Chao Cheng, Donna Spiegelman, and Fan Li

An evaluation of the product method for estimation and inference in mediation analysis

https://doi.org/..., Received ...; accepted ...

Abstract: The purpose of this article is to comprehensively investigate the product method for mediation analysis. We considered data types in which the mediator and outcome variables were binary and continuous. In the binary outcome scenarios, we newly derived exact expressions for two commonly used mediation measures, the natural indirect effect (NIE) and the mediation proportion (MP), without a need to invoke the rare outcome assumption as previously. We conducted an extensive simulation study to examine the performance of the product method for calculating point and interval estimates of these two measures, and compared the performance of the exact and approximate estimators when the rare outcome assumption was violated. Two approaches to obtain the confidence interval were evaluated: the multivariate delta method and percentile bootstrap approach. The simulation studies established that: (1) A sample size of at least 500 is needed to obtain accurate NIE point and interval estimates; (2) In order to obtain accurate MP point estimates, a sample size of 500 is required for the continuous outcome scenarios and a sample size of 5000 or at least 200 cases are required for the binary outcome scenarios; (3) The multivariate delta method provides satisfactory interval estimates when the sample size $\geq$ 500 for the cases of continuous outcome, and sample size $\geq$ 20,000 and number of cases $\geq$ 500 for the cases of binary outcome; (4) The rare outcome assumption generally performs well when the outcome prevalence is less than 5%, but not otherwise. An *R-software package* is provided for researchers to apply the mediation models considered in this paper.

Keywords: Mediation analysis, mediation proportion, natural indirect effect, product method, simulation study

Chao Cheng, Donna Spiegelman, Fan Li, Department of Biostatistics, Yale School of Public Health, New Haven, CT

Chao Cheng, Donna Spiegelman, Fan Li, Center for Methods in Implementation and Prevention Science, Yale School of Public Health, New Haven, CT

PACS: ...

Communicated by: ...
Dedicated to ...

1 Introduction

Biomedical and epidemiological studies evaluating the impact of exposure on health outcomes have received considerable attention in recent years. In addition to estimating the total effect of the exposure on the disease, it is also of interest to explore potential pathways and mechanisms underlying these exposure-disease relationships. An important tool for exploring these pathways is mediation analysis (Baron and Kenny, 1986; VanderWeele, 2015), which decomposes the *total effect* (TE) of the exposure into two components, a *natural indirect effect* (NIE) through a pre-specified intermediate variable (i.e., the mediator) and a *natural direct effect* (NDE) whose impact derives solely from the exposure. The NIE tells us how much we could exploit the exposure-disease mechanism through interventions targeting the mediator, in cases when it is difficult to manipulate the exposure (Barfield et al, 2017).

In mediation analysis, researchers sometimes calculate the ratio of the NIE and the TE to capture the relative importance of the mediator in explaining the pathway through which the exposure has an effect on the outcome (VanderWeele, 2013). This ratio is often called *mediation proportion* (MP). There has been an increasing number of epidemiological studies that report the MP to explain the exposure-disease mechanism when conducting mediation analysis (e.g., Bebu et al., 2017; Bowe et al., 2020; Chang et al. 2019; Huang et al, 2017; Inzaule et al., 2018; Parker et al., 2016), and some of these studies include only a small to moderate sample size not exceeding 1000 (e.g., Chang et al., 2019; Huang et al, 2017; Parker et al., 2016). Given the increasing use of MP in epidemiological studies, a comprehensive evaluation of the finite-sample operating characteristics of the point and interval estimators for MP is needed to better inform practice.

There are two popular methods for carrying out mediation analysis, the *difference method* and the *product method* (MacKinnon, Warsi and Dwyer, 1995; VanderWeele, 2016). The difference method evaluates two separate exposure-outcome models with and without the mediator, and quantifies the NIE by the difference in two coefficients of both models. The product method evaluates an exposure-outcome model and an exposure-mediator model, and quantifies the NIE by the product of two coefficients of these models. When both the outcome and mediator variables are continuous and ordinary least-squares regression is applied to the models, the difference method and product method are algebraically equivalent (MacKinnon, Warsi and Dwyer, 1995). Under the scenarios of binary outcome or binary mediator, however, the product method generally does not coincide with the difference method. However, it should be noted that both the product and difference methods could produce unbiased estimates of NIE and MP under their respective assumptions. While there have been extensive evaluations of the difference method (Baron and Kenny, 1986; MacKinnon, Warsi and Dwyer, 1995; Nevo, Liao and Spiegelman, 2017), empirical evaluations of the product method are relatively limited.

The purpose of this paper is to evaluate the empirical performance of the point and interval estimators of NIE and MP obtained by the product method, which has been increasingly used in epidemilogic studies in recent years (e.g., Bebu et al., 2017; Agerbo et al., 2015; Janssen et al., 2015; Dadvand et al., 2014; InterAct Consortium; 2013). Because biased estimates or inaccurate interval estimates could lead to misleading interpretations of mediation results, an understanding of the empirical performance of the product method is important to justify its use. Different from several prior studies that evaluated the performance of mediation tests (Barfield et al, 2017; Fritz and MacKinnon, 2007; MacKinnon et al., 2002), our focus is on bias and coverage for estimating the mediation measures. Our work builds on several existing work with a similar objective. For example, MacKinnon, Warsi, and Dwyer (1995) examined the performance of the point and interval estimators of NIE and MP, but restricted their studies to the scenario of continuous outcome and mediator; MacKinnon, Lockwood, and Williams (2004) investigated the interval estimates for NIE and demonstrated that more accurate confidence limits are obtained by the bootstrap; Rijnhart et al. (2019) compared the bias and relative efficiency across seveal methods for estimating the NIE and MP in the case of a binary outcome. To the best of the authors' knowledge, however, the empirical performance of the interval estimators, especially for MP, with either a binary outcome or a binary mediator has not been previously investigated. To provide a comprehensive assessment of the product method in epidemiological mediation studies, we examine the point and interval NIE and MP estimators under the following four scenarios: **Case #1**, continuous outcome and continuous mediator; **Case #2**, continuous outcome and binary mediator; **Case #3**, binary outcome and continuous mediator; and **Case #4**, binary outcome and binary mediator. We compared the multivariate delta method and percentile bootstrap approach for calculating the confidence intervals of mediation measures. Table 1 summarizes the scenarios that have been previously considered, so readers can easily identify the novel contributions of this paper.

When the outcome is binary, mediation analyses frequently make the rare disease or outcome assumption and compute the approximate NIE and MP (VanderWeele, 2015). The SAS macro given by Valeri and VanderWeele (2013) produces these approximate estimates with a binary outcome. However, it remains to be explored to what extent such approximation is accurate, especially when the outcome or disease becomes less rare. In this paper, we derived the exact NIE and MP expressions without rare outcome assumption, and compare the performance of the exact and approximate expressions under rare and common outcome prevalences. We also developed a R-software package `mediateP` (see Appendix A for more details) to facilitate the implementation of the mediation models considered in this work.

The structure of this paper is as follows. First, we describe the use of the product method in mediation analysis to obtain point and interval estimators of NIE and MP in different cases with different types of variables. Second, we describe a comprehensive Monte Carlo simulation study to investigate the performance of the point and interval estimators for NIE and MP. Third, we present the results from the simulation study. As an illustrative example, we perform a mediation analysis for the effect of an anti-retroviral delivery intervention on 12-month retention in HIV care by 6-month visit adherence among participants in the MaxART study (Khan et al., 2020). Finally, we offer a brief discussion and provide practical guidance for researchers when applying the product method in mediation analysis.

2 The product method for mediation analysis

2.1 Continuous mediator

Assume the following conditional mean model for the disease outcome (Y),

$$g(E(Y|X, M, \boldsymbol{W})) = \beta_0 + \beta_1 X + \beta_2 M + \boldsymbol{\beta}_3^T \boldsymbol{W}, \tag{1}$$

where $g(.)$ is a link function, X is a binary or continuous exposure, M is a continuous mediator, $\boldsymbol{W}$ is a vector of covariates associated with outcome and measured before the exposure, some of which may be confounders of the estimated exposure-outcome association and/or the mediator-outcome association. Here, β_1 is the exposure effect on the outcome conditional on the effects of the mediator and confounders, β_2 represents the relationship between the mediator and outcome conditional on the effect of the exposure and confounders. Common link functions include the identity function when the outcome is continuous, and the log and logistic functions when the outcome is binary. In addition to the outcome model (1), with a continuous mediator, the product method for assessing mediation effect additionally requires fitting the following model for the mediator:

$$E(M|X, \boldsymbol{W}) = \gamma_0 + \gamma_1 X + \boldsymbol{\gamma}_2^T \boldsymbol{W}, \tag{2}$$

where γ_1 represents the association between the exposure to the mediator conditional on the effects of the covariates. For simplicity of notation but with no loss of generality, we assume that the mediator model and the outcome model share the same set of covariates. We can set some elements in $\boldsymbol{\beta}_3$ or $\boldsymbol{\gamma}_2$ to zero in the scenarios that the covariates in the outcome and mediator model are not exactly the same. A directed acyclic graph illustrating the causal relationship between Y, X, M and $\boldsymbol{W}$ is shown in Figure 1.

First, we consider Case #1, where the outcome is continuous and an identity link function is assumed in the outcome model (1). If the outcome and mediator models are correctly specified and the identifiability assumptions (Valeri and VanderWeele, 2013) hold, the NIE, NDE and TE, defined for X in change from x^* to x conditional on $\boldsymbol{W} = \boldsymbol{w}$, can be expressed as $\beta_2\gamma_1(x - x^*)$, $\beta_1(x - x^*)$ and $(\beta_2\gamma_1 + \beta_1)(x - x^*)$, respectively. The identifiability assumptions include (A.1) no unmeasured confounding of the exposure-outcome relationship, (A.2) no unmeasured confounding of mediator-outcome relationship, (A.3) no unmeasured confounding of the exposure-mediator relationship, and (A.4) no mediator-outcome confounders affected by the exposure. The mediation proportion, given by the ratio of NIE to TE ($\text{MP}=\frac{\text{NIE}}{\text{TE}} = \frac{\text{NIE}}{\text{NIE+NDE}}$), is $\frac{\beta_2\gamma_1}{\beta_2\gamma_1+\beta_1}$. Finally, the estimators of NIE and MP are obtained by replacing the parameters with the corresponding maximum likelihood or ordinary least-squares estimators.

The most commonly used approach for estimating the variances of $\widehat{\text{NIE}}$ and $\widehat{\text{MP}}$ is the first-order multivariate delta method (Oehlert, 1992):

$$\widehat{\text{Var}}(\widehat{\text{NIE}}) = \left(\frac{\partial \mathscr{N}\mathscr{I}\mathscr{E}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\bigg|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}}\right)^T \widehat{\boldsymbol{\Sigma}}_{\boldsymbol{\theta}} \frac{\partial \mathscr{N}\mathscr{I}\mathscr{E}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\bigg|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}} \text{ and}$$
$$\widehat{\text{Var}}(\widehat{\text{MP}}) = \left(\frac{\partial \mathscr{M}\mathscr{P}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\bigg|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}}\right)^T \widehat{\boldsymbol{\Sigma}}_{\boldsymbol{\theta}} \frac{\partial \mathscr{M}\mathscr{P}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\bigg|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}},$$

where $\boldsymbol{\theta} = [\gamma_0, \gamma_1, \boldsymbol{\gamma}_2^T, \beta_0, \beta_1, \beta_2, \boldsymbol{\beta}_3^T]^T$ contains all the unknown parameters in the mediator and outcome models, $\mathscr{N}\mathscr{I}\mathscr{E}(\boldsymbol{\theta})$ and $\mathscr{M}\mathscr{P}(\boldsymbol{\theta})$ are the parametric expressions of NIE and MP, respectively (e.g., $\mathscr{N}\mathscr{I}\mathscr{E}(\boldsymbol{\theta}) = \beta_2\gamma_1(x - x^*)$ and $\mathscr{M}\mathscr{P}(\boldsymbol{\theta}) = \frac{\beta_2\gamma_1}{\beta_2\gamma_1+\beta_1}$). Here, $\hat{\boldsymbol{\Sigma}}_{\boldsymbol{\theta}} = \begin{bmatrix} \widehat{\boldsymbol{\Sigma}}_{\boldsymbol{\gamma}} & \mathbf{0} \\ \mathbf{0} & \widehat{\boldsymbol{\Sigma}}_{\boldsymbol{\beta}} \end{bmatrix}$ is the estimated variance-covariance matrix of $\hat{\boldsymbol{\theta}}$, $\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{\gamma}}$ and $\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{\beta}}$ are the estimated variance-covariance matrix of $\hat{\boldsymbol{\gamma}} = [\hat{\gamma}_0, \hat{\gamma}_1, \hat{\boldsymbol{\gamma}}_2^T]^T$ and $\hat{\boldsymbol{\beta}} = [\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2, \hat{\boldsymbol{\theta}}_3^T]^T$, which can be obtained by standard regression output in (1) and (2) directly. Note that the covariance of $\hat{\boldsymbol{\gamma}}$ and $\hat{\boldsymbol{\beta}}$ in $\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{\theta}}$ is set as $\mathbf{0}$ because they are asymptotically uncorrelated. This result applies to all four cases we consider, and further explanations are provided in the Supplementary Material Appendix A. If NIE is defined for a one unit increase in X (i.e., $x - x^* = 1$), $\widehat{\text{NIE}} = \hat{\beta}_2\hat{\gamma}_1$ only depends on two elements in $\hat{\boldsymbol{\theta}}$, and the expression of $\widehat{\text{Var}}(\widehat{\text{NIE}})$ becomes $\hat{\beta}_2^2\widehat{\text{Var}}(\hat{\gamma}_1) + \hat{\gamma}_1^2\widehat{\text{Var}}(\hat{\beta}_2)$, which is also known as the Sobel estimator of variance (Sobel, 1982). If $x - x^* = \delta$ for some pre-specified magnitude δ, the point and variance estimators of NIE becomes to $\delta\hat{\beta}_2\hat{\gamma}_1$ and $\delta^2\hat{\beta}_2^2\widehat{\text{Var}}(\hat{\gamma}_1) + \delta^2\hat{\gamma}_1^2\widehat{\text{Var}}(\hat{\beta}_2)$, respectively. Given these variance estimators, 95% confidence intervals of NIE and MP can be computed by normal approximation as $\left\{\widehat{\text{NIE}} - 1.96 \times \sqrt{\widehat{\text{Var}}(\widehat{\text{NIE}})}, \widehat{\text{NIE}} + 1.96 \times \sqrt{\widehat{\text{Var}}(\widehat{\text{NIE}})}\right\}$ and $\left\{\widehat{\text{MP}} - 1.96 \times \sqrt{\widehat{\text{Var}}(\widehat{\text{MP}})}, \widehat{\text{MP}} + 1.96 \times \sqrt{\widehat{\text{Var}}(\widehat{\text{MP}})}\right\}$, respectively.

Alternatively, the percentile bootstrap approach (Davison and Hinkley, 1997; Efron and Tibshirani, 1993) can be used to approximate the empirical distributions of $\widehat{\text{NIE}}$ by resampling the dataset with replacement and re-estimating all model parameters. The values for NIE are then calculated for each bootstrap dataset. This step of resampling and calculating NIE is repeated for a large of times (generally 1,000 or more), and then the bootstrap distribution of NIE is obtained from the collection of estimates based on resampled datasets. Finally, the percentile bootstrap approach employs the 2.5% and 97.5% percentiles of the bootstrap sample distribution to obtain a 95% confidence interval of NIE. Similarly, the percentile bootstrap approach can be used to obtain a 95% confidence interval of MP. When implementing the product method, several studies (Bollen and Stine, 1990; MacKinnon, Fairchild, and Fritz, 2007; MacKinnon, Lockwood and Williams, 2004; Shrout and Bolger, 2002) suggested using the bootstrap approach to calculate confidence intervals for the mediation measures, because the finite sample distribution of the mediation measure estimators, especially the MP estimator, usually exhibit skewness and may not be well approximated by the normal distribution, a critical assumption involved by the multivariate delta method in calculating the confidence interval. Because the bootstrap does not require the normality approximation, its confidence interval may be more accurate than that of the delta method, especially for smaller sample sizes.

In model (1), if the outcome is binary and modeled by a logistic link function (Case #3), and the error term in model (2) follows a normal distribution, the mediation measures, defined on a log odds ratio scale for X in change from x^* to x conditional on $\boldsymbol{W} = \boldsymbol{w}$, involve logistic-normal integrals (See Table 2 for the expressions and Supplementary Material Appendix B for the derivations). These integrals do not have closed-form solutions. Gaynor et al. (2019) uses a probit function to approximate the logistic function in the integral and obtained closed-form expressions for the mediation measures (Table 2). However, the probit approximation could be inaccurate as the outcome prevalence deviates from 50% (Gaynor et al., 2019; Supplementary Material Appendix B). In this paper, instead of approximating the logistic-normal integrals, we use Gauss-Hermite Quadrature (GHQ) approach (Liu and Pierce, 1994) to numerically calculate the integrals. Then, $\widehat{\text{NIE}}$ and $\widehat{\text{MP}}$ are obtained by plugging $\hat{\boldsymbol{\theta}}$ in their expressions shown in Table

2, in which all the integrals are numerically computed by the GHQ. Similar to Case #1, a multivariate delta method or percentile bootstrap approach can be used to calculate their 95% confidence intervals. If the outcome is rare, we can approximate the NIE and TE as $\beta_2\gamma_1(x-x^*)$ and $(\beta_2\gamma_1+\beta_1)(x-x^*)$, respectively (VanderWeele and Vansteelandt, 2010; Supplementary Material Appendix B). It follows that $\mathrm{MP} \approx \frac{\beta_2\gamma_1}{\beta_2\gamma_1+\beta_1}$. Note that the approximate causal mediation measures shown here coincide with the mediation measures in Case #1. In order to distinguish the approximate expressions under a rare outcome assumption from the exact expressions, we write both approximate mediation measures as $\mathrm{NIE}^{(a)}$ and $\mathrm{MP}^{(a)}$, respectively.

2.2 Binary Mediator

In this section, we define the product method estimators when the mediator is a binary variable. First, consider the case when outcome is continuous (Case #2 in the introduction) and the following two models are assumed for mediation analysis:

$$E(Y|X, M, \boldsymbol{W}) = \beta_0 + \beta_1 X + \beta_2 M + \boldsymbol{\beta}_3^T \boldsymbol{W}, \tag{3}$$

$$\operatorname{logit}\Big(P(M=1|X, \boldsymbol{W})\Big) = \gamma_0 + \gamma_1 X + \boldsymbol{\gamma}_2^T \boldsymbol{W}, \tag{4}$$

where X, M and $\boldsymbol{W}$ are defined in Section 2.1, and $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ are the corresponding regression coefficients. If the models are correctly specified and the identifiability assumptions (A.1)–(A.4) in the previous section hold, the NIE and NDE conditional on $\boldsymbol{W}=\boldsymbol{w}$ for a change in exposure from level x^* to level x can be expressed as $\mathrm{NIE} = \beta_2\Big(\frac{e^{\gamma_0+\gamma_1 x+\boldsymbol{\gamma}_2^T\boldsymbol{w}}}{1+e^{\gamma_0+\gamma_1 x+\boldsymbol{\gamma}_2\boldsymbol{w}}} - \frac{e^{\gamma_0+\gamma_1 x^*+\boldsymbol{\gamma}_2^T\boldsymbol{w}}}{1+e^{\gamma_0+\gamma_1 x^*+\boldsymbol{\gamma}_2^T\boldsymbol{w}}}\Big)$ and $\mathrm{NDE} = \beta_1(x-x^*)$, respectively, as shown in Barfield et al. (2017). As a result, we find that

$$\mathrm{MP} = \frac{\mathrm{NIE}}{\mathrm{NDE}+\mathrm{NIE}} = \frac{\beta_2\Big(\frac{e^{\gamma_0+\gamma_1 x+\boldsymbol{\gamma}_2^T\boldsymbol{w}}}{1+e^{\gamma_0+\gamma_1 x+\boldsymbol{\gamma}_2^T\boldsymbol{w}}} - \frac{e^{\gamma_0+\gamma_1 x^*+\boldsymbol{\gamma}_2^T\boldsymbol{w}}}{1+e^{\gamma_0+\gamma_1 x^*+\boldsymbol{\gamma}_2^T\boldsymbol{w}}}\Big)}{\beta_1(x-x^*) + \beta_2\Big(\frac{e^{\gamma_0+\gamma_1 x+\boldsymbol{\gamma}_2^T\boldsymbol{w}}}{1+e^{\gamma_0+\gamma_1 x+\boldsymbol{\gamma}_2^T\boldsymbol{w}}} - \frac{e^{\gamma_0+\gamma_1 x^*+\boldsymbol{\gamma}_2^T\boldsymbol{w}}}{1+e^{\gamma_0+\gamma_1 x^*+\boldsymbol{\gamma}_2^T\boldsymbol{w}}}\Big)}.$$

Substituting in parameter estimators from those regression models leads to $\widehat{\mathrm{NIE}}$ and $\widehat{\mathrm{MP}}$. Again, the multivariate delta method or the percentile bootstrap approach can be used to obtain the 95% confidence interval of both mediation measures.

Now, consider binary mediator and outcome (Case #4 in the introduction). We fit the mediator model (4) and the following logistic regression model for the outcome

$$\operatorname{logit}\Big(P(Y=1|X, M, \boldsymbol{W})\Big) = \beta_0 + \beta_1 X + \beta_2 M + \boldsymbol{\beta}_3^T \boldsymbol{W}. \tag{5}$$

Then, if the models are correctly specified and the identifiability assumptions (A.1)–(A.4) hold, the NIE, NDE on a log odds ratio scale for a change in exposure from level x^* to level x have complex expressions (Gaynor et al., 2019; Supplementary Appendix C), which were provided in Table 2. Now, given NIE and NDE, the MP is given by $\frac{\text{NIE}}{\text{NIE+NDE}}$. If the outcome is rare, the following approximate NIE and NDE have been proposed in VanderWeele (2015): $\text{NIE}^{(a)} = \log\left(\frac{(1+e^{\gamma_0+\gamma_1 x^*+\boldsymbol{\gamma}_2^T \boldsymbol{w}})(1+e^{\beta_2+\gamma_0+\gamma_1 x+\boldsymbol{\gamma}_2^T \boldsymbol{w}})}{(1+e^{\gamma_0+\gamma_1 x+\boldsymbol{\gamma}_2^T \boldsymbol{w}})(1+e^{\beta_2+\gamma_0+\gamma_1 x^*+\boldsymbol{\gamma}_2^T \boldsymbol{w}})}\right)$, $\text{NDE}^{(a)} = \beta_1(x - x^*)$. As a result,

$$\text{MP}^{(a)} = \frac{\log\left(\frac{(1+e^{\gamma_0+\gamma_1 x^*+\boldsymbol{\gamma}_2^T \boldsymbol{w}})(1+e^{\beta_2+\gamma_0+\gamma_1 x+\boldsymbol{\gamma}_2^T \boldsymbol{w}})}{(1+e^{\gamma_0+\gamma_1 x+\boldsymbol{\gamma}_2^T \boldsymbol{w}})(1+e^{\beta_2+\gamma_0+\gamma_1 x^*+\boldsymbol{\gamma}_2^T \boldsymbol{w}})}\right)}{\beta_1(x-x^*) + \log\left(\frac{(1+e^{\gamma_0+\gamma_1 x^*+\boldsymbol{\gamma}_2^T \boldsymbol{w}})(1+e^{\beta_2+\gamma_0+\gamma_1 x+\boldsymbol{\gamma}_2^T \boldsymbol{w}})}{(1+e^{\gamma_0+\gamma_1 x+\boldsymbol{\gamma}_2^T \boldsymbol{w}})(1+e^{\beta_2+\gamma_0+\gamma_1 x^*+\boldsymbol{\gamma}_2^T \boldsymbol{w}})}\right)}.$$

Supplementary Material Appendix C shows that the exact expressions approach to the approximate expressions when the outcome is rare. Again, we can obtain exact or approximate NIE and MP estimates by substituting in the parameter estimators from the mediator and outcome regressions. Similarly, the multivariate delta method or percentile bootstrap approach can be employed to calculate their 95% confidence intervals.

2.3 Simulation Study

We conducted simulation studies under a range of scenarios likely to be encountered in practice to assess the performance of the point and interval estimators of NIE and MP. **R** software (R Core Team, 2019) was used for this simulation study. In the simulation evaluation, we considered the four data types introduced earlier. We considered 4 levels of sample sizes at 150, 500, 1,000, and 5,000 for the continuous outcome cases (Cases #1 and #2). With a baseline outcome prevalence of 3%, to obtain 15, 30, 150, 600 expected cases, we considered sample sizes of 500, 1,000, 5,000, 20,000 for the binary outcome cases (Cases #3 and #4). For each data type and sample size, we considered the exposure (X) as a binary variable and for simplicity assumed that there were no confounders in both the mediator and outcome models. When the outcome was continuous, we set TE $\in (0.25, 0.5, 1)$, indicating small, medium and large total effects; otherwise, we set the TE to $(\log(1.2), \log(1.5), \log(2))$. For each value of TE, we considered MP $\in (0.05, 0.2, 0.5)$. All the mediation measures are defined for a change in X from 0 to 1. We conducted a factorial design of all combinations of TEs, MPs, and sample sizes. For each set of these factors, 5,000 replications were performed. Now, we summarize the data generation procedure used for each data type.

Case #1, continuous outcome and continuous mediator. We first generated the exposure $X \sim Bernoulli(0.5)$. Then, we simulated the mediator $M \sim N(\gamma_0 + \gamma_1 X, 1)$, where $\gamma_0 = 0$. γ_1 was chosen to 0.408, corresponding to the exposure-mediator correlation, $\mathrm{Corr}(X, M)$, 0.2. Finally, given X and M, we simulated $Y \sim N(\beta_0 + \beta_1 X + \beta_2 M, 1)$, where β_0 was fixed as 0. We let $\beta_1 = (1-\mathrm{MP}) \times \mathrm{TE}$ and $\beta_2 = \frac{\mathrm{MP}\times\mathrm{TE}}{\gamma_1}$ based on the relationships that $\mathrm{NDE} = (1 - \mathrm{MP}) \times \mathrm{TE} = \beta_1$ and $\mathrm{NIE} = \mathrm{MP} \times \mathrm{TE} = \beta_2\gamma_1$.

Case #2, continuous outcome and binary mediator. First, we simulated $X \sim Bernoulli(0.5)$. Then, we generated M conditional on X by the logistic regression model $\mathrm{logit}(P(M = 1|X)) = \gamma_0 + \gamma_1 X$. Noting that $\gamma_0 = \log\left(\frac{P(M=1|X=0)}{1-P(M=1|X=0)}\right)$, we chose the values of γ_0 such that the baseline frequency of M, i.e., $P(M = 1|X = 0)$, equaled 0.2. Similar to Case #1, we chose $\gamma_1 = 0.903$ to give a exposure-mediator correlation 0.2. Finally, we simulated $Y \sim N(\beta_0 + \beta_1 X + \beta_2 M, 1)$, where $\beta_0 = 0$, $\beta_1 = (1 - \mathrm{MP}) \times \mathrm{TE}$ from the difinition $\mathrm{NDE} = (1 - \mathrm{MP}) \times \mathrm{TE} = \beta_1$, and $\beta_2 = \frac{\mathrm{MP}\times\mathrm{TE}}{\frac{e^{\gamma_0+\gamma_1}}{1+e^{\gamma_0+\gamma_1}} - \frac{e^{\gamma_0}}{1+e^{\gamma_0}}}$ by noticing $\mathrm{MP} \times \mathrm{TE} = \mathrm{NIE} = \beta_2(\frac{e^{\gamma_0+\gamma_1}}{1+e^{\gamma_0+\gamma_1}} - \frac{e^{\gamma_0}}{1+e^{\gamma_0}})$.

Case #3, binary outcome and continuous mediator. We first generated X and M using the same procedure in Case #1. Now, given X and M, we used the following logistic regression model to simulate Y,

$$\mathrm{logit}\Big(P(Y = 1|X, M)\Big) = \beta_0 + \beta_1 X + \beta_2 M. \tag{6}$$

Since $\beta_0 = \log(\frac{P(Y=1|X=0,M=0)}{1-P(Y=1|X=0,M=0)})$, we selected β_0 so the baseline outcome frequency was 3%. Then, we selected β_1 and β_2 by numerically solving the system of equations $\mathrm{MP} \times \mathrm{TE} = \mathscr{NIE}(\beta_0, \beta_1, \beta_2, \gamma_0, \gamma_1)$ and $(1 - \mathrm{MP}) \times \mathrm{TE} = \mathscr{NDE}(\beta_0, \beta_1, \beta_2, \gamma_0, \gamma_1)$, where $\mathscr{NIE}(\beta_0, \beta_1, \beta_2, \gamma_0, \gamma_1)$ and $\mathscr{NDE}(\beta_0, \beta_1, \beta_2, \gamma_0, \gamma_1)$ refer to the exact expressions of NIE and NDE, given in Supplementary Material Appendix B.

Case #4, binary outcome and binary mediator. We first generated X and M using the same procedure in Case #2, then generated the outcome Y using the logistic regression model (6). The values of β_0, β_1 and β_2 were obtained as follows. We chose β_0 such that the baseline outcome frequency equaled 3%. We then chose β_1 and β_2 by solving the system of equations $\mathrm{MP} \times \mathrm{TE} = \mathscr{NIE}(\beta_0, \beta_1, \beta_2, \gamma_0, \gamma_1)$ and $(1 - \mathrm{MP}) \times \mathrm{TE} = \mathscr{NDE}(\beta_0, \beta_1, \beta_2, \gamma_0, \gamma_1)$, where $\mathscr{NIE}(\beta_0, \beta_1, \beta_2, \gamma_0, \gamma_1)$ and $\mathscr{NDE}(\beta_0, \beta_1, \beta_2, \gamma_0, \gamma_1)$ refer to the exact expressions of NIE and NDE in Supplementary Material Appendix C.

For each data type, we obtained $\widehat{\mathrm{NIE}}$ and $\widehat{\mathrm{MP}}$, variance estimates of $\widehat{\mathrm{NIE}}$ and $\widehat{\mathrm{MP}}$ by the multivariate delta method, and 95% confidence intervals of NIE and MP by the multivariate delta method and percentile bootstrap approach. We did not use the bootstrap approach for the sample size of 20,000 in Cases #3 and

#4, due to the computational cost of performing the number of bootstrapping samples with a large dataset across 5,000 replications. For binary outcome cases, we at first evaluated the performance of the estimators based on the approximate expressions (i.e., $\text{NIE}^{(a)}$ and $\text{MP}^{(a)}$), and next compared the performance between the approximate and the exact mediation measures by varying the baseline outcome prevalence from 1% to 50%, in a separate section.

The percent bias (Bias(%)) was used to evaluate the accuracy of the point estimates of NIE and MP. It was calculated as the median bias relative to the true value over 5,000 replications, i.e., $\text{Bias}(\%) = median(\frac{\hat{p}-p}{p}) \times 100\%$, where p denotes the true value of the causal mediation measure, and $\hat{p}$ is the point estimate of the simulated causal mediation measure. We employed the "median" rather than the "average" value over all the replications to avoid the undue influence of outliers on the results when the sample sizes were not large. The variance ratio was defined as the the ratio between the median of the estimated variance and the empirical variance, and is used to determined the accuracy of the variance estimator obtained by the multivariate delta method. The accuracy of the interval estimator is determined by calculating the empirical coverage rate (CR) of 95% confidence interval across 5,000 replications.

3 Simulation Results

The simulation results for the point, variance and interval estimates of NIE and MP under the four data types are shown in Table 3 (Case #1), Supplementary Material Table 1 (Case #2), Supplementary Material Table 2 (Case #3) and Table 4 (Case #4), respectively. Detailed findings from the simulation study are reported in this Section.

3.1 Estimates of NIE

When the outcome is continuous, the point estimates of NIE generally had minimal bias for sample sizes equal or greater than 500, for all values of TE and MP considered. When the sample was as small as 150, the NIE point estimates had relatively small bias (percent bias within $\pm 10\%$) as long as $\text{MP} \geq 0.2$ and $\text{TE} \geq 0.5$. With binary outcome, however, the point estimates did not show sufficiently low percent bias until the sample size was 1,000, as shown in Table 4 (Case #4) and Supplementary Material Table 2 (Case #3). When the outcome was relatively rare and sample size was not large, there was not enough data to accurately estimate the NIE by the product method. In Case #3 (Supplementary Material Table 2), we also observed that the percent bias of $\widehat{\text{NIE}}^{(a)}$ was notable when TE=log(2) and MP=0.5 even when the sample size was 20,000, indicating that under the rare outcome assumption, bias persisted when TE and MP were also large.

We observed that the variance ratio of the NIE under all data types and sample sizes were very close to their nominal value of 1, indicating that the variance estimator derived by the multivariate delta method was generally accurate. Compared to the multivariate delta method, the bootstrap provided more accurate NIE interval estimates under all sample sizes and data types considered, but its advantage was most notable when the sample size is small. When the outcome was continuous, the bootstrap coverage rates were close to or greater than 95% even when the sample size was 150. However, both the bootstrap and the multivariate delta method had confidence interval coverage rates very close to the nominal value when the sample size $\geq$ 500, except for the scenario when both TE and MP were large in Case #3, in which case both the multivariate delta and bootstrap method exhibited lower coverage rates than nominal as sample size increased, because the rare outcome assumption was inadequate here and so the resulting bias of $\widehat{\text{NIE}}^{(a)}$ are more pronounced. These findings suggested that the multivariate delta method is a valid approach for estimating the variance and confidence intervals for the NIE when the sample size is at least 500 with a continuous outcome.

For the cases of binary outcome, we also evaluated the performance of the NIE estimates based on the exact expressions (i.e., $\widehat{\text{NIE}}$). The results are shown in Supplementary Material Tables 2 and 3 for the results in Cases #3 and #4, respectively. As long as the sample sizes are greater than or equal to 1,000, the $\widehat{\text{NIE}}$ always presented satisfactory percent bias and coverage rate.

3.2 Estimates of MP

In contrast to the NIE estimates, the point estimates of MP often diverged from the true MP values with smaller sample sizes. When the outcome was binary, the percent bias in MP was usually larger than 30% for sample sizes $\leq$ 1,000, and sometimes even larger than 70% for sample size $\leq$ 500, as shown in Table 4 (Case #4) and Supplementary Material Table 2 (Case #3). The MP point estimates were more stable with a continuous outcome, in which case the bias is negligible when the sample size was at least 500. For the binary outcome scenarios, the MP percent bias was not close to 0 until the sample size was 5,000 or the number of cases $\geq$ 200. For all data types, the percent bias of $\widehat{\text{MP}}$ appears to become higher when the TE is small. For the same TE, a smaller percent bias occurred when the MP was larger.

The variance estimators of $\widehat{\text{MP}}$ obtained from the multivariate delta method were smaller than their corresponding empirical variances, and this phenomenon became more noticeable with small sample sizes and small TEs. In general, the multivariate delta method provided accurate variance estimators when the sample size was at least 500 for the continuous outcome scenarios. When the outcome was binary, the multivariate delta method only provided accurate variance estimators when the sample size was at least 5000 and TE $\geq \log(1.5)$.

Similar to the NIE interval estimates, the bootstrap provided more accurate MP interval estimates than the multivariate delta method, especially with smaller sample sizes, because the distribution of $\widehat{\text{MP}}$ with small sample sizes exhibited high skewness, not following a normal distribution, as required by the multivariate delta method. For all data types and sample sizes, the bootstrap approach generally had coverage rates higher than 95% and began to provide close to nominal coverage rates for sample size of 500 for scenarios with continuous outcome and 5,000 for the scenarios with binary outcome. On the other hand, the multivariate delta method did not provide accurate confidence intervals for smaller sample sizes, and its coverage rates sometimes dropped below 80% in some parameter settings in binary outcome scenarios. For continuous outcome scenarios, a sample size of 1,000 was needed for the multivariate delta method to provide satisfactory interval estimates, whereas the sample size requirement for the binary outcome scenarios was substantially larger. As shown in Table 4 (Case #4) and Supplementary Material Table 2 (Case #3), the multivariate delta method provided satisfactory coverage rates when TE $\geq \log(1.5)$ and sample size $\geq 5,000$. In the small TE (TE$= \log(1.2)$) scenarios, a sample size of at least 20,000 was needed for the delta method to provide satisfactory MP interval estimates. The performance of the delta method was also sensitive to the magnitude of TE. When the TE was small, the MP confidence interval tended to be much wider than it should be when the MP was also small, but tends to be narrower than it should be for larger MP.

3.3 The Impact of Outcome Prevalence for Binary Outcome Data

Here, we conducted additional simulations to compare the mediation analyses based on the approximate expressions and exact expressions ($\widehat{\text{NIE}}^{(a)}$ v.s. $\widehat{\text{NIE}}$ and $\widehat{\text{MP}}^{(a)}$ v.s. $\widehat{\text{MP}}$) when the baseline outcome prevalence was varied from 1% to 50%.

We considered TE$\in (\log(1.2), \log(2))$, MP$\in (0.1, 0.5)$ and a large sample size of 20,000.

Figures 2 and Supplementary Material Figure 1 present the results for MP and NIE estimates, respectively, for the binary outcome and binary mediator scenario (i.e., Case #4). The estimates based on the exact causal mediation expressions, $\widehat{\text{MP}}$ and $\widehat{\text{NIE}}$, provided accurate point and interval estimates when the outcome prevalence $> 1\%$. When TE $= \log(1.2)$ and the baseline outcome prevalence was 1%, $\widehat{\text{MP}}$ and $\widehat{\text{NIE}}$ showed significant negative percent bias, as the number of cases (about 200) were quite small. The NIE and MP estimates based on the approximate expressions did not generally exhibit satisfactory performance when the outcome prevalence was high. Specifically, when TE$> \log(1.2)$, the percent bias of $\widehat{\text{MP}}^{(a)}$ diverged from 0 and the coverage rate of $\widehat{\text{MP}}^{(a)}$ by the delta method substantially decreased from 95%. Similarly, $\widehat{\text{NIE}}^{(a)}$ was also quite different from its true value when the baseline outcome prevalence $\geq 10\%$ (See Supplementary Material Figure 1).

With a binary outcome and continuous mediator (Case #3), $\widehat{\text{MP}}$ and $\widehat{\text{NIE}}$ provided accurate point and interval estimates among all baseline outcome prevalences considered, as shown in Supplementary Material Figures 2 and 3, respectively. The $\widehat{\text{MP}}^{(a)}$ also provided very robust point and interval estimates (Supplementary Material Figure 2) for the common outcome scenarios, where its percent bias was less than 1% among all TEs, MPs and outcome prevalences considered. On the other hand, the performance of $\widehat{\text{NIE}}^{(a)}$ was sensitive to the baseline prevalence and MP estimator. For example, when MP=0.5, the percent bias of $\widehat{\text{NIE}}^{(a)}$ increased as baseline outcome prevalence increased, and the confidence interval coverage rates rapidly decreased. When the true MP=0.1, the percent bias of $\widehat{\text{NIE}}^{(a)}$ was negligible over the range of outcome prevalences considered.

3.4 Normality Assumption in Case #3

In the simulation studies above, we simulated $M|X$ under a normal distribution, as assumed in deriving the mediation measure expressions in Case #3. In this section, we evaluate the performance of the product method in Case #3, when M is not assumed to follow the normal distribution. Here, we simulated $M = \gamma_0 + \gamma_1 X + \epsilon$, where ϵ was specified as a gamma distribution. Specifically, $\epsilon \sim \frac{b - E[b]}{\sqrt{\text{Var}(b)}}$, where b follows a gamma distribution with density $f(b) = \frac{1}{\Gamma(k)\theta^k} b^{k-1} e^{-b/\theta}$ $(b > 0)$, k and θ are tuning parameters. We subtract b with its expectation $E[b] = k\theta$ and then divide it by its standard deviation $\sqrt{k}\theta$ in order to fix the mean and variance of ϵ

at 0 and 1, matching the first two moments of the standard normal distribution. Here, we chose $k = (2/s)^2$ and $\theta = s/2$ such that the coefficient of skewness of ϵ was s. In this simulation study, s was set as 1, 1.5 and 2, representing different degrees of skewness. Simulation results are presented in Supplementary Material Table 4. We also added a scenario that ϵ follows a standard normal distribution as a benchmark. We observed that both the exact and approximate method showed robustness with regard to violations of normality assumption, where all the point and variance estimators and confidence interval coverage rates are comparable across the two data generating processes.

4 Application to the MaxART study

We performed a mediation analysis in the MaxART study (Khan et al., 2020), which is a stepped-wedge cluster randomized trial among HIV-positive participants in Eswatini. The primary objective of the study was to understand the impact of early access to antiretroviral therapy (EAAA) versus standard of care (SoC). From September 2014 to August 2017, the MaxART Consortium randomly assigned 14 participating clinics in pairs to shift from SoC to EAAA at randomly chosen pre-specified dates. Further details of the design of this MaxART study were shown at Walsh et al. (2017). The MaxART study previously found that EAAA improved retention in HIV care (Khan et al., 2020), but the mechanisms underlying the intervention-retention relationship is unknown. In this illustrative example, we investigated the extent to which the effect of the intervention (SoC v.s. EAAA) on 12-month retention in HIV care was mediated by visit adherence at 6 months. Participants were classified as retained in HIV care for 12 months if, at the end of the 12th month post enrollment, the participant was alive and had not discontinued treatment, where either the last clinic visit was less than 90 days from the end of study or next scheduled visit date was within 30 days from the end of study. In order to obtain 12-month retention, we required (1) the participant's enrollment date to be longer than 12 months from end of the study and/or (2) if initially receiving SoC treatment, the participant's transition date to EAAA was longer than 12 months from enrollment. Participants who did not meet the above two requirements were excluded. Finally, 1,731 participants were used in our illustrative analysis, with 1,335 individuals retained in care for 12 months and 396 individuals not retained, of whom 1,014 individuals received SoC and 717 individuals received EAAA. Basic characteristics of these participants in our analysis are given in Supplementary Material Table 5. For illustrative purposes, we do not consider possible of clustering of the data, because minimal clustering had been reported previously in these data (Khan et al. 2020).

The hypothesized mediator considered here was 6-month visit adherence, which measures whether a participant's frequency of clinical visits coincides with the MaxART protocol over the first 6 months following enrollment. According to the MaxART protocol, participants are expected to have a follow-up visit in every 30 days. It follows that at the end of the 6th month the participants should have completed 6 or more visits. Here, the definition of 6-month visit adherence completion 5 or more clinical visits by the end of the 6th month after enrollment (yes=1, no=0). Based on this definition, 831 participants adhered to the MaxART visit schedule din the first 6 months, whereas 900 participants did not.

We considered two scenarios for confounding adjustment (i.e., $\boldsymbol{W}$) in the outcome and mediator models. In Scenario I, we only adjusted for the steptime. In Scenario II, we adjusted for all factors that may have been confounders of the intervention-retention relationship, visit adherence-retention relationship, or intervention-visit adherence relationship. From clinical knowledge and prior analyses of these data, the comprehensive set of potential confounders included steptime, age at study enrollment (< 20 yrs, $[20, 30)$ yrs, $[30, 40)$ yrs, $[40, 50)$ yrs, $[50, 60)$ yrs, ≥ 60 yrs), sex, marital status (married, devoiced/widowed, single), education (illiterate/primary, secondary, high school, and tertiary), CD4 counts (< 350 cells/ul, $[350, 500]$ cells/ul, > 500 cells/ul), WHO stage (I, II, III and IV stages), BMI (< 18.5, $[18.5, 25)$, $[25, 30)$, ≥ 30 kg/m^2), screened for TB symptoms (yes, no), viral load (< 5000 copies/ml, $[5000, 30000]$ copies/ml, > 30000 copies/ml), treatment support (yes, no), level of clinic (hospital, clinic with maternity ward, clinic without maternity ward), time from HIV tested positive to enrollment (< 1 yr, 1-3 yrs, > 3 yrs), and clinic volume (low: $<$ median, high $\geq$ median). For simplicity and consistency with the primary analysis of the MaxART study (Khan et al., 2020), the missing indicator method (Groenwold et al., 2012) was used to account for missing covariates, where missing data was treated as a separate group for each of the confounding variables in the models.

We first implemented the product method based on the approximate mediation measure expressions assuming a rare outcome. We coded non-retention as 1 and retention as 0. We calculated the NIE, TE and MP comparing EAAA to SoC, conditional on the mode for each model covariate. Although we estimated the mediation measures at the mode of each model covariate, these measures could also be calculated at other values of the covariates as well. Results are given in Table 5. In Scenario I, the steptime-adjusted model, we found that the intervention was protective against 12-month non-retention, with odds ratios of 0.23 and 0.55 for $\widehat{\mathrm{TE}}^{(a)}$ and $\widehat{\mathrm{NIE}}^{(a)}$ respectively. Because the 95% confidence intervals, either by the delta method or bootstrap, for both parameters excluded the null, we conclude that both effects were significantly different from zero. The steptime-adjusted $\widehat{\text{MP}}^{(a)}$ was 40.8% (95% CI by delta method: (0.27, 0.54)), implying that over 40% of the intervention effect was mediated by 6-month visit adherence. In multivariate-adjusted analyses (Scenario II), stronger NIE and TE effects were obtained, corresponding to odds ratios of 0.08 and 0.38, respectively and the multivariate-adjusted MP estimate was also around 40%. The bootstrap and multivariate confidence intervals were very close in this analysis, although the width of the bootstrap confidence interval was slightly smaller than that using the delta method variance for NIE and TE, but slightly larger than the delta method for MP.

At 23%, the outcome prevalence in this example is not low, the above analysis based on the rare outcome approximation could be biased. Thus, we repeated the mediation analysis using the exact expressions given in Table 5. Generally speaking, the results of the mediation analysis accounting for common outcome prevalence were similar to those obtained in the approximate previous analysis. In both the steptime-adjusted model and the multivariate-adjusted model, the adjusted $\widehat{\text{TE}}$ were slightly weaker than previous results assuming a rare outcome (steptime-adjusted $\widehat{\text{TE}} = 0.24$ and multivariate-adjusted $\widehat{\text{TE}} = 0.10$ on the odds ratio scale, compared to 0.23 and 0.08, respectively). As a result, $\widehat{\text{MP}}$ slightly increased in the adjusted analyses (steptime-adjust ed $\widehat{\text{MP}} = 44\%$, multivariate-adjusted $\widehat{\text{MP}} = 42\%$), compared to 41% and 40%, respectively, with the rare outcome approximation. In summary, for both the results based on and not based on the rare disease assumption, over 40% of the association between intervention and 12-month retention in care was mediated by 6-month visit adherence.

5 Discussion

The difference and product methods are two approaches for estimating NIE and MP in causal mediation analysis (VanderWeele, 2015). While there has been comprehensive empirical evaluations of the difference method (Nevo, Liao and Spiegelman, 2017), the empirical performance of the product method has not been extensively evaluated. For this reason, we conducted a comprehensive simulation study to evaluate the performance of $\widehat{\text{NIE}}$ and $\widehat{\text{MP}}$ obtained by the product method under various scenarios likely to be encountered in practice. We provided the $\widehat{\text{NIE}}$ and $\widehat{\text{MP}}$ estimators without the rare outcome assumption for binary outcome scenarios, and examined extent to which the approximate mediation analysis was robust to violations of the rare outcome assumption. The R package `mediateP` for implementing the mediation methods in this paper can be found at github page https://github.com/chaochengstat/mediateP, and installations and usage instructions for the R package are given in Appendix A.

We demonstrated that $\widehat{\mathrm{NIE}}$ had very little bias and the variance estimate for $\widehat{\mathrm{NIE}}$ were quite close to the true values estimated under all scenarios considered from Monte Carlo simulations. In general, the multivariate delta method provided accurate variance estimates and valid interval estimates once the sample size was at least 500, and the bootstrap remained accurate even when the same size was even smaller. We found that larger sample sizes were needed to obtain valid MP point and interval estimates. Specifically, when the outcome was continuous, a sample size of 500 was required for valid point and interval estimates. In the binary outcome scenarios under a rare disease, sample size of 5000, and 200 cases or more, were required to obtain satisfactory MP point estimate and bootstrap interval estimates. We observed that the multivariate delta method provided valid MP confidence intervals when sample size $\geq$ 20,000 and number of cases $\geq$ 500 in binary outcome scenario.

We confirmed that the bootstrap method provided better interval estimates compared the multivariate delta method in smaller sample size scenarios, as may be found in some social science applications. However, the bootstrap method requires substantially more computational time to fit the mediation models in order to obtain the empirical $\widehat{\mathrm{NIE}}$ or $\widehat{\mathrm{MP}}$ distribution, which may be computationally burdensome in large epidemiological cohort studies. While we recommend bootstrap for the interval estimation with small sample size, when the sample size is larger, sample size $\geq$ 500 for the continuous outcome scenarios and sample size $\geq$ 20,000 and number of cases $\geq$ 500 for studies of binary outcome, we recommend the multivariate delta method for obtaining valid and computationally efficient confidence intervals. To facilitate application of methods examined in our article, our R package `mediateP` implements both variance estimators.

In addition, our simulation study also showed that the accuracy of $\widehat{\mathrm{MP}}$ also depended on the effect size of the TE. When the sample size is too small, a smaller TE is associated with a more biased MP point estimates and interval estimates with under-coverage, especially in binary outcome scenarios. In many epidemiological studies, when there is reason to delieve that the NIE or NDE is not close to zero, a relatively smaller sample size may be adequate for obtaining valid point and interval estimates of the MP. For example, when the outcome is binary, we suggest that a sample size $\geq 20,000$ and number of cases ≥ 500 is needed for the multivariate delta method to obtain satisfactory MP confidence intervals with close to nominal coverage rate. If there is reason to believe that the TE is not too small (TE$\geq \log(1.5)$), we found that the product method could accurately estimate MP with a sample size of at least 5000 and number of cases at least 150.

In the binary outcome scenarios (Cases #3 and #4), the approximate NIE and MP estimators under a rare disease outcome have been commonly used in epidemiological studies (Agerbo et al., 2015; Dadvand et al., 2014; InterAct Consortium; 2013). Our simulation study suggests that, when the outcome prevalence was less than 5%, the rare outcome assumption worked well for $\widehat{\mathrm{NIE}}^{(a)}$ and $\widehat{\mathrm{MP}}^{(a)}$. When the outcome prevalence $\geq 5\%$ and with a binary mediator (Case #4), the percent bias of $\widehat{\mathrm{NIE}}^{(a)}$ and $\widehat{\mathrm{MP}}^{(a)}$ can be substantial. However, with a binary outcome and continuous mediator (Case #3), we found that $\widehat{\mathrm{MP}}^{(a)}$ always provides satisfactory point and interval estimates even when the outcome was common.

Our simulation study has several limitations and future work is needed to supplement the conclusions in this manuscript. For simplicity, our simulation study did not include confounders. However, because epidemiological studies need to adjust for all potential confounding factors to obtain valid results, future research is needed to examine the product method in the presence of confounders, potentially high-dimensional. In addition, when the outcome is binary, the mediation measures can be defined on the odds ratio scale (VanderWeele, 2015). Supplementary Material Appendix D shows the relationship between mediation measures defined on a log odds ratio scale and odds ratio scale.

In summary, we have found that asymptotic inference performs well for the product method in sample sizes typically found in epidemiology and public health settings. In addition, for common binary outcomes, exact expressions are needed to obtain unbiased estimates and have been provided here.

Acknowledgment: This work was funded in part by NIH grant DP1ES025459.

References

[1] Agerbo, E., Sullivan, P. F., Vilhjalmsson, B. J., Pedersen, C. B., Mors, O., Borglum, A. D., ... & Ripke, S. (2015). Polygenic risk score, parental socioeconomic status, family history of psychiatric disorders, and the risk for schizophrenia: a Danish population-based study and meta-analysis. JAMA psychiatry, 72(7), 635-641.

[2] Barfield, R., Shen, J., Just, A. C., Vokonas, P. S., Schwartz, J., Baccarelli, A. A., ... & Lin, X. (2017). Testing for the indirect effect under the null for genome-wide mediation analyses. Genetic epidemiology, 41(8), 824-833.

[3] Baron, R. M., & Kenny, D. A. (1986). The moderator–mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. Journal of personality and social psychology, 51(6), 1173.

[4] Bebu, I., Braffett, B. H., Pop-Busui, R., Orchard, T. J., Nathan, D. M., Lachin, J. M., & DCCT/EDIC Research Group. (2017). The relationship of blood glucose with cardiovascular disease is mediated over time by traditional risk factors in type 1 diabetes: the DCCT/EDIC study. Diabetologia, 60(10), 2084-2091.

[5] Biesanz, J. C., Falk, C. F., & Savalei, V. (2010). Assessing mediational models: Testing and interval estimation for indirect effects. Multivariate Behavioral Research, 45(4), 661-701.

[6] Bollen, K. A., & Stine, R. (1990). Direct and indirect effects: Classical and bootstrap estimates of variability. Sociological methodology, 115-140.

[7] Bowe, B., Xie, Y., Yan, Y., Xian, H., & Al-Aly, Z. (2020). Diabetes Minimally Mediated the Association Between MP 2.5 Air Pollution and Kidney Outcomes. Scientific reports, 10(1), 1-9.

[8] Chang, C. H., Huang, Y. F., Wang, P. W., Lai, C. H., Huang, L. W., Chen, H. C., ... & Chen, M. L. (2019). Associations between prenatal exposure to bisphenol a and neonatal outcomes in a Taiwanese cohort study: Mediated through oxidative stress?. Chemosphere, 226, 290-297.

[9] Dadvand, P., Ostro, B., Figueras, F., Foraster, M., Basagana, X., Valentin, A., ... & Jerrett, M. (2014). Residential proximity to major roads and term low birth weight: the roles of air pollution, heat, noise, and road-adjacent trees. Epidemiology, 518-525.

[10] Davison, A. C., & Hinkley, D. V. (1997). Bootstrap methods and their application (Vol. 1). Cambridge university press.

[11] Efron, B., & Tibshirani, T.J. (1993). An introduction to the bootstrap. New York: Chapman & Hall.

[12] Fasanelli, F., Baglietto, L., Ponzi, E., Guida, F., Campanella, G., Johansson, M., ... & Chadeau-Hyam, M. (2015). Hypomethylation of smoking-related genes is associated with future lung cancer in four prospective cohorts. Nature communications, 6(1), 1-9.

[13] Fritz, M. S., & MacKinnon, D. P. (2007). Required sample size to detect the mediated effect. Psychological science, 18(3), 233-239.

[14] Gaynor, S. M., Schwartz, J., & Lin, X. (2019). Mediation analysis for common binary outcomes. Statistics in medicine, 38(4), 512-529.

[15] Groenwold, R. H., White, I. R., Donders, A. R. T., Carpenter, J. R., Altman, D. G., & Moons, K. G. (2012). Missing covariate data in clinical research: when and when not to use the missing-indicator method for analysis. CMAj, 184(11), 1265-1269.

[16] Huang, L., Wei, Y., Shen, S., Shi, Q., Bai, J., Li, J., ... & Chen, F. (2017). Therapeutic effect of apatinib on overall survival is mediated by prolonged progression-free survival in advanced gastric cancer patients. Oncotarget, 8(17), 29346.

[17] InterAct Consortium. (2013). The link between family history and risk of type 2 diabetes is not explained by anthropometric, lifestyle or genetic risk factors: the EPIC-InterAct study. Diabetologia, 56(1), 60-69.

[18] Janssen, B. G., Byun, H. M., Gyselaers, W., Lefebvre, W., Baccarelli, A. A., & Nawrot, T. S. (2015). Placental mitochondrial methylation and exposure to airborne particulate matter in the early life environment: an ENVIR ON AGE birth cohort study. Epigenetics, 10(6), 536-544.

[19] Khan, S., Spiegelman, D., Walsh, F., Mazibuko, S., Pasipamire, M., Chai, B., . . . & Okello, V. (2020). Early access to antiretroviral therapy versus standard of care among HIV-positive participants in Eswatini in the public health sector: the MaxART stepped-wedge randomized controlled trial. Journal of the International AIDS Society, to appear.

[20] Liu, Q., & Pierce, D. A. (1994). A note on Gauss—Hermite quadrature. Biometrika, 81(3), 624-629.

[21] Inzaule, S. C., Kityo, C. M., Siwale, M., Akanmu, A. S., Wellington, M., de Jager, M., ... & Ondoa, P. (2018). Previous antiretroviral drug use compromises standard first-line HIV therapy and is mediated through drug-resistance. Scientific reports, 8(1), 1-7.

[22] Parker, S. E., Collett, B. R., Speltz, M. L., & Werler, M. M. (2016). Prenatal smoking and childhood behavior problems: is the association mediated by birth weight?. Journal of developmental origins of health and disease, 7(3), 273-281.

[23] MacKinnon, D. P., Fairchild, A. J., & Fritz, M. S. (2007). Mediation analysis. Annu. Rev. Psychol., 58, 593-614.

[24] MacKinnon, D. P., Lockwood, C. M., & Williams, J. (2004). Confidence limits for the indirect effect: Distribution of the product and resampling methods. Multivariate behavioral research, 39(1), 99-128.

[25] MacKinnon, D. P., Lockwood, C. M., Hoffman, J. M., West, S. G., & Sheets, V. (2002). A comparison of methods to test mediation and other intervening variable effects. Psychological methods, 7(1), 83.

[26] MacKinnon, D. P., Warsi, G., & Dwyer, J. H. (1995). A simulation study of mediated effect measures. Multivariate behavioral research, 30(1), 41-62.

[27] Nevo, D., Liao, X., & Spiegelman, D. (2017). Estimation and inference for the mediation proportion. The international journal of biostatistics, 13(2).

[28] Oehlert, G. W. (1992). A note on the delta method. The American Statistician, 46(1), 27-29.

[29] Piessens, R., de Doncker-Kapenga, E., Uberhuber, C. W., & Kahaner, D. K. (2012). Quadpack: a subroutine package for automatic integration (Vol. 1). Springer Science & Business Media.

[30] R Core Team. (2019). R: A language and environment for statistical computing.

R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

[31] Rijnhart, J. J., Twisk, J. W., Eekhout, I., & Heymans, M. W. (2019). Comparison of logistic-regression based methods for simple mediation analysis with a dichotomous outcome variable. BMC medical research methodology, 19(1), 19.

[32] Shrout, P. E., & Bolger, N. (2002). Mediation in experimental and nonexperimental studies: new procedures and recommendations. Psychological methods, 7(4), 422.

[33] Sobel, M. E. (1982). Asymptotic confidence intervals for indirect effects in structural equation models. Sociological methodology, 13, 290-312.

[34] Valeri, L., & VanderWeele, T. J. (2013). Mediation analysis allowing for exposure–mediator interactions and causal interpretation: theoretical assumptions and implementation with SAS and SPSS macros. Psychological methods, 18(2), 137.

[35] VanderWeele, T. J. (2016). Mediation analysis: a practitioner's guide. Annual review of public health, 37, 17-32.

[36] VanderWeele, T. J. (2015). Explanation in causal inference: methods for mediation and interaction. Oxford University Press.

[37] VanderWeele, T. J. (2013). Policy-relevant proportions for direct effects. Epidemiology (Cambridge, Mass.), 24(1), 175.

[38] VanderWeele, T. J., & Vansteelandt, S. (2010). Odds ratios for mediation analysis for a dichotomous outcome. American journal of epidemiology, 172(12), 1339-1348.

[39] Walsh, F. J., Bärnighausen, T., Delva, W., Fleming, Y., Khumalo, G., Lejeune, C. L., ... & Zwane, M. (2017). Impact of early initiation versus national standard of care of antiretroviral therapy in Swaziland's public sector health system: study protocol for a stepped-wedge randomized trial. Trials, 18(1), 383.

Appendix A: Instructions for the `mediateP` package

The `mediateP` package calculates the point and 95% interval estimates for the NIE, TE and MP, based on the product method, as described in this paper. The source files for the `mediateP` package was provided on the github webpage https://github.com/chaochengstat/mediateP.

First, use the following statements to install the `mediateP` package

```
> devtools::install_github("chaochengstat/mediateP")
>## install devtools at first if it is not installed
> library("mediateP")
```

The main function of the "mediateP" package is `mediate()`, which provides the mediation analysis results. It can be called with,

```
mediate(data, outcome, mediator, exposure, binary.outcome,
        binary.mediator, covariate.outcome, covariate.mediator,
        x0, x1, c.outcome, c.mediator, boot, R)
```

The function has 14 arguments. These are

`data=` (Required) The name of the dataset.

`outcome=` (Required) Name of the outcome variable, which should be either a continuous or binary datatype.

`mediator=` (Required) Name of the mediator variable, which should be either a continuous or binary datatype.

`exposure=` (Required) Name of the exposure variable, which should be either a continuous or binary datatype.

`binary.outcome=` (Required) If the outcome is binary, set to 1. If the outcome is continuous, set to 0. The default value is 0.

`binary.mediator=` (Required) If the mediator is binary, set to 1. If the mediator is continuous, set to 0. The default value is 0.

`covariate.outcome=` A vector of names showing the confounding variables used in the outcome model. The default value is `NULL`, which represents no confounding variables. We only accepted continuous and binary confounding variables, if one confounding variable is categorical, please set it to a series of binary variables in advance.

`covariate.mediator=` A vector of names showing the confounding variables used in the mediator model. The default value is `NULL`, which represents no confounding variables. We only accepted continuous and binary confounding variables, if one confounding variable is categorical, please set it to a series of binary variables in advance.

`x0=` (Required) The baseline exposure level (i.e., x^*). The default value is 0.

`x1=` (Required) The new exposure level (i.e., x). The default value is 1.

`c.outcome=` A vector of numbers representing the conditional level of the confounding variables in the outcome model. The default value is a vector of 0.

`c.mediator=` A vector of numbers representing the conditional level of the confounding variables in the mediator model. The default value is a vector of 0.

`boot=` If a percentile bootstrap confidence interval needed to be added, set to 1. Otherwise, set to 0. The default value is 0.

`R=` (Required if `boot=1`) The number of replications when apply the percentile bootstrap method to calculate the confidence interval. The default value is 2,000.

We now illustrate the usage of the `mediate` function. First, using the following statements to simulate a dataset

```
> C1=rnorm(2000)>0 # Confounder 1
> C2=rnorm(2000) # Confounder 2
> X = rnorm(2000) # exposure
> M= as.numeric(runif(2000)< 1/(1+exp(0-0.9*X+0.1*C1))) # mediator
> # outcome
> Y= as.numeric(runif(2000)< 1/(1+exp(-(-2+0.5*X+0.5*M+0.2*C1+0.2*C2))))
> mydata=as.data.frame(cbind(Y,M,X,C1,C2)) # summarize into a dataset
```

This dataset, named `mydata`, includes a continuous exposure (`X`), a binary mediator (`M`), a binary outcome (`Y`), as well as two confounding variables (`C1` and `C2`). `mydata` has 2,000 observations, where the first 6 observations are shown as follows

```
> head(mydata)
  Y M          X C1          C2
1 0 0 -1.1346302  0 -0.88614959
2 0 1  0.7645571  1 -1.92225490
3 0 1  0.5707101  0  1.61970074
4 0 0 -1.3516939  1  0.51926990
5 0 1 -2.0298855  1 -0.05584993
6 0 0  0.5904787  0  0.69641761
```

We conducted a mediation analysis using `mediate()`. In the outcome model, we adjusted for `C1` and `C2`. In the mediator model, we only adjusted for `C1`. We calculated the NIE, TE and MP for exposure in change from 0 to 1, conditional on `C1=0` and `C2=1`, as follows

```
> result=mediate(data=mydata, outcome="Y", mediator="M", exposure="X",
                 binary.outcome=1, binary.mediator=1,
                 covariate.outcome=c("C1","C2"),
                 covariate.mediator=c("C1"),
                 x0=0, x1=1, c.outcome=c(0,1), c.mediator=c(0),
                 boot=1, R=2000)
```

Finally, we got the following mediation analysis output

```
> print.mediateP(result)
                      Point (S.E.)  95% CI by Delta Approach  95% CI by Bootstrap
NIE: Approximate   0.1069 (0.0224)           (0.0631,0.1508)      (0.0632,0.1510)
NIE:       Exact   0.4568 (0.0645)           (0.3304,0.5832)      (0.3364,0.5840)
TE:  Approximate   0.2341 (0.0588)           (0.1190,0.3493)      (0.1300,0.3710)
TE:        Exact   0.1118 (0.0246)           (0.0635,0.1601)      (0.0647,0.1614)
MP:  Approximate   0.4568 (0.0638)           (0.3318,0.5818)      (0.3402,0.5827)
MP:        Exact   0.2447 (0.0631)           (0.1211,0.3683)      (0.1333,0.3901)
```

Tab. 1: A comparison of the current work with several previous literature evaluating the empirical performance of the product method in mediation analysis under four data types: Case #1, continuous outcome and continuous mediator; Case #2, continuous outcome and binary mediator; Case #3, binary outcome and continuous mediator; and Case #4, binary outcome and binary mediator.

Literatures	Natural Indirect Effect							Mediation Proportion						
	Case #1	Case #2	Case #3			Case #4		Case #1	Case #2	Case #3			Case #4	
			Approx.	Exact	P.A.	Approx.	Exact			Approx.	Exact	P.A.	Approx.	Exact
Current work	B.V.I	B.V.I	B.V.I	B.V.I		B.V.I	B.V.I	B.V.I	B.V.I	B.V.I	B.V.I		B.V.I	B.V.I
Barfield et al (2017)	T	T	T			T								
Biesanz, Falk, and Savalei (2010)	I.T													
Fritz and MacKinnon (2007)	T													
Gaynor et al. (2019)					B.I		B.I							
MacKinnon, Warsi, and Dwyer (1995)	B.V							B.V						
MacKinnon et al. (2002)	T													
MacKinnon, Lockwood, and Williams (2004)	I													
Rijnhart et al. (2019)			B.V			B.V				B.V			B.V	

Note: The B, V, I, and T denote the bias, variance, confidence interval, and hypothesis testing, respectively. If one of those appears in one cell, it indicates that this operating characteristic has been covered in this literature. In Cases #3 and #4, Approx., Exact, and P.A. denote the approximate expression, exact expression, and the probit approximation expression, respectively (See Table 2 for their specific formulas).

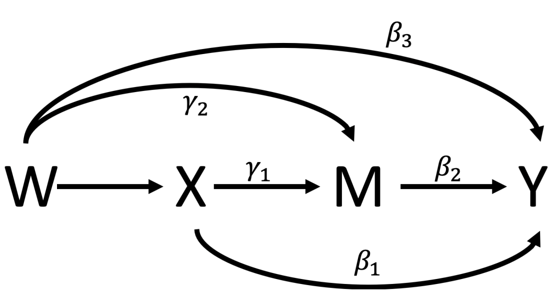

Fig. 1: Mediation directed acyclic graph, where Y, X, M and $\boldsymbol{W}$ denote the outcome, exposure, mediator, and confounders of the exposure-outcome and exposure-mediator relationships.

Tab. 2: Expressions of mediation measures under four different datatypes of the outcome and mediator. Case #1, continuous outcome and continuous mediator; Case #2, continuous outcome and binary mediator; Case #3, binary outcome and continuous mediator; and Case #4, binary outcome and binary mediator.

Datatypes		NIE	NDE	Depend on W	Reference(s)
Case #1		$\beta_2\gamma_1(x-x^*)$	$\beta_1(x-x^*)$	No	Valeri and VanderWeele (2013)
Case #2		$\beta_2\left(\frac{e^{\gamma_0+\gamma_1 x+\boldsymbol{\gamma}_2^T\boldsymbol{w}}}{1+e^{\gamma_0+\gamma_1 x+\boldsymbol{\gamma}_2\boldsymbol{w}}}-\frac{e^{\gamma_0+\gamma_1 x^*+\boldsymbol{\gamma}_2^T\boldsymbol{w}}}{1+e^{\gamma_0+\gamma_1 x^*+\boldsymbol{\gamma}_2^T\boldsymbol{w}}}\right)$	$\beta_1(x-x^*)$	Yes	Barfield et al. (2017)
Case #3	Approx.	$\beta_2\gamma_1(x-x^*)$	$\beta_1(x-x^*)$	No	VanderWeele and Vansteelandt (2010)
	Exact	$\log\left(\frac{\int_m \frac{\exp(-\frac{m^2-2(\sigma^2\beta_2+\gamma_0+\gamma_1 x+\boldsymbol{\gamma}_2^T\boldsymbol{w})m}{2\sigma^2})}{1+\exp(\beta_0+\beta_1 x+\beta_2 m+\boldsymbol{\beta}_3^T\boldsymbol{w})}\mathbf{d}m}{\int_m \frac{\exp(-\frac{m^2-2(\gamma_0+\gamma_1 x+\boldsymbol{\gamma}_2^T\boldsymbol{w})m}{2\sigma^2})}{1+\exp(\beta_0+\beta_1 x+\beta_2 m+\boldsymbol{\beta}_3^T\boldsymbol{w})}\mathbf{d}m}\right)-\log\left(\frac{\int_m \frac{\exp(-\frac{m^2-2(\sigma^2\beta_2+\gamma_0+\gamma_1 x^*+\boldsymbol{\gamma}_2^T\boldsymbol{w})m}{2\sigma^2})}{1+\exp(\beta_0+\beta_1 x+\beta_2 m+\boldsymbol{\beta}_3^T\boldsymbol{w})}\mathbf{d}m}{\int_m \frac{\exp(-\frac{m^2-2(\gamma_0+\gamma_1 x^*+\boldsymbol{\gamma}_2^T\boldsymbol{w})m}{2\sigma^2})}{1+\exp(\beta_0+\beta_1 x+\beta_2 m+\boldsymbol{\beta}_3^T\boldsymbol{w})}\mathbf{d}m}\right)$	$\beta_1(x-x^*)+\log\left(\frac{\int_m \frac{\exp(-\frac{m^2-2(\sigma^2\beta_2+\gamma_0+\gamma_1 x^*+\boldsymbol{\gamma}_2^T\boldsymbol{w})m}{2\sigma^2})}{1+\exp(\beta_0+\beta_1 x+\beta_2 m+\boldsymbol{\beta}_3^T\boldsymbol{w})}\mathbf{d}m}{\int_m \frac{\exp(-\frac{m^2-2(\gamma_0+\gamma_1 x^*+\boldsymbol{\gamma}_2^T\boldsymbol{w})m}{2\sigma^2})}{1+\exp(\beta_0+\beta_1 x+\beta_2 m+\boldsymbol{\beta}_3^T\boldsymbol{w})}\mathbf{d}m}\right)-\log\left(\frac{\int_m \frac{\exp(-\frac{m^2-2(\sigma^2\beta_2+\gamma_0+\gamma_1 x^*+\boldsymbol{\gamma}_2^T\boldsymbol{w})m}{2\sigma^2})}{1+\exp(\beta_0+\beta_1 x^*+\beta_2 m+\boldsymbol{\beta}_3^T\boldsymbol{w})}\mathbf{d}m}{\int_m \frac{\exp(-\frac{m^2-2(\gamma_0+\gamma_1 x^*+\boldsymbol{\gamma}_2^T w)m}{2\sigma^2})}{1+\exp(\beta_0+\beta_1 x^*+\beta_2 m+\boldsymbol{\beta}_3^T\boldsymbol{w})}\mathbf{d}m}\right)$	Yes	Supplementary Material Appendix B
	Probit Approx.	$\mathbf{logit}\left(\Phi\left\{\frac{s\beta_0+s\beta_1 x+s\beta_3^T\boldsymbol{w}+s\beta_2(\gamma_0+\gamma_1 x+\gamma_2^T\boldsymbol{w})}{\sqrt{1+s^2\beta_2^2\sigma^2}}\right\}\right)-\mathbf{logit}\left(\Phi\left\{\frac{s\beta_0+s\beta_1 x+s\beta_3^T\boldsymbol{w}+s\beta_2(\gamma_0+\gamma_1 x^*+\gamma_2^T\boldsymbol{w})}{\sqrt{1+s^2\beta_2^2\sigma^2}}\right\}\right)$	$\mathbf{logit}\left(\Phi\left\{\frac{s\beta_0+s\beta_1 x+s\beta_3^T\boldsymbol{w}+s\beta_2(\gamma_0+\gamma_1 x^*+\gamma_2^T\boldsymbol{w})}{\sqrt{1+s^2\beta_2^2\sigma^2}}\right\}\right)-\mathbf{logit}\left(\Phi\left\{\frac{s\beta_0+s\beta_1 x^*+s\beta_3^T\boldsymbol{w}+s\beta_2(\gamma_0+\gamma_1 x^*+\gamma_2^T\boldsymbol{w})}{\sqrt{1+s^2\beta_2^2\sigma^2}}\right\}\right)$	Yes	Gaynor et al. (2019)
Case #4	Approx.	$\log\left(\frac{(1+e^{\gamma_0+\gamma_1 x^*+\boldsymbol{\gamma}_2^T\boldsymbol{w}})(1+e^{\beta_2+\gamma_0+\gamma_1 x+\boldsymbol{\gamma}_2^T\boldsymbol{w}})}{(1+e^{\gamma_0+\gamma_1 x+\boldsymbol{\gamma}_2^T\boldsymbol{w}})(1+e^{\beta_2+\gamma_0+\gamma_1 x^*+\boldsymbol{\gamma}_2^T\boldsymbol{w}})}\right)$	$\beta_1(x-x^*)$	Yes	VanderWeele (2015)
	Exact	$\log\left(\frac{1+e^{\beta_0+\beta_1 x+\beta_2+\boldsymbol{\beta}_3^T w}+e^{\gamma_0+\gamma_1 x^*+\boldsymbol{\gamma}_2^T w}(1+e^{\beta_0+\beta_1 x+\boldsymbol{\beta}_3^T w})}{1+e^{\beta_0+\beta_1 x+\beta_2+\boldsymbol{\beta}_3^T w}+e^{\gamma_0+\gamma_1 x+\boldsymbol{\gamma}_2^T w}(1+e^{\beta_0+\beta_1 x+\boldsymbol{\beta}_3^T w})}\right)+\log\left(\frac{1+e^{\beta_0+\beta_1 x+\beta_2+\boldsymbol{\beta}_3^T\boldsymbol{w}}+e^{\beta_2+\gamma_0+\gamma_1 x+\boldsymbol{\gamma}_2^T\boldsymbol{w}}(1+e^{\beta_0+\beta_1 x+\boldsymbol{\beta}_3^T\boldsymbol{w}})}{1+e^{\beta_0+\beta_1 x+\beta_2+\boldsymbol{\beta}_3^T\boldsymbol{w}}+e^{\beta_2+\gamma_0+\gamma_1 x^*+\boldsymbol{\gamma}_2\boldsymbol{w}}(1+e^{\beta_0+\beta_1 x+\boldsymbol{\beta}_3^T\boldsymbol{w}})}\right)$	$\beta_1(x-x^*)+\log\left(\frac{1+e^{\beta_0+\beta_1 x^*+\beta_2+\boldsymbol{\beta}_3^T w}+e^{\gamma_0+\gamma_1 x^*+\boldsymbol{\gamma}_2^T w}(1+e^{\beta_0+\beta_1 x^*+\boldsymbol{\beta}_3^T w})}{1+e^{\beta_0+\beta_1 x+\beta_2+\boldsymbol{\beta}_3^T w}+e^{\gamma_0+\gamma_1 x^*+\boldsymbol{\gamma}_2^T w}(1+e^{\beta_0+\beta_1 x+\boldsymbol{\beta}_3^T w})}\right)+\log\left(\frac{1+e^{\beta_0+\beta_1 x+\beta_2+\boldsymbol{\beta}_3^T\boldsymbol{w}}+e^{\beta_2+\gamma_0+\gamma_1 x^*+\boldsymbol{\gamma}_2^T\boldsymbol{w}}(1+e^{\beta_0+\beta_1 x+\boldsymbol{\beta}_3^T\boldsymbol{w}})}{1+e^{\beta_0+\beta_1 x^*+\beta_2+\boldsymbol{\beta}_3^T\boldsymbol{w}}+e^{\beta_2+\gamma_0+\gamma_1 x^*+\boldsymbol{\gamma}_2^T\boldsymbol{w}}(1+e^{\beta_0+\beta_1 x^*+\boldsymbol{\beta}_3^T\boldsymbol{w}})}\right)$	Yes	Gaynor et al. (2019) and Supplementary Material Appendix C

Note: NIE and NDE denote the natural indirect effect and natural direct effect, respectively, which are defined for X in change from x^* to x conditional on $\boldsymbol{W}=\boldsymbol{w}$, on an identity scale in Cases #1 and #2 and a log risk ratio scale in Cases #3 and #4. Given NIE and NDE, the mediation proportion (MP) can be obtained by $\frac{\text{NIE}}{\text{NIE+NDE}}$. Note that the NIE and MP does not depend on the level of $\boldsymbol{W}$ with Case #1 and the approximate method in Case #3. In the probit approximation method, $s=1/1.6$, $\text{logit}(x)=\log(\frac{x}{1-x})$, and $\Phi(.)$ is the cumulative density function for the standard normal distribution.

Tab. 3: Simulation results for Case #1: continuous outcome and continuous mediator.

			$\widehat{\text{NIE}}$				$\widehat{\text{MP}}$			
N	**MP**	**TE**	**Bias(%)**	**CR$^{(d)}$**	**CR$^{(b)}$**	**VR**	**Bias(%)**	**CR$^{(d)}$**	**CR$^{(b)}$**	**VR**
	0.05	0.25	-16.4	**99.2**	**96.4**	0.998	-27.5	**99.2**	**98.8**	0.000
		0.5	-15.1	**96.5**	**95.7**	0.992	-14.1	**96.9**	**96.5**	0.058
		1	-10.5	**92.2**	94.7	0.993	-9.8	**92.6**	95.1	0.949
	0.2	0.25	-10.5	**92.2**	94.7	0.993	-19.7	**90.6**	**97.3**	0.000
150		0.5	-6.2	**92.5**	94.8	1.001	-2.8	**92.9**	**96.4**	0.092
		1	-2.1	94.8	94.9	1.012	-1.7	**95.7**	95.4	0.944
	0.5	0.25	-4.7	**93.2**	94.8	1.007	-14.1	**89.1**	**97.0**	0.000
		0.5	-1.6	95.1	95.0	1.012	-0.4	**96.0**	**97.0**	0.026
		1	-0.1	95.5	94.9	1.013	0.2	**97.5**	95.3	0.710
	0.05	0.25	-4.4	**96.0**	94.4	0.997	-3.5	**97.8**	**96.6**	0.002
		0.5	-3.5	94.7	**94.2**	0.988	-3.4	95.3	94.5	0.855
		1	-2.6	**93.1**	**94.2**	0.983	-2.8	**93.4**	**94.4**	0.948
	0.2	0.25	-2.6	**93.1**	**94.2**	0.983	-1.5	**92.1**	**97.0**	0.001
500		0.5	-1.5	**93.6**	94.5	0.982	-0.8	**94.4**	94.8	0.801
		1	-0.2	94.5	94.5	0.981	-0.1	94.7	94.6	0.945
	0.5	0.25	-0.9	**94.0**	94.5	0.980	-0.4	**92.3**	**97.5**	0.000
		0.5	-0.1	94.7	**94.3**	0.982	0.1	95.5	94.9	0.770
		1	0.6	94.5	94.5	0.983	0.0	95.6	94.7	0.915
	0.05	0.25	-4.0	95.4	94.7	0.969	-3.2	**97.0**	95.4	0.724
		0.5	-2.2	94.8	94.7	0.969	-2.8	95.2	95.1	0.924
		1	-2.1	**94.1**	94.8	0.971	-1.9	**94.0**	94.7	0.967
	0.2	0.25	-2.1	**94.1**	94.8	0.971	-1.1	**93.5**	**96.3**	0.580
1000		0.5	-1.1	**94.2**	94.6	0.981	-1.2	**94.2**	94.8	0.923
		1	-0.9	94.9	94.8	0.986	-0.8	94.6	94.4	0.981
	0.5	0.25	-1.1	94.4	94.6	0.983	-0.5	**92.8**	**96.6**	0.441
		0.5	-0.5	95.0	94.9	0.987	-0.5	94.9	94.6	0.916
		1	-0.3	94.8	94.7	0.988	-0.4	95.0	94.5	0.966
	0.05	0.25	-0.5	94.8	94.8	0.990	0.3	95.5	94.8	0.961
		0.5	-0.2	94.9	95.0	0.995	-0.3	95.4	95.0	0.990
		1	-0.2	95.0	95.1	1.007	-0.3	95.0	95.1	1.006
	0.2	0.25	-0.2	95.0	95.1	1.007	0.1	95.6	95.1	0.936
5000		0.5	-0.1	95.0	95.2	1.027	-0.1	95.1	95.3	1.004
		1	0.0	95.3	95.2	1.038	0.0	95.1	95.1	1.031
	0.5	0.25	0.0	95.3	95.5	1.032	0.1	95.0	94.7	0.920
		0.5	-0.1	95.3	95.2	1.040	0.0	95.2	95.0	0.995
		1	0.0	95.4	95.2	1.041	0.0	95.3	95.4	1.022

Note: Bias(%), CR$^{(d)}$, CR$^{(b)}$, and VR denote the median percent bias, 95% confidence interval coverage rate of multivariate delta method, 95% confidence interval coverage rate of percentile bootstrap method, and mediation variance ratio, respectively. The coverage rates outside the 95% confidence boundary, i.e., $q \pm 1.96 \times \sqrt{\frac{q(1-q)}{B}}$, were highlighted in bold, where q denotes the nominal confidence interval threshold (95%) and B denotes the number of replication (5,000). The median percent bias was calculated as the median of the ratio of bias to the true value over 5,000 replications, i.e., Bias(%) $= median(\frac{\hat{p}-p}{p}) \times 100\%$, where p denotes the true value of the causal mediation measure, and $\hat{p}$ is the point estimate of the simulated causal mediation measure. The median variance ratio is defined by the ratio of median delta-method variance estimators across 5,000 replications to the empirical variance of causal mediation measure estimates from the 5,000 replications.

Tab. 4: Simulation results for Case #4: binary outcome and binary mediator, where the MP and NIE estimates based on the approximate expressions were applied.

				$\widehat{\text{NIE}}^{(a)}$				$\widehat{\text{MP}}^{(a)}$			
N	**Ncase (%)**	**MP**	**TE**	**Bias(%)**	**CR$^{(d)}$**	**CR$^{(b)}$**	**VR**	**Bias(%)**	**CR$^{(d)}$**	**CR$^{(b)}$**	**VR**
	16 (3.3%)	0.05	log(1.2)	-50.0	**93.1**	**93.7**	0.867	-92.1	**99.5**	**99.4**	0.000
	18 (3.7%)	0.05	log(1.5)	-17.7	**93.9**	**94.1**	0.892	-61.3	**99.2**	**99.3**	0.000
	22 (4.6%)	0.05	log(2)	-6.0	**94.3**	**94.1**	0.899	-22.9	**98.1**	**98.3**	0.000
	17 (3.4%)	0.2	log(1.2)	-10.0	**92.8**	**93.8**	0.868	-88.8	**89.9**	**97.5**	0.000
500	20 (4.0%)	0.2	log(1.5)	-3.5	**93.2**	94.5	0.912	-50.5	**90.3**	**97.3**	0.000
	26 (5.2%)	0.2	log(2)	-2.8	**94.0**	94.9	0.940	-13.9	**90.1**	**96.8**	0.000
	18 (3.6%)	0.5	log(1.2)	-5.5	**92.6**	**94.3**	0.908	-84.8	**70.8**	**94.2**	0.000
	24 (4.8%)	0.5	log(1.5)	-1.7	**93.1**	94.9	0.949	-37.1	**81.9**	**95.7**	0.000
	37 (7.5%)	0.5	log(2)	-0.5	**93.9**	95.4	0.998	-6.1	**88.7**	**96.6**	0.000
	33 (3.3%)	0.05	log(1.2)	-7.5	**94.2**	94.9	0.922	-89.5	**99.4**	**99.4**	0.000
	37 (3.7%)	0.05	log(1.5)	-2.5	**94.3**	94.7	0.935	-32.6	**98.8**	**98.7**	0.000
	45 (4.6%)	0.05	log(2)	-4.4	**94.4**	94.7	0.959	-5.5	**97.3**	**97.0**	0.010
	34 (3.4%)	0.2	log(1.2)	0.8	**94.3**	94.6	0.936	-78.4	**91.4**	**97.9**	0.000
1,000	40 (4.0%)	0.2	log(1.5)	-1.5	**94.0**	94.7	0.941	-25.8	**90.9**	**97.2**	0.000
	52 (5.2%)	0.2	log(2)	-0.7	**94.2**	94.8	0.957	-1.8	**91.6**	**96.6**	0.005
	36 (3.6%)	0.5	log(1.2)	-0.7	**94.4**	95.0	0.944	-70.8	**75.6**	95.5	0.000
	48 (4.8%)	0.5	log(1.5)	0.4	**94.2**	95.1	0.970	-14.0	**87.1**	**97.2**	0.000
	75 (7.5%)	0.5	log(2)	0.4	94.8	95.2	0.990	-2.8	**90.4**	**96.7**	0.004
	166 (3.3%)	0.05	log(1.2)	-14.5	95.1	95.2	0.995	-42.4	**99.3**	**98.2**	0.002
	190 (3.7%)	0.05	log(1.5)	-3.4	94.9	95.1	0.996	-4.7	**97.7**	**96.9**	0.011
	229 (4.6%)	0.05	log(2)	-2.2	94.9	95.0	0.998	-1.6	**96.3**	95.4	0.890
	171 (3.4%)	0.2	log(1.2)	-4.1	94.8	95.1	0.991	-31.3	**92.4**	**97.8**	0.001
5,000	204 (4.0%)	0.2	log(1.5)	-0.9	95.1	95.2	0.999	-1.2	**93.5**	**97.5**	0.049
	261 (5.2%)	0.2	log(2)	0.5	95.0	95.1	0.991	0.4	95.2	95.3	0.832
	183 (3.6%)	0.5	log(1.2)	-1.4	94.8	95.1	0.980	-21.8	**85.5**	**97.4**	0.000
	242 (4.8%)	0.5	log(1.5)	0.2	94.9	95.0	0.981	0.5	**92.1**	**97.6**	0.000
	378 (7.5%)	0.5	log(2)	1.0	94.7	94.7	0.979	-1.9	**93.7**	95.1	0.851
	663 (3.3%)	0.05	log(1.2)	-5.1	95.1	\	1.022	-7.6	**98.4**	\	0.000
	759 (3.7%)	0.05	log(1.5)	-1.2	94.8	\	1.008	-1.5	**96.3**	\	0.886
	916 (4.6%)	0.05	log(2)	-0.1	95.3	\	1.016	-0.5	**95.8**	\	0.973
	684 (3.4%)	0.2	log(1.2)	-1.1	94.6	\	0.994	-2.0	**93.1**	\	0.002
20,000	815 (4.0%)	0.2	log(1.5)	0.5	94.7	\	1.007	0.2	95.1	\	0.793
	1043 (5.2%)	0.2	log(2)	0.9	95.1	\	0.999	0.3	95.2	\	0.931
	731 (3.6%)	0.5	log(1.2)	0.3	94.7	\	0.996	-0.7	**91.0**	\	0.000
	969 (4.8%)	0.5	log(1.5)	1.0	94.8	\	0.994	0.1	**94.3**	\	0.796
	1514 (7.5%)	0.5	log(2)	1.1	94.7	\	0.995	-1.7	**92.7**	\	0.927

Note: Ncase, Bias(%), CR$^{(d)}$, CR$^{(b)}$, and VR denote the average number of cases under each setting across 5,000 replications, the median percent bias, 95% confidence interval coverage rate of multivariate delta method, 95% confidence interval coverage rate of percentile bootstrap method, and mediation variance ratio, respectively. The coverage rates outside the 95% confidence boundary, i.e., $q \pm 1.96 \times \sqrt{\frac{q(1-q)}{B}}$, were highlighted in bold, where q denotes the nominal confidence interval threshold (95%) and B denotes the number of replication (5,000). The median percent bias was calculated as the median of the ratio of bias to the true value over 5,000 replications, i.e., Bias(%) $= median(\frac{\hat{p}-p}{p}) \times 100\%$, where p denotes the true value of the causal mediation measure, and $\hat{p}$ is the point estimate of the simulated causal mediation measure. The median variance ratio is defined by the ratio of median delta-method variance estimators across 5,000 replications to the empirical variance of causal mediation measure estimates from the 5,000 replications.

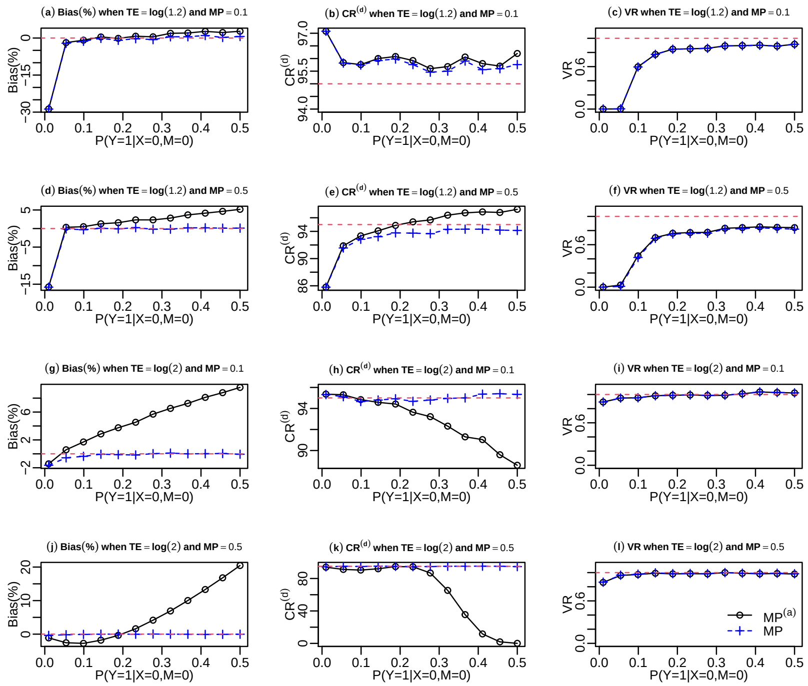

Fig. 2: Performance of MP$^{(a)}$ estimates (black line) and MP estimates (blue dotted line) when changing baseline outcome prevalence from 1% to 50% in Case #4, where sample size is 20,000. Bias(%), CR$^{(d)}$, and VR denote the percent bias, coverage rate by the multivariate delta method, and variance ratio. Upper row: results for TE=log(1.2) and MP=0.1; second row: results for TE=log(1.2) and MP=0.5; third row: results for TE=log(2) and MP=0.1; bottom row: results for TE=log(2) and MP=0.5.

Tab. 5: Mediation analysis of MaxART (Khan et al., 2020). (n=1731)

Expression	Scenario	Parameter	Point	S.E.	Delta 95% CI	Bootstrap 95% CI
Steptime adjusted	Approximate	NIE$^{(a)}$	-0.601	0.091	(-0.779,-0.424)	(-0.782,-0.445)
		TE$^{(a)}$	-1.472	0.226	(-1.915,-1.030)	(-1.973,-1.031)
		MP$^{(a)}$	0.408	0.069	(0.273,0.544)	(0.305,0.559)
	Exact	NIE	-0.630	0.093	(-0.813,-0.448)	(-0.816,-0.474)
		TE	-1.444	0.213	(-1.862,-1.027)	(-1.922,-1.023)
		MP	0.437	0.073	(0.293,0.580)	(0.327,0.597)
Multivariate adjusted	Approximate	NIE$^{(a)}$	-0.972	0.121	(-1.208,-0.735)	(-1.287,-0.775)
		TE$^{(a)}$	-2.520	0.287	(-3.082,-1.958)	(-3.282,-1.975)
		MP$^{(a)}$	0.386	0.050	(0.288,0.483)	(0.292,0.494)
	Exact	NIE	-0.970	0.120	(-1.205,-0.735)	(-1.282,-0.772)
		TE	-2.316	0.267	(-2.841,-1.792)	(-3.033,-1.819)
		MP	0.419	0.054	(0.313,0.525)	(0.322,0.539)

Note: All the mediation measures, including NIE, TE, and MP, are defined on a log odds ratio scale for the intervention in change from SoC to EAAA, conditional on the most frequent level of the confounding variables. S.E. denotes the standard error of the point estimates, which is calculated by the multivariate delta method. We implemented the delta method and bootstrap method with 1,000 replications to calculate the 95% confidence interval (95% CI) of each mediation measure.