Web Material

Is the product method more efficient than the difference method for assessing mediation?

Chao Cheng*, Donna Spiegelman, and Fan Li

Contents

*correspondence to Chao Cheng, MS, Department of Biostatistics, Yale School of Public Health, New Haven, CT, USA. E-mail: c.cheng@yale.edu.

Web Appendix 1: Proofs for the efficiency of the product method compared to the difference method under the four data types

Throughout, we will use $\text{NIE}_{a^*,a|\boldsymbol{c}}$, $\text{NDE}_{a^*,a|\boldsymbol{c}}$, $\text{TE}_{a^*,a|\boldsymbol{c}}$, and $\text{MP}_{a^*,a|\boldsymbol{c}}$ to denote the NIE, NDE, TE, and MP, respectively, defined for the exposure in change from a^* to a conditional on $\boldsymbol{C} = \boldsymbol{c}$. Meanwhile, we use $\widehat{\text{NIE}}^{(p)}_{a^*,a|\boldsymbol{c}}$, $\widehat{\text{NDE}}^{(p)}_{a^*,a|\boldsymbol{c}}$, $\widehat{\text{TE}}^{(p)}_{a^*,a|\boldsymbol{c}}$, and $\widehat{\text{MP}}^{(p)}_{a^*,a|\boldsymbol{c}}$, respectively, to denote estimators of $\text{NIE}_{a^*,a|\boldsymbol{c}}$, $\text{NDE}_{a^*,a|\boldsymbol{c}}$, $\text{TE}_{a^*,a|\boldsymbol{c}}$, and $\text{MP}_{a^*,a|\boldsymbol{c}}$ given by the product method, where the superscript '(p)' stands for the first letter of 'product method'. Similarly, we define $\widehat{\text{NIE}}^{(d)}_{a^*,a|\boldsymbol{c}}$, $\widehat{\text{NDE}}^{(d)}_{a^*,a|\boldsymbol{c}}$, $\widehat{\text{TE}}^{(d)}_{a^*,a|\boldsymbol{c}}$, and $\widehat{\text{MP}}^{(d)}_{a^*,a|\boldsymbol{c}}$ to be the corresponding estimators given by the difference method, where the superscript '(d)' stands for the first letter of 'difference method'.

Appendix 1.1 Proof of result 1

1. The product and difference methods are compatible.

Here and throughout, we say that the product and difference methods are compatible if the regression models under both methods can hold simultaneously. In Case $Y_c M_c$, the outcome regressions and mediator regression are assumed to be

$$\begin{aligned}
&\text{Model I: } \mathbb{E}[Y|A,\boldsymbol{C}] = \beta_0^* + \beta_1^* A + \boldsymbol{\beta}_3^{*T}\boldsymbol{C}, \\
&\text{Model II: } \mathbb{E}[Y|A,M,\boldsymbol{C}] = \beta_0 + \beta_1 A + \beta_2 M + \boldsymbol{\beta}_3^T\boldsymbol{C}, \\
&\text{Model III: } \mathbb{E}[M|A,\boldsymbol{C}] = \gamma_0 + \gamma_1 A + \boldsymbol{\gamma}_2^T\boldsymbol{C},
\end{aligned} \tag{s1}$$

where the difference method involves Models I and II and the product method involves Models II and III. Given Models II and III, we can show

$$\begin{aligned}
\mathbb{E}[Y|A,\boldsymbol{C}] &= \mathbb{E}\left[\mathbb{E}[Y|A,M,\boldsymbol{C}]|A,\boldsymbol{C}\right], \\
&= \beta_0 + \beta_1 A + \beta_2\mathbb{E}[M|A,\boldsymbol{C}] + \boldsymbol{\beta}_3^T\boldsymbol{C}, \\
&= \beta_0 + \beta_2\gamma_0 + (\beta_1 + \beta_2\gamma_1)A + (\boldsymbol{\beta}_3^T + \beta_2\boldsymbol{\gamma}_2^T)\boldsymbol{C}.
\end{aligned}$$

Therefore, Model I holds given Models II and III, and we have parametric relationships, $\beta_0^* = \beta_0 + \beta_2\gamma_0$, $\beta_1^* = \beta_1 + \beta_2\gamma_1$, and $\boldsymbol{\beta}_3^* = \boldsymbol{\beta}_3 + \beta_2\boldsymbol{\gamma}_2$. It follows that the above three regression models are compatible and it is easy to verify the product and difference method provide same mediation measures (e.g., NIE $= \beta_2\gamma_1(a-a^*) = (\beta_1^* - \beta_1)(a-a^*)$ and MP $= \frac{\beta_2\gamma_1}{\beta_2\gamma_1+\beta_1} = 1 - \frac{\beta_1}{\beta_1^*}$).

2. The product method and difference method are algebraically equivalent for estimating mediation measures.

MacKinnon et al. (1995) [4] showed that the product and difference method are algebraically equivalent for estimating mediation measures when $\boldsymbol{C}$ is null. Now, we show that this result extends when $\boldsymbol{C}$ exists.

The proof of MacKinnon et al. (1995) [4] can be easily adapted to the scenario where $\boldsymbol{C}$ exists. In Case Y_cM_c, ordinary least squares (OLS) is used to solve for the unknown regression parameters in Models I to III, separately. We reorganize the three linear regression models as follows

$$\begin{aligned}
&\text{Model I: } \mathbb{E}[Y|\boldsymbol{H}] = \beta_0^* + \bar{\boldsymbol{\beta}}_1^{*T}\boldsymbol{H}, \\
&\text{Model II: } \mathbb{E}[Y|\boldsymbol{H},M] = \beta_0 + \bar{\boldsymbol{\beta}}_1^T\boldsymbol{H} + \beta_2 M, \\
&\text{Model III: } \mathbb{E}[M|\boldsymbol{H}] = \gamma_0 + \bar{\boldsymbol{\gamma}}_1^T\boldsymbol{H},
\end{aligned}$$

where $\boldsymbol{H} = [A, \boldsymbol{C}^T]^T$, $\bar{\boldsymbol{\beta}}_1^* = [\beta_1^*, {\boldsymbol{\beta}_3^*}^T]^T$, $\bar{\boldsymbol{\beta}}_1 = [\beta_1, \boldsymbol{\beta}_3^T]^T$, and $\bar{\boldsymbol{\gamma}}_1 = [\gamma_1, \boldsymbol{\gamma}_2^T]^T$. By Model I, we have

$$\bar{\boldsymbol{\beta}}_1^* = [\mathrm{Var}(\boldsymbol{H})]^{-1}\mathrm{Cov}(\boldsymbol{H}, Y). \tag{s2}$$

Taking covariance of $\boldsymbol{H}$ and Y from Model II, we have

$$\mathrm{Cov}(\boldsymbol{H}, Y) = \mathrm{Cov}(\boldsymbol{H}, \beta_0 + \bar{\boldsymbol{\beta}}_1^T\boldsymbol{H} + \beta_2 M) = \mathrm{Var}(\boldsymbol{H})\bar{\boldsymbol{\beta}}_1 + \beta_2\mathrm{Cov}(\boldsymbol{H}, M). \tag{s3}$$

Substituting (s3) into (s2) leads to

$$\begin{aligned}\bar{\boldsymbol{\beta}}_1^* &= [\mathrm{Var}(\boldsymbol{H})]^{-1}\mathrm{Var}(\boldsymbol{H})\bar{\boldsymbol{\beta}}_1 + [\mathrm{Var}(\boldsymbol{H})]^{-1}\beta_2\mathrm{Cov}(\boldsymbol{H}, M)\\ &= \bar{\boldsymbol{\beta}}_1 + \beta_2[\mathrm{Var}(\boldsymbol{H})]^{-1}\mathrm{Cov}(\boldsymbol{H}, M).\end{aligned}$$

From Model III, we know $\bar{\boldsymbol{\gamma}}_1 = [\mathrm{Var}(\boldsymbol{H})]^{-1}\mathrm{Cov}(\boldsymbol{H}, M)$ and then we can show

$$\bar{\boldsymbol{\beta}}_1^* = \bar{\boldsymbol{\beta}}_1 + \beta_2\bar{\boldsymbol{\gamma}}_1.$$

Replacing $\mathrm{Cov}(\boldsymbol{H}, M)$ and $\mathrm{Var}(\boldsymbol{H})$ with their empirical versions and replacing the unknown regression parameters with OLS estimators, the above identity still holds, i.e., $\widehat{\bar{\boldsymbol{\beta}}}_1^* = \widehat{\bar{\boldsymbol{\beta}}}_1 + \hat{\beta}_2\widehat{\bar{\boldsymbol{\gamma}}}_1$. This indicates $\hat{\beta}_1^* - \hat{\beta}_1 = \hat{\beta}_2\hat{\gamma}_1$. Therefore, $\widehat{\mathrm{NIE}}^{(d)}_{a^*,a|\boldsymbol{c}} = (\hat{\beta}_1^* - \hat{\beta}_1)(a - a^*) = \hat{\beta}_2\hat{\gamma}_1(a - a^*) = \widehat{\mathrm{NIE}}^{(p)}_{a^*,a|\boldsymbol{c}}$ and $\widehat{\mathrm{TE}}^{(d)}_{a^*,a|\boldsymbol{c}} = \hat{\beta}_1^*(a - a^*) = (\hat{\beta}_1 + \hat{\beta}_2\hat{\gamma}_1)(a - a^*) = \widehat{\mathrm{TE}}^{(p)}_{a^*,a|\boldsymbol{c}}$. Also,

$$\hat{\beta}_1^* - \hat{\beta}_1 = \hat{\beta}_2\hat{\gamma}_1 \Longleftrightarrow \frac{\hat{\beta}_1^* - \hat{\beta}_1}{\hat{\beta}_1^*} = \frac{\hat{\beta}_2\hat{\gamma}_1}{\hat{\beta}_1^*} \Longleftrightarrow \frac{\hat{\beta}_1^* - \hat{\beta}_1}{\hat{\beta}_1^*} = \frac{\hat{\beta}_2\hat{\gamma}_1}{\hat{\beta}_1 + \hat{\beta}_2\hat{\gamma}_1},$$

which leads to $\widehat{\mathrm{MP}}^{(p)}_{a^*,a|\boldsymbol{c}} = \widehat{\mathrm{MP}}^{(d)}_{a^*,a|\boldsymbol{c}}$. Finally it is straightforward to show that $\widehat{\mathrm{NDE}}^{(p)}_{a^*,a|\boldsymbol{c}} = \widehat{\mathrm{NDE}}^{(d)}_{a^*,a|\boldsymbol{c}}$ as both have a same expression, $\hat{\beta}_1(a - a^*)$. Therefore, the difference and product methods are algebraically equivalent for estimating mediation measures given $\boldsymbol{C}$.

Appendix 1.2 Proof of result 2

1. The product and difference methods are compatible if and only if A is binary and $\boldsymbol{\gamma}_2 = \mathbf{0}$.

In Cases Y_cM_b, if A is binary and $\boldsymbol{\gamma}_2 = \mathbf{0}$, the three models simplify to

$$\begin{aligned}
&\text{Model I: } \mathbb{E}[Y|A,\boldsymbol{C}] = \beta_0^* + \beta_1^* A + \boldsymbol{\beta}_3^{*T}\boldsymbol{C},\\
&\text{Model II: } \mathbb{E}[Y|A,M,\boldsymbol{C}] = \beta_0 + \beta_1 A + \beta_2 M + \boldsymbol{\beta}_3^T\boldsymbol{C},\\
&\text{Model III: } \mathbb{E}[M|A,\boldsymbol{C}] = \frac{e^{\gamma_0+\gamma_1 A}}{1+e^{\gamma_0+\gamma_1 A}} = \frac{e^{\gamma_0}}{1+e^{\gamma_0}} + \left(\frac{e^{\gamma_0+\gamma_1}}{1+e^{\gamma_0+\gamma_1}} - \frac{e^{\gamma_0}}{1+e^{\gamma_0}}\right)A,
\end{aligned} \tag{s4}$$

and, given Models II and III, we have

$$\begin{aligned}
\mathbb{E}[Y|A,\boldsymbol{C}] &= \mathbb{E}\left[\mathbb{E}[Y|A,M,\boldsymbol{C}]|A,\boldsymbol{C}\right] = \beta_0 + \beta_1 A + \beta_2\mathbb{E}[M|A] + \boldsymbol{\beta}_3^T\boldsymbol{C}\\
&= \beta_0 + \beta_2\frac{e^{\gamma_0}}{1+e^{\gamma_0}} + \left[\beta_1 + \beta_2\left(\frac{e^{\gamma_0+\gamma_1}}{1+e^{\gamma_0+\gamma_1}} - \frac{e^{\gamma_0}}{1+e^{\gamma_0}}\right)\right]A + \boldsymbol{\beta}_3^T\boldsymbol{C}.
\end{aligned}$$

Therefore, above three models are compatible when A is binary and $\boldsymbol{\gamma}_2 = \mathbf{0}$, and we have $\beta_0^* = \beta_0 + \beta_2\frac{e^{\gamma_0}}{1+e^{\gamma_0}}$, $\beta_1^* = \beta_1 + \beta_2\left(\frac{e^{\gamma_0+\gamma_1}}{1+e^{\gamma_0+\gamma_1}} - \frac{e^{\gamma_0}}{1+e^{\gamma_0}}\right)$, and $\boldsymbol{\beta}_3^* = \boldsymbol{\beta}_3$. Again, one can check the product and difference methods share the same mediation measure estimands by substituting the parametric relationships into mediation measure expressions given by the difference method or the product method.

Of note, a binary A and $\boldsymbol{\gamma}_2 \neq \mathbf{0}$ are also necessary conditions to ensure compatibility between the product and difference method. To show this, assuming A has multiple values, Model III, i.e., $\mathbb{E}[M|A,\boldsymbol{C}] = \frac{e^{\gamma_0+\gamma_1 A+\boldsymbol{\gamma}_2^T\boldsymbol{C}}}{1+e^{\gamma_0+\gamma_1 A+\boldsymbol{\gamma}_2^T\boldsymbol{C}}}$, must be a non-linear function with regard to A and then, given Models II and III, $\mathbb{E}[Y|A,\boldsymbol{C}]$ will not have a linear representation with respect to A and $\boldsymbol{C}$. Therefore, A should be binary. If A is binary, then we can show that

$$\mathbb{E}[M|A,\boldsymbol{C}] = \frac{e^{\gamma_0+\boldsymbol{\gamma}_2^T\boldsymbol{C}}}{1+e^{\gamma_0+\boldsymbol{\gamma}_2^T\boldsymbol{C}}} + \left(\frac{e^{\gamma_0+\gamma_1+\boldsymbol{\gamma}_2^T\boldsymbol{C}}}{1+e^{\gamma_0+\gamma_1+\boldsymbol{\gamma}_2^T\boldsymbol{C}}} - \frac{e^{\gamma_0+\boldsymbol{\gamma}_2^T\boldsymbol{C}}}{1+e^{\gamma_0+\boldsymbol{\gamma}_2^T\boldsymbol{C}}}\right)A$$

According to the previous formula, it can be straightforwardly observed that $\mathbb{E}[M|A,\boldsymbol{C}]$ will always have an interaction term between A and $\boldsymbol{C}$ if $\boldsymbol{\gamma}_2 \neq \mathbf{0}$ such that $\mathbb{E}[M|A,\boldsymbol{C}]$ (and also $\mathbb{E}[Y|A,\boldsymbol{C}]$) cannot be presented as a linear function of A and $\boldsymbol{C}$. This concludes that Models I, II, and III are not compatible if either A is continuous or $\boldsymbol{\gamma}_2 \neq \mathbf{0}$.

2. When the error term in the conditional outcome model (2) follows a homoscedastic normal distribution, then the product method yields maximum likelihood estimation (MLE) and is asymptotically at least as efficient as the difference method.

Let $\{Y_i, A_i, M_i, \boldsymbol{C}_i\}_{i=1}^n$ be n observations of $\{Y, A, \boldsymbol{M}, \boldsymbol{C}\}$. The log-likelihood function for $\{Y_i, A_i, M_i, \boldsymbol{C}_i\}_{i=1}^n$ is

$$\begin{aligned}\log L(\boldsymbol{\beta}, \boldsymbol{\gamma}; Y, A, M, \boldsymbol{C}) &= \sum_{i=1}^n \log\{P_{\boldsymbol{\beta},\boldsymbol{\gamma}}(Y_i, A_i, M_i, \boldsymbol{C}_i)\} \\ &= \sum_{i=1}^n \log\{P_{\boldsymbol{\beta},\boldsymbol{\gamma}}(Y_i|A_i, M_i, \boldsymbol{C}_i) P_{\boldsymbol{\beta},\boldsymbol{\gamma}}(M_i|A_i, \boldsymbol{C}_i) P_{\boldsymbol{\beta},\boldsymbol{\gamma}}(A_i, \boldsymbol{C}_i)\} \\ &= \sum_{i=1}^n \log P_{\boldsymbol{\beta},\boldsymbol{\gamma}}(Y_i|A_i, M_i, \boldsymbol{C}_i) + \sum_{i=1}^n \log P_{\boldsymbol{\beta},\boldsymbol{\gamma}}(M_i|A_i, \boldsymbol{C}_i) + \sum_{i=1}^n \log P_{\boldsymbol{\beta},\boldsymbol{\gamma}}(A_i, \boldsymbol{C}_i),\end{aligned} \tag{s5}$$

where $P_{\boldsymbol{\beta},\boldsymbol{\gamma}}(Y_i|A_i, M_i, \boldsymbol{C}_i)$, $P_{\boldsymbol{\beta},\boldsymbol{\gamma}}(M_i|A_i, \boldsymbol{C}_i)$ and $P_{\boldsymbol{\beta},\boldsymbol{\gamma}}(A_i, \boldsymbol{C}_i)$ are density functions that may depend on unknown parameter $\{\boldsymbol{\beta}, \boldsymbol{\gamma}\}$. The joint distribution $\{A_i, \boldsymbol{C}_i\}$ does not depend on any unknown parameters, therefore $P_{\boldsymbol{\beta},\boldsymbol{\gamma}}(A_i, \boldsymbol{C}_i)$ simplifies to $P(A_i, \boldsymbol{C}_i)$. If we assume ϵ_2, or equivalently $Y_i|A_i, M_i, \boldsymbol{C}_i$, follows a homoscedastic normal distribution, we have

$$\sum_{i=1}^n \log P_{\boldsymbol{\beta},\boldsymbol{\gamma}}(Y_i|A_i, M_i, \boldsymbol{C}_i) \propto -\sum_{i=1}^n \left(Y_i - (\beta_0 + \beta_1 A_i + \beta_2 M_i + \boldsymbol{\beta}_3^T \boldsymbol{C}_i)\right)^2,$$

which implies that $P_{\boldsymbol{\beta},\boldsymbol{\gamma}}(Y_i|A_i, M_i, \boldsymbol{C}_i)$ is only a function of $\boldsymbol{\beta}$; i.e., $\sum_{i=1}^n \log P_{\boldsymbol{\beta},\boldsymbol{\gamma}}(Y_i|A_i, M_i, \boldsymbol{C}_i) = \sum_{i=1}^n \log P_{\boldsymbol{\beta}}(Y_i|A_i, M_i, \boldsymbol{C}_i)$. Furthermore, the OLS algorithm used to estimate the parameter of Model II are maximizing $\sum_{i=1}^n \log P_{\boldsymbol{\beta}}(Y_i|A_i, M_i, \boldsymbol{C}_i)$. Similarly, we have $\sum_{i=1}^n \log P_{\boldsymbol{\beta},\boldsymbol{\gamma}}(M_i|A_i, \boldsymbol{C}_i)$ only depends on $\boldsymbol{\gamma}$ and solving the logistic regression model III is equivalent to maximizing the log-likelihood function $\sum_{i=1}^n \log P_{\boldsymbol{\gamma}}(M_i|A_i, \boldsymbol{C}_i)$.

Following the above discussion, formula (s5) can be simplified to

$$\log L(\boldsymbol{\beta}, \boldsymbol{\gamma}; Y, A, M, \boldsymbol{C}) = \sum_{i=1}^n \log P_{\boldsymbol{\beta}}(Y_i|A_i, M_i, \boldsymbol{C}_i) + \sum_{i=1}^n \log P_{\boldsymbol{\gamma}}(M_i|A_i, \boldsymbol{C}_i) + \sum_{i=1}^n \log P(A_i, \boldsymbol{C}_i).$$

Solving Model II and III, separately, maximizes $\sum_{i=1}^n \log P_{\boldsymbol{\beta}}(Y_i|A_i, M_i, \boldsymbol{C}_i)$ and $\sum_{i=1}^n \log P_{\boldsymbol{\gamma}}(M_i|A_i, \boldsymbol{C}_i)$, respectively. Because $\sum_{i=1}^{n} \log P(A_i, \boldsymbol{C}_i)$ is not a function of $\boldsymbol{\beta}$ or $\boldsymbol{\gamma}$, solving Models II and III separately maximizes the full likelihood function $\log L(\boldsymbol{\beta}, \boldsymbol{\gamma}; Y, A, M, \boldsymbol{C})$. This indicates that the product method provides the maximum likelihood estimation (MLE) of the regression parameters, $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$. Noticing that mediation measures are functions of the above regression parameters, we have that the product method coincides with the MLE for estimating mediation measures by the invariance property of MLE and therefore is asymptotically efficient for estimating mediation measures, and $\text{Avar}(\hat{\theta}^{(p)}) \leq \text{Avar}(\hat{\theta}^{(d)})$, where θ denotes a mediation measure, e.g., NIE, and $\hat{\theta}^{(p)}$ and $\hat{\theta}^{(d)}$ are the point estimators obtained by the product and difference method, respectively.

3. If there are no confounders in the mediation analysis, i.e., $\boldsymbol{C}$ is null, then the product method and difference method are algebraically equivalent for estimating mediation measures.

When $\boldsymbol{C}$ is null, $\hat{\boldsymbol{\beta}}^* = [\hat{\beta}_0^*, \hat{\beta}_1^*]^T$ and $\hat{\boldsymbol{\beta}} = [\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2]^T$ are obtained by OLS method and $\hat{\boldsymbol{\gamma}} = [\hat{\gamma}_0, \hat{\gamma}_1]^T$ is obtained by solving the logistic regression given by Model III. Since A is binary and $\boldsymbol{C}$ is null, Model III becomes a saturated model, which indicate that the model-based estimators for $P(M = 1|A)$ equal to their corresponding nonparametric empirical estimators; that is, $\frac{e^{\hat{\gamma}_0+\hat{\gamma}_1}}{1+e^{\hat{\gamma}_0+\hat{\gamma}_1}} = \hat{P}(M = 1|A = 1)$ and $\frac{e^{\hat{\gamma}_0}}{1+e^{\hat{\gamma}_0}} = \hat{P}(M = 1|A = 0)$, where $\hat{P}(M = 1|A = a)$ denoted the empirical probabilities of $P(M = 1|A = a)$ for $x = 0, 1$. Based on discussions in Web Appendix 1.1, we know the following identity holds if outcome models are based on linear regression,

$$\hat{\beta}_1^* = \hat{\beta}_1 + \hat{\beta}_2 \frac{\widehat{\text{Cov}}(A, M)}{\widehat{\text{Var}}(A)},$$

where $\widehat{\text{Cov}}(A, M)$ and $\widehat{\text{Var}}(A)$ denote the empirical covariance and variance based on the dataset. Because both A and M are binary, we have

$$\begin{aligned}
\frac{\widehat{\text{Cov}}(A, M)}{\widehat{\text{Var}}(A)} &= \frac{\widehat{\mathbb{E}}[AM] - \widehat{\mathbb{E}}[A]\widehat{\mathbb{E}}[M]}{\hat{P}(A=1)\left\{1 - \hat{P}(A=1)\right\}} = \frac{\hat{P}(A=1, M=1) - \hat{P}(A=1)\hat{P}(M=1)}{\hat{P}(A=1)\hat{P}(A=0)} \\
&= \frac{\hat{P}(M=1|A=1)\hat{P}(A=1) - \hat{P}(A=1)\hat{P}(M=1)}{\hat{P}(A=1)\hat{P}(A=0)} = \frac{\hat{P}(M=1|A=1) - \hat{P}(M=1)}{\hat{P}(A=0)} \\
&= \frac{\hat{P}(M=1|A=1) - \hat{P}(M=1|A=1)\hat{P}(A=1) - \hat{P}(M=1|A=0)\hat{P}(A=0)}{\hat{P}(A=0)} \\
&= \hat{P}(M=1|A=1) - \hat{P}(M=1|A=0) \\
&= \frac{e^{\hat{\gamma}_0+\hat{\gamma}_1}}{1+e^{\hat{\gamma}_0+\hat{\gamma}_1}} - \frac{e^{\hat{\gamma}_0}}{1+e^{\hat{\gamma}_0}}.
\end{aligned}$$

Therefore, we have

$$\hat{\beta}_1^* = \hat{\beta}_1 + \hat{\beta}_2\left(\frac{e^{\hat{\gamma}_0+\hat{\gamma}_1}}{1+e^{\hat{\gamma}_0+\hat{\gamma}_1}} - \frac{e^{\hat{\gamma}_0}}{1+e^{\hat{\gamma}_0}}\right),$$

i.e., $\widehat{\text{TE}}^{(p)}_{a^*,a|\boldsymbol{c}} = \widehat{\text{TE}}^{(d)}_{a^*,a|\boldsymbol{c}}$. In addition, the previous formula indicates $\hat{\beta}_1^* - \hat{\beta}_1 = \hat{\beta}_2\left(\frac{e^{\hat{\gamma}_0+\hat{\gamma}_1}}{1+e^{\hat{\gamma}_0+\hat{\gamma}_1}} - \frac{e^{\hat{\gamma}_0}}{1+e^{\hat{\gamma}_0}}\right)$ and $\frac{\hat{\beta}_1^*-\hat{\beta}_1}{\hat{\beta}_1^*} = \frac{\hat{\beta}_2\left(\frac{e^{\hat{\gamma}_0+\hat{\gamma}_1}}{1+e^{\hat{\gamma}_0+\hat{\gamma}_1}} - \frac{e^{\hat{\gamma}_0}}{1+e^{\hat{\gamma}_0}}\right)}{\hat{\beta}_1+\hat{\beta}_2\left(\frac{e^{\hat{\gamma}_0+\hat{\gamma}_1}}{1+e^{\hat{\gamma}_0+\hat{\gamma}_1}} - \frac{e^{\hat{\gamma}_0}}{1+e^{\hat{\gamma}_0}}\right)}$, i.e., $\widehat{\text{NIE}}^{(p)}_{a^*,a|\boldsymbol{c}} = \widehat{\text{NIE}}^{(d)}_{a^*,a|\boldsymbol{c}}$ and $\widehat{\text{MP}}^{(p)}_{a^*,a|\boldsymbol{c}} = \widehat{\text{MP}}^{(d)}_{a^*,a|\boldsymbol{c}}$. Finally it is straightforward to show that $\widehat{\text{NDE}}^{(p)}_{a^*,a|\boldsymbol{c}} = \widehat{\text{NDE}}^{(d)}_{a^*,a|\boldsymbol{c}}$ as both have the same expression, $\hat{\beta}_1(a - a^*)$. Therefore, the difference method and product method are algebraically equivalent for estimating mediation measures when there is no confounding.

Appendix 1.3 Proof of result 3

1. The product and difference methods are compatible.

In Case Y_bM_c , the three models become

$$\begin{aligned}
&\text{Model I: } \mathbb{E}[Y|A, \boldsymbol{C}] = e^{\beta_0^*+\beta_1^*A+\boldsymbol{\beta}_3^{*T}\boldsymbol{C}}, \\
&\text{Model II: } \mathbb{E}[Y|A, M, \boldsymbol{C}] = e^{\beta_0+\beta_1A+\beta_2M+\boldsymbol{\beta}_3^T\boldsymbol{C}}, \qquad \text{(s6)} \\
&\text{Model III: } M = \gamma_0 + \gamma_1A + \boldsymbol{\gamma}_2^T\boldsymbol{C} + \epsilon_3,
\end{aligned}$$

where $\epsilon_3 \sim N(0, \sigma_3^2)$. Given Models II and III, we have

$$\begin{aligned}\mathbb{E}[Y|A,\boldsymbol{C}] = \mathbb{E}\left[\mathbb{E}[Y|A,M,\boldsymbol{C}]|A,\boldsymbol{C}\right] &= e^{\beta_0+\beta_1 A+\boldsymbol{\beta}_3^T\boldsymbol{C}}\mathbb{E}[e^{\beta_2 M}|A,\boldsymbol{C}] \\ &= e^{\beta_0+\beta_1 A+\boldsymbol{\beta}_3^T\boldsymbol{C}}\int e^{\beta_2 M}\frac{1}{\sqrt{2\pi\sigma_3^2}}e^{-\frac{\left(M-(\gamma_0+\gamma_1 A+\boldsymbol{\gamma}_2^T\boldsymbol{C})\right)^2}{2\sigma_3^2}}\mathrm{d}M \\ &= e^{\beta_0+\beta_2\gamma_0+0.5\sigma^2\beta_2^2+(\beta_1+\beta_2\gamma_1)A+(\boldsymbol{\beta}_3^T+\beta_2\boldsymbol{\gamma}_2^T)\boldsymbol{C}},\end{aligned}$$

It follows that the three models in (s6) are compatible and $\beta_0^* = \beta_0+\beta_2\gamma_0+0.5\sigma_3^2\beta_2^2$, $\beta_1^* = \beta_1+\beta_2\gamma_1$, and $\boldsymbol{\beta}_3^* = \boldsymbol{\beta}_3 + \beta_2\boldsymbol{\gamma}_2$.

2. The product method yields maximum likelihood estimation (MLE) and is asymptotically at least as efficient as the difference method.

Similar with Web Appendix 1.2, we can decompose the log-likelihood function for $\{Y_i, A_i, M_i, \boldsymbol{C}_i\}_{i=1}^n$ as

$$\log L(\boldsymbol{\beta},\boldsymbol{\gamma};Y,A,M,\boldsymbol{C}) = \sum_{i=1}^n \log P_{\boldsymbol{\beta}}(Y_i|A_i,M_i,\boldsymbol{C}_i) + \sum_{i=1}^n \log P_{\boldsymbol{\gamma}}(M_i|A_i,\boldsymbol{C}_i) + \sum_{i=1}^n \log P(A_i,\boldsymbol{C}_i),$$

where $P_{\boldsymbol{\beta}}(Y_i|A_i,M_i,\boldsymbol{C}_i)$ and $P_{\boldsymbol{\gamma}}(M_i|A_i,\boldsymbol{C}_i)$ are a log-binomial and normal density function that depends on unknown parameters $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$, respectively. Solving the log-binomial model II maximizes the log-likelihood function $\sum_{i=1}^n \log P_{\boldsymbol{\beta}}(Y_i|A_i,M_i,\boldsymbol{C}_i)$. Also, because $\epsilon_3 \sim N(0,\sigma_3^2)$, we have

$$\sum_{i=1}^n \log P_{\boldsymbol{\gamma}}(M_i|A_i,\boldsymbol{C}_i) \propto -\sum_{i=1}^n \left(M_i - (\gamma_0+\gamma_1 A_i+\boldsymbol{\gamma}_2^T\boldsymbol{C}_i)\right)^2,$$

and therefore the OLS estimator that solves linear model III maximizes $\sum_{i=1}^n \log P_{\boldsymbol{\gamma}}(M_i|A_i,\boldsymbol{C}_i)$. Noting that $\sum_{i=1}^n \log P(A_i,\boldsymbol{C}_i)$ is not a function of $\boldsymbol{\beta}$ or $\boldsymbol{\gamma}$, solving Models II and III maximizes the full log-likelihood function $\log L(\boldsymbol{\beta},\boldsymbol{\gamma};Y,A,M,\boldsymbol{C})$. Therefore the product method yields MLE and therefore is asymptotically efficient for estimating mediation measures, which implies that $\mathrm{Avar}(\hat{\theta}^{(p)}) \leq \mathrm{Avar}(\hat{\theta}^{(d)})$, where θ denote the mediation measure. Therefore, the product method is asymptotically at least as efficient as the difference method. This completes the proof.

Appendix 1.4 Proof of result 4

1. The difference method and product method are compatible if and only if A is binary and $\boldsymbol{\gamma}_2 = \mathbf{0}$.

Similar to Case Y_cM_b, if A is binary and $\boldsymbol{\gamma}_2 = \mathbf{0}$, the three models are compatible. Specifically, the three models become

$$\begin{aligned}
&\text{Model I: } \mathbb{E}[Y|A,\boldsymbol{C}] = e^{\beta_0^* + \beta_1^* A + \boldsymbol{\beta}_3^{*T}\boldsymbol{C}},\\
&\text{Model II: } \mathbb{E}[Y|A,M,\boldsymbol{C}] = e^{\beta_0 + \beta_1 A + \beta_2 M + \boldsymbol{\beta}_3^T\boldsymbol{C}},\\
&\text{Model III: } \mathbb{E}[M|A,\boldsymbol{C}] = \mathbb{E}[M|A] = \frac{e^{\gamma_0+\gamma_1 A}}{1+e^{\gamma_0+\gamma_1 A}}.
\end{aligned} \tag{s7}$$

Now, given Models II and III, we can show

$$\begin{aligned}
\log\{\mathbb{E}[Y|A,\boldsymbol{C}]\} &= \log\{\mathbb{E}\left[\mathbb{E}[Y|A,M,\boldsymbol{C}]|A,\boldsymbol{C}\right]\}\\
&= \beta_0 + \beta_1 A + \log\left\{\mathbb{E}[e^{\beta_2 M}|A,\boldsymbol{C}]\right\} + \boldsymbol{\beta}_3^T\boldsymbol{C}\\
&= \beta_0 + \beta_1 A + \log\left\{1 + (e^{\beta_2}-1)\frac{e^{\gamma_0+\gamma_1 A}}{1+e^{\gamma_0+\gamma_1 A}}\right\} + \boldsymbol{\beta}_3^T\boldsymbol{C}\\
&= \beta_0 + \log\left\{1+(e^{\beta_2}-1)\frac{e^{\gamma_0}}{1+e^{\gamma_0}}\right\} + \left(\beta_1 + \log\left\{\frac{1+(e^{\beta_2}-1)\frac{e^{\gamma_0+\gamma_1}}{1+e^{\gamma_0+\gamma_1}}}{1+(e^{\beta_2}-1)\frac{e^{\gamma_0}}{1+e^{\gamma_0}}}\right\}\right)A + \boldsymbol{\beta}_3^T\boldsymbol{C}.
\end{aligned}$$

Thus the product and difference methods are compatible and we have $\beta_0^* = \beta_0 + \log\left\{1+(e^{\beta_2}-1)\frac{e^{\gamma_0}}{1+e^{\gamma_0}}\right\}$, $\beta_1^* = \beta_1 + \log\left\{\frac{1+(e^{\beta_2}-1)\frac{e^{\gamma_0+\gamma_1}}{1+e^{\gamma_0+\gamma_1}}}{1+(e^{\beta_2}-1)\frac{e^{\gamma_0}}{1+e^{\gamma_0}}}\right\}$, and $\boldsymbol{\beta}_3^* = \boldsymbol{\beta}_3$ by comparing the coefficients in the above formula with those in Model I.

Of note, a binary A and $\boldsymbol{\gamma}_2 \neq \mathbf{0}$ are also necessary conditions to ensure compatibility between the product and difference method. To show this, assuming A has more than two values, Model III, i.e., $\mathbb{E}[M|A,\boldsymbol{C}] = \frac{e^{\gamma_0+\gamma_1 A+\boldsymbol{\gamma}_2^T\boldsymbol{C}}}{1+e^{\gamma_0+\gamma_1 A+\boldsymbol{\gamma}_2^T\boldsymbol{C}}}$, must be a non-linear function with regard to A and then, given Models II and III,

$$\log\left(\mathbb{E}[Y|A,\boldsymbol{C}]\right) = \beta_0 + \beta_1 A + \log\left\{1+(e^{\beta_2}-1)\frac{e^{\gamma_0+\gamma_1 A+\boldsymbol{\gamma}_2^T\boldsymbol{C}}}{1+e^{\gamma_0+\gamma_1 A+\boldsymbol{\gamma}_2^T\boldsymbol{C}}}\right\} + \boldsymbol{\beta}_3^T\boldsymbol{C}$$

will not have a linear representation with respect to A and $\boldsymbol{C}$. Therefore, A should be binary. If

A is binary, then we can show that $\mathbb{E}[M|A,\boldsymbol{C}] = \frac{e^{\gamma_0+\boldsymbol{\gamma}_2^T\boldsymbol{C}}}{1+e^{\gamma_0+\boldsymbol{\gamma}_2^T\boldsymbol{C}}} + \left(\frac{e^{\gamma_0+\gamma_1+\boldsymbol{\gamma}_2^T\boldsymbol{C}}}{1+e^{\gamma_0+\gamma_1+\boldsymbol{\gamma}_2^T\boldsymbol{C}}} - \frac{e^{\gamma_0+\boldsymbol{\gamma}_2^T\boldsymbol{C}}}{1+e^{\gamma_0+\boldsymbol{\gamma}_2^T\boldsymbol{C}}}\right) A$ and then

$$\begin{aligned}
\log\left(\mathbb{E}[Y|A,\boldsymbol{C}]\right) &= \beta_0 + \beta_1 A + \log\left\{1 + (e^{\beta_2}-1)\frac{e^{\gamma_0+\gamma_1 A+\boldsymbol{\gamma}_2^T\boldsymbol{C}}}{1+e^{\gamma_0+\gamma_1 A+\boldsymbol{\gamma}_2^T\boldsymbol{C}}}\right\} + \boldsymbol{\beta}_3^T\boldsymbol{C} \\
&= \beta_0 + \log\left\{1 + (e^{\beta_2}-1)\frac{e^{\gamma_0+\boldsymbol{\gamma}_2^T\boldsymbol{C}}}{1+e^{\gamma_0+\boldsymbol{\gamma}_2^T\boldsymbol{C}}}\right\} + \left(\beta_1 + \log\left\{\frac{1+(e^{\beta_2}-1)\frac{e^{\gamma_0+\gamma_1+\boldsymbol{\gamma}_2^T\boldsymbol{C}}}{1+e^{\gamma_0+\gamma_1+\boldsymbol{\gamma}_2^T\boldsymbol{C}}}}{1+(e^{\beta_2}-1)\frac{e^{\gamma_0+\boldsymbol{\gamma}_2^T\boldsymbol{C}}}{1+e^{\gamma_0+\boldsymbol{\gamma}_2^T\boldsymbol{C}}}}\right\}\right) A + \boldsymbol{\beta}_3^T\boldsymbol{C}
\end{aligned}$$

According to the previous formula, it can be straightforwardly observed that $\log\left(\mathbb{E}[Y|A,\boldsymbol{C}]\right)$ will always have an interaction term between A and $\boldsymbol{C}$ if $\boldsymbol{\gamma}_2 \neq \boldsymbol{0}$ and therefore cannot be presented as a linear function of A and $\boldsymbol{C}$. This concludes that Models I, II, and III are not compatible if either A is continuous or $\boldsymbol{\gamma}_2 \neq \boldsymbol{0}$.

2. The product method yields MLE and is asymptotically at least as efficient as the difference method.

After noticing that $P_{\boldsymbol{\gamma}}(M_i|A_i,\boldsymbol{C}_i) = P_{\boldsymbol{\gamma}}(M_i|A_i)$ if $\boldsymbol{\gamma}_2 = \boldsymbol{0}$, the full log-likelihood function for $\{Y_i, A_i, M_i, \boldsymbol{C}_i\}_{i=1}^n$ can be decomposed to

$$\begin{aligned}
\log L(\boldsymbol{\beta},\boldsymbol{\gamma};Y,A,M,\boldsymbol{C}) &= \sum_{i=1}^n \log P_{\boldsymbol{\beta}}(Y_i|A_i,M_i,\boldsymbol{C}_i) + \sum_{i=1}^n \log P_{\boldsymbol{\gamma}}(M_i|A_i,\boldsymbol{C}_i) + \sum_{i=1}^n \log P(A_i,\boldsymbol{C}_i), \\
&= \sum_{i=1}^n \log P_{\boldsymbol{\beta}}(Y_i|A_i,M_i,\boldsymbol{C}_i) + \sum_{i=1}^n \log P_{\boldsymbol{\gamma}}(M_i|A_i) + \sum_{i=1}^n \log P(A_i,\boldsymbol{C}_i),
\end{aligned}$$

where $P_{\boldsymbol{\beta}}(Y_i|A_i,M_i,\boldsymbol{C}_i)$ and $P_{\boldsymbol{\gamma}}(M_i|A_i)$ are binomial probability mass functions (with log and logistic links) that depends on unknown parameters $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$, respectively. Solving the log-binomial model II and logistic regression model III maximizes the log-likelihood function $\sum_{i=1}^n \log P_{\boldsymbol{\beta}}(Y_i|A_i,M_i,\boldsymbol{C}_i)$ and $\sum_{i=1}^n \log P_{\boldsymbol{\gamma}}(M_i|A_i)$, respectively. Noting that $\sum_{i=1}^n \log P(A_i,\boldsymbol{C}_i)$ is not a function of $\boldsymbol{\beta}$ or $\boldsymbol{\gamma}$, solving Models II and III maximizes the full log-likelihood function. Therefore the product method agrees with MLE and therefore is asymptotically efficient for estimating mediation measures, which implies that $\text{Avar}(\hat{\theta}^{(p)}) \leq \text{Avar}(\hat{\theta}^{(d)})$, where θ denote the mediation measure.

Web Appendix 2: Numerical studies of the asymptotic relative efficiency (ARE) of the product method compared to the difference method

Appendix 2.1 Procedures for calculating the ARE

In Case Y_bM_c and Y_bM_b, the regression models in (s6) and (s7) were used, respectively. Given the intuitive parameters in Table 2 in the manuscript, the regression parameters were calculated based on the procedures shown in Web Appendix 2.2. Next, we calculated the asymptotic variances of NIE and MP estimators obtained from the product and difference method, i.e., $\text{Avar}(\widehat{\text{NIE}}_{0,1}^{(p)})$ v.s. $\text{Avar}(\widehat{\text{NIE}}_{0,1}^{(d)})$ and $\text{Avar}(\widehat{\text{MP}}_{0,1}^{(p)})$ v.s. $\text{Avar}(\widehat{\text{MP}}_{0,1}^{(d)})$, through the asymptotic variance formulas shown in Web Appendix 2.3. Finally, the AREs for the NIE and MP were given by $\text{ARE(NIE)} = \frac{\text{Avar}(\widehat{\text{NIE}}_{0,1}^{(p)})}{\text{Avar}(\widehat{\text{NIE}}_{0,1}^{(d)})}$ and $\text{ARE(MP)} = \frac{\text{Avar}(\widehat{\text{MP}}_{0,1}^{(p)})}{\text{Avar}(\widehat{\text{MP}}_{0,1}^{(d)})}$, respectively.

We calculated the AREs based on a factorial design of all combinations of the intuitive parameters shown in Table 2 in the manuscript. In summary, a total of 21,600 and 135,000 scenarios were considered in Cases Y_bM_c and Y_bM_b, respectively. Because the intuitive parameters were not always compatible, 2,522 and 73,392 of those scenarios were excluded from analysis. The details for the reasons why we excluded those scenarios were shown as follows.

In Case Y_bM_c , 2,730 among 21,600 scenarios were removed from analysis because the probabilities $P(Y = 1|M, A)$ are larger than 1. Noticing that $P(Y = 1|A, M) = e^{\beta_0+\beta_1 A+\beta_2 M}$ is unbounded with respect to the regression coefficients, $P(Y = 1|A, M)$ can be greater than 1 if the regression coefficients are large and the values of A and M are extreme. Therefore, we need exclude scenarios where $P(Y = 1|M, A)$ may be greater than 1 from our numerical analysis. Ideally, we need to substitute all possible combinations of A and M into $P(Y = 1|A, M) = e^{\beta_0+\beta_1 A+\beta_2 M}$ to check if it is bigger than 1. If it is true, then we remove its corresponding scenario from analysis. However, $f(M|A = a)$ follows a normal distribution in Case Y_bM_c, therefore M can be any value between $(-\infty, +\infty)$, which can always make $P(Y = 1|A, M)$ greater than 1 if we numerate all possible M. However, since M near $-\infty$ or $+\infty$ is too rare to happen in practice, we consider all possible M lying within the 0.001% to 99.999% quantile of $M|A = a$, which should cover all possible values of M that could appear in a finite sample data set. Specifically, in the simulation study, we excluded the scenarios that $P(Y = 1|A = 1, M = m)$ or $P(Y = 1|A = 0, M = m)$ is greater than 1 for any possible m in (i) 0.001% to 99.999% quantile of the distribution of $M|A=1$ or (ii) 0.001% to 99.999% quantile of the distribution of $M|A=0$.

In Case Y_bM_b, 73,392 among 135,000 scenarios were removed from analysis and reasons are summarized as below:

- 1. In 68,808 scenarios, β_2 cannot be specified given the intuitive parameters. Specifically, the equation $\exp(\text{MP} \times \text{TE}) = \frac{(1+e^{\gamma_0})(1+e^{\beta_2+\gamma_0+\gamma_1})}{(1+e^{\gamma_0+\gamma_1})(1+e^{\beta_2+\gamma_0})}$ for calculating β_2 given the TE, MP, γ_0, and γ_1 does not have a real root, when (i) $\exp(\text{MP} \times \text{TE}) > \frac{1+e^{\gamma_0}}{1+e^{\gamma_0+\gamma_1}}$, $\gamma_0 < 0$ and $\gamma_1 < 0$ and (ii) $\exp(\text{MP} \times \text{TE}) < \frac{1+e^{\gamma_0}}{1+e^{\gamma_0+\gamma_1}}$, $\gamma_0 > 0$ and $\gamma_1 > 0$. Under condition (i), it is straightforward to show that the right hand of the equation $\frac{(1+e^{\gamma_0})(1+e^{\beta_2+\gamma_0+\gamma_1})}{(1+e^{\gamma_0+\gamma_1})(1+e^{\beta_2+\gamma_0})} \in (0, \frac{1+e^{\gamma_0}}{1+e^{\gamma_0+\gamma_1}})$ for all $\beta_2 \in \mathbb{R}$ when $\gamma_0 < 0$ and $\gamma_1 < 0$, which contradicts the left hand of the equation $\exp(\text{MP} \times \text{TE}) > \frac{1+e^{\gamma_0}}{1+e^{\gamma_0+\gamma_1}}$. Similarly, under condition (ii) we can show that the right hand of the equation $\frac{(1+e^{\gamma_0})(1+e^{\beta_2+\gamma_0+\gamma_1})}{(1+e^{\gamma_0+\gamma_1})(1+e^{\beta_2+\gamma_0})} \in (\frac{1+e^{\gamma_0}}{1+e^{\gamma_0+\gamma_1}}, +\infty)$ for all $\beta_2 \in \mathbb{R}$ when $\gamma_0 > 0$ and $\gamma_1 > 0$, which contradicts the left hand of the equation $\exp(\text{MP} \times \text{TE}) < \frac{1+e^{\gamma_0}}{1+e^{\gamma_0+\gamma_1}}$.
- 2. In 4,584 scenarios, the probabilility $P(Y=1|A=a, M=m) = e^{\beta_0+\beta_1 a+\beta_2 m}$ is greater than 1, where $a \in \{0,1\}$ and $m \in \{0,1\}$. This may happen when the total effect is large and the outcome prevalence is common (e.g., TE=log(2) and $P(Y=1) = 32\%$).

Appendix 2.2 Relationships between the intuitive parameters and regression parameters

Case Y_bM_c , binary outcome and continuous mediator without consideration of confounders

In Case Y_bM_b, the intuitive parameters include TE, MP, $P(Y=1)$, $P(A=1)$, $\text{Var}(M)$, $\mathbb{E}[M]$, and $\text{Corr}(A, M)$. Here, we show how to calculate the regression parameters based on those intuitive parameters. First, calculate the γ_1, γ_0, and σ_3 of the mediator model, based on the following equations,

$$\text{Corr}(A, M) = \frac{\gamma_1 P(A=1)(1-P(A=1))}{\sqrt{P(A=1)(1-P(A=1))\text{Var}(M)}},$$
$$\mathbb{E}[M] = \gamma_0 + \gamma_1 P(A=1),$$
$$\text{Var}(M) = \gamma_1^2 P(A=1)\,(1-P(A=1)) + \sigma_3^2.$$

Then, we calculate β_1, β_2, and β_0, the regression parameters in the conditional outcome model. Specifically, we calculate β_1, β_2, and β_0, sequentially and separately, based on the following equations:

$$(1 - \text{MP}) \times \text{TE} = \beta_1,$$
$$\text{MP} \times \text{NIE} = \beta_2\gamma_1,$$
$$P(Y=1) = P(A=1)\exp\Big(\beta_0 + \beta_1 + \beta_2(\gamma_0+\gamma_1) + \tfrac{\beta_2^2\sigma_3^2}{2}\Big) + (1 - P(A=1))\exp\Big(\beta_0 + \beta_2\gamma_0 + \tfrac{\beta_2^2\sigma_3^2}{2}\Big),$$

The third equation for solving β_0 is non-linear and does not have an explicit root, therefore we use the **nleqslv** function in R software to obtain a numerical solution. Finally, β_0^* and β_1^*, the regression parameters in the marginal outcome model, were obtained by the parametric relationships shown in Web Appendix 1.3.

Remark: The formula for $P(Y=1)$ and $\text{Corr}(A, M)$ were derived as follows

$$\begin{aligned}
P(Y=1) =& \mathbb{P}(Y=1|A=1)P(A=1) + \mathbb{P}(Y=1|A=0)\left(1-P(A=1)\right) \\
=& P(A=1)\int \mathbb{P}(Y=1|M=m, A=1)p(M=m|A=1)dM \\
&+ (1-P(A=1))\int \mathbb{P}(Y=1|M=m, A=0)p(M=m|A=0)dM \\
=& P(A=1)\int e^{\beta_0+\beta_1+\beta_2 m}\frac{1}{\sqrt{2\pi}\sigma_3}e^{-\frac{(m-\gamma_0-\gamma_1)^2}{2\sigma_3^2}}dm + (1-P(A=1))\int e^{\beta_0+\beta_2 m}\frac{1}{\sqrt{2\pi}\sigma_3}e^{-\frac{(m-\gamma_0)^2}{2\sigma_3^2}}dm \\
=& P(A=1)\exp\Big(\beta_0+\beta_1+\beta_2(\gamma_0+\gamma_1)+\frac{\beta_2^2\sigma_3^2}{2}\Big) + (1-P(A=1))\exp\Big(\beta_0+\beta_2\gamma_0+\frac{\beta_2^2\sigma_3^2}{2}\Big).
\end{aligned}$$

$$\begin{aligned}
\text{Corr}(A, M) &= \frac{\text{Cov}(M, A)}{\sqrt{\text{Var}(A)\text{Var}(M)}} = \frac{\mathbb{E}(AM) - P(A=1)\mathbb{E}M}{\sqrt{P(A=1)\left(1-P(A=1)\right)\text{Var}(M)}} \\
&= \frac{\mathbb{E}(\gamma_0 A + \gamma_1 A^2) - P(A=1)\mathbb{E}(\gamma_0+\gamma_1 A)}{\sqrt{P(A=1)\left(1-P(A=1)\right)\text{Var}(M)}} \\
&= \frac{\gamma_1 P(A=1)\left(1-P(A=1)\right)}{\sqrt{P(A=1)\left(1-P(A=1)\right)\text{Var}(M)}}.
\end{aligned}$$

Case Y_bM_b, binary outcome and binary mediator without consideration of confounders

In Case Y_bM_b, the intuitive parameters are TE, MP, $P(Y=1)$, $P(A=1)$, $P(M=1)$, and $OR(M|A)$. Here, we show how to calculate the regression parameters based on those intuitive parameters. We first calculate γ_1 and γ_0, the regression parameters in the mediator model, based on the following two equations:

$$OR(M|A) = \exp(\gamma_1)$$
$$\mathbb{E}[M] = P(A=1)\tfrac{e^{\gamma_0+\gamma_1}}{1+e^{\gamma_0+\gamma_1}} + (1-P(A=1))\,\tfrac{e^{\gamma_0}}{1+e^{\gamma_0}},$$

Then, we calculate β_1, β_2, and β_0, the unknown parameters in the conditional outcome model, based on the following three equations, separately,

$$(1-\text{MP}) \times \text{TE} = \beta_1,$$
$$\text{MP} \times \text{TE} = \log\left(\tfrac{(1+e^{\gamma_0})(1+e^{\beta_2+\gamma_0+\gamma_1})}{(1+e^{\gamma_0+\gamma_1})(1+e^{\beta_2+\gamma_0})}\right),$$
$$P(Y=1) = e^{\beta_0}p_{00} + e^{\beta_0+\beta_1}p_{01} + e^{\beta_0+\beta_2}p_{10} + e^{\beta_0+\beta_1+\beta_2}p_{11},$$

where $p_{am} = P(A=a, M=m) = P(M=m|A=a)P(A=a)$; i.e., $p_{11} = \frac{e^{\gamma_0+\gamma_1}}{1+e^{\gamma_0+\gamma_1}}p_a$, $p_{10} = \frac{e^{\gamma_0}}{1+e^{\gamma_0}}(1-p_a)$, $p_{01} = (1-\frac{e^{\gamma_0+\gamma_1}}{1+e^{\gamma_0+\gamma_1}})p_a$, and $p_{00} = (1-\frac{e^{\gamma_0}}{1+e^{\gamma_0}})(1-p_a)$. Finally, β_0^* and β_1^*, the regression parameters in the marginal outcome model, were obtained by the parametric relationships shown in Web Appendix 1.4.

Appendix 2.3 Asymptotic variance formulas

The estimating equations for the marginal and conditional outcome model can be shown as

$$\mathbb{U}^{(1)}(\boldsymbol{\beta}^*) = \sum_{i=1}^{n} \boldsymbol{U}_{i1}V_{i1}^{-1}\Big[Y_i - \mathbb{E}[Y_i|A_i,\boldsymbol{C}_i]\Big] = \boldsymbol{0} \tag{s8}$$

$$\mathbb{U}^{(2)}(\boldsymbol{\beta}) = \sum_{i=1}^{n} \boldsymbol{U}_{i2}V_{i2}^{-1}\Big[Y_i - \mathbb{E}[Y_i|A_i,M_i,\boldsymbol{C}_i]\Big] = \boldsymbol{0} \tag{s9}$$

where $\boldsymbol{U}_{i1} = \frac{\partial\mathbb{E}[Y_i|A_i,\boldsymbol{C}_i]}{\partial\boldsymbol{\beta}^*}$, and $\boldsymbol{U}_{i2} = \frac{\partial\mathbb{E}[Y_i|A_i,M_i,\boldsymbol{C}_i]}{\partial\boldsymbol{\beta}}$. If the outcome is continuous, $V_{i1} = \sigma_1^2$ and $V_{i2} = \sigma_2^2$; if the outcome is binary, we have $V_{i1} = e^{\beta_0^*+\beta_1^*A_i+\boldsymbol{\beta}_3^{*T}\boldsymbol{C}_i} \times (1-e^{\beta_0^*+\beta_1^*A_i+\boldsymbol{\beta}_3^{*T}\boldsymbol{C}_i})$ and $V_{i2} = e^{\beta_0+\beta_1A_i+\beta_2M_i+\boldsymbol{\beta}_3^T\boldsymbol{C}_i} \times (1-e^{\beta_0+\beta_1A_i+\beta_2M_i+\boldsymbol{\beta}_3^T\boldsymbol{C}_i})$. The estimating equation for the mediator model is

$$\mathbb{U}^{(3)}(\boldsymbol{\gamma}) = \sum_{i=1}^{n} \boldsymbol{U}_{i3}V_{i3}^{-1}\Big[M_i - \mathbb{E}[M_i|A_i,\boldsymbol{C}_i]\Big] = \boldsymbol{0}, \tag{s10}$$

where $\boldsymbol{U}_{i3} = \frac{\partial\mathbb{E}[M_i|A_i,\boldsymbol{C}_i]}{\partial\boldsymbol{\gamma}}$ and V_{i3} is σ_3^2 and $\frac{e^{\gamma_0+\gamma_1A_i+\boldsymbol{\gamma}_2^T\boldsymbol{C}_i}}{(1+e^{\gamma_0+\gamma_1A_i+\boldsymbol{\gamma}_2^T\boldsymbol{C}_i})^2}$ if M is continuous and binary, respectively. Next we derive the asymptotic variance formulas for the NIE and MP estimators obtained by the product and difference method.

Difference Method

For the difference method, the estimating equation for unknown regression parameters $\boldsymbol{\theta} = [\boldsymbol{\beta}^{*T}, \boldsymbol{\beta}^T]^T$ is $\mathbb{U}(\boldsymbol{\theta}) = \begin{pmatrix} \mathbb{U}^{(1)}(\boldsymbol{\beta}^*) \\ \mathbb{U}^{(2)}(\boldsymbol{\beta}) \end{pmatrix} = \mathbf{0}$. By theory of unbiased estimating equations [3, 6], we know $\sqrt{n}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})$ is asymptotically multivariate normal with zero mean and covariance matrix is given by

$$\boldsymbol{\Omega}_{\boldsymbol{\theta}} = \left(\mathbb{E}\left[\boldsymbol{D}^T\boldsymbol{V}^{-1}\boldsymbol{D}\right]\right)^{-1}\left\{\mathbb{E}\left[\boldsymbol{D}^T\boldsymbol{V}^{-1}\boldsymbol{\epsilon}\boldsymbol{\epsilon}^T\boldsymbol{V}^{-1}\boldsymbol{D}\right]\right\}\left(\mathbb{E}\left[\boldsymbol{D}^T\boldsymbol{V}^{-1}\boldsymbol{D}\right]\right)^{-1}, \tag{s11}$$

where $\boldsymbol{D} = \left[\frac{\partial\mathbb{E}[Y|A,\boldsymbol{C}]}{\partial\boldsymbol{\theta}}, \frac{\partial\mathbb{E}[Y|A,M,\boldsymbol{C}]}{\partial\boldsymbol{\theta}}\right]$, $\boldsymbol{\epsilon} = (Y - \mathbb{E}[Y|A,\boldsymbol{C}], Y - \mathbb{E}[Y|A,M,\boldsymbol{C}])^T$ includes the error terms in the marginal and conditional outcome models, and the expectations in (s11) are taken over $\{Y, A, M, \boldsymbol{C}\}$. Here, $\boldsymbol{V}$ is a variance-covariance matrix that equals to $diag(\sigma_1^2, \sigma_2^2)$ and $diag\Big(e^{\beta_0^*+\beta_1^*A+\boldsymbol{\beta}_3^{*T}\boldsymbol{C}} \times (1 - e^{\beta_0^*+\beta_1^*A+\boldsymbol{\beta}_3^{*T}\boldsymbol{C}}), e^{\beta_0+\beta_1A+\beta_2M+\boldsymbol{\beta}_3^T\boldsymbol{C}} \times (1 - e^{\beta_0+\beta_1A+\beta_2M+\boldsymbol{\beta}_3^T\boldsymbol{C}})\Big)$ when the outcome is continuous and binary, respectively. Formula (s11) can be further simplified to

$$\boldsymbol{\Omega}_{\boldsymbol{\theta}} = \left(\mathbb{E}\left[\boldsymbol{D}^T\boldsymbol{V}^{-1}\boldsymbol{D}\right]\right)^{-1}\left\{\mathbb{E}\left[\boldsymbol{D}^T\boldsymbol{V}^{-1}\mathbb{E}[\boldsymbol{\epsilon}\boldsymbol{\epsilon}^T|A,M,\boldsymbol{C}]\boldsymbol{V}^{-1}\boldsymbol{D}\right]\right\}\left(\mathbb{E}\left[\boldsymbol{D}^T\boldsymbol{V}^{-1}\boldsymbol{D}\right]\right)^{-1},$$

where $\mathbb{E}[\boldsymbol{\epsilon}\boldsymbol{\epsilon}^T|A,M,\boldsymbol{C}]$ is a function of $\{M, A, \boldsymbol{C}\}$ shown in Web Appendix 2.4.

Next we show how to numerically calculate $\boldsymbol{\Omega}_{\boldsymbol{\theta}}$. We assume A and $\boldsymbol{C}$ are binary variables. Note that the formula for calculating $\boldsymbol{\Omega}_{\boldsymbol{\theta}}$ depends on $\mathbb{E}\left[\boldsymbol{D}^T\boldsymbol{V}^{-1}\mathbb{E}[\boldsymbol{\epsilon}\boldsymbol{\epsilon}^T|A,M,\boldsymbol{C}]\boldsymbol{V}^{-1}\boldsymbol{D}\right]$ and $\mathbb{E}\left[\boldsymbol{D}^T\boldsymbol{V}^{-1}\boldsymbol{D}\right]$, two expectations with respect to variables $\{A, M, \boldsymbol{C}\}$. We use the following method to numerically calculate the expectations with respect to $\{A, M, \boldsymbol{C}\}$. When M is also binary (Case Y_bM_b), the density function of $\{A, M, \boldsymbol{C}\}$ is discrete and we can obtain the expectation for any function of $\{M, A, \boldsymbol{C}\}$, say $f(M, A, \boldsymbol{C})$, through the following formula

$$\begin{aligned}
\mathbb{E}[f(M, A, \boldsymbol{C})] &= \sum_{m,a,\boldsymbol{c}} f(M = m, A = a, \boldsymbol{C} = c)P(M = m, A = a, \boldsymbol{C} = c), \\
&= \sum_{m,a,\boldsymbol{c}} f(M = m, A = a, \boldsymbol{C} = c)P(M = m|A = a, \boldsymbol{C} = c)P(A = a, \boldsymbol{C} = c),
\end{aligned}$$

where $P(M = m|A = a, \boldsymbol{C} = c)$ refers to the mediator model and $P(A = a, \boldsymbol{C} = c)$ is an explicit function describing the A-$\boldsymbol{C}$ relationship. If M is continuous (Case Y_bM_c), we can calculate

$\mathbb{E}[f(M,A,\boldsymbol{C})]$ by

$$\mathbb{E}[f(A,M,\boldsymbol{C})]=\sum_{a,\boldsymbol{c}}P(A=a,\boldsymbol{C}=\boldsymbol{c})\times\int_{-\infty}^{+\infty}f(M,A=a,\boldsymbol{C}=\boldsymbol{c})\phi(\frac{M-\gamma_0-\gamma_1 a-\boldsymbol{\gamma}_2^T\boldsymbol{c}}{\sigma_3})dM,$$

where $\phi(.)$ denotes the density function of standard normal distribution. We use the Gauss-Hermite Quadrature (GHQ) method to numerically calculate the above integral.

Now, we know how to numerically calculate the asymptotic variance-covariance matrix of $\hat{\boldsymbol{\theta}}$. Finally, by multivariate delta method, the asymptotic variances of $\widehat{\text{NIE}}^{(d)}_{a^*,a|\boldsymbol{c}}$ and $\widehat{\text{MP}}^{(d)}_{a^*,a|\boldsymbol{c}}$ can be shown as

$$\text{Avar}(\widehat{\text{NIE}}^{(d)}_{a^*,a|\boldsymbol{c}})=\frac{1}{n}\left[\frac{\partial NIE^{(d)}_{a^*,a|\boldsymbol{c}}}{\partial\boldsymbol{\theta}}\right]^T\boldsymbol{\Omega}_{\boldsymbol{\theta}}\left[\frac{\partial NIE^{(d)}_{a^*,a|\boldsymbol{c}}}{\partial\boldsymbol{\theta}}\right],$$

$$\text{Avar}(\widehat{\text{MP}}^{(d)}_{a^*,a|\boldsymbol{c}})=\frac{1}{n}\left[\frac{\partial MP^{(d)}_{a^*,a|\boldsymbol{c}}}{\partial\boldsymbol{\theta}}\right]^T\boldsymbol{\Omega}_{\boldsymbol{\theta}}\left[\frac{\partial MP^{(d)}_{a^*,a|\boldsymbol{c}}}{\partial\boldsymbol{\theta}}\right],$$

where $\text{NIE}^{(d)}_{a^*,a|\boldsymbol{c}}=(\beta_1^*-\beta_1)(a-a^*)$ and $\text{MP}^{(d)}_{a^*,a|\boldsymbol{c}}=1-\frac{\beta_1}{\beta_1^*}$ are the NIE and MP expressions given by the difference method. It is straightforward to show $\frac{\partial NIE^{(d)}_{a^*,a|\boldsymbol{c}}}{\partial\boldsymbol{\theta}}=[0,a-a^*,\mathbf{0},0,-(a-a^*),0,\mathbf{0}]^T$ and $\frac{\partial MP^{(d)}_{a^*,a|\boldsymbol{c}}}{\partial\boldsymbol{\theta}}=[0,\frac{\beta_1}{\beta_1^{*2}},\mathbf{0},0,-\frac{1}{\beta_1^*},0,\mathbf{0}]^T$.

Product Method

For the product method, the estimating equation for unknown regression parameters $\boldsymbol{\theta}^*=[\boldsymbol{\beta}^T,\boldsymbol{\gamma}^T]^T$ is $\mathbb{U}(\boldsymbol{\theta})=\begin{pmatrix}\mathbb{U}^{(2)}(\boldsymbol{\beta})\\ \mathbb{U}^{(3)}(\boldsymbol{\gamma})\end{pmatrix}=\mathbf{0}$. By the theory of unbiased estimating equation [3, 6], we know $\sqrt{n}(\hat{\boldsymbol{\theta}}^*-\boldsymbol{\theta}^*)$ is asymptotically multivariate normal with zero mean and covariance matrix is given by

$$\boldsymbol{\Omega}_{\boldsymbol{\theta}^*}=\left(\mathbb{E}\left[\boldsymbol{D^*}^T\boldsymbol{V^*}^{-1}\boldsymbol{D}^*\right]\right)^{-1}\left\{\mathbb{E}\left[\boldsymbol{D^*}^T\boldsymbol{V^*}^{-1}\boldsymbol{\epsilon}^*\boldsymbol{\epsilon^*}^T\boldsymbol{V}^{*-1}\boldsymbol{D}^*\right]\right\}\left(\mathbb{E}\left[\boldsymbol{D^*}^T\boldsymbol{V^*}^{-1}\boldsymbol{D}^*\right]\right)^{-1},\quad\text{(s12)}$$

where $\boldsymbol{D}^*=\left[\frac{\partial\mathbb{E}[Y|A,M,\boldsymbol{C}]}{\partial\boldsymbol{\theta}^*},\frac{\partial\mathbb{E}[M|A,\boldsymbol{C}]}{\partial\boldsymbol{\theta}^*}\right]$, $\boldsymbol{V}^*=diag(V_2,V_3)$ is the working variance-covariance matrix, $\boldsymbol{\epsilon}^*=[\epsilon_2,\epsilon_3]^T=[Y-\mathbb{E}[Y|A,M,\boldsymbol{C}],M-\mathbb{E}[M|A,\boldsymbol{C}]]^T$ includes the error terms in the conditional outcome model and mediator model, and the expectations in equation (s12) are taken over $\{Y,A,M,\boldsymbol{C}\}$. Here, V_3 is σ_3^2 and $\frac{e^{\gamma_0+\gamma_1 A+\boldsymbol{\gamma}_2^T\boldsymbol{C}}}{(1+e^{\gamma_0+\gamma_1 A+\boldsymbol{\gamma}_2^T\boldsymbol{C}})^2}$ if M is continuous and binary, respectively. We can further simplify formula (s12) after noticing

$$\begin{aligned}\mathbb{E}\left[\epsilon_2\epsilon_3|A,M,\boldsymbol{C}\right] &= \mathbb{E}\left[(Y-\mathbb{E}[Y|A,M,\boldsymbol{C}])\times(M-\mathbb{E}[M|A,\boldsymbol{C}])|A,M,\boldsymbol{C}\right]\\ &= \mathbb{E}\left[Y-\mathbb{E}[Y|A,M,\boldsymbol{C}]|A,M,\boldsymbol{C}\right]\times(M-\mathbb{E}[M|A,\boldsymbol{C}])\\ &= 0\times(M-\mathbb{E}[M|A,\boldsymbol{C}])=0.\end{aligned}$$

Therefore,

$$\begin{aligned}&\mathbb{E}\Big[\boldsymbol{D^*}^T\boldsymbol{V^*}^{-1}\boldsymbol{\epsilon}^*\boldsymbol{\epsilon^*}^T\boldsymbol{V}^{*-1}\boldsymbol{D}^*\Big]\\ =&\mathbb{E}\Big[\boldsymbol{D^*}^T\boldsymbol{V^*}^{-1}\mathbb{E}[\boldsymbol{\epsilon}^*\boldsymbol{\epsilon^*}^T|A,M,\boldsymbol{C}]\boldsymbol{V^*}^{-1}\boldsymbol{D}^*\Big]\\ =&\mathbb{E}\Big[\boldsymbol{D^*}^T\boldsymbol{V^*}^{-1}\mathbb{E}\left[\begin{pmatrix}\epsilon_2^2 & \epsilon_2\epsilon_3\\ \epsilon_2\epsilon_3 & \epsilon_3^2\end{pmatrix}|A,M,\boldsymbol{C}\right]\boldsymbol{V^*}^{-1}\boldsymbol{D}^*\Big]\\ =&\mathbb{E}\Big[\begin{pmatrix}\frac{\partial\mathbb{E}[Y|A,M,\boldsymbol{C}]}{\partial\boldsymbol{\beta}^T} & \mathbf{0}\\ \mathbf{0} & \frac{\partial\mathbb{E}[M|A,\boldsymbol{C}]}{\partial\boldsymbol{\gamma}^T}\end{pmatrix}\begin{pmatrix}V_2^{-1} & 0\\ 0 & V_3^{-1}\end{pmatrix}\begin{pmatrix}V_2 & 0\\ 0 & V_3\end{pmatrix}\begin{pmatrix}V_2^{-1} & 0\\ 0 & V_3^{-1}\end{pmatrix}\begin{pmatrix}\frac{\partial\mathbb{E}[Y|A,M,\boldsymbol{C}]}{\partial\boldsymbol{\beta}} & \mathbf{0}\\ \mathbf{0} & \frac{\partial\mathbb{E}[M|A,\boldsymbol{C}]}{\partial\boldsymbol{\gamma}}\end{pmatrix}\Big]\\ =&\begin{pmatrix}\mathbb{E}[\boldsymbol{U}_2^TV_2^{-1}\boldsymbol{U}_2] & \mathbf{0}\\ \mathbf{0} & \mathbb{E}[\boldsymbol{U}_3^TV_3^{-1}\boldsymbol{U}_3]\end{pmatrix}.\end{aligned}$$

Also, we can show

$$\begin{aligned}\mathbb{E}\Big[\boldsymbol{D^*}^T\boldsymbol{V^*}^{-1}\boldsymbol{D}^*\Big] &= \mathbb{E}\left[\begin{pmatrix}\frac{\partial\mathbb{E}[Y|A,M,\boldsymbol{C}]}{\partial\boldsymbol{\beta}^T} & \mathbf{0}\\ \mathbf{0} & \frac{\partial\mathbb{E}[M|A,\boldsymbol{C}]}{\partial\boldsymbol{\gamma}^T}\end{pmatrix}\begin{pmatrix}V_2^{-1} & 0\\ 0 & V_3^{-1}\end{pmatrix}\begin{pmatrix}\frac{\partial\mathbb{E}[Y|A,M,\boldsymbol{C}]}{\partial\boldsymbol{\beta}} & \mathbf{0}\\ \mathbf{0} & \frac{\partial\mathbb{E}[M|A,\boldsymbol{C}]}{\partial\boldsymbol{\gamma}}\end{pmatrix}\right]\\ &= \begin{pmatrix}\mathbb{E}[\boldsymbol{U}_2^TV_2^{-1}\boldsymbol{U}_2] & \mathbf{0}\\ \mathbf{0} & \mathbb{E}[\boldsymbol{U}_3^TV_3^{-1}\boldsymbol{U}_3]\end{pmatrix}.\end{aligned}$$

It follows that

$$\begin{aligned}\boldsymbol{\Omega}_{\boldsymbol{\theta}^*} &= \begin{pmatrix}\mathbb{E}[\boldsymbol{U}_2^TV_2^{-1}\boldsymbol{U}_2] & \mathbf{0}\\ \mathbf{0} & \mathbb{E}[\boldsymbol{U}_3^TV_3^{-1}\boldsymbol{U}_3]\end{pmatrix}^{-1}\begin{pmatrix}\mathbb{E}[\boldsymbol{U}_2^TV_2^{-1}\boldsymbol{U}_2] & \mathbf{0}\\ \mathbf{0} & \mathbb{E}[\boldsymbol{U}_3^TV_3^{-1}\boldsymbol{U}_3]\end{pmatrix}\begin{pmatrix}\mathbb{E}[\boldsymbol{U}_2^TV_2^{-1}\boldsymbol{U}_2] & \mathbf{0}\\ \mathbf{0} & \mathbb{E}[\boldsymbol{U}_3^TV_3^{-1}\boldsymbol{U}_3]\end{pmatrix}^{-1}\\ &= \begin{pmatrix}\mathbb{E}[\boldsymbol{U}_2^TV_2^{-1}\boldsymbol{U}_2]^{-1} & \mathbf{0}\\ \mathbf{0} & \mathbb{E}[\boldsymbol{U}_3^TV_3^{-1}\boldsymbol{U}_3]^{-1}\end{pmatrix}.\end{aligned}$$

Now, we know the asymptotic variance of $\sqrt{n}(\hat{\boldsymbol{\theta}}^* - \boldsymbol{\theta}^*)$ is $\boldsymbol{\Omega}_{\boldsymbol{\theta}^*} = \begin{bmatrix} \mathbb{E}[\boldsymbol{U}_2^T V_2^{-1} \boldsymbol{U}_2]^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbb{E}[\boldsymbol{U}_3^T V_3^{-1} \boldsymbol{U}_3]^{-1} \end{bmatrix}$. The formula for $\boldsymbol{\Omega}_{\boldsymbol{\theta}^*}$ takes expectations with respect to $\{A, M, \boldsymbol{C}\}$ and we can again use the numerical method shown in the previous subsection to numerically calculate $\boldsymbol{\Omega}_{\boldsymbol{\theta}^*}$. Finally, using the multivariate delta method, we have

$$\begin{aligned}
\text{Avar}(\widehat{\text{NIE}}^{(p)}_{a^*,a|\boldsymbol{c}}) &= \frac{1}{n} \left[\frac{\partial \text{NIE}^{(p)}_{a^*,a|\boldsymbol{c}}}{\partial \boldsymbol{\theta}^*} \right]^T \boldsymbol{\Omega}_{\boldsymbol{\theta}^*} \left[\frac{\partial \text{NIE}^{(p)}_{a^*,a|\boldsymbol{c}}}{\partial \boldsymbol{\theta}^*} \right], \\
\text{Avar}(\widehat{\text{MP}}^{(p)}_{a^*,a|\boldsymbol{c}}) &= \frac{1}{n} \left[\frac{\partial \text{MP}^{(p)}_{a^*,a|\boldsymbol{c}}}{\partial \boldsymbol{\theta}^*} \right]^T \boldsymbol{\Omega}_{\boldsymbol{\theta}^*} \left[\frac{\partial \text{MP}^{(p)}_{a^*,a|\boldsymbol{c}}}{\partial \boldsymbol{\theta}^*} \right],
\end{aligned}$$

where $\text{NIE}^{(p)}_{a^*,a|\boldsymbol{c}}$ and $\text{MP}^{(p)}_{a^*,a|\boldsymbol{c}}$ are given in Table 1 in the manuscript.

Appendix 2.4 Expression of $E(\epsilon\epsilon^T|A, M, \mathbf{C})$

In the numerical studies in this paper, we only considered a binary outcome (Cases $Y_b M_c$ and $Y_b M_b$). Therefore, we further derive more explicit results for those Cases. Specifically, we have $\mathbb{E}(\boldsymbol{\epsilon}\boldsymbol{\epsilon}^T|A, M, \boldsymbol{C}) = \begin{bmatrix} a & b \\ b & d \end{bmatrix}$, with

$$\begin{aligned}
d &= \mathbb{E}\left[(Y - e^{\beta_0+\beta_1 A+\beta_2 M+\boldsymbol{\beta}_3^T \boldsymbol{C}})(Y - e^{\beta_0+\beta_1 A+\beta_2 M+\boldsymbol{\beta}_3^T \boldsymbol{C}})|A, M, \boldsymbol{C} \right] \\
&= \mathbb{E}[Y|A, M, \boldsymbol{C}] - 2e^{\beta_0+\beta_1 A+\beta_2 M+\boldsymbol{\beta}_3^T \boldsymbol{C}} \mathbb{E}[Y|A, M, \boldsymbol{C}] + e^{2\beta_0+2\beta_1 A+2\beta_2 M+2\boldsymbol{\beta}_3^T \boldsymbol{C}} \\
&= e^{\beta_0+\beta_1 A+\beta_2 M+\boldsymbol{\beta}_3^T \boldsymbol{C}} - e^{2\beta_0+2\beta_1 A+2\beta_2 M+2\boldsymbol{\beta}_3^T \boldsymbol{C}} \\
&= e^{\beta_0+\beta_1 A+\beta_2 M+\boldsymbol{\beta}_3^T \boldsymbol{C}}(1 - e^{\beta_0+\beta_1 A+\beta_2 M+\boldsymbol{\beta}_3^T \boldsymbol{C}}),
\end{aligned}$$

$$
\begin{aligned}
b &= \mathbb{E}\left[(Y - e^{\beta_0^* + \beta_1^* A + \boldsymbol{\beta}_3^{*T}\boldsymbol{C}})(Y - e^{\beta_0 + \beta_1 A + \beta_2 M + \boldsymbol{\beta}_3^T\boldsymbol{C}})|A, M, \boldsymbol{C}\right] \\
&= \mathbb{E}[Y|A, M, \boldsymbol{C}] - e^{\beta_0^* + \beta_1^* A + \boldsymbol{\beta}_3^{*T}\boldsymbol{C}}\mathbb{E}[Y|A, M, \boldsymbol{C}] - e^{\beta_0 + \beta_1 A + \beta_2 M + \boldsymbol{\beta}_3^T\boldsymbol{C}}\mathbb{E}[Y|A, M, \boldsymbol{C}] \\
&\quad + e^{\beta_0 + \beta_1 A + \beta_2 M + \boldsymbol{\beta}_3^T\boldsymbol{C} + \beta_0^* + \beta_1^* A + \boldsymbol{\beta}_3^{*T}\boldsymbol{C}} \\
&= e^{\beta_0 + \beta_1 A + \beta_2 M + \boldsymbol{\beta}_3^T\boldsymbol{C}} - e^{\beta_0^* + \beta_1^* A + \boldsymbol{\beta}_3^{*T}\boldsymbol{C}}e^{\beta_0 + \beta_1 A + \beta_2 M + \boldsymbol{\beta}_3^T\boldsymbol{C}} - e^{\beta_0 + \beta_1 A + \beta_2 M + \boldsymbol{\beta}_3^T\boldsymbol{C}}e^{\beta_0 + \beta_1 A + \beta_2 M + \boldsymbol{\beta}_3^T\boldsymbol{C}} \\
&\quad + e^{\beta_0 + \beta_1 A + \beta_2 M + \boldsymbol{\beta}_3^T\boldsymbol{C} + \beta_0^* + \beta_1^* A + \boldsymbol{\beta}_3^{*T}\boldsymbol{C}} \\
&= e^{\beta_0 + \beta_1 A + \beta_2 M + \boldsymbol{\beta}_3^T\boldsymbol{C}}(1 - e^{\beta_0 + \beta_1 A + \beta_2 M + \boldsymbol{\beta}_3^T\boldsymbol{C}}) = d,
\end{aligned}
$$

and,

$$
\begin{aligned}
a &= \mathbb{E}\left[(Y - e^{\beta_0^* + \beta_1^* A + \boldsymbol{\beta}_3^{*T}\boldsymbol{C}})(Y - e^{\beta_0^* + \beta_1^* A + \boldsymbol{\beta}_3^{*T}\boldsymbol{C}})|A, M, \boldsymbol{C}\right] \\
&= \mathbb{E}[Y|A, M, \boldsymbol{C}] - 2e^{\beta_0^* + \beta_1^* A + \boldsymbol{\beta}_3^{*T}\boldsymbol{C}}\mathbb{E}[Y|A, M, \boldsymbol{C}] + e^{2\beta_0^* + 2\beta_1^* A + 2\boldsymbol{\beta}_3^{*T}\boldsymbol{C}} \\
&= e^{\beta_0 + \beta_1 A + \beta_2 M + \boldsymbol{\beta}_3^T\boldsymbol{C}} - 2e^{\beta_0 + \beta_1 A + \beta_2 M + \boldsymbol{\beta}_3^T\boldsymbol{C} + \beta_0^* + \beta_1^* A + \boldsymbol{\beta}_3^{*T}\boldsymbol{C}} + e^{2\beta_0^* + 2\beta_1^* A + 2\boldsymbol{\beta}_3^{*T}\boldsymbol{C}} \\
&= e^{\beta_0 + \beta_1 A + \beta_2 M + \boldsymbol{\beta}_3^T\boldsymbol{C}}(1 - e^{\beta_0 + \beta_1 A + \beta_2 M + \boldsymbol{\beta}_3^T\boldsymbol{C}}) + (e^{\beta_0^* + \beta_1^* A + \boldsymbol{\beta}_3^{*T}\boldsymbol{C}} - e^{\beta_0 + \beta_1 A + \beta_2 M + \boldsymbol{\beta}_3^T\boldsymbol{C}})^2 \\
&= d + (e^{\beta_0^* + \beta_1^* A + \boldsymbol{\beta}_3^{*T}\boldsymbol{C}} - e^{\beta_0 + \beta_1 A + \beta_2 M + \boldsymbol{\beta}_3^T\boldsymbol{C}})^2.
\end{aligned}
$$

Therefore, it is straightforward to observe that $\mathbb{E}(\boldsymbol{\epsilon}\boldsymbol{\epsilon}^T|A, M, \boldsymbol{C})$ is an explicit function of $\{A, M, \boldsymbol{C}\}$.

Appendix 2.5 In Case $Y_b M_c$, the ARE does not depend on the expectation or the variance of the mediator

We consider two settings: In Setting I, we consider $\mathbb{E}(M) = 0$ and $\text{Var}(M) = 1$; and in Setting II, we set $\mathbb{E}(M) = u_m$ and $\text{Var}(M) = v_m$, but all other intuitive parameters are otherwise the same as those in Setting I, where u_m and $v_m > 0$ are any numbers other than 0 and 1. If we can show the regression models in both settings are algebraically equivalent, then both settings will have the same AREs and we complete the proof.

Let $\boldsymbol{\beta}^{*(1)}$, $\boldsymbol{\beta}^{(1)}$, $\boldsymbol{\gamma}^{(1)}$, and $\sigma_3^{2^{(1)}}$ be the regression parameters correspond to setting I. And let $\boldsymbol{\beta}^{*(1)}$, $\boldsymbol{\beta}^{(2)}$, $\boldsymbol{\gamma}^{(2)}$, and $\sigma_3^{2^{(2)}}$ denote the regression parameters correspond to II. According to the relationships between the parametric relationships between regression and intuitive parameters shown in Web Appendix 2.2., we have

$$\gamma_0^{(2)} = u_m + \sqrt{v_m}\gamma_0^{(1)}; \gamma_1^{(2)} = \gamma_1^{(1)}\sqrt{v_m}; \sigma_3^{2^{(2)}} = v_m\sigma_3^{2^{(1)}};$$
$$\beta_1^{(2)} = \beta_1^{(1)}; \beta_2^{(2)} = \frac{\beta_2^{(1)}}{\sqrt{v_m}}; \beta_0^{(2)} = \beta_0^{(1)} - \frac{\beta_2^{(1)}u_m}{\sqrt{v_m}};$$
$$\beta_0^{*(2)} = \beta_0^{*(1)}; \beta_1^{*(2)} = \beta_1^{*(2)}.$$

Therefore, the three regression models used in Setting II is

$$\text{Model I: } \mathbb{E}[Y|A] = e^{\beta_0^{*(1)} + \beta_1^{*(1)}A},$$
$$\text{Model II: } \mathbb{E}[Y|A, M] = e^{\beta_0^{(1)} - \frac{\beta_2^{(1)}u_m}{\sqrt{v_m}} + \beta_1^{(1)}A + \frac{\beta_2^{(1)}}{\sqrt{v_m}}M},$$
$$\text{Model III: } M = u_m + \sqrt{v_m}\gamma_0^{(1)} + \gamma_1^{(1)}\sqrt{v_m}A + \epsilon_3, \quad \text{where } \epsilon_3 \sim N(0, v_m\sigma_3^{2^{(1)}}).$$

Next, let $M^* = \frac{M - u_m}{\sqrt{v_m}}$ and substitute it into the previous regression models, then we have

$$\text{Model I: } \mathbb{E}[Y|A] = e^{\beta_0^{*(1)} + \beta_1^{*(1)}A},$$
$$\text{Model II: } \mathbb{E}[Y|A, M^*] = e^{\beta_0^{(1)} + \beta_1^{(1)}A + \beta_2^{(1)}M^*},$$
$$\text{Model III: } M^* = \gamma_0^{(1)} + \gamma_1^{(1)}A + \epsilon_3^*, \quad \text{where } \epsilon_3^* \sim N(0, \sigma_3^{2^{(1)}}).$$

Notice that now M^* is mean zero and variance 1 and the above three models are actually the regression models used in Setting I. This indicates the regression models used in Settings I and II are algebraically equivalent and therefore share same variances for NIE and MP.

Web Appendix 3: Numerical studies of ARE in the presence of confounding

Appendix 3.1 Results

Here, we report the results of numerical studies to investigate how AREs change in the presence of a binary confounder M. We consider the following regression models

$$\begin{aligned}
&\mathbb{E}[Y|A,C] = e^{\beta_0^*+\beta_1^* A+\beta_3^* C},\\
&\mathbb{E}[Y|A,M,C] = e^{\beta_0+\beta_1 A+\beta_2 M+\beta_3 C},\\
&M = \gamma_0+\gamma_1 A+\gamma_2 C+\epsilon_3,\\
&P(A=1|C) = \frac{\exp(\psi_0+\psi_1 C)}{1+\exp(\psi_0+\psi_1 C)}
\end{aligned} \tag{s13}$$

where C is a binary confounder and $\epsilon_3 \sim N(0,\sigma_3^2)$. The last regression model defines A-C relationship. The intuitive parameters include those used previously (See Table 2 in the manuscript) and the following four parameters in relation to C: (i) the prevalence of C; (ii) the odds ratio of A for a change in C from 0 to 1, i.e., $OR(A|C) = e^{\phi_1}$; (iii) the correlation of M and C, $\mathrm{Corr}(C, M)$; and (iv) the risk ratio of Y for C corresponding to the change from 0 to 1, i.e., $RR(Y|C) = e^{\beta_3}$. The choices of the intuitive parameters were given in Web Table 1. We conducted a factorial design of all combinations of the intuitive parameters shown above. For each of combinations, we first calculated the regression parameters based on the intuitive parameters according to the procedures shown in Web Appendix 3.2, and then calculated the AREs based on the formulas shown in Web Appendix 2.3.

A total of 19,200 scenarios were included for comparison, where 5,319 (27.7%) of those scenarios were excluded from analysis because the intuitive parameters were not compatible. The detailed reasons were shown below

- 1. In 1,440 scenarios, the specified value of $\mathrm{Var}(M)$ was smaller than the specified value of $\mathrm{Var}(\gamma_0+\gamma_1 A+\gamma_2 C)$, which indicated that the variance of the error term in the mediator model, i.e., σ_3^2, was less than 0.
- 2. In 3,879 scenarios, $P(Y=1|A,M,C)$ may be greater than 1. Noticing that $P(Y=1|A,M,C) = e^{\beta_0+\beta_1 A+\beta_2 M+\beta_3 C}$ is unbounded with respect to the regression coefficients,

$P(Y = 1|A, M, C)$ can be greater than 1 if the regression coefficients are large and the values of A, M, and C are extreme. Therefore, we need exclude scenarios where $P(Y = 1|A, M, C)$ may be greater than 1 from our numerical analysis. Ideally, we need to substitute all possible combinations of A, M, and C into $P(Y = 1|A, M, C) = e^{\beta_0+\beta_1 A+\beta_2 M+\beta_3 C}$ to check if it is bigger than 1. If it is true, then we remove its corresponding scenario from analysis. However, $f(M|A = a, C = c)$ follows a normal distribution in Case $Y_b M_c$, therefore M can be any value between $(-\infty, +\infty)$, which can always make $P(Y = 1|A, M, C)$ greater than 1 if we numerate all possible M. However, since M near $-\infty$ or $+\infty$ is too rare to happen in practice, we consider all possible M lying within the 0.001% to 99.999% quantile of $M|A = a, C = c$, which should cover all possible values of M that could appear in a finite sample data set. Specifically, in the simulation study, we excluded the scenarios that $P(Y = 1|A = 0, M = m, C = 0)$, $P(Y = 1|A = 0, M = m, C = 1)$, $P(Y = 1|A = 1, M = m, C = 0)$, or $P(Y = 1|A = 1, M = m, C = 1)$ is greater than 1 for any possible m in (i) 0.001% to 99.999% quantile of the distribution of $M|A = 1, C = 1$, or (ii) 0.001% to 99.999% quantile of the distribution of $M|A = 1, C = 0$, or (iii) 0.001% to 99.999% quantile of the distribution of $M|A = 0, C = 1$, or (iv) 0.001% to 99.999% quantile of the distribution of $M|A = 0, C = 0$.

Thus, we calculated the AREs in the remaining 13,881 scenarios. The distributions of the AREs under those scenarios are shown in Web Table 2. The distributions of the AREs were slightly more dispersed compared to those without confounding effects, but most of the AREs were still near 1, where more than 95% of the ARE(NIE) and ARE(MP) were greater than 0.7 and 0.95, respectively. Moreover, we observed that increasing the confounder-mediator association decreases the AREs of the NIE estimator. The relationships between the AREs and the intuitive parameters are visualized in Web Figures 5 and 6 for aspects of NIE and MP, respectively. It is shown that changing the A-C and Y-C associations has minimal influence on the AREs, but strengthening the C-M association lowered the AREs, favoring the product method.

Appendix 3.2 Relationships between the intuitive parameters and regression parameters

The relationship between the parameters in regression models (s13) and the intuitive parameters $P(Y=1)$, TE, MP, $RR(Y|C)$, $\mathbb{E}(M)$, $\text{Var}(M)$, $\text{Corr}(M,A)$, $\text{Corr}(M,C)$, $P(A=1)$, $OR(A|C)$ are derived below.

First, we obtain ψ_0 and ψ_1 by solving

$$OR(A|C) = \exp(\psi_1) \quad \text{and}$$
$$P(A=1) = \tfrac{\exp(\psi_0+\psi_1)}{1+\exp(\psi_0+\psi_1)} P(C=1) + \tfrac{\exp(\psi_0)}{1+\exp(\psi_0)}(1-P(C=1)).$$

Next, we calculate γ_1 and γ_2 by solving the following two linear equations simultaneously,

$$\text{Corr}(M,A) = \tfrac{\gamma_1 \text{Var}(A) + \gamma_2 \text{Corr}(C,A)\sqrt{\text{Var}(A)\text{Var}(C)}}{\sqrt{\text{Var}(A)\text{Var}(M)}},$$
$$\text{Corr}(M,C) = \tfrac{\gamma_1 \text{Corr}(C,A)\sqrt{\text{Var}(A)\text{Var}(C)} + \gamma_2 \text{Var}(C)}{\sqrt{\text{Var}(A)\text{Var}(M)}},$$

where $\text{Corr}(C,A) = \frac{P(A=1,C=1)-P(A=1)P(C=1)}{\sqrt{\text{Var}(A)\text{Var}(M)}} = \frac{\frac{e^{\psi_0+\psi_1}}{1+e^{\psi_0+\psi_1}}P(C=1)-P(A=1)P(C=1)}{\sqrt{\text{Var}(A)\text{Var}(M)}}$, $\text{Var}(A) = P(A=1)(1-P(A=1))$ and $\text{Var}(C) = P(C=1)(1-P(C=1))$. Then, we obtain γ_0 and σ_3^2 by solving the following two equations

$$\mathbb{E}(M) = \gamma_0 + \gamma_1 P(A=1) + \gamma_2 P(C=1),$$
$$\text{Var}(M) = \gamma_1^2 \text{Var}(A) + \gamma_2^2 \text{Var}(C) + 2\gamma_1\gamma_2 \text{Corr}(C,A) \times \sqrt{\text{Var}(A)\text{Var}(M)} + \sigma_3^2.$$

Next, we obtain β_1, β_3, β_2 by solving

$$(1-\text{MP}) \times \text{TE} = \beta_1,$$
$$\text{MP} \times \text{NIE} = \beta_2\gamma_1,$$
$$RR(Y|C) = e^{\beta_3}.$$

Finally, after noticing

$$\begin{aligned}P(Y=1) &= \sum_{a=0}^{1}\sum_{c=0}^{1} P(A=a, C=c)\int P(Y=1|A=a, M=m, C=c) f(M=m|A=a, C=c) dm \\ &= \sum_{a=0}^{1}\sum_{c=0}^{1} P(A=a, C=c)\int e^{\beta_0+\beta_1 a+\beta_2 m+\beta_3 c}\frac{1}{\sqrt{2\pi}\sigma_3}e^{-\frac{(m-\gamma_0-\gamma_1 a-\gamma_2 c)^2}{2\sigma_3^2}} dm \\ &= \sum_{a=0}^{1}\sum_{c=0}^{1} P(A=a, C=c) e^{\beta_0+\beta_1 a+\beta_3 c+\beta_2(\gamma_0+\gamma_1 a+\gamma_2 c)+\frac{\beta_2^2\sigma_3^2}{2}},\end{aligned}$$

we obtain β_0 by solving

$$P(Y=1) = \sum_{a=0}^{1}\sum_{c=0}^{1} P(A=a, C=c)\exp\Big(\beta_0+\beta_1 a+\beta_3 c+\beta_2(\gamma_0+\gamma_1 a+\gamma_2 c)+\frac{\beta_2^2\sigma_3^2}{2}\Big),$$

where $P(A=a, C=c) = P(A=a|C=c)P(C=c) = \frac{a\exp(\psi_0+\psi_1 c)}{1+\exp(\psi_0+\psi_1 c)}P(C=c)$. Finally, β_0^*, β_1^*, and β_3^*, the regression parameters in the marginal outcome model, were obtained by the parametric relationships given in Web Appendix 1.3.

Web Appendix 4: An example that the difference method may be more robust than the product method when the exposure-mediator relationship is misspecified

Appendix 4.1 The example

We provide an example under Case Y_bM_c to illustrate that the product method has estimation bias if the exposure-mediator relationship is misspecified, but the difference method remains unbiased and can be more robust. In Case Y_bM_c, the product method requires the following mediator model,

$$M = \gamma_0 + \gamma_1 A + \boldsymbol{\gamma}_2^T \boldsymbol{C} + \epsilon_3,$$

where the error term, ϵ_3, follows a homoscedastic normal distribution $N(0, \sigma_3^2)$ such that $\text{Var}(\epsilon_3) = \text{Var}(\epsilon_3|A, \boldsymbol{C}) = \sigma_3^2$. In reality, $M|A, \boldsymbol{C}$ may be heteroskedastic such that $\text{Var}(\epsilon_3|A, \boldsymbol{C})$ is a function of A and $\boldsymbol{C}$. Assume now $\text{Var}(\epsilon_3|A, \boldsymbol{C}) = \eta_0 + \eta_1 A + \boldsymbol{\eta}_2^T \boldsymbol{C}$ but we misspecify the error term to follow a homoscedastic normal distribution. In Web Appendix 4.2, we prove that given the outcome model adjusted for the mediator (i.e., model (6) in manuscript) and the heteroskedastic mediator model, the outcome model without adjustment for the mediator (i.e., model (5) in manuscript) still hold, which indicates that the product and difference methods are compatible even under heteroskedasticity.

It is well known that heteroskedasticity will not affect the consistency of the ordinary least squares estimator; i.e., estimator of $\boldsymbol{\gamma} = [\gamma_0, \gamma_1, \boldsymbol{\gamma}_2^T]^T$ is still consistent even under misspecification of ϵ_3. In addition, regression coefficient estimators for the outcome models (5) and (6) are also unbiased as those models are correctly specified. However, we claim that $\widehat{\text{NIE}}^{(p)}_{a^*,a|\boldsymbol{c}}$, $\widehat{\text{TE}}^{(p)}_{a^*,a|\boldsymbol{c}}$, and $\widehat{\text{MP}}^{(p)}_{a^*,a|\boldsymbol{c}}$ are biased, though $\widehat{\text{NDE}}^{(p)}_{a^*,a|\boldsymbol{c}}$ is consistent. That is because the expression of the mediation measures given by the product method changes. Under the counterfactual framework, Web Appendix 4.2 shows that the correct expressions for the NIE and NDE are

$$\text{NIE}_{a^*,a|\boldsymbol{c}} = (\beta_2\gamma_1 + 0.5\beta_2^2\eta_1)(a - a^*) \quad \text{and} \quad \text{NDE}_{a^*,a|\boldsymbol{c}} = \beta_1(a - a^*),$$

where the NDE expression is exactly identical with its shown in Table 1 in manuscript but the NIE expression is different. Therefore, $\text{TE}_{a^*,a|\boldsymbol{c}} = (\beta_1 + \beta_2\gamma_1 + 0.5\beta_2^2\eta_1)(a - a^*)$ and $\text{MP}_{a^*,a|\boldsymbol{c}} = \frac{\beta_2\gamma_1 + 0.5\beta_2^2\eta_1}{\beta_1 + \beta_2\gamma_1 + 0.5\beta_2^2\eta_1}$. It follows that the asymptotic bias[1] of the NIE estimator given by the product method is

$$\begin{aligned}
ABIAS(\widehat{\text{NIE}}^{(p)}_{a^*,a|\boldsymbol{c}}) &= \lim_{n\to\infty} \widehat{\text{NIE}}^{(p)}_{a^*,a|\boldsymbol{c}} - \text{NIE}_{a^*,a|\boldsymbol{c}} \\
&= \lim_{n\to\infty} \hat{\beta}_2\hat{\gamma}_1(a - a^*) - (\beta_2\gamma_1 + 0.5\beta_2^2\eta_1)(a - a^*) \\
&= \beta_2\gamma_1(a - a^*) - (\beta_2\gamma_1 + 0.5\beta_2^2\eta_1)(a - a^*) \\
&= -0.5\beta_2^2\eta_1(a - a^*)
\end{aligned}$$

where the second to the third row holds since $\hat{\beta}_2$ and $\hat{\gamma}_1$ are consistent estimator of β_2 and γ_1.

[1]The asymptotic bias is defined as $\lim_{n\to\infty} \hat{p} - p$, where p is a mediation measure and $\lim_{n\to\infty} \hat{p}$ is its estimator's asymptotic limit when sample size is sufficient large.

Using the same strategy, we can show the asymptotic biases for the NDE, TE, and MP are

$$
\begin{aligned}
ABIAS(\widehat{\text{NDE}}^{(p)}_{a^*,a|\boldsymbol{c}}) &= 0, \\
ABIAS(\widehat{\text{TE}}^{(p)}_{a^*,a|\boldsymbol{c}}) &= -0.5\beta_2^2\eta_1(a-a^*), \\
ABIAS(\widehat{\text{MP}}^{(p)}_{a^*,a|\boldsymbol{c}}) &= -\frac{0.5\beta_1\beta_2^2\eta_1}{(\beta_1+\beta_2\gamma_1)(\beta_1+\beta_2\gamma_1+0.5\beta_2^2\eta_1)},
\end{aligned}
$$

respectively. From the above formula, we observe that the NDE estimator is always unbiased and biases of NIE and NDE estimators are functions of β_2, the mediator effect on the outcome, and η_1, degree of heteroskedasticity. The bias of the MP estimator is also affected by β_1 and γ_1. In Web Figure 8, we illustrate the asymptotic biases when varying β_2 and η_1. The NIE, TE, and MP present minimal bias when η_1 and β_2 are small but their biases dramatically increase when η_1 or β_2 increases.

On the other hand, the expressions of the mediation measures given by the difference methods are not affected (see Web Appendix 4.2). Since the two outcome mean models are also correctly specified under heteroskedasticity, the difference method can still provide consistent estimators for all mediation measures. That means the asymptotic biases for all the estimators given by the difference method are 0.

Of note, the bias analysis results shown here differ from the results in Cheng et al. (2021) [1]. In Cheng et al. (2021) [1], they found that the product method exhibited minimal bias even under heteroskedasticity in Case Y_bM_c. That is because they only evaluated the bias of the respective regression parameters but did not address the additional complication that the expressions of the mediation measures (e.g. NIE and MP) will change if the error term in the mediator model is heteroscedastic. In this study, we clarified that in Case Y_bM_c the expressions for the mediation measures given by the product method can be different when the error term in the mediator model is in fact heteroscedastic. Therefore, by analytically deriving the asymptotic relative bias formulas, we showed that the estimators for mediation measures given by the product method incorrectly assuming homoscedastic variance can subject to bias.

Appendix 4.2 Several analytic relationships for when $\epsilon_3 \sim N(0, \eta_0 + \eta_1 A + \eta_2^T \mathbf{C})$

i) In Case $Y_b M_c$, the product and difference methods are compatible when $\epsilon_3 \sim N(0, \eta_0 + \eta_1 A + \boldsymbol{\eta}_2^T \boldsymbol{C})$

.

Given the heteroskedastic mediator model and the conditional outcome model as shown in equation (6) in manuscript, we have that

$$
\begin{aligned}
\mathbb{E}[Y|A,\boldsymbol{C}] = \mathbb{E}\left[\mathbb{E}[Y|A,M,\boldsymbol{C}]|A,\boldsymbol{C}\right] &= e^{\beta_0+\beta_1 A+\boldsymbol{\beta}_3^T\boldsymbol{C}}\mathbb{E}[e^{\beta_2 M}|A,\boldsymbol{C}] \\
&= e^{\beta_0+\beta_1 A+\boldsymbol{\beta}_3^T\boldsymbol{C}} \int e^{\beta_2 M} \frac{1}{\sqrt{2\pi(\eta_0+\eta_1 A+\boldsymbol{\eta}_2^T\boldsymbol{C})}} e^{-\frac{\left(M-(\gamma_0+\gamma_1 A+\boldsymbol{\gamma}_2^T\boldsymbol{C})\right)^2}{2(\eta_0+\eta_1 A+\boldsymbol{\eta}_2^T\boldsymbol{C})}} \mathrm{d}M \\
&= e^{\beta_0+\beta_1 A+\boldsymbol{\beta}_3^T\boldsymbol{C}} \times e^{\beta_2(\gamma_0+\gamma_1 A+\boldsymbol{\gamma}_2^T\boldsymbol{C})+0.5\beta_2^2(\eta_0+\eta_1 A+\boldsymbol{\eta}_2^T\boldsymbol{C})} \\
&= e^{\beta_0+\beta_2\gamma_0+0.5\beta_2^2\eta_0+(\beta_1+\beta_2\gamma_1+0.5\beta_2^2\eta_1)A+(\boldsymbol{\beta}_3^T+\beta_2\boldsymbol{\gamma}_2^T+0.5\beta_2^2\boldsymbol{\eta}_2^T)\boldsymbol{C}},
\end{aligned}
$$

This concludes that the outcome model without adjustment for mediator (as shown in equation (5) in manuscript) still holds and $\beta_0^* = \beta_0 + \beta_2\gamma_0 + 0.5\beta_2^2\eta_0$, $\beta_1^* = \beta_1 + \beta_2\gamma_1 + 0.5\beta_2^2\eta_1$, and $\boldsymbol{\beta}_2^* = \boldsymbol{\beta}_3 + \beta_2\boldsymbol{\gamma}_2 + 0.5\beta_2^2\boldsymbol{\eta}_2$. This completes the proof.

ii) Expressions of the mediation measures under the product method when $\epsilon_3 \sim N(0, \eta_0 + \eta_1 A + \boldsymbol{\eta}_2^T \boldsymbol{C})$

.

Under the counterfactual framework, we have that

$$
\begin{aligned}
\mathbb{E}[Y(a,M(a))|\boldsymbol{C}=\boldsymbol{c}] &= \int P(Y=1|A=a,M=m,\boldsymbol{C}=\boldsymbol{c})f(M=m|A=a,\boldsymbol{C}=\boldsymbol{c})dm \\
&= \int e^{\beta_0+\beta_1 a+\beta_2 m+\boldsymbol{\beta}_3^T\boldsymbol{c}} \frac{1}{\sqrt{2\pi(\eta_0+\eta_1 a+\boldsymbol{\eta}_2^T\boldsymbol{c})}} e^{-\frac{\left(M-(\gamma_0+\gamma_1 a+\boldsymbol{\gamma}_2^T\boldsymbol{c})\right)^2}{2(\eta_0+\eta_1 a+\boldsymbol{\eta}_2^T\boldsymbol{c})}} dm \\
&= e^{\beta_0+\beta_2\gamma_0+0.5\beta_2^2\eta_0+(\beta_1+\beta_2\gamma_1+0.5\beta_2^2\eta_1)a+(\boldsymbol{\beta}_3^T+\beta_2\boldsymbol{\gamma}_2^T+0.5\beta_2^2\boldsymbol{\eta}_2^T)\boldsymbol{c}}.
\end{aligned}
$$

Similarly, we can show that

$$
\begin{aligned}
\mathbb{E}[Y(a, M(a^*))|\boldsymbol{C}=\boldsymbol{c}] &= \int P(Y=1|A=a, M=m, \boldsymbol{C}=\boldsymbol{c}) f(M=m|A=a^*, \boldsymbol{C}=\boldsymbol{c}) dm \\
&= e^{\beta_0+\beta_2\gamma_0+0.5\beta_2^2\eta_0+\beta_1 a+(\beta_2\gamma_1+0.5\beta_2^2\eta_1)a^*+(\boldsymbol{\beta}_3^T+\beta_2\boldsymbol{\gamma}_2^T+0.5\beta_2^2\boldsymbol{\eta}_2^T)\boldsymbol{c}}, \\
\mathbb{E}[Y(a^*, M(a^*))|\boldsymbol{C}=\boldsymbol{c}] &= \int P(Y=1|A=a^*, M=m, \boldsymbol{C}=\boldsymbol{c}) f(M=m|A=a^*, \boldsymbol{C}=\boldsymbol{c}) dm \\
&= e^{\beta_0+\beta_2\gamma_0+0.5\beta_2^2\eta_0+(\beta_1+\beta_2\gamma_1+0.5\beta_2^2\eta_1)a^*+(\boldsymbol{\beta}_3^T+\beta_2\boldsymbol{\gamma}_2^T+0.5\beta_2^2\boldsymbol{\eta}_2^T)\boldsymbol{c}}.
\end{aligned}
$$

Finally, by definition of NIE, we have that

$$
\begin{aligned}
\text{NIE}_{a^*,a|\boldsymbol{c}} &= \log\left(\mathbb{E}[Y(a, M(a))|\boldsymbol{C}=\boldsymbol{c}]\right) - \log\left(\mathbb{E}[Y(a, M(a^*))|\boldsymbol{C}=\boldsymbol{c}]\right) \\
&= (\beta_2\gamma_1 + 0.5\beta_2^2\eta_1)(a-a^*).
\end{aligned}
$$

Similarly, we have that $\text{NDE}_{a^*,a|\boldsymbol{c}} = \beta_1(a-a^*)$, $\text{TE}_{a^*,a|\boldsymbol{c}} = (\beta_1+\beta_2\gamma_1+0.5\beta_2^2\eta_1)(a-a^*)$ and $\text{MP}_{a^*,a|\boldsymbol{c}} = \frac{\beta_2\gamma_1+0.5\beta_2^2\eta_1}{\beta_1+\beta_2\gamma_1+0.5\beta_2^2\eta_1}$.

iii) The difference method presents correct mediation measure expressions when $\epsilon_3 \sim N(0, \eta_0 + \eta_1 A + \boldsymbol{\eta}_2^T \boldsymbol{C})$

.

First, because the NDE expression given by the product method is correct and same with the NDE expression given by difference method, the difference method offers correct NIE expression. Second, as derived in i), we have the following parametric link $\beta_1^* - \beta = \beta_2\gamma_1 + 0.5\beta_2^2\eta_1$. Therefore, the NIE expression under the difference method is $(\beta_1^* - \beta)(a-a^*) = (\beta_2\gamma_1 + 0.5\beta_2^2\eta_1)(a-a^*)$, which is exactly the $\text{NIE}_{a^*,a|\boldsymbol{c}}$ under heteroskedasticity (see derivations in ii)). It follows that the difference method provides correct NIE expression. By noticing TE and MP are functions of NIE and NDE, the difference method also provides correct TE and MP expressions.

Web Table and Figures

Table 1: Specifications of the intuitive parameters under Case $Y_b M_c$ in the presence of a binary confounder.

Intuitive Parameters	Notation	Values
Total effect (log risk ratio scale)	TE	$\log(1.2), \log(1.6), \log(2)$
Mediation proportion	MP	0.1, 0.3, 0.5, 0.7, 0.9
Prevalence of the outcome	$P(Y = 1)$	1%, 4%, 8%, 32%
Prevalence of the exposure	$P(A = 1)$	50%
Expectation of mediator	$\mathbb{E}(M)$	0
Variance of mediator	$\text{Var}(M)$	1
Correlation of the mediator and exposure	$\text{Corr}(A, M)$	0.2, 0.4, 0.6, 0.8
Prevalence of the confounder	$P(C = 1)$	50%
Odds ratio of the confounder on exposure	$OR(A\|C)$	1, 1.2, 1.5, 2
Correlation of the confounder and mediator	$\text{Corr}(C, M)$	0, 0.2, 0.4, 0.6, 0.8
Risk ratio of confounder on outcome	$RR(Y\|C)$	1, 1.2, 1.5, 2

Table 2: The AREs for NIE and MP under Case $Y_b M_c$ in the presence of a binary confounder. ($n = 13,881$ scenarios)

Index	Min	Percentiles							Max
		5%	10%	25%	50%	75%	90%	95%	
NIE	0.011	0.734	0.895	0.985	0.999	1.000	1.000	1.000	1.000
MP	0.100	0.967	0.989	0.999	1.000	1.000	1.000	1.000	1.000

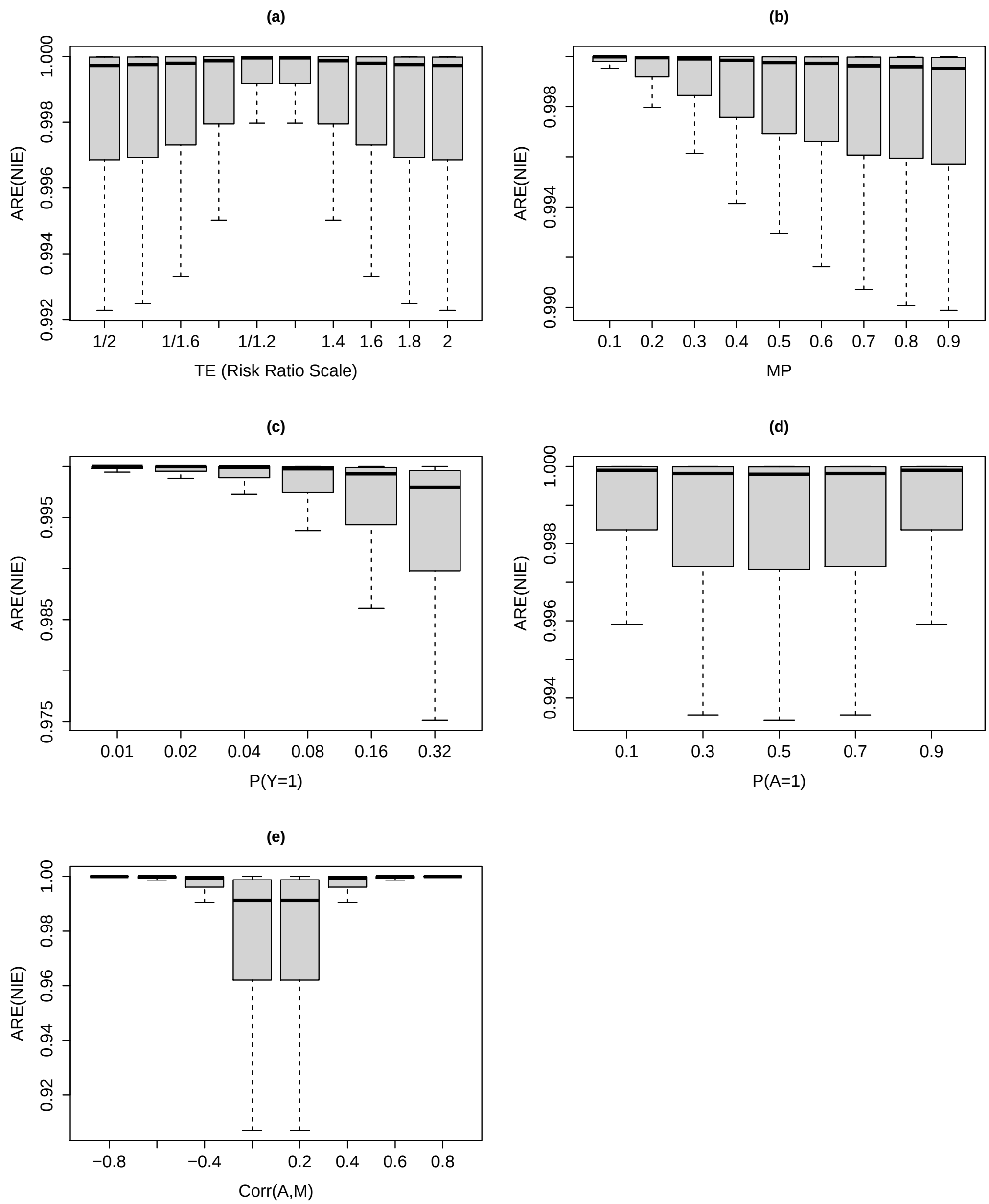

Figure 1: Relative Asymptotic Efficiency of $\widehat{\text{NIE}}$ (denoted as ARE(NIE)) under Case $Y_b M_c$ (binary outcome and continuous mediator). The outliers, defined as AREs outside 1.5 times the interquartile range above the third quartile and below the first quartile, are not shown [5].

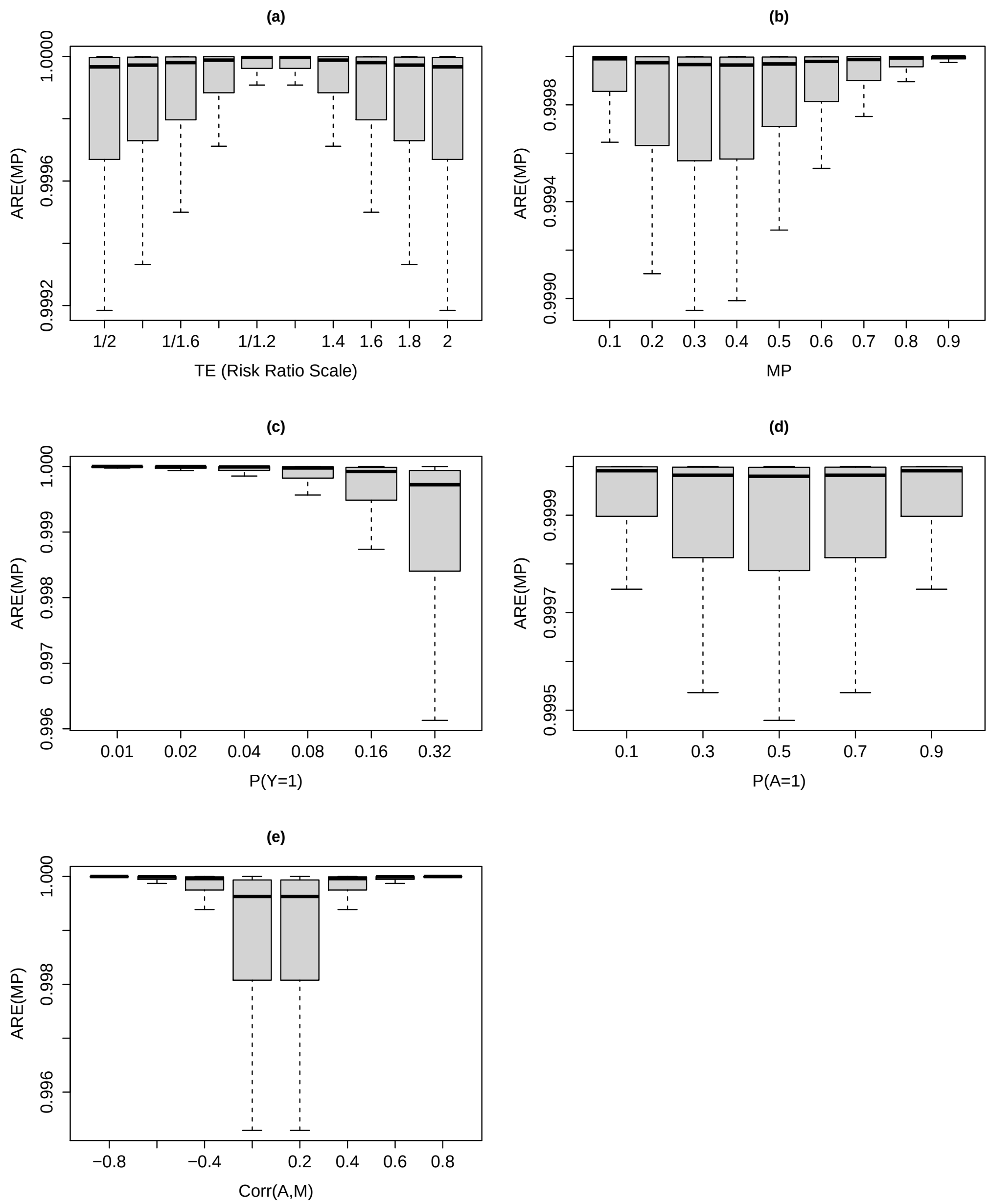

Figure 2: Relative Asymptotic Efficiency of $\widehat{\mathrm{MP}}$ (denoted as ARE(MP)) under Case $Y_b M_c$ (binary outcome and continuous mediator). The outliers, defined as AREs outside 1.5 times the interquartile range above the third quartile and below the first quartile, are not shown in the box plots.

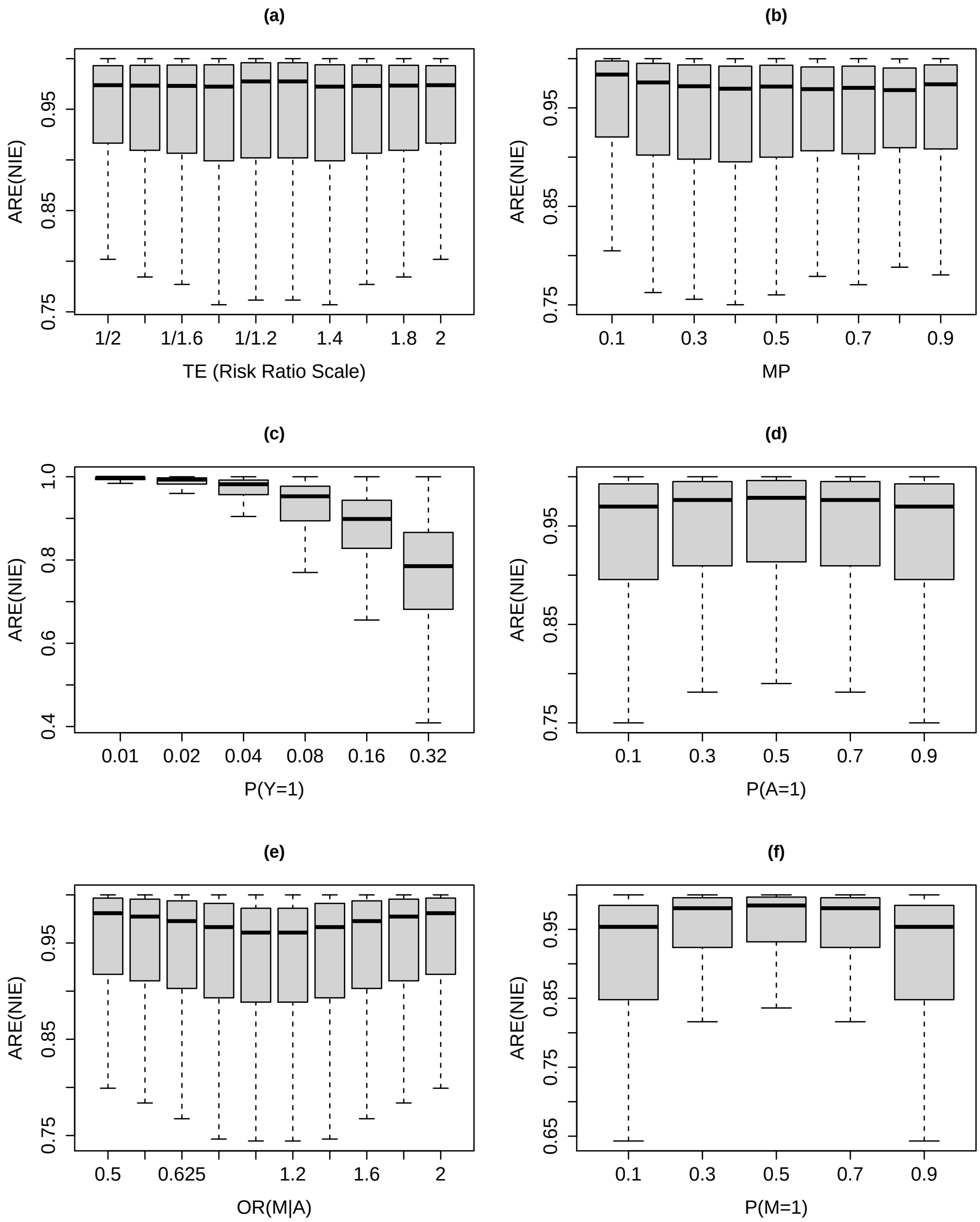

Figure 3: Relative Asymptotic Efficiency of $\widehat{\text{NIE}}$ (denoted as ARE(NIE)) under Case $Y_b M_b$(binary outcome and binary mediator). The outliers, defined as AREs outside 1.5 times the interquartile range above the third quartile and below the first quartile, are not shown [5].

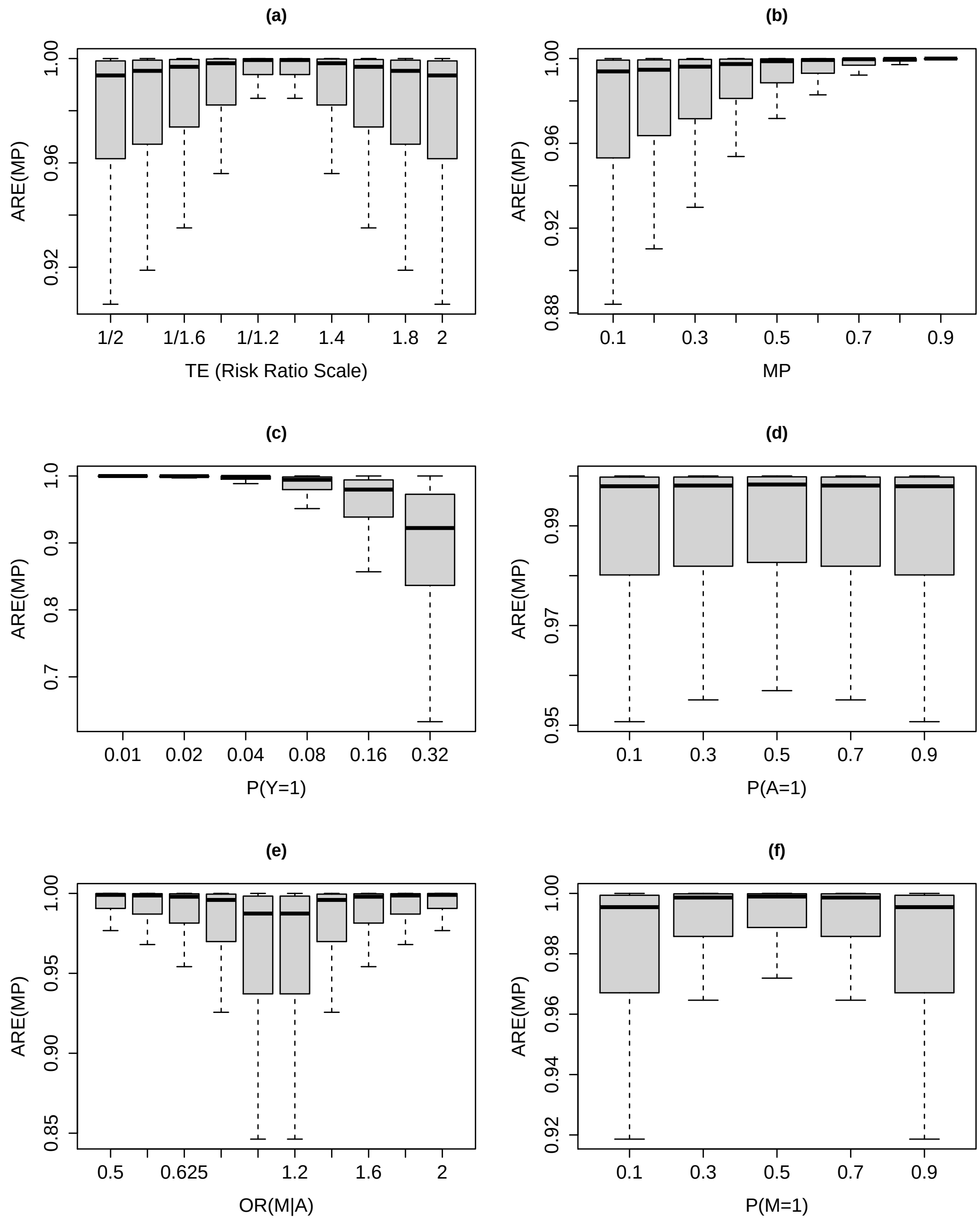

Figure 4: Relative Asymptotic Efficiency of $\widehat{\text{MP}}$ (denoted as ARE(MP)) under Case Y_bM_b(binary outcome and binary mediator). The outliers, defined as AREs outside 1.5 times the interquartile range above the third quartile and below the first quartile, are not shown [5].

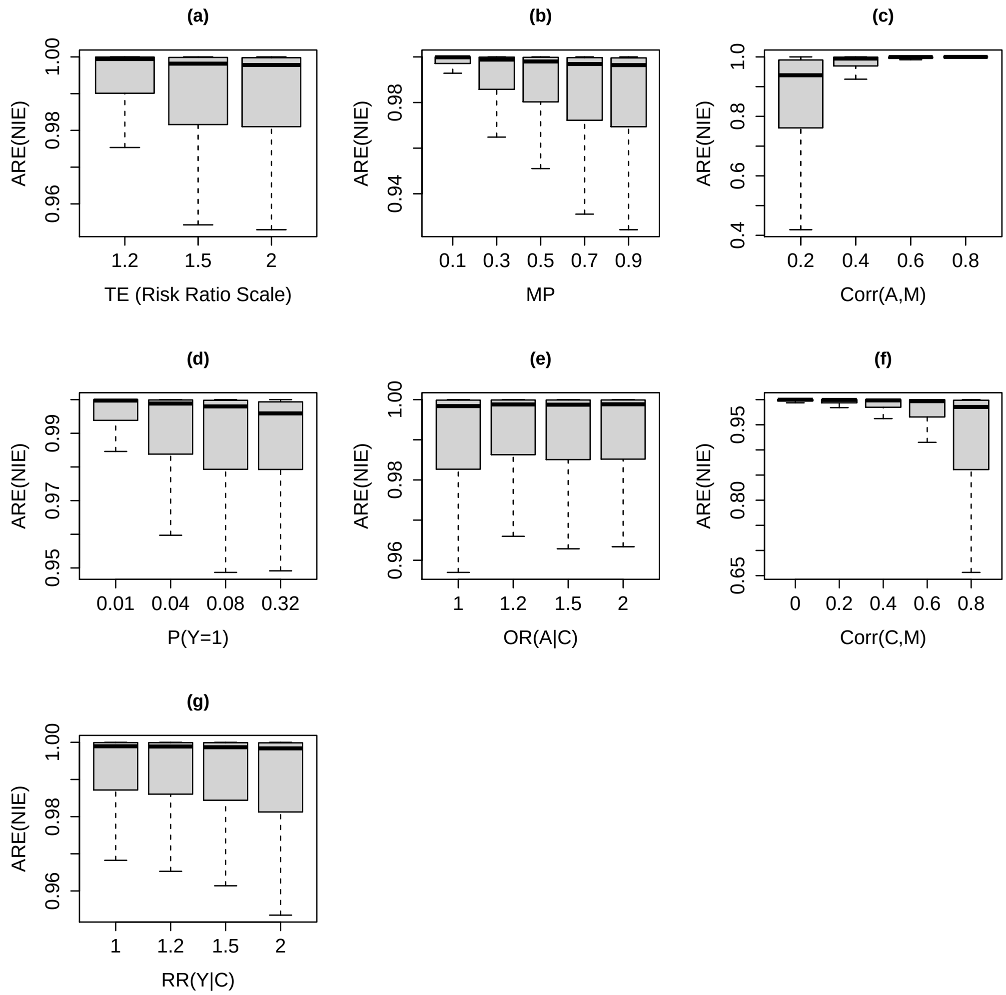

Figure 5: Relative Asymptotic Efficiency of $\widehat{\text{NIE}}$ (denoted as ARE(NIE)) under Case Y_bM_c (binary outcome and continuous mediator) in the presence of a binary confounder. The outliers, defined as AREs outside 1.5 times the interquartile range above the third quartile and below the first quartile, are not shown [5].

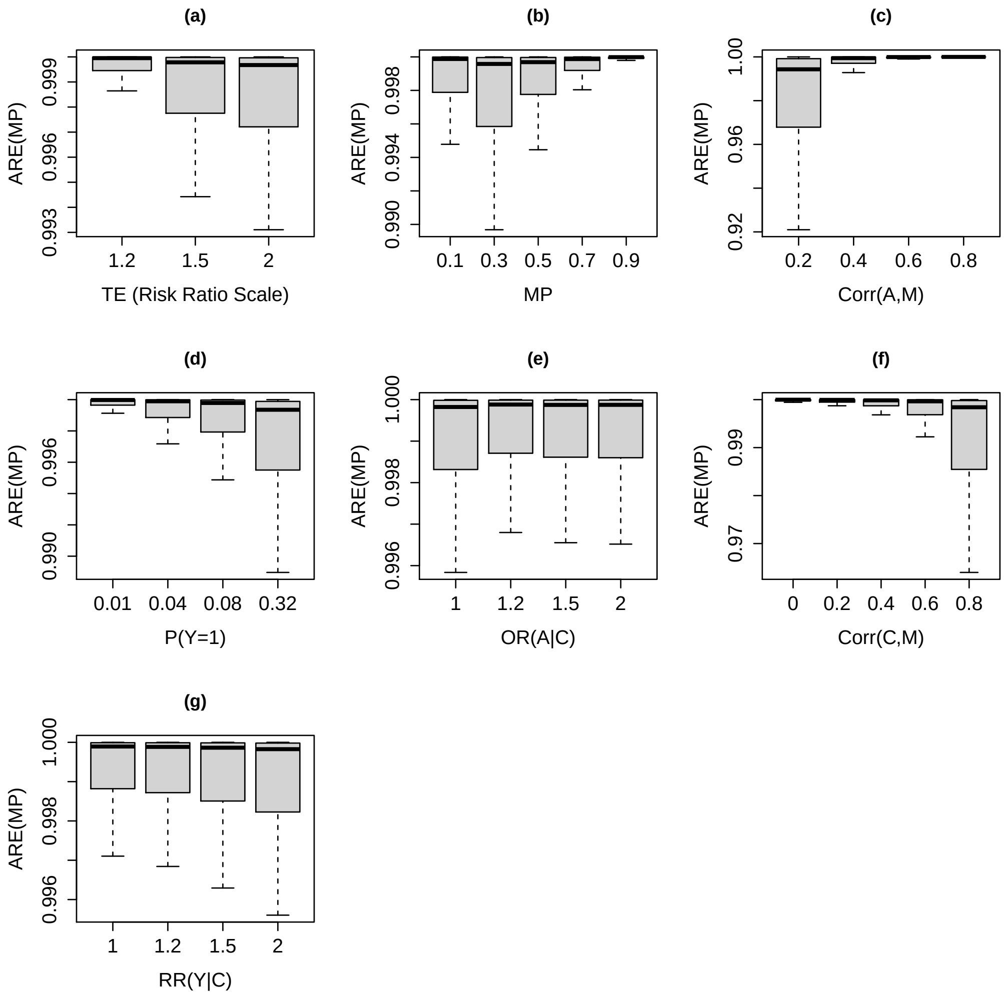

Figure 6: Relative Asymptotic Efficiency of $\widehat{\text{MP}}$ (denoted as ARE(MP)) under Case $Y_b M_c$ (binary outcome and continuous mediator) in the presence of a binary confounder. The outliers, defined as AREs outside 1.5 times the interquartile range above the third quartile and below the first quartile, are not shown [5].

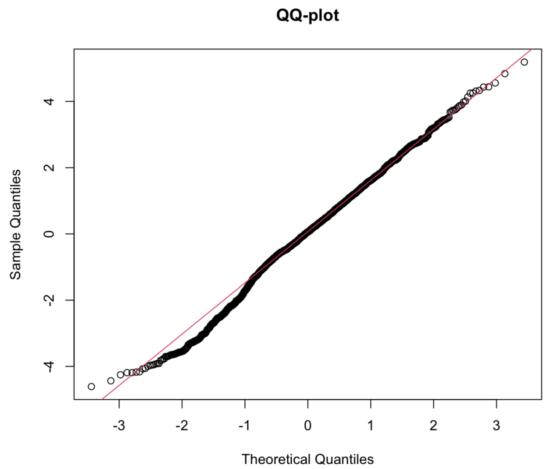

Figure 7: QQ-plot for the residuals in the linear regression model for the 6-month visit adherence, MaxART [2]

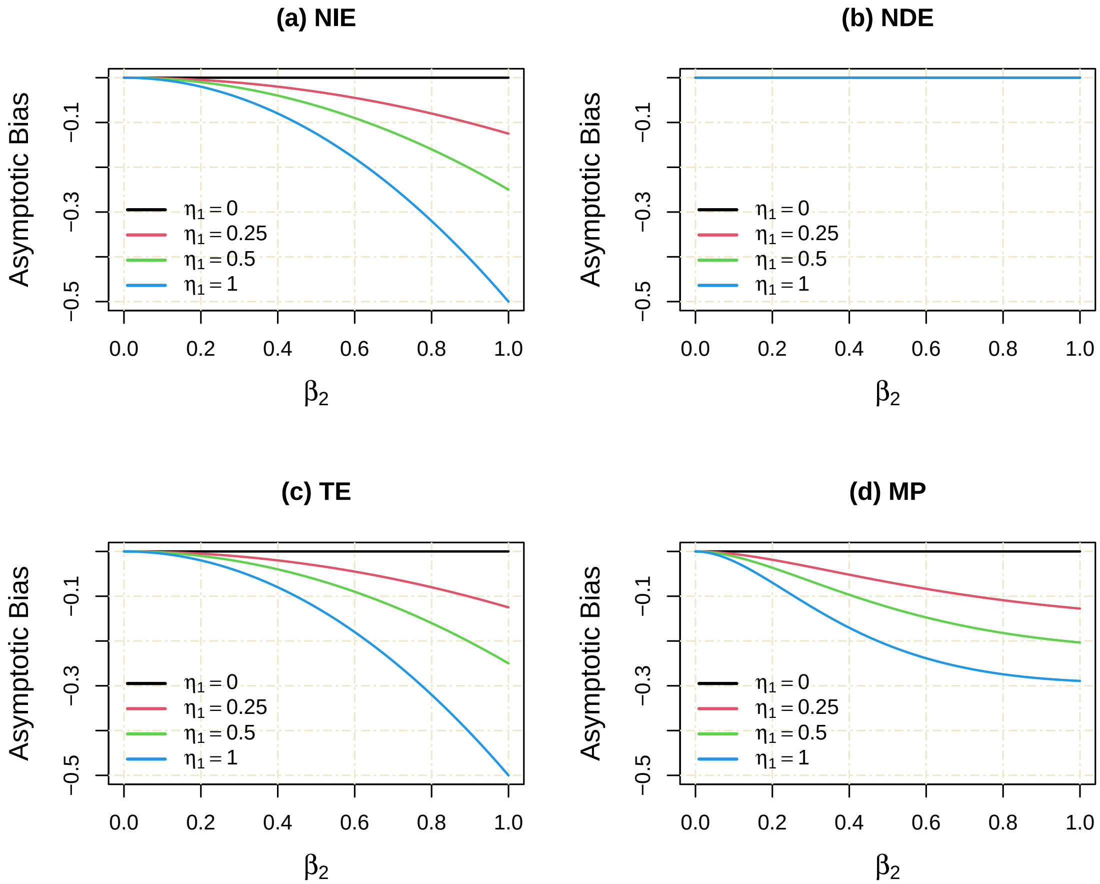

Figure 8: Asymptotic biases of the NIE, NDE, TE, and MP estmators given by the product method when varying η_1 and β_2 from 0 to 1. The mediation measures were evaluated for one unit increase of the exposure. Other parameters affecting the asmptotic bias of MP, including β_1 and η_1, were fixed to log(1.2).

References

[1] C. Cheng, D. Spiegelman, and F. Li. Estimating the natural indirect effect and the mediation proportion via the product method. BMC Medical Research Methodology, 21(1):1–20, 2021.

[2] S. Khan, D. Spiegelman, F. Walsh, S. Mazibuko, M. Pasipamire, B. Chai, R. Reis, K. Mlambo,

W. Delva, G. Khumalo, et al. 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, 23(9):e25610, 2020.

[3] K.-Y. Liang and S. L. Zeger. Longitudinal data analysis using generalized linear models. Biometrika, 73(1):13–22, 1986.

[4] D. P. MacKinnon, G. Warsi, and J. H. Dwyer. A simulation study of mediated effect measures. Multivariate Behavioral Research, 30(1):41–62, 1995.

[5] F. Mosteller, J. W. Tukey, et al. Data analysis and regression: a second course in statistics. 1977.

[6] L. A. Stefanski and D. D. Boos. The calculus of m-estimation. The American Statistician, 56 (1):29–38, 2002.