Supplement for "Effect Measure Modification by Covariates in Mediation: Extending the Regression-Based Closed-Form Causal Mediation Analysis"

Yi Li[1,2], Maya B. Mathur[3], Daniel H. Solomon[2,4,5], Paul M. Ridker[4,6,7], Robert J. Glynn[4,5,6,8], and Kazuki Yoshida[2,4]

[1]Department of Epidemiology, Biostatistics and Occupational Health, School of Population and Global Health, Faculty of Medicine, McGill University, Montreal, QC, Canada.
[2]Division of Rheumatology, Inflammation, and Immunity, Department of Medicine, Brigham and Women's Hospital, Boston, MA, USA.
[3]Quantitative Science Unit, Department of Medicine, Stanford University, Palo Alto, CA, USA.
[4]Department of Medicine, Harvard Medical School, Boston, MA, USA.
[5]Division of Pharmacoepidemiology and Pharmacoeconomics, Department of Medicine, Brigham and Women's Hospital, Boston, MA, USA.
[6]Division of Preventive Medicine, Department of Medicine, Brigham and Women's Hospital, Boston, MA, USA.
[7]Division of Cardiovascular Medicine, Department of Medicine, Brigham and Women's Hospital, Boston, MA, USA.
[8]Department of Biostatistics, Harvard T.H. Chan School of Public Health, Boston, MA, USA.

Contents

1 Brief overview of causal mediation analysis

The literature on causal mediation analysis is vast (VanderWeele, 2015). Thus, only the pieces relevant for the current software are reviewed here. This section discusses the decompositions of total treatment effect, controlled effect, natural effect, and identification assumptions.

1.1 Decomposition of total effect

Let Y be the outcome variable of interest, A be the treatment variable of interest, M be the mediator variable of interest, and $\mathbf{C}$ be a vector of the covariates that need to be adjusted. The treatment contrast of interest is a vs a^*, where a^* is the baseline/reference level. The potential outcome $Y_{a,m}$ is the value of Y for an individual when, possibly contrary to the fact, if the treatment level is a and mediator level is m.

On the difference scale, the total effect (TE), the causal effect of moving the treatment level from the reference level a^* to the level of interest a, is defined as follows at the covariate level $\mathbf{C} = \mathbf{c}$:

$$TE = \mathrm{E}[Y_a|\mathbf{C} = \mathbf{c}] - \mathrm{E}[Y_{a^*}|\mathbf{C} = \mathbf{c}]$$

The conditional direct effect ($CDE(m)$), the effect of moving the treatment level from the reference level a^* to the level of interest a while fixing the mediator at m is the following at the covariate level $\mathbf{C} = \mathbf{c}$:

$$CDE(m) = \mathrm{E}[Y_{a,m}|\mathbf{C} = \mathbf{c}] - \mathrm{E}[Y_{a^*,m}|\mathbf{C} = \mathbf{c}]$$

TE can be decomposed into direct effect and indirect effect in two different ways (Robins & Greenland, 1992), although the sum of pure natural direct effect (PNDE) and total natural indirect effect (TNIE) is the more usual way of decomposition (Pearl, 2001). Note that the treatment value indexing the mediator M is fixed at a^* in the PNDE, whereas the treatment value indexing the outcome Y is fixed at a in the TNIE.

$$PNDE = \mathrm{E}[Y_{a,\utilde{M_{a^*}}}|\mathbf{C} = \mathbf{c}] - \mathrm{E}[Y_{a^*,\utilde{M_{a^*}}}|\mathbf{C} = \mathbf{c}]$$
$$TNIE = \mathrm{E}[Y_{a,M_a}|\mathbf{C} = \mathbf{c}] - \mathrm{E}[Y_{a,M_{a^*}}|\mathbf{C} = \mathbf{c}]$$

For a general link function g (e.g., the logit function), these expressions on the difference scale generalize to(Valeri & VanderWeele, 2013; VanderWeele, 2015; Lange, Vansteelandt, & Bekaert, 2012):

$$TE = g(\mathrm{E}[Y_a|\mathbf{C} = \mathbf{c}]) - g(\mathrm{E}[Y_{a^*}|\mathbf{C} = \mathbf{c}])$$
$$CDE(m) = g(\mathrm{E}[Y_{a,m}|\mathbf{C} = \mathbf{c}]) - g(\mathrm{E}[Y_{a^*,m}|\mathbf{C} = \mathbf{c}])$$
$$PNDE = g(\mathrm{E}[Y_{a,\utilde{M_{a^*}}}|\mathbf{C} = \mathbf{c}]) - g(\mathrm{E}[Y_{a^*,\utilde{M_{a^*}}}|\mathbf{C} = \mathbf{c}])$$
$$TNIE = g(\mathrm{E}[Y_{a,M_a}|\mathbf{C} = \mathbf{c}]) - g(\mathrm{E}[Y_{a,M_{a^*}}|\mathbf{C} = \mathbf{c}])$$

The other decomposition of TE is the sum of pure natural indirect effect (PNIE) and total natural direct effect (TNDE) (Robins & Greenland, 1992). Note that the treatment value indexing the mediator M is fixed at a in the TNDE, whereas the treatment value indexing the outcome Y is fixed at a^* in the PNIE.

$$TNDE = g(\mathrm{E}[Y_{a,\utilde{M_a}}|\mathbf{C} = \mathbf{c}]) - g(\mathrm{E}[Y_{a^*,\utilde{M_a}}|\mathbf{C} = \mathbf{c}])$$
$$PNIE = g(\mathrm{E}[Y_{\utilde{a^*},M_a}|\mathbf{C} = \mathbf{c}]) - g(\mathrm{E}[Y_{\utilde{a^*},M_{a^*}}|\mathbf{C} = \mathbf{c}])$$

More intuitively, the subtle difference between two decompositions is in the cross-world counterfactual state (the subscripts in the definition of natural direct and indirect effects). In natural direct effects, the treatment levels moves from a^* to a, and the mediator level is fixed. In PNDE, the mediator level is fixed to the natural level that it would be if treatment level is the reference level (a^*), while in TNDE, the mediator level is fixed to the natural level that it would be if treatment level is the level of interest (a^*). In natural indirect effects, the treatment level is fixed, and the mediator level changes. In TNIE, the treatment level is fixed to the level of interest (a), while in PNIE, the treatment level is fixed to the reference level (a^*).

In either case, the effect that has the component of $Y_{a^*} = Y_{a^*,M_{a^*}}$ is the pure effect (PNDE and PNIE) and the effect that has the component of $Y_a = Y_{a,M_a}$ is the total natural effect (TNDE and TNIE). See (VanderWeele, 2013, 2014) for the meaning of these two decompositions in terms of causal interaction.

1.2 Identification of natural effects

Several conditional exchangeabilities must be assumed for identification of effects in the causal mediation framework. See (VanderWeele, 2015) (p463) for details.

$$\begin{aligned}
&\mathbf{A1}.\ Y_{a,m} \perp\!\!\!\perp A|\mathbf{C}\\
&\mathbf{A2}.\ Y_{a,m} \perp\!\!\!\perp M|\{A,\mathbf{C}\}\\
&\mathbf{A3}.\ M_a \perp\!\!\!\perp A|\mathbf{C}\\
&\mathbf{A4}.\ Y_{a,m} \perp\!\!\!\perp M_{a^*}|\mathbf{C}
\end{aligned}$$

Particularly, the cross-world **A4** means that there is no $M - Y$ confounder that is affected by A.

The controlled direct effect (CDE) can be identified with only A1 and A2. The identification of natural effects (PNDE, TNIE, TNDE, PNIE) require all four assumptions.

Given these four assumptions, the mean counterfactual with different treatment values indexing the outcome Y and the mediator M can be identified as follows. For similicity, we assume M is categorical. For a continuous M, $\sum$ is replaced with $\int$, and $\mathrm{P}(M = m)$ is replaced with $\mathrm{d}F(M)$.

$$\begin{aligned}
&\mathrm{E}[Y_{a,M_{a^*}} \mid \mathbf{C}=\mathbf{c}] = \mathrm{E}\left[\mathrm{E}(Y_{a,M_{a^*}} \mid M_{a^*}, \mathbf{C}=\mathbf{c}) \mid \mathbf{C}=\mathbf{c}\right] && \text{(Law of iterative expectation)}\\
&= \sum_m \mathrm{E}[Y_{a,M_{a^*}} \mid M_{a^*}=m, \mathbf{C}=\mathbf{c}]\mathrm{P}(M_{a^*}=m \mid \mathbf{C}=\mathbf{c})\\
&= \sum_m \mathrm{E}[Y_{a,m} \mid M_{a^*}=m, \mathbf{C}=\mathbf{c}]\mathrm{P}(M_{a^*}=m \mid \mathbf{C}=\mathbf{c}) && (M_{a^*}=m)\\
&= \sum_m \mathrm{E}[Y_{a,m} \mid \mathbf{C}=\mathbf{c}]\mathrm{P}(M_{a^*}=m \mid \mathbf{C}=\mathbf{c}) && (\text{A4. } Y_{a,m} \perp\!\!\!\perp M_{a^*} \mid \mathbf{C})\\
&= \sum_m \mathrm{E}[Y_{a,m} \mid \mathbf{C}=\mathbf{c}]\mathrm{P}(M_{a^*}=m \mid A=a^*, \mathbf{C}=\mathbf{c}) && (\text{A3. } M_a \perp\!\!\!\perp A \mid \mathbf{C})\\
&= \sum_m \mathrm{E}[Y_{a,m} \mid \mathbf{C}=\mathbf{c}]\mathrm{P}(M=m \mid A=a^*, \mathbf{C}=\mathbf{c}) && (\text{Consistency for } M)\\
&= \sum_m \mathrm{E}[Y_{a,m} \mid A=a, \mathbf{C}=\mathbf{c}]\mathrm{P}(M=m \mid A=a^*, \mathbf{C}=\mathbf{c}) && (\text{A1. } Y_{a,m} \perp\!\!\!\perp A \mid \mathbf{C})\\
&= \sum_m \mathrm{E}[Y_{a,m} \mid A=a, M=m, \mathbf{C}=\mathbf{c}]\mathrm{P}(M=m \mid A=a^*, \mathbf{C}=\mathbf{c}) && (\text{A3. } Y_{a,m} \perp\!\!\!\perp M \mid \{A,\mathbf{C}\})\\
&= \sum_m \mathrm{E}[Y \mid A=a, M=m, \mathbf{C}=\mathbf{c}]\mathrm{P}(M=m \mid A=a^*, \mathbf{C}=\mathbf{c}) && (\text{Consistency for } Y)\\
&= \mathrm{E}\left[\mathrm{E}(Y \mid A=a, M, \mathbf{C}=\mathbf{c}) \mid A=a^*, \mathbf{C}=\mathbf{c}\right]
\end{aligned}$$

The resulting short-hand identification formula for the counterfactual expectation is the following:

$$\mathrm{E}[Y_{a,M_{a^*}} \mid \mathbf{c}] = \mathrm{E}\left[\mathrm{E}(Y \mid a, M, \mathbf{c}) \mid a^*, \mathbf{c}\right]$$

Similarly, we obtain expressions for the other three potential outcomes:

$$\mathrm{E}[Y_{a,M_a} \mid \mathbf{C}=\mathbf{c}] = \mathrm{E}\left[\mathrm{E}(Y \mid a, M, \mathbf{c}) \mid a, \mathbf{c}\right]$$

$$\mathrm{E}[Y_{a^*,M_{a^*}} \mid \mathbf{C} = \mathbf{c}] = \mathrm{E}\left[\mathrm{E}(Y \mid a^*, M, \mathbf{c}) \mid a^*, \mathbf{c}\right]$$

$$\mathrm{E}[Y_{a,M_{a^*}} \mid \mathbf{C} = \mathbf{c}] = \mathrm{E}\left[\mathrm{E}(Y \mid a, M, \mathbf{c}) \mid a^*, \mathbf{c}\right]$$

2 Effect measure modification (EMM)-extended regression-based causal mediation analysis method

2.1 Mediator and outcome model specifications

Two models are involved in the identification of natural effects: the outcome model $\mathrm{E}[Y|A=a, M=m, \mathbf{C}=\mathbf{c}]$ and the mediator model $\mathrm{P}(M|A=a, \mathbf{C}=\mathbf{c})$. The identification formulas do not specify any particular model structure (non-parametric). In the method described in (VanderWeele, 2015; Valeri & VanderWeele, 2013, 2015), a simple parametric model is proposed for each.

The mediator model with a link function g_M is parametrized as follows:

$$g_M(E[M|A=a, \mathbf{C}=\mathbf{c}]) = \beta_0 + \beta_1 a + \boldsymbol{\beta}_2^T \mathbf{c} + \boldsymbol{\beta}_3^T a\mathbf{c}.$$

Note that the mediator model can only be either linear or logistic, regardless of the prevalence of the mediator (VanderWeele, 2015).
The outcome model with a link function g_Y is parametrized as follows:

$$g_Y(E[Y|A=a, M=m, \mathbf{C}=\mathbf{c}]) = \theta_0 + \theta_1 a + \theta_2 m + \theta_3 am + \boldsymbol{\theta}_4^T \mathbf{c} + \boldsymbol{\theta}_5^T a\mathbf{c} + \boldsymbol{\theta}_6^T m\mathbf{c}.$$

Since $\mathbf{c}$ is a vector, $\beta_2,\ \beta_3,\ \theta_4,\ \theta_5,\ \theta_6$ are vectors as well.

Under these parametric modeling assumptions, each effect of interest can be written as a function of the parameters (coefficients) of the the mediator model (β, σ^2) and the outcome model (θ).

Note: the non-linear g_Y can accommodate logistic, log-linear, Poisson, negative binomial, accelerated failure time (AFT), and Cox proportional hazards models. It is just that the scale of effect estimates varies by the link function. The following table shows the scale of the effect estimates under each type of outcome model (Valeri & VanderWeele, 2013, 2015).

Table 1: Outcome model and corresponding effect scale

Outcome model	Scale
Logistic (binary)	Odds ratio
Log-linear (binary)	Risk ratio
Poisson, negative binomial	Rate ratio
Accelerated failure time	Mean survival ratio
Cox's proportional hazards	Hazard ratio

2.2 Proportion mediated

Proportion mediated (PM) is defined as follows.
For linear outcome model:

$$PM = \frac{TNIE}{PNDE + TNIE}$$

For non-linear outcome model:

$$PM = \frac{\exp(PNDE) \cdot [\exp(TNIE) - 1]}{\exp(PNDE) \cdot \exp(TNIE) - 1}$$

Please note that when PNDE and TNIE are in opposite directions, PM may be outside the range of $[0, 1]$ and is therefore not interpretable.

2.3 Standard errors via multivariate delta method

We can obtain the standard error of each effect estimate using the variance-covariance matrix for the coefficients and multivariate delta method (Ver Hoef, 2012).

Let the scalar quantity of interest be Q, a function of parameter vector $(\boldsymbol{\beta}^T, \boldsymbol{\theta}^T)^T$. Then, its gradient (vector of partial derivatives) with respect to the parameter vector $(\boldsymbol{\beta}^T, \boldsymbol{\theta}^T)^T$ is the following.

$$\nabla Q = \frac{\partial Q}{\partial(\boldsymbol{\beta}^T, \boldsymbol{\theta}^T)^T} = \begin{bmatrix} \frac{\partial Q}{\partial \beta_0} \\ \frac{\partial Q}{\partial \beta_1} \\ \frac{\partial Q}{\partial \boldsymbol{\beta}_2} \\ \frac{\partial Q}{\partial \boldsymbol{\beta}_3} \\ \frac{\partial Q}{\partial \theta_0} \\ \frac{\partial Q}{\partial \theta_1} \\ \frac{\partial Q}{\partial \theta_2} \\ \frac{\partial Q}{\partial \theta_3} \\ \frac{\partial Q}{\partial \boldsymbol{\theta}_4} \\ \frac{\partial Q}{\partial \boldsymbol{\theta}_5} \\ \frac{\partial Q}{\partial \boldsymbol{\theta}_6} \end{bmatrix}$$

In the case of a linear mediator model and a non-linear outcome model, there is an additional element $\frac{\partial Q}{\partial \sigma^2}$ at the bottom of the gradient vector.

By the large sample approximation using the multivariate delta method, the variance of the quantity of interest evaluated at the MLEs $(\hat{\boldsymbol{\beta}}^T, \hat{\boldsymbol{\theta}}^T)^T$ is the following.

$$\underbrace{Var\left[Q\left\{(\hat{\boldsymbol{\beta}}^T, \hat{\boldsymbol{\theta}}^T)^T\right\}\right]}_{\text{scalar}} \approx \underbrace{\left[\nabla Q\left((\hat{\boldsymbol{\beta}}^T, \hat{\boldsymbol{\theta}}^T)^T\right)\right]^T}_{\text{row vector}} \underbrace{Var((\hat{\boldsymbol{\beta}}^T, \hat{\boldsymbol{\theta}}^T)^T)}_{\text{matrix}} \underbrace{\left[\nabla Q\left((\hat{\boldsymbol{\beta}}^T, \hat{\boldsymbol{\theta}}^T)^T\right)\right]}_{\text{column vector}}$$

This expression is abbreviated as $\Gamma\Sigma\Gamma'$ in (VanderWeele, 2015; Valeri & VanderWeele, 2013, 2015). In these references, the treatment contrast $(a - a^*)$ is factored out from $\nabla Q\left((\hat{\boldsymbol{\beta}}^T, \hat{\boldsymbol{\theta}}^T)^T\right)$ when possible. In the following, we define Γ as a column vector to be consistent with the implementation of `regmedint`, thus, the corresponding expression appears as $\Gamma^T\Sigma\Gamma$.

$\Gamma_{PNDE}, \Gamma_{TNIE}, \Gamma_{TNDE}, \Gamma_{PNIE}$ can be calculated by straightforward partial derivatives with respect to $(\boldsymbol{\beta}^T, \boldsymbol{\theta}^T)^T$. Γ_{TE} and Γ_{PM} can be calculated by $\frac{\partial TE}{\partial(\beta^T, \theta^T)^T}$ and $\frac{\partial PM}{\partial(\beta^T, \theta^T)^T}$, where

$$\frac{\partial TE}{\partial(\beta^T, \theta^T)^T} = \frac{\partial(PNDE + TNIE)}{\partial(\boldsymbol{\beta}^T, \boldsymbol{\theta}^T)^T} = \frac{\partial PNDE}{\partial(\boldsymbol{\beta}^T, \boldsymbol{\theta}^T)^T} + \frac{\partial TNIE}{\partial(\boldsymbol{\beta}^T, \boldsymbol{\theta}^T)^T}$$

$$\frac{\partial PM}{\partial(\boldsymbol{\beta}^T, \boldsymbol{\theta}^T)^T} = \frac{\partial PM}{\partial PNDE}\frac{\partial PNDE}{\partial(\boldsymbol{\beta}^T, \boldsymbol{\theta}^T)^T} + \frac{\partial PM}{\partial TNIE}\frac{\partial TNIE}{\partial(\boldsymbol{\beta}^T, \boldsymbol{\theta}^T)^T}$$

$$\boldsymbol{\Sigma} = \begin{bmatrix} \boldsymbol{\Sigma}_{\boldsymbol{\beta}} & 0 \\ 0 & \boldsymbol{\Sigma}_{\boldsymbol{\theta}} \end{bmatrix}$$

More calculation details are given in Section 3.

3 Estimation formulas for four model types

In this section, we give formulas for CDE(m), PNDE, TNIE, TNDE, PNIE, TE, PM. Proofs and derivations are included.

Please note that to be succinct in presenting the formulas, in this whole section, we only consider the case where there is only one covariate in the covariate vector $\mathbf{C}$. Therefore $\beta_{\mathbf{2}}{}^T$, $\beta_{\mathbf{3}}{}^T$, $\theta_{\mathbf{4}}{}^T$, $\theta_{\mathbf{5}}{}^T$ and $\theta_{\mathbf{6}}{}^T$ are all scalars, so we omit the transpose symbols.

3.1 Linear mediator model, linear outcome model

3.1.1 Point estimate

$$\begin{aligned} \mathrm{E}[M \mid a, c] &= \beta_0 + \beta_1 a + \beta_2 c + \beta_3 ac \\ \mathrm{E}[Y \mid a, m, c] &= \theta_0 + \theta_1 a + \theta_2 m + \theta_3 am + \theta_4 c + \theta_5 ac + \theta_6 mc \end{aligned}$$

$$CDE(m) = (\theta_1 + \theta_3 m + \theta_5 c)(a - a^*) \tag{1}$$

$$\begin{aligned} PNDE &= \mathrm{E}[Y_{a,M^{a^*}} - Y_{a^*,M^{a^*}} \mid c] \\ &= \mathrm{E}\left[\mathrm{E}(Y \mid a, M, c) - \mathrm{E}(Y \mid a^*, M, c) \mid a^*, c\right] \\ &= \mathrm{E}\left[\mathrm{E}(Y \mid a, M, c) \mid a^*, c\right] - \mathrm{E}\left[\mathrm{E}(Y \mid a^*, M, c) \mid a^*, c\right] \\ &= \mathrm{E}\left[(\theta_0 + \theta_1 a + \theta_2 M + \theta_3 aM + \theta_4 c + \theta_5 ac + \theta_6 Mc) \mid a^*, c\right] \\ &- \mathrm{E}\left[(\theta_0 + \theta_1 a^* + \theta_2 M + \theta_3 a^* M + \theta_4 c + \theta_5 a^* c + \theta_6 Mc) \mid a^*, c\right] \\ &= \theta_1(a - a^*) + \theta_3 \cdot \mathrm{E}(M \mid a^*, c)(a - a^*) + \theta_5 c(a - a^*) \\ &= (a - a^*)[\theta_1 + \theta_3(\beta_0 + \beta_1 a^* + \beta_2 c + \beta_3 a^* c) + \theta_5 c] \end{aligned} \tag{2}$$

$$\begin{aligned} TNDE &= \mathrm{E}[Y_{a,M^{a}} - Y_{a^*,M^{a}} \mid c] \\ &= \mathrm{E}\left[\mathrm{E}(Y \mid a, M, c) - \mathrm{E}(Y \mid a^*, M, c) \mid a, c\right] \\ &= \mathrm{E}\left[\mathrm{E}(Y \mid a, M, c) \mid a, c\right] - \mathrm{E}\left[\mathrm{E}(Y \mid a^*, M, c) \mid a, c\right] \\ &= \mathrm{E}\left[(\theta_0 + \theta_1 a + \theta_2 M + \theta_3 aM + \theta_4 c + \theta_5 ac + \theta_6 Mc) \mid a, c\right] \\ &- \mathrm{E}\left[(\theta_0 + \theta_1 a^* + \theta_2 M + \theta_3 a^* M + \theta_4 c + \theta_5 a^* c + \theta_6 Mc) \mid a, c\right] \\ &= \theta_1(a - a^*) + \theta_3 \cdot \mathrm{E}(M \mid a, c)(a - a^*) + \theta_5 c(a - a^*) \\ &= (a - a^*)[\theta_1 + \theta_3(\beta_0 + \beta_1 a + \beta_2 c + \beta_3 ac) + \theta_5 c] \end{aligned} \tag{3}$$

$$\begin{aligned} TNIE &= \mathrm{E}[Y_{a,M^{a}} - Y_{a,M^{a^*}} \mid c] \\ &= \mathrm{E}\left[\mathrm{E}(Y \mid a, M, c) \mid a, c\right] - \mathrm{E}\left[\mathrm{E}(Y \mid a, M, c) \mid a^*, c\right] \\ &= \mathrm{E}\left[(\theta_0 + \theta_1 a + \theta_2 M + \theta_3 aM + \theta_4 c + \theta_5 ac + \theta_6 Mc) \mid a, c\right] \\ &- \mathrm{E}\left[(\theta_0 + \theta_1 a + \theta_2 M + \theta_3 aM + \theta_4 c + \theta_5 ac + \theta_6 Mc) \mid a^*, c\right] \\ &= (\theta_0 + \theta_1 a + \theta_4 c + \theta_5 ac) + \mathrm{E}\left[(\theta_2 M + \theta_3 aM + \theta_6 Mc) \mid a, c\right] \\ &- (\theta_0 + \theta_1 a + \theta_4 c + \theta_5 ac) - \mathrm{E}\left[(\theta_2 M + \theta_3 aM + \theta_6 Mc) \mid a^*, c\right] \\ &= (\theta_2 + \theta_3 a + \theta_6 c) \cdot [\mathrm{E}(M \mid a, c) - \mathrm{E}(M \mid a^*, c)] \\ &= (\theta_2 + \theta_3 a + \theta_6 c)[(\beta_0 + \beta_1 a + \beta_2 c + \beta_3 ac) - (\beta_0 + \beta_1 a^* + \beta_2 c + \beta_3 a^* c)] \\ &= (\theta_2 + \theta_3 a + \theta_6 c)(\beta_1 + \beta_3 c)(a - a^*) \end{aligned} \tag{4}$$

$$PNIE = \mathrm{E}[Y_{a^*,M^{a}} - Y_{a^*,M^{a^*}} \mid c]$$

$$
\begin{aligned}
&= \mathrm{E}\left[\mathrm{E}(Y \mid a^*, M, c) \mid a, c\right] - \mathrm{E}\left[(Y \mid a^*, M, c) \mid a^*, c\right] \\
&= \mathrm{E}\left[(\theta_0 + \theta_1 a^* + \theta_2 M + \theta_3 a^* M + \theta_4 c + \theta_5 a^* c + \theta_6 M c) \mid a, c\right] \\
&- \mathrm{E}\left[(\theta_0 + \theta_1 a^* + \theta_2 M + \theta_3 a^* M + \theta_4 c + \theta_5 a^* c + \theta_6 M c) \mid a^*, c\right] \\
&= (\theta_0 + \theta_1 a^* + \theta_4 c + \theta_5 a^* c) + \mathrm{E}\left[(\theta_2 M + \theta_3 a^* M + \theta_6 M c) \mid a, c\right] \\
&- (\theta_0 + \theta_1 a^* + \theta_4 c + \theta_5 a^* c) - \mathrm{E}\left[(\theta_2 M + \theta_3 a^* M + \theta_6 M c) \mid a^*, c\right] \\
&= (\theta_2 + \theta_3 a^* + \theta_6 c) \cdot \left[\mathrm{E}(M \mid a, c) - \mathrm{E}(M \mid a^*, c)\right] \\
&= (\theta_2 + \theta_3 a^* + \theta_6 c)[(\beta_0 + \beta_1 a + \beta_2 c + \beta_3 ac) - (\beta_0 + \beta_1 a^* + \beta_2 c + \beta_3 a^* c)] \\
&= (\theta_2 + \theta_3 a^* + \theta_6 c)(\beta_1 + \beta_3 c)(a - a^*) \qquad (5)
\end{aligned}
$$

3.1.2 Standard error

$$
(a - a^*)\Gamma_{CDE(m)} = \frac{\partial CDE(m)}{\partial(\beta^T, \theta^T)^T} = (a - a^*) \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 1 \\ 0 \\ m \\ 0 \\ c \\ 0 \end{bmatrix} \qquad (6)
$$

$$
(a - a^*)\Gamma_{PNDE} = \frac{\partial PNDE}{\partial(\beta^T, \theta^T)^T} = (a - a^*) \begin{bmatrix} \theta_3 \\ \theta_3 a^* \\ \theta_3 c \\ \theta_3 a^* c \\ 0 \\ 1 \\ 0 \\ \beta_0 + \beta_1 a^* + \beta_2 c + \beta_3 a^* c \\ 0 \\ c \\ 0 \end{bmatrix} \qquad (7)
$$

$$
(a - a^*)\Gamma_{TNDE} = \frac{\partial TNDE}{\partial(\beta^T, \theta^T)^T} = (a - a^*) \begin{bmatrix} \theta_3 \\ \theta_3 a \\ \theta_3 c \\ \theta_3 ac \\ 0 \\ 1 \\ 0 \\ \beta_0 + \beta_1 a + \beta_2 c + \beta_3 ac \\ 0 \\ c \\ 0 \end{bmatrix} \qquad (8)
$$

$$(a-a^*)\Gamma_{TNIE}=\frac{\partial TNIE}{\partial(\beta^T,\theta^T)^T}=(a-a^*)\begin{bmatrix}0\\ \theta_2+\theta_3 a+\theta_6 c\\ 0\\ c(\theta_2+\theta_3 a+\theta_6 c)\\ 0\\ 0\\ \beta_1+\beta_3 c\\ a(\beta_1+\beta_3 c)\\ 0\\ 0\\ c(\beta_1+\beta_3 c)\end{bmatrix} \tag{9}$$

$$(a-a^*)\Gamma_{PNIE}=\frac{\partial PNIE}{\partial(\beta^T,\theta^T)^T}=(a-a^*)\begin{bmatrix}0\\ \theta_2+\theta_3 a^*+\theta_6 c\\ 0\\ c(\theta_2+\theta_3 a^*+\theta_6 c)\\ 0\\ 0\\ \beta_1+\beta_3 c\\ a^*(\beta_1+\beta_3 c)\\ 0\\ 0\\ c(\beta_1+\beta_3 c)\end{bmatrix} \tag{10}$$

$$\begin{aligned}(a-a^*)\Gamma_{TE}&=\frac{\partial TE}{\partial(\boldsymbol{\beta}^T,\boldsymbol{\theta}^T,\sigma^2)^T}\\ &=\frac{\partial(PNDE+TNIE)}{\partial(\boldsymbol{\beta}^T,\boldsymbol{\theta}^T,\sigma^2)^T}\\ &=(a-a^*)(\Gamma_{PNDE}+\Gamma_{TNIE})\end{aligned}$$

$$\begin{aligned}(a-a^*)\Gamma_{PM}&=\frac{\partial PM}{\partial(\boldsymbol{\beta}^T,\boldsymbol{\theta}^T,\sigma^2)^T}\\ &=\frac{\partial PM}{\partial PNDE}\frac{\partial PNDE}{\partial(\boldsymbol{\beta}^T,\boldsymbol{\theta}^T,\sigma^2)^T}+\frac{\partial PM}{\partial TNIE}\frac{\partial TNIE}{\partial(\boldsymbol{\beta}^T,\boldsymbol{\theta}^T,\sigma^2)^T}\\ &=\frac{\partial PM}{\partial PNDE}(a-a^*)\Gamma_{PNDE}+\frac{\partial PM}{\partial TNIE}(a-a^*)\Gamma_{TNIE}\\ &=-\frac{\exp(PNDE)\left\{\exp(TNIE)-1\right\}}{\left\{\exp(PNDE)\exp(TNIE)-1\right\}^2}(a-a^*)\Gamma_{PNDE}\\ &\quad+\frac{\exp(PNDE)\exp(TNIE)\left\{\exp(PNDE)-1\right\}}{\left\{\exp(PNDE)\exp(TNIE)-1\right\}^2}(a-a^*)\Gamma_{TNIE}\end{aligned}$$

$$\boldsymbol{\Sigma}=\begin{bmatrix}\boldsymbol{\Sigma}_{\boldsymbol{\beta}} & 0\\ 0 & \boldsymbol{\Sigma}_{\boldsymbol{\theta}}\end{bmatrix}$$

Hence,

$$\begin{aligned}SE(\widehat{CDE}(m))&=\sqrt{\Gamma_{CDE(m)}^T\ \boldsymbol{\Sigma}\ \Gamma_{CDE(m)}}\ |a-a^*|\\ SE(\widehat{PNDE})&=\sqrt{\Gamma_{PNDE}^T\ \boldsymbol{\Sigma}\ \Gamma_{PNDE}}\ |a-a^*|\\ SE(\widehat{TNIE})&=\sqrt{\Gamma_{TNIE}^T\ \boldsymbol{\Sigma}\ \Gamma_{TNIE}}\ |a-a^*|\end{aligned}$$

$$SE(\widehat{TNDE}) = \sqrt{\Gamma_{TNDE}^T \, \mathbf{\Sigma} \, \Gamma_{TNDE}} \; |a - a^*|$$

$$SE(\widehat{PNIE}) = \sqrt{\Gamma_{PNIE}^T \, \mathbf{\Sigma} \, \Gamma_{PNIE}} \; |a - a^*|$$

$$SE(\widehat{TE}) = \sqrt{\Gamma_{TE}^T \, \mathbf{\Sigma} \, \Gamma_{TE}} \; |a - a^*|$$

$$SE(\widehat{PM}) = \sqrt{\Gamma_{PM}^T \, \mathbf{\Sigma} \, \Gamma_{PM}} \; |a - a^*|$$

3.2 Logistic mediator model, linear outcome model

3.2.1 Point estimate

$$\begin{aligned}
\text{logit } \mathrm{P}[M=1 \mid a,c] &= \beta_0+\beta_1 a+\beta_2 c+\beta_3 ac \\
\Rightarrow \mathrm{P}[M=1 \mid a,c] &= \frac{\exp(\beta_0+\beta_1 a+\beta_2 c+\beta_3 ac)}{1+\exp(\beta_0+\beta_1 a+\beta_2 c+\beta_3 ac)} \\
\mathrm{E}[Y \mid a,m,c] &= \theta_0+\theta_1 a+\theta_2 m+\theta_4 c+\theta_3 am+\theta_5 ac+\theta_6 mc
\end{aligned}$$

$$CDE(m) = (\theta_1+\theta_3 m+\theta_5 c)(a-a^*) \tag{11}$$

$$\begin{aligned}
PNDE &= \mathrm{E}[Y_{a,M^{a^*}} - Y_{a^*,M^{a^*}} \mid c] \\
&= \mathrm{E}\left[\mathrm{E}(Y \mid a,M,c) - \mathrm{E}(Y \mid a^*,M,c) \mid a^*,c\right] \\
&= \mathrm{E}\left[\theta_1(a-a^*)+\theta_3 M(a-a^*)+\theta_5 c(a-a^*) \mid a^*,c\right] \\
&= \theta_1(a-a^*)+\theta_3\cdot\mathrm{E}(M \mid a^*,c)(a-a^*)+\theta_5 c(a-a^*) \\
&= (a-a^*)\left[\theta_1+\theta_3\frac{\exp(\beta_0+\beta_1 a^*+\beta_2 c+\beta_3 a^* c)}{1+\exp(\beta_0+\beta_1 a^*+\beta_2 c+\beta_3 a^* c)}+\theta_5 c\right] \\
&= (a-a^*)\left[\theta_1+\theta_3\cdot\mathrm{expit}(\beta_0+\beta_1 a^*+\beta_2 c+\beta_3 a^* c)+\theta_5 c\right]
\end{aligned} \tag{12}$$

$$\begin{aligned}
TNDE &= \mathrm{E}[Y_{a,M^{a}} - Y_{a^*,M^{a}} \mid c] \\
&= \mathrm{E}\left[\mathrm{E}(Y \mid a,M,c) - \mathrm{E}(Y \mid a^*,M,c) \mid a,c\right] \\
&= \mathrm{E}[\theta_1(a-a^*)+\theta_3 M(a-a^*)+\theta_5 c(a-a^*) \mid a,c] \\
&= \theta_1(a-a^*)+\theta_3\cdot\mathrm{E}(M \mid a,c)(a-a^*)+\theta_5 c(a-a^*) \\
&= (a-a^*)\left[\theta_1+\theta_3\frac{\exp(\beta_0+\beta_1 a+\beta_2 c+\beta_3 ac)}{1+\exp(\beta_0+\beta_1 a+\beta_2 c+\beta_3 ac)}+\theta_5 c\right] \\
&= (a-a^*)\left[\theta_1+\theta_3\cdot\mathrm{expit}(\beta_0+\beta_1 a+\beta_2 c+\beta_3 ac)+\theta_5 c\right]
\end{aligned} \tag{13}$$

$$\begin{aligned}
TNIE &= \mathrm{E}[Y_{a,M^{a}} - Y_{a,M^{a^*}} \mid c] \\
&= \mathrm{E}\left[\mathrm{E}(Y \mid a,M,c) \mid a,c\right] - \mathrm{E}\left[\mathrm{E}(Y \mid a,M,c) \mid a^*,c\right] \\
&= \mathrm{E}\left[(\theta_0+\theta_1 a+\theta_2 M+\theta_3 aM+\theta_4 c+\theta_5 ac+\theta_6 Mc) \mid a,c\right] \\
&- \mathrm{E}\left[(\theta_0+\theta_1 a+\theta_2 M+\theta_3 aM+\theta_4 c+\theta_5 ac+\theta_6 Mc) \mid a^*,c\right] \\
&= (\theta_2+\theta_3 a+\theta_6 c)\cdot\mathrm{E}(M \mid a,c) - \mathrm{E}(M \mid a^*,c)] \\
&= (\theta_2+\theta_3 a+\theta_6 c)\left[\frac{\exp(\beta_0+\beta_1 a+\beta_2 c+\beta_3 ac)}{1+\exp(\beta_0+\beta_1 a+\beta_2 c+\beta_3 ac)} - \frac{\exp(\beta_0+\beta_1 a^*+\beta_2 c+\beta_3 a^* c)}{1+\exp(\beta_0+\beta_1 a^*+\beta_2 c+\beta_3 a^* c)}\right] \\
&= (\theta_2+\theta_3 a+\theta_6 c)\left[\mathrm{expit}(\beta_0+\beta_1 a+\beta_2 c+\beta_3 ac) - \mathrm{expit}(\beta_0+\beta_1 a^*+\beta_2 c+\beta_3 a^* c)\right]
\end{aligned} \tag{14}$$

$$\begin{aligned}
PNIE &= \mathrm{E}[Y_{a^*,M^{a}} - Y_{a^*,M^{a^*}} \mid c] \\
&= \mathrm{E}\left[\mathrm{E}(Y \mid a^*,M,c) \mid a,c\right] - \mathrm{E}\left[(Y \mid a^*,M,c) \mid a^*,c\right] \\
&= \mathrm{E}\left[(\theta_0+\theta_1 a^*+\theta_2 M+\theta_3 a^* M+\theta_4 c+\theta_5 a^* c+\theta_6 Mc) \mid a,c\right] \\
&- \mathrm{E}\left[(\theta_0+\theta_1 a^*+\theta_2 M+\theta_3 a^* M+\theta_4 c+\theta_5 a^* c+\theta_6 Mc) \mid a^*,c\right] \\
&= (\theta_2+\theta_3 a^*+\theta_6 c)\cdot[\mathrm{E}(M \mid a,c) - \mathrm{E}(M \mid a^*,c)] \\
&= (\theta_2+\theta_3 a^*+\theta_6 c)\left[\frac{\exp(\beta_0+\beta_1 a+\beta_2 c+\beta_3 ac)}{1+\exp(\beta_0+\beta_1 a+\beta_2 c+\beta_3 ac)} - \frac{\exp(\beta_0+\beta_1 a^*+\beta_2 c+\beta_3 a^* c)}{1+\exp(\beta_0+\beta_1 a^*+\beta_2 c+\beta_3 a^* c)}\right] \\
&= (\theta_2+\theta_3 a^*+\theta_6 c)\left[\mathrm{expit}(\beta_0+\beta_1 a+\beta_2 c+\beta_3 ac) - \mathrm{expit}(\beta_0+\beta_1 a^*+\beta_2 c+\beta_3 a^* c)\right]
\end{aligned} \tag{15}$$

3.2.2 Standard error

$$(a-a^*)\Gamma_{CDE(m)} = \frac{\partial CDE(m)}{\partial(\beta^T,\theta^T)^T} = (a-a^*)\begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 1 \\ 0 \\ m \\ 0 \\ c \\ 0 \end{bmatrix} \tag{16}$$

$$\begin{aligned}(a-a^*)\Gamma_{PNDE} &= \frac{\partial PNDE}{\partial(\beta^T,\theta^T)^T} \\ &= (a-a^*)\begin{bmatrix} \theta_3 \cdot \text{expit}\left[a^*(\beta_1+\beta_3 c)+\beta_0+\beta_2 c\right] \cdot \{1-\text{expit}[a^*(\beta_1+\beta_3 c)+\beta_0+\beta_2 c]\} \\ a^* \cdot \theta_3 \cdot \text{expit}\left[a^*(\beta_1+\beta_3 c)+\beta_0+\beta_2 c\right] \cdot \{1-\text{expit}[a^*(\beta_1+\beta_3 c)+\beta_0+\beta_2 c]\} \\ c \cdot \theta_3 \cdot \text{expit}\left[a^*(\beta_1+\beta_3 c)+\beta_0+\beta_2 c\right] \cdot \{1-\text{expit}[a^*(\beta_1+\beta_3 c)+\beta_0+\beta_2 c]\} \\ a^* \cdot c \cdot \theta_3 \cdot \text{expit}\left[a^*(\beta_1+\beta_3 c)+\beta_0+\beta_2 c\right] \cdot \{1-\text{expit}[a^*(\beta_1+\beta_3 c)+\beta_0+\beta_2 c]\} \\ 0 \\ 1 \\ 0 \\ \text{expit}\left[a^*(\beta_1+\beta_3 c)+\beta_0+\beta_2 c\right] \\ 0 \\ c \\ 0 \end{bmatrix}\end{aligned} \tag{17}$$

$$\begin{aligned}(a-a^*)\Gamma_{TNDE} &= \frac{\partial TNDE}{\partial(\beta^T,\theta^T)^T} \\ &= (a-a^*)\begin{bmatrix} \theta_3 \cdot \text{expit}\left[a(\beta_1+\beta_3 c)+\beta_0+\beta_2 c\right] \cdot \{1-\text{expit}[a(\beta_1+\beta_3 c)+\beta_0+\beta_2 c]\} \\ a \cdot \theta_3 \cdot \text{expit}\left[a(\beta_1+\beta_3 c)+\beta_0+\beta_2 c\right] \cdot \{1-\text{expit}[a(\beta_1+\beta_3 c)+\beta_0+\beta_2 c]\} \\ c \cdot \theta_3 \cdot \text{expit}\left[a(\beta_1+\beta_3 c)+\beta_0+\beta_2 c\right] \cdot \{1-\text{expit}[a(\beta_1+\beta_3 c)+\beta_0+\beta_2 c]\} \\ a \cdot c \cdot \theta_3 \cdot \text{expit}\left[a(\beta_1+\beta_3 c)+\beta_0+\beta_2 c\right] \cdot \{1-\text{expit}[a(\beta_1+\beta_3 c)+\beta_0+\beta_2 c]\} \\ 0 \\ 1 \\ 0 \\ \text{expit}\left[a(\beta_1+\beta_3 c)+\beta_0+\beta_2 c\right] \\ 0 \\ c \\ 0 \end{bmatrix}\end{aligned} \tag{18}$$

$$\Gamma_{TNIE} = \frac{\partial TNIE}{\partial(\beta^T, \theta^T)^T}$$
$$= \begin{bmatrix} (\theta_2+\theta_3 a+\theta_6 c)\cdot\Big(\text{expit}\left[a(\beta_1+\beta_3 c)+\beta_0+\beta_2 c\right]\cdot\{1-\text{expit}[a(\beta_1+\beta_3 c)+\beta_0+\beta_2 c]\} \\ -\text{expit}\left[a^*(\beta_1+\beta_3 c)+\beta_0+\beta_2 c\right]\cdot\{1-\text{expit}[a^*(\beta_1+\beta_3 c)+\beta_0+\beta_2 c]\}\Big) \\ \\ (\theta_2+\theta_3 a+\theta_6 c)\cdot\Big(a\cdot\{\text{expit}\left[a(\beta_1+\beta_3 c)+\beta_0+\beta_2 c\right]\}\cdot\{1-\text{expit}[a(\beta_1+\beta_3 c)+\beta_0+\beta_2 c]\} \\ -a^*\cdot\{\text{expit}\left[a^*(\beta_1+\beta_3 c)+\beta_0+\beta_2 c\right]\}\cdot\{1-\text{expit}[a^*(\beta_1+\beta_3 c)+\beta_0+\beta_2 c]\}\Big) \\ \\ c\cdot(\theta_2+\theta_3 a+\theta_6 c)\cdot\Big(\text{expit}\left[a(\beta_1+\beta_3 c)+\beta_0+\beta_2 c\right]\cdot\{1-\text{expit}[a(\beta_1+\beta_3 c)+\beta_0+\beta_2 c]\} \\ -\text{expit}\left[a^*(\beta_1+\beta_3 c)+\beta_0+\beta_2 c\right]\cdot\{1-\text{expit}[a^*(\beta_1+\beta_3 c)+\beta_0+\beta_2 c]\}\Big) \\ \\ c\cdot(\theta_2+\theta_3 a+\theta_6 c)\cdot\Big(a\cdot\{\text{expit}\left[a(\beta_1+\beta_3 c)+\beta_0+\beta_2 c\right]\}\cdot\{1-\text{expit}[a(\beta_1+\beta_3 c)+\beta_0+\beta_2 c]\} \\ -a^*\cdot\{\text{expit}\left[a^*(\beta_1+\beta_3 c)+\beta_0+\beta_2 c\right]\}\cdot\{1-\text{expit}[a^*(\beta_1+\beta_3 c)+\beta_0+\beta_2 c]\}\Big) \\ 0 \\ 0 \\ \text{expit}[a(\beta_1+\beta_3 c)+\beta_0+\beta_2 c]-\text{expit}[a^*(\beta_1+\beta_3 c)+\beta_0+\beta_2 c] \\ a\cdot\Big(\text{expit}[a(\beta_1+\beta_3 c)+\beta_0+\beta_2 c]-\text{expit}[a^*(\beta_1+\beta_3 c)+\beta_0+\beta_2 c]\Big) \\ 0 \\ 0 \\ c\cdot\Big(\text{expit}[a(\beta_1+\beta_3 c)+\beta_0+\beta_2 c]-\text{expit}[a^*(\beta_1+\beta_3 c)+\beta_0+\beta_2 c]\Big) \end{bmatrix} \quad (19)$$

$$\Gamma_{PNIE} = \frac{\partial PNIE}{\partial(\beta^T, \theta^T)^T}$$
$$= \begin{bmatrix} (\theta_2+\theta_3 a^*+\theta_6 c)\cdot\Big(\text{expit}\left[a(\beta_1+\beta_3 c)+\beta_0+\beta_2 c\right]\cdot\{1-\text{expit}[a(\beta_1+\beta_3 c)+\beta_0+\beta_2 c]\} \\ -\text{expit}\left[a^*(\beta_1+\beta_3 c)+\beta_0+\beta_2 c\right]\cdot\{1-\text{expit}[a^*(\beta_1+\beta_3 c)+\beta_0+\beta_2 c]\}\Big) \\ \\ (\theta_2+\theta_3 a^*+\theta_6 c)\cdot\Big(a\cdot\{\text{expit}\left[a(\beta_1+\beta_3 c)+\beta_0+\beta_2 c\right]\}\cdot\{1-\text{expit}[a(\beta_1+\beta_3 c)+\beta_0+\beta_2 c]\} \\ -a^*\cdot\{\text{expit}\left[a^*(\beta_1+\beta_3 c)+\beta_0+\beta_2 c\right]\}\cdot\{1-\text{expit}[a^*(\beta_1+\beta_3 c)+\beta_0+\beta_2 c]\}\Big) \\ \\ c\cdot(\theta_2+\theta_3 a^*+\theta_6 c)\cdot\Big(\text{expit}\left[a(\beta_1+\beta_3 c)+\beta_0+\beta_2 c\right]\cdot\{1-\text{expit}[a(\beta_1+\beta_3 c)+\beta_0+\beta_2 c]\} \\ -\text{expit}\left[a^*(\beta_1+\beta_3 c)+\beta_0+\beta_2 c\right]\cdot\{1-\text{expit}[a^*(\beta_1+\beta_3 c)+\beta_0+\beta_2 c]\}\Big) \\ \\ c\cdot(\theta_2+\theta_3 a^*+\theta_6 c)\cdot\Big(a\cdot\{\text{expit}\left[a(\beta_1+\beta_3 c)+\beta_0+\beta_2 c\right]\}\cdot\{1-\text{expit}[a(\beta_1+\beta_3 c)+\beta_0+\beta_2 c]\} \\ -a^*\cdot\{\text{expit}\left[a^*(\beta_1+\beta_3 c)+\beta_0+\beta_2 c\right]\}\cdot\{1-\text{expit}[a^*(\beta_1+\beta_3 c)+\beta_0+\beta_2 c]\}\Big) \\ 0 \\ 0 \\ \text{expit}[a(\beta_1+\beta_3 c)+\beta_0+\beta_2 c]-\text{expit}[a^*(\beta_1+\beta_3 c)+\beta_0+\beta_2 c] \\ a^*\cdot\Big(\text{expit}[a(\beta_1+\beta_3 c)+\beta_0+\beta_2 c]-\text{expit}[a^*(\beta_1+\beta_3 c)+\beta_0+\beta_2 c]\Big) \\ 0 \\ 0 \\ c\cdot\Big(\text{expit}[a(\beta_1+\beta_3 c)+\beta_0+\beta_2 c]-\text{expit}[a^*(\beta_1+\beta_3 c)+\beta_0+\beta_2 c]\Big) \end{bmatrix} \quad (20)$$

$$\Gamma_{TE} = \frac{\partial TE}{\partial(\boldsymbol{\beta}^T, \boldsymbol{\theta}^T)^T}$$

$$= \frac{\partial(PNDE + TNIE)}{\partial(\boldsymbol{\beta}^T, \boldsymbol{\theta}^T)^T}$$
$$= (a - a^*)\Gamma_{PNDE} + \Gamma_{TNIE}$$

$$\begin{aligned}
\Gamma_{PM} &= \frac{\partial PM}{\partial(\boldsymbol{\beta}^T, \boldsymbol{\theta}^T)^T} \\
&= \frac{\partial PM}{\partial PNDE} \frac{\partial PNDE}{\partial(\boldsymbol{\beta}^T, \boldsymbol{\theta}^T)^T} + \frac{\partial PM}{\partial TNIE} \frac{\partial TNIE}{\partial(\boldsymbol{\beta}^T, \boldsymbol{\theta}^T)^T} \\
&= \frac{\partial PM}{\partial PNDE}(a - a^*)\Gamma_{PNDE} + \frac{\partial PM}{\partial TNIE}\Gamma_{TNIE} \\
&= \frac{-TNIE}{(PNDE + TNIE)^2}(a - a^*)\Gamma_{PNDE} + \frac{PNDE}{(PNDE + TNIE)^2}\Gamma_{TNIE} \\
&= \frac{-TNIE\ (a - a^*)\Gamma_{PNDE} + PNDE\ \Gamma_{TNIE}}{(PNDE + TNIE)^2}
\end{aligned}$$

$$\boldsymbol{\Sigma} = \begin{bmatrix} \boldsymbol{\Sigma}_{\boldsymbol{\beta}} & 0 \\ 0 & \boldsymbol{\Sigma}_{\boldsymbol{\theta}} \end{bmatrix}$$

Hence,

$$\begin{aligned}
SE(\widehat{CDE}(m)) &= \sqrt{\Gamma_{CDE(m)}^T\ \boldsymbol{\Sigma}\ \Gamma_{CDE(m)}}\ |a - a^*| \\
SE(\widehat{PNDE}) &= \sqrt{\Gamma_{PNDE}^T\ \boldsymbol{\Sigma}\ \Gamma_{PNDE}}\ |a - a^*| \\
SE(\widehat{TNIE}) &= \sqrt{\Gamma_{TNIE}^T\ \boldsymbol{\Sigma}\ \Gamma_{TNIE}} \\
SE(\widehat{TNDE}) &= \sqrt{\Gamma_{TNDE}^T\ \boldsymbol{\Sigma}\ \Gamma_{TNDE}}\ |a - a^*| \\
SE(\widehat{PNIE}) &= \sqrt{\Gamma_{PNIE}^T\ \boldsymbol{\Sigma}\ \Gamma_{PNIE}} \\
SE(\widehat{TE}) &= \sqrt{\Gamma_{TE}^T\ \boldsymbol{\Sigma}\ \Gamma_{TE}} \\
SE(\widehat{PM}) &= \sqrt{\Gamma_{PM}^T\ \boldsymbol{\Sigma}\ \Gamma_{PM}}
\end{aligned}$$

3.3 Linear mediator model, non-linear outcome model

3.3.1 Point estimate

$$
\begin{aligned}
\mathrm{E}[M \mid a, c] &= \beta_0 + \beta_1 a + \beta_2 c + \beta_3 ac \\
\text{logit } \mathrm{P}[Y = 1 \mid a, m, c] &= \theta_0 + \theta_1 a + \theta_2 m + \theta_4 c + \theta_3 am + \theta_5 ac + \theta_6 mc
\end{aligned}
$$

For demonstration, we assume the outcome is a logistic model. For demonstration, we assume the outcome is a logistic model. For other types of outcome model, the effect will be on the log of the corresponding scale in Table 1.

$$
OR^{CDE}(m) = (\theta_1 + \theta_3 m + \theta_5 c)(a - a^*) \tag{21}
$$

$$
\begin{aligned}
OR^{PNDE} &= \exp\left\{\log\frac{\mathrm{P(Y_{a,M^{a^*}} = 1 \mid c)/[1 - P(Y_{a,M^{a^*}} = 1 \mid c)]}}{\mathrm{P(Y_{a^*,M^{a^*}} = 1 \mid c)/[1 - P(Y_{a^*,M^{a^*}} = 1 \mid c)]}}\right\} \\
&= \exp\left\{\text{logit } \mathrm{P}(Y_{a,M^{a^*}} = 1 \mid c) - \text{logit } \mathrm{P}(Y_{a^*,M^{a^*}} = 1 \mid c)\right\} \\
\text{If } Y \text{ is rare, } &\approx \exp\left\{\log \mathrm{P}(Y_{a,M^{a^*}} = 1 \mid c) - \log \mathrm{P}(Y_{a^*,M^{a^*}} = 1 \mid c)\right\} \\
&= \exp\left\{\log\frac{\mathrm{P(Y_{a,M^{a^*}} = 1 \mid c)}}{\mathrm{P(Y_{a^*,M^{a^*}} = \mid c)}}\right\}
\end{aligned}
$$

$$
\begin{aligned}
\mathrm{E}(Y_{a,M^{a^*}} \mid c) &= \mathrm{E}\left[\mathrm{E}(Y \mid a, M, c) \mid a^*, c\right] \\
&\approx \mathrm{E}[\exp(\theta_0 + \theta_1 a + \theta_2 M + \theta_4 c + \theta_3 aM + \theta_5 ac + \theta_6 Mc) \mid a^*] \\
&\approx \exp(\theta_0 + \theta_1 a + \theta_4 c + \theta_5 ac) \cdot \mathrm{E}\left\{\exp[(\theta_2 + \theta_3 a + \theta_6 c)M] \mid a^*, c\right\}
\end{aligned}
$$

$\mathrm{E}\left\{\exp[(\theta_2 + \theta_3 a + \theta_6 c)M] \mid a^*, c\right\}$ can be calculated using moment generating function (MGF).
Since

$$
\begin{aligned}
&M \sim N(\mu, \sigma^2), \\
&\mathrm{E}(e^{tM}) = MGF_M(t) = e^{t\mu + \frac{1}{2}\sigma^2 t^2}
\end{aligned}
$$

where

$$
\begin{aligned}
&\sigma^2 \text{ is the variance of error in the linear mediator model,} \\
&\mu = \mathrm{E}(M \mid a^*, c) = \beta_0 + \beta_1 a^* + \beta_2 c + \beta_3 a^* c, \\
&t = \theta_2 + \theta_3 a + \theta_6 c
\end{aligned}
$$

Plugging in and we have

$$
\mathrm{E}\left\{\exp[(\theta_2 + \theta_3 a + \theta_6 c)M] \mid a^*, c\right\} = \exp\left[(\theta_2 + \theta_3 a + \theta_6 c)(\beta_0 + \beta_1 a^* + \beta_2 c + \beta_3 a^* c) + \frac{1}{2}\sigma^2(\theta_2 + \theta_3 a + \theta_6 c)^2\right]
$$

Hence,

$$
\begin{aligned}
\mathrm{E}(Y_{a,M^{a^*}} \mid c) &= \exp\Big[(\theta_0 + \theta_1 a + \theta_4 c + \theta_5 ac) + (\theta_2 + \theta_3 a + \theta_6 c)(\beta_0 + \beta_1 a^* + \beta_2 c + \beta_3 a^* c) + \frac{1}{2}\sigma^2(\theta_2 + \theta_3 a + \theta_6 c)^2\Big] \\
\therefore \log\ \mathrm{E}(Y_{a,M^{a^*}} \mid c) &= \theta_0 + \theta_1 a + \theta_4 c + \theta_5 ac + (\theta_2 + \theta_3 a + \theta_6 c)(\beta_0 + \beta_1 a^* + \beta_2 c + \beta_3 a^* c) + \frac{1}{2}\sigma^2(\theta_2 + \theta_3 a + \theta_6 c)^2
\end{aligned}
$$

Similarly,

$$
\mathrm{E}(Y_{a^*,M^{a^*}} \mid c) = \exp\Big[(\theta_0 + \theta_1 a^* + \theta_4 c + \theta_5 a^* c) + (\theta_2 + \theta_3 a^* + \theta_6 c)(\beta_0 + \beta_1 a^* + \beta_2 c + \beta_3 a^* c) + \frac{1}{2}\sigma^2(\theta_2 + \theta_3 a^* + \theta_6 c)^2\Big]
$$

$$\therefore \log\ \mathrm{E}(Y_{a^*,M^{a^*}} \mid c) = \theta_0 + \theta_1 a^* + \theta_4 c + \theta_5 a^* c + (\theta_2 + \theta_3 a^* + \theta_6 c)(\beta_0 + \beta_1 a^* + \beta_2 c + \beta_3 a^* c) + \frac{1}{2}\sigma^2(\theta_2 + \theta_3 a^* + \theta_6 c)^2$$

Hence,

$$OR^{PNDE} \approx \exp\Big([a - a^*] \cdot [\theta_1 + \theta_5 c + \theta_3\,(\beta_0 + \beta_1 a^* + \beta_2 c + \beta_3 a^* c) + \frac{1}{2}\sigma^2\theta_3(2\theta_2 + \theta_3 a + \theta_3 a^* + 2\theta_6 c)]\Big) \tag{22}$$

$$\begin{aligned} OR^{TNIE} &= \exp\left\{\log\frac{\mathrm{P}(Y_{a,M^a} = 1 \mid c)/[1 - \mathrm{P}(Y_{a,M^a} = 1 \mid c)]}{\mathrm{P}(Y_{a,M^{a^*}} = 1 \mid c)/[1 - \mathrm{P}(Y_{a,M^{a^*}} = 1 \mid c)]}\right\} \\ &= \exp\left\{\mathrm{logit}\ \mathrm{P}(Y_{a,M^a} = 1 \mid c) - \mathrm{logit}\ \mathrm{P}(Y_{a,M^{a^*}} = 1 \mid c)\right\} \\ \text{If } Y \text{ is rare, } &\approx \exp\left\{\log\frac{\mathrm{P}(Y_{a,M^a} = 1 \mid c)}{\mathrm{P}(Y_{a,M^{a^*}} = 1 \mid c)}\right\} \end{aligned} \tag{23}$$

Using the same approach (MGF),

$$OR^{TNIE} \approx \exp\{(\theta_2 + \theta_3 a + \theta_6 c)(\beta_1 + \beta_3 c)(a - a^*)\}$$

For OR^{TNDE} and OR^{PNIE}, notice that the first exponential term in OR^{PNDE} and OR^{TNIE} are just the exponentiated PNDE and TNIE in 2.1, we can easily write OR^{TNDE} and OR^{PNIE}.

$$\begin{aligned} OR^{TNDE} &\approx \exp\Big([a - a^*] \cdot [\theta_1 + \theta_5 c + \theta_3\,(\beta_0 + \beta_1 a + \beta_2 c + \beta_3 a c) \\ &\quad + \frac{1}{2}\sigma^2\theta_3(2\theta_2 + \theta_3 a + \theta_3 a^* + 2\theta_6 c)]\Big) \end{aligned} \tag{24}$$

$$OR^{PNIE} = \exp\{(\theta_2 + \theta_3 a^* + \theta_6 c)(\beta_1 + \beta_3 c)(a - a^*)\} \tag{25}$$

3.3.2 Standard error

Reminder: The formulas are on log odds ratio scale.

For the convenience of programming, the parameter vector in the following variance formulas is $(\beta^T, \theta^T, \sigma^2)$, although NIE does not have σ^2.

$$(a - a^*)\Gamma_{CDE(m)} = \frac{\partial\ \log(OR^{CDE(m)})}{\partial(\beta^T, \theta^T, \sigma^2)^T} = (a - a^*)\begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 1 \\ 0 \\ m \\ 0 \\ c \\ 0 \\ 0 \end{bmatrix} \tag{26}$$

$$(a - a^*)\Gamma_{PNDE} = \frac{\partial\ \log(OR^{PNDE})}{\partial(\beta^T, \theta^T, \sigma^2)^T}$$

$$= (a - a^*) \begin{bmatrix} \theta_3 \\ \theta_3 a^* \\ \theta_3 c \\ \theta_3 a^* c \\ 0 \\ 1 \\ \theta_3 \sigma^2 \\ a^*(\beta_1 + \beta_3 c) + \beta_0 + \beta_2 c + \sigma^2[(\theta_2 + \theta_6 c) + \theta_3(a^* + a)] \\ 0 \\ c \\ \theta_3 c \sigma^2 \\ \frac{1}{2}\theta_3 \cdot [2(\theta_2 + \theta_6 c) + \theta_3(a^* + a)] \end{bmatrix} \quad (27)$$

$$(a - a^*)\Gamma_{TNDE} = \frac{\partial \ \log(OR^{TNDE})}{\partial(\beta^T, \theta^T, \sigma^2)^T}$$

$$= (a - a^*) \begin{bmatrix} \theta_3 \\ \theta_3 a \\ \theta_3 c \\ \theta_3 a c \\ 0 \\ 1 \\ \theta_3 \sigma^2 \\ a(\beta_1 + \beta_3 c) + \beta_0 + \beta_2 c + \sigma^2[(\theta_2 + \theta_6 c) + \theta_3(a^* + a)] \\ 0 \\ c \\ \theta_3 c \sigma^2 \\ \frac{1}{2}\theta_3 \cdot [2(\theta_2 + \theta_6 c) + \theta_3(a^* + a)] \end{bmatrix} \quad (28)$$

$$(a - a^*)\Gamma_{TNIE} = \frac{\partial \ \log(OR^{TNIE})}{\partial(\beta^T, \theta^T, \sigma^2)^T}$$

$$= (a - a^*) \begin{bmatrix} 0 \\ \theta_2 + \theta_3 a + \theta_6 c \\ 0 \\ c(\theta_2 + \theta_3 a + \theta_6 c) \\ 0 \\ 0 \\ \beta_1 + \beta_3 c \\ a(\beta_1 + \beta_3 c) \\ 0 \\ 0 \\ c(\beta_1 + \beta_3 c) \\ 0 \end{bmatrix} \quad (29)$$

$$(a - a^*)\Gamma_{PNIE} = \frac{\partial \ \log(OR^{PNIE})}{\partial(\beta^T, \theta^T, \sigma^2)^T} = (a - a^*) \begin{bmatrix} 0 \\ \theta_2 + \theta_3 a^* + \theta_6 c \\ 0 \\ c(\theta_2 + \theta_3 a^* + \theta_6 c) \\ 0 \\ 0 \\ \beta_1 + \beta_3 c \\ a^*(\beta_1 + \beta_3 c) \\ 0 \\ 0 \\ c(\beta_1 + \beta_3 c) \\ 0 \end{bmatrix} \quad (30)$$

$$
\begin{aligned}
(a-a^*)\Gamma_{TE} &= \frac{\partial TE}{\partial(\boldsymbol{\beta}^T, \boldsymbol{\theta}^T, \sigma^2)^T} \\
&= \frac{\partial(PNDE + TNIE)}{\partial(\boldsymbol{\beta}^T, \boldsymbol{\theta}^T, \sigma^2)^T} \\
&= (a-a^*)(\Gamma_{PNDE} + \Gamma_{TNIE})
\end{aligned}
$$

$$
\begin{aligned}
(a-a^*)\Gamma_{PM} &= \frac{\partial PM}{\partial(\boldsymbol{\beta}^T, \boldsymbol{\theta}^T, \sigma^2)^T} \\
&= \frac{\partial PM}{\partial PNDE}\frac{\partial PNDE}{\partial(\boldsymbol{\beta}^T, \boldsymbol{\theta}^T, \sigma^2)^T} + \frac{\partial PM}{\partial TNIE}\frac{\partial TNIE}{\partial(\boldsymbol{\beta}^T, \boldsymbol{\theta}^T, \sigma^2)^T} \\
&= \frac{\partial PM}{\partial PNDE}(a-a^*)\Gamma_{PNDE} + \frac{\partial PM}{\partial TNIE}(a-a^*)\Gamma_{TNIE} \\
&= -\frac{\exp(PNDE)\left\{\exp(TNIE)-1\right\}}{\left\{\exp(PNDE)\exp(TNIE)-1\right\}^2}(a-a^*)\Gamma_{PNDE} \\
&\quad + \frac{\exp(PNDE)\exp(TNIE)\left\{\exp(PNDE)-1\right\}}{\left\{\exp(PNDE)\exp(TNIE)-1\right\}^2}(a-a^*)\Gamma_{TNIE}
\end{aligned}
$$

$$
\begin{aligned}
\boldsymbol{\Sigma} &= \begin{bmatrix} \boldsymbol{\Sigma_{\beta}} & 0 & 0 \\ 0 & \boldsymbol{\Sigma_{\theta}} & 0 \\ 0 & 0 & \Sigma_{\sigma^2} \end{bmatrix} \\
\Sigma_{\sigma^2} &= \frac{2(\sigma^2)^2}{n-p} \text{ , where } p = \text{length}(\boldsymbol{\beta})
\end{aligned}
$$

Hence,

$$
\begin{aligned}
SE(\widehat{CDE}(m)) &= \sqrt{\Gamma_{CDE(m)}^T \ \boldsymbol{\Sigma} \ \Gamma_{CDE(m)}} \ |a-a^*| \\
SE(\widehat{PNDE}) &= \sqrt{\Gamma_{PNDE}^T \ \boldsymbol{\Sigma} \ \Gamma_{PNDE}} \ |a-a^*| \\
SE(\widehat{TNIE}) &= \sqrt{\Gamma_{TNIE}^T \ \boldsymbol{\Sigma} \ \Gamma_{TNIE}} \ |a-a^*| \\
SE(\widehat{TNDE}) &= \sqrt{\Gamma_{TNDE}^T \ \boldsymbol{\Sigma} \ \Gamma_{TNDE}} \ |a-a^*| \\
SE(\widehat{PNIE}) &= \sqrt{\Gamma_{PNIE}^T \ \boldsymbol{\Sigma} \ \Gamma_{PNIE}} \ |a-a^*| \\
SE(\widehat{TE}) &= \sqrt{\Gamma_{TE}^T \ \boldsymbol{\Sigma} \ \Gamma_{TE}} \ |a-a^*| \\
SE(\widehat{PM}) &= \sqrt{\Gamma_{PM}^T \ \boldsymbol{\Sigma} \ \Gamma_{PM}} \ |a-a^*|
\end{aligned}
$$

3.4 Logistic mediator model, non-linear outcome model

3.4.1 Point estimate

$$\begin{aligned}
\text{logit}\ \mathrm{E}[M \mid a, c] &= \beta_0 + \beta_1 a + \beta_2 c + \beta_3 ac \\
\Rightarrow \mathrm{P}[M = 1 \mid a, c] &= \frac{\exp(\beta_0 + \beta_1 a + \beta_2 c + \beta_3 ac)}{1 + \exp(\beta_0 + \beta_1 a + \beta_2 c + \beta_3 ac)} \\
\text{logit}\ \mathrm{P}[Y = 1 \mid a, m, c] &= \theta_0 + \theta_1 a + \theta_2 m + \theta_4 c + \theta_3 am + \theta_5 ac + \theta_6 mc
\end{aligned}$$

For demonstration, we assume the outcome is a logistic model. For other types of outcome model, the effect will be on the log of the corresponding scale in Table 1.

$$OR^{CDE}(m) = (\theta_1 + \theta_3 m + \theta_5 c)(a - a^*) \tag{31}$$

$$\text{If } Y \text{ is rare, } OR^{PNDE} \approx \exp\left\{\log \frac{\mathrm{P}(Y_{a,M^{a^*}} = 1 \mid c)}{\mathrm{P}(Y_{a^*,M^{a^*}} = 1 \mid c)}\right\}$$

$$\begin{aligned}
\mathrm{E}(Y_{a,M^{a^*}} \mid c) &= \mathrm{E}\left[\mathrm{E}(Y \mid a, M, c) \mid a^*, c\right] \\
&\approx \mathrm{E}[\exp(\theta_0 + \theta_1 a + \theta_2 M + \theta_4 c + \theta_3 aM + \theta_5 ac + \theta_6 Mc) \mid a^*, c] \\
&= \exp(\theta_0 + \theta_1 a + \theta_4 c + \theta_5 ac) \cdot \mathrm{E}\left\{\exp[(\theta_2 + \theta_3 a + \theta_6 c)M] \mid a^*, c\right\}
\end{aligned}$$

$\mathrm{E}\left\{\exp[(\theta_2 + \theta_3 a + \theta_6 c)M] \mid a^*, c\right\}$ can be calculated using MGF.
Since

$$\begin{aligned}
&M \sim Ber(p), \\
&\mathrm{E}(\mathrm{e}^{tM}) = MGF_M(t) = 1 - p + p\mathrm{e}^t
\end{aligned}$$

where

$$\begin{aligned}
p &= \mathrm{E}(M \mid a^*, c) = \frac{\exp(\beta_0 + \beta_1 a^* + \beta_2 c + \beta_3 a^* c)}{1 + \exp(\beta_0 + \beta_1 a^* + \beta_2 c + \beta_3 a^* c)}, \\
t &= \theta_2 + \theta_3 a + \theta_6 c
\end{aligned}$$

Plugging in and we have

$$\mathrm{E}\left\{\exp[(\theta_2 + \theta_3 a + \theta_6 c)M] \mid a^*, c\right\} = \frac{1 + \exp(\beta_0 + \beta_1 a^* + \beta_2 c + \beta_3 a^* c + \theta_2 + \theta_3 a + \theta_6 c)}{1 + \exp(\beta_0 + \beta_1 a^* + \beta_2 c + \beta_3 a^* c)}$$

Hence,

$$\begin{aligned}
\mathrm{E}(Y_{a,M^{a^*}} \mid c) &= \mathrm{E}\left[\mathrm{E}(Y \mid a, M, c) \mid a^*, c\right] \\
&= \exp(\theta_0 + \theta_1 a + \theta_4 c + \theta_5 ac) \cdot \frac{1 + \exp(\beta_0 + \beta_1 a^* + \beta_2 c + \beta_3 a^* c + \theta_2 + \theta_3 a + \theta_6 c)}{1 + \exp(\beta_0 + \beta_1 a^* + \beta_2 c + \beta_3 a^* c)}
\end{aligned}$$

Similarly,

$$\begin{aligned}
\mathrm{E}(Y_{a^*,M^{a^*}} \mid c) &= \mathrm{E}\left[\mathrm{E}(Y \mid a^*, M, c) \mid a^*, c\right] \\
&= \exp(\theta_0 + \theta_1 a^* + \theta_4 c + \theta_5 a^* c) \cdot \frac{1 + \exp(\beta_0 + \beta_1 a^* + \beta_2 c + \beta_3 a^* c + \theta_2 + \theta_3 a^* + \theta_6 c)}{1 + \exp(\beta_0 + \beta_1 a^* + \beta_2 c + \beta_3 a^* c)}
\end{aligned}$$

Hence,

$$OR^{PNDE} \approx \exp[(\theta_1 + \theta_5 c)(a - a^*)] \frac{1 + \exp(\beta_0 + \beta_1 a^* + \beta_2 c + \beta_3 a^* c + \theta_2 + \theta_3 a + \theta_6 c)}{1 + \exp(\beta_0 + \beta_1 a^* + \beta_2 c + \beta_3 a^* c + \theta_2 + \theta_3 a^* + \theta_6 c)} \tag{32}$$

For OR^{TNIE},

$$\text{If } Y \text{ is rare, } OR^{TNIE} \approx \exp\left\{\log \frac{\mathrm{P}(Y_{a,M^a}=1 \mid c)}{\mathrm{P}(Y_{a,M^{a^*}}=1 \mid c)}\right\}$$

$$OR^{TNIE} \approx \frac{1+\exp(\beta_0+\beta_1 a+\beta_2 c+\beta_3 ac+\theta_2+\theta_3 a+\theta_6 c)}{1+\exp(\beta_0+\beta_1 a^*+\beta_2 c+\beta_3 a^* c+\theta_2+\theta_3 a+\theta_6 c)} \cdot \frac{1+\exp(\beta_0+\beta_1 a^*+\beta_2 c+\beta_3 a^* c)}{1+\exp(\beta_0+\beta_1 a+\beta_2 c+\beta_3 ac)} \tag{33}$$

For OR^{TNDE},

$$\text{If } Y \text{ is rare, } OR^{TNDE} \approx \exp\left\{\log \frac{\mathrm{P}(Y_{a,M^a}=1 \mid c)}{\mathrm{P}(Y_{a^*,M^a}=1 \mid c)}\right\}$$

$$\begin{aligned}
\mathrm{E}(Y_{a,M^a} \mid c) &= \mathrm{E}\left[\mathrm{E}(Y \mid a, M, c) \mid a, c\right] \\
&= \exp(\theta_0+\theta_1 a+\theta_4 c+\theta_5 ac) \cdot \mathrm{E}\left\{\exp[(\theta_2+\theta_3 a+\theta_6 c)M] \mid a, c\right\} \\
&= \exp(\theta_0+\theta_1 a+\theta_4 c+\theta_5 ac) \cdot \frac{1+\exp(\beta_0+\beta_1 a+\beta_2 c+\beta_3 ac+\theta_2+\theta_3 a+\theta_6 c)}{1+\exp(\beta_0+\beta_1 a+\beta_2 c+\beta_3 ac)}
\end{aligned}$$

Similarly,

$$\begin{aligned}
\mathrm{E}(Y_{a^*,M^a} \mid c) &= \mathrm{E}\left[\mathrm{E}(Y \mid a^*, M, c) \mid a, c\right] \\
&= \exp(\theta_0+\theta_1 a^*+\theta_4 c+\theta_5 a^* c) \cdot \frac{1+\exp(\beta_0+\beta_1 a+\beta_2 c+\beta_3 ac+\theta_2+\theta_3 a^*+\theta_6 c)}{1+\exp(\beta_0+\beta_1 a+\beta_2 c+\beta_3 ac)}
\end{aligned}$$

Hence,

$$OR^{TNDE} \approx \exp[(\theta_1+\theta_5 c)(a-a^*)] \cdot \frac{1+\exp(\beta_0+\beta_1 a+\beta_2 c+\beta_3 ac+\theta_2+\theta_3 a+\theta_6 c)}{1+\exp(\beta_0+\beta_1 a+\beta_2 c+\beta_3 ac+\theta_2+\theta_3 a^*+\theta_6 c)} \tag{34}$$

For OR^{PNIE},

$$\text{If } Y \text{ is rare, } OR^{PNIE} \approx \exp\left\{\log \frac{\mathrm{P}(Y_{a^*,M^a}=1 \mid c)}{\mathrm{P}(Y_{a^*,M^{a^*}}=1 \mid c)}\right\}$$

$$\begin{aligned}
\mathrm{E}(Y_{a^*,M^{a^*}} \mid c) &= \mathrm{E}\left[\mathrm{E}(Y \mid a^*, M, c) \mid a^*, c\right] \\
&= \exp(\theta_0+\theta_1 a^*+\theta_4 c+\theta_5 a^* c) \cdot \mathrm{E}\left\{\exp[(\theta_2+\theta_3 a^*+\theta_6 c)M] \mid a^*, c\right\} \\
&= \exp(\theta_0+\theta_1 a^*+\theta_4 c+\theta_5 a^* c) \cdot \frac{1+\exp(\beta_0+\beta_1 a^*+\beta_2 c+\beta_3 a^* c+\theta_2+\theta_3 a^*+\theta_6 c)}{1+\exp(\beta_0+\beta_1 a^*+\beta_2 c+\beta_3 a^* c)}
\end{aligned}$$

$$\begin{aligned}
\mathrm{E}(Y_{a^*,M^a} \mid c) &= \mathrm{E}\left[\mathrm{E}(Y \mid a^*, M, c) \mid a, c\right] \\
&= \exp(\theta_0+\theta_1 a^*+\theta_4 c+\theta_5 a^* c) \cdot \mathrm{E}\left\{\exp[(\theta_2+\theta_3 a^*+\theta_6 c)M] \mid a, c\right\} \\
&= \exp(\theta_0+\theta_1 a^*+\theta_4 c+\theta_5 a^* c) \cdot \frac{1+\exp(\beta_0+\beta_1 a+\beta_2 c+\beta_3 ac+\theta_2+\theta_3 a^*+\theta_6 c)}{1+\exp(\beta_0+\beta_1 a+\beta_2 c+\beta_3 ac)}
\end{aligned}$$

Therefore,

$$OR^{PNIE} \approx \frac{1+\exp(\beta_0+\beta_1 a+\beta_2 c+\beta_3 ac+\theta_2+\theta_3 a^*+\theta_6 c)}{1+\exp(\beta_0+\beta_1 a^*+\beta_2 c+\beta_3 a^* c+\theta_2+\theta_3 a^*+\theta_6 c)} \cdot \frac{1+\exp(\beta_0+\beta_1 a^*+\beta_2 c+\beta_3 a^* c)}{1+\exp(\beta_0+\beta_1 a+\beta_2 c+\beta_3 ac)} \tag{35}$$

3.4.2 Standard error

Reminder: The formulas are on log odds ratio scale.

$$(a-a^*)\Gamma_{CDE(m)} = \frac{\partial \ \log(OR^{CDE(m)})}{\partial(\beta^T,\theta^T)^T} = (a-a^*)\begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 1 \\ 0 \\ m \\ 0 \\ c \\ 0 \end{bmatrix} \tag{36}$$

$$\Gamma_{PNDE} = \frac{\partial \ \log(OR^{PNDE})}{\partial(\beta^T,\theta^T)^T} = \begin{bmatrix} \begin{gathered}\text{expit}[a^*(\beta_1+\beta_3 c)+a\theta_3+\beta_0+c(\beta_2+\theta_6)+\theta_2] \\ -\text{expit}[a^*(\beta_1+\beta_3 c)+a^*\theta_3+\beta_0+c(\beta_2+\theta_6)+\theta_2]\end{gathered} \\ \\ \begin{gathered}a^*\cdot\Big(\text{expit}[a^*(\beta_1+\beta_3 c)+a\theta_3+\beta_0+c(\beta_2+\theta_6)+\theta_2] \\ -\text{expit}[a^*(\beta_1+\beta_3 c)+a^*\theta_3+\beta_0+c(\beta_2+\theta_6)+\theta_2]\Big)\end{gathered} \\ \\ \begin{gathered}c\cdot\Big(\text{expit}[a^*(\beta_1+\beta_3 c)+a\theta_3+\beta_0+c(\beta_2+\theta_6)+\theta_2] \\ -\text{expit}[a^*(\beta_1+\beta_3 c)+a^*\theta_3+\beta_0+c(\beta_2+\theta_6)+\theta_2]\Big)\end{gathered} \\ \\ \begin{gathered}a^*c\cdot\Big(\text{expit}[a^*(\beta_1+\beta_3 c)+a\theta_3+\beta_0+c(\beta_2+\theta_6)+\theta_2] \\ -\text{expit}[a^*(\beta_1+\beta_3 c)+a^*\theta_3+\beta_0+c(\beta_2+\theta_6)+\theta_2]\Big)\end{gathered} \\ \\ 0 \\ a-a^* \\ \\ \begin{gathered}\text{expit}[a^*(\beta_1+\beta_3 c)+a\theta_3+\beta_0+c(\beta_2+\theta_6)+\theta_2] \\ -\text{expit}[a^*(\beta_1+\beta_3 c)+a^*\theta_3+\beta_0+c(\beta_2+\theta_6)+\theta_2]\end{gathered} \\ \\ \begin{gathered}a\cdot\text{expit}[a^*(\beta_1+\beta_3 c)+a\theta_3+\beta_0+c(\beta_2+\theta_6)+\theta_2] \\ -a^*\cdot\text{expit}[a^*(\beta_1+\beta_3 c)+a^*\theta_3+\beta_0+c(\beta_2+\theta_6)+\theta_2]\end{gathered} \\ \\ 0 \\ c(a-a^*) \\ \\ \begin{gathered}c\cdot\Big(\text{expit}[a^*(\beta_1+\beta_3 c)+a\theta_3+\beta_0+c(\beta_2+\theta_6)+\theta_2] \\ -\text{expit}[a^*(\beta_1+\beta_3 c)+a^*\theta_3+\beta_0+c(\beta_2+\theta_6)+\theta_2]\Big)\end{gathered} \end{bmatrix} \tag{37}$$

$$\Gamma_{TNDE} = \frac{\partial\ \log(OR^{TNDE})}{\partial(\beta^T, \theta^T)^T} = \begin{bmatrix} \text{expit}[a(\beta_1+\beta_3 c)+a\theta_3+\beta_0+c(\beta_2+\theta_6)+\theta_2] \\ -\text{expit}[a(\beta_1+\beta_3 c)+a^*\theta_3+\beta_0+c(\beta_2+\theta_6)+\theta_2] \\ \\ a\cdot\Big(\text{expit}[a(\beta_1+\beta_3 c)+a\theta_3+\beta_0+c(\beta_2+\theta_6)+\theta_2] \\ -\text{expit}[a(\beta_1+\beta_3 c)+a^*\theta_3+\beta_0+c(\beta_2+\theta_6)+\theta_2]\Big) \\ \\ c\cdot\Big(\text{expit}[a(\beta_1+\beta_3 c)+a\theta_3+\beta_0+c(\beta_2+\theta_6)+\theta_2] \\ -\text{expit}[a(\beta_1+\beta_3 c)+a^*\theta_3+\beta_0+c(\beta_2+\theta_6)+\theta_2]\Big) \\ \\ ac\cdot\Big(\text{expit}[a(\beta_1+\beta_3 c)+a\theta_3+\beta_0+c(\beta_2+\theta_6)+\theta_2] \\ -\text{expit}[a(\beta_1+\beta_3 c)+a^*\theta_3+\beta_0+c(\beta_2+\theta_6)+\theta_2]\Big) \\ \\ 0 \\ a-a^* \\ \\ \text{expit}[a(\beta_1+\beta_3 c)+a\theta_3+\beta_0+c(\beta_2+\theta_6)+\theta_2] \\ -\text{expit}[a(\beta_1+\beta_3 c)+a^*\theta_3+\beta_0+c(\beta_2+\theta_6)+\theta_2] \\ \\ a\cdot\text{expit}[a(\beta_1+\beta_3 c)+a\theta_3+\beta_0+c(\beta_2+\theta_6)+\theta_2] \\ -a^*\cdot\text{expit}[a(\beta_1+\beta_3 c)+a^*\theta_3+\beta_0+c(\beta_2+\theta_6)+\theta_2] \\ \\ 0 \\ c(a-a^*) \\ \\ c\cdot\Big(\text{expit}[a(\beta_1+\beta_3 c)+a\theta_3+\beta_0+c(\beta_2+\theta_6)+\theta_2] \\ -\text{expit}[a(\beta_1+\beta_3 c)+a^*\theta_3+\beta_0+c(\beta_2+\theta_6)+\theta_2]\Big) \end{bmatrix} \tag{38}$$

$$
\Gamma_{TNIE} = \frac{\partial \ \log(OR^{TNIE})}{\partial(\beta^T, \theta^T)^T}
$$

$$
= \begin{bmatrix}
\text{expit}[a^*(\beta_1+\beta_3 c)+\beta_0+\beta_2 c] - \text{expit}[a^*(\beta_1+\beta_3 c)+a\theta_3+\beta_0+c(\beta_2+\theta_6)+\theta_2] \\
-\text{expit}[a(\beta_1+\beta_3 c)+\beta_0+\beta_2 c] + \text{expit}[a(\beta_1+\beta_3 c)+a\theta_3+\beta_0+c(\beta_2+\theta_6)+\theta_2] \\
\\
a^* \cdot \Big(\text{expit}[a^*(\beta_1+\beta_3 c)+\beta_0+\beta_2 c] - \text{expit}[a^*(\beta_1+\beta_3 c)+a\theta_3+\beta_0+c(\beta_2+\theta_6)+\theta_2]\Big) \\
+a \cdot \Big(-\text{expit}[a(\beta_1+\beta_3 c)+\beta_0+\beta_2 c] + \text{expit}[a(\beta_1+\beta_3 c)+a\theta_3+\beta_0+c(\beta_2+\theta_6)+\theta_2]\Big) \\
\\
c \cdot \Big(\text{expit}[a^*(\beta_1+\beta_3 c)+\beta_0+\beta_2 c] - \text{expit}[a^*(\beta_1+\beta_3 c)+a\theta_3+\beta_0+c(\beta_2+\theta_6)+\theta_2] \\
-\text{expit}[a(\beta_1+\beta_3 c)+\beta_0+\beta_2 c] + \text{expit}[a(\beta_1+\beta_3 c)+a\theta_3+\beta_0+c(\beta_2+\theta_6)+\theta_2]\Big) \\
\\
c \cdot \Bigg(a^* \cdot \Big(\text{expit}[a^*(\beta_1+\beta_3 c)+\beta_0+\beta_2 c] - \text{expit}[a^*(\beta_1+\beta_3 c)+a\theta_3+\beta_0+c(\beta_2+\theta_6)+\theta_2]\Big) \\
+a \cdot \Big(-\text{expit}[a(\beta_1+\beta_3 c)+\beta_0+\beta_2 c] + \text{expit}[a(\beta_1+\beta_3 c)+a\theta_3+\beta_0+c(\beta_2+\theta_6)+\theta_2]\Big)\Bigg) \\
\\
0 \\
0 \\
\\
\text{expit}[a(\beta_1+\beta_3 c)+a\theta_3+\beta_0+c(\beta_2+\theta_6)+\theta_2] \\
-\text{expit}[a^*(\beta_1+\beta_3 c)+a\theta_3+\beta_0+c(\beta_2+\theta_6)+\theta_2] \\
\\
a \cdot \Big(\text{expit}[a(\beta_1+\beta_3 c)+a\theta_3+\beta_0+c(\beta_2+\theta_6)+\theta_2] \\
-\text{expit}[a^*(\beta_1+\beta_3 c)+a\theta_3+\beta_0+c(\beta_2+\theta_6)+\theta_2]\Big) \\
\\
0 \\
0 \\
\\
c\Big(\text{expit}[a(\beta_1+\beta_3 c)+a\theta_3+\beta_0+c(\beta_2+\theta_6)+\theta_2] \\
-\text{expit}[a^*(\beta_1+\beta_3 c)+a\theta_3+\beta_0+c(\beta_2+\theta_6)+\theta_2]\Big)
\end{bmatrix} \tag{39}
$$

$$\Gamma_{PNIE} = \frac{\partial \ \log(OR^{PNIE})}{\partial(\beta^T, \theta^T)^T}$$

$$= \begin{bmatrix} \begin{gathered} \text{expit}[a^*(\beta_1+\beta_3 c)+\beta_0+\beta_2 c] - \text{expit}[a^*(\beta_1+\beta_3 c)+a^*\theta_3+\beta_0+c(\beta_2+\theta_6)+\theta_2] \\ -\text{expit}[a(\beta_1+\beta_3 c)+\beta_0+\beta_2 c] + \text{expit}[a(\beta_1+\beta_3 c)+a^*\theta_3+\beta_0+c(\beta_2+\theta_6)+\theta_2] \end{gathered} \\ \\ \begin{gathered} a^* \cdot \Big(\text{expit}[a^*(\beta_1+\beta_3 c)+\beta_0+\beta_2 c] - \text{expit}[a^*(\beta_1+\beta_3 c)+a^*\theta_3+\beta_0+c(\beta_2+\theta_6)+\theta_2]\Big) \\ +a \cdot \Big(-\text{expit}[a(\beta_1+\beta_3 c)+\beta_0+\beta_2 c] + \text{expit}[a(\beta_1+\beta_3 c)+a^*\theta_3+\beta_0+c(\beta_2+\theta_6)+\theta_2]\Big) \end{gathered} \\ \\ \begin{gathered} c \cdot \Big(\text{expit}[a^*(\beta_1+\beta_3 c)+\beta_0+\beta_2 c] - \text{expit}[a^*(\beta_1+\beta_3 c)+a^*\theta_3+\beta_0+c(\beta_2+\theta_6)+\theta_2] \\ -\text{expit}[a(\beta_1+\beta_3 c)+\beta_0+\beta_2 c] + \text{expit}[a(\beta_1+\beta_3 c)+a^*\theta_3+\beta_0+c(\beta_2+\theta_6)+\theta_2]\Big) \end{gathered} \\ \\ \begin{gathered} c \cdot \Bigg(a^* \cdot \Big(\text{expit}[a^*(\beta_1+\beta_3 c)+\beta_0+\beta_2 c] - \text{expit}[a^*(\beta_1+\beta_3 c)+a^*\theta_3+\beta_0+c(\beta_2+\theta_6)+\theta_2]\Big) \\ +a \cdot \Big(-\text{expit}[a(\beta_1+\beta_3 c)+\beta_0+\beta_2 c] + \text{expit}[a(\beta_1+\beta_3 c)+a^*\theta_3+\beta_0+c(\beta_2+\theta_6)+\theta_2]\Big)\Bigg) \end{gathered} \\ \\ 0 \\ 0 \\ \\ \begin{gathered} \text{expit}[a(\beta_1+\beta_3 c)+a^*\theta_3+\beta_0+c(\beta_2+\theta_6)+\theta_2] \\ -\text{expit}[a^*(\beta_1+\beta_3 c)+a^*\theta_3+\beta_0+c(\beta_2+\theta_6)+\theta_2] \end{gathered} \\ \\ \begin{gathered} a^* \cdot \Big(\text{expit}[a(\beta_1+\beta_3 c)+a^*\theta_3+\beta_0+c(\beta_2+\theta_6)+\theta_2] \\ -\text{expit}[a^*(\beta_1+\beta_3 c)+a^*\theta_3+\beta_0+c(\beta_2+\theta_6)+\theta_2]\Big) \end{gathered} \\ \\ 0 \\ 0 \\ \\ \begin{gathered} c \cdot \Big(\text{expit}[a(\beta_1+\beta_3 c)+a^*\theta_3+\beta_0+c(\beta_2+\theta_6)+\theta_2] \\ -\text{expit}[a^*(\beta_1+\beta_3 c)+a^*\theta_3+\beta_0+c(\beta_2+\theta_6)+\theta_2]\Big) \end{gathered} \end{bmatrix} \tag{40}$$

$$\begin{aligned} \Gamma_{TE} &= \frac{\partial TE}{\partial(\boldsymbol{\beta}^T, \boldsymbol{\theta}^T)^T} \\ &= \frac{\partial(PNDE + TNIE)}{\partial(\boldsymbol{\beta}^T, \boldsymbol{\theta}^T)^T} \\ &= \Gamma_{PNDE} + \Gamma_{TNIE} \end{aligned}$$

$$\begin{aligned} \Gamma_{PM} &= \frac{\partial PM}{\partial(\boldsymbol{\beta}^T, \boldsymbol{\theta}^T)^T} \\ &= \frac{\partial PM}{\partial PNDE}\frac{\partial PNDE}{\partial(\boldsymbol{\beta}^T, \boldsymbol{\theta}^T)^T} + \frac{\partial PM}{\partial TNIE}\frac{\partial TNIE}{\partial(\boldsymbol{\beta}^T, \boldsymbol{\theta}^T)^T} \\ &= \frac{\partial PM}{\partial PNDE}\Gamma_{PNDE} + \frac{\partial PM}{\partial TNIE}\Gamma_{TNIE} \\ &= -\frac{\exp(PNDE)\left\{\exp(TNIE) - 1\right\}}{\left\{\exp(PNDE)\exp(TNIE) - 1\right\}^2}\Gamma_{PNDE} \\ &\quad + \frac{\exp(PNDE)\exp(TNIE)\left\{\exp(PNDE) - 1\right\}}{\left\{\exp(PNDE)\exp(TNIE) - 1\right\}^2}\Gamma_{TNIE} \end{aligned}$$

$$\mathbf{\Sigma} = \begin{bmatrix} \mathbf{\Sigma}_{\boldsymbol{\beta}} & 0 \\ 0 & \mathbf{\Sigma}_{\boldsymbol{\theta}} \end{bmatrix}$$

Hence,

$$SE(\widehat{CDE}(m)) = \sqrt{\Gamma_{CDE(m)}^T \, \mathbf{\Sigma} \, \Gamma_{CDE(m)}}$$
$$SE(\widehat{PNDE}) = \sqrt{\Gamma_{PNDE}^T \, \mathbf{\Sigma} \, \Gamma_{PNDE}}$$
$$SE(\widehat{TNIE}) = \sqrt{\Gamma_{TNIE}^T \, \mathbf{\Sigma} \, \Gamma_{TNIE}}$$
$$SE(\widehat{TNDE}) = \sqrt{\Gamma_{TNDE}^T \, \mathbf{\Sigma} \, \Gamma_{TNDE}}$$
$$SE(\widehat{PNIE}) = \sqrt{\Gamma_{PNIE}^T \, \mathbf{\Sigma} \, \Gamma_{PNIE}}$$
$$SE(\widehat{TE}) = \sqrt{\Gamma_{TE}^T \, \mathbf{\Sigma} \, \Gamma_{TE}}$$
$$SE(\widehat{PM}) = \sqrt{\Gamma_{PM}^T \, \mathbf{\Sigma} \, \Gamma_{PM}}$$

4 Visual demonstration

In this section, we use a simulated dataset and demonstrate how NDE and NIE depend on covariate levels. The codes of generating the simulated dataset and fitting the models can be found at: `https://github.com/einsley1993/emm-ext-med`.

Data Generating Process

We randomly generate data for A, M, C and Y. For simplicity, we assume there is only one continuous covariate:

$$C \sim N(0, 2^2).$$

We also assume exposure is binary:

$$A \sim Ber\ (\text{expit}(C + C^2)).$$

M and Y can be either continuous or binary, and are generated as the following:

$$Mcont = 0.2 + 0.4A + 0.5C + \beta_3 AC + \epsilon,\ \text{ where } \epsilon \sim N(0, 0.5^2),$$

$$Mbin \sim Ber\ \Big(\text{expit}(0.2 + 0.4A + 0.5C + \beta_3 AC)\Big),$$

$$Ycont = 0.5 + 0.3A + 0.2M + \theta_3 AM + 0.1C + \theta_5 AC + \theta_6 MC + \epsilon, \text{where } \epsilon\ N(0, 0.5^2),$$

$$Ybin \sim Ber\Big(\text{expit}(-5 + 0.3A + 0.2M + \theta_3 AM + 0.1C + \theta_5 AC + \theta_6 MC)\Big),\ \text{Models 1, 2 and 4,}$$

$$Ybin \sim Ber\Big(\text{expit}(-10 + 0.3A + 0.2M + \theta_3 AM + 0.1C + \theta_5 AC + \theta_6 MC)\Big),\ \text{Model 3.}$$

We fit the following mediator and outcome models, depending on the types of Y and M in the DGP. Note that the prevalence of $Ybin$ in the DGP all $< 10\%$, so we fit the logistic outcome model.

Model	Mediator model	Outcome model
1	linear	linear
2	logistic	linear
3	linear	logistic
4	logistic	logistic

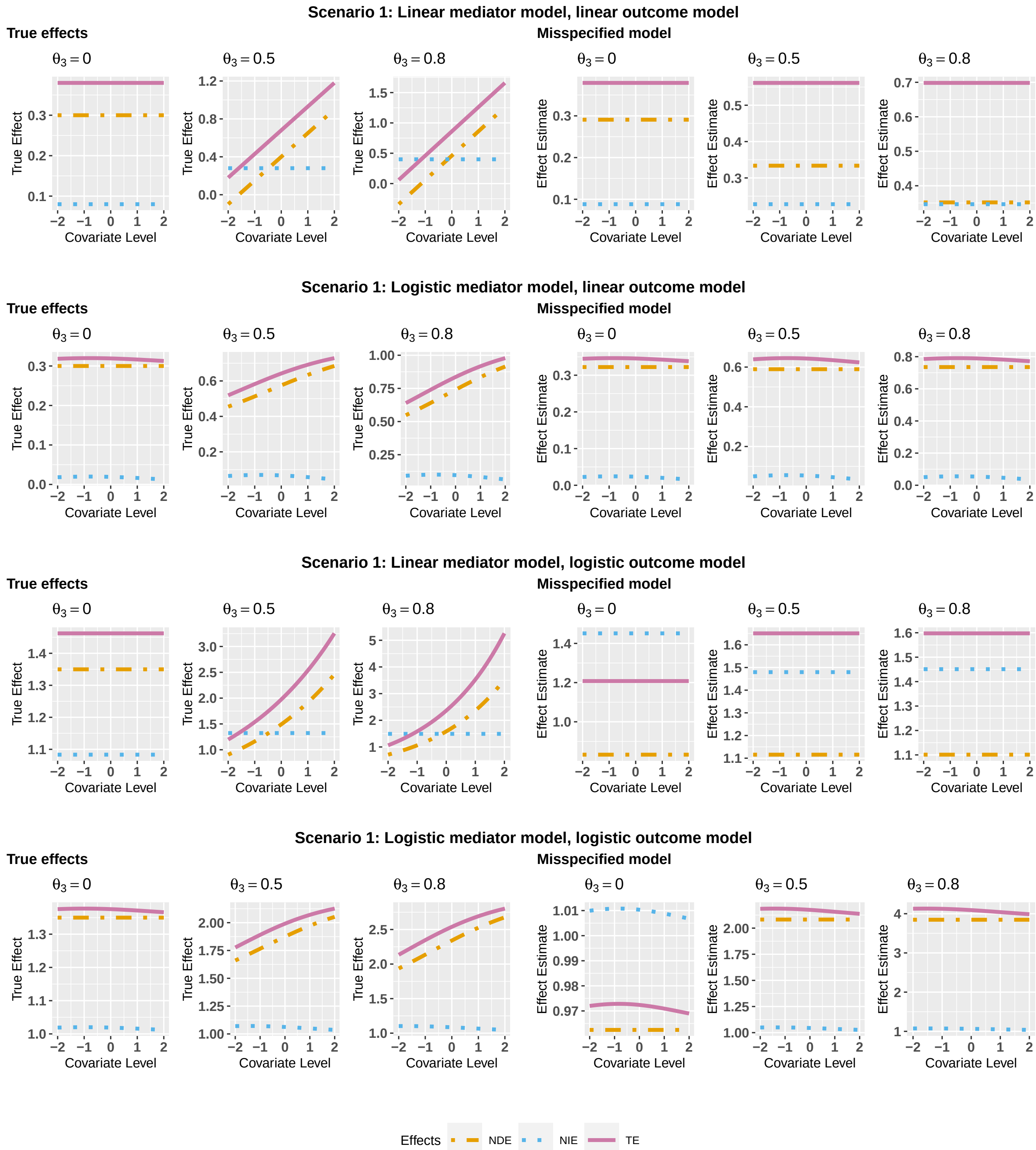
We are mainly interested in how different levels of C modify NDE and NIE. For plotting, we loop over the range of C = [-2, 2], with a step size = 0.01. We are also interested in how the magnitudes of the coefficients of A×M, A×C, M×C modify NDE and NIE, so we vary coefficients $\beta_3, \theta_3, \theta_5, \theta_6$ as well.

Scenario	Product Terms	θ_3	β_3	θ_5	θ_6
1	A×M in mediator	0, 0.5, 0.8	0	0	0
2	A×C in mediator; A×M in outcome	0.5	0.1, 0.4, 0.7	0	0
3	A×C in mediator; A×M and A×C in outcome	0.5	0.2	0.2, 0.5, 0.8	0
4	A×C in mediator; A×M and A×C and M×C in outcome	0.5	0.2	0.3	0.3, 0.6, 0.9

In Sections 4.1 - 4.4, we show how NDE, NIE and TE change as the level of covariate C changes. For each scenario and each model, the three plots on the left are generated from correct model specifications. The three plots on the right are generated from misspecified models, where in Scenario 1, we omit $A \times M$ in outcome model. In Scenarios 2-4, we only keep $A \times M$ in outcome model, but omit $A \times C$ and $M \times C$.

4.1 Scenario 1: only $A \times M$ in outcome model

When only $A \times M$ interaction is present, we vary θ_3(the coefficient of $A \times M$ in outcome model). The 3 columns on the left show the correctly specified model, whereas the 3 columns on the right show the misspecified model, where $A \times M$ interaction is omitted.

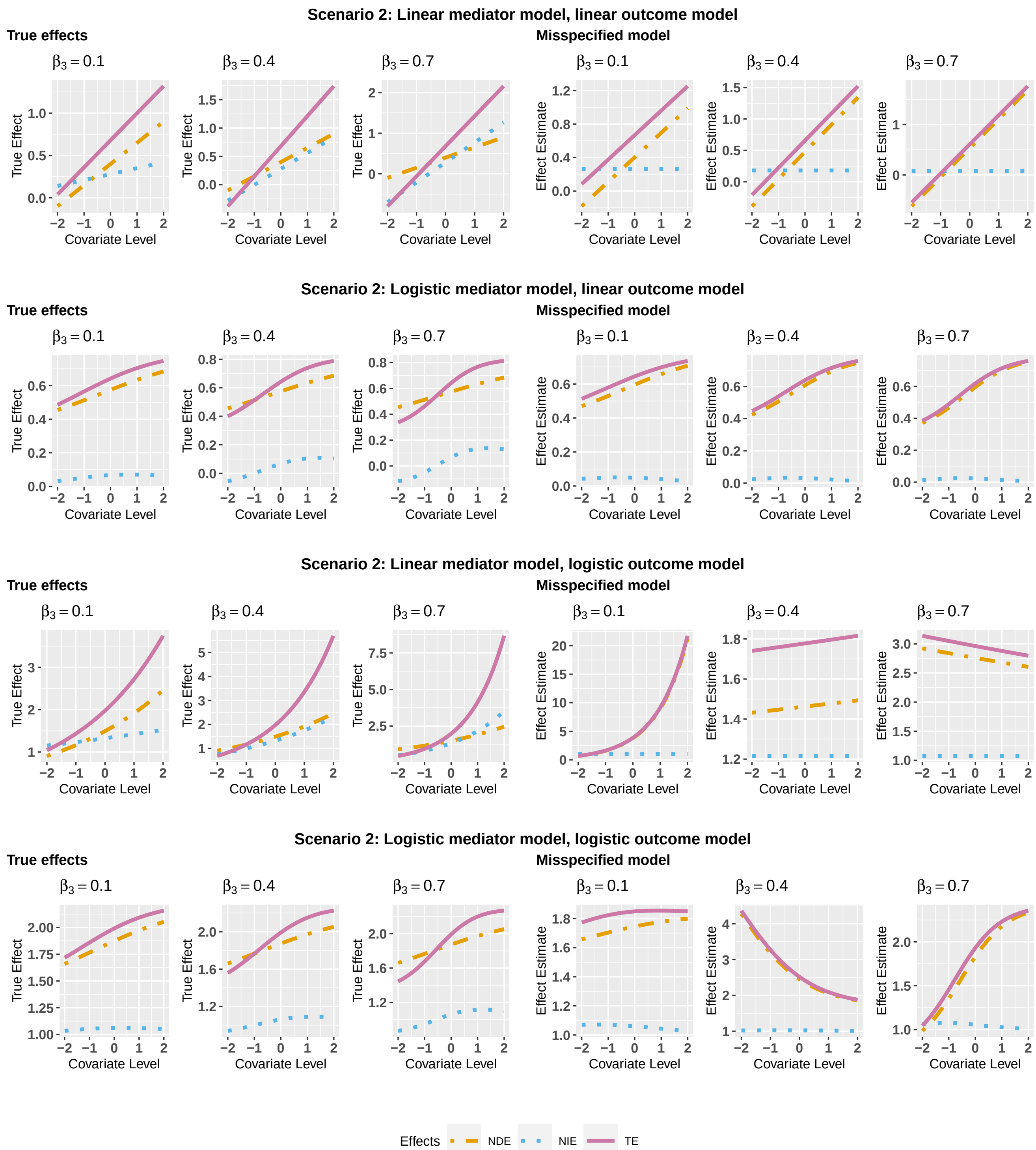


Misspecified models: in Scenario 1, we omit $A \times M$ in outcome model. In Scenarios 2-4, we only keep $A \times M$ in outcome model, but omit $A \times C$ and $M \times C$.

Note that the plots generated from misspecified models oversimplify the patterns in NDE, NIE and TE. For instance, across all four rows, when θ_3 takes either 0.5 or 0.8, the true NDE is not constant (either linear or non-linear, against x-axis), while the biased NDE is.

4.2 Scenario 2: $A \times M$ in outcome model and $A \times C$ in mediator model

When only $A \times M$ interaction and $A \times C$ term in mediator model are present, we fix θ_3 to 0.5 and vary β_3(the coefficient of $A \times C$ in mediator model). The 3 columns on the left shows the correctly specified model, whereas the 3 columns on the right show the misspecified model, where $A \times C$ term is omitted.

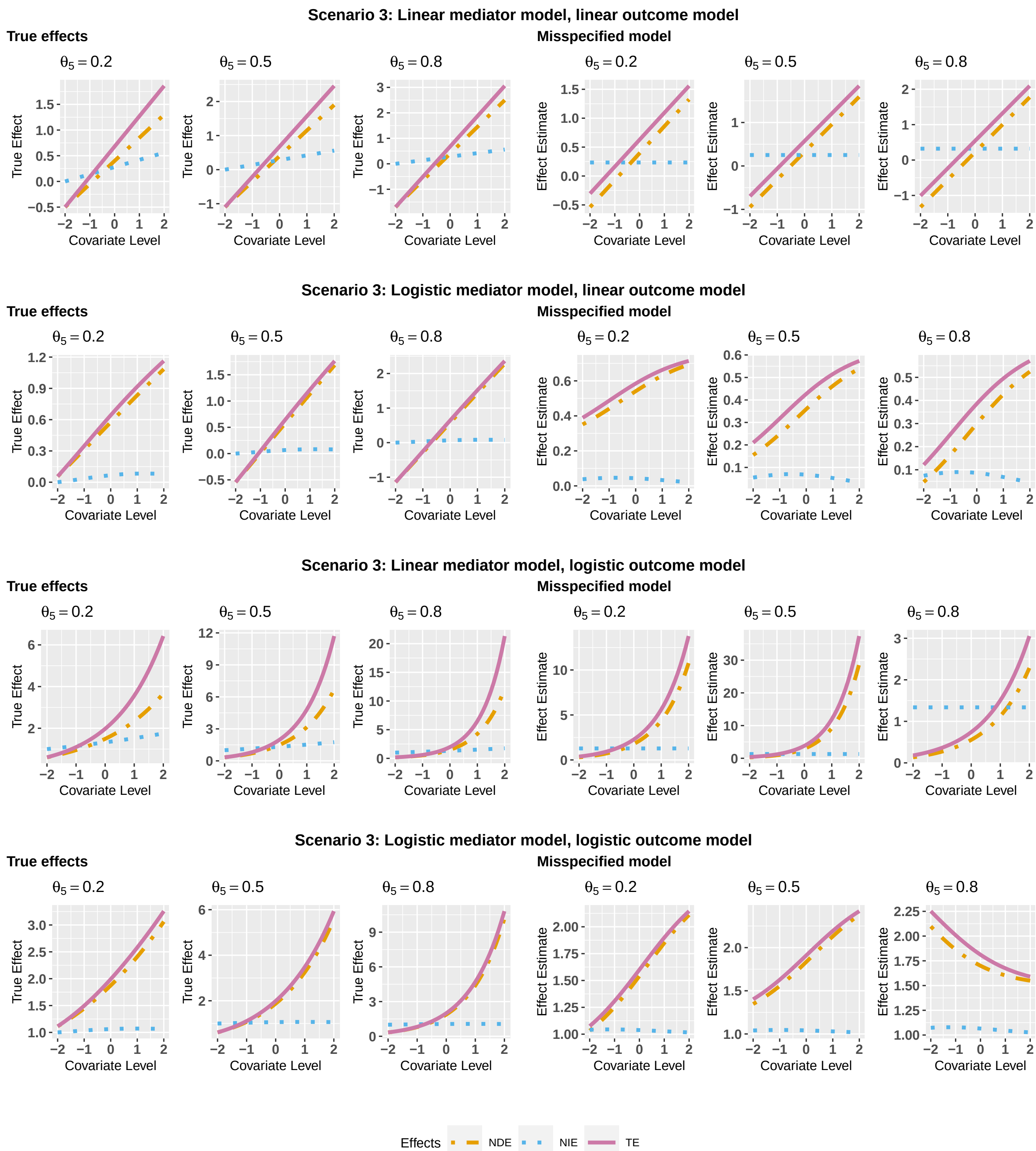


Misspecified models: in Scenario 1, we omit $A \times M$ in outcome model. In Scenarios 2-4, we only keep $A \times M$ in outcome model, but omit $A \times C$ and $M \times C$.

Note that the plots generated from misspecified models oversimplify the patterns in NDE, NIE and TE. For instance, in all four rows, the true NIE changes more dramatically as covariate level changes than the biased NIE.

4.3 Scenario 3: $A \times M$ in outcome model, and $A \times C$ in both mediator and outcome models

When only $A \times M$ interaction, $A \times C$ terms in both mediator and outcome models are present, we fix θ_3 to 0.5, β_3 to 0.2, and vary θ_5(the coefficient of $A \times C$ in outcome model). The 3 columns on the left show the correctly specified model, whereas the 3 columns on the right show the misspecified model, where both $A \times C$ terms are omitted.

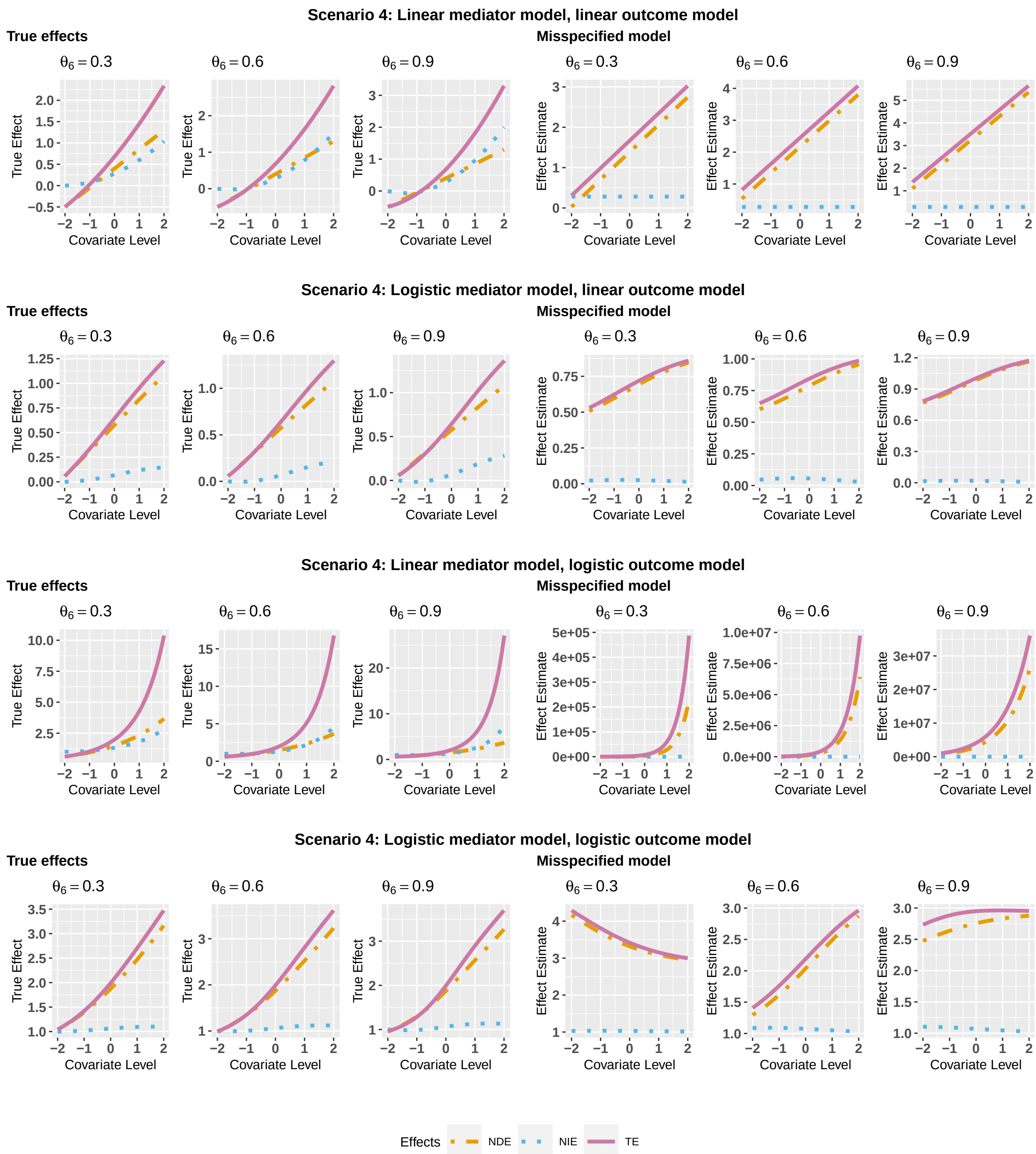


Misspecified models: in Scenario 1, we omit $A \times M$ in outcome model. In Scenarios 2-4, we only keep $A \times M$ in outcome model, but omit $A \times C$ and $M \times C$.

Note that the plots generated from misspecified models oversimplify the patterns in NDE, NIE and TE. For instance, in the first and third rows, the true NIE is not constant when covariate level changes, while the biased NIE is.

4.4 Scenario 4: $A \times M$ in outcome model, $A \times C$ terms in both mediator model and outcome models, and $M \times C$ in outcome model

When $A \times M$ interaction and all three EMM terms are present, we fix θ_3 to 0.5, β_3 to 0.2, and θ_5 to 0.3, and vary θ_6(the coefficient of $M \times C$ in outcome model). The 3 columns on the left show the correctly specified model, whereas the 3 columns on the right show the misspecified model, where all EMM terms are omitted.



Misspecified models: in Scenario 1, we omit $A \times M$ in outcome model. In Scenarios 2-4, we only keep $A \times M$ in outcome model, but omit $A \times C$ and $M \times C$.

Note that the plots generated from misspecified models oversimplify the patterns in NDE, NIE and TE. Note that the plots generated from misspecified models oversimplify the patterns in NDE, NIE and TE. For instance, in all four rows, the true NIE changes more dramatically as covariate level changes than the biased NIE does. In addition to the pattern, omitting EMM can also result in the opposite direction of the relationship between NDE or NIE and covariate. For the plots in the first row, in the bottom left corner, when covariate level is less than -1, true NDE and TE are negative, while the biased NDE and TE are positive.

5 Empirical data example and R package tutorial

In this section, we use an empirical dataset and demonstrate how to use our R package `regmedint` to conduct causal mediation analysis. The R code and the output file can be found at `https://github.com/einsley1993/emm-ext-med-cantos`.

We use the data from a post-hoc mediation analysis from Canakinumab Anti-Inflammatory Thrombosis Outcome Study (CANTOS) trial (Vallurupalli et al., 2020). The dataset consists of 8,302 subjects. We examine the effect of cholesterol level on the risk of heart disease. The binary exposure is canakinumab use (all doses combined vs. placebo), the continuous mediator is the change in log hsCRP from baseline to the 3rd month, and the outcome is time to incident anemia. Hence, we fit a linear mediator model, and an accelerated failure time outcome model (assuming Weibull distribution). The estimated effects will be on mean survival ratio. We adjusted for the same baseline covariates as the published paper (age, sex, log hsCRP, heart failure status, diabetes status, and hypertension status)(Vallurupalli et al., 2020).

The main function in regmedint package is `regmedint()`. Before fitting `regmedint()`, all variables that will be used in analysis (the exposure, mediator, outcome, and confounders) should be coded as numeric, which is already the case in our example. We then specify the statements in `regmedint()` function:

```
fit0.0 <- regmedint(data = anemia3,
                    yvar = "anemiayrs",
                    eventvar = "anemiaev",
                    avar = "canakinumab",
                    mvar = "diff_logCRP",
                    cvar = cvar,
                    ## EMM
                    emm_ac_mreg = NULL,
                    emm_ac_yreg = NULL,
                    emm_mc_yreg = NULL,
                    ##
                    mreg = "linear",
                    yreg = "survAFT_weibull",
                    interaction = TRUE,
                    casecontrol = FALSE,
                    a0 = 0,
                    a1 = 1,
                    m_cde = mean(anemia3$diff_logCRP),
                    c_cond = c_cond0)
summary(fit0.0)
```

Figure 1: Fit `regmedint()` when no EMM terms are included

The above Figure 1 is the fit of `regmedint()` when no EMM term is included. The name of the dataset, "anemia3", is specified in the argument `data`. The names of the outcome, exposure and mediator ("canakinumab", "anemiayrs", "diff_logCRP", respectively) are specified in the arguments `yvar`, `avar`, and `mvar`. The names of the covariates are specified in `cvar`, which can take one or more variable names. We also need to specify the reference level and new level of the exposure that we want to compare using the arguments **a0** and **a1**, respectively. If the exposure is binary, as in our example, then we would simply specify **a1** $= 1$ (i.e., exposed) and **a0** $= 0$ (unexposed). If the exposure is continuous, the two levels specified define the size of the contrast in the exposure that is of interest. For instance, we can choose the 1st quartile as the reference level (**a0**) and the 3rd quartile as the new level (**a1**).

The estimated TE, NDE and NIE can sometimes differ for individuals with different levels of the covariates. Hence, we must specify what levels of covariates we want to condition on, using the argument **c_cond**. A reasonable default choice is to use the sample mean levels of the continuous covariates. For example, if we are interested in the population whose age and baseline log(hsCRP) are both at the mean levels in the study population, we could use the sample mean in the argument **c_cond**. For binary covariates ("female", "chf", "htn", and "dm") in **cvar**, if we chose to obtain mediation estimates for male population with no heart failure and no diabetes, but have hypertension, we set both "female", "chf", "dm" to 0, and "htn" to 1. Note that the values in **c_cond** should correspond to the order that variables are listed in **cvar**. The binary and continuous covariate levels are specified in one vector **c_cond0** in the argument **c_cond** before fitting `regmedint()`, as shown in Figure 2. We must also specify what level of mediator we are interested in. This is used to estimate conditional direct effect (CDE(m)), but not the NDE or NIE. Again, we use the sample mean here.

```
# Condition on other non-EMM covariates:
# female = 0, chf (heart failure) = 0, htn (hyptertention) = 1
c_cond0 <- c(mean(anemia3$age), 0, mean(anemia3$base_logCRP), 0, 1, 0)
c_cond0
```

Figure 2: Covariate levels conditioned on when no EMM terms are included

The statements **mreg** and **yreg** allow us to specify the type of mediator and outcome regression models. The types of mediator models supported are "linear" and "logistic". The types of outcome models supported are "linear", "logistic", "loglinear", "poisson", "negbin" (negative-binomial), "survCox" (Cox proportional hazards), 'survAFT_exp" (accelerated failure time model using an exponential distribution), and "survAFT_weibull" (accelerated failure time model using a Weibull distribution). For survival outcomes, i.e., if **yreg** = "survCox", "survAFT_exp" or "survAFT_weibull", **yvar** will be the time variable. To include an exposure-mediator interaction term in the outcome model, the statement interaction should be set to TRUE, and otherwise should be set to FALSE. Since it is known that including the exposure-mediator interaction captures the possibility that direction or strength of mediation differs by levels of the exposure (VanderWeele, 2015), we would suggest setting **interaction** = TRUE by default. Please also note that including such an interaction term makes the outcome model more flexible and thus imposes fewer assumptions. On the contrary, not including the interaction term requires strong evidence that there is no interaction for all individuals in the study population.

The statement **na_omit** controls whether any missing data, should be removed prior to analysis. This argument is set to FALSE by default, so if there is missing data, the function will return an error message indicating that missing data are not allowed. This is designed to encourage users to check if there is missingness in the main variables of interest. If we instead specify **na_omit** = TRUE, the function will print a message indicating the number of missing values in the dataset, and a complete-case analysis will then be performed. In our example, there are no missing values. Other arguments are available, but are not applicable to our example. The argument **eventvar** is required for time-to-event outcomes. If the event indicator variable in the dataset is called "event", users need to specify **eventvar** = "event". The argument **casecontrol** defaults to FALSE. if the study is a case-control study, it should be set to TRUE.

The modeling and estimation approach discussed so far are applicable to non case-control studies. For case-control studies, special modifications are needed by leaving the outcome model as is but fitting the mediator model only among controls. This requires the outcome to be rare in the population from which the controls are sampled. An alternative way (not implemented in `regmedint`) is to run a weighted mediator regression (different weights for cases and controls). Details of these two modification approaches are discussed in the literature (VanderWeele, 2015).

To see the output of `regmedint()`, we call `summary()`, like in `lm()` and `glm()` functions. The standard outputs of fitted mediator and outcome models will be printed out first, followed by the mediation analysis results.

The results are shown in the output PDF file at `https://github.com/einsley1993/emm-ext-med-cantos/blob/main/EMM_CANTOS.pdf`, **cde** is the controlled direct effect, CDE(m), estimated for the specified mediator value. **pnde** and **tnie** are the NDE and NIE. **tnde** and **pnie** represent an alternative decomposition of total effect, which is not used as frequently as **pnde** and **tnie**; details on the subtleties between the two effect decompositions can be found in (Robins & Greenland, 1992; VanderWeele, 2015). **te** is the total effect, and **pm** is the proportion mediated. If TNDE and PNIE are in opposite directions, PM will be outside the range of $[0, 1]$ and therefore not interpretable (VanderWeele, 2015).

The analysis above allows for exposure-mediator interaction, as we recommend doing by default. If we instead wished to assume there is no such interaction, we could set the argument **interaction** = FALSE.

Now suppose age, baseline log(hsCRP) and diabetes are effect measure modifiers, we can specify the new arguments **emm_ac_mreg**, **emm_ac_yreg**, **emm_mc_yreg**. For the purpose of demonstration, we assume baseline log(hsCRP) modifies the exposure's effect in mediator model, and age and diabetes modify the mediator's effect in outcome model. Therefore, **emm_ac_mreg** is set to c("base_logCRP"), **emm_mc_yreg** is set to c("age", "dm"). As mentioned earlier in this section, **c_cond** specifies the levels of covariates we want to condition on. Suppose we compare two subgroups. Group 1 is the "lower anemia risk" group, containing those who are younger (within 1st quartile of age), having lower baseline log hsCRP (within 1st quartile of baseline log hsCRP), and having no diabetes. Group 2 is the "higher anemia risk" group, containing those who are older (within 3rd quartile of age), having higher log hsCRP (within 3rd quartile of log hsCRP), and having diabetes.

In Figure 3, we specify the covariate levels for both groups.

```
c_cond_low_all <- c_cond_high_all <- c_cond0
c_cond_low_all[c(1, 3, 6)] <- c(quantile(anemia3$age, 0.25),
                                quantile(anemia3$base_logCRP, 0.25),
                                0)
c_cond_high_all[c(1, 3, 6)] <- c(quantile(anemia3$age, 0.75),
                                 quantile(anemia3$base_logCRP, 0.75),
                                 1)
```

Figure 3: Fit `regmedint()` when EMM by age is included in both mediator and outcome models

Now we fit `regmedint()` to examine the mediation effects in those two subgroups. In Figure 4, we set **c_cond** = **c_cond_low_all** as we want to condition on "lower anemia risk" group. By calling `summary()`, we obtain the mediation results in Figure 5.

```
fit.emm <- regmedint(data = anemia3,
                     yvar = "anemiayrs",
                     eventvar = "anemiaev",
                     avar = "canakinumab",
                     mvar = "diff_logCRP",
                     cvar = cvar,
                     ## EMM
                     emm_ac_mreg = c("base_logCRP"),
                     emm_ac_yreg = NULL,
                     emm_mc_yreg = c("age", "dm"),
                     ##
                     mreg = "linear",
                     yreg = "survAFT_weibull",
                     interaction = TRUE,
                     casecontrol = FALSE,
                     a0 = 0,
                     a1 = 1,
                     m_cde = mean(anemia3$diff_logCRP),
                     c_cond = c_cond_low_all)

summary(fit.emm)
```

Figure 4: Fit `regmedint()` when EMM terms are included, subgroup = lower anemia risk

```
## ### Mediation analysis
##            est         se        Z            p        lower     upper
## cde  0.1780520 0.08776683 2.028693 4.248956e-02  0.006032135 0.3500718
## pnde 0.1387042 0.09665386 1.435061 1.512698e-01 -0.050733912 0.3281422
## tnie 0.2068696 0.03998078 5.174225 2.288590e-07  0.128508667 0.2852305
## tnde 0.1671901 0.08655933 1.931508 5.342023e-02 -0.002463110 0.3368432
## pnie 0.1783837 0.04636945 3.847009 1.195687e-04  0.087501223 0.2692661
## te   0.3455737 0.09103109 3.796216 1.469213e-04  0.167156063 0.5239914
## pm   0.6395734 0.17300163 3.696921 2.182302e-04  0.300496404 0.9786503
##
## Evaluated at:
## avar: canakinumab
##  a1 (intervened value of avar) = 1
##  a0 (reference value of avar)  = 0
## mvar: diff_logCRP
##  m_cde (intervend value of mvar for cde) = -2.137102
## cvar: age female base_logCRP chf htn dm
##  c_cond (covariate vector value) = 54 0 1.0116 0 1 0
##
## Note that effect estimates can vary over m_cde and c_cond values when interaction = TRUE.
```

Figure 5: Mediation analyses results for lower anemia risk group

For "higher anemia risk" group, this time we do not need to refit `regmedint()` and re-specify **c_cond**, but just need to call `summary()` and directly set **c_cond = c_cond_high_all** (Figure 6). The results are shown in (Figure 7).

```
summary(fit.emm, c_cond = c_cond_high_all)
```

Figure 6: Fit `regmedint()` when EMM terms are included, subgroup = higher anemia risk

```
## ### Mediation analysis
##            est         se        Z            p       lower     upper
## cde  0.1780520 0.08776683 2.028693 0.0424895616 0.006032135 0.3500718
## pnde 0.1905550 0.09129110 2.087334 0.0368579399 0.011627762 0.3694823
## tnie 0.1521701 0.04534835 3.355582 0.0007919818 0.063288972 0.2410512
## tnde 0.2202737 0.11154304 1.974787 0.0482923802 0.001653350 0.4388940
## pnie 0.1224514 0.05256308 2.329609 0.0198268189 0.019429688 0.2254732
## te   0.3427251 0.09821450 3.489557 0.0004838213 0.150228244 0.5352220
## pm   0.4864715 0.14075463 3.456167 0.0005479164 0.210597461 0.7623455
##
## Evaluated at:
## avar: canakinumab
##  a1 (intervened value of avar) = 1
##  a0 (reference value of avar)  = 0
## mvar: diff_logCRP
##  m_cde (intervend value of mvar for cde) = -2.137102
## cvar: age female base_logCRP chf htn dm
##  c_cond (covariate vector value) = 67 0 1.8871 0 1 1
##
## Note that effect estimates can vary over m_cde and c_cond values when interaction = TRUE.
```

Figure 7: Mediation analyses results for higher anemia risk group

Note that `summary()` gives individual model fits for mediator and outcome models before printing the mediation results. We omit individual model fits in this Supplement material, but readers can find the full outputs at `https://github.com/einsley1993/emm-ext-med-cantos/blob/main/EMM_CANTOS.pdf`.

References

Lange, T., Vansteelandt, S., & Bekaert, M. (2012). A simple unified approach for estimating natural direct and indirect effects. American journal of epidemiology, 176(3), 190–195.

Pearl, J. (2001). Direct and indirect effects. In Proceedings of the seventeenth conference on uncertainty in artificial intelligence (pp. 411–420).

Robins, J. M., & Greenland, S. (1992). Identifiability and exchangeability for direct and indirect effects. Epidemiology, 143–155.

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.

Valeri, L., & VanderWeele, T. J. (2015). Sas macro for causal mediation analysis with survival data. Epidemiology, 26(2), e23–e24.

Vallurupalli, M., MacFadyen, J. G., Glynn, R. J., Thuren, T., Libby, P., Berliner, N., & Ridker, P. M. (2020). Effects of interleukin-1β inhibition on incident anemia: exploratory analyses from a randomized trial. Annals of internal medicine, 172(8), 523–532.

VanderWeele, T. J. (2013). A three-way decomposition of a total effect into direct, indirect, and interactive effects. Epidemiology (Cambridge, Mass.), 24(2), 224.

VanderWeele, T. J. (2014). A unification of mediation and interaction: a four-way decomposition. Epidemiology (Cambridge, Mass.), 25(5), 749.

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

Ver Hoef, J. M. (2012). Who invented the delta method? The American Statistician, 66(2), 124–127.