

MS# EDE25-0319

Invited Commentary

***L* or *M*₁—Critical Challenges in Mediation Analysis**

Etsuji Suzuki

Department of Epidemiology, Graduate School of Medicine, Dentistry and Pharmaceutical
Sciences, Okayama University, Okayama, Japan

Corresponding author: Etsuji Suzuki, Department of Epidemiology, Graduate School of
Medicine, Dentistry and Pharmaceutical Sciences, Okayama University, 2-5-1 Shikata-cho, Kita-
ku, Okayama 700-8558, Japan. Tel: +81-86-235-7174, E-mail: etsuji-s@cc.okayama-u.ac.jp

Running head: Critical Challenges in Mediation Analysis

Conflicts of Interest: The author declares no conflicts of interest.

Source of Funding: ES is supported by the Japan Society for the Promotion of Science (KAKENHI
Grant Number JP23K09740).

Acknowledgments: The author is grateful to Eiji Yamamoto for helpful discussions. The author
also thanks Edanz (<https://jp.edanz.com/ac>) for editing a draft of this commentary.

Methods for causal mediation analysis have developed dramatically over the past few decades.¹⁻

⁷ In the causal mediation literature, several causal quantities—or estimands—have been proposed, including natural direct and indirect effects, interventional direct and indirect effects, and separable direct and indirect effects. As another possible causal estimand, Chen and Lin⁸ proposed separable path-specific effects, which is an extension of the separable effects framework to cases that involve multiple ordered mediators. In this commentary, I briefly discuss the newly proposed method from a broader perspective on causal mediation analysis. For readers less familiar with common causal mediation approaches, please see related literature.^{1-3,9-11}

L: A CRITICAL CHALLENGE IN MEDIATION ANALYSIS

First, let us consider the causal diagram in Figure panel A. Presumably, even without explicit explanations of the notation, many readers familiar with the causal mediation literature would intuitively—and appropriately—interpret A , M , and Y as representing an exposure, mediator, and outcome of interest, respectively. In this setting, researchers may be interested in the direct effect of A on Y that is not mediated by M , as well as the indirect effect of A on Y through M . Then, if certain assumptions about confounding, positivity, and consistency are met, they may

aim to identify natural direct and indirect effects in nonparametric structural equation models with independent errors.¹²⁻¹⁴ Note that, throughout this commentary, I omit the baseline confounders C following Chen and Lin.⁸ (I believe that, while this is technically correct, few epidemiologists would interpret A as a confounder of the M - Y relationship in panel A. Similarly, few would consider Y as a selection variable for the A - M relationship.)

Next, let us consider the causal diagram with L in Figure panel B. Again, many readers would intuitively—and appropriately—consider that L denotes an exposure-induced mediator–outcome confounder. When there is an exposure-induced mediator–outcome confounder, the so-called cross-world independence assumption does not generally hold, with or without conditioning on L .¹⁵ Indeed, figures like panel B often appear to illustrate the limitations of natural direct and indirect effects. As has been noted, these effects are generally not identifiable in this scenario, regardless of whether data on L are available, except under strong assumptions,¹⁶ such as the absence of an interaction between the exposure and the mediator at the individual level.¹⁷

Indeed, L poses a critical challenge in mediation analysis. To partially circumvent this challenge, VanderWeele et al.¹⁸ described three approaches to effect decomposition that yield quantities interpretable as direct and indirect effects, and identifiable from data even in the

presence of an exposure-induced mediator–outcome confounder: (i) joint mediators; (ii) path-specific effects; and (iii) interventional effects. Although these approaches do not yield natural direct and indirect effects, they provide causal quantities that can offer useful insights into mediation, provided our research questions align with the assumptions that underlie these methods. Alternatively, some may treat natural direct and indirect effects as targets of interest and apply sensitivity analysis methods to address issues of non-identifiability.¹⁹⁻²²

THE SAME CAUSAL STRUCTURE: IS IT L OR M_1 ?

Now, let us consider another causal diagram in Figure panel C. In this case, many readers would intuitively—and appropriately—interpret M_1 and M_2 as representing two causally ordered mediators of interest from exposure A to outcome Y . Although it may seem obvious, it is important to emphasize that panels B and C illustrate exactly the same causal structures, particularly when no explicit explanation of the notation is provided. What distinguishes these two panels is our (implicit) interpretation of the meaning, or simply, the conventional use of L in panel B and M_1 in panel C. Note that panel C is a scenario discussed by Chen and Lin⁸ to propose separable path-specific effects (see their Figure 1B).

This point highlights how the interpretation of a causal diagram—and ultimately the

conclusions drawn from it—can depend heavily on the specific research question being asked.^{23,24} Even when researchers are presented with the same underlying causal structure, different assumptions or interpretations about what each variable represents can lead to very different causal questions and analytic strategies. This underscores the importance of carefully defining the research question and the role of each variable when applying causal inference methods in practice.

A KEY PREMISE UNDERLYING SEPARABLE EFFECTS (AND THEIR EXTENSIONS)

As noted by Chen and Lin,⁸ the separable effects method in the context of mediation relies on a key premise: exposure can potentially be separated into two components—one component exerts no population-level effect on the outcome without going through the mediator and the other has no population-level effect on the mediator.^{25,26} Then, the total effect of the exposure on the outcome can be decomposed into separable direct and indirect effects.

For example, we first consider the scenario in Figure panel A in which there are no exposure-induced mediator–outcome confounders. In the separable effects approach, we assume that exposure A is decomposed into two separable components A_M and A_Y , as shown in panel D. The bold arrows from A to A_M and A_Y indicate deterministic relationships.²⁵ Importantly,

the two separable components are not observed; only the value of A is observed. Although $A \equiv A_M \equiv A_Y$ holds in the observed data, we must assume that a future trial could be designed in which interventions are applied separately to the two separable components. As noted by Chen and Lin,⁸ the key premise underlying separable effects requires a deep understanding of the nature of the exposure, which may limit the scope of causal mediation analysis. Indeed, the premise itself also poses a critical challenge.

Even when an exposure-induced mediator–outcome confounder is present, separable direct and indirect effects can still be identified from the data, provided certain assumptions hold.^{25,26} We consider the scenario in Figure panel B when an exposure-induced mediator–outcome confounder L is present. Even in this scenario, it has been shown that separable effects are identified when A_M is a parent of L , but A_Y is not, as described in panel E. Similarly, these effects are identified when A_Y is a parent of L , but A_M is not, as described in panel F. However, when both A_M and A_Y are parents of L , as described in panel G, separable effects are generally not identifiable.

Recall that Chen and Lin⁸ considered the scenario described in Figure panel C, in which M_1 is not treated as an exposure-induced mediator–outcome confounder; rather both M_1 and M_2 represent mediators of interest. Accordingly, Chen and Lin⁸ assumed that exposure A was

decomposed into three separable components, A_1 , A_2 , and A_Y , as described in panel H. Once again, the three separable components are not observed; only the value of A is observed.

Although $A \equiv A_1 \equiv A_2 \equiv A_Y$ holds in the observed data, we must assume that a future trial could be designed in which interventions are applied separately to the three separable components. This is a key premise underlying their approach.

A STEP FORWARD ON THE PATH?

To illustrate the causal interpretation of separable path-specific effects, Chen and Lin⁸ used a COVID-19 example, where exposure A represented the confirmation of COVID-19 infection status (1/0) and outcome Y indicated the risk of death (1/0). The causally ordered mediators of interest were isolation precaution (M_1) and depression status (M_2), both of which were binary variables. Then, Chen and Lin⁸ assumed that exposure A could be decomposed into the following three components: a positive viral test result (A_1), neurologic impact (A_2), and respiratory and other physiologic impacts (A_Y).

They provided interpretations of each of the three separable path-specific effects in their eTable 1.⁸ For example, note that the separable path-specific effect, SE_2 , is defined as the contrast between the average potential outcome of Y if (A_Y, A_1, A_2) were set to (1, 0, 0) and the

average potential outcome of Y if (A_Y, A_1, A_2) were set to $(1, 0, 1)$. Under certain assumptions, SE_2 captured the effect caused by $A \rightarrow M_2 \rightarrow Y$. In the context of the COVID-19 example, Chen and Lin⁸ described the interpretation of the estimated SE_2 as follows: “if an individual had not undergone a viral test ($A_1 [= 0]$) but had been affected by the respiratory and other physiologic impacts ($A_Y [= 1]$) of the virus, the neurologic damages (A_2) caused by COVID-19 would, on average, increase the odds of mortality (Y) by 14.8% through depression (M_2).” Recall, however, that the three components, A_1 , A_2 , and A_Y , could be, at least conceptually, intervened upon separately in a future trial. Furthermore, the distinction between A and A_1 may be blurred. Thus, if we take the premise seriously, we must ask ourselves whether such separate interventions can realistically be envisaged in practice. If the answer is yes, the interpretation above should be revised to more clearly describe the distinct interventions on each of the separable components. This may pose a critical challenge for researchers using separable path-specific effects, even from the perspective of a target trial.²⁷

Chen and Lin⁸ should be commended for their efforts in proposing separable path-specific effects. However, their practical applicability is yet to be demonstrated in empirical research. Future studies should also carefully consider the key premise underlying the method to provide a clear explanation and envisage the relevant hypothetical intervention(s). Whether

this represents a meaningful step forward is ultimately a question for the broader research community to consider. As with any new methodologic direction, the path goes ever on.

ABOUT THE AUTHOR

ETSUJI SUZUKI is Research Associate Professor of Epidemiology at Okayama University. He has aimed to improve epidemiologic and statistical theories with the goal of advancing our understanding of causation in population health. In 2023, he was honored with awards from both the Japan Epidemiological Association and the Japan Medical Association in recognition of his contributions to causal inference research.

Figure legend

FIGURE. Causal diagrams illustrating mediation. Let A and Y denote an exposure and outcome of interest, respectively. Let M denote a mediator of interest in panels A, B, D, E, F, and G, whereas let M_1 and M_2 denote causally ordered mediators of interest in panels C and H. Panel C shows the same causal structure as Figure 1B of Chen and Lin.⁸ In panels B, E, F, and G, let L denote an exposure-induced mediator–outcome confounder. In panels D, E, F, and G, we assume that exposure A is decomposed into two separable components A_M and A_Y . The bold arrows from A

to A_M and A_Y indicate deterministic relationships. In panel H, we assume that exposure A is decomposed into three separable components A_1 , A_2 , and A_Y . The bold arrows from A to A_1 , A_2 , and A_Y indicate deterministic relationships. See Figure 2 of Chen and Lin.⁸

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.
3. VanderWeele TJ. Mediation analysis. In: Lash TL, VanderWeele TJ, Haneuse S, Rothman KJ, eds. *Modern Epidemiology*. 4th ed. Philadelphia, PA: Wolters Kluwer; 2021:655-675.
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;26:255-271.
5. Nguyen TQ, Schmid I, Ogburn EL, Stuart EA. Clarifying causal mediation analysis: Effect identification via three assumptions and five potential outcomes. *J Causal Inference*. 2022;10:246-279. doi: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. *Statist Surv.* 2023;17:1-41.
7. Imai K, Keele L, Tingley D. A general approach to causal mediation analysis. *Psychol Methods.* 2010;15:309-334.
8. Chen YL, Lin SH. On definition and interpretation of separable path-specific effects with multiple ordered mediators. *Epidemiology.* 2025: (in press).
9. VanderWeele TJ, Vansteelandt S. Conceptual issues concerning mediation, interventions and composition. *Stat Interface.* 2009;2:457-468.
10. Shrier I, Suzuki E. The primary importance of the research question: Implications for understanding natural versus controlled direct effects. *Int J Epidemiol.* 2022;51:1041-1046.
11. Schuler MS, Coffman DL, Stuart EA, Nguyen TQ, Vegetabile B, McCaffrey DF. Practical challenges in mediation analysis: a guide for applied researchers. *Health Serv Outcomes Res Methodol.* 2024.
12. Robins JM, Greenland S. Identifiability and exchangeability for direct and indirect effects. *Epidemiology.* 1992;3:143-155.

13. Pearl J. Direct and indirect effects. Proceedings of the Seventeenth Conference on Uncertainty in Artificial Intelligence. San Francisco, CA: Morgan Kaufmann, 2001;411-420.
14. Pearl J. *Causality: Models, Reasoning, and Inference*. 2nd ed. New York, NY: Cambridge University Press; 2009.
15. Andrews RM, Didelez V. Insights into the cross-world independence assumption of causal mediation analysis. *Epidemiology*. 2021;32:209-219.
16. Tchetgen Tchetgen EJ, VanderWeele TJ. Identification of natural direct effects when a confounder of the mediator is directly affected by exposure. *Epidemiology*. 2014;25:282-291.
17. Robins JM. Semantics of causal DAG models and the identification of direct and indirect effects. In: Green PJ, Hjort NL, Richardson S, eds. *Highly Structured Stochastic Systems*. Oxford Statistical Science Series. New York, NY: Oxford University Press; 2003:70-81.
18. VanderWeele TJ, Vansteelandt S, Robins JM. Effect decomposition in the presence of an exposure-induced mediator–outcome confounder. *Epidemiology*. 2014;25:300-306.
19. Imai K, Yamamoto T. Identification and sensitivity analysis for multiple causal mechanisms: Revisiting evidence from framing experiments. *Political Analysis*. 2013;21:141-171.

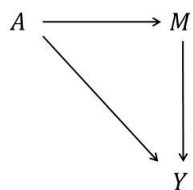
20. Tchetgen Tchetgen EJ, Shpitser I. Semiparametric theory for causal mediation analysis: efficiency bounds, multiple robustness and sensitivity analysis. *Annals of Statistics*. 2012;40:1816-1845.
21. VanderWeele TJ, Chiba Y. Sensitivity analysis for direct and indirect effects in the presence of exposure-induced mediator–outcome confounders. *Epidemiol Biostat Public Health*. 2014;11:e9027. doi:10.2427/9027.
22. Vansteelandt S, VanderWeele TJ. Natural direct and indirect effects on the exposed: effect decomposition under weaker assumptions. *Biometrics*. 2012;68:1019-1027.
23. Greenland S, Pearl J, Robins JM. Causal diagrams for epidemiologic research. *Epidemiology*. 1999;10:37-48.
24. Suzuki E, Shinozaki T, Yamamoto E. Causal diagrams: pitfalls and tips. *J Epidemiol*. 2020;30:153-162.
25. Robins JM, Richardson TS. Alternative graphical causal models and the identification of direct effects. In: Shrout PE, Keyes KM, Ornstein K, eds. *Causality and Psychopathology: Finding the Determinants of Disorders and Their Cures*. New York, NY: Oxford University Press; 2011:103-158.
26. Robins JM, Richardson TS, Shpitser I. An interventionist approach to mediation analysis.

In: Geffner H, Dechter R, Halpern JY, eds. *Probabilistic and Causal Inference: The Works of Judea Pearl*. Vol. 36 Association for Computing Machinery; 2022:713–764.

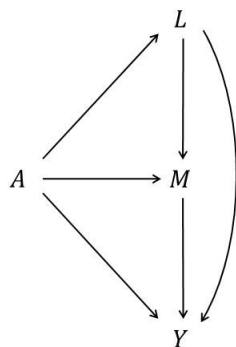
27. Hernán MA, Robins JM. Using big data to emulate a target trial when a randomized trial is not available. *Am J Epidemiol*. 2016;183:758-764.

Figure

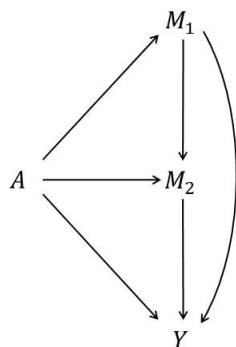
A



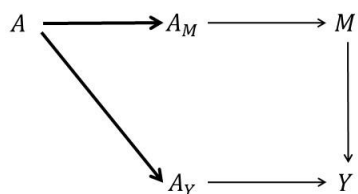
B



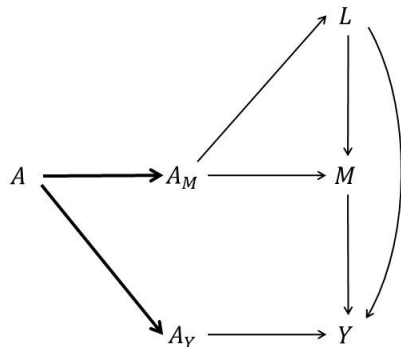
C



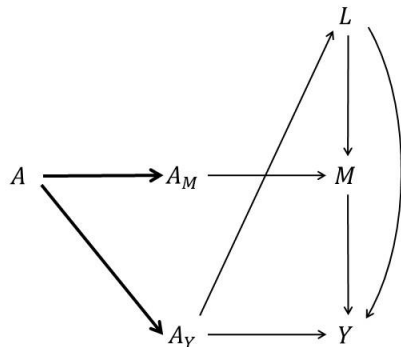
D



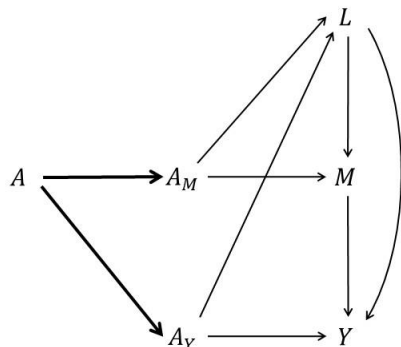
E



F



G



H

