P(H=h) \quad (\because)\ Y(a,m) \perp\!\!\!\perp M(a)|(A=a,H) \\
&= \sum_h E[Y|A=a, M=m, H=h]P(H=h). \quad (\because) \text{ consistency, positivity}
\end{aligned}$$

See Nguyen et al. [1] for an explanation of the positivity assumption. If we apply this result and replace a with a', we get

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

Thus, the CDE(*m*) is given by

$$E[Y(a,m)] - E[Y(a',m)] = \sum_h (E[Y|A=a, M=m, H=h] - E[Y|A=a', M=m, H=h])P(H=h). \quad \blacksquare$$

Note that this becomes identical to the formula in Proposition 12 if $A \perp\!\!\!\perp H$ holds, as in Figure 2b.

Proposition 14

If assumptions (1) and (10) hold, the CDE that sets the mediator *M* at *m*, or CDE(*m*), is identified and given by

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

Proof:

We have

$$\begin{aligned}
E[Y(a,m)] &= E[Y(a,m)|A=a] \quad (\because)\, Y(a,m) \perp\!\!\!\perp A \\
&= \sum_{l} E[Y(a,m)|A=a, L(a)=l]P(L(a)=l|A=a) \quad (\because) \text{ law of total expectation} \\
&= \sum_{l} E[Y(a,m)|A=a, M(a)=m, L(a)=l]P(L(a)=l|A=a) \quad (\because)\, Y(a,m) \perp\!\!\!\perp M(a)|\big(A=a, L(a)\big) \\
&= \sum_{l} E[Y|A=a, M=m, L=l]P(L=l|A=a). \quad (\because) \text{ consistency, positivity}
\end{aligned}$$

See Nguyen et al. [1] for an explanation of the positivity assumption. If we apply this result and replace a with a', we get

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

Thus, the CDE(*m*) is given by

$$E[Y(a,m)] - E[Y(a',m)] = \sum_{l} \big(E[Y|A=a, M=m, L=l]P(L=l|A=a) - E[Y|A=a', M=m, L=l]P(L=l|A=a')\big). \quad \blacksquare$$

Online Appendix L: Identification of interventional direct and indirect effects

Proposition 15

If assumptions (1), (3), and (10) hold, the interventional direct effect (in Definition 6) is identified and given by

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

and the interventional indirect effect (in Definition 6) is identified and given by

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

Proof:

We have

$$\begin{aligned}
E\big[Y\big(a,G(a')\big)\big]&=\sum_{m}E\big[Y\big(a,G(a')\big)|G(a')=m\big]P(G(a')=m)\quad (\because)\text{ law of total expectation}\\
&=\sum_{m}E[Y(a,m)|G(a')=m]P(G(a')=m)\\
&=\sum_{m}E[Y(a,m)|G(a')=m]P(M(a')=m)\quad (\because)\ G(a')\sim M(a')\text{ identically distributed}\\
&=\sum_{m}E[Y(a,m)]P(M(a')=m)\quad (\because)\ Y(a,m)\perp\!\!\!\perp G(a')\ \text{ independent random samples}\\
&=\sum_{m}E[Y(a,m)|A=a]P(M(a')=m|A=a')\quad (\because)\ Y(a,m)\perp\!\!\!\perp A,\ \ M(a')\perp\!\!\!\perp A\\
&=\sum_{l}\sum_{m}E[Y(a,m)|A=a,L(a)=l]P(L(a)=l|A=a)P(M(a')=m|A=a')\quad (\because)\text{ law of total expectation}\\
&=\sum_{l}\sum_{m}E[Y(a,m)|A=a,M(a)=m,L(a)=l]P(L(a)=l|A=a)P(M(a')=m|A=a')\quad (\because)\ Y(a,m)\perp\!\!\!\perp M(a)|\big(A=a,L(a)\big)
\end{aligned}$$

$$= \sum_l \sum_m E[Y|A = a, M = m, L = l]P(L = l|A = a)P(M = m|A = a'). \quad (\because) \text{ consistency, positivity}$$

See Nguyen et al. [1] for an explanation of the positivity assumption. If we apply this result and replace a with a', we get

$$E\big[Y\big(a', G(a')\big)\big] = \sum_l \sum_m E[Y|A = a', M = m, L = l]P(L = l|A = a')P(M = m|A = a').$$

Thus, the interventional direct effect is given by

$$E\big[Y\big(a, G(a')\big)\big] - E\big[Y\big(a', G(a')\big)\big] = \sum_l \sum_m \big(E[Y|A = a, M = m, L = l]P(L = l|A = a) - E[Y|A = a', M = m, L = l]P(L = l|A = a')\big)P(M = m|A = a').$$

If we apply the result and replace a' with a, we get

$$E\big[Y\big(a, G(a)\big)\big] = \sum_l \sum_m E[Y|A = a, M = m, L = l]P(L = l|A = a)P(M = m|A = a).$$

Thus, the interventional indirect effect is given by

$$E\big[Y\big(a, G(a)\big)\big] - E\big[Y\big(a, G(a')\big)\big] = \sum_l \sum_m E[Y|A = a, M = m, L = l]P(L = l|A = a)\big(P(M = m|A = a) - P(M = m|A = a')\big). \quad \blacksquare$$

Proposition 16

If assumption (4) holds, the interventional direct effect (in Definition 6) becomes identical to the natural direct effect (in Definition 1), such that

$$E\big[Y\big(a, G(a')\big)\big] - E\big[Y\big(a', G(a')\big)\big] = E\big[Y\big(a, M(a')\big)\big] - E\big[Y\big(a', M(a')\big)\big],$$

and the interventional indirect effect (in Definition 6) becomes identical to the natural indirect effect (in Definition 1), such that

$$E\big[Y\big(a, G(a)\big)\big] - E\big[Y\big(a, G(a')\big)\big] = E\big[Y\big(a, M(a)\big)\big] - E\big[Y\big(a, M(a')\big)\big].$$

Proof:

We have

$$\begin{aligned}
E\big[Y\big(a, G(a')\big)\big] &= \sum_m E\big[Y\big(a, G(a')\big)|G(a') = m\big]P(G(a') = m) \quad (\because) \text{ law of total expectation} \\
&= \sum_m E[Y(a,m)|G(a') = m]P(G(a') = m) \\
&= \sum_m E[Y(a,m)|G(a') = m]P(M(a') = m) \quad (\because)\ G(a') \sim M(a') \text{ identically distributed} \\
&= \sum_m E[Y(a,m)]P(M(a') = m) \quad (\because)\ Y(a,m) \perp\!\!\!\perp G(a') \text{ independent random samples} \\
&= \sum_m E[Y(a,m)|M(a') = m]P(M(a') = m) \quad (\because)\ Y(a,m) \perp\!\!\!\perp M(a') \\
&= \sum_m E\big[Y\big(a, M(a')\big)|M(a') = m\big]P(M(a') = m) \\
&= E\big[Y\big(a, M(a')\big)\big].
\end{aligned}$$

If we apply this result and replace a with a', we get

$$E\big[Y\big(a', G(a')\big)\big] = E\big[Y\big(a', M(a')\big)\big].$$

Thus, the interventional direct effect (in Definition 6) becomes identical to the natural direct effect (in Definition 1), such that

$$E\big[Y\big(a, G(a')\big)\big] - E\big[Y\big(a', G(a')\big)\big] = E\big[Y\big(a, M(a')\big)\big] - E\big[Y\big(a', M(a')\big)\big].$$

If we apply the result and replace a' with a, we get

$$E\big[Y\big(a, G(a)\big)\big] = E\big[Y\big(a, M(a)\big)\big].$$

Thus, the interventional indirect effect (in Definition 6) becomes identical to the natural indirect effect (in Definition 1), such that

$$E\big[Y\big(a, G(a)\big)\big] - E\big[Y\big(a, G(a')\big)\big] = E\big[Y\big(a, M(a)\big)\big] - E\big[Y\big(a, M(a')\big)\big]. \quad \blacksquare$$

Therefore, even in the presence of an exposure-induced mediator–outcome confounder *L*, if—in addition to assumptions (1), (3), and (10)—assumption (4) holds, the interventional direct and indirect effects are identified and given by the formulae in Proposition 15, which become identical to the natural direct and indirect effects, respectively. Additionally, the overall effect becomes identical to the total effect. However, recall that assumption (4) does not *generally* hold in Figure 3b. 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). Indeed, scenarios in which assumptions (1) to (4) do not hold, yet assumptions (1), (3), (4), and (10) are satisfied, may be unrealistic.

A similar discussion applies when considering a mediator–outcome confounder that is not affected by the exposure. If assumptions (1), (3), and (6) hold, as in Figure 2b, the interventional direct and indirect effects are identified. If assumption (4) holds in addition to assumptions (1), (3), and (6), the interventional direct and indirect effects are identified, which become identical to the natural direct and indirect effects, respectively. Recall that, as shown in Proposition 4, assumptions (1), (3), (4), and (6) are sufficient conditions to identify the natural direct and indirect effects. However, assumption (4) does not *generally* hold in Figure 2b. As shown in Proposition 3, we may identify the natural direct and indirect effects under assumptions (5) to (8) instead, all of which hold in Figure 2b. These findings underscore a significant distinction between scenarios where the mediator–outcome confounder is affected by the exposure and those where it is not.

References

1 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.

2 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.

3 Pearl J. Causality: Models, Reasoning, and Inference. 2nd ed. New York, NY: Cambridge University Press 2009.

4 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.

5 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.

6 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.

7 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. pp. 713–64. https://doi.org/10.1145/3501714.3501754.

8 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.