frontiers in Psychology

GENERAL COMMENTARY
published: 09 February 2017
doi: 10.3389/fpsyg.2017.00151

Commentary: Causal Effects in Mediation Modeling: An Introduction with Applications to Latent Variables

Emil N. Coman, Felix Thoemmes and Judith Fifield*

Health Disparities Institute, UConn Health, Farmington, CT, USA

Keywords: mediation, causal mediation, potential outcomes, causal inference, counterfactuals

A commentary on

Causal Effects in Mediation Modeling: An Introduction with Applications to Latent Variables
by Muthén, B., and Asparouhov, T. (2015). Struct. Equation Model. 22, 12–23. doi: 10.1080/10705511.2014.935843

Causal mediation[1] is an increasingly popular analysis, as recently described by Muthén and Asparouhov (2015, M&A)[2]. We suggest a simplified notation for causal mediation effects, $\mathbf{i}_{\mathrm{T}}/i_{P=BK}$ and $\mathbf{d}_{\mathrm{T}}/d_P$, provide a graphical view of potential outcomes (PO) and expand the M&A approach by using VanderWeele's (2014) mediation decomposition.

An intuitive way to label and see causal in/direct effects is to directly display POs, as in **Figure 1** below. POs are values that could be observed, but have not been realized (yet). They reveal themselves partially once nature or researchers assign people to specific experimental conditions, or when people make choices. POs are useful in defining causal total effects (TE), as differences between the same individual's (i) two POs, $\mathrm{Y}_{i1} - \mathrm{Y}_{i0}$, had the person been treated (subscript 1), and alternatively (but *simultaneously*) not treated (0); evidently, in our reality one of these has to be "contrary-to-fact" (CF).

The indirect effect of X on Y through a mediator M is the part of the total effect that "flows through" M, or the contribution of the path X->M->Y to the observed association between X and Y, which is an open path because causal association flows through it (Elwert, 2013). The key problem in intuitively grasping causal in/direct effects is the "nesting" of the POs due to the double role of the mediator as a cause *and* an effect[3]: the PO "Y if X was set to x," or Y_{x}, can be combined with "Y if M was set to m," or Y^{m} (we suggest using a superscript for scenarios involving M). So $^*\mathrm{Y}_1^{\mathrm{M0}}$, for example, labeled Y(1, M(0)) in M&A, is the PO of the outcome Y if a person was treated ($_1$), but his/her mediator took on the value had s/he would belonged to the opposite (control) condition ($^{\mathrm{M0}}$). This PO is clearly contrary-to-fact (CF), never observable, a "cross-worlds" quantity (Lok, 2016), hence our * sign. Y_0^0 and Y_1^1 are in principle realizable, only one of them at a time for the same person, however.

[1]The label "causal" mediation reflects more than the expansion of the original Baron and Kenny model to allow for X-by-M interaction, and does not suddenly make any three-variable model causal in the profound sense. Causal mediation relies on meeting other assumptions, like the no-confounder assumption of M and Y, and would require causal investigations like those afforded by Direct Acyclic Graphs (DAGs, Greenland et al., 1999).

[2]Other dominant causal mediation "schools" are led by the Imai (Imai et al., 2010) and Pearl (Pearl, 2001) teams, first centered on R and Stata implementations, the latter more theoretical and non-parametrical. They differ also in terms of formulating the assumptions for identification of the causal in/direct effects.

[3]Pearl (2013) calls them nested counterfactuals; the key insight Sewall Wright foresaw when proposing the path analytic method may have been that the change in Y in relation to the change in X (the slope δY/δX), traced on the path through an intermediary M, is linked to the slopes δM/δX and δY/δM following the composite function chain rule of derivatives: $\delta y/\delta x = \delta y/\delta m \cdot \delta m/\delta x$, which mirrors the Baron and Kenny $i = a \cdot b$. Adding the contributions of all such X-to-Y open paths yields the model predicted association between X and Y (see the "tracing rule," Loehlin, 2004).

OPEN ACCESS

Edited by:
Pietro Cipresso,
Istituto Auxologico Italiano (IRCCS),
Italy

Reviewed by:
Ali Ünlü,
Technische Universität München,
Germany

***Correspondence:**
Emil N. Coman
comanus@gmail.com

Specialty section:
This article was submitted to
Quantitative Psychology and
Measurement,
a section of the journal
Frontiers in Psychology

Received: 08 October 2016
Accepted: 23 January 2017
Published: 09 February 2017

Citation:
Coman EN, Thoemmes F and Fifield J
(2017) Commentary: Causal Effects in
Mediation Modeling: An Introduction
with Applications to Latent Variables.
Front. Psychol. 8:151.
doi: 10.3389/fpsyg.2017.00151

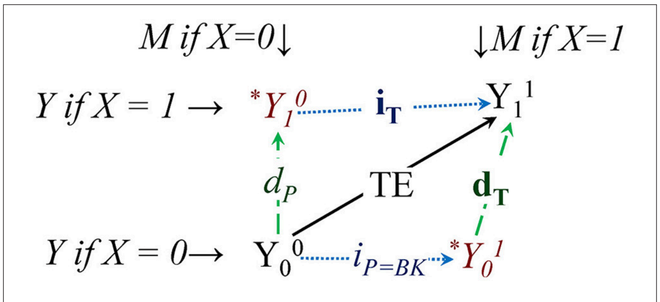

FIGURE 1 | Causal mediation represented with potential outcomes (POs). $^*Y_0^1$ is the PO of Y is X was set to 0 (subscript) but M took the value it would attain if X was instead set to 1 (superscript); * means that PO is unobservable/contrary-to-fact (CF). The upper level is the potential world if treated (X = 1), the bottom if not treated (X = 0); $i_{P=BK}$/$\mathbf{i}_\mathrm{T}$ and d_P/$\mathbf{d}_\mathrm{T}$ are pure/total indirect and direct effects (BK comes from the "classic" Baron-Kenny). The total effects ($\mathbf{i}_\mathrm{T}$ and $\mathbf{d}_\mathrm{T}$, in bold) are shown as longer than their pure counterparts, $i_{P=BK}$ and d_P (in *italics*), both by exactly INTMed, the mediated interaction. The arrows from lower to upper PO worlds are the two causal direct effects d_P and $\mathbf{d}_\mathrm{T}$, those from the left-side POs to the right-side POs capture the indirect effects $i_{P=BK}$ and $\mathbf{i}_\mathrm{T}$, while the diagonal up and to right the total effect TE. TE can be decomposed then as $d_P + \mathbf{i}_\mathrm{T}$, or $i_{P=BK} + \mathbf{d}_\mathrm{T}$.

The four key POs involved in understanding causal in/direct effects are shown in **Figure 1**. The total effect is decomposable into direct and indirect causal effects, possibly in two ways, through one of two fully contrary-to-fact POs: $^*\mathrm{Y}_1^0$ or $^*\mathrm{Y}_0^1$.

Both decompositions of TE can be obtained by adding and subtracting a fully CF intermediary term; e.g., through $^*\mathrm{Y}_1^0$:

$$\mathrm{TE} = \mathrm{Y}_1^1 - \mathrm{Y}_0^0 = (\mathrm{Y}_1^1 - {}^*\mathrm{Y}_1^0) + ({}^*\mathrm{Y}_1^0 - \mathrm{Y}_0^0) = \mathbf{i}_\mathbf{T} + \mathrm{d}_\mathrm{P} \qquad (1)$$

Intuitively, one can see that the two vertical arrows are direct effects, because they capture the "change" in Y (in the PO world), marked by subscript/superscript changes: when "changing" only X, i.e., while (un-naturally) holding the mediator at a "constant" PO-value. The causal pure direct effect d_P is often referred to as natural (or pure natural direct effect, PNDE, in M&A), because the mediator takes on the same value under the control condition, which would be the "natural" course of action without any change in nature.

Similarly, the two horizontal arrows are indirect effects, because they are the result of "changing" only M, while keeping X constant (at 0, or 1)[4]. The "upper" indirect effect is called also natural, but is in fact a *total* indirect effect (total natural indirect effect, TNIE, in M&A); it is *total* because it is a sum, of its *pure* kind, which we label $i_{P=BK}$, and an interacted mediation component, see Equation (3) below; here X is kept "unchanged" at the treated level (1), yet the mediator "changes" its (potential) value, from its natural (control) value to the value "if treated."

We suggest to label the *pure* indirect effect $i_{P=BK}$, because its estimate for continuous M and Y matches the classic no interaction and no confounder Baron and Kenny (1986) indirect effect "a · b" (see Equation 8 in M&A, when an interaction X-by-M is specified).

The relation between the key causal effects $\mathbf{d}_\mathrm{T}$ and d_P and $\mathbf{i}_\mathrm{T}$ and $i_{P=BK}$ has been revealed by VanderWeele's decomposition (VanderWeele, 2014), hence the *total* labels we proposed:

$$\mathbf{d}_\mathrm{T} = d_P + \mathrm{INT}_\mathrm{Med} \quad \text{and} \quad \mathbf{i}_\mathrm{T} = i_{P=BK} + \mathrm{INT}_\mathrm{Med}, \qquad (2)$$

where $\mathrm{INT}_\mathrm{Med}$ is the mediated interaction component[5], which is the product of the interaction estimate and the X->M linear effect, $\beta_{X*M} \cdot a$, labeled $\gamma_1 \cdot \beta_3$ in M&A, see their Equations (5) and (9); INTMed is non-zero when X impacts M, and X and M interact in how they impact Y.

Because the Mplus software code in M&A for computing causal in/direct effects did not estimate the effects proposed by VanderWeele's "decomposition" (mediated interaction, controlled direct effect, proportion attributable to interaction, and portion eliminated), we expand the Mplus code for continuous M and Y to estimate them (see the online appendix at https://bit.ly/pos_frontiers); we present an expanded VanderWeele SAS code too, which estimates the Mplus additional effects: pure direct, total indirect and total direct.

To illustrate, we estimated effects from a weight-loss randomized intervention data (SisterTalk Hartford, Burleson et al., 2008; de-identified data for replication available in appendix), which was meant to improve food habits and consequently reduce BMI in African-American women; effects are shown in Equation (3) (following VanderWeele's Figure 4, 2013, which is an expanded online version of the published (VanderWeele, 2014); * signals statistically significant at $p < 0.05$, NS signifies non-significant):

$$\underset{-0.663^*}{TE} = \underbrace{\overbrace{\underset{-0.507^*(76\%)}{{}^{m}CDE} + \underset{0.012^{NS}(-2\%)}{{}^{m}INT_{\mathrm{Ref}}}}^{d_P = -0.495^*(75\%)} + \underset{0.021^{NS}(-3\%)}{INT_{\mathrm{Med}}}}_{\mathbf{d}_\mathrm{T} = -0.474^*(71\%)} + \underbrace{\underset{-0.189^*(29\%)}{BK}}_{i_{P=BK} = -0.189^*(29\%)} \qquad \mathbf{i}_\mathrm{T} = -0.168^*(25\%) \qquad (3)$$

where INT_{Med} is the mediated interaction, BK is the "Baron and Kenny" causal indirect effect, ^{m}CDE is the controlled direct effect, $^{m}\mathrm{INT}_\mathrm{Ref}$ reference interaction, with superscript *m* signaling that those effects depend on what value *m* the analyst decided to estimate them at.

[4]The fact that there is possibly more than one indirect (and hence direct) effect to estimate follows from the interaction of X and M in causing Y, which makes the effect of M on Y vary with X (or the effect of X on Y vary with M).

[5]The INTMed key component is defined in VanderWeele (2013) as $(\mathrm{Y}_1^1 - {}^*\mathrm{Y}_1^0 - {}^*\mathrm{Y}_0^1 + \mathrm{Y}_0^0)$ $(\mathrm{M}_1 - \mathrm{M}_0)$ which for continuous M and Y becomes either $\mathbf{i}_\mathrm{T}$-$i_{P=BK}$ or $\mathbf{d}_\mathrm{T}$-d_P.

The total effect TE was −0.66 BMI units (approx. −3.9 lbs. for an average 64 inch woman). The mediated interaction effect INT_{Med} is about 3% of the TE, and statistically non-significant, hence statistically $\mathbf{i}_{\mathrm{T}} \overset{stat.}{=} i_{P=BK}$ and $\mathbf{d}_{\mathrm{T}} \overset{stat.}{=} d_P$ ("stat." signals statistical, not mathematical, equality), so one can report the classic i_{BK}[6]: the weight loss achieved through improving one's food habits is about 25% of the total effect, while the residual direct effect is about 75% of it.

While POs are central to "causal" mediation, visually "seeing" them is challenging, yet, when achieved, it helps uncover the mechanics behind causal direct and indirect effect estimation. Intuitive graphical displays could aid in visualizing some assumptions, many of which refer to relations between POs, and not their observed cousins (e.g., ignorability, or unconfoundedness, Imai et al., 2010); such assumptions ensure identifiability of in/direct causal effects.

We hope that the simplified notation and a visual display of how causal in/direct effects emerge from a mix of the POs of the mediator and the final outcome can contribute to a more intuitive understanding and reporting of causal mediation, as presented in the seminal paper we commented on. The notational bridge and cross-pollination of software syntaxes we suggested should facilitate such an improved understanding.

[6]While we label the pure indirect effect $i_{P=BK}$, as being the Baron and Kenny classic indirect effect, a · b, its estimate in the "causal" specification, with the X-by-M interaction term included, will not coincide of course with the estimate from the simpler model without interaction; in our case the classic BK estimate was $\mathrm{i}_{\mathrm{BK}} = -0.179$ ($SE = 0.054$), $p < 0.001$, while $i_{P=BK}$ was -0.189 ($SE = 0.069$), $p = 0.006$.

AUTHOR CONTRIBUTIONS

ENC has developed the idea, FT has verified the claims, expanded, and revised the manuscript extensively, JF has worked on the theoretical and design portion of the original study and has revised and edited the manuscript.

FUNDING

The Sistertalk Hartford project was funded by the Patrick and Catherine Weldon Donaghue Medical Research Foundation.

ACKNOWLEDGMENTS

We are grateful to Judea Pearl for his useful suggestions that were instrumental in clarifying and simplifying our language, and to Ali Ünlü, who provided us with comprehensive yet friendly constructive criticism and suggestions.

REFERENCES

Baron, R. M., and Kenny, D. A. (1986). The moderator–mediator variable distinction in social psychological research: conceptual, strategic, and statistical considerations. *J. Pers. Soc. Psychol.* 51, 1173. doi: 10.1037/0022-3514.51.6.1173

Burleson, J., Curry, L., Dauser-Forrest, D., Henderson, C., McKinney, M., Peele, M. et al. (2008). "Testing the Social Action Theory framework of a faith-based weight control program collaboratively translated with African-American and Black Women: SisterTalk Hartford," in *Paper Presented at the Annual Meeting of North American Primary Care Research Group* (Rio Grande).

Elwert, F. (2013). "Graphical causal models," in *Handbook of Causal Analysis for Social Research,* ed S. L. Morgan (New York, NY: Springer), 245–273.

Greenland, S., Robins, J. M., and Pearl, J. (1999). Confounding and collapsibility in causal inference. *Stat. Sci.* 14, 29–46.

Imai, K., Keele, L., and Tingley, D. (2010). A general approach to causal mediation analysis. *Psychol. Methods* 15, 309–334. doi: 10.1037/a0020761

Loehlin, J. C. (2004). *Latent Variable Models: An introduction to Factor, Path, and Structural Equation Analysis*. Mahwah, NJ: Lawrence Erlbaum.

Lok, J. J. (2016). Defining and estimating causal direct and indirect effects when setting the mediator to specific values is not feasible. *Stat. Med.* 35, 4008–4020. doi: 10.1002/sim.6990

Muthén, B., and Asparouhov, T. (2015). Causal effects in mediation modeling: an introduction with applications to latent variables. *Struct. Equation Model.* 22, 12–23. doi: 10.1080/10705511.2014.935843

Pearl, J. (2001). "Direct and indirect effects" in *Proceedings of the Seventeenth Conference on Uncertainty in Artificial Intelligence.* Available online at: ftp://ftp.cs.ucla.edu/pub/stat_ser/R273-U.pdf

Pearl, J. (2013). Structural counterfactuals: a brief introduction. *Cogn. Sci.* 37, 977–985. doi: 10.1111/cogs.12065

VanderWeele, T. J. (2013). A Unification of Mediation and Interaction. Harvard University Biostatistics. Working Paper Series. Available online at: http://biostats.bepress.com/harvardbiostat/paper164

VanderWeele, T. J. (2014). A unification of mediation and interaction: a 4-way decomposition. *Epidemiology* 25, 749–761. doi: 10.1097/EDE.0000000000000121

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.