American Journal of Epidemiology

Vol. 182, No. 2
DOI: 10.1093/aje/kwv059
Advance Access publication:
May 5, 2015

Practice of Epidemiology

When Is the Difference Method Conservative for Assessing Mediation?

Zhichao Jiang* and Tyler J. VanderWeele

* Correspondence to Zhichao Jiang, School of Mathematical Sciences, Peking University, No. 5 Yiheyuan Road, Beijing 100871, China (e-mail: zhichaojiang@pku.edu.cn).

Initially submitted March 26, 2014; accepted for publication September 24, 2014.

Assessment of indirect effects is useful for epidemiologists interested in understanding the mechanisms of exposure-outcome relationships. A traditional way of estimating indirect effects is to use the "difference method," which is based on regression analysis in which one adds a possible mediator to the regression model and examines whether the coefficient for the exposure changes. The difference method has been criticized for lacking a causal interpretation when it is used with logistic regression. In this article, we use the counterfactual framework to define the natural indirect effect (NIE) and assess the relationship between the NIE and the difference method. We show that under appropriate assumptions, the difference method consistently estimates the NIE for continuous outcomes and is always conservative for binary outcomes. Thus, the difference method can be used to provide evidence for the presence of mediation but not for the absence of mediation.

difference method; epidemiologic methods; mediation analysis; natural indirect effect

Abbreviations: NDE, natural direct effect; NIE, natural indirect effect.

Editor's note: *An invited commentary on this article appears on page 109, and the authors' response appears on page 115.*

Mediation analysis is a useful tool in epidemiologic studies. Investigators are sometimes not satisfied with knowing only the total effect of exposure on an outcome and want to gain insight into the mechanisms that explain this effect. For assessment of mediation, a common approach in the epidemiologic literature is using regression analysis to calculate the indirect effect as a comparison between the total effect of exposure on the outcome and the effect of exposure adjusted for an intermediate variable, which is sometimes referred to as the "difference method" (1–3).

However, the difference method has been criticized for lacking a causal interpretation (4). VanderWeele and Vansteelandt (5) have argued that for a binary outcome that is rare, use of the difference method estimator for logistic regression in the absence of exposure-mediator interaction is approximately consistent with calculation of an indirect effect. However, when the outcome is not rare or when there is exposure-mediator interaction, using the difference method may lead to biased estimators of indirect effects and incorrect conclusions concerning mediation. In this paper, we consider the relationships between the difference method and the counterfactual-based definitions of the indirect effect, sometimes called the natural indirect effect (NIE).

THE DIFFERENCE METHOD FOR MEDIATION ANALYSIS

In the epidemiologic literature, investigators often use a "difference method" to estimate a mediated or indirect effect. The difference method uses the contrast between the effects of exposure on the outcome with and without adjustment for 1 or more variables potentially lying on the pathway from exposure to outcome. Thus, the difference method is thought to capture the effect that operates through the specified intermediate variables. We will let A be an exposure of interest, M be a mediator, Y be an outcome of interest, and X be a set of baseline covariates not affected by the treatment.

For a continuous outcome, we consider a linear regression model for the outcome that is constructed both with and

without adjustment for the mediator M:

$$E(Y|a,x) = \phi_0 + \phi_1 a + \phi_2^\top x, \quad (1)$$

$$E(Y|a,m,x) = \theta_0 + \theta_1 a + \theta_2 m + \theta_3^\top x. \quad (2)$$

Then the indirect effect estimated from the difference method, $\phi_1 - \theta_1$, is a comparison between the effects of exposure A on the outcome Y with and without adjustment for the mediator M. The traditional "proportion explained" method (6–9) is closely related to the difference method and uses $(\phi_1 - \theta_1)/\phi_1$ as the measure of interest, which likewise relies on the difference between ϕ_1 and θ_1. The "proportion explained" method is less stable than the difference method, $\phi_1 - \theta_1$, when the effect of exposure on the outcome is small (10). A related approach consists of a linear model for M along with model 2 (equation 2 above), and this is sometimes referred to as the "Baron-Kenny" approach (2) or the "product-of-coefficients approach" or more simply the "product method." One of the advantages of the difference method is that it places no restrictions on the distribution of the mediator M. The mediator can be binary, discrete, or continuous, and we do not need to specify a model for M. In contrast, the Baron-Kenny method traditionally relies on a linear model for the mediator, which is sometimes too restrictive.

Similar approaches are sometimes used in logistic regression with a binary outcome. Consider 2 logistic regression models for the binary outcome Y, both with and without adjustment for the mediator M:

$$\text{logit}\{P(Y|a,x)\} = \phi_0 + \phi_1 a + \phi_2^\top x, \quad (3)$$

$$\text{logit}\{P(Y|a,m,x)\} = \theta_0 + \theta_1 a + \theta_2 m + \theta_3^\top x. \quad (4)$$

Then the indirect effect estimated from the difference method on a log odds ratio scale is $\phi_1 - \theta_1$.

Next we will consider the causal interpretation of the difference method.

RELATIONSHIP BETWEEN THE DIFFERENCE METHOD AND THE NIE

Robins and Greenland (11) and Pearl (12) propose definitions of natural direct and indirect effects based on the counterfactual framework (13, 14). The total effect compares the average outcome that would be observed if the exposure were present for everyone in the population with the average outcome that would be observed if the exposure were absent for everyone. The natural direct effect (NDE) compares the effect of being exposed with the effect of being unexposed while in both cases fixing the mediator to the level at which it would have been naturally in the absence of exposure. Thus, the NDE captures the effect of exposure on the outcome via pathways that do not involve the mediator. The NIE or mediated effect compares average outcomes that would be observed if we were to set the exposure as present and change the mediator for each individual from the level it would have been at in the absence of exposure to the level it would have been at in the presence of exposure. Therefore, the NIE captures the effect of exposure on the outcome operating through the mediator.

In order for the direct and indirect effect estimates to have a causal interpretation, control must be made for the confounding variables. Four assumptions are needed: 1) no unmeasured confounding of the exposure-outcome relationship; 2) no unmeasured confounding of the mediator-outcome relationship; 3) no unmeasured confounding of the exposure-mediator relationship; and 4) no mediator-outcome confounder which is affected by the exposure (12, 15). We describe the formal causal framework in Web Appendix 1 (available at http://aje.oxfordjournals.org/), including notation, definitions of the causal effects, and the confounding assumptions.

Next we describe the relationship between the difference method and the NIE for both continuous and binary outcomes. Proofs of the results are given in Web Appendix 2.

For continuous outcomes, we have the following result:

Result 1. *If the confounding assumptions 1–4 hold and the regression models 1 and 2 are correctly specified, then, on the difference scale, the total effect is ϕ_1 and the NDE is θ_1, and the difference method is consistent for estimating the NIE; that is, the NIE is $\phi_1 - \theta_1$.*

Result 1 shows that the difference method coincides with the counterfactual approach for continuous outcomes and thus provides a formal causal interpretation of the difference method. This result for continuous outcomes also follows from previous literature on mediation (5, 10, 16, 17). Result 1 is correct only under correct specification of models 1 and 2 (equations 1 and 2). When interactions or other nonlinear terms are present, model 2 is not correctly specified, and then the difference $\phi_1 - \theta_1$ will no longer be equal to the NIE and will not have a straightforward interpretation. However, we can still use more advanced methods to estimate the NIE. In practice, before implementing the difference method, investigators should evaluate whether the confounding assumptions hold and the models are correctly specified. If the confounding assumptions do not hold, sensitivity analysis should be conducted (16, 18).

For binary outcomes, when the confounding assumptions hold with correctly specified models 3 and 4 (equations 3 and 4), if the outcome is rare and M follows a linear regression model, the difference method is approximately equal to the NIE on the log odds ratio scale (5, 15). When the outcome is rare, the odds ratio is approximately equal to the risk ratio, which is "collapsible" (19), and the difference method can be used if the models are correctly specified.

However, when the outcome is not rare, the odds ratio no longer approximates the risk ratio, and problems arise because the odds ratio is "noncollapsible." Fortunately, the difference method will often be conservative for the NIE. We state this formally in the following result:

Result 2. *If the confounding assumptions hold and logistic regression models 3 and 4 are correctly specified, then on the odds ratio scale the total effect is e^{ϕ_1}, and the difference method is always conservative for estimating the NIE; that is, if $\theta_1 \geq 0$, the NIE on the odds ratio scale is greater than $e^{\phi_1 - \theta_1}$.*

Result 2 shows that the causal interpretation for the difference method is a lower bound for the NIE. Neuhaus and Jewell (20) show that estimates of parameters will move away from the null when entering additional covariates into the logistic regression, even if the covariates are not associated with

one another. Since the NDE is a quantity obtained after adding the mediator to the regression model, the estimated θ_1 will overestimate the NDE. Therefore, the difference method will underestimate the NIE because ϕ_1 is a consistent estimator of the total effect. Result 2 can help us to obtain qualitative conclusions about the NIE. In the case of positive θ_1, when the quantity $(\phi_1 - \theta_1)$ estimated by the difference method is positive, we can conclude that the NIE must be positive. However, when the quantity estimated by the difference method is zero or negative, we cannot draw any conclusions about the sign of the NIE or about mediation.

In the case of negative θ_1, we can obtain analogous results. When the quantity estimated by the difference method is negative, we can conclude that the NIE must be negative. However, when the quantity estimated by the difference method is zero or positive, we cannot draw any conclusions about the sign of the NIE or about mediation. In practice, we may also want to obtain statistical evidence of the sign of the NIE. If we believe that θ_1 is greater than 0 based on some prior knowledge and obtain the result that the lower 95% confidence bound of the difference method is positive, we can conclude that the lower 95% confidence bound of the NIE is also positive. If we do not have prior knowledge about the sign of θ_1, we can still draw a conclusion about the NIE based on the confidence region of θ_1 and $\phi_1 - \theta_1$. We present the formal result in Web Appendix 3. Other qualitative conclusions are presented in Web Appendix 4.

ILLUSTRATION

We shall provide an example to illustrate the applicability of result 2 in the use of the difference method with logistic regression. In their study, Douglas et al. (21) consider the effect of adverse childhood events on substance dependence, mediated by mood and anxiety disorders. The primary outcome variable is a dichotomized indicator of lifetime substance dependence diagnosis, and the exposure is a cumulative variable, "number of types of violent crime/abuse experiences," which varies from 0 to 3, depending on whether the subject reported a history of violent crime victimization, physical abuse, or sexual abuse. The mediator is an index of lifetime mood and anxiety disorders. Two logistic regression models are fitted to the data, one with adjustment for the mediator and one without adjustment for the mediator. The estimated odds ratios for the exposure, number of types of violent crime/abuse experiences, are 1.76 (95% confidence interval: 1.35, 2.28) without controlling for the mediator and 1.42 ($P = 0.02$) when controlling for the mediator (21); the ratio between the 2 values for the indirect effect estimate equals approximately 1.24 (i.e., 1.76/1.42) and is statistically significant ($P < 0.01$). From result 2, if we have prior knowledge that the log odds ratio for the treatment controlling for the mediator is positive, then the 95% lower bound of the NIE is larger than 1, and we have evidence that the NIE is positive; thus, there is mediation. We can therefore conclude that the effect of adverse childhood events on substance dependence is mediated by mood and anxiety disorders.

DISCUSSION

In this paper, we have provided a new perspective on the causal interpretation of the difference method for logistic regression analysis. In this case, the difference method is always conservative for estimating the NDE on the log odds ratio scale and thus gives a lower bound of the NIE, which can be helpful in drawing qualitative conclusions about mediation.

Result 2 requires the logistic models to be correctly specified for Y both with and without the mediator, and these models may not be compatible with each other, because marginalizing a conditional logistic data distribution may lead to a distribution that is no longer logistic. However, Davidson and MacKinnon (22) point out that in most cases, the only real difference between the probit and logit models is the way in which the parameters are scaled. Thus, the logistic regression models and the probit regression models are approximately equivalent. Since it is common for the probit model to hold both marginally and conditionally, the logistic regression models with and without the mediator may both hold at least approximately.

When interaction terms for interaction between exposure and mediator are included in the regression models, the difference method can no longer be used to draw causal conclusions about mediation. Other estimators for direct and indirect effects, based on the regression analysis, can be used when there is exposure-mediator interaction (5, 16, 19).

The results presented here make it clear that the approach typically used by epidemiologists to assess mediation, the difference method, leads to conservative inferences in the case of logistic regression. Provided that control has been made for the relevant confounding relationships, if the difference method indicates mediation, then there is indeed evidence for mediation. However, if the difference method does not indicate mediation, this does not help us reason about the presence or absence of mediation because the difference method is conservative. Understanding what we can and cannot learn from the difference method is important in ensuring correct reasoning about mechanisms. It would be of interest to extend the results shown here to settings beyond logistic regression.

ACKNOWLEDGMENTS

Author affiliations: School of Mathematical Sciences, Peking University, Beijing, China (Zhichao Jiang); and Departments of Epidemiology and Biostatistics, Harvard School of Public Health, Boston, Massachusetts (Tyler J. VanderWeele).

The research was supported by National Institutes of Health grant ES017876.

Conflict of interest: none declared.

REFERENCES

1. Judd CM, Kenny DA. Process analysis: estimating mediation in treatment evaluations. *Eval Rev*. 1981;5(5):602–619.
2. Baron RM, Kenny DA. The moderator-mediator variable distinction in social psychological research: conceptual, strategic, and statistical considerations. *J Pers Soc Psychol*. 1986;51(6):1173–1182.
3. MacKinnon DP. *Introduction to Statistical Mediation Analysis*. New York, NY: Lawrence Erlbaum Associates; 2008.

4. 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(1):4.
5. VanderWeele TJ, Vansteelandt S. Conceptual issues concerning mediation, interventions and composition. *Stat Interface*. 2009;2(4):457–468.
6. Freedman LS, Graubard BI, Schatzkin A. Statistical validation of intermediate endpoints for chronic diseases. *Stat Med*. 1992; 11(2):167–178.
7. Lin DY, Fleming TR, De Gruttola V. Estimating the proportion of treatment effect explained by a surrogate marker. *Stat Med*. 1997;16(13):1515–1527.
8. Li Z, Meredith MP, Hoseyni MS. A method to assess the proportion of treatment effect explained by a surrogate endpoint. *Stat Med*. 2001;20(21):3175–3188.
9. Chen C, Wang H, Snapinn SM. Proportion of treatment effect (PTE) explained by a surrogate marker. *Stat Med*. 2003;22(22): 3449–3459.
10. MacKinnon DP, Fairchild AJ, Fritz MS. Mediation analysis. *Annu Rev Psychol*. 2007;58:593–614.
11. Robins JM, Greenland S. Identifiability and exchangeability for direct and indirect effects. *Epidemiology*. 1992;3(2): 143–155.
12. Pearl J. Direct and indirect effects. In: *Proceedings of the Seventeenth Conference on Uncertainty in Artificial Intelligence*. San Francisco, CA: Morgan Kaufmann Publishers; 2001:411–420.
13. Rubin DB. Formal mode of statistical inference for causal effects. *J Stat Plan Infer*. 1990;25(3):279–292.
14. Hernán MA. A definition of causal effect for epidemiological research. *J Epidemiol Community Health*. 2004;58(4):265–271.
15. Tchetgen Tchetgen EJ. A note on formulae for causal mediation analysis in an odds ratio context. *Epidemiol Method*. 2014;2(1): 21–31.
16. Imai K, Keele L, Yamamoto T. Identification, inference and sensitivity analysis for causal mediation effects. *Stat Sci*. 2010; 25(1):51–71.
17. VanderWeele TJ. *Explanation in Causal Inference: Methods for Mediation and Interaction*. New York, NY: Oxford University Press; 2015.
18. VanderWeele TJ. Bias formulas for sensitivity analysis for direct and indirect effects. *Epidemiology*. 2010;21(4):540–551.
19. Valeri L, VanderWeele TJ. Mediation analysis allowing for exposure-mediator interactions and causal interpretation: theoretical assumptions and implementation with SAS and SPSS macros. *Psychol Methods*. 2013;18(2):137–150.
20. Neuhaus JM, Jewell NP. A geometric approach to assess bias due to omitted covariates in generalized linear models. *Biometrika*. 1993;80(4):807–815.
21. Douglas KR, Chan G, Gelernter J, et al. Adverse childhood events as risk factors for substance dependence: partial mediation by mood and anxiety disorders. *Addict Behav*. 2010;35(1):7–13.
22. Davidson R, MacKinnon JG. *Econometric Theory and Methods*. New York, NY: Oxford University Press; 2004.