Received: 3 January 2022 | Revised: 3 September 2022 | Accepted: 21 November 2022

DOI: 10.1002/sim.9621

RESEARCH ARTICLE

Statistics in Medicine WILEY

An exact regression-based approach for the estimation of natural direct and indirect effects with a binary outcome and a continuous mediator

Mariia Samoilenko[1] | **Geneviève Lefebvre**[1,2]

[1]Département de mathématiques, Université du Québec à Montréal, Montréal, Québec, Canada

[2]Faculté de pharmacie, Université de Montréal, Montréal, Québec, Canada

Correspondence
Mariia Samoilenko, Département de mathématiques, Université du Québec à Montréal, 201, Président Kennedy Avenue, PK-5151, Montréal, QC H2X 3Y7, Canada.
Email: samoilenko.mariia@uqam.ca

Funding information
Fonds de Recherche du Québec - Santé, Grant/Award Number: 268860; Natural Sciences and Engineering Research Council of Canada, Grant/Award Number: RGPIN-2020-05473

In the causal mediation framework, a number of parametric regression-based approaches have been introduced in recent years for estimating natural direct and indirect effects for a binary outcome in an exact manner, without invoking simplifying assumptions based on the rareness or commonness of the outcome. However, most of these works have focused on a binary mediator. In this article, we aim at a continuous mediator and introduce an exact approach for the estimation of natural effects on the odds ratio, risk ratio, and risk difference scales. Our approach relies on logistic and linear models for the outcome and mediator, respectively, and uses numerical integration to calculate the nested counterfactual probabilities underlying the definition of natural effects. Formulas for the delta method standard errors for all effects estimators are provided. The performance of our proposed exact estimators was evaluated in simulation studies that featured scenarios with different levels of outcome rareness/commonness, including a marginally but not conditionally rare outcome scenario. Furthermore, we evaluated the merit of Firth's penalization to mitigate the bias in the logistic regression coefficients estimators for the smallest outcome prevalences and sample sizes investigated. Using a SAS macro provided, we implemented our approach to assess the effect of placental abruption on low birth weight mediated by gestational age. We found that our exact natural effects estimators worked properly in both simulated and real data applications.

KEYWORDS

binary outcome, continuous mediator, exact natural effects estimators, outcome rareness/commonness, regression-based causal mediation analysis

1 | INTRODUCTION

Natural direct and indirect effects are the cornerstone of causal mediation analysis. These well-known quantities are expressed using contrasts of counterfactual outcomes. Specifically, define the nested counterfactual outcome $Y(a, M(a^*))$ as the outcome that would be observed if exposure A has been set to a and mediator M has been set to the value it

Abbreviations: CI, confidence interval; EPV, number of events per variable; NDE, natural direct effect; NEM, natural effect model; NIE, natural indirect effect; OR, odds ratio; RD, risk difference; ROA, rare outcome assumption; RR, risk ratio; TE, total effect.

would have taken if the exposure had been set to a^*. Then the natural direct effect (NDE) compares $E[Y(a, M(a^*))]$ with $E[Y(a^*, M(a^*))]$, while the natural indirect effect (NIE) compares $E[Y(a, M(a))]$ with $E[Y(a, M(a^*))]$. Under several assumptions, the nested counterfactual expectation $E[Y(a, M(a^*))]$ is non-parametrically identified using the *mediation formula*.[1-3] The mediation formula allows for a univocal definition of the natural effects since it is not tied to specific models for the outcome Y and mediator M. However this conceptual flexibility comes at a price since computing $E[Y(a, M(a^*))]$ can be challenging even for standard outcome and mediator models.[4]

When logistic and linear regression models are respectively used to model a binary outcome and a continuous mediator, the natural effects are expressed using integrals that do not have closed-form expressions.[5] To circumvent this inconvenience, VanderWeele and Vansteelandt[6] used a series of approximations invoking the so-called rare outcome assumption (ROA) to derive closed-form expressions for the NDE and NIE on the odds ratio (OR) scale (OR^{NDE} and OR^{NIE}). In practice, a 10% threshold for the outcome prevalence (that is, $P(Y = 1) = 0.1$) is often used for qualifying an outcome as rare. VanderWeele and Vansteelandt's approximate approach was subsequently implemented in the well-known SAS macro `mediation` by Valeri and VanderWeele,[7] the SAS procedure PROC CAUSALMED[8] and in the novel R package `CMAverse`.[9]

Under the above standard specification of the binary outcome and continuous mediator models, Gaynor et al[5] instead focused on a common outcome and derived approximate formulas for the NDE and NIE by exploiting a relationship between logit and probit models. More precisely, they suggested approximating $\exp(x)/(1 + \exp(x))$ by $\Phi(sx)$ for $s > 0$, where $\Phi(\cdot)$ is the normal cumulative distribution function, thereby allowing for closed-form formulas for the OR^{NDE} and OR^{NIE}. The parameter s can be chosen to minimize the distance between the logistic and normal cumulative distribution functions (eg, minimax solution[10]) or using some statistical criteria (eg, equality of variances[11] or Kullback-Leibler information criterion[12]). In the literature, proposed values for this scaling constant range between 0.551 and 0.625.[12] In Gaynor et al,[5] the authors proposed estimating s from the data by comparing the regression coefficients obtained by fitting a probit outcome model to those obtained from a logistic outcome model. A simulation study performed in Gaynor et al[5] demonstrated the adequate performance of their approximate approach for outcome prevalences between 20% and 60%.

Two practical questions can immediately be raised in this binary outcome and continuous mediator context. First, as only a subrange in outcome prevalence is covered by these two approximate approaches, one may ask which strategy to adopt when the prevalence is outside these bounds (eg, when $P(Y = 1) = 0.15$ for instance). Second, one can also legitimately ask whether it is sufficient that the outcome be rare marginally. One possible answer to the first question is to use either approximate approach to obtain natural effects estimates, and potentially both to assess the robustness of the results. However, this solution is not completely satisfying since each may have a suboptimal performance when underlying hypotheses are not respected. Second, in the context of a binary outcome and a binary mediator, Samoilenko, Blais and Lefebvre[13] have brought to attention that it is not enough that the outcome be rare marginally, that is, it also requires that the outcome be rare conditionally to the mediator. Indeed, these authors have shown that large biases in the natural effects estimates can be obtained when using a mediation regression-based approach that invokes the ROA when the outcome is only marginally rare. We can therefore suspect that similar problems prevail using VanderWeele and Vansteelandt's approach with a continuous mediator.[6]

Recently, there has been a strong interest in the development of so-called exact regression-based estimators for natural effects, where the term *exact* refers to estimators that are developed without any theoretical simplifying assumptions and whose accuracy is defined, beyond sample size consideration, by the numeric precision of the software tools utilized and the default/user-specified tolerance of the routines applied. However most effort has been concentrated on the case of a binary outcome and a binary mediator.[13-16] Notably, Samoilenko and Lefebvre[14] proposed exact estimators for the natural effects without invoking the rareness or commonness of the outcome, thereby addressing the inherent difficulty in assessing the adequacy of the ROA in a mediation setting. These natural effects estimators are based on the specification of a logistic model for both the outcome and mediator, and are expressed on the OR, risk ratio (RR), and risk difference (RD) scales. Samoilenko and Lefebvre's exact approach[14] was found well-performing in simulation scenarios ranging from a rare to a common outcome, including a marginally but not conditionally rare outcome. In a very recent paper, Cheng, Spiegelman and Li[16] compared exact and approximate natural effects estimators[5-7] on the (log) OR scale in different simulation scenarios when the mediator was either binary or continuous. For the case of a continuous mediator, Cheng et al[16] studied the performance and numerical stability of the proposed exact natural effects estimators through a simulation study without covariates in which scenarios varied as a function of the number of outcome cases and sample size; in all the scenarios considered, the outcome was rare marginally (maximal prevalence $\approx$ 6.1%). The authors globally concluded favorably regarding exact estimators, although for a continuous mediator, few differences between

the exact and approximate estimators by VanderWeele and Vansteelandt[6] were observed in the simulation settings with a rare marginal outcome they considered.

The present work is a follow up of Samoilenko and Lefebvre[14] and Cheng et al.[16] As in Cheng et al,[16] we propose to numerically solve the integrals underlying the natural effects estimators when the outcome and mediator are respectively modeled using logistic and linear regressions. In our article however, we allow the outcome model to include an interaction term between the exposure and mediator—unlike in Cheng et al[16]—so to align with a more common specification of this model in causal mediation. Moreover, we introduce exact estimators for the natural effects on the OR, RR, and RD scales, thereby going beyond the typical (log) OR scale. In our simulation study, we examine the performance of the proposed exact natural effects estimators in scenarios with different levels of outcome rareness/commonness, both with and without covariates. For the OR scale, we compare our proposed approach with the approximate approaches by VanderWeele and Vansteelandt[6] and Gaynor et al.[5] To provide additional benchmarks, we compare, when possible, our exact regression based-approach to two other approaches which also do not rely on the rareness or commonness of the outcome, namely the natural effect model (NEM) approach by Lange et al[17] and Imai et al's[2] parametric inference algorithm based on quasi-Bayesian Monte Carlo approximations.

When applying a logistic regression to small samples and/or sparse data, conventional maximum likelihood estimation methods are prone to be biased or to produce infinite coefficients estimates because of complete or quasi-complete separation.[18,19] In Cheng et al,[16] numerical problems were observed for small sample sizes and low outcome prevalences. Therefore, an additional objective of this article was to investigate the impact of Firth's penalization, a popular approach widely implemented in statistical software to deal with aforementioned estimation problems,[20] on the exact natural effects estimators proposed.

2 | METHODS

2.1 | Models and nested counterfactual outcome probabilities

Let us note A the exposure (binary or continuous). As in VanderWeele and Vansteelandt,[6] Valeri and VanderWeele,[7] and Gaynor et al,[5] we assume the following linear and logistic regression models for the continuous mediator M and binary outcome Y, respectively:

$$E\{M|A=a,\boldsymbol{C}=\boldsymbol{c}\}=\beta_0+\beta_1 a+\boldsymbol{\beta_2'c},\quad \epsilon\sim\mathcal{N}(0,\sigma^2), \tag{1}$$

$$\text{logit}\{P(Y=1|A=a,M=m,\boldsymbol{C}=\boldsymbol{c})\}=\theta_0+\theta_1 a+\theta_2 m+\theta_3 am+\boldsymbol{\theta_4'c}, \tag{2}$$

where $\boldsymbol{C}$ is a set of pre-exposure covariates sufficient to control for exposure-outcome, mediator-outcome, and exposure-mediator confounding.

Under identification assumptions (see Appendix A) and modeling assumptions in Equations (1) and (2), the conditional nested counterfactual outcome probabilities $P(Y(a,M(a^*)|\boldsymbol{C}=\boldsymbol{c})$ for all possible values of a and a^* can be expressed as follows:

$$\begin{aligned}P(Y(a,M(a^*))=1|\boldsymbol{C}=\boldsymbol{c})&=\int_{-\infty}^{\infty}P(Y=1|A=a,M=m,\boldsymbol{C}=\boldsymbol{c})\,d\Phi_{M|A=a^*,\boldsymbol{C}=\boldsymbol{c}}(m)\\&=\frac{1}{\sqrt{2\pi\sigma^2}}\int_{-\infty}^{\infty}\text{expit}\left(\theta_0+\theta_1 a+\theta_2 m+\theta_3 am+\boldsymbol{\theta_4'c}\right)\exp\left(-\frac{\left(m-\left(\beta_0+\beta_1 a^*+\boldsymbol{\beta_2'c}\right)\right)^2}{2\sigma^2}\right)dm,\end{aligned} \tag{3}$$

where $\Phi_{M|A=a^*,\boldsymbol{C}=\boldsymbol{c}}$ is the cumulative distribution function of the normal distribution $\mathcal{N}\left(\beta_0+\beta_1 a^*+\boldsymbol{\beta_2'c},\sigma^2\right)$ and $\text{expit}(\alpha)=\exp(\alpha)/(1+\exp(\alpha))$.

The mediation formula[1-3] allows expressing the NDE and NIE ORs (OR^{NDE}, OR^{NIE}), the NDE and NIE RRs (RR^{NDE}, RR^{NIE}), as well as the NDE and NIE RDs (RD^{NDE}, RD^{NIE}) in terms of the nested counterfactual outcome probabilities (3) in an exact manner; for a change in the exposure level from $A=a^*$ to $A=a$, these effects are:

$$OR^{\text{NDE}}_{a,a^*|\boldsymbol{c}}=\frac{g(a,a^*,\boldsymbol{c})/(1-g(a,a^*,\boldsymbol{c}))}{g(a^*,a^*,\boldsymbol{c})/(1-g(a^*,a^*,\boldsymbol{c}))},\quad OR^{\text{NIE}}_{a,a^*|\boldsymbol{c}}=\frac{g(a,a,\boldsymbol{c})/(1-g(a,a,\boldsymbol{c}))}{g(a,a^*,\boldsymbol{c})/(1-g(a,a^*,\boldsymbol{c}))}, \tag{4}$$

$$RR^{\text{NDE}}_{a,a^*|\boldsymbol{c}} = \frac{g(a,a^*,\boldsymbol{c})}{g(a^*,a^*,\boldsymbol{c})}, \quad RR^{\text{NIE}}_{a,a^*|\boldsymbol{c}} = \frac{g(a,a,\boldsymbol{c})}{g(a,a^*,\boldsymbol{c})}, \tag{5}$$

$$RD^{\text{NDE}}_{a,a^*|\boldsymbol{c}} = g(a,a^*,\boldsymbol{c}) - g(a^*,a^*,\boldsymbol{c}), \quad RD^{\text{NIE}}_{a,a^*|\boldsymbol{c}} = g(a,a,\boldsymbol{c}) - g(a,a^*,\boldsymbol{c}), \tag{6}$$

where

$$g(a,a^*,\boldsymbol{c}) = \frac{1}{\sqrt{2\pi\sigma^2}}\int_{-\infty}^{\infty} \text{expit}\left(\theta_0 + \theta_1 a + \theta_2 m + \theta_3 am + \boldsymbol{\theta_4'c}\right) \exp\left(-\frac{\left(m - \left(\beta_0 + \beta_1 a^* + \boldsymbol{\beta_2'c}\right)\right)^2}{2\sigma^2}\right) dm. \tag{7}$$

The improper integral in Equation (7) reduces to the first moment of the logit-normal distribution for which no closed-form expression exists; therefore, numerical integration is needed to evaluate $g(a,a^*,\boldsymbol{c})$.

The total effect (TE) odds and risk ratios, $OR^{\text{TE}}_{a,a^*|\boldsymbol{c}}$ and $RR^{\text{TE}}_{a,a^*|\boldsymbol{c}}$, are defined as the product of the NDE and NIE on their respective scale:

$$OR^{\text{TE}}_{a,a^*|\boldsymbol{c}} = OR^{\text{NDE}}_{a,a^*|\boldsymbol{c}} \times OR^{\text{NIE}}_{a,a^*|\boldsymbol{c}}, \quad RR^{\text{TE}}_{a,a^*|\boldsymbol{c}} = RR^{\text{NDE}}_{a,a^*|\boldsymbol{c}} \times RR^{\text{NIE}}_{a,a^*|\boldsymbol{c}}. \tag{8}$$

On the RD scale, the TE, $RD^{\text{TE}}_{a,a^*|\boldsymbol{c}}$, is defined as the sum of the NDE and NIE:

$$RR^{\text{TE}}_{a,a^*|\boldsymbol{c}} = RR^{\text{NDE}}_{a,a^*|\boldsymbol{c}} + RR^{\text{NIE}}_{a,a^*|\boldsymbol{c}}. \tag{9}$$

For each scale, the exact NDE and NIE estimators are obtained from Equations (4)-(7) by replacing the parameters involved in Equation (7) with corresponding estimators: least squares estimators for β_0, β_1, $\boldsymbol{\beta_2'}$, maximum likelihood estimators for θ_0, θ_1, θ_2, θ_3, $\boldsymbol{\theta_4'}$, and mean squared error for σ^2. Note that our approach requires consistent estimators for all of the above population regression parameters, which can be readily achieved with cohort data, but not with case-control data. The integration involved in Equation (7) can then be performed via numerical quadrature or other techniques devised for solving one-dimensional integrals. To perform numerical integration in this study, we used the SAS QUAD subroutine[21] that implements adaptive Romberg-type integration techniques and which is devised to deal with singularities, functions with large derivatives, and infinite domains.[21,22] Adaptive Romberg-type integration techniques are notably known to have advantages over Gauss-Hermite and Gauss-Laguerre quadratures for infinite intervals[21] (such as in Equation (7)).

The formulas for calculating the standard errors via the first-order multivariate delta method[23] are provided in Appendix B. The theoretical convergence of the improper integrals involved in the delta method is established in Appendix C.

2.2 | Simulation studies

2.2.1 | Main simulation studies

We performed two simulation studies to examine the performance of the proposed exact natural effects estimators. In the first simulation study, no covariates $\boldsymbol{C}$ were included for the sake of simplicity (*Crude simulation study*), while two covariates were included in the second study (*Adjusted simulation study*).

Data-generating mechanisms and simulation design

In the crude simulation study, the binary exposure A and the continuous mediator M were simulated from a *Bernoulli* (p_A) and a $\mathcal{N}\left(\beta_0 + \beta_1 a,\ \sigma^2\right)$ distributions, respectively, where $p_A = 0.3$, $\beta_0 = 0.1$, $\beta_1 = 0.5$, and $\sigma^2 = 0.5^2$. The binary outcome Y was simulated as a *Bernoulli* (p_Y), with $p_Y = \text{expit}\,(\theta_0 + \theta_1 a + \theta_2 m + \theta_3 am)$, $\theta_1 = 0.4$, $\theta_2 = 0.5$, and $\theta_3 = 0.15$. We considered five simulation scenarios (*Crude scenarios 1-5*) corresponding to $\theta_0 \in \{-3,\ -2,\ -0.5,\ 1,\ 2\}$; these values of θ_0 were chosen to allow for different levels of outcome rareness/commonness. More precisely, the five specified θ_0 values yielded the following estimated marginal outcome prevalences: 6.66%, 15.95%, 44.48%, 77.24%, and 90.03%, respectively. These prevalences were estimated from large datasets of 10^7 observations simulated using the data-generating mechanisms described above.

In the adjusted simulation study, a binary variable C_1 and a continuous variable C_2 were first generated independently from a *Bernoulli* $\left(p_{C_1}\right)$ and a $\mathcal{N}(\mu_{C_2},\ \sigma^2_{C_2})$ distributions, respectively, where $p_{C_1} = 0.5$, $\mu_{C_2} = 0$ and $\sigma^2_{C_2} = 0.75^2$. Second, we generated the binary exposure A as a *Bernoulli* (p_A), where $p_A = \text{expit}\,(\alpha_0 + \alpha_1 c_1 + \alpha_2 c_2)$, $\alpha_0 = -0.85$, $\alpha_1 = 0.1$, and $\alpha_2 =$

−0.15. Then, the continuous mediator M was generated from a $\mathcal{N}\left(\beta_0 + \beta_1 a + \beta_{21} c_1 + \beta_{22} c_2,\ \sigma^2\right)$ distribution, where $\beta_0 = 0.1$, $\beta_1 = 0.5$, $\beta_{21} = 0.1$, $\beta_{22} = 0.2$, and $\sigma^2 = 0.5^2$. Finally, the binary outcome Y was generated as a *Bernoulli* (p_Y), where $p_Y = \text{expit}\,(\theta_0 + \theta_1 a + \theta_2 m + \theta_3 am + \theta_{41} c_1 + \theta_{42} c_2)$, $\theta_1 = 0.4$, $\theta_2 = 0.5$, $\theta_3 = 0.15$, $\theta_{41} = 0.2$, and $\theta_{42} = 0.1$. We considered the same five values of θ_0 as in the crude simulation study (*Adjusted scenarios 1-5*), which yielded the following estimated marginal outcome prevalences: 7.64%, 17.94%, 47.71%, 79.28%, and 91.03%.

For both simulation studies, the known values of the simulation parameters were used to numerically calculate the true natural effects on the OR, RR, and RD scales according to Equations (4)-(9), where $a = 1$, $a^* = 0$, $\boldsymbol{\beta_2'} = \boldsymbol{\theta_4'} = \boldsymbol{c'} = (0, 0)$ for the crude scenarios, and $\boldsymbol{\beta_2'} = (\beta_{21}, \beta_{22})$, $\boldsymbol{\theta_4'} = (\theta_{41}, \theta_{42})$, and $\boldsymbol{c'} = (p_{C_1}, \mu_{C_2})$ for the adjusted scenarios. For these calculations, we also used the SAS QUAD subroutine.[21]

Description of analyzes

For each simulation scenario, 1000 independent samples of size $n = 5000$ were generated using the SAS/IML software (SAS Institute Inc., Cary, NC, USA) with initial seed 1234. For each sample generated, correctly specified linear and logistic regressions were fitted for the mediator and the outcome, respectively. Natural effects were then obtained on the OR, RR, and RD scales using the corresponding exact estimators.

For the OR scale, the proposed exact mediation approach was compared to the regression-based approaches by VanderWeele and Vansteelandt[6] and Gaynor et al.[5] For the approach by Gaynor et al,[5] we used these authors' strategy for the choice of s, namely taking the median of the ratios of the coefficients of a probit outcome model compared to respective coefficients of a logistic outcome model. We implemented both approximate approaches using the same linear model for the mediator and logistic model for the outcome. We also compared the exact approach to the NEM approach by Lange et al[17] and Imai et al's[2] parametric inference algorithm based on quasi-Bayesian Monte Carlo approximations, which, recall, also do not rely on the outcome rareness/commonness. The NEM approach was implemented using the R package `medflex`[24] for both multiplicative scales. Two technical procedures to handle the missingness in the counterfactual outcomes, namely weighting and imputation, are implemented in `medflex`; we used the imputation-based approach as recommended by Steen et al[24] when dealing with a continuous mediator. For the OR scale, NEMs were fitted using a logistic regression with a predictor function that matched that in the logistic outcome model used for the exact approach. In particular, in the adjusted simulation study, the NEMs were fitted with covariates as main effect terms. For the RR scale, a Poisson regression was used for the NEMs since numeric problems were obtained for the log-linear regression. For both multiplicative scales, the working model for the imputation was logistic and mimicked the specification of the corresponding NEM. For the RD scale, Imai et al's[2] approach was implemented with the R package `mediation`,[25] and using the same mediator and outcome models as in the exact approach. The number of quasi-Bayesian Monte Carlo simulations was set to 1000 for each sample generated.

For each simulation scenario, the mean value, bias, relative bias, standard deviation, and root mean squared error of all estimators considered were obtained over the 1000 samples simulated. In the adjusted simulation study, we estimated all natural effects at the sample-specific mean values of C_1 and C_2. It should be noted that, in the absence of the exposure-covariate interaction terms in the NEM, the conditional natural effects returned by `medflex` are the same for any level of adjustment covariates.[24] The coverage probabilities of the 95% confidence intervals (CIs) were computed by calculating the proportion of times when CIs included corresponding true values of the natural effects. For the exact approach and the approximate approach by VanderWeele and Vansteelandt,[6] 95% CIs were constructed by percentile bootstrap[26] based on 500 resamples with replacement and using the first-order delta method.[23] For the approximate approach by Gaynor et al,[5] 95% CIs were obtained by percentile bootstrap only. For the NEM approach, 95% CIs were constructed using robust SEs based on the sandwich estimator.[27] For the parametric inference algorithm by Imai et al,[2] 95% CIs were based on White's heteroskedasticity-consistent estimator for the covariance matrix.[28]

2.2.2 | Simulation study with a marginal but not conditional rare outcome

As in Samoilenko and Lefebvre,[13] we also performed a simulation study to examine the performance of the proposed exact estimators when the ROA was markedly violated in some strata formed by the exposure and mediator, while the ROA was satisfied marginally. In order to accomplish this, we repeated the main adjusted simulation study, but under the following specification for the simulation parameters values: $p_{C_1} = 0.036$, $\mu_{C_2} = 27.896$, $\sigma_{C_2} = 5.690$, $\alpha_0 = -3.831$, $\alpha_1 = 0.589$, $\alpha_2 = 0.018$, $\beta_0 = 39.003$, $\beta_1 = -1.823$, $\beta_{21} = -0.579$, $\beta_{22} = -0.013$, $\sigma = 2.003$, $\theta_0 = 36.950$, $\theta_1 = 1.484$, $\theta_2 = -1.066$, $\theta_3 = -0.025$, $\theta_{41} = -0.792$, and $\theta_{42} = 0.014$. These values yielded

estimated marginal and conditional outcome prevalences of: $\hat{P}(Y=1)=9.10\%$, $\hat{P}(Y=1|A=0,M\in S_1)=28.29\%$, $\hat{P}(Y=1|A=0,M\in S_2)=4.72\%$, $\hat{P}(Y=1|A=0,M\in S_3)=1.20\%$, $\hat{P}(Y=1|A=0,M\in S_4)=0.21\%$, $\hat{P}(Y=1|A=1,M\in S_1)=50.53\%$, $\hat{P}(Y=1|A=1,M\in S_2)=8.34\%$, $\hat{P}(Y=1|A=1,M\in S_3)=2.06\%$, $\hat{P}(Y=1|A=1,M\in S_4)=0.46\%$, where $S_1=\{m:\ m\le Q_1\}, S_2=\{m:\ Q_1<m\le Q_2\}, S_3=\{m:\ Q_2<m\le Q_3\}, S_4=\{m:\ m>Q_3\}$, and Q_1, Q_2, and Q_3 are respectively the first, second and third quartiles of the mediator distribution. As in the main simulation study, these prevalences were estimated from large datasets of 10^7 observations simulated using the aforementioned data-generating mechanisms. Therefore, in this additional simulation study, the outcome Y was rare marginally but not rare conditionally: the ROA was significantly violated in two strata defined by the exposure and mediator (more precisely, when $A=0,1$, and $M\in S_1$). As reference, in *Adjusted scenario 1* of the main simulation study with covariates, the two largest conditional probabilities obtained were $\hat{P}(Y=1|A=1,M\in S_3)=10.16\%$ and $\hat{P}(Y=1|A=1,M\in S_4)=14.84\%$.

2.2.3 | Simulation study with Firth's penalization

It is well known that conventional maximum likelihood estimation of logistic regression is impaired with small-sample or sparse-data biases. Sparse-data bias can appear even in large data studies and is characterized by a lack of adequate number of events (ie, cases when $Y=1$) for some combinations of regressors levels.[29,30] In logistic regression analyzes of small or sparse datasets, maximum likelihood estimates may diverge to infinity (so-called separation phenomenon occurring when the outcome is perfectly predicted by a linear combination of the regressors).[31,32] Separation problems are likely to be associated with a low number of events per variable (EPV).[33] Furthermore, low EPV is one of the data features contributing to sparse-data bias.[30,33] Firth's penalization is generally considered as a effective tool to deal with small-sample or sparse-data biases.[29,34] This penalization modifies the likelihood by multiplying it by the square root of the determinant of the Fisher information matrix, which entails the removal of the first-order term in the asymptotic bias expansion of the maximum likelihood coefficients estimates.[34] Firth's penalization is also a default choice to handle separation issues in logistic regression analyzes as this method yields finite and consistent estimates when separation occurs.[18,31,34]

To explore the impact of Firth's penalization[34,35] on the exact natural effects estimators proposed, we repeated the main simulation study with covariates for *Adjusted scenarios 1* and *2* ($\theta_0\in\{-3,\ -2\}$) with samples sizes of $n=150$, $n=250$, and $n=500$. The outcome was either marginally rare or relatively rare in these scenarios since the specified θ_0 values yielded, as mentioned above, estimated marginal outcome prevalences of 7.64% and 17.94%, respectively. For each scenario, we estimated the EPV as the product of the corresponding estimated marginal outcome prevalence and sample size divided by the number of variables in the outcome model (5 with the exposure-mediator interaction term). The estimated EPVs were 2.29, 3.82, and 7.64 for *Adjusted scenario 1* with $n=150$, $n=250$, and $n=500$, respectively; for *Adjusted scenario 2*, the values were 5.38, 8.97, and 17.94. The results of the exact approach with Firth's penalization in the outcome logistic model were compared to those obtained without penalization.

2.2.4 | Simulation study with omitted exposure-mediator interaction term

We performed an additional simulation study to assess the impact of omitting the exposure-mediator interaction term in the fitted outcome model of proposed exact mediation approach when such an interaction exists. More precisely, we repeated the main simulation study with covariates for *Adjusted scenarios 1*, *3*, and *5* ($\theta_0\in\{-3,\ -0.5,\ 2\}$) with a coefficient value for the exposure-mediator interaction term of $\theta_3=0.15$ (refer to Equation (2)). As in the main simulation study, 1000 independent samples of size $n=5000$ were generated. For each sample generated, we fitted a correctly specified linear regression model for the mediator, but we misspecified the outcome logistic regression model by excluding the exposure-mediator interaction term. Natural effects were then estimated on the OR scale using the exact estimators proposed (Equations 4 and 7) with the $\theta_3 am$ term omitted. We then redid all these steps but when using a larger value for the interaction coefficient in the data-generating mechanism: $\theta_3=0.30$.

2.2.5 | Simulation study with a non-normal mediator error term

We also performed a simulation study to examine how the non-normality of the error term in the mediator model (Equation 1) affects the performance of the exact estimators proposed. Two cases were considered.

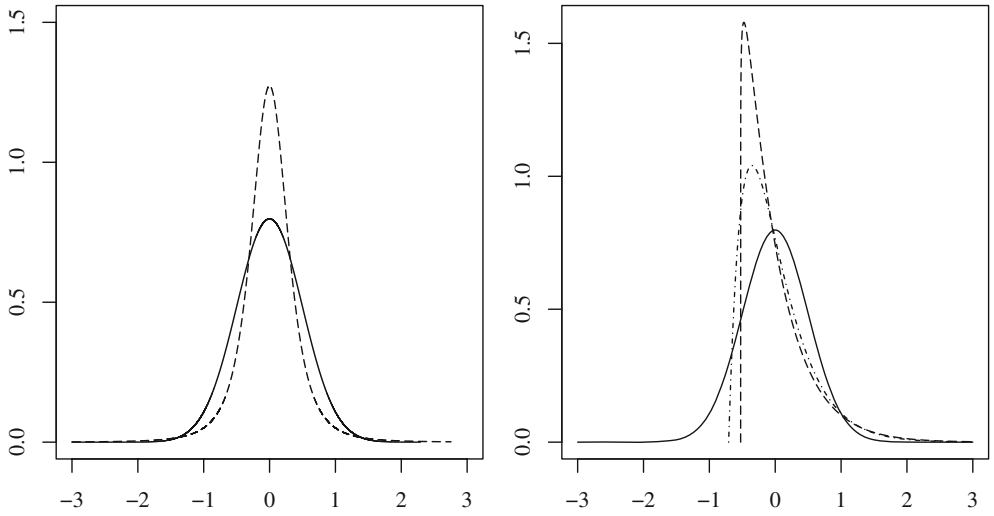

FIGURE 1 Left graph: density function of the generalized t-distribution with $\nu = 3$ degrees of freedom, position parameter $\mu = 0$ and scale parameter $\sigma = \frac{1}{2\sqrt{3}}$ (dashed line). Right graph: density functions of the gamma distributions with *shape* = 1.1025, *scale* = 0.4762 (dashed line; *skewness* = 1.9048) and *shape* = 2, *scale* = 0.3536 (dot-dashed line; *skewness* = 1.4142); both gamma distributions were centered to have an expectation equal to zero). For both graphs, the solid line depicts the density function of the normal distribution with $\mu = 0$ and *variance* = 0.5^2

First, the main simulation study with covariates for *Adjusted scenarios 1*, *3*, and *5* was rerun, but the error terms were generated from a generalized t-distribution with $\nu = 3$ degrees of freedom (smallest integer value of ν for which its variance is finite), position parameter $\mu = 0$ and scale parameter $\sigma = \frac{1}{2\sqrt{3}}$ (*Case 1*; see Figure 1, left graph). The scale parameter value was chosen to ensure that the variance of this distribution ($\sigma^2\nu/(\nu-2) = 0.5^2$) coincided with the variance of the normal distribution specified for the error terms in the main simulation study. The corresponding density function for the generalized t-distribution considered was thus $f(x) = \frac{4}{\pi}\left(1+4x^2\right)^{-2},\ x \in (-\infty, \infty)$.

Second, error terms were generated from a gamma distribution (*Case 2*; see Figure 1, right graph). In order to allow different degrees of skewness while keeping the variance approximately equal to 0.5^2, two sets of shape and scale parameters were considered: *shape* = 1.1025, *scale* = 0.4762 and *shape* = 2, *scale* = 0.3536, yielding skewness values of 1.9048 and 1.4142, respectively. We then centered the error terms by subtracting the expected value of the corresponding gamma distribution (to yield an expectation equal to zero).

For each mediator error term distribution, 1000 independent samples of size $n = 5000$ were generated; natural effects were then obtained on the OR scale using the exact estimators, which assume a normal distribution for this error term. The true values of the natural effects were calculated by replacing the second product term in the integrand of Equation (7) by the corresponding true density function (generalized t-distribution or gamma) and using the true (simulation) parameters values.

3 | RESULTS

3.1 | Results of the main simulation studies

Tables 1-3 summarize the performance of the proposed exact natural effects estimators on the OR, RR, and RD scales in the main adjusted simulation study; for space purposes, the results for the crude simulation study are deferred to Tables D1-D3 (Appendix D). For each scenario and all scales, the mean values of the exact NDE, NIE, and TE estimates were very close to corresponding true values, with relative bias values ranging between −1.15% and 2.06%. All exact interval estimators, using the delta method or the bootstrap, yielded coverage probability values close to 95% (the smallest value was 93.0%).

For both multiplicative scales, the results returned by the exact approach in the crude simulation study were almost identical to those obtained using the NEM approach,[17] while they were very close in the adjusted simulation study (see

TABLE 1 Adjusted simulation study: Odds ratio scale (1000 samples of size $n = 5000$)

Effect	Estimator	True value	Mean	Bias	Relative bias (%)	SD	RMSE	Delta/robust SE CP (%)[a]	Bootstrap CP (%)[b]
Adjusted scenario 1									
NDE OR	Exact	1.550	1.565	0.015	0.98	0.217	0.218	94.6	94.1
	Gaynor et al		1.529	−0.021	−1.35	0.202	0.203	-	94.4
	VV		1.574	0.024	1.55	0.215	0.217	94.8	94.1
	NEM		1.571	0.021	1.37	0.214	0.215	94.6	-
NIE OR	Exact	1.380	1.386	0.006	0.43	0.110	0.110	94.2	94.5
	Gaynor et al		1.350	−0.029	−2.13	0.108	0.112	-	93.2
	VV		1.391	0.011	0.83	0.114	0.115	94.3	94.6
	NEM		1.384	0.004	0.30	0.109	0.109	94.2	-
TE OR	Exact	2.139	2.155	0.016	0.76	0.240	0.241	94.7	94.6
	Gaynor et al		2.052	−0.086	−4.03	0.223	0.240	-	92.9
	VV		2.176	0.037	1.74	0.244	0.247	95.0	94.8
	NEM		2.161	0.022	1.05	0.241	0.242	95.3	-
Adjusted scenario 2									
NDE OR	Exact	1.539	1.545	0.006	0.41	0.150	0.150	94.6	93.9
	Gaynor et al		1.532	−0.007	−0.44	0.146	0.146	-	93.9
	VV		1.565	0.026	1.69	0.147	0.149	94.9	93.7
	NEM		1.550	0.011	0.71	0.148	0.148	94.7	-
NIE OR	Exact	1.376	1.377	0.002	0.11	0.080	0.080	94.4	94.4
	Gaynor et al		1.357	−0.019	−1.38	0.078	0.080	-	92.7
	VV		1.387	0.011	0.77	0.085	0.085	94.5	94.2
	NEM		1.376	−0.000	−0.02	0.078	0.078	94.2	-
TE OR	Exact	2.118	2.121	0.004	0.18	0.165	0.165	95.0	94.0
	Gaynor et al		2.072	−0.045	−2.13	0.159	0.165	-	93.1
	VV		2.163	0.046	2.16	0.172	0.178	93.9	93.6
	NEM		2.126	0.008	0.37	0.165	0.166	94.6	-
Adjusted scenario 3									
NDE OR	Exact	1.515	1.520	0.005	0.34	0.115	0.115	95.4	95.5
	Gaynor et al		1.516	0.002	0.16	0.114	0.114	-	95.4
	VV		1.565	0.050	3.32	0.108	0.119	94.4	94.4
	NEM		1.521	0.007	0.43	0.115	0.115	95.4	-
NIE OR	Exact	1.373	1.375	0.002	0.12	0.068	0.068	94.6	94.5
	Gaynor et al		1.374	0.002	0.11	0.068	0.068	-	94.6
	VV		1.387	0.014	1.00	0.074	0.075	94.4	93.9
	NEM		1.373	−0.000	−0.01	0.066	0.066	94.7	-
TE OR	Exact	2.080	2.084	0.005	0.23	0.125	0.125	96.1	96.5
	Gaynor et al		2.080	0.001	0.04	0.124	0.124	-	96.4
	VV		2.167	0.087	4.19	0.142	0.166	93.8	92.9
	NEM		2.084	0.004	0.20	0.125	0.125	96.0	-

(Continues)

TABLE 1 (Continued)

Effect	Estimator	True value	Mean	Bias	Relative bias (%)	SD	RMSE	Delta/robust SE CP (%)[a]	Bootstrap CP (%)[b]
Adjusted scenario 4									
NDE OR	Exact	1.499	1.511	0.013	0.84	0.168	0.168	94.9	94.3
	Gaynor et al		1.523	0.024	1.62	0.173	0.175	-	94.4
	VV		1.576	0.078	5.19	0.150	0.169	92.9	91.6
	NEM		1.509	0.010	0.66	0.169	0.169	95.0	-
NIE OR	Exact	1.378	1.384	0.006	0.46	0.097	0.097	95.3	95.2
	Gaynor et al		1.408	0.030	2.21	0.102	0.106	-	94.7
	VV		1.392	0.014	1.02	0.102	0.103	95.7	95.0
	NEM		1.382	0.004	0.31	0.095	0.095	95.3	-
TE OR	Exact	2.065	2.082	0.017	0.83	0.182	0.182	95.4	94.4
	Gaynor et al		2.135	0.070	3.39	0.199	0.211	-	91.2
	VV		2.190	0.124	6.02	0.220	0.252	92.6	90.9
	NEM		2.075	0.010	0.50	0.181	0.181	95.5	-
Adjusted scenario 5									
NDE OR	Exact	1.495	1.525	0.030	2.02	0.247	0.249	94.0	93.8
	Gaynor et al		1.527	0.032	2.15	0.250	0.252	-	94.2
	VV		1.599	0.104	6.98	0.217	0.241	93.9	92.3
	NEM		1.521	0.026	1.73	0.250	0.252	94.2	-
NIE OR	Exact	1.381	1.393	0.011	0.82	0.149	0.149	95.6	95.7
	Gaynor et al		1.406	0.025	1.78	0.149	0.151	-	95.9
	VV		1.397	0.016	1.15	0.153	0.154	95.7	95.8
	NEM		1.390	0.009	0.66	0.146	0.146	95.3	-
TE OR	Exact	2.065	2.102	0.037	1.78	0.266	0.268	95.4	95.2
	Gaynor et al		2.127	0.062	2.98	0.288	0.294	-	94.9
	VV		2.227	0.161	7.82	0.345	0.381	94.6	92.5
	NEM		2.092	0.027	1.30	0.264	0.265	94.7	-

Abbreviations: CP, coverage probability; NDE, natural direct effect; NEM, natural effect model (imputation approach); NIE, natural indirect effect; OR, odds ratio; RMSE, root mean squared error; SD, standard deviation; SE, standard error; TE, total effect; VV, VanderWeele and Vansteelandt's approach.
[a]Delta method[23] for exact and VanderWeele and Vansteelandt's estimators; robust standard errors based on the sandwich estimator[27] for NEMs.
[b]Percentile bootstrap[26] based on 500 resamples with replacement.

Tables D1,D2,1, and 2). For the RD scale, the exact results were close to those obtained using Imai et al's[2] quasi-Bayesian approach.

Compared to the exact OR estimators, the approximate OR estimators by Gaynor et al[5] resulted in smaller relative bias values in *Crude* and *Adjusted scenarios 3*. However, we generally observed greater relative bias values for the other scenarios, especially for the NIE and TE estimators (see Tables D1 and 1).

It is noteworthy to emphasize the key role of the parameter s in the Gaynor et al approach.[5] To get a better understanding of its impact on the quality of the approximation of the logistic function by the normal cumulative distribution function, we present in Table E1 the distribution of the s estimates in each crude and adjusted simulation scenario (see Appendix E). We observed that the relative bias values were related to the distribution of the s estimates with respect to the recommended s values reported in the literature. For example, the smallest relative bias values were observed for both *Crude* and *Adjusted scenarios 3*, for which the corresponding s estimates were close to the s value recommended by Amemiya.[36] An increase in relative biases for *Crude* and *Adjusted scenarios 4*,

TABLE 2 Adjusted simulation study: Risk ratio scale (1000 samples of size $n = 5000$)

Effect	Estimator	True value	Mean	Bias	Relative bias (%)	SD	RMSE	Delta/robust SE CP (%)[a]	Bootstrap CP (%)[b]
Adjusted scenario 1									
NDE RR	Exact	1.503	1.515	0.012	0.81	0.194	0.194	94.4	94.1
	NEM		1.517	0.014	0.96	0.189	0.190	94.7	-
NIE RR	Exact	1.336	1.341	0.005	0.39	0.097	0.098	94.2	94.4
	NEM		1.337	0.001	0.11	0.095	0.095	94.1	-
TE RR	Exact	2.007	2.020	0.013	0.64	0.206	0.206	95.1	94.8
	NEM		2.018	0.011	0.55	0.204	0.204	95.2	-
Adjusted scenario 2									
NDE RR	Exact	1.430	1.433	0.003	0.22	0.113	0.113	94.5	93.9
	NEM		1.432	0.002	0.13	0.110	0.110	94.6	-
NIE RR	Exact	1.279	1.280	0.001	0.08	0.059	0.059	94.1	94.0
	NEM		1.276	−0.003	−0.22	0.057	0.058	94.0	-
TE RR	Exact	1.828	1.830	0.001	0.07	0.113	0.113	94.7	94.5
	NEM		1.823	−0.005	−0.30	0.111	0.112	94.7	-
Adjusted scenario 3									
NDE RR	Exact	1.245	1.246	0.001	0.04	0.047	0.047	95.7	95.5
	NEM		1.243	−0.002	−0.18	0.046	0.046	95.3	-
NIE RR	Exact	1.149	1.149	0.001	0.04	0.027	0.027	95.1	94.8
	NEM		1.148	−0.001	−0.05	0.027	0.027	95.0	-
TE RR	Exact	1.430	1.431	0.000	0.02	0.040	0.040	96.4	96.3
	NEM		1.426	−0.004	−0.29	0.039	0.040	96.2	-
Adjusted scenario 4									
NDE RR	Exact	1.086	1.086	0.000	0.00	0.022	0.022	95.1	94.2
	NEM		1.086	−0.000	−0.04	0.022	0.022	94.8	-
NIE RR	Exact	1.050	1.051	0.001	0.06	0.013	0.013	94.8	95.1
	NEM		1.051	0.001	0.14	0.013	0.013	94.5	-
TE RR	Exact	1.141	1.141	0.001	0.05	0.016	0.016	95.0	94.2
	NEM		1.141	0.001	0.05	0.016	0.016	95.5	-
Adjusted scenario 5									
NDE RR	Exact	1.036	1.036	−0.000	−0.01	0.013	0.013	94.0	94.1
	NEM		1.036	0.000	0.00	0.013	0.013	94.0	-
NIE RR	Exact	1.020	1.021	0.000	0.04	0.008	0.008	95.0	95.2
	NEM		1.021	0.001	0.10	0.008	0.008	94.7	-
TE RR	Exact	1.057	1.057	0.000	0.03	0.009	0.009	95.4	95.2
	NEM		1.058	0.001	0.09	0.009	0.009	95.0	-

Abbreviations: CP, coverage probability; NDE, natural direct effect; NEM, natural effect model (imputation approach); NIE, natural indirect effect; RMSE, root mean squared error; RR, risk ratio; SD, standard deviation; SE, standard error; TE, total effect.

[a]Delta method[23] for exact estimators; robust standard errors based on the sandwich estimator[27] for NEMs.

[b]Percentile bootstrap[26] based on 500 resamples with replacement.

TABLE 3 Adjusted simulation study: Risk difference scale (1000 samples of size $n = 5000$)

Effect	Estimator	True value	Mean	Bias	Relative bias (%)	SD	RMSE	Delta/robust SE CP (%)	Bootstrap CP (%)
Adjusted scenario 1									
NDE RD	Exact	0.0289	0.0291	0.0002	0.75	0.0100	0.0100	94.9	94.1
	Imai et al		0.0295	0.0006	2.04	0.0100	0.0100	94.8	-
NIE RD	Exact	0.0290	0.0289	−0.0001	−0.37	0.0066	0.0066	94.3	94.0
	Imai et al		0.0287	−0.0003	−0.90	0.0067	0.0067	94.0	-
TE RD	Exact	0.0579	0.0580	0.0001	0.19	0.0092	0.0092	93.0	93.3
	Imai et al		0.0582	0.0003	0.56	0.0092	0.0092	93.3	-
Adjusted scenario 2									
NDE RD	Exact	0.0608	0.0610	0.0002	0.25	0.0146	0.0146	94.2	93.9
	Imai et al		0.0613	0.0005	0.75	0.0146	0.0146	94.4	-
NIE RD	Exact	0.0564	0.0561	−0.0003	−0.52	0.0093	0.0093	94.6	94.1
	Imai et al		0.0562	−0.0005	−0.91	0.0093	0.0093	94.5	-
TE RD	Exact	0.1172	0.1171	−0.0001	−0.09	0.0128	0.0128	94.3	93.9
	Imai et al		0.1172	−0.0000	−0.01	0.0128	0.0128	93.7	-
Adjusted scenario 3									
NDE RD	Exact	0.1031	0.1032	0.0001	0.19	0.0188	0.0188	95.4	95.5
	Imai et al		0.1032	0.0002	0.10	0.0187	0.0187	95.9	-
NIE RD	Exact	0.0778	0.0778	−0.0001	−0.08	0.0013	0.0013	94.5	94.4
	Imai et al		0.0776	−0.0002	−0.29	0.0012	0.0012	94.3	-
TE RD	Exact	0.1809	0.1810	0.0000	0.02	0.0144	0.0144	96.1	96.3
	Imai et al		0.1808	−0.0002	0.29	0.0144	0.0144	96.1	-
Adjusted scenario 4									
NDE RD	Exact	0.0657	0.0657	−0.0001	−0.11	0.0164	0.0164	95.1	93.9
	Imai et al		0.0652	−0.0005	−0.83	0.0163	0.0163	94.7	-
NIE RD	Exact	0.0412	0.0416	0.0004	0.95	0.0102	0.0102	94.8	95.1
	Imai et al		0.0418	0.0006	1.39	0.0101	0.0101	94.3	-
TE RD	Exact	0.1070	0.1073	0.0003	0.30	0.0115	0.0115	95.0	94.3
	Imai et al		0.1070	0.0000	0.03	0.0115	0.0115	94.8	-
Adjusted scenario 5									
NDE RD	Exact	0.0320	0.0320	−0.0001	−0.25	0.0113	0.0113	95.1	93.9
	Imai et al		0.0313	−0.0007	2.42	0.0112	0.0112	94.4	-
NIE RD	Exact	0.0189	0.0192	0.0003	1.71	0.0072	0.0072	94.8	95.1
	Imai et al		0.0196	0.0007	3.74	0.0072	0.0072	95.1	-
TE RD	Exact	0.0509	0.0512	0.0002	0.48	0.0076	0.0076	95.2	94.3
	Imai et al		0.0508	−0.0001	0.14	0.0076	0.0076	95.4	-

Abbreviations: CP, coverage probability; NDE, natural direct effect; NIE, natural indirect effect; RD, risk difference; RMSE, root mean squared error; SD, standard deviation; SE, standard error; TE, total effect.

[a]Delta method[23] for exact estimators; White's heteroskedasticity-consistent estimator for the covariance matrix[28] for approach by Imai et al.

[b]Percentile bootstrap[26] based on 500 resamples with replacement.

corresponding to estimated marginal probabilities of ≈80%, could be explained by the fact that the observed s estimates were closer to the minimax solution,[10] which performs better at the middle abscissas (ie, argmax) of the logistic density.[12] Therefore, the procedure to estimate s proposed by Gaynor et al[5] can produce suboptimal values for this parameter.

The approximate OR estimators by VanderWeele and Vansteelandt[6] demonstrated small relative bias values (between 0.66% and 1.74%) for *Crude* and *Adjusted scenarios 1*, corresponding to the marginally rare outcome (see Tables D1 and 1). However, we observed an increase in relative bias values with increased outcome marginal prevalence (up to 7.82% in *Adjusted scenario 5*). We found that the approximate OR estimators were systematically more biased than the corresponding exact OR estimators in each scenario.

3.2 | Results of the simulation study with a marginal but not conditional rare outcome

The results of the simulation study conducted to examine the impact of a conditional but not marginal violation of the ROA are presented in Table F1 (see Appendix F). For all scales, the mean values of the exact NDE, NIE, and TE estimates were very close to the corresponding true values, with relative bias values varying between −0.65% and 1.54%; all exact interval estimators showed coverage probabilities close to 95%. The results returned by the exact approach were similar to those obtained by the NEM approach[17] for the multiplicative scales and were almost identical to those returned by Imai et al's[2] quasi-Bayesian approach for the RD scale.

Compared to the approximate OR estimators by Gaynor et al,[5] the exact OR estimators demonstrated smaller absolute relative bias values (0.29% vs 0.63%, 1.54% vs 2.68%, and 0.79% vs 4.36% for NDE, NIE, and TE, respectively).

Finally, we observed that the approximate OR estimators by VanderWeele and Vansteelandt[6] were impaired by the conditional ROA violation: the relative bias values ranged between 68.18% and 426.72%, and a significant decrease in coverage probabilities was observed for the NIE and TE interval estimators.

3.3 | Results of the simulation study with Firth's penalization

The results of the simulation study conducted to examine the impact of Firth's penalization on the exact estimators are presented in Tables G1-G3 (see Appendix G).

For *Adjusted scenario 1* with a sample size of $n = 150$ and EPV=2.29, the relative bias values were between 9.60% and 30.32% for the multiplicative scales (Tables G1 and G2) and between −22.22% and 32.98% for the RD scale when applying the Firth's penalization. For that sample size, no results are reported for the conventional (unpenalized) exact estimators since we observed 42 cases of quasi-separation, resulting in meaningless performance metrics (eg, mean NDE OR = 72.98, corresponding relative bias = 4608.6%). Quasi-separation was not observed in the other analyzes.

For the multiplicative scales, we found that the exact estimators with Firth's penalization were either less biased or equivalent to their conventional counterparts. For instance, a bias reduction due to Firth's penalization was uniformly observed for the exact NIE estimators. The differences between the penalized and unpenalized estimators were minor for *Adjusted scenarios 1* and *2* with $n = 500$ (EPV = 7.64 and EPV = 17.94, respectively).

Adding Firth's penalization to the exact approach did not generally result in bias reduction in the natural effects estimates on the RD scale, as compared to the conventional approach.

3.4 | Results of the simulation study with omitted exposure-mediator interaction term

In Table H1 of Appendix H, we present the results of the simulation study that examined the impact of incorrectly omitting the exposure-mediator interaction term in the fitted outcome logistic regression model.

When the data-generating mechanism considered $\theta_3 = 0.15$ in Equation (2), the relative biases (in absolute values) increased along with the outcome marginal prevalence (from 5.18% to 7.43% for NDE OR and from 3.67% to 5.53% for NIE OR) and were uniformly larger than those obtained for the exact estimators with a correctly specified outcome model (see Table 1 to compare). For *Adjusted scenarios 3* and *5*, the omission of the exposure-mediator interaction term yielded a

remarkable decrease in the coverage probability of the exact NIE OR interval estimators (delta method: 59.7% and 78.5%; bootstrap: 60.9% and 78.7% respectively).

The misspecified outcome model resulted in biases which increased when increasing θ_3 from 0.15 to 0.30 (relative biases in absolute values ranged from 9.39% to 13.69% for NDE OR and from 7.12% to 11.10% for NIE OR). The undercoverage of the NDE OR and NIE OR interval estimators also increased, most notably for the NIE.

For both values of θ_3 considered, the TE OR relative bias values from the misspecified outcome models were comparable to those obtained for the exact estimators using correctly specified outcome models (presented in Table 1).

3.5 | Results of the simulation study with a non-normal mediator error term

We present in Tables I1 and I2 (see Appendix I) the results of the simulation study performed to examine the impact of assuming that the mediator is normally distributed when it is not.

When the mediator error terms were generated from a generalized *t*-distribution (*Case 1*), the NDE OR, NIE OR and TE OR returned by the proposed exact estimators were close to the true values for all scenarios considered, with relative biases ranging from 0.08% to 2.27%. Error terms derived from the gamma distributions (*Case 2*) also yielded small or negligible relative biases for these effects estimators (between 0.19% and 2.02%) for both sets of shape and scale parameters considered.

All exact interval estimators, via the delta method or by bootstrap, resulted in coverage probability values close to 95% (the smallest value was 92.9%), for all error term distributions and all scenarios considered.

4 | REAL DATA EXAMPLE

We applied the proposed exact approach to the cohort data studied in our previous works.[13,14] More specifically, we considered the data of 6197 singleton pregnancies in asthmatic women who gave birth between 1998 and 2008 in the province of Québec, Canada. Gestational age (in weeks) and low birth weight were treated as the continuous mediator and binary outcome, respectively, and placental abruption was considered as the binary exposure. The outcome was rare marginally (ie, the observed prevalence of low birth weight was 7.7%). Both unadjusted (crude) and adjusted analyzes were performed; in the latter, we examined the same set of adjustment variables as in Samoilenko and Lefebvre:[14] maternal age at the beginning of pregnancy (<18 years, 18-34 years, or >34 years), baby's sex, diabetes mellitus and gestational diabetes. We used our SAS macro `bin_cont_exactmed` (see Supporting Information section and Appendix J) to estimate exact NDE and NIE on the OR scale. The results were compared to those obtained by the approximate approaches by VanderWeele and Vansteelandt[6] and Gaynor et al.[5] Natural effects were also estimated using NEMs.[17,24] For all approaches, exposure-mediator interaction was allowed in the analysis. In the adjusted analysis, natural effects were estimated at the sample-specific mean values of covariates. We used percentile bootstrap[26] based on 5000 resamples with replacement to calculate 95% CIs.

The results are presented in Table 4. In both crude and adjusted analyzes, the results obtained by our exact approach were close to those returned by the NEM approach. However, in the adjusted analysis, the differences in the NDE OR and NIE OR estimates between the exact and NEM approaches were larger than those between the exact approach and the approximate approach by Gaynor et al.[5] Interestingly, the results obtained using Gaynor et al[5] were relatively close to those obtained by the exact approach in the adjusted analysis, but not in the crude analysis. Upon closer examination, we found that Gaynor et al's approach[5] used a *s* parameter value of $\hat{s} = 0.532$ in the adjusted analysis, but of $\hat{s} = 1.242$ in the crude analysis, which suggests that results in the crude analysis may have been driven by a suboptimal choice for *s*. Indeed based on the extant literature, the best scaling constant *s* according to different criteria ranges between 0.551 and 0.625.[12] One explanation we put forward is the relatively small number of ratios of coefficients used for determining the *s* value in the crude analysis as compared to the adjusted one (4 vs 9). Lastly, results from the approximate approach by VanderWeele and Vansteelandt[6] generally stood apart from the results returned by the other approaches in both crude and adjusted analyzes.

As a final comment, it should be noted that the point estimates returned by the exact and NEM approaches for the TE in the adjusted analysis were close to those obtained when we considered gestational age as a binary variable (preterm

TABLE 4 Real data example with placental abruption as the exposure, gestational age as the mediator and low birth weight as the outcome ($n = 6197$)

Effect	Exact	95% CI[a]	VV	95% CI[a]	Gaynor et al	95% CI[a]	NEM	95% CI[a]
Crude analyzes								
NDE OR	1.478	1.027, 2.028	1.851	0.891, 8.910	1.719	0.722, 2.258	1.395	0.988, 1.897
NIE OR	3.489	2.555, 4.881	7.393	3.961, 20.228	4.844	0.340, 5.972	3.678	2.794, 4.972
TE OR	5.157	3.344, 7.995	13.685	4.334, 157.114	8.326	0.230, 10.851	5.132	3.663, 6.906
Adjusted analyzes								
NDE OR	1.485	1.020, 2.082	1.800	0.893, 8.394	1.478	1.014, 2.028	1.407	0.982, 1.925
NIE OR	3.463	2.548, 4.896	7.358	4.054, 19.832	3.354	2.425, 4.717	3.644	2.757, 4.986
TE OR	5.144	3.338, 8.029	13.245	4.565, 143.038	4.957	3.197, 7.562	5.127	3.659, 6.906

Abbreviations: CI, confidence interval; NDE, natural direct effect; NEM, natural effect model (imputation approach); NIE, natural indirect effect; OR, odds ratio; TE, total effect; VV, VanderWeele and Vansteelandt's approach.
[a]Percentile bootstrap[26] based on 5000 resamples with replacement.

birth) instead of a continuous variable (see our previous analysis in Samoilenko et al[14]). This is a reassuring finding since, conceptually, the total effect should not be affected by the mediator and its type.

5 | DISCUSSION

In this work, we expanded the exact mediation approach for a binary outcome and a continuous mediator that was proposed by Cheng et al[16] for estimating natural direct and indirect effects on the (log) OR scale. As in Cheng et al,[16] our approach is based on logistic and linear regression models for the binary outcome and the continuous mediator, respectively. A first contribution was to introduce exact point estimators to express natural effects on three standard effect measures for binary outcomes, namely the OR, RR, and RD scales. Exact formulas for the standard errors were also derived for each scale considered using the multivariate first-order delta method. The exact point and interval estimators result in improper integrals for which no closed-form expressions exist. As these integrals must be approximated, this is accomplished using numerical quadrature. Another contribution was to allow the exact approach to feature an exposure-mediator interaction term in the outcome model (Equation 2). This addition is worthwhile since outcome model misspecification by omitting the exposure-mediator interaction term when such an interaction exists can affect the performance of the natural effects estimators. In our simulation study, a decrease in performance was observed to go along the magnitude of the interaction coefficient value in the data-generating mechanism. As a practical contribution, our proposed exact mediation approach is available for general uses in the SAS macro `bin_cont_exactmed` (see Supporting Information section).

Our main simulation studies showed an adequate performance of the exact estimators proposed independently of the outcome marginal rareness or commonness. More precisely, we obtained very small values of relative bias, suggesting that our exact estimators are unbiased for large enough sample sizes. Furthermore, both the delta method and the percentile bootstrap resulted in coverage probabilities close to the nominal level. The exact estimators also performed well in the additional simulation study in which the outcome was rare marginally while the ROA was markedly violated in some strata defined by the exposure and the mediator. Conversely, the approximate OR estimators by VanderWeele and Vansteelandt[6] resulted in large bias and variance for both the NDE and NIE; a low coverage probability was observed for the NIE. These results reinforce the argument that assessing the outcome rareness in terms of the marginal outcome probability can be misleading in the context of causal mediation analysis.

When it was possible, we compared our exact approach to the parametric inference estimation algorithm by Imai et al[2] and the NEM approach by Lange et al.[17] These benchmark approaches also do not rely on the outcome rareness or commonness. The R package `mediation`,[25] which implements the approach by Imai et al examined (and its non-parametric bootstrap version),[2] returns natural effects estimates for binary outcomes on the RD scale only, while our SAS macro `bin_cont_exactmed` provides estimates on the OR, RR, and RD scales with associated delta method and bootstrap

standard errors. To obtain estimates on all these binary scales, our SAS macro `bin_cont_exactmed` uses the same fitted outcome logistic model. This contrasts with the R package `medflex`,[24] implementing the NEM approach by Lange et al,[17] which requires different NEM specifications in order to obtain natural effects estimates on these scales (eg, logistic model for OR, log-binomial or Poisson model for RR).

As Tchetgen Tchetgen[37] pointed out, the modeling assumption encoded in Equation (1) is often violated in epidemiology, for example, in practical applications characterized by a skewed mediator distribution. In our simulation study assessing the impact of deviating from the normality of mediator errors, we observed that our exact estimators were robust to non-normality under the scenarios and mediator distributions considered (generalized *t*-distribution and gamma distributions). In the presence of more extreme deviation from normality, a transformation (eg, log) of the continuous mediator could be considered as a potential solution. However, this solution presupposes that the linearity of the effect of the transformed mediator on the outcome on the logit scale (see Equation (2) for reference) is reasonably satisfied.

Our exact approach is conceptually straightforward and a logical choice when both the mediator and outcome models (1), (2) are correctly specified. When it is not reasonable to assume the linearity of the effect of M on Y on the logit scale, the NEM weighting-based approach implemented in `medflex`[24] may be more adequate since it does not require modeling the outcome given the mediator. Alternatively, the NEM imputation-based approach could be implemented with a sufficiently rich and flexible imputation model based on generalized additive models or machine learning techniques (by applying the `SuperLearner` function). In general, these more sophisticated imputation strategies can also be used whenever there is a concern regarding potential incoherence (uncongeniality) of the imputation procedure and NEM analysis procedure,[24,38] which is prone to occur with nonlinear outcome models.[24] While the approach by Imai et al[2] used the same logistic and linear models as the exact approach in our simulations, the R package `mediation`[25]—which implements this approach—allows going beyond standard regression models for the mediator and outcome models (eg, generalized additive models) if desired.

It is well known that classical maximum likelihood estimation can result in unreliable point and interval estimates in logistic regression analyzes of small and/or sparse data because of complete separation or quasi-complete separation.[18,39] Taking in consideration this problem, our SAS macro `bin_cont_exactmed` allows for Firth's penalization in the outcome model. However, despite the fact that Firth's penalization is generally considered as an effective solution to address separation problems in logistic regression models,[18,19,39] some authors have mentioned that this penalization can introduce bias in both average and individual predicted probabilities[19] (of note, the expit function in the integrand of Equation (3) can be considered as a predicted probability). The latter observation and the significant bias values obtained for the multiplicative and/or RD scales in some of our simulation scenarios suggest that further studies are needed to examine the impact of Firth's penalization in exact mediation settings.

Selection of the adjustment covariates is a crucial step to address causal questions from observational data. In a causal mediation context, Diop et al[40] recommended to adjust for pure predictors of the outcome, in addition to true confounders, to reduce the standard errors of the natural effects estimators. Moreover they suggested to avoid adjusting for pure predictors of the exposure since adjustment for such covariates tends to increase the standard errors of these estimators. Furthermore, considering that adjustment for pure predictors of the mediator was found to increase the standard error of the NDE estimators and could either increase or decrease the variance of the NIE estimators, Diop et al[40] advised to avoid adjusting for such predictors. We recommend for applying Diop et al[40] strategy when selecting the covariates to be adjusted for in the proposed exact approach. However, for large adjustment sets or when covariates types are not fully known, LASSO-based methods (eg, outcome-adaptive LASSO)[41] could be used to select covariates to control for confounding.

The SAS macro `bin_cont_exactmed` was first developed for the estimation of natural effects. However, a controlled direct effect is often considered as a more relevant concept regarding the evaluation of public health policies.[7,42] Therefore, this macro also returns controlled direct effects estimated on the OR, RR, and RD scales at a user-defined mediator level and calculated according to the formulas presented in Samoilenko and Lefebvre.[14] As the VanderWeele and Vansteelandt,[6] Gaynor et al,[5] and Cheng et al[16] approaches, our exact approach targets conditional natural effects. By default, our SAS macro `bin_cont_exactmed` evaluates the natural indirect, natural direct and controlled direct effects at the sample-specific mean values of the adjustment covariates, but it also allows to estimate these effects at user-defined covariates levels (that is, stratum-specific effects). If the user-defined values are not specified for some adjustment covariates, our macro `bin_cont_exactmed` sets them at the default sample-specific mean values.

If marginal (population) natural effects are desired, the exact estimators could be modified by averaging the estimated conditional effects over the distribution of the adjustment covariates in the sample, and corresponding standard errors of estimates be obtained using the bootstrap. Marginal natural effects estimation is not currently available in `bin_cont_exactmed`.

Regarding the execution time, our SAS macro returns results almost immediately when using the delta method to obtain interval estimates. Considerably more time is required to obtain exact interval estimates using the bootstrap. For example, in our real data adjusted analysis ($n = 6197$, 5000 bootstrap resamples), 8 min were required to obtain results on all three binary scales considered with a machine having a CPU speed of 3.40 GHz and a RAM of 12 GB. Corresponding execution times were approximately 8 and 4.5 minutes when using the R packages `medflex` (OR scale only) and `mediation` (RD scale only), respectively.

In conclusion, our exact mediation approach does not rely on any assumptions on the outcome rareness or commonness and, consequently, does not require to assess the adequacy of these assumptions. It thus eases implementation for practitioners aiming to perform causal mediation analysis based on the standard outcome logistic and mediator linear models. Moreover, our SAS macro `bin_cont_exactmed` returns the natural and controlled direct effects on all standard binary scales (ie, OR, RR, and RD), thereby facilitating a direct comparison with results returned by other mediation approaches for binary outcomes. Lastly, as our approach was developed for cohort studies in this article, it will be worthwhile to extend it to case-control designs in order to increase its practical applicability.

AUTHOR CONTRIBUTIONS

Mariia Samoilenko and Geneviève Lefebvre devised the project, developed the main conceptual ideas, and planned the simulations and real data analyzes. Mariia Samoilenko performed the simulations and analyzes. Mariia Samoilenko and Geneviève Lefebvre wrote the article. Mariia Samoilenko wrote the SAS macro. Both authors have reviewed the article and approved its submission.

ACKNOWLEDGEMENTS

This work was funded by grants from the Fonds de recherche du Québec–Santé (FRQ-S; # 268860) and the Natural Sciences and Engineering Research Council of Canada (# RGPIN-2020-05473). G.L. is an FRQ-S Research Scholar. The authors thank Miguel Caubet Fernandez for a prior review of the article and Dr. Lucie Blais for the real-example data.

CONFLICT OF INTEREST

The authors declare no potential conflict of interests.

DATA AVAILABILITY STATEMENT

The data that support the findings in the REAL DATA EXAMPLE section are not publicly available because of privacy and ethical restrictions.

ORCID

Geneviève Lefebvre https://orcid.org/0000-0002-7155-8218

REFERENCES

1. Pearl J. Direct and indirect effects. Proceedings of the 17th Conference on Uncertainty in Artificial Intelligence; August 2-5, 2001:411-420; Morgan Kaufmann Publishers Inc.
2. Imai K, Keele L, Tingley D. A general approach to causal mediation analysis. *Psychol Methods*. 2010;15(4):309-334.
3. Pearl J. The causal mediation formula: a guide to the assessment of pathways and mechanisms. *Prev Sci*. 2012;13(4):426-436.
4. Loeys T, Moerkerke B, De Smet O, Buysse A, Steen J, Vansteelandt S. Flexible mediation analysis in the presence of nonlinear relations: beyond the mediation formula. *Multivar Behav Res*. 2013;48(6):871-894.
5. Gaynor SM, Schwartz J, Lin X. Mediation analysis for common binary outcomes. *Stat Med*. 2019;38(4):512-529.
6. VanderWeele TJ, Vansteelandt S. Odds ratios for mediation analysis for a dichotomous outcome. *Am J Epidemiol*. 2010;172(12):1339-1348.
7. Valeri L, VanderWeele TJ. Mediation analysis allowing for exposure-mediator interactions and causal interpretation: theoretical assumptions and implementation with SAS and SPSS macros. *Psychol Methods*. 2013;18(2):137-150.

8. Yung YF, Lamm M, Wei Z. Causal mediation analysis with the CAUSALMED procedure. Proceedings of the SAS Global Forum 2018 Conference; 2018; SAS Institute Inc.
9. Shi B, Choirat C, Coull BA, VanderWeele TJ, Valeri L. CMAverse: a suite of functions for reproducible causal mediation analyses. *Epidemiology*. 2021;32(5):e20-e22.
10. Camilli G. Origin of the scaling constant d = 1.7 in item response theory. *J Educ Stat*. 1994;19(3):293-295.
11. Cox DR, Snell EJ. *Analysis of Binary Data*. 2nd ed. London: Chapman & Hall; 1989.
12. Savalei V. Logistic approximation to the normal: the KL rationale. *Psychometrika*. 2006;71:763-767.
13. Samoilenko M, Blais L, Lefebvre G. Comparing logistic and log-binomial models for causal mediation analyses of binary mediators and rare binary outcomes: evidence to support cross-checking of mediation results in practice. *Obs Stud*. 2018;4:1846-1858.
14. Samoilenko M, Lefebvre G. Parametric-regression–based causal mediation analysis of binary outcomes and binary mediators: moving beyond the rareness or commonness of the outcome. *Am J Epidemiol*. 2021;190(9):193-216.
15. Doretti M, Raggi M, Stanghellini E. Exact parametric causal mediation analysis for a binary outcome with a binary mediator. *Stat Methods Appl*. 2021;31:87-108.
16. Cheng C, Spiegelman D, Li F. Estimating the natural indirect effect and the mediation proportion via the product method. *BMC Med Res Methodol*. 2021;21(1):253.
17. Lange T, Vansteelandt S, Bekaert M. A simple unified approach for estimating natural direct and indirect effects. *Am J Epidemiol*. 2012;176(3):190-195.
18. Allison PD. *Logistic Regression Using SAS: Theory and Application*. 2nd ed. Cary, NC: SAS Institute Inc.; 2012.
19. Mansournia MA, Geroldinger A, Greenland S, Heinze G. Separation in logistic regression: causes, consequences, and control. *Am J Epidemiol*. 2017;187(4):864-870.
20. Greenland S, Mansournia MA. Penalization, bias reduction, and default priors in logistic and related categorical and survival regressions. *Stat Med*. 2015;34(23):3133-3143.
21. SAS Institute Inc. *SAS/IML 9.22 User's Guide*. Cary, NC: SAS Institute Inc; 2010.
22. Wicklin R. How to numerically integrate a function in SAS the Do Loop SAS blog; 2011. https://blogs.sas.com/content/iml/2011/05/06/how-to-numerically-integrate-a-function-in-sas.html. Accessed September 7, 2021.
23. Casella G, Berger RL. *Statistical Inference*. 2nd ed. Pacific Grove, CA: Duxbury Thomson Learning; 2002.
24. Steen J, Loeys T, Moerkerke B, Vansteelandt S. medflex: an R package for flexible mediation analysis using natural effect models. *J Stat Softw*. 2017;76:1-46.
25. Tingley D, Yamamoto TI, Hirose K, Keele L, Imai K. mediation: R package for causal mediation analysis. *J Stat Softw*. 2014;59(5):1-38.
26. Chernick MR. *Bootstrap Methods: A Guide for Practitioners and Researchers*. 2nd ed. Hoboken, NJ: John Wiley & Sons; 2011.
27. Liang K-Y, Zeger SL. Longitudinal data analysis using generalized linear models. *Biometrika*. 1986;73(1):13-22.
28. White H. A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica*. 1980;48(4):817-838.
29. Cole SR, Chu H, Greenland S. Maximum likelihood, profile likelihood, and penalized likelihood: a primer. *Am J Epidemiol*. 2014;179(2):252-260.
30. Greenland S, Mansournia MA, Altman DG. Sparse data bias: a problem hiding in plain sight. *BMJ*. 2016;352:i1981.
31. Šinkovec H, Geroldinger A, Heinze G. Bring more data! — A good advice? Removing separation in logistic regression by increasing sample size. *Int J Environ Res Public Health*. 2019;16(23):4658.
32. Gelman A, Jakulin A, Pittau MG, Su Y-S. A weakly informative default prior distribution for logistic and other regression models. *Ann Appl Stat*. 2008;2(4):1360-1383.
33. Van Smeden M, de Groot JAH, Moons KGM, et al. No rationale for 1 variable per 10 events criterion for binary logistic regression analysis. *BMC Med Res Methodol*. 2016;16(1):1-12.
34. Heinze G, Schemper M. A solution to the problem of separation in logistic regression. *Stat Med*. 2002;21(16):2409-2419.
35. Firth D. Bias reduction of maximum likelihood estimates. *Biometrika*. 1993;80(1):27-38.
36. Amemiya T. Qualitive response models: a survey. *J Econ Lit*. 1981;19:1483-1536.
37. Tchetgen ET. A note on formulae for causal mediation analysis in an odds ratio context. *Epidemiol Methods*. 2014;2(1):21-31.
38. Bartlett JW, Hughes RA. Bootstrap inference for multiple imputation under uncongeniality and misspecification. *Stat Methods Med Res*. 2020;29(12):3533-3546.
39. Heinze G. A comparative investigation of methods for logistic regression with separated or nearly separated data. *Stat Med*. 2006;25(24):4216-4226.
40. Diop A, Lefebvre G, Duchaine CS, Laurin D, Talbot D. The impact of adjusting for pure predictors of exposure, mediator, and outcome on the variance of natural direct and indirect effect estimators. *Stat Med*. 2021;40(10):2339-2354.
41. Ye Z, Zhu Y, Coffman DL. Variable selection for causal mediation analysis using LASSO-based methods. *Stat Methods Med Res*. 2021;30(6):1413-1427.
42. Naimi AI, Kaufman JS, MacLehose RF. Mediation misgivings: ambiguous clinical and public health interpretations of natural direct and indirect effects. *Int J Epidemiol*. 2014;43(5):1656-1661.
43. VanderWeele TJ, Vansteelandt S. Conceptual issues concerning mediation, interventions and composition. *Stat Interface*. 2009;2:457-468.

44. Lange T, Hansen KW, Sørensen R, Galatius S. Applied mediation analyses: a review and tutorial. *Epidemiol Health*. 2017;39: e2017035.
45. Andrews RM, Didelez V. Insights into the cross-world independence assumption of causal mediation analysis. *Epidemiology*. 2021;32(2):209-219.
46. Gao X, Luo L. An improvement in estimation of the standard error for the natural direct effect in causal mediation analysis. *Epidemiology*. 2019;30(4):e25-e26.
47. Khuri AD. *Advanced Calculus with Applications in Statistics*. 2nd ed. New York, NY: John Wiley & Sons; 2002.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Samoilenko M, Lefebvre G. An exact regression-based approach for the estimation of natural direct and indirect effects with a binary outcome and a continuous mediator. *Statistics in Medicine*. 2023;42(3):353-387. doi: 10.1002/sim.9621

APPENDIX A. IDENTIFICATION ASSUMPTIONS

The identification of the natural effects from observed data requires that the following assumptions hold for all possible values of a, a^*, m, and $\boldsymbol{c}$: (1) if $A = a$ then the observed value of the mediator M is almost surely equal to $M(a)$; (2) if $A = a$ and $M = m$ then the observed value of the outcome Y is almost surely equal to $Y(a, m)$; (3) $P(A = a|\boldsymbol{C} = \boldsymbol{c}) > 0$; (4) $P(M = m|A = a, \boldsymbol{C} = \boldsymbol{c}) > 0$; (5) $Y(a, m) \perp\!\!\!\perp A|\boldsymbol{C}$; (6) $M(a) \perp\!\!\!\perp A|\boldsymbol{C}$; (7) $Y(a, m) \perp\!\!\!\perp M|A, \boldsymbol{C}$; (8) $Y(a, m) \perp\!\!\!\perp M(a^*)|\boldsymbol{C}$. Assumptions 1 and 2 are so-called *consistency assumptions*.[43,44] Assumptions 3 and 4, also known as *positivity assumptions*, mean that all exposure values have a non-zero probability for every possible values of confounders, and that all mediator values have a non-zero probability for every possible values of confounders and exposure (for A or M continuous, the corresponding assumption is expressed in terms of a density function). Assumptions 5-7, or *no unmeasured confounding assumptions*, formally express the postulates that there are no unmeasured confounders for the exposure-outcome, exposure-mediator and mediator-outcome relationships. Assumption 8, or *cross-world independence assumption*, is impossible to verify using observed data alone; this assumption generally holds, but not always, if there are no measured or unmeasured confounders for the mediator-outcome relationship affected by the exposure.[44,45]

APPENDIX B. DELTA METHOD FOR EXACT MEDIATION APPROACH

Let us note:

$$\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_{21}, \beta_{22}, \ \dots\ , \beta_{2k})', \quad \boldsymbol{\theta} = (\theta_0, \theta_1, \theta_2, \theta_3, \theta_{41}, \theta_{42}, \ \dots\ , \theta_{4k})'.$$

In Equation (7), $g(a, a^*, \boldsymbol{c})$ is a function of the vector

$$(\boldsymbol{\beta}, \boldsymbol{\theta}, \sigma^2) = (\beta_0, \beta_1, \beta_{21}, \beta_{22}, \ \dots\ , \beta_{2k}, \theta_0, \theta_1, \theta_2, \theta_3, \theta_{41}, \theta_{42}, \ \dots\ , \theta_{4k}, \sigma^2)'; \tag{B1}$$

a, a^*, and $\boldsymbol{c}$ are fixed parameters.

Thus

$$\begin{aligned}
&\frac{\partial g(a, a^*, \boldsymbol{c})}{\partial \beta_0} \\
&= \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} \operatorname{expit}\left(\theta_0 + \theta_1 a + \theta_2 m + \theta_3 a m + \boldsymbol{\theta_4'}\boldsymbol{c}\right) \exp\left(-\frac{\left(m - \left(\beta_0 + \beta_1 a^* + \boldsymbol{\beta_2'}\boldsymbol{c}\right)\right)^2}{2\sigma^2}\right) \frac{m - (\beta_0 + \beta_1 a^* + \boldsymbol{\beta_2'}\boldsymbol{c})}{\sigma^2} dm \\
&= \frac{1}{\sigma^3\sqrt{2\pi}} \int_{-\infty}^{\infty} m \operatorname{expit}\left(\theta_0 + \theta_1 a + \theta_2 m + \theta_3 a m + \boldsymbol{\theta_4'}\boldsymbol{c}\right) \exp\left(-\frac{\left(m - \left(\beta_0 + \beta_1 a^* + \boldsymbol{\beta_2'}\boldsymbol{c}\right)\right)^2}{2\sigma^2}\right) dm \\
&\quad - \frac{\beta_0 + \beta_1 a^* + \boldsymbol{\beta_2'}\boldsymbol{c}}{\sigma^2} g(a, a^*, \boldsymbol{c}),
\end{aligned} \tag{B2}$$

$$\frac{\partial g(a,a^*,\boldsymbol{c})}{\partial \beta_1} = a^* \frac{\partial g(a,a^*,\boldsymbol{c})}{\partial \beta_0}, \tag{B3}$$

$$\frac{\partial g(a,a^*,\boldsymbol{c})}{\partial \beta_{2i}} = c_i \frac{\partial}{\partial \beta_0} g(a,a^*,\boldsymbol{c}), \quad i = 1,2,\ldots,k, \tag{B4}$$

$$\frac{\partial g(a,a^*,\boldsymbol{c})}{\partial \theta_0} = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} \frac{\exp\left(\theta_0 + \theta_1 a + \theta_2 m + \theta_3 am + \boldsymbol{\theta_4'c}\right)}{\left(1+\exp\left(\theta_0 + \theta_1 a + \theta_2 m + \theta_3 am + \boldsymbol{\theta_4'c}\right)\right)^2} \exp\left(-\frac{\left(m-\left(\beta_0+\beta_1 a^* + \boldsymbol{\beta_2'c}\right)\right)^2}{2\sigma^2}\right) dm, \tag{B5}$$

$$\frac{\partial g(a,a^*,\boldsymbol{c})}{\partial \theta_1} = a \frac{\partial g(a,a^*,\boldsymbol{c})}{\partial \theta_0}, \tag{B6}$$

$$\frac{\partial g(a,a^*,\boldsymbol{c})}{\partial \theta_2} = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} m \frac{\exp\left(\theta_0 + \theta_1 a + \theta_2 m + \theta_3 am + \boldsymbol{\theta_4'c}\right)}{\left(1+\exp\left(\theta_0 + \theta_1 a + \theta_2 m + \theta_3 am + \boldsymbol{\theta_4'c}\right)\right)^2} \exp\left(-\frac{\left(m-\left(\beta_0+\beta_1 a^* + \boldsymbol{\beta_2'c}\right)\right)^2}{2\sigma^2}\right) dm, \tag{B7}$$

$$\frac{\partial g(a,a^*,\boldsymbol{c})}{\partial \theta_3} = a \frac{\partial g(a,a^*,\boldsymbol{c})}{\partial \theta_2}, \tag{B8}$$

$$\frac{\partial g(a,a^*,\boldsymbol{c})}{\partial \theta_{4i}} = c_i \frac{\partial g(a,a^*,\boldsymbol{c})}{\partial \theta_0}, \quad i = 1,2,\ldots,k, \tag{B9}$$

$$\begin{aligned}
\frac{\partial g(a,a^*,\boldsymbol{c})}{\partial \sigma^2} &= [t := \sigma^2] \\
&= \frac{\partial}{\partial t}\left(\frac{1}{\sqrt{2\pi t}} \int_{-\infty}^{\infty} \operatorname{expit}\left(\theta_0 + \theta_1 a + \theta_2 m + \theta_3 am + \boldsymbol{\theta_4'c}\right) \exp\left(-\frac{\left(m-\left(\beta_0+\beta_1 a^* + \boldsymbol{\beta_2'c}\right)\right)^2}{2t}\right) dm\right) \\
&= -\frac{1}{2}\frac{1}{\sqrt{2\pi t^3}} \int_{-\infty}^{\infty} \operatorname{expit}\left(\theta_0 + \theta_1 a + \theta_2 m + \theta_3 am + \boldsymbol{\theta_4'c}\right) \exp\left(-\frac{\left(m-\left(\beta_0+\beta_1 a^* + \boldsymbol{\beta_2'c}\right)\right)^2}{2t}\right) dm \\
&\quad + \frac{1}{\sqrt{2\pi t}} \int_{-\infty}^{\infty} \operatorname{expit}\left(\theta_0 + \theta_1 a + \theta_2 m + \theta_3 am + \boldsymbol{\theta_4'c}\right) \exp\left(-\frac{\left(m-\left(\beta_0+\beta_1 a^* + \boldsymbol{\beta_2'c}\right)\right)^2}{2t}\right) \frac{\left(m-\left(\beta_0+\beta_1 a^* + \boldsymbol{\beta_2'c}\right)\right)^2}{2t^2} dm \\
&= -\frac{1}{2t} g(a,a^*,\boldsymbol{c}) \\
&\quad + \frac{1}{2t^2\sqrt{2\pi t}} \int_{-\infty}^{\infty} \operatorname{expit}\left(\theta_0 + \theta_1 a + \theta_2 m + \theta_3 am + \boldsymbol{\theta_4'c}\right) \exp\left(-\frac{\left(m-\left(\beta_0+\beta_1 a^* + \boldsymbol{\beta_2'c}\right)\right)^2}{2t}\right) \left(m-\left(\beta_0+\beta_1 a^* + \boldsymbol{\beta_2'c}\right)\right)^2 dm \\
&= -\frac{1}{2\sigma^2} g(a,a^*,\boldsymbol{c}) \\
&\quad + \frac{1}{2\sigma^4\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} \operatorname{expit}\left(\theta_0 + \theta_1 a + \theta_2 m + \theta_3 am + \boldsymbol{\theta_4'c}\right) \exp\left(-\frac{\left(m-\left(\beta_0+\beta_1 a^* + \boldsymbol{\beta_2'c}\right)\right)^2}{2\sigma^2}\right) \\
&\quad \times \left(m-\left(\beta_0+\beta_1 a^* + \boldsymbol{\beta_2'c}\right)\right)^2 dm.
\end{aligned} \tag{B10}$$

The gradient of the scalar function $g(a,a^*,\boldsymbol{c})$ with respect to the vector $(\boldsymbol{\beta},\boldsymbol{\theta},\sigma^2)$ (Equation B1) is

$$\nabla\left(g(a,a^*,\boldsymbol{c})\right) = \left(\frac{\partial g(a,a^*,\boldsymbol{c})}{\partial \boldsymbol{\beta}}, \frac{\partial g(a,a^*,\boldsymbol{c})}{\partial \boldsymbol{\theta}}, \frac{\partial g(a,a^*,\boldsymbol{c})}{\partial \sigma^2}\right)', \tag{B11}$$

where

$$\frac{\partial g(a,a^*,\boldsymbol{c})}{\partial \boldsymbol{\beta}} = \left(\frac{\partial g(a,a^*,\boldsymbol{c})}{\partial \beta_0}, \frac{\partial g(a,a^*,\boldsymbol{c})}{\partial \beta_1}, \frac{\partial g(a,a^*,\boldsymbol{c})}{\partial \beta_{21}}, \frac{\partial g(a,a^*,\boldsymbol{c})}{\partial \beta_{22}}, \ldots, \frac{\partial g(a,a^*,\boldsymbol{c})}{\partial \beta_{2k}}\right),$$

$$\frac{\partial g(a,a^*,\boldsymbol{c})}{\partial \boldsymbol{\theta}} = \left(\frac{\partial g(a,a^*,\boldsymbol{c})}{\partial \theta_0}, \frac{\partial g(a,a^*,\boldsymbol{c})}{\partial \theta_1}, \frac{\partial g(a,a^*,\boldsymbol{c})}{\partial \theta_2}, \frac{\partial g(a,a^*,\boldsymbol{c})}{\partial \theta_3}, \frac{\partial g(a,a^*,\boldsymbol{c})}{\partial \theta_{41}}, \frac{\partial g(a,a^*,\boldsymbol{c})}{\partial \theta_{42}}, \ldots, \frac{\partial g(a,a^*,\boldsymbol{c})}{\partial \theta_{4k}}\right).$$

Let us note

$$\hat{g}(a,a^*,\boldsymbol{c}) = g(a,a^*,\boldsymbol{c})\Big|_{(\boldsymbol{\beta},\theta,\sigma^2)=\left(\hat{\boldsymbol{\beta}},\hat{\theta},\widehat{\sigma^2}\right)},$$
$$\nabla\left(\hat{g}(a,a^*,\boldsymbol{c})\right) = \nabla\left(g(a,a^*,\boldsymbol{c})\right)\Big|_{(\boldsymbol{\beta},\theta,\sigma^2)=\left(\hat{\boldsymbol{\beta}},\hat{\theta},\widehat{\sigma^2}\right)},$$

where $\nabla(g(a,a^*,\boldsymbol{c}))$ is defined in (B11).

B.1 Delta method for exact natural effects odds ratios

We can express the exact $OR^{\text{NDE}}_{a,a^*|\boldsymbol{c}}$ and $OR^{\text{NIE}}_{a,a^*|\boldsymbol{c}}$ in terms of $g(a,a^*,\boldsymbol{c})$ defined in (7) as follows:

$$OR^{\text{NDE}}_{a,a^*|\boldsymbol{c}} = \frac{g(a,a^*,\boldsymbol{c})/(1-g(a,a^*,\boldsymbol{c}))}{g(a^*,a^*,\boldsymbol{c})/(1-g(a^*,a^*,\boldsymbol{c}))}, \quad OR^{\text{NIE}}_{a,a^*|\boldsymbol{c}} = \frac{g(a,a,\boldsymbol{c})/(1-g(a,a,\boldsymbol{c}))}{g(a,a^*,\boldsymbol{c})/(1-g(a,a^*,\boldsymbol{c}))}.$$

To construct the 95% confidence intervals (CIs) for $OR^{\text{NDE}}_{a,a^*|\boldsymbol{c}}$ and $OR^{\text{NIE}}_{a,a^*|\boldsymbol{c}}$ by the first-order multivariate delta method,[23] we expressed standard errors (*se*) for $\ln\left(\widehat{OR^{\text{NDE}}_{a,a^*|\boldsymbol{c}}}\right)$ and $\ln\left(\widehat{OR^{\text{NIE}}_{a,a^*|\boldsymbol{c}}}\right)$ according to the following approximate formulas:

$$se\left(\ln\left(\widehat{OR^{\text{NDE}}_{a,a^*|\boldsymbol{c}}}\right)\right) \approx \sqrt{\nabla\left(\ln\left(\widehat{OR^{\text{NDE}}_{a,a^*|\boldsymbol{c}}}\right)\right)' \Sigma\, \nabla\left(\ln\left(\widehat{OR^{\text{NDE}}_{a,a^*|\boldsymbol{c}}}\right)\right)},$$
$$se\left(\ln\left(\widehat{OR^{\text{NIE}}_{a,a^*|\boldsymbol{c}}}\right)\right) \approx \sqrt{\nabla\left(\ln\left(\widehat{OR^{\text{NIE}}_{a,a^*|\boldsymbol{c}}}\right)\right)' \Sigma\, \nabla\left(\ln\left(\widehat{OR^{\text{NIE}}_{a,a^*|\boldsymbol{c}}}\right)\right)},$$

where $\Sigma = \text{diag}\left\{\Sigma_{\hat{\boldsymbol{\beta}}}, \Sigma_{\hat{\theta}}, \Sigma_{\widehat{\sigma^2}}\right\}$ is a block matrix; $\Sigma_{\hat{\boldsymbol{\beta}}}$, $\Sigma_{\hat{\theta}}$, and $\Sigma_{\widehat{\sigma^2}}$ are the covariance matrices for $\hat{\boldsymbol{\beta}}$, $\hat{\theta}$, and $\widehat{\sigma^2}$, respectively. To estimate $\Sigma_{\widehat{\sigma^2}} = \text{Var}\left(\widehat{\sigma^2}\right)$, we used the following unbiased estimator proposed by Gao and Luo:[46]

$$W_l = \frac{2\text{MSE}^2}{l+2},$$

where MSE is the residual mean squared error from the mediator model (Equation 1), $l = n - p$, and n and p are the number of observations and the number of regression coefficients (including intercept) in the mediator model (Equation 1), respectively; see Gao and Luo[46] for more information on the properties of W_l.

The gradients of $\ln\left(\widehat{OR^{\text{NDE}}_{a,a^*|\boldsymbol{c}}}\right)$ and $\ln\left(\widehat{OR^{\text{NIE}}_{a,a^*|\boldsymbol{c}}}\right)$ are expressed using $\nabla(\hat{g}(a,a^*,\boldsymbol{c}))$ as follows:

$$\nabla\left(\ln\left(\widehat{OR^{\text{NDE}}_{a,a^*|\boldsymbol{c}}}\right)\right) = \frac{\nabla(\hat{g}(a,a^*,\boldsymbol{c}))}{\hat{g}(a,a^*,\boldsymbol{c})(1-\hat{g}(a,a^*,\boldsymbol{c}))} - \frac{\nabla(\hat{g}(a^*,a^*,\boldsymbol{c}))}{\hat{g}(a^*,a^*,\boldsymbol{c})(1-\hat{g}(a^*,a^*,\boldsymbol{c}))},$$
$$\nabla\left(\ln\left(\widehat{OR^{\text{NIE}}_{a,a^*|\boldsymbol{c}}}\right)\right) = \frac{\nabla(\hat{g}(a,a,\boldsymbol{c}))}{\hat{g}(a,a,\boldsymbol{c})(1-\hat{g}(a,a,\boldsymbol{c}))} - \frac{\nabla(\hat{g}(a,a^*,\boldsymbol{c}))}{\hat{g}(a,a^*,\boldsymbol{c})(1-\hat{g}(a,a^*,\boldsymbol{c}))}.$$

Thus,

$$\ln\left(\widehat{OR^{\text{NDE}}_{a,a^*|\boldsymbol{c}}}\right) \pm \Phi^{-1}(0.975)\cdot se\left(\ln\left(\widehat{OR^{\text{NDE}}_{a,a^*|\boldsymbol{c}}}\right)\right),$$
$$\ln\left(\widehat{OR^{\text{NIE}}_{a,a^*|\boldsymbol{c}}}\right) \pm \Phi^{-1}(0.975)\cdot se\left(\ln\left(\widehat{OR^{\text{NIE}}_{a,a^*|\boldsymbol{c}}}\right)\right),$$

are the approximate 95% CIs for $\ln\left(OR^{\text{NDE}}_{a,a^*|\boldsymbol{c}}\right)$ and $\ln\left(OR^{\text{NIE}}_{a,a^*|\boldsymbol{c}}\right)$, respectively, and the corresponding 95% CIs for $OR^{\text{NDE}}_{a,a^*|\boldsymbol{c}}$ and $OR^{\text{NIE}}_{a,a^*|\boldsymbol{c}}$ are

$$\widehat{OR^{\text{NDE}}_{a,a^*|\boldsymbol{c}}}\cdot\exp\left(\pm\Phi^{-1}(0.975)\cdot se\left(\ln\left(\widehat{OR^{\text{NDE}}_{a,a^*|\boldsymbol{c}}}\right)\right)\right),$$
$$\widehat{OR^{\text{NIE}}_{a,a^*|\boldsymbol{c}}}\cdot\exp\left(\pm\Phi^{-1}(0.975)\cdot se\left(\ln\left(\widehat{OR^{\text{NIE}}_{a,a^*|\boldsymbol{c}}}\right)\right)\right).$$

Finally, we have for the total effect odds ratio $OR^{\text{TE}}_{a,a^*|\boldsymbol{c}}$:

$$\ln\left(\widehat{OR^{\text{TE}}_{a,a^*|\boldsymbol{c}}}\right) = \ln\left(\widehat{OR^{\text{NDE}}_{a,a^*|\boldsymbol{c}}}\right) + \ln\left(\widehat{OR^{\text{NIE}}_{a,a^*|\boldsymbol{c}}}\right),$$
$$\nabla\left(\ln\left(\widehat{OR^{\text{TE}}_{a,a^*|\boldsymbol{c}}}\right)\right) = \nabla\left(\ln\left(\widehat{OR^{\text{NDE}}_{a,a^*|\boldsymbol{c}}}\right)\right) + \nabla\left(\ln\left(\widehat{OR^{\text{NIE}}_{a,a^*|\boldsymbol{c}}}\right)\right),$$
$$se\left(\ln\left(\widehat{OR^{\text{TE}}_{a,a^*|\boldsymbol{c}}}\right)\right) \approx \sqrt{\nabla\left(\ln\left(\widehat{OR^{\text{TE}}_{a,a^*|\boldsymbol{c}}}\right)\right)' \Sigma\, \nabla\left(\ln\left(\widehat{OR^{\text{TE}}_{a,a^*|\boldsymbol{c}}}\right)\right)}.$$

Thus,

$$\ln\left(\widehat{OR^{\text{TE}}_{a,a^*|\boldsymbol{c}}}\right) \pm \Phi^{-1}(0.975) \cdot se\left(\ln\left(\widehat{OR^{\text{TE}}_{a,a^*|\boldsymbol{c}}}\right)\right)$$

is the approximate 95% CI for $\ln\left(OR^{\text{TE}}_{a,a^*|\boldsymbol{c}}\right)$, and the 95% CI for $OR^{\text{TE}}_{a,a^*|\boldsymbol{c}}$ is approximated by

$$\widehat{OR^{\text{TE}}_{a,a^*|\boldsymbol{c}}} \cdot \exp\left(\pm\Phi^{-1}(0.975) \cdot se\left(\ln\left(\widehat{OR^{\text{TE}}_{a,a^*|\boldsymbol{c}}}\right)\right)\right).$$

B.2 Delta method for exact natural effects risk ratios

For the RR scale, we have that

$$RR^{\text{NDE}}_{a,a^*|\boldsymbol{c}} = \frac{g(a,a^*,\boldsymbol{c})}{g(a^*,a^*,\boldsymbol{c})}, \quad RR^{\text{NIE}}_{a,a^*|\boldsymbol{c}} = \frac{g(a,a,\boldsymbol{c})}{g(a,a^*,\boldsymbol{c})},$$

and, correspondingly,

$$\nabla\left(\ln\left(\widehat{RR^{\text{NDE}}_{a,a^*|\boldsymbol{c}}}\right)\right) = \frac{\nabla\left(\hat{g}(a,a^*,\boldsymbol{c})\right)}{\hat{g}(a,a^*,\boldsymbol{c})} - \frac{\nabla\left(\hat{g}(a^*,a^*,\boldsymbol{c})\right)}{\hat{g}(a^*,a^*,\boldsymbol{c})},$$
$$\nabla\left(\ln\left(\widehat{RR^{\text{NIE}}_{a,a^*|\boldsymbol{c}}}\right)\right) = \frac{\nabla\left(\hat{g}(a,a,\boldsymbol{c})\right)}{\hat{g}(a,a,\boldsymbol{c})} - \frac{\nabla\left(\hat{g}(a,a^*,\boldsymbol{c})\right)}{\hat{g}(a,a^*,\boldsymbol{c})}.$$

Thus,

$$se\left(\ln\left(\widehat{RR^{\text{NDE}}_{a,a^*|\boldsymbol{c}}}\right)\right) \approx \sqrt{\nabla\left(\ln\left(\widehat{RR^{\text{NDE}}_{a,a^*|\boldsymbol{c}}}\right)\right)' \Sigma\, \nabla\left(\ln\left(\widehat{RR^{\text{NDE}}_{a,a^*|\boldsymbol{c}}}\right)\right)},$$
$$se\left(\ln\left(\widehat{RR^{\text{NIE}}_{a,a^*|\boldsymbol{c}}}\right)\right) \approx \sqrt{\nabla\left(\ln\left(\widehat{RR^{\text{NIE}}_{a,a^*|\boldsymbol{c}}}\right)\right)' \Sigma\, \nabla\left(\ln\left(\widehat{RR^{\text{NIE}}_{a,a^*|\boldsymbol{c}}}\right)\right)},$$

and the 95% CIs for $RR^{\text{NDE}}_{a,a^*|\boldsymbol{c}}$, and $RR^{\text{NIE}}_{a,a^*|\boldsymbol{c}}$ can be approximated by

$$\widehat{RR^{\text{NDE}}_{a,a^*|\boldsymbol{c}}} \cdot \exp\left(\pm\Phi^{-1}(0.975) \cdot se\left(\ln\left(\widehat{RR^{\text{NDE}}_{a,a^*|\boldsymbol{c}}}\right)\right)\right),$$
$$\widehat{RR^{\text{NIE}}_{a,a^*|\boldsymbol{c}}} \cdot \exp\left(\pm\Phi^{-1}(0.975) \cdot se\left(\ln\left(\widehat{RR^{\text{NIE}}_{a,a^*|\boldsymbol{c}}}\right)\right)\right).$$

Finally, we have for the total effect risk ratio $RR^{\text{TE}}_{a,a^*|\boldsymbol{c}}$:

$$\ln\left(\widehat{RR^{\text{TE}}_{a,a^*|\boldsymbol{c}}}\right) = \ln\left(\widehat{RR^{\text{NDE}}_{a,a^*|\boldsymbol{c}}}\right) + \ln\left(\widehat{RR^{\text{NIE}}_{a,a^*|\boldsymbol{c}}}\right),$$
$$\nabla\left(\ln\left(\widehat{RR^{\text{TE}}_{a,a^*|\boldsymbol{c}}}\right)\right) = \nabla\left(\ln\left(\widehat{RR^{\text{NDE}}_{a,a^*|\boldsymbol{c}}}\right)\right) + \nabla\left(\ln\left(\widehat{RR^{\text{NIE}}_{a,a^*|\boldsymbol{c}}}\right)\right),$$
$$se\left(\ln\left(\widehat{RR^{\text{TE}}_{a,a^*|\boldsymbol{c}}}\right)\right) \approx \sqrt{\nabla\left(\ln\left(\widehat{RR^{\text{TE}}_{a,a^*|\boldsymbol{c}}}\right)\right)' \Sigma\, \nabla\left(\ln\left(\widehat{RR^{\text{TE}}_{a,a^*|\boldsymbol{c}}}\right)\right)}.$$

Thus, the 95% CI for $RR^{\mathrm{TE}}_{a,a^*|\boldsymbol{c}}$ can be approximated by

$$\widehat{RR^{\mathrm{TE}}_{a,a^*|\boldsymbol{c}}} \cdot \exp\left(\pm\Phi^{-1}(0.975) \cdot se\left(\ln\left(\widehat{RR^{\mathrm{TE}}_{a,a^*|\boldsymbol{c}}}\right)\right)\right).$$

B.3 Delta method for exact natural effects risk differences

For the RD scale, we have:

$$RD^{\mathrm{NDE}}_{a,a^*|\boldsymbol{c}} = g(a,a^*,\boldsymbol{c}) - g(a^*,a^*,\boldsymbol{c}), \quad RD^{\mathrm{NIE}}_{a,a^*|\boldsymbol{c}} = g(a,a,\boldsymbol{c}) - g(a,a^*,\boldsymbol{c}),$$
$$RD^{\mathrm{TE}}_{a,a^*|\boldsymbol{c}} = RD^{\mathrm{NDE}}_{a,a^*|\boldsymbol{c}} + RD^{\mathrm{NIE}}_{a,a^*|\boldsymbol{c}}$$

and, correspondingly,

$$\nabla\left(\widehat{RD^{\mathrm{NDE}}_{a,a^*|\boldsymbol{c}}}\right) = \nabla\left(\hat{g}(a,a^*,\boldsymbol{c})\right) - \nabla\left(\hat{g}(a^*,a^*,\boldsymbol{c})\right),$$
$$\nabla\left(\widehat{RD^{\mathrm{NIE}}_{a,a^*|\boldsymbol{c}}}\right) = \nabla\left(\hat{g}(a,a,\boldsymbol{c})\right) - \nabla\left(\hat{g}(a,a^*,\boldsymbol{c})\right),$$
$$\nabla\left(\widehat{RD^{\mathrm{TE}}_{a,a^*|\boldsymbol{c}}}\right) = \nabla\left(\widehat{RD^{\mathrm{NDE}}_{a,a^*|\boldsymbol{c}}}\right) + \nabla\left(\widehat{RD^{\mathrm{NIE}}_{a,a^*|\boldsymbol{c}}}\right).$$

Thus,

$$se\left(\widehat{RD^{\mathrm{NDE}}_{a,a^*|\boldsymbol{c}}}\right) \approx \sqrt{\nabla\left(\widehat{RD^{\mathrm{NDE}}_{a,a^*|\boldsymbol{c}}}\right)' \Sigma\, \nabla\left(\widehat{RD^{\mathrm{NDE}}_{a,a^*|\boldsymbol{c}}}\right)},$$
$$se\left(\widehat{RD^{\mathrm{NIE}}_{a,a^*|\boldsymbol{c}}}\right) \approx \sqrt{\nabla\left(\widehat{RD^{\mathrm{NIE}}_{a,a^*|\boldsymbol{c}}}\right)' \Sigma\, \nabla\left(\widehat{RD^{\mathrm{NIE}}_{a,a^*|\boldsymbol{c}}}\right)},$$
$$se\left(\widehat{RD^{\mathrm{TE}}_{a,a^*|\boldsymbol{c}}}\right) \approx \sqrt{\nabla\left(\widehat{RD^{\mathrm{TE}}_{a,a^*|\boldsymbol{c}}}\right)' \Sigma\, \nabla\left(\widehat{RD^{\mathrm{TE}}_{a,a^*|\boldsymbol{c}}}\right)},$$

and the 95% CIs for $RD^{\mathrm{NDE}}_{a,a^*|\boldsymbol{c}}$, $RD^{\mathrm{NIE}}_{a,a^*|\boldsymbol{c}}$, and $RD^{\mathrm{TE}}_{a,a^*|\boldsymbol{c}}$ can be approximated by

$$\widehat{RD^{\mathrm{NDE}}_{a,a^*|\boldsymbol{c}}} \pm \Phi^{-1}(0.975) \cdot se\left(\widehat{Rd^{\mathrm{NDE}}_{a,a^*|\boldsymbol{c}}}\right),$$
$$\widehat{RD^{\mathrm{NIE}}_{a,a^*|\boldsymbol{c}}} \pm \Phi^{-1}(0.975) \cdot se\left(\widehat{RD^{\mathrm{NIE}}_{a,a^*|\boldsymbol{c}}}\right),$$
$$\widehat{RD^{\mathrm{TE}}_{a,a^*|\boldsymbol{c}}} \pm \Phi^{-1}(0.975) \cdot se\left(\widehat{RD^{\mathrm{TE}}_{a,a^*|\boldsymbol{c}}}\right).$$

APPENDIX C. CONVERGENCE OF IMPROPER INTEGRALS IN THE EXACT APPROACH

All improper integrals used to construct the exact point and interval estimators (see Equations (7), (B2)-(B10)) are (absolutely) convergent since the corresponding integrands are majorated on $(-\infty,\infty)$ by the zeroth-, first-, or second-order moments (or their linear combination) of the normal distribution:[47]

$$0 < \frac{1}{\sqrt{2\pi\sigma^2}}\,\mathrm{expit}\,(a+bm)\,\exp\left(-\frac{(m-c)^2}{2\sigma^2}\right) < \frac{1}{\sqrt{2\pi\sigma^2}}\,\exp\left(-\frac{(m-c)^2}{2\sigma^2}\right),$$
$$0 < \frac{1}{\sqrt{2\pi\sigma^2}}\,\frac{\exp(a+bm)}{(1+\exp(a+bm))^2}\,\exp\left(-\frac{(m-c)^2}{2\sigma^2}\right) < \frac{1}{\sqrt{2\pi\sigma^2}}\,\exp\left(-\frac{(m-c)^2}{2\sigma^2}\right),$$
$$\left|\frac{m}{\sqrt{2\pi\sigma^2}}\,\mathrm{expit}\,(a+bm)\,\exp\left(-\frac{(m-c)^2}{2\sigma^2}\right)\right| < \frac{|m|}{\sqrt{2\pi\sigma^2}}\,\exp\left(-\frac{(m-c)^2}{2\sigma^2}\right) < \frac{m^2+1}{\sqrt{2\pi\sigma^2}}\,\exp\left(-\frac{(m-c)^2}{2\sigma^2}\right),$$
$$\left|\frac{m}{\sqrt{2\pi\sigma^2}}\,\frac{\exp(a+bm)}{(1+\exp(a+bm))^2}\,\exp\left(-\frac{(m-c)^2}{2\sigma^2}\right)\right| < \frac{|m|}{\sqrt{2\pi\sigma^2}}\,\exp\left(-\frac{(m-c)^2}{2\sigma^2}\right) < \frac{m^2+1}{\sqrt{2\pi\sigma^2}}\,\exp\left(-\frac{(m-c)^2}{2\sigma^2}\right),$$
$$0 \le \frac{m^2}{\sqrt{2\pi\sigma^2}}\,\mathrm{expit}\,(a+bm)\,\exp\left(-\frac{(m-c)^2}{2\sigma^2}\right) \le \frac{m^2}{\sqrt{2\pi\sigma^2}}\,\exp\left(-\frac{(m-c)^2}{2\sigma^2}\right), \qquad \forall m \in (-\infty,\infty).$$

APPENDIX D. RESULTS OF THE CRUDE SIMULATION STUDY

The results of the crude simulation study are presented in Tables D1-D3.

TABLE D1 Crude simulation study: Odds ratio scale (1000 samples of size $n = 5000$)

Effect	Estimator	True value	Mean	Bias	Relative bias (%)	SD	RMSE	Delta/robust SE[a] CP (%)	Bootstrap[b] CP (%)
Crude scenario 1									
NDE OR	Exact	1.539	1.555	0.015	0.98	0.230	0.231	94.0	93.9
	Gaynor et al		1.534	−0.001	−0.38	0.216	0.216	-	94.4
	VV		1.562	0.023	1.47	0.229	0.230	94.0	94.2
	NEM		1.555	0.015	0.98	0.230	0.231	94.1	-
NIE OR	Exact	1.380	1.384	0.004	0.30	0.121	0.121	94.5	94.4
	Gaynor et al		1.361	−0.019	−1.40	0.126	0.127	-	94.0
	VV		1.389	0.009	0.66	0.125	0.125	94.6	94.4
	NEM		1.384	0.004	0.27	0.121	0.121	94.1	-
TE OR	Exact	2.125	2.136	0.011	0.52	0.256	0.256	94.4	94.7
	Gaynor et al		2.072	−0.053	−2.47	0.244	0.250	-	94.3
	VV		2.154	0.030	1.39	0.260	0.262	93.7	93.7
	NEM		2.135	0.011	0.50	0.256	0.256	94.5	-
Crude scenario 2									
NDE OR	Exact	1.530	1.536	0.001	0.46	0.155	0.156	94.8	94.4
	Gaynor et al		1.531	0.002	0.11	0.153	0.153	-	94.4
	VV		1.554	0.025	1.61	0.153	0.155	93.9	93.7
	NEM		1.536	0.001	0.46	0.155	0.156	94.8	-
NIE OR	Exact	1.376	1.378	0.002	0.15	0.084	0.084	94.5	94.6
	Gaynor et al		1.362	−0.014	−1.02	0.084	0.085	-	94.2
	VV		1.387	0.011	0.78	0.089	0.089	94.6	94.2
	NEM		1.378	0.002	0.13	0.084	0.084	94.7	-
TE OR	Exact	2.105	2.110	0.005	0.24	0.172	0.172	94.5	94.3
	Gaynor et al		2.078	−0.027	−1.27	0.167	0.169	-	93.7
	VV		2.148	0.043	2.05	0.179	0.184	94.1	94.1
	NEM		2.110	0.005	0.23	0.172	0.172	94.4	-
Crude scenario 3									
NDE OR	Exact	1.506	1.509	0.003	0.23	0.121	0.121	94.5	93.9
	Gaynor et al		1.508	0.002	0.16	0.121	0.121	-	93.8
	VV		1.551	0.046	3.05	0.113	0.122	93.6	93.0
	NEM		1.509	0.003	0.23	0.121	0.121	94.4	-
NIE OR	Exact	1.373	1.376	0.004	0.26	0.074	0.074	93.3	93.7
	Gaynor et al		1.376	0.003	0.24	0.078	0.078	-	93.6
	VV		1.389	0.016	1.18	0.080	0.082	93.5	93.6
	NEM		1.377	0.004	0.27	0.073	0.073	93.6	-
TE OR	Exact	2.067	2.071	0.005	0.22	0.129	0.129	96.0	95.6
	Gaynor et al		2.069	0.003	0.13	0.129	0.131	-	95.6
	VV		2.151	0.084	4.06	0.146	0.169	92.5	91.4
	NEM		2.072	0.005	0.23	0.129	0.130	96.0	-

(Continues)

TABLE D1 (Continued)

Effect	Estimator	True value	Mean	Bias	Relative bias (%)	SD	RMSE	Delta/robust SE[a] CP (%)	Bootstrap[b] CP (%)
Crude scenario 4									
NDE OR	Exact	1.489	1.500	0.012	0.79	0.163	0.164	95.4	94.6
	Gaynor et al		1.516	0.028	1.84	0.170	0.173	-	94.8
	VV		1.565	0.076	5.12	0.142	0.161	92.7	90.5
	NEM		1.502	0.012	0.79	0.163	0.164	95.1	-
NIE OR	Exact	1.377	1.386	0.009	0.64	0.100	0.101	94.7	95.2
	Gaynor et al		1.413	0.035	2.57	0.105	0.111	-	94.5
	VV		1.394	0.017	1.26	0.106	0.108	94.8	95.1
	NEM		1.387	0.009	0.66	0.100	0.101	95.0	-
TE OR	Exact	2.050	2.069	0.019	0.91	0.168	0.169	95.3	95.3
	Gaynor et al		2.130	0.080	3.91	0.182	0.199	-	92.6
	VV		2.177	0.126	6.16	0.204	0.240	91.6	89.9
	NEM		2.069	0.019	0.93	0.168	0.169	95.1	-
Crude scenario 5									
NDE OR	Exact	1.484	1.515	0.031	2.06	0.257	0.259	94.9	94.3
	Gaynor et al		1.533	0.048	3.26	0.280	0.285	-	94.7
	VV		1.587	0.103	6.93	0.219	0.242	93.3	91.1
	NEM		1.515	0.031	2.06	0.257	0.259	94.8	-
NIE OR	Exact	1.381	1.390	0.009	0.68	0.153	0.153	94.1	94.2
	Gaynor et al		1.410	0.029	2.08	0.151	0.154	-	93.3
	VV		1.395	0.015	1.05	0.158	0.159	94.6	93.7
	NEM		1.391	0.010	0.71	0.153	0.154	94.4	-
TE OR	Exact	2.050	2.080	0.031	1.50	0.258	0.260	95.2	94.4
	Gaynor et al		2.134	0.084	4.11	0.292	0.304	-	93.6
	VV		2.203	0.153	7.48	0.321	0.356	94.4	91.6
	NEM		2.081	0.031	1.52	0.258	0.260	95.1	-

Abbreviations: CP, coverage probability; NDE, natural direct effect; NEM, natural effect model (imputation approach); NIE, natural indirect effect; OR, odds ratio; RMSE, root mean squared error; SD, standard deviation; SE, standard error; TE, total effect; VV, VanderWeele and Vansteelandt's approach.

[a]Delta method[23] for exact and VanderWeele and Vansteelandt's estimators; robust standard errors based on the sandwich estimator[27] for NEMs.

[b]Percentile bootstrap[26] based on 500 resamples with replacement.

TABLE D2 Crude simulation study: Risk ratio scale (1000 samples of size $n = 5000$)

Effect	Estimator	True value	Mean	Bias	Relative bias (%)	SD	RMSE	Delta/robust SE CP (%)[a]	Bootstrap CP (%)[b]
Crude scenario 1									
NDE RR	Exact	1.498	1.510	0.012	0.81	0.208	0.209	94.0	94.0
	NEM		1.510	0.012	0.81	0.208	0.209	94.1	-
NIE RR	Exact	1.341	1.345	0.004	0.28	0.109	0.109	94.2	94.3
	NEM		1.344	0.003	0.25	0.108	0.108	94.0	-
TE RR	Exact	2.009	2.018	0.008	0.42	0.223	0.223	94.1	94.7
	NEM		2.017	0.008	0.40	0.222	0.223	94.5	-
Crude scenario 2									
NDE RR	Exact	1.433	1.437	0.004	0.28	0.121	0.121	94.5	94.4
	NEM		1.437	0.004	0.28	0.121	0.121	94.6	-
NIE RR	Exact	1.288	1.290	0.002	0.12	0.064	0.064	94.6	94.8
	NEM		1.290	0.001	0.11	0.064	0.064	94.6	-
TE RR	Exact	1.846	1.848	0.003	0.14	0.122	0.122	94.5	94.2
	NEM		1.848	0.002	0.13	0.122	0.122	94.4	-
Crude scenario 3									
NDE RR	Exact	1.257	1.257	−0.000	−0.01	0.053	0.053	94.2	93.9
	NEM		1.257	−0.000	−0.00	0.053	0.053	94.1	-
NIE RR	Exact	1.160	1.162	0.002	0.13	0.032	0.032	93.4	94.0
	NEM		1.162	0.002	0.13	0.032	0.032	93.8	-
TE RR	Exact	1.459	1.459	0.001	0.04	0.045	0.045	95.6	95.0
	NEM		1.459	0.001	0.05	0.045	0.045	95.4	-
Crude scenario 4									
NDE RR	Exact	1.094	1.094	−0.000	−0.00	0.024	0.024	94.8	94.5
	NEM		1.094	−0.000	−0.00	0.024	0.024	94.8	-
NIE RR	Exact	1.056	1.057	0.001	0.11	0.015	0.015	94.1	94.5
	NEM		1.057	0.001	0.11	0.015	0.015	94.1	-
TE RR	Exact	1.156	1.156	0.001	0.08	0.017	0.017	94.8	94.9
	NEM		1.156	0.001	0.08	0.017	0.017	94.8	-
Crude scenario 5									
NDE RR	Exact	1.039	1.039	−0.000	−0.01	0.015	0.015	94.2	94.8
	NEM		1.039	−0.000	−0.01	0.015	0.015	93.8	-
NIE RR	Exact	1.023	1.023	0.000	0.05	0.009	0.009	93.8	93.7
	NEM		1.024	0.000	0.05	0.009	0.009	93.6	-
TE RR	Exact	1.063	1.064	0.000	0.02	0.010	0.010	95.2	94.8
	NEM		1.064	0.000	0.02	0.010	0.010	95.1	-

Abbreviations: CP, coverage probability; NDE, natural direct effect; NEM, natural effect model (imputation approach); NIE, natural indirect effect; RMSE, root mean squared error; RR, risk ratio; SD, standard deviation; SE, standard error; TE, total effect.

[a]Delta method[23] for exact estimators; robust standard errors based on the sandwich estimator[27] for NEMs.

[b]Percentile bootstrap[26] based on 500 resamples with replacement.

TABLE D3 Crude simulation study: Risk difference scale (1000 samples of size $n = 5000$)

Effect	Estimator	True value	Mean	Bias	Relative bias (%)	SD	RMSE	Delta/robust SE CP (%)[a]	Bootstrap CP (%)[b]
Crude scenario 1									
NDE RD	Exact	0.0254	0.0256	0.0002	0.68	0.0096	0.0096	93.9	94.1
	Imai et al		0.0260	0.0006	2.31	0.0096	0.0096	94.3	-
NIE RD	Exact	0.0261	0.0258	−0.0003	−1.15	0.0065	0.0065	93.7	93.3
	Imai et al		0.0257	−0.0004	−1.69	0.0065	0.0066	93.4	-
TE RD	Exact	0.0515	0.0514	−0.0001	−0.25	0.0091	0.0091	94.1	93.8
	Imai et al		0.0517	0.0001	0.28	0.0090	0.0090	93.7	-
Crude scenario 2									
NDE RD	Exact	0.0550	0.0551	0.0001	0.20	0.0140	0.0140	95.2	94.7
	Imai et al		0.0555	0.0005	0.84	0.0140	0.0140	94.9	-
NIE RD	Exact	0.0524	0.0522	−0.0003	−0.52	0.0093	0.0093	94.2	94.1
	Imai et al		0.0520	−0.0005	−0.91	0.0093	0.0093	94.5	-
TE RD	Exact	0.1075	0.1073	−0.0002	−0.15	0.0124	0.0124	94.5	94.5
	Imai et al		0.1075	−0.0000	−0.01	0.0124	0.0124	94.5	-
Crude scenario 3									
NDE RD	Exact	0.1005	0.1003	−0.0003	−0.26	0.0199	0.0199	94.5	93.7
	Imai et al		0.1003	−0.0002	−0.20	0.0199	0.0199	94.2	-
NIE RD	Exact	0.0788	0.0790	0.0002	0.31	0.0013	0.0013	93.3	93.6
	Imai et al		0.0788	0.0001	0.08	0.0013	0.0013	93.5	-
TE RD	Exact	0.1793	0.1792	−0.0000	−0.01	0.0152	0.0152	95.8	95.5
	Imai et al		0.1791	−0.0001	−0.07	0.0152	0.0152	95.9	-
Crude scenario 4									
NDE RD	Exact	0.0694	0.0693	−0.0001	−0.15	0.0175	0.0175	94.9	94.7
	Imai et al		0.0688	−0.0006	−0.87	0.0174	0.0174	95.4	-
NIE RD	Exact	0.0450	0.0457	0.0007	1.52	0.0115	0.0115	94.5	94.7
	Imai et al		0.0459	0.0009	1.96	0.0115	0.0115	95.2	-
TE RD	Exact	0.1144	0.1150	0.0006	0.51	0.0118	0.0118	95.1	94.9
	Imai et al		0.1147	0.0003	0.24	0.0118	0.0118	94.8	-
Crude scenario 5									
NDE RD	Exact	0.0349	0.0347	−0.0002	−0.49	0.0129	0.0129	94.0	94.8
	Imai et al		0.0339	−0.0010	−2.79	0.0129	0.0129	94.7	-
NIE RD	Exact	0.0211	0.0215	0.0003	1.64	0.0084	0.0084	93.9	93.6
	Imai et al		0.0219	0.0008	3.79	0.0085	0.0085	94.7	-
TE RD	Exact	0.0560	0.0562	0.0002	0.31	0.0083	0.0083	95.3	94.8
	Imai et al		0.0559	−0.0002	−0.31	0.0083	0.0083	95.3	-

Abbreviations: CP, coverage probability; NDE, natural direct effect; NIE, natural indirect effect; RD, risk difference; RMSE, root mean squared error; SD, standard deviation; SE, standard error; TE, total effect.

[a]Delta method[23] for exact estimators; White's heteroskedasticity-consistent estimator for the covariance matrix[28] for approach by Imai et al.

[b]Percentile bootstrap[26] based on 500 resamples with replacement.

APPENDIX E. ESTIMATION OF THE PARAMETER *S* IN THE APPROACH BY GAYNOR ET AL

Table E1 presents the distribution of the parameter *s* involved in the approximate approach by Gaynor et al[5] for each crude and adjusted simulation scenario.

TABLE E1 Distribution of the parameter *s* estimates in the approach by Gaynor et al[5]

Simulation scenario	Estimated outcome prevalence (%)	Mean[a]	SD	Min	Max	Median	First quartile	Third quartile
Crude scenario 1	6.66	0.498	0.023	0.415	0.621	0.508	0.473	0.513
Crude scenario 2	15.95	0.558	0.011	0.522	0.603	0.562	0.558	0.565
Crude scenario 3	44.48	0.622	0.001	0.616	0.625	0.622	0.621	0.623
Crude scenario 4	77.24	0.593	0.007	0.575	0.609	0.592	0.588	0.598
Crude scenario 5	90.03	0.530	0.017	0.463	0.573	0.524	0.518	0.538
Adjusted scenario 1	7.64	0.495	0.019	0.437	0.580	0.492	0.481	0.507
Adjusted scenario 2	17.94	0.559	0.011	0.523	0.594	0.559	0.552	0.567
Adjusted scenario 3	47.71	0.621	0.001	0.616	0.625	0.621	0.620	0.622
Adjusted scenario 4	79.28	0.582	0.007	0.555	0.611	0.582	0.578	0.587
Adjusted scenario 5	91.03	0.515	0.012	0.470	0.562	0.515	0.508	0.521

Abbreviation: SD, standard deviation.

[a]Values of the scaling parameter *s* according to the literature: $s \approx 0.551$ by Cox[11] based on the equality of variances; $s \approx 0.572$ using Kullback-Leibler information criterion;[12] $s \approx 0.588$ from the minimax solution;[10] $s \approx 0.625$ using comparative approach by Amemiya.[12,36]

APPENDIX F. RESULTS OF THE SIMULATION STUDY WITH MARGINALLY BUT NOT CONDITIONALLY RARE OUTCOME

Table F1 reports the results of the adjusted simulation study with a marginally but not conditionally rare outcome.

TABLE F1 Adjusted simulation study where the outcome is rare marginally but not conditionally (1000 samples of size $n = 5000$)

Effect	Estimator	True value	Mean	Bias	Relative bias (%)	SD	RMSE	CP (%)[a]
OR scale								
NDE OR	Exact	1.477	1.482	0.004	0.29	0.255	0.255	94.6
	Gaynor et al		1.468	−0.009	−0.63	0.256	0.256	95.3
	VV		2.485	1.007	68.18	2.116	2.344	97.7
	NEM		1.481	0.003	0.23	0.254	0.254	94.3
NIE OR	Exact	3.452	3.505	0.053	1.54	0.474	0.477	94.5
	Gaynor et al		3.359	−0.093	−2.68	0.573	0.580	95.0
	VV		8.513	5.062	146.66	3.918	6.401	37.8
	NEM		3.494	0.043	1.24	0.494	0.496	94.7
TE OR	Exact	5.099	5.139	0.040	0.79	0.843	0.844	95.0
	Gaynor et al		4.877	−0.222	−4.36	0.935	0.961	95.7
	VV		26.859	21.760	426.72	52.480	56.813	79.4
	NEM		5.120	0.022	0.42	0.864	0.864	95.2
RR scale								
NDE RR	Exact	1.421	1.421	−0.000	−0.02	0.215	0.215	94.6
	NEM		1.419	−0.002	−0.15	0.215	0.215	94.5
NIE RR	Exact	2.681	2.714	0.032	1.21	0.307	0.309	93.9
	NEM		2.694	0.012	0.46	0.313	0.313	93.5
TE RR	Exact	3.811	3.814	0.003	0.08	0.441	0.441	95.5
	NEM		3.781	−0.030	−0.79	0.449	0.450	95.4
RD scale								
NDE RD	Exact	0.035	0.035	−0.000	−0.65	0.017	0.017	94.1
	Imai et al		0.037	0.002	5.85	0.017	0.018	95.5
NIE RD	Exact	0.197	0.197	−0.000	−0.12	0.026	0.026	95.1
	Imai et al		0.196	−0.001	−0.71	0.025	0.025	95.5
TE RD	Exact	0.232	0.231	−0.000	−0.20	0.033	0.033	95.2
	Imai et al		0.232	0.006	0.28	0.033	0.033	95.9

Abbreviations: CP, coverage probability; NDE, natural direct effect; NEM, natural effect model (imputation approach); NIE, natural indirect effect; OR, odds ratio; RD, risk difference; RMSE, root mean squared error; RR, risk ratio; SD, standard deviation; TE, total effect; VV, VanderWeele and Vansteelandt's approach.

[a]Delta method[23] for exact and VanderWeele and Vansteelandt's estimators; percentile bootstrap[26] based on 500 resamples with replacement for approach by Gaynor et al; robust standard errors based on the sandwich estimator[27] for NEMs; White's heteroskedasticity-consistent estimator for the covariance matrix[28] for approach by Imai et al.

APPENDIX G. RESULTS OF THE SIMULATION STUDY WITH FIRTH'S PENALIZATION

Tables G1-G3 present the results of the simulation study performed to examine the impact of Firth's penalization on the exact estimators.

TABLE G1 Adjusted simulation study with Firth's penalization: Odds ratio scale (1000 samples of sizes $n = 150,250, 500$)

Effect	Estimation method	True value	Mean	Bias	Relative bias (%)	SD	RMSE	Delta CP (%)
Adjusted scenario 1, n = 150, estimated EPV = 2.29								
NDE OR	Penalized	1.550	2.020	0.470	30.32	1.898	1.955	95.7
NIE OR	Penalized	1.380	1.529	0.149	10.80	0.814	0.828	94.7
TE OR	Penalized	2.139	2.437	0.298	13.94	1.686	1.712	96.8
Adjusted scenario 1, n = 250, estimated EPV = 3.82								
NDE OR	Conventional	1.550	1.918	0.368	23.77	1.364	1.413	94.6
	Penalized		1.878	0.328	21.16	1.173	1.218	95.5
NIE OR	Conventional	1.380	1.534	0.155	11.20	0.706	0.723	93.8
	Penalized		1.464	0.085	6.14	0.591	0.597	94.7
TE OR	Conventional	2.139	2.508	0.369	17.25	1.480	1.525	95.8
	Penalized		2.404	0.266	12.43	1.205	1.234	96.5
Adjusted scenario 1, n = 500, estimated EPV = 7.64								
NDE OR	Conventional	1.550	1.696	0.146	9.41	0.774	0.788	94.9
	Penalized		1.697	0.147	9.51	0.736	0.751	95.1
NIE OR	Conventional	1.380	1.434	0.054	3.91	0.399	0.402	94.6
	Penalized		1.410	0.031	2.21	0.373	0.375	94.8
TE OR	Conventional	2.139	2.265	0.126	5.89	0.816	0.826	95.9
	Penalized		2.241	0.102	4.79	0.761	0.768	96.2
Adjusted scenario 2, n = 150, estimated EPV = 5.38								
NDE OR	Conventional	1.539	1.799	0.260	16.92	1.090	1.120	94.4
	Penalized		1.782	0.243	15.79	0.986	1.015	95.9
NIE OR	Conventional	1.376	1.506	0.130	9.47	0.611	0.625	94.3
	Penalized		1.437	0.061	4.44	0.520	0.524	95.7
TE OR	Conventional	2.118	2.379	0.261	12.35	1.178	1.207	95.5
	Penalized		2.298	0.180	8.52	1.040	1.055	96.1
Adjusted scenario 2, n = 250, estimated EPV = 8.97								
NDE OR	Conventional	1.539	1.736	0.197	12.77	0.810	0.833	93.6
	Penalized		1.734	0.194	12.63	0.768	0.792	94.8
NIE OR	Conventional	1.376	1.443	0.067	4.89	0.434	0.439	94.4
	Penalized		1.408	0.032	2.35	0.398	0.399	95.5
TE OR	Conventional	2.118	2.316	0.198	9.37	0.864	0.886	94.4
	Penalized		2.275	0.158	7.44	0.806	0.821	95.3
Adjusted scenario 2, n = 500, estimated EPV = 17.94								
NDE OR	Conventional	1.539	1.617	0.077	5.03	0.509	0.514	94.4
	Penalized		1.621	0.082	5.32	0.497	0.503	94.7
NIE OR	Conventional	1.376	1.398	0.022	1.62	0.260	0.261	96.3
	Penalized		1.383	0.007	0.50	0.249	0.249	96.8
TE OR	Conventional	2.118	2.184	0.067	3.14	0.539	0.543	95.9
	Penalized		2.170	0.052	2.47	0.522	0.525	96.1

Abbreviations: CP, coverage probability; EPV, number of events per variable; NDE, natural direct effect; NIE, natural indirect effect; OR, odds ratio; RMSE, root mean squared error; SD, standard deviation; TE, total effect.

TABLE G2 Adjusted simulation study with Firth's penalization: Risk ratio scale (1000 samples of sizes $n = 150,250, 500$)

Effect	Estimation method	True value	Mean	Bias	Relative bias (%)	SD	RMSE	Delta CP (%)
Adjusted scenario 1, n = 150, estimated EPV = 2.29								
NDE RR	Penalized	1.503	1.836	0.334	22.23	1.504	1.541	96.1
NIE RR	Penalized	1.336	1.464	0.128	9.60	0.726	0.737	94.5
TE RR	Penalized	2.007	2.207	0.200	9.96	1.375	1.390	97.2
Adjusted scenario 1, n = 250, estimated EPV = 3.82								
NDE RR	Conventional	1.503	1.801	0.298	19.83	1.170	1.207	94.6
	Penalized		1.755	0.252	16.79	0.985	1.017	95.4
NIE RR	Conventional	1.336	1.477	0.141	10.56	0.641	0.656	93.8
	Penalized		1.408	0.073	5.43	0.527	0.532	94.4
TE RR	Conventional	2.007	2.312	0.305	15.19	1.259	1.295	96.1
	Penalized		2.205	0.198	9.86	1.005	1.024	96.8
Adjusted scenario 1, n = 500, estimated EPV = 7.64								
NDE RR	Conventional	1.503	1.618	0.116	7.70	0.674	0.684	94.9
	Penalized		1.616	0.113	7.53	0.636	0.646	95.1
NIE RR	Conventional	1.336	1.385	0.050	3.71	0.357	0.360	94.3
	Penalized		1.362	0.026	1.94	0.332	0.333	94.4
TE RR	Conventional	2.007	2.109	0.101	5.05	0.700	0.703	95.9
	Penalized		2.080	0.073	3.64	0.643	0.647	96.4
Adjusted scenario 2, n = 150, estimated EPV = 5.38								
NDE RR	Conventional	1.430	1.571	0.141	9.87	0.763	0.776	94.2
	Penalized		1.552	0.122	8.56	0.678	0.689	95.9
NIE RR	Conventional	1.279	1.378	0.100	7.83	0.469	0.480	93.0
	Penalized		1.321	0.043	3.33	0.390	0.393	94.8
TE RR	Conventional	1.828	1.974	0.146	7.98	0.772	0.785	95.6
	Penalized		1.905	0.076	4.17	0.669	0.673	96.4
Adjusted scenario 2, n = 250, estimated EPV = 8.97								
NDE RR	Conventional	1.430	1.545	0.115	8.07	0.579	0.590	94.1
	Penalized		1.539	0.109	7.63	0.542	0.553	94.5
NIE RR	Conventional	1.279	1.330	0.051	4.02	0.328	0.332	93.2
	Penalized		1.301	0.022	1.74	0.298	0.298	94.6
TE RR	Conventional	1.828	1.947	0.118	6.48	0.576	0.588	95.0
	Penalized		1.909	0.081	4.41	0.531	0.537	96.2
Adjusted scenario 2, n = 500, estimated EPV = 17.94								
NDE RR	Conventional	1.430	1.474	0.044	3.09	0.373	0.375	94.6
	Penalized		1.475	0.045	3.16	0.362	0.364	94.7
NIE RR	Conventional	1.279	1.297	0.018	1.42	0.196	0.197	95.8
	Penalized		1.284	0.005	0.40	0.187	0.187	96.3
TE RR	Conventional	1.828	1.867	0.039	2.13	0.367	0.369	96.1
	Penalized		1.853	0.024	1.33	0.353	0.354	96.0

Abbreviations: CP, coverage probability; EPV, number of events per variable; NDE, natural direct effect; NIE, natural indirect effect; RMSE, root mean squared error; RR, risk ratio; SD, standard deviation; TE, total effect.

TABLE G3 Adjusted simulation study with Firth's penalization: Risk difference scale (1000 samples of sizes $n = 150,250, 500$)

Effect	Estimation method	True value	Mean	Bias	Relative bias (%)	SD	RMSE	Delta CP (%)
Adjusted scenario 1, n = 150, estimated EPV = 2.29								
NDE RD	Penalized	0.0289	0.0384	0.0095	32.98	0.0624	0.0632	93.7
NIE RD	Penalized	0.0290	0.0226	−0.0064	−22.22	0.0441	0.0446	98.2
TE RD	Penalized	0.0579	0.0610	0.0031	5.32	0.0520	0.0521	94.3
Adjusted scenario 1, n = 250, estimated EPV = 3.82								
NDE RD	Conventional	0.0289	0.0322	0.0034	11.66	0.0435	0.0436	92.3
	Penalized		0.0370	0.0081	27.97	0.0449	0.0457	94.6
NIE RD	Conventional	0.0290	0.0245	−0.0045	−15.67	0.0311	0.0315	98.3
	Penalized		0.0253	−0.0037	−12.73	0.0317	0.0319	98.3
TE RD	Conventional	0.0579	0.0567	−0.0012	−2.03	0.0380	0.0380	93.5
	Penalized		0.0623	0.0044	7.58	0.0381	0.0384	95.7
Adjusted scenario 1, n = 500, estimated EPV = 7.64								
NDE RD	Conventional	0.0289	0.0301	0.0012	4.23	0.0310	0.0310	93.3
	Penalized		0.0325	0.0037	12.64	0.0316	0.0318	94.3
NIE RD	Conventional	0.0290	0.0261	−0.0029	−9.94	0.0210	0.0212	97.9
	Penalized		0.0266	−0.0024	−8.27	0.0213	0.0214	97.9
TE RD	Conventional	0.0579	0.0562	−0.0017	−2.87	0.0276	0.0277	94.5
	Penalized		0.0591	0.0013	2.17	0.0276	0.0277	95.8
Adjusted scenario 2, n = 150, estimated EPV = 5.38								
NDE RD	Conventional	0.0608	0.0643	0.0034	5.62	0.0845	0.0846	92.6
	Penalized		0.0691	0.0083	13.57	0.0827	0.0832	94.2
NIE RD	Conventional	0.0564	0.0519	−0.0045	−7.99	0.0566	0.0568	96.3
	Penalized		0.0495	−0.0069	−12.29	0.0541	0.0545	96.9
TE RD	Conventional	0.1172	0.1161	−0.0011	−0.92	0.0742	0.0742	94.3
	Penalized		0.1185	0.0013	1.13	0.0715	0.0715	95.5
Adjusted scenario 2, n = 250, estimated EPV = 8.97								
NDE RD	Conventional	0.0608	0.0666	0.0057	9.40	0.0662	0.0664	92.6
	Penalized		0.0697	0.0089	14.59	0.0654	0.0660	93.7
NIE RD	Conventional	0.0564	0.0529	−0.0034	−6.09	0.0442	0.0444	96.2
	Penalized		0.0513	−0.0051	−9.01	0.0431	0.0434	96.9
TE RD	Conventional	0.1172	0.1195	0.0023	1.95	0.0565	0.0565	94.4
	Penalized		0.1210	0.0038	3.24	0.0553	0.0554	95.2
Adjusted scenario 2, n = 500, estimated EPV = 17.94								
NDE RD	Conventional	0.0608	0.0624	0.0016	2.57	0.0457	0.0458	94.2
	Penalized		0.0642	0.0033	5.48	0.0455	0.0456	94.7
NIE RD	Conventional	0.0564	0.0539	−0.0025	−4.37	0.0289	0.0290	96.7
	Penalized		0.0530	−0.0034	−5.96	0.0285	0.0287	97.0
TE RD	Conventional	0.1172	0.1163	−0.0009	−0.77	0.0390	0.0390	95.0
	Penalized		0.1172	−0.0000	−0.02	0.0386	0.0386	95.4

Abbreviations: CP, coverage probability; EPV, number of events per variable; NDE, natural direct effect; NIE, natural indirect effect; RD, risk difference; RMSE, root mean squared error; SD, standard deviation; TE, total effect.

APPENDIX H. RESULTS OF THE SIMULATION STUDY WITH OMITTED EXPOSURE-MEDIATOR INTERACTION TERM

Table H1 presents the results of the simulation study conducted to examine the impact of incorrectly omitting the exposure-mediator interaction term in the fitted outcome logistic regression model.

TABLE H1 Adjusted simulation study with omitted exposure-mediator interaction term (odds ratio scale; 1000 samples of size $n = 5000$)

Effect	True value	Mean	Bias	Relative bias (%)	SD	RMSE	Delta CP (%)	Bootstrap CP (%)[a]
Adjusted scenario 1, $\theta_3 = 0.15$								
NDE OR	1.550	1.630	0.080	5.18	0.200	0.216	93.4	92.3
NIE OR	1.380	1.329	−0.051	−3.67	0.071	0.087	89.3	89.4
TE OR	2.139	2.161	0.022	1.04	0.241	0.242	95.2	94.9
Adjusted scenario 3, $\theta_3 = 0.15$								
NDE OR	1.515	1.594	0.079	5.25	0.106	0.132	90.1	89.8
NIE OR	1.373	1.307	−0.066	−4.80	0.038	0.076	59.7	60.9
TE OR	2.080	2.082	0.002	0.12	0.125	0.125	95.9	96.5
Adjusted scenario 5, $\theta_3 = 0.15$								
NDE OR	1.495	1.606	0.111	7.43	0.216	0.243	93.0	92.1
NIE OR	1.381	1.305	−0.076	−5.53	0.066	0.101	78.5	78.7
TE OR	2.065	2.091	0.026	1.24	0.263	0.265	94.5	95.0
Adjusted scenario 1, $\theta_3 = 0.30$								
NDE OR	1.620	1.772	0.152	9.39	0.208	0.258	89.3	88.3
NIE OR	1.483	1.377	−0.106	−7.12	0.072	0.128	70.9	71.8
TE OR	2.401	2.434	0.032	1.35	0.262	0.264	94.5	94.7
Adjusted scenario 3, $\theta_3 = 0.30$								
NDE OR	1.546	1.706	0.160	10.36	0.114	0.197	72.2	72.0
NIE OR	1.470	1.334	−0.136	−9.26	0.038	0.141	8.6	9.5
TE OR	2.273	2.275	0.002	0.07	0.138	0.138	96.4	96.5
Adjusted scenario 5, $\theta_3 = 0.30$								
NDE OR	1.495	1.699	0.205	13.69	0.235	0.312	87.8	85.8
NIE OR	1.486	1.321	−0.165	−11.10	0.068	0.178	35.5	37.1
TE OR	2.221	2.239	0.018	0.81	0.289	0.290	95.2	94.7

Abbreviations: CP, coverage probability; NDE, natural direct effect; NIE, natural indirect effect; OR, odds ratio; RMSE, root mean squared error; SD, standard deviation; TE, total effect.

[a]Percentile bootstrap[26] based on 500 resamples with replacement.

APPENDIX I. RESULTS OF THE SIMULATION STUDY WITH A NON-NORMAL MEDIATOR ERROR TERM

Tables I1-I2 present the results of the simulation study performed to examine the impact of incorrectly assuming that the mediator is normally distributed when it is not.

TABLE I1 Adjusted simulation study with a non-normal mediator: *Case 1*, generalized *t*-distribution for the error term (odds ratio scale; 1000 samples of size $n = 5000$)

Effect	True value	Mean	Bias	Relative bias (%)	SD	RMSE	Delta CP (%)	Bootstrap CP (%)[a]
Adjusted scenario 1								
NDE OR	1.546	1.563	0.017	1.10	0.210	0.211	95.3	94.9
NIE OR	1.379	1.388	0.009	0.64	0.114	0.115	93.6	92.9
TE OR	2.131	2.155	0.023	1.09	0.236	0.238	94.9	94.7
Adjusted scenario 3								
NDE OR	1.516	1.524	0.008	0.50	0.127	0.127	95.3	95.2
NIE OR	1.375	1.376	0.001	0.08	0.076	0.076	95.3	94.8
TE OR	2.085	2.091	0.006	0.29	0.137	0.137	93.4	94.1
Adjusted scenario 5								
NDE OR	1.497	1.531	0.034	2.27	0.261	0.263	93.7	93.8
NIE OR	1.380	1.382	0.002	0.14	0.143	0.143	94.7	94.7
TE OR	2.066	2.094	0.028	1.36	0.284	0.285	93.8	93.4

Abbreviations: CP, coverage probability; NDE, natural direct effect; NIE, natural indirect effect; OR, odds ratio; RMSE, root mean squared error; SD, standard deviation; TE, total effect.

[a]Percentile bootstrap[26] based on 500 resamples with replacement.

TABLE I2 Adjusted simulation study with a non-normal mediator: *Case 2*, gamma distribution for the error term (odds ratio scale; 1000 samples of size $n = 5000$)

Effect	True value	Mean	Bias	Relative bias (%)	SD	RMSE	Delta CP (%)	Bootstrap CP (%)[a]
Gamma (shape = 1.1025, scale = 0.4762), skewness = 1.9048								
Adjusted scenario 1								
NDE OR	1.552	1.566	0.014	0.88	0.199	0.200	94.8	94.7
NIE OR	1.377	1.383	0.006	0.40	0.088	0.088	94.6	94.4
TE OR	2.137	2.156	0.019	0.87	0.236	0.237	95.0	94.7
Adjusted scenario 3								
NDE OR	1.511	1.516	0.005	0.34	0.127	0.127	93.7	93.3
NIE OR	1.374	1.377	0.003	0.23	0.075	0.075	94.9	94.4
TE OR	2.077	2.083	0.006	0.29	0.136	0.136	93.8	93.5
Adjusted scenario 5								
NDE OR	1.501	1.505	0.004	0.30	0.279	0.279	94.9	94.8
NIE OR	1.382	1.410	0.028	2.02	0.200	0.201	94.5	93.6
TE OR	2.074	2.084	0.010	0.48	0.272	0.272	94.3	94.2
Gamma (shape = 2, scale = 0.3536), skewness = 1.4142								
Adjusted scenario 1								
NDE OR	1.552	1.565	0.013	0.86	0.201	0.201	95.1	95.0
NIE OR	1.378	1.384	0.007	0.48	0.093	0.093	95.7	95.4
TE OR	2.138	2.158	0.020	0.92	0.238	0.239	94.9	94.2
Adjusted scenario 2								
NDE OR	1.512	1.521	0.009	0.57	0.120	0.121	94.7	94.4
NIE OR	1.374	1.376	0.003	0.19	0.072	0.072	94.4	94.2
TE OR	2.077	2.088	0.011	0.51	0.135	0.136	94.0	94.0
Adjusted scenario 3								
NDE OR	1.499	1.521	0.022	1.46	0.267	0.268	95.5	95.0
NIE OR	1.382	1.399	0.017	1.21	0.185	0.186	93.8	93.8
TE OR	2.072	2.095	0.023	1.11	0.273	0.274	95.1	94.9

Abbreviations: CP, coverage probability; NDE, natural direct effect; NIE, natural indirect effect; OR, odds ratio; RMSE, root mean squared error; SD, standard deviation; TE, total effect.

[a]Percentile bootstrap[26] based on 500 resamples with replacement.

APPENDIX J. COMMENTS ON THE SAS MACRO EXECUTION

Use of our SAS macro `bin_cont_exactmed` developed to perform exact mediation analysis for a binary outcome and a continuous mediator (see Supporting Information section) requires the specification of macro variables. We provide some examples to illustrate how to specify values for these variables.

By default, our SAS macro reports natural effects on the OR, RR, and RD scales simultaneously; the point estimates are accompanied by 95% CIs based on the delta method. For example, the following statement returns crude (unadjusted) NDE, NIE, and TE estimates on the OR, RR, and RD scales for a change in the exposure (binary or continuous) from level t_0 to level t_1 assuming there is no exposure-mediator interaction and using the delta method to construct 95% CIs:

```
%bin_cont_exactmed(mydata=data, A=exposure, M=mediator, Y=outcome, interaction=0,
adjusted=0, a1=t1, a0=t0)
```

To perform an adjusted and penalized (ie, based on Firth's penalization of the outcome model) mediation analysis allowing for an exposure-mediator interaction and using the percentile bootstrap based on 5000 resamples with initial random seed = 1234 to construct 95% CIs (in addition to the default 95% CIs based on the delta method), our SAS macro `bin_cont_exactmed` should be executed as follows:

```
%bin_cont_exactmed(mydata=data, A=exposure, M=mediator, Y=outcome, interaction=1,
adjusted=1, cvar_M=Mvar1 Mvar2 ... Mvarp, cvar_Y=Yvar1 Yvar2 ... Yvars, a1=t1, a0=t0, boot=1,
bootseed=1234, nboot=5000, Firth=1)
```

where $\texttt{Mvar}_1$ $\texttt{Mvar}_2$ … $\texttt{Mvar}_p$ and $\texttt{Yvar}_1$ $\texttt{Yvar}_2$ … $\texttt{Yvar}_s$ are the sets of adjustment covariates for the mediator and outcome models, respectively.

If the user specifies `cde=1`, $\texttt{mcontrol=m}_c$, our SAS macro `bin_cont_exactmed` additionally returns the controlled direct effect estimated at the level m_c of the mediator.

By default, our SAS macro reports natural and controlled direct effects evaluated at the sample-specific mean values of the adjustment covariates. In order to estimate mediation effects at specific values of some covariates (that is, stratum-specific effects), the user needs to provide SAS datasets `DATA_M` and `DATA_Y` containing those values **before** executing the SAS macro `bin_cont_exactmed`. For example, to estimate mediation effects corresponding to $\texttt{Mvar}_1\texttt{=Cm}_1$, $\texttt{Mvar}_2\texttt{=Cm}_2$, $\texttt{Mvar}_3\texttt{=Cm}_3$ (ie, at user-defined values for the first three adjustment covariates in the mediator model), and $\texttt{Yvar}_3\texttt{=Cy}_3$, $\texttt{Yvar}_4\texttt{=Cy}_4$ (ie, at user-defined values for the third and fourth covariates in the outcome model), datasets `DATA_M` and `DATA_Y` can be constructed using SAS `datalines` statements as follows:

```
data DATA_M; input Mvar1 Mvar2 Mvar3; datalines;
Cm1 Cm2 Cm3
;
data DATA_Y; input Yvar3 Yvar4; datalines;
Cy3 Cy4
;
```

Hence, the following statement returns natural and controlled direct effects evaluated at the covariate values specified in `DATA_M` and `DATA_Y`, assuming an exposure-mediator interaction, requiring to evaluate the controlled direct effect at the mediator level m_c, and using the delta method to construct 95% CIs:

```
%bin_cont_exactmed(mydata=data, A=exposure, M=mediator, Y=outcome, interaction=1,
adjusted=1, cvar_M=Mvar1 Mvar2 ... Mvarp, cvar_Y=Yvar1 Yvar2 ... Yvars, a1=t1, a0=t0,
Firth=0, stratum=1, cvar_M_data=DATA_M, cvar_Y_data=DATA_Y, cde=1, mcontrol=mc)
```

Common adjustment covariates in `DATA_M` and `DATA_Y` must have the same values; otherwise, the macro execution will be aborted and a warning will be displayed in the SAS log. Moreover, the list of variables with unequal values will be shown in the SAS Results Viewer window.

If the covariates specified in DATA_M (DATA_Y) constitute some proper subset of {$Mvar_1$, $Mvar_2$, …, $Mvar_p$} ({$Yvar_1$, $Yvar_2$, …, $Yvar_s$}), then the other covariates will be set to their sample-specific mean levels.

If, for example, $Mvar_1$, $Mvar_2$, are two dummy variables coding some categorical covariate V_{cat} with three levels, we can estimate mediation effects at the reference level by constructing DATA_M as follows:

```
data DATA_M; input Mvar1 Mvar2;datalines;
0 0
;
```

In order to estimate mediation effects corresponding to the second level of V_{cat}, the user has to provide DATA_M as

```
data DATA_M; input Mvar1 Mvar2;datalines;
1 0
;
```

Finally, to estimate mediation effects corresponding to the third level of V_{cat}, DATA_M should be provided as

```
data DATA_M; input Mvar1 Mvar2;datalines;
0 1
;
```

The same strategy can be applied to the construction of DATA_Y.

In some cases, the user can obtain an error message informing that numerical convergence is not attained (eg, when the integrand values in Equation (7) and/or Equations (B2)-(B10) are close to zero on $(-\infty, \infty)$, possibly except on some small intervals). To overcome these problems, the user can change the default value (ie, 1) of the scale parameter in the QUAD subroutine. For example, the following statement returns crude natural effects estimates for a change in the exposure from t_0 to t_1, assuming an exposure-mediator interaction and using the delta method to construct 95% CIs; the default scale value in the QUAD subroutine to calculate the nested counterfactual outcome probabilities based on Equation (7) is replaced by 0.001 by specifying pscale = 1 (specification required to replace the default value of the scale parameter by some user-defined value) and pscalevalue = 0.001 (user-defined value):

```
%bin_cont_exactmed(mydata=data, A=exposure, M=mediator, Y=outcome, interaction=1,
adjusted=0, a1=t1,a0=t0,pscale=1,pscalevalue=0.001)
```

In order to also replace the default scale value by 0.0001 when calculating the integrals needed for the delta method (Equations B2-B10), our SAS macro bin_cont_exactmed should be executed as follows:

```
%bin_cont_exactmed(mydata=data, A=exposure, M=mediator, Y=outcome, interaction=1,
adjusted=0, a1=t1,a0=t0,pscale=1,pscalevalue=0.001,dpscale=1,dscalevalue=0.0001)
```

The recommended values for pscalevalue and dscalevalue are 0.01, 0.001, 0.0001 and so forth.