UNIVERSITÉ DU QUÉBEC À MONTRÉAL

APPROCHES DE MÉDIATION BASÉES SUR LA RÉGRESSION AVEC RÉPONSE BINAIRE : CONTOURNER L'HYPOTHÈSE DE LA RÉPONSE RARE OU COMMUNE

THÈSE

PRÉSENTÉE

COMME EXIGENCE PARTIELLE

DU DOCTORAT EN MATHÉMATIQUES

PAR

MARIIA SAMOILENKO

SEPTEMBRE 2023

REMERCIEMENTS

Je tiens à exprimer une sincère gratitude à ma directrice de recherche Geneviève Lefebvre pour m'avoir encouragée à commencer mes études doctorales, pour son soutien et son encadrement hors pair tout au long de mon doctorat ainsi que pour m'avoir communiqué sa passion pour la recherche en statistique causale.

Je tiens aussi à exprimer mes remerciements aux professeurs du Département de mathématiques de l'UQÀM pour leurs contributions inestimables à ma formation en statistique.

J'aimerais exprimer ma profonde reconnaissance à Miguel Caubet Fernandez pour sa collaboration et pour toute son aide apportée à mon projet de recherche. Je remercie également Jill Vandermeerschen et Jesse Gervais pour nos discussions sur de nombreux sujets statistiques.

J'aimerais remercier mon fils Dmytro pour son soutien moral.

Finalement, je voudrais remercier profondément le Conseil de recherches en sciences naturelles et en génie du Canada, le Fonds de recherche du Québec - Nature et technologies ainsi que le Centre de recherche en statistique et science des données STATQAM de la Faculté des sciences de l'UQÀM pour leurs soutiens financiers.

TABLE DES MATIÈRES

LISTE DES TABLEAUX vi
LISTE DES FIGURES ix
ACRONYMES xi
RÉSUMÉ xii
INTRODUCTION 1
CHAPITRE I APPROCHE CONTREFACTUELLE DE L'ANALYSE DE MÉDIATION . 9
1.1 Cadre contrefactuel de l'inférence causale 9
1.2 Définition de l'effet causal moyen populationnel sous le cadre contrefactuel 11
1.3 Identification de l'effet causal moyen populationnel sous le cadre contrefactuel 11
1.3.1 Confusion dans le contexte des études observationnelles 12
1.3.2 Application des hypothèses d'identification pour établir un lien entre les réponses contrefactuelles et les données observées 19
1.4 Cadre contrefactuel dans la médiation causale 20
1.4.1 Définition des effets direct et indirect dans le cadre de l'approche contrefactuelle 21
1.4.2 Identification des effets naturels direct et indirect dans les études observationnelles 23
1.4.3 Décomposition de l'effet total 26
1.4.4 Effet direct contrôlé et son identification 27
CHAPITRE II MÉDIATION CAUSALE POUR RÉPONSE BINAIRE : TECHNIQUES D'ESTIMATION 28
2.1 Techniques d'estimation approximatives pour un médiateur continu et une réponse binaire 28
2.1.1 Approche de VanderWeele et Vansteelandt (2010) 28
2.1.2 Approche de Gaynor *et al.* (2019) 34
2.2 Approche d'estimation approximative de Valeri et VanderWeele (2013) pour un médiateur et une réponse binaires 37
2.3 Approche d'estimation d'Imai *et al.* (2010) 41
2.4 Analyse de médiation causale par des modèles à effets naturels 44
CHAPITRE III RÉPONSE ET MÉDIATEUR BINAIRES : COMPARAISON DES APPROCHES PAR MODÈLES DE RÉGRESSION LOGISTIQUE ET LOG-BINOMIAL 48
3.1 Expressions pour les effets naturels direct et indirect sur l'échelle du rapport de risques . 48

3.2 ARTICLE 1. Point: Risk ratio equations for natural direct and indirect effects in causal mediation analysis of a binary mediator and a binary outcome – A fresh look at the formulas . . . 53

3.2.1 Appendix 1: Correct natural direct and indirect effects' risk ratio expressions . . 57

3.2.2 Appendix 2: Comparison between natural direct and indirect effects' risk ratio expressions with logistic and log-binomial mediator models . . . 61

3.2.3 Appendix 3: Comparison between logistic and log-binomial probabilities . . . 65

3.3 Hypothèse de la réponse rare dans les analyses de médiation causale . . . 72

CHAPITRE IV ARTICLE 2. PARAMETRIC-REGRESSION—BASED CAUSAL MEDIATION ANALYSIS OF BINARY OUTCOMES AND BINARY MEDIATORS : MOVING BEYOND THE RARENESS OR COMMONNESS OF THE OUTCOME . . . 73

4.1 Introduction . . . 74

4.2 Methods . . . 76

4.2.1 Models and nested counterfactual outcome probabilities . . . 76

4.2.2 Natural direct and indirect effects on the OR, RR and RD scales . . . 77

4.2.3 Valeri and VanderWeele (2013) approximate NDE and NIE approach . . . 78

4.2.4 Simulation studies . . . 79

4.3 Results . . . 83

4.4 Real-data example . . . 93

4.5 Discussion . . . 98

4.6 Appendix 1 . . . 99

4.6.1 General formulas used in the delta method for the exact regression-based mediation effects . . . 99

4.6.2 Nested probability formulas based on the mediator and outcome logistic models . 100

4.6.3 Delta method for mediation exact regression-based odds ratios . . . 102

4.6.4 Delta method for exact mediation regression-based risk ratios . . . 104

4.6.5 Delta method for exact mediation regression-based risk differences . . . 105

4.6.6 Mediation controlled direct effects . . . 106

4.6.7 Decomposition property of the exact total effect estimator . . . 109

4.6.8 Comments on estimation procedures . . . 113

4.7 Appendix 2 . . . 114

4.8 Appendix 3 . . . 118

CHAPITRE V ARTICLE 3. AN EXACT REGRESSION-BASED APPROACH FOR THE ESTIMATION OF NATURAL DIRECT AND INDIRECT EFFECTS WITH A BINARY OUTCOME AND A CONTINUOUS MEDIATOR 130
5.1 Introduction 131
5.2 Methods 134
5.2.1 Models and nested counterfactual outcome probabilities 134
5.2.2 Simulation studies 136
5.3 Results 142
5.3.1 Results of the main simulation studies 142
5.3.2 Results of the simulation study with a marginal but not conditional rare outcome 143
5.3.3 Results of the simulation study with Firth's penalization 149
5.3.4 Results of the simulation study with omitted exposure-mediator interaction term 149
5.3.5 Results of the simulation study with a non-normal mediator error term 150
5.4 Real data example 150
5.5 Discussion 151
5.6 Appendices 156
5.6.1 Identification assumptions 156
5.6.2 Delta method for exact mediation approach 157
5.6.3 Convergence of improper integrals in the exact approach 163
5.6.4 Results of the crude simulation study 164
5.6.5 Estimation of the parameter s in the approach by Gaynor et al. 169
5.6.6 Results of the simulation study with marginally but not conditionally rare outcome 169
5.6.7 Results of the simulation study with Firth's penalization 169
5.6.8 Results of the simulation study with omitted exposure-mediator interaction term 169
5.6.9 Results of the simulation study with a non-normal mediator error term 169
5.6.10 Comments on the SAS macro execution 177
5.6.11 SAS macro `bin_cont_exactmed` 180
CHAPITRE VI IMPLÉMENTATION DE L'APPROCHE EXACTE D'ANALYSE DE MÉDIATION CAUSALE POUR UNE RÉPONSE BINAIRE EN SAS ET R 194
CONCLUSION 196
ANNEXE A DISTRIBUTION LOGIT NORMALE 202
BIBLIOGRAPHIE 204

LISTE DES TABLEAUX

Tableau Page

1.1 Exemple hypothétique sur la réussite professionnelle en fonction de l'éducation : espérances des réponses contrefactuelles 17

4.1 Data-generating mechanisms for a simulation study without covariates conducted to evaluate proposed exact estimators 79

4.2 Data-generating mechanisms for the simulation study with covariates: Outcome simulation parameters 82

4.3 Exact and approximate natural-effects estimators on the odds ratio scale in scenarios with increasing levels of outcome commonness (simulation study without covariates based on 1000 independent samples of size $n = 5000$) 85

4.4 Exact and approximate natural-effects estimators on the risk ratio scale in scenarios with increasing levels of outcome commonness (simulation study without covariates based on 1000 independent samples of size $n = 5000$) 86

4.5 Natural-effects estimators on the risk difference scale in scenarios with increasing levels of outcome commonness (simulation study without covariates based on 1000 independent samples of size $n = 5000$) 87

4.6 Natural-effects estimators on the odds ratio scale in scenarios with increasing levels of outcome commonness (simulation study with covariates based on 1000 independent samples of size $n = 5000$) 88

4.6 Natural-effects estimators on the odds ratio scale in scenarios with increasing levels of outcome commonness (simulation study with covariates based on 1000 independent samples of size $n = 5000$; continuation) 89

4.7 Natural-effects estimators on the risk ratio scale in scenarios with increasing levels of outcome commonness (simulation study with covariates based on 1000 independent samples of size $n = 5000$) 90

4.7 Natural-effects estimators on the risk ratio scale in scenarios with increasing levels of outcome commonness (simulation study with covariates based on 1000 independent samples of size $n = 5000$; continuation) 91

4.8 Natural-effects estimators on the risk difference scale in scenarios with increasing levels of outcome commonness (simulation study with covariates based on 1000 independent samples of size $n = 5000$) 92

4.8 Natural-effects estimators on the risk difference scale in scenarios with increasing levels of outcome commonness (simulation study with covariates based on 1000 independent samples of size $n = 5000$; continuation) 93

4.9 Comparison between natural direct and indirect effect estimates on the odds ratio, risk ratio, and risk difference scales obtained from the exact estimator and existing estimators available in various software packages (real-data example) 95

4.10 Simulation study without covariates: Total effect multiplicative decomposition property of exact and approximate natural-effects estimators by scenarios with increasing levels of outcome commonness 112

4.11 Simulation study with covariates: Total effect multiplicative decomposition property of exact and approximate natural-effects estimators by scenarios with increasing levels of outcome commonness 113

5.1 Adjusted simulation study: odds ratio scale (1000 samples of size $n = 5000$) 145

5.1 Adjusted simulation study: odds ratio scale (1000 samples of size $n = 5000$; continuation) 146

5.2 Adjusted simulation study: risk ratio scale (1000 samples of size $n = 5000$) 147

5.3 Adjusted simulation study: risk difference scale (1000 samples of size $n = 5000$) . . . 148

5.4 Real data example with placental abruption as the exposure, gestational age as the mediator and low birth weight as the outcome ($n = 6197$) 152

5.5 Crude simulation study: odds ratio scale (1000 samples of size $n = 5000$) 165

5.5 Crude simulation study: odds ratio scale (1000 samples of size $n = 5000$; continuation) 166

5.6 Crude simulation study: risk ratio scale (1000 samples of size $n = 5000$) 167

5.7 Crude simulation study: risk difference scale (1000 samples of size $n = 5000$) 168

5.8 Distribution of the parameter s estimates in the approach by Gaynor *et al.* (2019) . . 169

5.9 Adjusted simulation study where the outcome is rare marginally but not conditionally (1000 samples of size $n = 5000$) 170

5.10 Adjusted simulation study with Firth's penalization: odds ratio scale (1000 samples of sizes $n = 150, 250, 500$) 171

5.11 Adjusted simulation study with Firth's penalization: risk ratio scale (1000 samples of sizes $n = 150, 250, 500$) 172

5.12 Adjusted simulation study with Firth's penalization: risk difference scale (1000 samples of sizes $n = 150, 250, 500$) 173

5.13 Adjusted simulation study with omitted exposure-mediator interaction term (odds ratio scale; 1000 samples of size $n = 5000$) 174

5.14 Adjusted simulation study with a non-normal mediator: *Case 1*, generalized t-distribution for the error term (odds ratio scale; 1000 samples of size $n = 5000$) 175

5.15 Adjusted simulation study with a non-normal mediator: *Case 2*, gamma distribution for the error term (odds ratio scale; 1000 samples of size $n = 5000$) 176

LISTE DES FIGURES

Figure Page

0.1 Modèle de médiation simple avec l'exposition A, le médiateur M et la réponse Y . . 1

1.1 L'effet du traitement A sur la réponse Y en présence d'une variable confondante C . 15

1.2 Modèle de médiation avec l'exposition A, le médiateur M et la réponse Y en présence des variables confondantes des relations $A - Y$, $A - M$, et $M - Y$ 21

3.1 Comparing ratios of natural direct and indirect effects' risk ratio expressions derived from using a logistic model and a log-binomial model for the mediator (with a log-binomial outcome model) as a function of the outcome regression coefficients (θ_2, θ_3) for fixed scenarios of commonness of the mediator. 63

4.1 Comparison between natural direct effect (NDE), natural indirect effect (NIE), and total effect (TE) estimates on the odds ratio scale obtained from the exact estimator and existing estimators available in software (real-data example). A) Mediation analyses with use of inhaled corticosteroids as the exposure variable; B) mediation analyses with placental abruption as the exposure variable. The solid lines present 95% confidence intervals (CIs) obtained by the exact approach using the delta method. The dashed and dotted lines correspond to 95% CIs returned by the SAS PROC CAUSALMED procedure (via the delta method) and the R package `medflex` (via percentile bootstrap), respectively. The dotted-dashed line presents 95% CIs for the conventional (nonmediated) TE (CTE) by percentile bootstrap. The black circles show effect point estimates, and the white circles show the CI endpoints. 96

4.2 Exact natural direct effect (NDE), natural indirect effect (NIE), and total effect (TE) on the odds ratio (A) and risk difference (B) scales evaluated at particular levels of the adjustment covariates (real-data example with placental abruption as the exposure variable). Solid lines correspond to 95% confidence intervals (CIs) given the following set of covariate values: baby's sex = female, maternal age = 18–34 years, diabetes mellitus = no, and gestational diabetes = no. Dashed lines correspond to 95% CIs when the covariate values are specified as follows: baby's sex = male, maternal age <18 years, diabetes mellitus = no, and gestational diabetes = yes. The 95% CIs were constructed by percentile bootstrapping based on 1000 resamples with replacement. The black circles show effect point estimates, and the white circles show the CI endpoints. 97

5.1 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 $shape = 2$, $scale = 0.3636$ (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 $variance = 0.5^2$ 141

ACRONYMES

CDE effet direct contrôlé (*controlled direct effect*)

CI intervalle de confiance (*confidence interval*)

CP probabilité de couverture (*coverage probability*)

EPV nombre d'événements par variable (*number of events per variable*)

NDE effet naturel direct (*natural direct effect*)

NEM modèle à effets naturels (*natural effect model*)

NIE effet naturel indirect (*natural indirect effect*)

OR rapport de cotes (*odds ratio*)

RD différence de risques (*risk difference*)

ROA hypothèse de la réponse rare (*rare outcome assumption*)

RR rapport de risques (*risk ratio*)

TE effet total (*total effect*)

RÉSUMÉ

L'analyse de médiation causale désigne un ensemble de concepts théoriques et d'outils d'estimation conçus pour étudier dans quelle mesure l'effet d'une exposition sur une variable réponse se produit à travers des variables intermédiaires (médiateurs). Dans cette thèse, nous nous concentrons sur l'analyse de médiation causale qui porte sur un seul médiateur et nous cherchons à décomposer l'effet total d'une exposition selon les soi-disant effets naturels direct et indirect. Un nombre d'approches d'estimation paramétriques basées sur la régression logistique pour une réponse binaire ont été proposées dans la littérature pour l'estimation d'effets naturels. Ces approches ont invoqué lesdites hypothèses de la réponse rare ou commune (non rare) afin d'obtenir des expressions approximatives simples et fermées pour ces effets. Toutefois, dans les études appliquées, l'évaluation de l'hypothèse de la réponse rare représente un défi à cause de l'absence de lignes directrices explicites permettant de qualifier une réponse binaire comme rare dans le contexte de la médiation causale.

Dans cette thèse, nous présentons des estimateurs exacts des effets naturels pour une réponse binaire basés sur la régression. L'utilisation du terme « exact » est justifiée par le fait que nos estimateurs sont dérivés sans invoquer aucune hypothèse théorique simplificatrice, ce qui permet de surmonter la difficulté inhérente à l'évaluation de l'hypothèse de la réponse rare dans le cadre de la médiation causale d'une réponse binaire. Nos estimateurs sont développés pour des expositions et des médiateurs binaires et/ou continus, et ils accommodent trois échelles binaires standards, c'est-à-dire le rapport de cotes, le rapport de risques et la différence de risques. Notre approche exacte repose sur la spécification de modèles de régression logistique pour les réponses et médiateurs binaires et sur la régression linéaire pour les médiateurs continus. Des formules pour les erreurs standards basées sur la méthode delta sont proposées.

Nous avons évalué le comportement de nos estimateurs dans des études de simulation où la réponse était rare ou commune, y compris des scénarios où la réponse était rare marginalement, mais pas conditionnellement. Dans nos études de simulation, basées sur une taille échantillonnale relativement grande, la performance adéquate des estimateurs exacts proposés a été observée, indépendamment de l'échelle des effets et de la prévalence (marginale ou conditionnelle) de la réponse. Plus précisément, nous avons obtenu des valeurs faibles de biais relatif, ce qui suggère que nos estimateurs exacts sont sans biais pour des tailles échantillonnales suffisamment grandes. Les intervalles de confiance par la méthode delta ou par le bootstrap basé sur les percentiles ont démontré des probabilités de couverture proches du niveau nominal.

Notre contribution pratique consiste au développement de macros SAS conçues pour implémenter l'approche exacte proposée. Ces macros fournissent les estimations des effets naturels sur les trois échelles binaires standards facilitant ainsi une comparaison directe avec les résultats obtenus par d'autres approches de médiation causale pour une réponse binaire.

Mots-clés : analyse de médiation causale simple basée sur la régression, effets naturels direct et indirect, estimateurs exacts, hypothèse de la réponse rare, régression logistique, variable réponse binaire

INTRODUCTION

Dans un large éventail de domaines scientifiques, l'objectif principal de nombreuses recherches appliquées consiste à établir l'existence d'une relation de cause à effet entre une exposition A et une variable réponse Y étudiées. Toutefois, lorsque l'hypothèse sur une relation causale entre A et Y est évaluée, nous observons souvent un processus de raffinement des questions de recherche, particulièrement sur les mécanismes sous-jacents responsables de cette relation (Lange *et al.*, 2017). L'analyse de médiation désigne un ensemble de concepts théoriques et de techniques analytiques statistiques conçus pour aider les chercheurs appliqués à élucider le processus de transmission de l'effet d'une exposition A sur une variable réponse Y par un ou plusieurs chemins causaux (Lash *et al.*, 2021). Plus précisément, l'analyse de médiation examine dans quelle mesure l'effet de A sur Y se produit à travers des variables intermédiaires (également appelées *médiateurs*) qui se trouvent sur les chemins causaux entre A et Y. Ainsi, l'objectif principal de l'analyse de médiation est de décomposer l'effet total de l'exposition A sur la réponse Y comme une combinaison de l'effet indirect (c'est-à-dire la contribution d'un ensemble donné de médiateurs à la transmission de l'effet de A sur Y) et de l'effet direct (l'effet de A sur Y qui n'est pas expliqué par ces médiateurs) (Richiardi *et al.*, 2013; Zugna *et al.*, 2022). Lorsque l'analyse de médiation porte sur un seul médiateur M, les termes « modèle de médiation simple » ou « analyse de médiation simple » (*simple mediation model*, *simple mediation analysis*) sont utilisés (Hayes et Little, 2018; Preacher et Hayes, 2004; Rijnhart *et al.*, 2019; VanderWeele, 2015). La Figure 0.1 illustre un modèle de médiation simple où le chemin $A \to Y$ correspond à l'effet direct de l'exposition A sur la réponse Y, tandis que l'effet indirect de A sur Y via le médiateur M est reflété par le chemin $A \to M \to Y$.

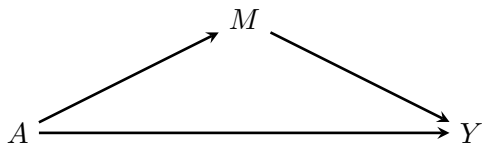


FIGURE 0.1 Modèle de médiation simple avec l'exposition A, le médiateur M et la réponse Y

La pratique de l'analyse de la médiation a été fortement influencée par l'article séminal de Baron et Kenny (1986) qui est généralement considéré comme l'origine de ladite *approche traditionnelle* de l'analyse de médiation. Cette approche s'appuie sur les définitions des effets direct et indirect en termes des coefficients des modèles linéaires pour le médiateur M (avec A comme variable explicative) et pour la réponse (avec A et M comme variables explicatives, sans le terme d'interaction $A*M$) :

$$M|A = a \sim \mathcal{N}\big(\beta_0 + \beta_1 a, \sigma_M^2\big), \tag{0.1}$$

$$Y|A = a, M = m \sim \mathcal{N}\big(\theta_0 + \theta_1 a + \theta_2 m, \sigma_Y^2\big). \tag{0.2}$$

Notamment, l'effet indirect est défini comme le produit $\beta_1\theta_2$ des coefficients de A et de M dans les modèles du médiateur (0.1) et de la réponse (0.2), respectivement ; l'effet direct est défini comme le coefficient θ_1 de A dans le modèle de la réponse (0.1) (Nguyen *et al.*, 2021; Valeri et VanderWeele, 2013). Cette technique est appelée « la méthode du produit » (*product method*). Une autre variation de l'approche traditionnelle consiste à ajuster, conjointement avec la régression donnée à l'équation (0.2), la régression linéaire suivante pour la réponse Y :

$$Y|A = a \sim \mathcal{N}\big(\theta_0^* + \theta_1^* a, \sigma_Y^{*2}\big). \tag{0.3}$$

Le coefficient θ_1^* dans le modèle (0.3) définit l'effet total de A sur Y, tandis que l'effet direct est défini, comme dans la méthode du produit, par le coefficient θ_1 du modèle (0.2). L'effet indirect est défini comme $\theta_1^* - \theta_1$ et, pour cette raison, cette technique est appelée « la méthode de la différence » (*difference method*). Sous l'approche traditionnelle, les effets total, direct et indirect sont définis pour l'augmentation d'une unité en A. Les méthodes du produit et de la différence sont équivalentes lorsque les modèles impliqués sont linéaires (Nguyen *et al.*, 2021; Vo *et al.*, 2020).

Comme nous pouvons le constater, l'approche traditionnelle ne fournit pas une définition générale des effets de médiation qui est applicable au-delà des modèles statistiques spécifiques (Imai *et al.*, 2010). Une autre limitation de l'approche traditionnelle est qu'elle n'est pas généralisable aux modèles non linéaires, y compris les modèles *logit* et *probit*, aux cas des médiateurs et des réponses discrets, ainsi qu'aux modèles non paramétriques ou semi-paramétriques (Imai *et al.*, 2010; Loeys *et al.*, 2013). Notamment, en commentant l'application de l'approche traditionnelle dans le contexte de la réponse binaire, Pearl (2012) a souligné : « By focusing on parameters of logistic and probit

estimators, instead of the target effect measures themselves, traditional methods produce ... biased estimates. » Vo *et al.* (2020) ont également précisé que, lorsqu'un modèle de régression logistique est utilisé pour la réponse binaire, la méthode de la différence a une tendance systématique à trouver des effets indirects non nuls dans des contextes où le traitement n'a pas d'effet sur le médiateur, et que l'approche du produit peut fournir des effets direct et indirect qui n'additionnent pas à l'effet causal total, ce qui ne permet pas d'obtenir une décomposition correcte de l'effet total. Toutefois, dans le domaine de la santé, les variables étudiées sont souvent représentées de manière binaire (Jewell, 2004, p. 9). Ainsi, le développement de nouvelles méthodes pour traiter des réponses binaires a été reconnue comme une question importante de l'analyse de médiation (Iacobucci, 2012).

L'intégration du cadre contrefactuel de l'inférence causale dans l'analyse de médiation s'est traduite par des avancées significatives méthodologiques et pratiques, tout d'abord en définissant précisément (en termes des réponses contrefactuelles) ce que l'on entend par les effets direct et indirect (De Stavola *et al.*, 2015). Contrairement à l'approche traditionnelle, ces définitions ne sont pas liées à des modèles statistiques spécifiques (Imai *et al.*, 2010).

Il est également important de souligner que l'analyse de médiation est *intrinsèquement* causale (Coffman *et al.*, 2023; Pirlott et MacKinnon, 2016). En d'autres termes, le but de ce type d'analyse statistique, sous l'approche traditionnelle ou sous l'approche causale, est toujours (inévitablement) d'estimer des effets causaux (Nguyen *et al.*, 2021). Conséquemment, des hypothèses d'identification de ces effets, y incluant celles liées au contrôle de la confusion (c'est-à-dire une distorsion de l'interprétation causale de l'association entre deux variables), doivent être rigoureusement élucidées. Par exemple, dans le cadre d'une analyse de médiation appliquée aux données observationnelles, le chercheur doit contrôler pour les variables confondantes (c'est-à-dire les variables responsables de la confusion) des relations entre l'exposition A et la réponse Y, entre l'exposition A et le médiateur M et entre M et Y. Il faut aussi noter qu'une randomisation du traitement A ne suffit pas pour éliminer la confusion dans une analyse de la médiation. Plus précisément, dans les expériences randomisées idéales, l'assignation aléatoire du traitement A élimine la confusion des relations entre A et Y et entre A et M. Toutefois, une randomisation de A ne garantit pas que l'hypothèse d'absence de confusion de la relation entre M et Y soit respectée puisque M n'est pas généralement randomisé (VanderWeele, 2015, p. 26). La présence de ce type de confusion dans les études randomisées est régulièrement ignorée par les chercheurs appliqués (Pirlott et MacKinnon, 2016). Ainsi, une contribution substantielle de l'approche causale consiste dans l'articulation explicite des postulats

nécessaires pour l'interprétation causale des estimateurs des effets direct et indirect (VanderWeele et Vansteelandt, 2014).

Finalement, des outils d'estimation des effets de médiation causale basés sur une gamme de méthodes paramétriques et non paramétriques, incluant celles permettant de traiter des interactions et des réponses et/ou médiateurs binaires, ont été mis à la disposition des chercheurs appliqués (voir, par exemple, Tingley *et al.* (2014) et Steen *et al.* (2017)). Il convient de remarquer que, dans le cadre de la médiation causale, nous cherchons généralement à décomposer l'effet total d'une exposition sur les soi-disant *effets naturels direct et indirect* (Lange *et al.*, 2017).

Même si des techniques d'estimation non paramétriques et semi-paramétriques ont été développées pour l'analyse de médiation causale (voir, par exemple, Tchetgen Tchetgen et Shpitser (2012)), les méthodes qui reposent sur des modèles paramétriques pour le médiateur et la réponse sont largement utilisées en pratique (Albert et Wang, 2015) ; des raisons pouvant être invoquées pour ceci sont leur lien naturel avec l'approche traditionnelle intrinsèquement paramétrique, ainsi que leur simplicité conceptuelle. Toutefois, dans le cadre des modèles non linéaires, la dérivation d'expressions simples et fermées pour les effets de médiation causale peut être problématique (Loeys *et al.*, 2013).

Pour une réponse binaire Y, un certain nombre d'approches basées sur la régression ont été proposées dans la littérature causale pour estimer les effets naturels sur l'échelle du rapport de cotes dans le contexte de l'analyse de médiation simple. Dans le cas d'un médiateur M continu, en se basant sur la régression linéaire pour M et sur la régression logistique pour Y, VanderWeele et Vansteelandt (2010) ont développé des expressions approximatives pour les effets naturels sous *l'hypothèse simplificatrice de la réponse rare*. Plus récemment, dans le cas de M binaire, Valeri et VanderWeele (2013) ont aussi appliqué cette hypothèse afin d'obtenir des estimateurs des effets naturels basés sur des régressions logistiques pour Y et M.

Les approches de VanderWeele et Vansteelandt (2010) et de Valeri et VanderWeele (2013) sont devenues populaires parmi les chercheurs appliqués grâce à leur implémention en SAS et en STATA [1]. Toutefois, il faut souligner que leur application demande une évaluation préalable de l'hypothèse de la réponse rare, ce qui est problématique compte tenu de la difficulté à proposer des lignes directrices explicites permettant de faire cette vérification en pratique. Bien que le seuil marginal de 10 % soit

1. Macro `mediation` (Valeri et VanderWeele, 2013) et procédure CAUSALMED (Yung *et al.*, 2018) en SAS, module PARAMED (Emsley et Liu, 2013) en STATA.

couramment utilisé en épidémiologie pour définir une maladie (réponse) rare, il ne fournit pas une justification valide pour appliquer les approches approximatives susmentionnées (Samoilenko *et al.*, 2018). Une définition plus précise, mais plus abstraite, de la réponse rare repose sur l'équivalence approximative entre le logit et le logarithme de la probabilité de Y. Ainsi, cette équivalence devrait être satisfaite dans chaque strate (combinaison des valeurs de l'exposition, du médiateur et des covariables) impliquée dans la dérivation des estimateurs approximatifs, ce qui entraîne une difficulté inévitable pour la vérification de l'hypothèse de la réponse rare dans ce contexte.

En se basant sur *l'hypothèse de la réponse commune*, Gaynor *et al.* (2019) ont dérivé des estimateurs approximatifs des effets naturels dans le cas de M continu, en utilisant la relation entre les modèles logit et probit. Cependant, leur étude de simulation n'a démontré une performance adéquate des estimateurs proposés que pour un intervalle de prévalences de Y assez limité (entre 20 % et 60 %).

Ainsi, *l'objectif principal de cette thèse* est de présenter des estimateurs exacts des effets naturels pour une réponse binaire basés sur la régression. Ici, le terme « exact » se réfère aux estimateurs dérivés sans aucune des hypothèses simplificatrices utilisées dans les approches approximatives susmentionnées, ce qui permet de surmonter la difficulté inhérente à l'évaluation de l'hypothèse de la réponse rare dans le cadre de la médiation causale d'une réponse binaire. Nos estimateurs exacts sont dérivés dans le cadre du modèle de médiation simple pour des expositions et des médiateurs binaires ou continus, et ils accommodent trois échelles binaires standards, c'est-à-dire le rapport de cotes, le rapport de risques et la différence de risques. Notre approche repose sur la spécification des modèles de régression logistique pour Y et M binaires ; la régression linéaire est utilisée pour modéliser M continu. Puisque la construction des intervalles de confiance correspondants par bootstrap peut devenir trop exigeante en termes de temps d'exécution et de mémoire de l'ordinateur, nous avons fourni les formules explicites et directes pour les erreurs standards basées sur la méthode delta. Notre approche exacte nécessite des estimateurs convergents pour les paramètres populationnels des régressions $M \sim A + \boldsymbol{C}$ et $Y \sim A + M + \boldsymbol{C}$ impliquées dans l'évaluation des réponses contrefactuelles [2]. Cela peut être réalisé dans le cadre d'une étude de cohorte vu que ce devis repose sur une sélection des unités sur la base de leur statut d'exposition (Stürmer et Brookhart, 2013).

Dans le Chapitre I de cette thèse, nous définissons les effets causaux communément utilisés dans la médiation causale et nous présentons aussi l'ensemble des hypothèses permettant d'identifier ces

2. $\boldsymbol{C}$ est un ensemble de covariables permettant l'interprétation causale des estimateurs exacts.

effets à partir des données observées.

Le Chapitre II porte sur les techniques d'estimation approximatives des effets naturels de VanderWeele et Vansteelandt (2010), Valeri et VanderWeele (2013) et Gaynor *et al.* (2019). Nous présentons aussi deux approches d'estimation fréquemment utilisées, qui ne s'appuient pas sur les hypothèses de la réponse rare ou commune, notamment celles d'Imai *et al.* (2010) et de Lange *et al.* (2012).

Le Chapitre III porte principalement sur l'étude des conséquences d'une erreur liée à la spécification du modèle du médiateur binaire sur les estimateurs des effets naturels sur l'échelle du rapport des risques présentés dans Ananth et VanderWeele (2011). La magnitude de cette erreur a été évaluée en fonction de la rareté du médiateur dans l'article *« Point : Risk ratio equations for natural direct and indirect effects in causal mediation analysis of a binary mediator and a binary outcome — A fresh look at the formulas »* de Samoilenko et Lefebvre (2019) publié dans *American Journal of Epidemiology* et présenté dans ce chapitre (Article 1). La contribution principale de cet article consiste dans l'illustration de l'impact du non-respect de l'hypothèse de la réponse rare (au sens large du terme « réponse », y incluant le médiateur) sur la qualité des estimateurs des effets naturels dérivés sous cette hypothèse. Notamment, nous avons montré qu'une erreur de la paramétrisation des effet naturels découlant de la violation de l'hypothèse de la rareté du médiateur peut engendrer des biais non négligeables. Une discussion sur les difficultés de déterminer si une réponse binaire peut être classifiée comme rare conclut ce chapitre.

Dans les Chapitres IV et V, nous présentons les estimateurs paramétriques des effets naturels direct et indirect pour une variable réponse binaire et un médiateur binaire (Chapitre IV) ou continu (Chapitre V) qui ne reposent pas sur les hypothèses de la réponse rare (marginalement ou conditionnellement) ou commune. De cette manière, nos estimateurs permettent d'éviter les difficultés liées à la vérification des hypothèses de la réponse rare ou commune dans le contexte de la médiation causale. Le Chapitre IV comporte l'article *« Parametric-regression–based causal mediation analysis of binary outcomes and binary mediators : Moving beyond the rareness or commonness of the outcome »* de Samoilenko et Lefebvre (2021) publié dans *American Journal of Epidemiology* (Article 2). Le Chapitre V est formé de l'article *« An exact regression-based approach for the estimation of natural direct and indirect effects with a binary outcome and a continuous mediator »* de Samoilenko et Lefebvre (2023) publié dans *Statistics in Medicine* (Article 3). La performance de nos estimateurs exacts a été évaluée dans des études de simulation où la réponse était rare ou commune, y compris des scénarios visant une réponse rare marginalement, mais pas conditionnellement. Il convient de

mentionner que l'utilisation de la régression logistique est un élément fondamental de notre approche. Cependant, lorsqu'un modèle de régression logistique est appliqué à de petits échantillons et/ou à des données éparses, les méthodes conventionnelles d'estimation du maximum de vraisemblance sont susceptibles d'être biaisées ou de produire des estimations infinies des coefficients en raison d'une séparation complète ou quasi-complète (Allison, 2012; Mansournia *et al.*, 2018). Ainsi, dans l'Article 3, nous avons effectué une étude de simulation afin d'évaluer l'impact de la pénalisation de Firth (1993), une approche populaire pour traiter les problèmes d'estimation susmentionnés (Greenland et Mansournia, 2015), sur les estimateurs exacts proposés.

Un nombre d'outils informatiques permettant d'implémenter l'analyse de médiation causale pour une réponse binaire ont été mis à la disposition des chercheurs appliqués au cours de la dernière décennie. Tingley *et al.* (2014) ont développé le paquet R `mediation` basé sur l'approche par simulation d'Imai *et al.* (2010). Alternativement, Steen *et al.* (2017) ont implémenté l'approche par imputation de Vansteelandt *et al.* (2012) et celle par pondération de Lange *et al.* (2012) dans le paquet R `medflex`. Si un modèle de régression logistique est spécifié pour la réponse binaire, les paquets R `mediation` et `medflex` ne fournissent que des effets naturels estimés sur les échelles de la différence de risques et du rapport de cotes, respectivement. Assez récemment, la procédure CAUSALMED a été incorporé en SAS (SAS Institute Inc., 2017; Yung *et al.*, 2018). Lorsqu'un modèle de régression logistique est utilisé pour la réponse binaire, cette procédure s'appuie sur les approches de VanderWeele et Vansteelandt (2010) et de Valeri et VanderWeele (2010) invoquant l'hypothèse de la réponse rare ; les estimations sont exprimées sur l'échelle du rapport de cotes. Les approches approximatives de VanderWeele et Vansteelandt (2010) et Valeri et VanderWeele (2010) ont été aussi récemment implémentées dans le paquet R `CMAverse` (Shi *et al.*, 2021). Il est aussi à noter que la sortie de la version 18 de STATA a été annoncée en avril de 2023. Cette version inclut des fonctionnalités permettant effectuer l'analyse de médiation causale basée sur la régression (StataCorp, 2023). Dans le contexte d'une réponse binaire, les formules exactes sont utilisées. Par défaut, les estimations des effets naturels sont fournies sur l'échelle de la différence de risques ; toutefois, les échelles du rapport de risques et du rapport de cotes sont aussi disponibles via certaines commandes *post estimation*. Cependant, contrairement à notre approche exacte, un modèle logit pour la réponse binaire ne peut pas être combiné avec un modèle linéaire pour le médiateur continu ; dans ces cas, l'utilisateur devrait spécifier un modèle probit pour la réponse (StataCorp, 2023, p. 184).

En pratique, le choix de la méthode d'estimation par un chercheur appliqué est souvent déterminé par

sa préférence individuelle pour un logiciel statistique spécifique (Starkopf *et al.*, 2017). Dans le Chapitre VI, nous présentons les fonctionnalités principales de nos macros SAS `mediation_estimates` et `bin_cont_exactmed`, ainsi que du paquet R `ExactMed`, conçus pour implémenter l'approche exacte de Samoilenko et Lefebvre (2021, 2023) pour une réponse binaire. Nos macros SAS et paquet R peuvent être utiles aux chercheurs et analystes pour qui SAS et/ou R sont le premier choix pour effectuer des analyses de médiation causale. De plus, les utilisateurs peuvent bénéficier du fait que nos outils proposent des estimations des effets naturels sur les trois échelles binaires standards. Également, une des fonctionnalités qui distinguent nos macros SAS et paquet R des outils informatiques présentés dans le paragraphe précédent est la possibilité d'appliquer la pénalisation de Firth aux données éparses et/ou à de petits échantillons.

Nous complétons cette thèse par une synthèse des méthodes d'estimation proposées. Nous discutons aussi des limites de notre approche exacte, conceptuelles ou détectées au cours du travail sur les articles constituant cette thèse. Certaines solutions potentielles sont proposées. Une discussion sur les généralisations possibles de notre approche et des difficultés anticipées est présentée.

CHAPITRE I

APPROCHE CONTREFACTUELLE DE L'ANALYSE DE MÉDIATION

Une analyse d'inférence causale comporte généralement, explicitement ou implicitement, trois étapes (Nguyen *et al.*, 2022) :

Étape 1. Définir l'effet causal ciblé ;

Étape 2. Évaluer son identifiabilité (c'est-à-dire déterminer quelles hypothèses devraient être satisfaites pour que nous puissions apprendre cet effet à partir des données observées) ;

Étape 3. Estimer l'effet causal ciblé (c'est-à-dire l'apprendre à partir des données).

Dans ce chapitre, nous définissons les effets causaux communément utilisés dans la médiation causale (*Étape 1*). Nous présentons aussi l'ensemble d'hypothèses permettant d'identifier ces effets à partir des données observationnelles (*Étape 2*).

1.1 Cadre contrefactuel de l'inférence causale

De multiples définitions de la causalité ont été proposées en épidémiologie. Une des premières définitions de la cause a été formulée par Lilienfeld (1957) : « A factor may be defined as a cause of a disease, if the incidence of the disease is diminished when exposure to this factor is likewise diminished. » À leur tour, MacMahon et Pugh (1970, p. 12) ont proposé la définition suivante de la relation de cause à effet : « An association may be classed as presumptively causal when it is believed that, had the cause been altered, the effect would have been changed [1]. » La particularité de cette dernière définition est dans l'utilisation explicite de la terminologie contrefactuelle, car il s'agit du résultat que nous aurions observé, contrairement à la réalité, si le traitement (cause) avait été différent de celui effectivement reçu. Les concepts contrefactuels sont les éléments fondamentaux du modèle causal de Neyman-Rubin (Holland, 1986), une approche largement utilisée dans la modélisation causale dont l'idée clé est que toute inférence est dérivée en utilisant non seulement les résultats

1. Selon l'interprétation de Greenland et Brumback (2003), le terme *effect* devrait s'interpréter ici comme *résultat*.

réalisés (observés) mais également les résultats non réalisés (non observés). Ce cadre statistique est initialement apparu dans l'analyse des expériences randomisées de Neyman (Splawa-Neyman, 1923; Splawa-Neyman *et al.*, 1990) ; Rubin (1974) l'a extrapolé aux études observationnelles (voir l'article de Greenland et Brumback (2003) pour un bref aperçu d'autres types principaux de modèles causaux).

Identifions les primitives du cadre contrefactuel de l'inférence causale. Soit Ω une population étudiée. Notons ω une unité (un élément) arbitraire de Ω, $\omega \in \Omega$. Les unités sont les objets primaires d'étude, par exemple, des sujets humains, des ménages, des parcelles de terrain, etc. Dans les chapitres qui suivent, nous présentons des études dont les unités sont des grossesses. Soit Y une variable [2] pour laquelle nous cherchons à expliquer pourquoi ces valeurs varient en fonction des unités de la population Ω. Nous appelons Y *la variable réponse* en raison de son statut de « variable à expliquer » (Holland, 1986). Soit A la variable de traitement [3]. Désignons le support de A par $\mathcal{A}$. Notons aussi $\boldsymbol{C}$ l'ensemble des covariables, c'est-à-dire un ensemble des variables définies sur Ω autres que Y et A. Dans le contexte d'une étude causale, les unités sont les objets sur lesquels le traitement A peut agir et le rôle de la réponse Y est de révéler l'effet du traitement A via un changement dans ses valeurs suite à un changement dans les valeurs du traitement A.

Notons A_ω la valeur du traitement A effectivement reçu par l'unité $\omega \in \Omega$. Notons Y_ω la valeur de la réponse Y observée pour l'unité ω. Soit $Y_\omega(a)$ la valeur contrefactuelle de la réponse Y pour l'unité ω qui est ou qui aurait été exposée au niveau de traitement $A = a$. Pour une unité ω, l'effet du niveau de traitement $A = a$ versus le niveau $A = a^*$ est un contraste entre $Y_\omega(a)$ et $Y_\omega(a^*)$ (par exemple, $Y_\omega(a) - Y_\omega(a^*)$; un contraste nul indique une absence d'effet individuel). Cependant, il est impossible d'observer simultanément pour la même unité ω les deux réponses contrefactuelles $Y_\omega(a)$ et $Y_\omega(a^*)$ car nous n'observons que le résultat qui correspond au niveau de traitement effectivement reçu A_ω. Par conséquent, pour une unité quelconque $\omega \in \Omega$, il est impossible d'observer l'effet *individuel* de $A = a$ versus $A = a^*$ (*Fundamental Problem of Causal Inference* (Holland, 1986)). Par contre, lorsque certaines conditions sont satisfaites, l'effet causal moyen populationnel est identifiable (estimable) à partir des données observées.

2. En suivant Holland (1986), nous définissons une *variable* comme une fonction définie sur tout ensemble Ω.

3. Dans tout ce que suit, nous utilisons les termes « traitement » et « exposition » de façon interchangeable.

1.2 Définition de l'effet causal moyen populationnel sous le cadre contrefactuel

L'effet causal moyen populationnel (*Average treatment effect ; ATE*) du niveau de traitement $A = a$ versus le niveau $A = a^*$, où $a, a^* \in \mathcal{A}$, est défini comme un contraste entre $E\{Y_\omega(a)\}$ et $E\{Y_\omega(a^*)\}$ [4] (Hernán, 2004). Nous disons qu'il existe un effet causal moyen populationnel de $A = a$ versus $A = a^*$, $a \neq a^*$, si

$$E\{Y(a)\} \neq E\{Y(a^*)\}. \tag{1.1}$$

Dans le cas d'une réponse Y binaire prenant comme valeurs les valeurs 0 ou 1,

$$E\{Y(t)\} = P(Y(t) = 1) \cdot 1 + P(Y(t) = 0) \cdot 0 = P(Y(t) = 1), \quad \forall t \in \mathcal{A},$$

et l'inégalité (1.1) s'exprime en termes des probabilités des réponses contrefactuelles :

$$P(Y(a) = 1) \neq P(Y(a^*) = 1).$$

Pour Y binaire, les mesures de l'effet causal communément utilisées sont le rapport de risques (*causal risk ratio*, $RR(a, a^*)$), le rapport de cotes (*causal odds ratio*, $OR(a, a^*)$) et la différence de risques (*causal risk difference*, $RD(a, a^*)$) (Hernán, 2004) :

$$\begin{aligned}
RR(a, a^*) &= \frac{P(Y(a) = 1)}{P(Y(a^*) = 1)}, \\
OR(a, a^*) &= \frac{P(Y(a) = 1)/(1 - P(Y(a) = 1))}{P(Y(a^*) = 1)/(1 - P(Y(a^*) = 1))}, \\
RD(a, a^*) &= P(Y(a) = 1) - P(Y(a^*) = 1).
\end{aligned}$$

1.3 Identification de l'effet causal moyen populationnel sous le cadre contrefactuel

Nous présentons maintenant un ensemble d'hypothèses qui sont généralement invoquées afin d'identifier l'effet moyen du traitement dans les études observationnelles.

Premièrement, sous le modèle contrefactuel de Neyman-Rubin, il est crucial pour chaque unité $\omega \in \Omega$ d'avoir le potentiel d'être exposée à chaque niveau du traitement A (l'hypothèse de positivité ; *positivity assumption*) (Holland, 1986).

4. Dans tout ce que suit, nous omettons l'indice ω dans $E\{Y_\omega(a)\}$, $E\{Y_\omega(a^*)\}$ et leurs dérivés ($E\{Y_\omega(a)|A = a^*\}$, $P(Y_\omega(a) = 1)$, etc.) pour faciliter la lecture.

Nous supposons aussi que la réponse Y satisfait l'hypothèse de valeur de traitement unitaire stable (*Stable Unit Treatment Value Assumption*, ou abréviativement *SUTVA*) (Rubin, 2005; VanderWeele et Hernan, 2013). Cette hypothèse consiste en deux hypothèses sous-jacentes : (1) l'hypothèse de non interférence entre les unités, c'est-à-dire pour une unité arbitraire $\omega \in \Omega$, la réponse contrefactuelle $Y_\omega(a)$, $a \in \mathcal{A}$, ne dépend pas des niveaux de traitement A reçus par les autres unités dans Ω (*no-interference assumption*) ; (2) l'hypothèse d'absence de versions multiples du traitement (*no-multiple-versions-of-treatment assumption*) stipulant que la manière dont l'unité ω a reçu le niveau de traitement $A_\omega = a$, $a \in \mathcal{A}$, n'influence pas $Y_\omega(a)$ et, par conséquent, $Y_\omega(a)$ est bien (uniquement) défini. De plus, Rubin (2005), dans sa formulation de l'hypothèse d'absence des versions multiples du traitement, requiert aussi que

$$Y_\omega(a) = Y_\omega \text{ si } A_\omega = a, \quad \forall a \in \mathcal{A}, \tag{1.2}$$

c'est-à-dire le résultat réellement observé est le même que le résultat contrefactuel correspondant au niveau de traitement observé. Ce dernier postulat lie les réponses contrefactuelles aux données observées et est souvent appelé dans la littérature causale l'hypothèse de cohérence (*consistency assumption*) (VanderWeele, 2011; VanderWeele et Hernan, 2013).

1.3.1 Confusion dans le contexte des études observationnelles

Dans cette sous-section, sans perte de généralité, nous considérons l'effet causal moyen basé sur la différence :

$$\tau = E\{Y(a)\} - E\{Y(a^*)\}. \tag{1.3}$$

Présumons aussi que le support de A est $\mathcal{A} = \{a, a^*\}$, $a \neq a^*$, c'est-à-dire le traitement A est binaire. Notons Ω_a la sous-population des unités réellement exposées au niveau de traitement $A = a$: $\Omega_a = \{\omega \in \Omega : A_\omega = a\}$. Similairement, soit $\Omega_{a^*} = \{\omega \in \Omega : A_\omega = a^*\}$. Nous avons que $\Omega_a \cap \Omega_{a^*} = \varnothing$ et $\Omega_a \cup \Omega_{a^*} = \Omega$. Notons aussi $\bar{Y}_a$ et $\bar{Y}_{a^*}$ les moyennes échantillonnales basées sur des échantillons des unités exposées à $A = a$ et $A = a^*$, respectivement.

Considérons le contraste suivant :

$$\tau_0 = E_{\Omega_a}\{Y|A = a\} - E_{\Omega_{a^*}}\{Y|A = a^*\}, \tag{1.4}$$

où $E_{\Omega_a}\{Y|A = a\}$ et $E_{\Omega_{a^*}}\{Y|A = a^*\}$ sont respectivement les espérances de la réponse Y dans

les sous-populations Ω_a et Ω_{a^*}. Dans la littérature, ce contraste est appelé l'estimand descriptif (Lundberg *et al.*, 2021). Les deux espérances $E_{\Omega_a}\{Y|A = a\}$ et $E_{\Omega_{a^*}}\{Y|A = a^*\}$ peuvent être estimées sans biais par les moyennes échantillonnales $\bar{Y}_a$ et $\bar{Y}_{a^*}$, respectivement. Ainsi, l'estimand descriptif τ_0 est estimable sans biais à partir des données observées par l'estimateur $\bar{Y}_a - \bar{Y}_{a^*}$.

Soit π la proportion des unités dans la sous-population Ω_a, $\pi = card(\Omega_a)/card(\Omega)$ où $card(\cdot)$ est le nombre cardinal. Ainsi, la proportion des unités dans la sous-population $\Omega_{a^*} = \Omega \setminus \Omega_a$ est $1 - \pi$. Notons $E_{\Omega_a}\{Y(a^*)|A = a\}$ l'espérance de la réponse contrefactuelle $Y(a^*)$ dans la sous-population Ω_a si ses unités avaient été exposées à $A = a^*$. L'espérance $E_{\Omega_{a^*}}\{Y(a)|A = a^*\}$ est définie de manière analogue. L'effet causal moyen populationnel τ (1.3) peut être décomposé comme la somme pondérée des effets causaux moyens dans les sous-populations Ω_a et Ω_{a^*} (Morgan et Winship, 2014, p. 57) :

$$\begin{aligned}\tau = \pi &\left[E_{\Omega_a}\{Y(a)|A = a\} - E_{\Omega_a}\{Y(a^*)|A = a\}\right] \\ &+ (1 - \pi)\left[E_{\Omega_{a^*}}\{Y(a)|A = a^*\} - E_{\Omega_{a^*}}\{Y(a^*)|A = a^*\}\right].\end{aligned} \tag{1.5}$$

Nous pouvons réécrire l'égalité (1.5) comme suit :

$$\begin{aligned}\tau = &\left[\pi E_{\Omega_a}\{Y(a)|A = a\} + (1 - \pi)E_{\Omega_{a^*}}\{Y(a)|A = a^*\}\right] \\ &- \left[\pi E_{\Omega_a}\{Y(a^*)|A = a\} + (1 - \pi)E_{\Omega_{a^*}}\{Y(a^*)|A = a^*\}\right].\end{aligned} \tag{1.6}$$

Si nous présumons que

$$\begin{aligned}E_{\Omega_a}\{Y(a)|A = a\} &= E_{\Omega_{a^*}}\{Y(a)|A = a^*\}, \\ E_{\Omega_a}\{Y(a^*)|A = a\} &= E_{\Omega_{a^*}}\{Y(a^*)|A = a^*\},\end{aligned} \tag{1.7}$$

nous avons que

$$\begin{aligned}&\pi E_{\Omega_a}\{Y(a)|A = a\} + (1 - \pi)E_{\Omega_{a^*}}\{Y(a)|A = a^*\} \\ &\quad = \pi E_{\Omega_a}\{Y(a)|A = a\} + (1 - \pi)E_{\Omega_a}\{Y(a)|A = a\} \\ &\quad = E_{\Omega_a}\{Y(a)|A = a\} \\ &\quad \overset{\text{par (1.2)}}{=} E_{\Omega_a}\{Y|A = a\},\end{aligned}$$

$$
\begin{aligned}
&\pi E_{\Omega_a}\{Y(a^*)|A=a\} + (1-\pi)E_{\Omega_{a^*}}\{Y(a^*)|A=a^*\} \\
&\quad = \pi E_{\Omega_{a^*}}\{Y(a^*)|A=a^*\} + (1-\pi)E_{\Omega_{a^*}}\{Y(a^*)|A=a^*\} \\
&\quad = E_{\Omega_{a^*}}\{Y(a^*)|A=a^*\} \\
&\quad \overset{\text{par (1.2)}}{=} E_{\Omega_{a^*}}\{Y|A=a^*\},
\end{aligned}
$$

et, par conséquent,

$$\tau = \tau_0 = E_{\Omega_a}\{Y|A=a\} - E_{\Omega_{a^*}}\{Y|A=a^*\}. \tag{1.8}$$

Dans les expériences randomisées, les hypothèses (1.7) sont plausibles puisque, dans ce type de devis, l'assignation du traitement est aléatoire pour chaque unité. Cela implique que, dans le cadre d'une randomisation parfaite [5], la variable de traitement est indépendante de toutes les caractéristiques de pré-traitement des unités, incluant celles qui ne sont pas observées, ainsi que des réponses contrefactuelles [6]. Hernán et Robins (2023) ont précisé qu'une randomisation parfaite entraîne que les réponses contrefactuelles sont indépendantes conjointement de la variable de traitement :

$$Y^{\mathcal{A}} \perp\!\!\!\perp A, \tag{1.9}$$

où $Y^{\mathcal{A}}$ est le vecteur de toutes les réponses contrefactuelles $Y(a)$, $a \in \mathcal{A}$. Le postulat (1.9) est appelé *l'hypothèse d'échangeabilité complète* (*full exchangeability* ; Hernán et Robins (2023), p. 15). Sous cette hypothèse, les égalités (1.7) sont satisfaites [7], ce qui implique, comme nous l'avons montré, l'égalité (1.8). Ainsi, dans les études randomisées bien conçues, nous pouvons utiliser l'estimateur sans biais $\bar{Y}_a - \bar{Y}_{a^*}$ de l'estimand descriptif τ_0 (1.4) pour estimer l'effet causal moyen populationnel τ (1.3).

Dans les études observationnelles [8] sur l'effet d'un traitement sur une réponse d'intérêt, le chercheur est plutôt un observateur qu'un agent qui assigne le traitement (Lash *et al.*, 2021, p. 106). En d'autres termes, l'assignation du traitement est hors de contrôle du chercheur (Morgan et Winship, 2014, p. 7). Dans ce type d'études, l'exposition des unités à un traitement est souvent déterminée par leurs

5. Voir la définition de *ideal randomized experiment* dans le livre de Hernán et Robins (2023), p. 14.

6. Greenland et Robins (2009) ont écrit : « Simple (complete) randomization of exposure means that exposure events occur independently of every event that precedes their occurrence. In particular, because the potential outcomes already exist at the times of exposure events, it implies that exposure events occur independently of the potential outcomes of interest. »

7. Hernán et Robins (2023) ont utilisé le terme *mean exchangeability* pour les conditions (1.7).

8. Le terme *études non expérimentales* est aussi utilisé (Lash *et al.*, 2021, p. 106)

caractéristiques initiales (variables de pré-traitement). Si ces variables sont aussi des causes de la variable réponse, nous les appelons des *variables confondantes*. Ainsi, une variable confondante est à la fois une cause de la variable de traitement et de la variable réponse (Hernán *et al.*, 2002). La Figure 1.1 représente visuellement un scénario où la variable C est une cause commune du traitement A et de la réponse Y, ce qui est reflété par les flèches $C \to A$ et $C \to Y$, respectivement. Ainsi, C est une variable confondante pour l'effet de A sur Y.

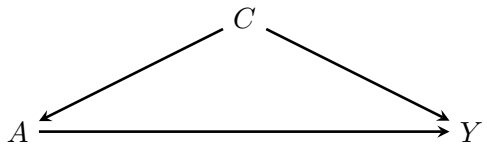


FIGURE 1.1 L'effet du traitement A sur la réponse Y en présence d'une variable confondante C

Dans les études observationnelles, contrairement aux études randomisées, les variables de pré-traitement sont rarement distribuées de façon homogène à travers des niveaux du traitement. Cela implique que les hypothèses (1.7) ne sont pas généralement satisfaites et

$$\tau \neq \tau_0.$$

Autrement dit,

$$E\{Y(a)\} - E\{Y(a^*)\} \neq E_{\Omega_a}\{Y|A=a\} - E_{\Omega_{a^*}}\{Y|A=a^*\}, \tag{1.10}$$

et donc, l'estimateur $\bar{Y}_a - \bar{Y}_{a^*}$ n'est pas un estimateur sans biais pour l'effet causal populationnel τ.

Pour toute situation où l'inégalité (1.10) se produit (c'est-à-dire lorsque l'estimand causal (1.3) ne coïncide pas avec l'estimand descriptif (1.4)), nous disons que la confusion est présente (Jewell, 2004, p. 98). Cette définition peut être tout aussi bien extrapolée pour d'autres mesures d'effets. Par exemple, l'inégalité (1.10) s'exprime sur l'échelle du rapport de risques comme

$$\frac{P(Y(a)=1)}{P(Y(a^*)=1)} \neq \frac{P_{\Omega_a}(Y=1|A=a)}{P_{\Omega_{a^*}}(Y=1|A=a^*)}.$$

Pour comprendre les sources de l'inégalité (1.10), nous réécrivons l'équation (1.6) comme suit (Mor-

gan et Winship, 2014, p. 59) :

$$
\begin{aligned}
\tau &= \left(1 + (\pi - 1)\right) E_{\Omega_a}\{Y(a)|A = a\} + (1 - \pi) E_{\Omega_{a^*}}\{Y(a)|A = a^*\} \\
&\quad - \pi E_{\Omega_a}\{Y(a^*)|A = a\} - (1 - \pi) E_{\Omega_{a^*}}\{Y(a^*)|A = a^*\} \\
&= E_{\Omega_a}\{Y(a)|A = a\} + (\pi - 1) E_{\Omega_a}\{Y(a)|A = a\} + (1 - \pi) E_{\Omega_{a^*}}\{Y(a)|A = a^*\} \\
&\quad - \pi E_{\Omega_a}\{Y(a^*)|A = a\} - E_{\Omega_{a^*}}\{Y(a^*)|A = a^*\} + \pi E_{\Omega_{a^*}}\{Y(a^*)|A = a^*\} \\
&= \left[E_{\Omega_a}\{Y(a)|A = a\} - E_{\Omega_{a^*}}\{Y(a^*)|A = a^*\}\right] \\
&\quad + (\pi - 1) E_{\Omega_a}\{Y(a)|A = a\} + (1 - \pi) E_{\Omega_{a^*}}\{Y(a)|A = a^*\} \\
&\quad - \left(1 + (\pi - 1)\right) E_{\Omega_a}\{Y(a^*)|A = a\} + \left(1 + (\pi - 1)\right) E_{\Omega_{a^*}}\{Y(a^*)|A = a^*\} \\
&= \tau_0 - E_{\Omega_a}\{Y(a^*)|A = a\} + E_{\Omega_{a^*}}\{Y(a^*)|A = a^*\} \\
&\quad + (\pi - 1) E_{\Omega_a}\{Y(a) - Y(a^*)|A = a\} + (1 - \pi) E_{\Omega_{a^*}}\{Y(a) - Y(a^*)|A = a^*\} \\
&= \tau_0 - \left[E_{\Omega_a}\{Y(a^*)|A = a\} - E_{\Omega_{a^*}}\{Y(a^*)|A = a^*\}\right] \\
&\quad - (1 - \pi)\left[E_{\Omega_a}\{Y(a) - Y(a^*)|A = a\} - E_{\Omega_{a^*}}\{Y(a) - Y(a^*)|A = a^*\}\right].
\end{aligned}
$$

Ainsi, nous avons que

$$
\begin{aligned}
\tau_0 - \tau = &\left[E_{\Omega_a}\{Y(a^*)|A = a\} - E_{\Omega_{a^*}}\{Y(a^*)|A = a^*\}\right] \\
&+ (1 - \pi)\left[E_{\Omega_a}\{Y(a) - Y(a^*)|A = a\} - E_{\Omega_{a^*}}\{Y(a) - Y(a^*)|A = a^*\}\right].
\end{aligned}
\tag{1.11}
$$

Dans cette décomposition de $\tau_0 - \tau$, le terme

$$E_{\Omega_a}\{Y(a^*)|A = a\} - E_{\Omega_{a^*}}\{Y(a^*)|A = a^*\}$$

correspond à la différence entre les espérances de $Y(a^*)$ chez les unités exposées à $A = a$ et chez celles qui ont été exposées à $A = a^*$; cette différence est appelée le biais de référence ou le biais de pré-traitement (*baseline bias*, *pre-treatment bias* ; Morgan et Winship (2014), p. 59 ; Breen *et al.* (2015)). Le deuxième terme

$$(1 - \pi)\left[E_{\Omega_a}\{Y(a) - Y(a^*)|A = a\} - E_{\Omega_{a^*}}\{Y(a) - Y(a^*)|A = a^*\}\right]$$

dans la décomposition (1.11) exprime une différence pondérée des effets causaux moyens chez les unités exposées à $A = a$ ($E_{\Omega_a}\{Y(a) - Y(a^*)|A = a\}$) et chez celles qui ont été exposées à $A = a^*$ ($E_{\Omega_{a^*}}\{Y(a) - Y(a^*)|A = a^*\}$) ; cette différence est appelée le biais de l'effet de traitement différentiel

(*differential treatment effect bias* ; Morgan et Winship (2014), p. 59 ; Breen *et al.* (2015)).

Considérons un exemple hypothétique d'une étude observationnelle examinant l'effet potentiel de l'éducation sur la réussite professionnelle d'un individu (Morgan et Winship, 2014, pp. 59-60). Considérons un traitement A binaire dont les niveaux sont $a =$ *avoir un diplôme universitaire* et $a^* =$ *ne pas avoir un diplôme universitaire*. Définissons la variable réponse continue Y comme un certain score de la réussite professionnelle (les valeurs plus grandes correspondant à une réussite plus grande). Présumons aussi que 30% de la population étudiée ont des diplômes universitaires (c'est-à-dire, $\pi = 0.3$). L'information sur les espérances des réponses contrefactuelles est résumée dans le Tableau 1.1.

TABLEAU 1.1 Exemple hypothétique sur la réussite professionnelle en fonction de l'éducation : espérances des réponses contrefactuelles

Sous-population	Réponse contrefactuelle	Espérance de la réponse contrefactuelle
Individus non diplômés, Ω_{a^*}	$Y(a^*)$	5
Individus non diplômés, Ω_{a^*}	$Y(a)$	8
Individus diplômés, Ω_a	$Y(a^*)$	6
Individus diplômés, Ω_a	$Y(a)$	10

Selon le Tableau 1.1,

$$E_{\Omega_a^*}\{Y(a^*)|A = a^*\} = 5, \quad E_{\Omega_a}\{Y(a)|A = a\} = 10,$$

et $\tau_0 = 10 - 5 = 5$.

Comparativement aux individus sans diplômes universitaires (les unités appartenant à Ω_a^*), leurs contreparties graduées (les unités appartenant à Ω_a) auraient mieux réussi en moyenne sur le marché du travail si elles n'avaient pas obtenu, contrairement à la réalité, des diplômes universitaires ($E_{\Omega_a}\{Y(a^*)|A = a\} = 6$ versus $E_{\Omega_{a^*}}\{Y(a^*)|A = a^*\} = 5$). Ainsi, nous pouvons constater la présence d'un biais de référence :

$$E_{\Omega_a}\{Y(a^*)|A = a\} - E_{\Omega_{a^*}}\{Y(a^*)|A = a^*\} = 6 - 5 = 1. \tag{1.12}$$

Ce biais strictement positif pourrait être expliqué, par exemple, par le fait que les individus diplômés sont peut-être plus intelligents intrinsèquement. En d'autres termes, une distribution non balancée des caractéristiques initiales liées aux fonctions cognitives peut être la source du biais de référence

dans cet exemple.

L'effet causal moyen chez les individus diplômés est

$$E_{\Omega_a}\{Y(a) - Y(a^*)|A = a\} = 10 - 6 = 4;$$

l'effet causal moyen chez les individus sans diplômes universitaires est

$$E_{\Omega_{a^*}}\{Y(a) - Y(a^*)|A = a^*\} = 8 - 5 = 3.$$

Puisque $\pi = 0.3$, nous avons que le biais de l'effet de traitement différentiel est

$$\begin{aligned}(1-\pi)\big[E_{\Omega_a}\{Y(a) - Y(a^*)|A = a\} - E_{\Omega_{a^*}}\{Y(a) - Y(a^*)|A = a^*\}\big] \\ = (1 - 0.3) \cdot (4 - 3) = 0.7.\end{aligned} \tag{1.13}$$

La présence d'un biais positif de l'effet de traitement différentiel signifie que l'effet causal moyen de l'éducation universitaire sur la réussite professionnelle est plus grand dans la sous-population des individus diplômés que chez les personnes non diplômées [9].

Ainsi, en utilisant la décomposition (1.11) et en tenant compte des valeurs des biais (1.12-1.13), nous avons que

$$\tau_0 - \tau = 1 + 0.7 = 1.7,$$

d'où nous pouvons constater la présence de la confusion dans cet exemple hypothétique.

La confusion est un problème présent dans presque toutes les études non randomisées (Lash *et al.*, 2021, p. 263). Pour identifier l'effet causal moyen populationnel dans le contexte des études observationnelles, en tenant compte de la présence potentielle de la confusion, nous avons besoin d'introduire une autre hypothèse fondamentale, à savoir l'hypothèse forte conditionnelle de l'ignorabilité de l'affectation du traitement (*strong conditional ignorable treatment assignment* ; Gao et Luo (2019), p. 29 ; Greenland et Robins (2009)). Cette hypothèse est aussi appelée *l'hypothèse forte d'échangeabilité conditionnelle* (*strong conditional exchangeability* ; Greenland et Robins (2009) ; voir d'autres termes pour cette hypothèse dans Gao et Luo (2019, p. 29)). Pour introduire cette hypothèse, nous définissons le concept de l'indépendance conditionnelle. On dit que deux événements E_1 et E_2 sont

9. Dans le cas général, il est facile de voir que, si les hypothèses (1.7) sont satisfaites, le biais de l'effet de traitement différentiel est nul.

conditionnellement indépendants selon F si « la probabilité conditionnelle de E_1, F étant réalisé, n'est pas affectée par l'information que E_2 est ou n'est pas survenu » (Ross, 2009, p. 114) :

$$P(E_1|E_2, F) = P(E_1|F).$$

Nous utilisons la notation suivante pour l'indépendance conditionnelle des événements E_1 et E_2 selon F :

$$E_1 \perp\!\!\!\perp E_2 \,|\, F.$$

L'hypothèse forte conditionnelle de l'ignorabilité de l'affectation du traitement est formulée comme suit :

$$Y^{\mathcal{A}} \perp\!\!\!\perp A \mid \boldsymbol{C}. \tag{1.14}$$

Cette hypothèse signifie que, conditionnellement aux variables $\boldsymbol{C}$, le vecteur $Y^{\mathcal{A}}$ de toutes les réponses contrefactuelles $Y(a)$, $a \in \mathcal{A}$, est indépendant de l'affectation au traitement A.

1.3.2 Application des hypothèses d'identification pour établir un lien entre les réponses contrefactuelles et les données observées

Dans le contexte des études observationnelles, l'hypothèse de positivité doit être ajustée (c'est-à-dire reformulée) à la présence des variables potentiellement confondantes $\boldsymbol{C}$. Dans le cas d'une exposition A discrète, nous présumons que

$$P(A = a|\boldsymbol{C} = \boldsymbol{c}) > 0, \quad \forall a \in \mathcal{A}, \quad \forall \boldsymbol{c} \in \mathcal{C}, \tag{1.15}$$

où $\mathcal{C}$ est le support du vecteur $\boldsymbol{C}$ (Westreich et Cole, 2010) [10]. La positivité exige que, pour chaque combinaison de variables potentiellement confondantes qui peut être observée dans la population étudiée, la probabilité d'être exposée à chaque niveau du traitement A est strictement positive pour chaque unité de cette population.

Sous les hypothèses de positivité, SUTVA, cohérence et d'ignorabilité de l'affectation du traitement, nous pouvons identifier les éléments $E\{Y(a)\}$ et $E\{Y(a^*)\}$ de l'effet causal populationnel de $A = a$ versus $A = a^*$:

10. Pour une exposition A continue, nous présumons que $f_{A|\boldsymbol{C}=\boldsymbol{c}}(a) > 0$, $\forall a \in \mathcal{A}$, $\forall \boldsymbol{c} \in \mathcal{C}$, où $f_{A|\boldsymbol{C}=\boldsymbol{c}}(a)$ est la fonction de densité conditionnelle de A lorsque $\boldsymbol{C} = \boldsymbol{c}$.

$$
\begin{aligned}
E\{Y(a)\} &= E\{E\{Y(a)\}|\boldsymbol{C}\} \\
&= \sum_{\boldsymbol{c}\in\mathcal{C}} E\{Y(a)|\boldsymbol{C}=\boldsymbol{c}\}P(\boldsymbol{C}=\boldsymbol{c}) \\
&\overset{\text{par (1.14,1.15)}}{=} \sum_{\boldsymbol{c}\in\mathcal{C}} E\{Y(a)|A=a,\boldsymbol{C}=\boldsymbol{c}\}P(\boldsymbol{C}=\boldsymbol{c}) \\
&\overset{\text{par (1.2)}}{=} \sum_{\boldsymbol{c}\in\mathcal{C}} E\{Y|A=a,\boldsymbol{C}=\boldsymbol{c}\}P(\boldsymbol{C}=\boldsymbol{c}),
\end{aligned}
$$

$$
\begin{aligned}
E\{Y(a^*)\} &= E\{E\{Y(a^*)\}|\boldsymbol{C}\} \\
&= \sum_{\boldsymbol{c}\in\mathcal{C}} E\{Y(a^*)|\boldsymbol{C}=\boldsymbol{c}\}P(\boldsymbol{C}=\boldsymbol{c}) \\
&\overset{\text{par (1.14,1.15)}}{=} \sum_{\boldsymbol{c}\in\mathcal{C}} E\{Y(a^*)|A=a^*,\boldsymbol{C}=\boldsymbol{c}\}P(\boldsymbol{C}=\boldsymbol{c}) \\
&\overset{\text{par (1.2)}}{=} \sum_{\boldsymbol{c}\in\mathcal{C}} E\{Y|A=a^*,\boldsymbol{C}=\boldsymbol{c}\}P(\boldsymbol{C}=\boldsymbol{c}).
\end{aligned}
$$

Dans les expressions finales de ces dérivations, tous les éléments sont estimables à partir des données observées. Ainsi, les hypothèses de positivité, SUTVA, cohérence et d'ignorabilité de l'affectation du traitement permettent un lien entre les données observées et les réponses contrefactuelles définissant les effets causaux.

1.4 Cadre contrefactuel dans la médiation causale

Pour l'ensemble de la présentation, nous notons M la variable médiatrice (ou simplement médiateur) entre le traitement A [11] et la variable réponse Y. Soit $\mathcal{M}$ le support du médiateur M. De plus, soit $\boldsymbol{C}$ l'ensemble des covariables. Dans le contexte de la médiation, le terme *covariables* est utilisé pour désigner les variables autres que A, M et Y qui sont définies sur la population étudiée Ω. Par exemple, la Figure 1.2 reflète le scénario où $\boldsymbol{C} = \boldsymbol{C}_{AY} \cup \boldsymbol{C}_{AM} \cup \boldsymbol{C}_{MY}$, et les ensembles $\boldsymbol{C}_{AY}$, $\boldsymbol{C}_{AM}$ et $\boldsymbol{C}_{MY}$ représentent les variables confondantes des relations $A-Y$, $A-M$, et $M-Y$, respectivement.

Notons M_ω la valeur du médiateur M observée pour l'unité $\omega \in \Omega$. Soit $M_\omega(a)$ la valeur contrefactuelle de M pour l'unité ω qui est ou aurait été exposée au traitement $A = a$. Désignons aussi

11. À moins d'indication contraire, nous considérons la variable de traitement A binaire ou continue.

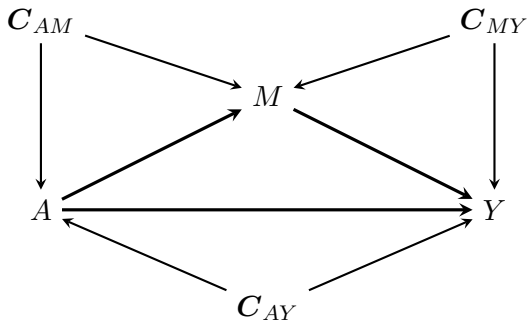


FIGURE 1.2 Modèle de médiation avec l'exposition A, le médiateur M et la réponse Y en présence des variables confondantes des relations $A - Y$, $A - M$, et $M - Y$

par $Y_\omega(a, m)$ la valeur contrefactuelle de Y pour l'unité ω si A et M étaient fixés, possiblement contrairement à la réalité, aux valeurs a et m. Notons aussi $Y_\omega(a, M(a^*))$ la valeur contrefactuelle de la réponse Y pour l'unité ω si l'exposition était fixée à $A = a$ et le médiateur M était fixé à la valeur que l'on aurait observée si l'exposition avait été fixée au niveau $A = a^*$.

1.4.1 Définition des effets direct et indirect dans le cadre de l'approche contrefactuelle

Nous utilisons le terme *effet total populationnel* (*populational total effect*, TE) pour l'effet causal moyen populationnel du niveau de traitement $A = a$ versus le niveau $A = a^*$, $a, a^* \in \mathcal{A}$, $a \neq a^*$, que nous avons défini dans la Section 1.3 (voir l'équation (1.3)).

Dans le cadre de la médiation causale, nous cherchons généralement à décomposer l'effet total d'une exposition sur la base des soi-disant effet naturel direct (*natural direct effect*, NDE) et effet naturel indirect (*natural indirect effect*, NIE).

Les effets naturels sont exprimés en termes des réponses contrefactuelles emboîtées $Y_\omega(a, M(a^*))$, $a, a^* \in \mathcal{A}$. Notamment, l'effet naturel direct populationnel du niveau de traitement $A = a$ versus le niveau $A = a^*$ est défini comme le contraste entre les espérances $E\{Y_\omega(a, M(a^*))\}$ et $E\{Y_\omega(a^*, M(a^*))\}$; l'effet naturel indirect populationnel de $A = a$ versus $A = a^*$ s'exprime comme le contraste entre $E\{Y_\omega(a, M(a))\}$ et $E\{Y_\omega(a, M(a^*))\}$. Autrement dit, l'effet naturel direct populationnel correspond au changement moyen que l'on aurait observé pour la réponse Y si le niveau de traitement était changé de $A = a$ à $A = a^*$ et si le médiateur était fixé pour toute la population Ω au niveau qu'il devrait être si le niveau de traitement était fixé à $A = a^*$. L'effet naturel indirect populationnel reflète le changement qu'on aurait observé pour la réponse Y si le traitement était fixé à $A = a$ pour chaque unité $\omega \in \Omega$ et que le médiateur varierait du niveau qu'il devrait être si

le traitement était fixé à $A = a^*$ au niveau qu'il devrait être si le traitement était fixé à $A = a$. Conceptuellement, l'effet naturel indirect populationnel capture l'effet du traitement A sur la réponse Y expliqué par l'effet de A sur le médiateur M et du médiateur M sur la réponse Y, et l'effet naturel direct populationnel capture la fraction de l'effet total populationnel qui n'est pas expliqué par l'effet naturel indirect populationnel.

Dans tout ce qui suit, nous omettons l'indice ω dans les espérances $E\{Y_\omega(a, M(a))\}$, $E\{Y_\omega(a, M(a^*))\}$, $E\{Y_\omega(a^*, M(a^*))\}$ et leurs dérivés pour faciliter la lecture. Dans le cas de la réponse binaire Y, l'effet naturel direct populationnel de $A = a$ versus $A = a^*$ est le contraste entre les probabilités contrefactuelles emboîtées $P(Y(a, M(a^*)) = 1)$ et $P(Y(a^*, M(a^*)) = 1)$; le contraste entre $P(Y(a, M(a)) = 1)$ et $P(Y(a, M(a^*)) = 1)$ exprime l'effet naturel indirect populationnel. Les effets naturels direct et indirect populationnels peuvent être présentés sur les échelles du rapport de risques (RR^{NDE}_{a,a^*}, RR^{NIE}_{a,a^*}), du rapport de cotes (OR^{NDE}_{a,a^*}, OR^{NIE}_{a,a^*}) et de la différence de risques (RD^{NDE}_{a,a^*}, RD^{NIE}_{a,a^*}) :

$$\begin{aligned} RR^{NDE}_{a,a^*} &= \frac{P(Y(a, M(a^*)) = 1)}{P(Y(a^*, M(a^*)) = 1)}, \\ RR^{NIE}_{a,a^*} &= \frac{P(Y(a, M(a)) = 1)}{P(Y(a, M(a^*)) = 1)}, \end{aligned} \tag{1.16}$$

$$\begin{aligned} OR^{NDE}_{a,a^*} &= \frac{\dfrac{P(Y(a, M(a^*)) = 1)}{1 - P(Y(a, M(a^*)) = 1)}}{\dfrac{P(Y(a^*, M(a^*)) = 1}{1 - P(Y(a^*, M(a^*)) = 1)}}, \\ OR^{NIE}_{a,a^*} &= \frac{\dfrac{P(Y(a, M(a)) = 1)}{1 - P(Y(a, M(a)) = 1)}}{\dfrac{P(Y(a, M(a^*)) = 1)}{1 - P(Y(a, M(a^*)) = 1)}}, \end{aligned} \tag{1.17}$$

$$\begin{aligned} RD^{NDE}_{a,a^*} &= P(Y(a, M(a^*)) = 1) - P(Y(a^*, M(a^*)) = 1), \\ RD^{NIE}_{a,a^*} &= P(Y(a, M(a)) = 1) - P(Y(a, M(a^*)) = 1). \end{aligned} \tag{1.18}$$

Conjointement aux effets populationnels (1.16-1.18), nous pouvons également définir les effets naturels spécifiques aux strates des valeurs particulières des covariables $\boldsymbol{C}$. Ainsi, pour la strate $\boldsymbol{C} = \boldsymbol{c}$,

ces effets s'expriment en termes des probabilités contrefactuelles emboîtées conditionnelles

$$P(Y(a, M(a)) = 1|\boldsymbol{C} = \boldsymbol{c}),$$
$$P(Y(a, M(a^*)) = 1|\boldsymbol{C} = \boldsymbol{c}),$$
$$P(Y(a^*, M(a^*)) = 1|\boldsymbol{C} = \boldsymbol{c}),$$

et par conséquent, nous les appelons « effets naturels conditionnels ». Par exemple, pour la strate $\boldsymbol{C} = \boldsymbol{c}$, les effets naturels direct et indirect conditionnels s'expriment sur les échelles standards binaires comme suit :

$$\begin{aligned} RR^{NDE}_{a,a^*|\boldsymbol{c}} &= \frac{P(Y(a, M(a^*)) = 1|\boldsymbol{C} = \boldsymbol{c})}{P(Y(a^*, M(a^*)) = 1|\boldsymbol{C} = \boldsymbol{c})}, \\ RR^{NIE}_{a,a^*|\boldsymbol{c}} &= \frac{P(Y(a, M(a)) = 1|\boldsymbol{C} = \boldsymbol{c})}{P(Y(a, M(a^*)) = 1|\boldsymbol{C} = \boldsymbol{c})}, \end{aligned} \tag{1.19}$$

$$\begin{aligned} OR^{NDE}_{a,a^*|\boldsymbol{c}} &= \frac{\frac{P(Y(a, M(a^*)) = 1|\boldsymbol{C} = \boldsymbol{c})}{1 - P(Y(a, M(a^*)) = 1|\boldsymbol{C} = \boldsymbol{c})}}{\frac{P(Y(a^*, M(a^*)) = 1|\boldsymbol{C} = \boldsymbol{c})}{1 - P(Y(a^*, M(a^*)) = 1|\boldsymbol{C} = \boldsymbol{c})}}, \\ OR^{NIE}_{a,a^*|\boldsymbol{c}} &= \frac{\frac{P(Y(a, M(a)) = 1|\boldsymbol{C} = \boldsymbol{c})}{1 - P(Y(a, M(a)) = 1|\boldsymbol{C} = \boldsymbol{c})}}{\frac{P(Y(a, M(a^*)) = 1|\boldsymbol{C} = \boldsymbol{c})}{1 - P(Y(a, M(a^*)) = 1|\boldsymbol{C} = \boldsymbol{c})}}, \end{aligned} \tag{1.20}$$

$$\begin{aligned} RD^{NDE}_{a,a^*|\boldsymbol{c}} &= P(Y(a, M(a^*)) = 1|\boldsymbol{C} = \boldsymbol{c}) - P(Y(a^*, M(a^*)) = 1|\boldsymbol{C} = \boldsymbol{c}), \\ RD^{NIE}_{a,a^*|\boldsymbol{c}} &= P(Y(a, M(a)) = 1|\boldsymbol{C} = \boldsymbol{c}) - P(Y(a, M(a^*)) = 1|\boldsymbol{C} = \boldsymbol{c}). \end{aligned} \tag{1.21}$$

1.4.2 Identification des effets naturels direct et indirect dans les études observationnelles

Dans le cadre de la médiation causale, nous généralisons l'hypothèse de cohérence (1.2) comme suit :

$$A_\omega = a \quad \Rightarrow \quad M_\omega(a) = M_\omega, \tag{1.22}$$
$$A_\omega = a, M_\omega = m \quad \Rightarrow \quad Y_\omega(a, m) = Y_\omega, \tag{1.23}$$
$$M_\omega(a^*) = m \quad \Rightarrow \quad Y_\omega(a, M(a^*)) = Y_\omega(a, m), \tag{1.24}$$

pour $\forall a$, $a^* \in \mathcal{A}$, $\forall m \in \mathcal{M}$, $\forall \omega \in \Omega$.

Nous présumons aussi que, dans le cas d'un médiateur M discret [12],

$$P(M = m|A = a, \boldsymbol{C} = \boldsymbol{c}) > 0, \quad \forall a \in \mathcal{A}, \quad \forall \boldsymbol{c} \in \mathcal{C}. \tag{1.25}$$

La dernière hypothèse est une hypothèse de positivité par rapport au médiateur (Nguyen *et al.*, 2022).

Les hypothèses suivantes, ainsi que les hypothèses (1.15), (1.22-1.25), sont suffisantes pour identifier les effets naturels sans biais de confusion à partir des données observées :

$$Y(a, m) \perp\!\!\!\perp A \,|\, \boldsymbol{C}, \tag{1.26}$$

$$Y(a, m) \perp\!\!\!\perp M \,|\, A, \boldsymbol{C}, \tag{1.27}$$

$$M(a) \perp\!\!\!\perp A \,|\, \boldsymbol{C}, \tag{1.28}$$

$$Y(a, m) \perp\!\!\!\perp M(a^*) \,|\, \boldsymbol{C}, \tag{1.29}$$

pour $\forall a,\, a^* \in \mathcal{A}$, $\forall m \in \mathcal{M}$. Les hypothèses (1.26), (1.27) et (1.28) expriment de façon formelle que l'ensemble $\boldsymbol{C}$ est suffisant pour contrôler la confusion des relations $A - Y$, $M - Y$, et $A - M$, respectivement (VanderWeele et Vansteelandt, 2009; VanderWeele, 2015). L'hypothèse (1.29) est souvent interprétée comme l'absence de variables confondantes (mesurées ou non mesurées) pour la relation $M - Y$ affectées par A (VanderWeele et Vansteelandt, 2009; VanderWeele, 2015) [13].

Considérons maintenant, sans perte de généralité, la variable réponse Y binaire et le médiateur M discret. Montrons comment la probabilité $P(Y(a, M(a^*)) = 1|\boldsymbol{C} = \boldsymbol{c})$ peut être estimée à partir des données observées :

$$\begin{aligned}
&P(Y(a, M(a^*)) = 1|\boldsymbol{C} = \boldsymbol{c}) \\
&\quad = \sum_{m \in \mathcal{M}} P(Y(a, M(a^*)) = 1|M(a^*) = m, \boldsymbol{C} = \boldsymbol{c}) P(M(a^*) = m|\boldsymbol{C} = \boldsymbol{c}) \\
&\quad \overset{\text{par (1.24)}}{=} \sum_{m \in \mathcal{M}} P(Y(a, m) = 1|M(a^*) = m, \boldsymbol{C} = \boldsymbol{c}) P(M(a^*) = m|\boldsymbol{C} = \boldsymbol{c})
\end{aligned}$$

12. Si M est continu, nous présumons que $f_{M|A=a,\boldsymbol{C}=\boldsymbol{c}}(m) > 0$, $\forall a \in \mathcal{A}$, $\forall \boldsymbol{c} \in \mathcal{C}$, où $f_{M|A=a^*,\boldsymbol{C}=\boldsymbol{c}}(m)$ est la fonction de densité conditionnelle de M lorsque $A = a^*$ et $\boldsymbol{C} = \boldsymbol{c}$.

13. Andrews et Didelez (2021) ont illustré que la violation de l'hypothèse (1.29) peut se produire autrement que par une confusion intermédiaire (le terme *confusion intermédiaire* signifie la présence de variables confondantes pour la relation $M - Y$ affectées par A).

$$
\begin{aligned}
&\overset{\text{par } (1.29)}{=} \sum_{m\in\mathcal{M}} P(Y(a,m)=1|\boldsymbol{C}=\boldsymbol{c})P(M(a^*)=m|\boldsymbol{C}=\boldsymbol{c}) \\
&\overset{\text{par } (1.26,1.15)}{=} \sum_{m\in\mathcal{M}} P(Y(a,m)=1|A=a,\boldsymbol{C}=\boldsymbol{c})P(M(a^*)=m|\boldsymbol{C}=\boldsymbol{c}) \\
&\overset{\text{par } (1.27,1.25)}{=} \sum_{m\in\mathcal{M}} P(Y(a,m)=1|A=a,M=m,\boldsymbol{C}=\boldsymbol{c})P(M(a^*)=m|\boldsymbol{C}=\boldsymbol{c}) \\
&\overset{\text{par } (1.23)}{=} \sum_{m\in\mathcal{M}} P(Y=1|A=a,M=m,\boldsymbol{C}=\boldsymbol{c})P(M(a^*)=m|\boldsymbol{C}=\boldsymbol{c}) \\
&\overset{\text{par } (1.28,1.15)}{=} \sum_{m\in\mathcal{M}} P(Y=1|A=a,M=m,\boldsymbol{C}=\boldsymbol{c})P(M(a^*)=m|A=a^*,\boldsymbol{C}=\boldsymbol{c}) \\
&\overset{\text{par } (1.22)}{=} \sum_{m\in\mathcal{M}} P(Y=1|A=a,M=m,\boldsymbol{C}=\boldsymbol{c})P(M=m|A=a^*,\boldsymbol{C}=\boldsymbol{c}). \qquad (1.30)
\end{aligned}
$$

Ainsi, l'expression (1.30) permet d'apprendre la probabilité contrefactuelle emboîtée $P(Y(a,M(a^*))=1|\boldsymbol{C}=\boldsymbol{c})$ à partir des données observées.

De la même façon, nous pouvons montrer que, sous les hypothèses (1.15) et (1.22-1.29),

$$P(Y(a^*,M(a^*))=1|\boldsymbol{C}=\boldsymbol{c}) = \sum_{m\in\mathcal{M}} P(Y=1|A=a^*,M=m,\boldsymbol{C}=\boldsymbol{c})P(M=m|A=a^*,\boldsymbol{C}=\boldsymbol{c})$$

et

$$P(Y(a,M(a))=1|\boldsymbol{C}=\boldsymbol{c}) = \sum_{m\in\mathcal{M}} P(Y=1|A=a,M=m,\boldsymbol{C}=\boldsymbol{c})P(M=m|A=a,\boldsymbol{C}=\boldsymbol{c}).$$

Ainsi, les effets naturels direct et indirect conditionnels (1.19-1.21) sont identifiables (estimables) à partir des données observées.

La dérivation (1.30) a été effectuée pour un cas particulier, précisément pour Y binaire et M discret. Dans le cas général, sous les hypothèses (1.15) et (1.22-1.29),

$$E\left\{Y(a,M(a^*))|\boldsymbol{C}=\boldsymbol{c}\right\} = \sum_{m\in\mathcal{M}} E\left\{Y|A=a,M=m,\boldsymbol{C}=\boldsymbol{c}\right\} P(M=m|A=a^*,\boldsymbol{C}=\boldsymbol{c}) \qquad (1.31)$$

pour M discret, et

$$E\left\{Y(a,M(a^*))|\boldsymbol{C}=\boldsymbol{c}\right\} = \int_{\mathcal{M}} E\left\{Y|A=a,M=m,\boldsymbol{C}=\boldsymbol{c}\right\} dF_{M|A=a^*,\boldsymbol{C}=\boldsymbol{c}}(m) \qquad (1.32)$$

pour M continu où $F_{M|A=a^*,\boldsymbol{C}=\boldsymbol{c}}(m)$ est la fonction de répartition conditionnelle de M lorsque $A = a^*$ et $\boldsymbol{C} = \boldsymbol{c}$.

1.4.3 Décomposition de l'effet total

Pour Y binaire et la strate $\boldsymbol{C} = \boldsymbol{c}$, l'effet total conditionnel du niveau de traitement $A = a$ versus le niveau $A = a^*$ sur l'échelle du rapport de risques est défini comme

$$RR^{TE}_{a,a^*|\boldsymbol{c}} = \frac{P(Y(a) = 1|\boldsymbol{C} = \boldsymbol{c})}{P(Y(a^*) = 1|\boldsymbol{C} = \boldsymbol{c})}.$$

Pour établir un lien entre cet effet total conditionnel et les effets naturels correspondants (1.19), nous présumons que

$$Y_\omega(a, M(a)) = Y_\omega(a) \quad \forall a \in \mathcal{A}, \quad \forall \omega \in \Omega. \tag{1.33}$$

Dans la littérature, cette hypothèse est appelée l'hypothèse de composition (*composition assumption* ; Nguyen *et al.* (2022)).

Sous cette hypothèse, nous pouvons exprimer l'effet total $RR^{TE}_{a,a^*|\boldsymbol{c}}$ comme suit :

$$\begin{aligned}
RR^{TE}_{a,a^*|\boldsymbol{c}} &= \frac{P(Y(a) = 1|\boldsymbol{C} = \boldsymbol{c})}{P(Y(a^*) = 1|\boldsymbol{C} = \boldsymbol{c})} \\
&\overset{\text{par } (1.33)}{=} \frac{P(Y(a, M(a)) = 1|\boldsymbol{C} = \boldsymbol{c})}{P(Y(a^*, M(a^*)) = 1|\boldsymbol{C} = \boldsymbol{c})} \\
&= \frac{P(Y(a, M(a)) = 1|\boldsymbol{C} = \boldsymbol{c})}{P(Y(a, M(a^*)) = 1|\boldsymbol{C} = \boldsymbol{c})} \cdot \frac{P(Y(a, M(a^*)) = 1|\boldsymbol{C} = \boldsymbol{c})}{P(Y(a^*, M(a^*)) = 1|\boldsymbol{C} = \boldsymbol{c})} \\
&= RR^{NIE}_{a,a^*|\boldsymbol{c}} \cdot RR^{NDE}_{a,a^*|\boldsymbol{c}}.
\end{aligned}$$

De la même façon, nous pouvons décomposer les effets totaux conditionnels sur l'échelle du rapport de cotes et de la différence de risques :

$$OR^{TE}_{a,a^*|\boldsymbol{c}} = OR^{NIE}_{a,a^*|\boldsymbol{c}} \cdot OR^{NDE}_{a,a^*|\boldsymbol{c}},$$

$$RD^{TE}_{a,a^*|\boldsymbol{c}} = RD^{NIE}_{a,a^*|\boldsymbol{c}} + RD^{NDE}_{a,a^*|\boldsymbol{c}}.$$

1.4.4 Effet direct contrôlé et son identification

Considérons maintenant un effet populationnel qui est défini comme le contraste entre les espérances $E\{Y_\omega(a,m)\}$ et $E\{Y_\omega(a^*,m)\}$ des réponses contrefactuelles $Y_\omega(a,m)$ et $Y_\omega(a^*,m)$, $\omega \in \Omega$, $m \in \mathcal{M}$. Ce contraste est un estimand d'intérêt lorsque nous cherchons à évaluer l'effet du niveau de traitement $A = a$ versus le niveau $A = a^*$ si le médiateur M est fixé par une intervention hypothétique à une valeur spécifique $m \in \mathcal{M}$ pour toute la population étudiée Ω. Cet estimand, appelé l'effet direct contrôlé (*controlled direct effect*, CDE), est souvent considéré comme un concept important pour l'évaluation des politiques de santé publique (Naimi *et al.*, 2014; VanderWeele, 2013).

Pour Y binaire et la strate $\boldsymbol{C} = \boldsymbol{c}$, l'effet direct contrôlé conditionnel de $A = a$ versus $A = a^*$, correspondant à $m \in \mathcal{M}$, est défini comme le contraste entre les probabilités $P(Y(a,m) = 1|\boldsymbol{C} = \boldsymbol{c})$ et $P(Y(a^*,m) = 1|\boldsymbol{C} = \boldsymbol{c})$. Plus précisément, cet effet s'exprime sur les échelles binaires standards de la manière suivante :

$$RR^{CDE}_{a,a^*|\boldsymbol{c}}(m) = \frac{P(Y(a,m) = 1|\boldsymbol{C} = \boldsymbol{c})}{P(Y(a^*,m) = 1|\boldsymbol{C} = \boldsymbol{c})},$$

$$OR^{CDE}_{a,a^*|\boldsymbol{c}}(m) = \frac{\dfrac{P(Y(a,m) = 1|\boldsymbol{C} = \boldsymbol{c})}{1 - P(Y(a,m) = 1|\boldsymbol{C} = \boldsymbol{c})}}{\dfrac{P(Y(a^*,m) = 1|\boldsymbol{C} = \boldsymbol{c})}{1 - P(Y(a^*,m) = 1|\boldsymbol{C} = \boldsymbol{c})}},$$

$$RD^{CDE}_{a,a^*|\boldsymbol{c}}(m) = P(Y(a,m) = 1|\boldsymbol{C} = \boldsymbol{c}) - P(Y(a^*,m) = 1|\boldsymbol{C} = \boldsymbol{c}).$$

Sous les hypothèses (1.15), (1.25), (1.26) et (1.27), la probabilité contrefactuelle $P(Y(a,m) = 1|\boldsymbol{C} = \boldsymbol{c})$ est estimable en utilisant les données observées :

$$\begin{aligned} & P(Y(a,m) = 1|\boldsymbol{C} = \boldsymbol{c}) \\ \overset{\text{par (1.26,1.15)}}{=} & P(Y(a,m) = 1|A = a, \boldsymbol{C} = \boldsymbol{c}) \\ \overset{\text{par (1.27,1.25)}}{=} & P(Y(a,m) = 1|A = a, M = m, \boldsymbol{C} = \boldsymbol{c}) \\ \overset{\text{par (1.23)}}{=} & P(Y = 1|A = a, M = m, \boldsymbol{C} = \boldsymbol{c}). \end{aligned}$$

De la même façon,

$$P(Y(a^*,m) = 1|\boldsymbol{C} = \boldsymbol{c}) = P(Y = 1|A = a^*, M = m, \boldsymbol{C} = \boldsymbol{c}).$$

CHAPITRE II

MÉDIATION CAUSALE POUR RÉPONSE BINAIRE : TECHNIQUES D'ESTIMATION

Pour les variables réponses binaires, un certain nombre d'approches d'analyse de médiation causale sont à la disposition des chercheurs appliqués. Dans ce chapitre, nous présentons les techniques d'estimation approximatives des effets naturels de VanderWeele et Vansteelandt (2010) pour un médiateur continu et de Valeri et VanderWeele (2013) pour un médiateur binaire ; ces approches sont basées sur la régression et développées sous *l'hypothèse de la réponse rare*. Pour le cas d'un médiateur continu, nous présentons les estimateurs approximatifs de Gaynor *et al.* (2019) pour les effets naturels direct et indirect, également basés sur la régression, mais dérivés sous *l'hypothèse de la réponse commune*. Deux approches d'estimation fréquemment utilisées, qui ne s'appuient pas sur les hypothèses de la réponse rare ou commune, notamment celles d'Imai *et al.* (2010) et de Lange *et al.* (2012), sont également présentées.

2.1 Techniques d'estimation approximatives pour un médiateur continu et une réponse binaire

2.1.1 Approche de VanderWeele et Vansteelandt (2010)

Dans le contexte d'une variable réponse binaire Y et d'un médiateur continu M, la formule (1.32) implique que la probabilité emboîtée conditionnelle $P(Y(a, M(a^*)) = 1|\boldsymbol{C} = \boldsymbol{c})$ s'exprime par l'équation suivante :

$$P(Y(a, M(a^*)) = 1|\boldsymbol{C} = \boldsymbol{c}) = \int_{\mathcal{M}} P(Y = 1|A = a, M = m, \boldsymbol{C} = \boldsymbol{c})\, dF_{M|A=a^*,\boldsymbol{C}=\boldsymbol{c}}(m), \qquad (2.1)$$

où $\mathcal{M}$ est le support du médiateur M et $F_{M|A=a^*,\boldsymbol{C}=\boldsymbol{c}}(m)$ est la fonction de répartition conditionnelle de M lorsque $A = a^*$ et $\boldsymbol{C} = \boldsymbol{c}$.

Sous les hypothèses de modélisation

$$M|A=a,\boldsymbol{C}=\boldsymbol{c}\sim\mathcal{N}\Big(\beta_0+\beta_1 a+\boldsymbol{\beta}_2^{'}\boldsymbol{c},\sigma^2\Big), \tag{2.2}$$

$$\text{logit}\{P(Y=1|A=a,M=m,\boldsymbol{C}=\boldsymbol{c})\}=\theta_0+\theta_1 a+\theta_2 m+\theta_3 am+\boldsymbol{\theta}_4^{'}\boldsymbol{c}, \tag{2.3}$$

nous avons l'expression suivante pour la probabilité emboîtée (2.1) :

$$\begin{aligned}P(Y(a,M(a^*))=1|\boldsymbol{C}=\boldsymbol{c})=\frac{1}{\sqrt{2\pi\sigma^2}}\int_{-\infty}^{\infty}&\text{expit}\left(\theta_0+\theta_1 a+\theta_2 m+\theta_3 am+\boldsymbol{\theta}_4^{'}\boldsymbol{c}\right)\\&\times\exp\left(-\frac{\Big(m-\Big(\beta_0+\beta_1 a^*+\boldsymbol{\beta}_2^{'}\boldsymbol{c}\Big)\Big)^2}{2\sigma^2}\right)dm,\end{aligned} \tag{2.4}$$

où

$$\text{expit}(x)=\frac{\exp(x)}{1+\exp(x)},\quad x\in\mathbb{R}.$$

La partie droite de l'égalité (2.4) présente l'espérance de la loi logit-normale (voir l'expression (A.4) dans l'Annexe A) pour laquelle une forme analytique fermée n'existe pas (Frederic et Lad, 2008). Ainsi, une intégration numérique est nécessaire [1].

Pour dériver des expressions fermées pour les effets naturels $OR^{NDE}_{a,a^*|\boldsymbol{c}}$ et $OR^{NIE}_{a,a^*|\boldsymbol{c}}$ (1.20), VanderWeele et Vansteelandt (2010) ont approximé des rapports de cotes par des rapports de risques, en utilisant ladite *l'hypothèse de la réponse rare* :

$$\begin{aligned}OR^{NDE}_{a,a^*|\boldsymbol{c}}&\approx\frac{P(Y(a,M(a^*))=1|\boldsymbol{C}=\boldsymbol{c})}{P(Y(a^*,M(a^*))=1|\boldsymbol{C}=\boldsymbol{c})},\\OR^{NIE}_{a,a^*|\boldsymbol{c}}&\approx\frac{P(Y(a,M(a))=1|\boldsymbol{C}=\boldsymbol{c})}{P(Y(a,M(a^*))=1|\boldsymbol{C}=\boldsymbol{c})}.\end{aligned} \tag{2.5}$$

Par la suite, toujours sous l'hypothèse de la réponse rare, le terme

$$\text{expit}\left(\theta_0+\theta_1 a+\theta_2 m+\theta_3 am+\boldsymbol{\theta}_4^{'}\boldsymbol{c}\right)$$

dans l'expression (2.4) a été remplacé par

$$\exp\left(\theta_0+\theta_1 a+\theta_2 m+\theta_3 am+\boldsymbol{\theta}_4^{'}\boldsymbol{c}\right),$$

1. Les résultats présentés par Holmes et Schofield (2022) permettent d'éviter l'intégration numérique pour calculer la probabilité emboîtée (2.4) ; voir l'Annexe A.

et, en conséquence, VanderWeele et Vansteelandt (2010) ont obtenu l'approximation suivante pour la probabilité (2.4) :

$$
\begin{aligned}
&app.P(Y(a, M(a^*)) = 1|\boldsymbol{C} = \boldsymbol{c}) \\
&\quad = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} \exp\left(\theta_0 + \theta_1 a + \theta_2 m + \theta_3 am + \boldsymbol{\theta}_4^{'}\boldsymbol{c}\right) \exp\left(-\frac{\left(m - \left(\beta_0 + \beta_1 a^* + \boldsymbol{\beta}_2^{'}\boldsymbol{c}\right)\right)^2}{2\sigma^2}\right) dm \\
&\quad = \frac{\exp\left(\theta_0 + \theta_1 a + \boldsymbol{\theta}_4^{'}\boldsymbol{c}\right)}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp\left((\theta_2 + \theta_3 a)m\right) \\
&\qquad \times \exp\left(-\frac{\left(m - \left(\beta_0 + \beta_1 a^* + \boldsymbol{\beta}_2^{'}\boldsymbol{c}\right)\right)^2}{2\sigma^2}\right) d\left(\frac{m - \left(\beta_0 + \beta_1 a^* + \boldsymbol{\beta}_2^{'}\boldsymbol{c}\right)}{\sigma}\right). \qquad (2.6)
\end{aligned}
$$

Le changement de la variable d'intégration

$$
t = \frac{m - \left(\beta_0 + \beta_1 a^* + \boldsymbol{\beta}_2^{'}\boldsymbol{c}\right)}{\sigma} \quad \Longrightarrow \quad m = \sigma t + \left(\beta_0 + \beta_1 a^* + \boldsymbol{\beta}_2^{'}\boldsymbol{c}\right)
$$

dans l'expression (2.6) implique que

$$
\begin{aligned}
&app.P(Y(a, M(a^*)) = 1|\boldsymbol{C} = \boldsymbol{c}) \\
&\quad = \frac{\exp\left(\theta_0 + \theta_1 a + \boldsymbol{\theta}_4^{'}\boldsymbol{c}\right)}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp\left((\theta_2 + \theta_3 a)\left(\sigma t + \beta_0 + \beta_1 a^* + \boldsymbol{\beta}_2^{'}\boldsymbol{c}\right)\right) \exp\left(-\frac{t^2}{2}\right) dt \\
&\quad = \frac{\exp\left(\theta_0 + \theta_1 a + \boldsymbol{\theta}_4^{'}\boldsymbol{c} + (\theta_2 + \theta_3 a)\left(\beta_0 + \beta_1 a^* + \boldsymbol{\beta}_2^{'}\boldsymbol{c}\right)\right)}{\sqrt{2\pi}} \\
&\qquad \times \int_{-\infty}^{\infty} \exp\left((\theta_2 + \theta_3 a)\sigma t\right) \exp\left(-\frac{t^2}{2}\right) dt \\
&\quad = \frac{\exp\left(\theta_0 + \theta_1 a + \boldsymbol{\theta}_4^{'}\boldsymbol{c} + (\theta_2 + \theta_3 a)\left(\beta_0 + \beta_1 a^* + \boldsymbol{\beta}_2^{'}\boldsymbol{c}\right) + \frac{(\theta_2 + \theta_3 a)^2\sigma^2}{2}\right)}{\sqrt{2\pi}} \\
&\qquad \times \int_{-\infty}^{\infty} \exp\left(-\frac{t^2 - 2(\theta_2 + \theta_3 a)\sigma t + (\theta_2 + \theta_3 a)^2\sigma^2}{2}\right) dt \\
&\quad = \exp\left(\theta_0 + \theta_1 a + \boldsymbol{\theta}_4^{'}\boldsymbol{c} + (\theta_2 + \theta_3 a)\left(\beta_0 + \beta_1 a^* + \boldsymbol{\beta}_2^{'}\boldsymbol{c}\right) + \frac{(\theta_2 + \theta_3 a)^2\sigma^2}{2}\right) \\
&\qquad \times \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp\left(-\frac{\left(t - (\theta_2 + \theta_3 a)\sigma\right)^2}{2}\right) dt
\end{aligned}
$$

$$= \exp\left(\theta_0 + \theta_1 a + \boldsymbol{\theta}_4^{'}\boldsymbol{c} + (\theta_2 + \theta_3 a)\left(\beta_0 + \beta_1 a^* + \boldsymbol{\beta}_2^{'}\boldsymbol{c}\right) + \frac{(\theta_2 + \theta_3 a)^2\sigma^2}{2}\right).$$

Ainsi, selon l'approche de VanderWeele et Vansteelandt (2010),

$$\begin{aligned} P(Y(a, M(a^*)) = 1|\boldsymbol{C} = \boldsymbol{c}) &\approx app.P(Y(a, M(a^*)) = 1|\boldsymbol{C} = \boldsymbol{c}) \\ &= \exp\left(\theta_0 + \theta_1 a + \boldsymbol{\theta}_4^{'}\boldsymbol{c} + (\theta_2 + \theta_3 a)\left(\beta_0 + \beta_1 a^* + \boldsymbol{\beta}_2^{'}\boldsymbol{c}\right) + \frac{(\theta_2 + \theta_3 a)^2\sigma^2}{2}\right), \end{aligned}$$

ce qui entraîne, conjointement avec (2.5), les approximations suivantes pour $OR^{NDE}_{a,a^*|\boldsymbol{c}}$ et $OR^{NIE}_{a,a^*|\boldsymbol{c}}$ (1.20) :

$$\begin{aligned} app.OR^{NDE}_{a,a^*|\boldsymbol{c}} &= \frac{app.P(Y(a, M(a^*)) = 1|\boldsymbol{C} = \boldsymbol{c})}{app.P(Y(a^*, M(a^*)) = 1|\boldsymbol{C} = \boldsymbol{c})} \\ &= \frac{\exp\left(\theta_0 + \theta_1 a + \boldsymbol{\theta}_4^{'}\boldsymbol{c} + (\theta_2 + \theta_3 a)\left(\beta_0 + \beta_1 a^* + \boldsymbol{\beta}_2^{'}\boldsymbol{c}\right) + \frac{(\theta_2 + \theta_3 a)^2\sigma^2}{2}\right)}{\exp\left(\theta_0 + \theta_1 a^* + \boldsymbol{\theta}_4^{'}\boldsymbol{c} + (\theta_2 + \theta_3 a^*)\left(\beta_0 + \beta_1 a^* + \boldsymbol{\beta}_2^{'}\boldsymbol{c}\right) + \frac{(\theta_2 + \theta_3 a^*)^2\sigma^2}{2}\right)} \\ &= \exp\left[\left(\theta_1 + \theta_3\left(\beta_0 + \beta_1 a^* + \boldsymbol{\beta}_2^{'}\boldsymbol{c}\right) + \theta_2\theta_3\sigma^2\right)(a - a^*) + 0.5\theta_3^2\sigma^2(a^2 - a^{*2})\right], \end{aligned} \tag{2.7}$$

$$\begin{aligned} app.OR^{NIE}_{a,a^*|\boldsymbol{c}} &= \frac{app.P(Y(a, M(a)) = 1|\boldsymbol{C} = \boldsymbol{c})}{app.P(Y(a, M(a^*)) = 1|\boldsymbol{C} = \boldsymbol{c})} \\ &= \frac{\exp\left(\theta_0 + \theta_1 a + \boldsymbol{\theta}_4^{'}\boldsymbol{c} + (\theta_2 + \theta_3 a)\left(\beta_0 + \beta_1 a + \boldsymbol{\beta}_2^{'}\boldsymbol{c}\right) + \frac{(\theta_2 + \theta_3 a)^2\sigma^2}{2}\right)}{\exp\left(\theta_0 + \theta_1 a + \boldsymbol{\theta}_4^{'}\boldsymbol{c} + (\theta_2 + \theta_3 a)\left(\beta_0 + \beta_1 a^* + \boldsymbol{\beta}_2^{'}\boldsymbol{c}\right) + \frac{(\theta_2 + \theta_3 a)^2\sigma^2}{2}\right)} \\ &= \exp\left[\beta_1(\theta_2 + \theta_3 a)(a - a^*)\right]. \end{aligned} \tag{2.8}$$

Les estimateurs approximatifs de VanderWeele et Vansteelandt (2010) sont obtenus directement des expressions (2.7-2.8) en remplaçant $\boldsymbol{\beta}$, $\boldsymbol{\theta}$ et σ^2 par les estimateurs $\widehat{\boldsymbol{\beta}}$, $\widehat{\boldsymbol{\theta}}$ et $\widehat{\sigma^2}$.

Les expressions (2.7-2.8) impliquent que

$$\ln\left(app.OR^{NDE}_{a,a^*|\boldsymbol{c}}\right) = \left(\theta_1 + \theta_3\left(\beta_0 + \beta_1 a^* + \boldsymbol{\beta}_2^{'}\boldsymbol{c}\right) + \theta_2\theta_3\sigma^2\right)(a - a^*) + 0.5\theta_3^2\sigma^2(a^2 - a^{*2})$$

et

$$\ln\left(app.OR^{NIE}_{a,a^*|\boldsymbol{c}}\right) = \beta_1(\theta_2 + \theta_3 a)(a - a^*).$$

Soit k la longueur du vecteur $\boldsymbol{c}$. Notons

$$L_{NDE} = \ln\left(app.OR^{NDE}_{a,a^*|\boldsymbol{c}}\right), \quad L_{NIE} = \ln\left(app.OR^{NIE}_{a,a^*|\boldsymbol{c}}\right).$$

Nous avons que

$$\begin{aligned}
&\frac{\partial L_{NDE}}{\partial \beta_0} = \theta_3(a - a^*),\\
&\frac{\partial L_{NDE}}{\partial \beta_1} = a^* \, \frac{\partial L_{NDE}}{\partial \beta_0},\\
&\frac{\partial L_{NDE}}{\partial \beta_{2i}} = c_i \, \frac{\partial L_{NDE}}{\partial \beta_0}, \quad i = 1, 2, \ldots, k,\\
&\frac{\partial L_{NDE}}{\partial \theta_0} \equiv 0,\\
&\frac{\partial L_{NDE}}{\partial \theta_1} = a - a^*,\\
&\frac{\partial L_{NDE}}{\partial \theta_2} = \theta_3\sigma^2(a - a^*),\\
&\frac{\partial L_{NDE}}{\partial \theta_3} = \left(\beta_0 + \beta_1 a^* + \boldsymbol{\beta}_2'\boldsymbol{c} + \theta_2\sigma^2\right)(a - a^*) + \theta_3\sigma^2(a^2 - a^{*2}),\\
&\frac{\partial L_{NDE}}{\partial \theta_{4i}} \equiv 0, \quad i = 1, 2, \ldots, k,\\
&\frac{\partial L_{NDE}}{\partial \sigma^2} = \theta_2\theta_3(a - a^*) + 0.5\theta_3^2(a^2 - a^{*2}),
\end{aligned}$$

et

$$\begin{aligned}
&\frac{\partial L_{NIE}}{\partial \beta_0} \equiv 0,\\
&\frac{\partial L_{NIE}}{\partial \beta_1} = (\theta_2 + \theta_3 a)(a - a^*),\\
&\frac{\partial L_{NIE}}{\partial \beta_{2i}} \equiv 0, \quad i = 1, 2, \ldots, k,\\
&\frac{\partial L_{NIE}}{\partial \theta_0} \equiv 0, \quad \frac{\partial L_{NIE}}{\partial \theta_1} \equiv 0,\\
&\frac{\partial L_{NIE}}{\partial \theta_2} = \beta_1(a - a^*),\\
&\frac{\partial L_{NIE}}{\partial \theta_3} = a \, \frac{\partial L_{NIE}}{\partial \theta_2},\\
&\frac{\partial L_{NIE}}{\partial \theta_{4i}} \equiv 0, \quad i = 1, 2, \ldots, k,\\
&\frac{\partial L_{NIE}}{\partial \sigma^2} \equiv 0.
\end{aligned}$$

Les gradients des fonctions $L_{NDE}(a, a^*, \boldsymbol{c})$ et $L_{NIE}(a, a^*, \boldsymbol{c})$ par rapport au vecteur $(\boldsymbol{\beta}, \boldsymbol{\theta}, \sigma^2)$ sont respectivement les vecteurs

$$\nabla\left(L_{NDE}\right) = \left(\frac{\partial L_{NDE}}{\partial \boldsymbol{\beta}}, \frac{\partial L_{NDE}}{\partial \boldsymbol{\theta}}, \frac{\partial L_{NDE}}{\partial \sigma^2}\right)'$$

et

$$\nabla\left(L_{NIE}\right) = \left(\frac{\partial L_{NIE}}{\partial \boldsymbol{\beta}}, \frac{\partial L_{NIE}}{\partial \boldsymbol{\theta}}, \frac{\partial L_{NIE}}{\partial \sigma^2}\right)',$$

où

$$\begin{aligned}
\frac{\partial L_{NDE}}{\partial \boldsymbol{\beta}} &= \left(\frac{\partial L_{NDE}}{\partial \beta_0}, \frac{\partial L_{NDE}}{\partial \beta_1}, \frac{\partial L_{NDE}}{\partial \beta_{21}}, \ldots, \frac{\partial L_{NDE}}{\partial \beta_{2k}}\right), \\
\frac{\partial L_{NDE}}{\partial \boldsymbol{\theta}} &= \left(\frac{\partial L_{NDE}}{\partial \theta_0}, \frac{\partial L_{NDE}}{\partial \theta_1}, \frac{\partial L_{NDE}}{\partial \theta_2}, \frac{\partial L_{NDE}}{\partial \theta_3}, \frac{\partial L_{NDE}}{\partial \theta_{41}}, \ldots, \frac{\partial L_{NDE}}{\partial \theta_{4k}}\right), \\
\frac{\partial L_{NIE}}{\partial \boldsymbol{\beta}} &= \left(\frac{\partial L_{NIE}}{\partial \beta_0}, \frac{\partial L_{NIE}}{\partial \beta_1}, \frac{\partial L_{NIE}}{\partial \beta_{21}}, \ldots, \frac{\partial L_{NIE}}{\partial \beta_{2k}}\right), \\
\frac{\partial L_{NIE}}{\partial \boldsymbol{\theta}} &= \left(\frac{\partial L_{NIE}}{\partial \theta_0}, \frac{\partial L_{NIE}}{\partial \theta_1}, \frac{\partial L_{NIE}}{\partial \theta_2}, \frac{\partial L_{NIE}}{\partial \theta_3}, \frac{\partial L_{NIE}}{\partial \theta_{41}}, \ldots, \frac{\partial L_{NIE}}{\partial \theta_{4k}}\right).
\end{aligned}$$

Notons

$$\nabla\left(\widehat{L_{NDE}}\right) = \nabla\left(L_{NDE}\right)\Big|_{(\boldsymbol{\beta},\boldsymbol{\theta},\sigma^2)=(\widehat{\boldsymbol{\beta}},\widehat{\boldsymbol{\theta}},\widehat{\sigma^2})}$$

et

$$\nabla\left(\widehat{L_{NIE}}\right) = \nabla\left(L_{NIE}\right)\Big|_{(\boldsymbol{\beta},\boldsymbol{\theta},\sigma^2)=(\widehat{\boldsymbol{\beta}},\widehat{\boldsymbol{\theta}},\widehat{\sigma^2})}.$$

VanderWeele et Vansteelandt (2010) ont exprimé des erreurs standards pour $\ln\left(\widehat{app.OR^{NDE}_{a,a^*|\boldsymbol{c}}}\right)$ et $\ln\left(\widehat{app.OR^{NIE}_{a,a^*|\boldsymbol{c}}}\right)$ en se basant sur la méthode delta du premier ordre (Casella et Berger, 2002; Fay et Brittain, 2022) :

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

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

où $\widehat{\Sigma} = diag\{\widehat{\Sigma}_{\widehat{\boldsymbol{\beta}}}, \widehat{\Sigma}_{\widehat{\boldsymbol{\theta}}}, \widehat{\Sigma}_{\widehat{\sigma^2}}\}$ est une matrice à blocs, dont les blocs $\widehat{\Sigma}_{\widehat{\boldsymbol{\beta}}}$, $\widehat{\Sigma}_{\widehat{\boldsymbol{\theta}}}$, et $\widehat{\Sigma}_{\widehat{\sigma^2}}$ sont respectivement les estimateurs des matrices de covariance de $\widehat{\boldsymbol{\beta}}$, $\widehat{\boldsymbol{\theta}}$, et $\widehat{\sigma^2}$.

Ainsi, les intervalles de confiance à 95% pour $\ln\left(app.OR^{NDE}_{a,a^*|\boldsymbol{c}}\right)$ et $\ln\left(app.OR^{NIE}_{a,a^*|\boldsymbol{c}}\right)$ sont respecti-

vement approximés par

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

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

où $\Phi(\cdot)$ est la fonction de répartition de la loi normale standard $\mathcal{N}(0,1)$.

Par conséquent, les intervalles de confiance à 95% approximés pour $app.OR^{NDE}_{a,a^*|\boldsymbol{c}}$ et $app.OR^{NIE}_{a,a^*|\boldsymbol{c}}$ sont

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

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

La méthode *bootstrap* (Efron et Tibshirani, 1994) peut alternativement être utilisée pour construire des intervalles de confiance à 95% pour $app.OR^{NDE}_{a,a^*|\boldsymbol{c}}$ et $app.OR^{NIE}_{a,a^*|\boldsymbol{c}}$ (2.7-2.8).

2.1.2 Approche de Gaynor *et al.* (2019)

Toujours dans le contexte d'une variable réponse binaire et d'un médiateur continu, Gaynor *et al.* (2019) ont dérivé les expressions approximatives en forme fermée pour les effets naturels conditionnels $OR^{NDE}_{a,a^*|\boldsymbol{c}}$ et $OR^{NIE}_{a,a^*|\boldsymbol{c}}$ (1.20) en utilisant le lien approximatif suivant entre la fonction logit et la fonction de répartition de la loi normale standard $\mathcal{N}(0,1)$:

$$\text{logit}^{-1}(x) = \text{expit}(x) \approx \Phi(sx), \quad \forall x \in \mathbb{R}, \tag{2.9}$$

où $\text{logit}(p) = \ln\big[p/(1-p)\big]$, $p \in (0,1)$, $s > 0$.

Pour la spécification du paramètre s, de nombreuses approches ont été proposées dans la littérature ; nous en faisons une revue descriptive, en incluant aussi l'approche empirique de Gaynor *et al.* (2019), à la fin de cette sous-section (voir aussi le Chapitre V).

En se basant sur le lien (2.9), Gaynor *et al.* (2019) ont approximé la probabilité emboîtée contrefactuelle (2.4) par

$$app.P(Y(a, M(a^*)) = 1|\boldsymbol{C} = \boldsymbol{c})$$

$$= \int_{-\infty}^{\infty} \Phi\left(s\left(\theta_0 + \theta_1 a + \theta_2 m + \theta_3 am + \boldsymbol{\theta}_4^{'}\boldsymbol{c}\right)\right) \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{\left(m - \left(\beta_0 + \beta_1 a^* + \boldsymbol{\beta}_2^{'}\boldsymbol{c}\right)\right)^2}{2\sigma^2}\right) dm$$

$$= \int_{-\infty}^{\infty} \Phi\left(s\left(\theta_0 + \theta_1 a + \boldsymbol{\theta}_4^{'}\boldsymbol{c}\right) + s\left(\theta_2 + \theta_3 a\right) m\right)$$

$$\times \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{\left(m - \left(\beta_0 + \beta_1 a^* + \boldsymbol{\beta}_2^{'}\boldsymbol{c}\right)\right)^2}{2\sigma^2}\right) d\left(\frac{m - \left(\beta_0 + \beta_1 a^* + \boldsymbol{\beta}_2^{'}\boldsymbol{c}\right)}{\sigma}\right). \tag{2.10}$$

Sous le changement de la variable d'intégration

$$t = \frac{m - \left(\beta_0 + \beta_1 a^* + \boldsymbol{\beta}_2^{'}\boldsymbol{c}\right)}{\sigma},$$

impliquant que

$$m = \sigma t + \left(\beta_0 + \beta_1 a^* + \boldsymbol{\beta}_2^{'}\boldsymbol{c}\right),$$

nous pouvons réécrire l'expression (2.10) comme suit :

$$app.P(Y(a, M(a^*)) = 1 | \boldsymbol{C} = \boldsymbol{c})$$

$$= \int_{-\infty}^{\infty} \Phi\left(s\left(\theta_0 + \theta_1 a + \boldsymbol{\theta}_4^{'}\boldsymbol{c}\right) + s\left(\theta_2 + \theta_3 a\right)\left(\sigma t + \beta_0 + \beta_1 a^* + \boldsymbol{\beta}_2^{'}\boldsymbol{c}\right)\right)$$

$$\times \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{\left(m - \left(\beta_0 + \beta_1 a^* + \boldsymbol{\beta}_2^{'}\boldsymbol{c}\right)\right)^2}{2\sigma^2}\right) d\left(\frac{m - \left(\beta_0 + \beta_1 a^* + \boldsymbol{\beta}_2^{'}\boldsymbol{c}\right)}{\sigma}\right)$$

$$= \int_{-\infty}^{\infty} \Phi\left(s\left(\theta_0 + \theta_1 a + \boldsymbol{\theta}_4^{'}\boldsymbol{c} + \left(\theta_2 + \theta_3 a\right)\left(\beta_0 + \beta_1 a^* + \boldsymbol{\beta}_2^{'}\boldsymbol{c}\right)\right) + s\sigma\left(\theta_2 + \theta_3 a\right) t\right) \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{t^2}{2}\right) dt.$$

L'équation suivante (Patel et Read, 1996, p. 36)

$$\int_{-\infty}^{\infty} \Phi\left(\alpha + \gamma m\right) \phi(m) dm = \Phi\left(\frac{\alpha}{\sqrt{1+\gamma^2}}\right)$$

permet d'exprimer $app.P(Y(a, M(a^*)) = 1 | \boldsymbol{C} = \boldsymbol{c})$ comme suit :

$$app.P(Y(a, M(a^*)) = 1 | \boldsymbol{C} = \boldsymbol{c}) = \Phi\left(\frac{s\left(\theta_0 + \theta_1 a + \boldsymbol{\theta}_4^{'}\boldsymbol{c} + \left(\theta_2 + \theta_3 a\right)\left(\beta_0 + \beta_1 a^* + \boldsymbol{\beta}_2^{'}\boldsymbol{c}\right)\right)}{\sqrt{1 + s^2\sigma^2\left(\theta_2 + \theta_3 a\right)^2}}\right).$$

En remplaçant les probabilités emboîtées contrefactuelles dans les expressions (1.20) par leurs contre-

parties approximatives, Gaynor *et al.* (2019) ont dérivé les expressions approximatives suivantes pour $OR^{NDE}_{a,a^*|\boldsymbol{c}}$ et $OR^{NIE}_{a,a^*|\boldsymbol{c}}$ (1.20) :

$$app.OR^{NDE}_{a,a^*|\boldsymbol{c}} = \frac{H\left\{\Phi\left(\frac{s\left(\theta_0+\theta_1 a+\boldsymbol{\theta}_4'\boldsymbol{c}+(\theta_2+\theta_3 a)\left(\beta_0+\beta_1 a^*+\boldsymbol{\beta}_2'\boldsymbol{c}\right)\right)}{\sqrt{1+s^2\sigma^2\left(\theta_2+\theta_3 a\right)^2}}\right)\right\}}{H\left\{\Phi\left(\frac{s\left(\theta_0+\theta_1 a^*+\boldsymbol{\theta}_4'\boldsymbol{c}+(\theta_2+\theta_3 a^*)\left(\beta_0+\beta_1 a^*+\boldsymbol{\beta}_2'\boldsymbol{c}\right)\right)}{\sqrt{1+s^2\sigma^2\left(\theta_2+\theta_3 a^*\right)^2}}\right)\right\}}, \tag{2.11}$$

$$app.OR^{NIE}_{a,a^*|\boldsymbol{c}} = \frac{H\left\{\Phi\left(\frac{s\left(\theta_0+\theta_1 a+\boldsymbol{\theta}_4'\boldsymbol{c}+(\theta_2+\theta_3 a)\left(\beta_0+\beta_1 a+\boldsymbol{\beta}_2'\boldsymbol{c}\right)\right)}{\sqrt{1+s^2\sigma^2\left(\theta_2+\theta_3 a\right)^2}}\right)\right\}}{H\left\{\Phi\left(\frac{s\left(\theta_0+\theta_1+\boldsymbol{\theta}_4'\boldsymbol{c}+(\theta_2+\theta_3 a)\left(\beta_0+\beta_1 a^*+\boldsymbol{\beta}_2'\boldsymbol{c}\right)\right)}{\sqrt{1+s^2\sigma^2\left(\theta_2+\theta_3 a\right)^2}}\right)\right\}}, \tag{2.12}$$

où $H(p) = p/(1-p)$, $p \in (0,1)$. Les estimateurs approximatifs de Gaynor *et al.* (2019) sont obtenus des expressions (2.11-2.12) en remplaçant $\boldsymbol{\beta}$, $\boldsymbol{\theta}$, σ^2 et s par les estimateurs correspondants $\widehat{\boldsymbol{\beta}}$, $\widehat{\boldsymbol{\theta}}$, $\widehat{\sigma^2}$ et $\widehat{s}$. Afin de construire les intervalles de confiance à 95% pour $app.OR^{NDE}_{a,a^*|\boldsymbol{c}}$ et $app.OR^{NIE}_{a,a^*|\boldsymbol{c}}$ (2.11-2.12), nous pouvons appliquer la méthode bootstrap. L'étude de simulation effectuée par Gaynor *et al.* (2019) a montré une performance adéquate de leurs estimateurs basés sur (2.11-2.12) dans les cas d'une *réponse binaire commune* (plus précisément, pour une réponse dont la prévalence était entre 20% et 60%).

Le paramètre s est crucial pour la qualité de l'approximation (2.9). Une approche numérique (Cox, 1970) pour évaluer s se base sur l'égalité approximative suivante des variances de la loi logistique généralisée, définie par la fonction de répartition $F_s(t) = 1/\left(1+exp(-t/s)\right)$, $s > 0$, et de la loi normale standard $\mathcal{N}(0,1)$:

$$\frac{\pi^2 s^2}{3} \approx 1 \quad \Longrightarrow \quad s \approx \sqrt{3}/\pi.$$

Une autre approche (Camilli, 1994) consiste à minimiser la distance maximale entre les courbes représentant les fonctions $F_s(t)$ et $\Phi(t)$, c'est-à-dire

$$s = \underset{s}{\operatorname{argmin}} \left\{ \max_t |F_s(t) - \Phi(t)| \right\}.$$

La solution minimax est $s \approx 1/1.70174$.

Savalei (2006) a proposé d'évaluer le paramètre s dans l'approximation (2.9) en minimisant la divergence de Kullback-Leibler (KL) de la loi logistique généralisée avec la fonction de répartition $F_s(t)$ par rapport à la loi normale standard $\mathcal{N}(0,1)$:

$$s = \underset{s}{\operatorname{argmin}}\, \mathrm{KL}(F_s, \Phi) = \underset{s}{\operatorname{argmin}} \int_{-\infty}^{\infty} \ln\left(\frac{\Phi'(t)}{F_s{}'(t)}\right) \Phi'(t)dt \quad \Longrightarrow \quad s \approx 1/1.749.$$

Savalei (2006) a constaté que, comparativement à la solution KL, la solution minimax avait mieux approximé la partie centrale de la courbe $\Phi(t)$ par la courbe $F_s(t)$, tandis que la solution KL avait entraîné la meilleure approximation des queues de la courbe $\Phi(t)$.

Gaynor *et al.* (2019) ont proposé l'algorithme empirique suivant afin d'estimer le paramètre s. Premièrement, on ajuste le modèle logistique (2.3) de la réponse ; deuxièmement, on ajuste le modèle probit

$$P(Y = 1|A = a, M = m, \boldsymbol{C} = \boldsymbol{c}) = \Phi(\gamma_0 + \gamma_1 a + \gamma_2 m + \gamma_3 am + \boldsymbol{\gamma}_4^{'}\boldsymbol{c}).$$

Finalement, l'estimateur de s est construit comme la médiane des ratios des estimateurs des coefficients du modèle probit aux estimateurs des coefficients correspondants du modèle logistique (2.3), c'est-à-dire

$$\widehat{s} = \text{médiane}\left\{\frac{\widehat{\gamma_0}}{\widehat{\theta_0}}, \frac{\widehat{\gamma_1}}{\widehat{\theta_1}}, \ldots, \frac{\widehat{\gamma_{41}}}{\widehat{\theta_{41}}}, \ldots, \frac{\widehat{\gamma_{4k}}}{\widehat{\theta_{4k}}}\right\}. \tag{2.13}$$

En plus du bootstrap, Gaynor *et al.* (2019) ont proposé d'utiliser la méthode delta afin de construire les intervalles de confiance à 95% pour $app.OR^{NDE}_{a,a^*|\boldsymbol{c}}$ et $app.OR^{NIE}_{a,a^*|\boldsymbol{c}}$ (2.11-2.12). Toutefois, leur procédure basée sur la méthode delta ne reflète pas le fait que le paramètre s est également estimé à partir des données en utilisant l'estimateur (2.13).

2.2 Approche d'estimation approximative de Valeri et VanderWeele (2013) pour un médiateur et une réponse binaires

La dérivation (1.30) implique que, dans le cas d'un médiateur M et d'une réponse Y binaires, la probabilité emboîtée conditionnelle $P(Y(a, M(a^*)) = 1|\boldsymbol{C} = \boldsymbol{c})$ s'exprime comme suit :

$$P(Y(a, M(a^*)) = 1|\boldsymbol{C} = \boldsymbol{c}) = \sum_{m=0,1} P(Y = 1|A = a, M = m, \boldsymbol{C} = \boldsymbol{c})P(M = m|A = a^*, \boldsymbol{C} = \boldsymbol{c}). \tag{2.14}$$

Sous les hypothèses de modélisation (2.3) et

$$\text{logit}\{P(M=1|A=a,\boldsymbol{C}=\boldsymbol{c})\}=\beta_0+\beta_1 a+\boldsymbol{\beta}_2^{'}\boldsymbol{c}, \tag{2.15}$$

la probabilité emboîtée (2.14) s'exprime comme suit :

$$\begin{aligned}
&P(Y(a,M(a^*))=1|\boldsymbol{C}=\boldsymbol{c})\\
&\quad=\text{expit}\left(\theta_0+\theta_1 a+\theta_2+\theta_3 a+\boldsymbol{\theta}_4^{'}\boldsymbol{c}\right)\text{expit}\left(\beta_0+\beta_1 a^*+\boldsymbol{\beta}_2^{'}\boldsymbol{c}\right)\\
&\qquad+\text{expit}\left(\theta_0+\theta_1 a+\boldsymbol{\theta}_4^{'}\boldsymbol{c}\right)\left(1-\text{expit}\left(\beta_0+\beta_1 a^*+\boldsymbol{\beta}_2^{'}\boldsymbol{c}\right)\right).
\end{aligned} \tag{2.16}$$

Valeri et VanderWeele (2013) ont utilisé les idées principales de l'approche approximative de VanderWeele et Vansteelandt (2010), présentée dans la Sous-section 2.1.1, afin d'obtenir des expressions fermées simples des effets naturels pour le cas d'un médiateur et d'une réponse binaires. Notamment, sous l'hypothèse de la réponse rare, les approximations (2.5) ont été appliquées, et les expressions

$$\text{expit}\left(\theta_0+\theta_1 am+\theta_2 m+\theta_3 a+\boldsymbol{\theta}_4^{'}\boldsymbol{c}\right),\quad m=0,1,$$

dans (2.16) ont été remplacées par

$$\exp\left(\theta_0+\theta_1 a+\theta_2 m+\theta_3 am+\boldsymbol{\theta}_4^{'}\boldsymbol{c}\right),\quad m=0,1.$$

En conséquence de ces simplifications, la probabilité (2.16) est approximée par

$$\begin{aligned}
&app.P(Y(a,M(a^*))=1|\boldsymbol{C}=\boldsymbol{c})\\
&\quad=\exp\left(\theta_0+\theta_1 a+\theta_2+\theta_3 a+\boldsymbol{\theta}_4^{'}\boldsymbol{c}\right)\text{expit}\left(\beta_0+\beta_1 a^*+\boldsymbol{\beta}_2^{'}\boldsymbol{c}\right)\\
&\qquad+\exp\left(\theta_0+\theta_1 a+\boldsymbol{\theta}_4^{'}\boldsymbol{c}\right)\left(1-\text{expit}\left(\beta_0+\beta_1 a^*+\boldsymbol{\beta}_2^{'}\boldsymbol{c}\right)\right)\\
&\quad=\frac{\exp\left(\theta_0+\theta_1 a+\boldsymbol{\theta}_4^{'}\boldsymbol{c}\right)\left[\exp\left(\theta_2+\theta_3 a+\beta_0+\beta_1 a^*+\boldsymbol{\beta}_2^{'}\boldsymbol{c}\right)+1\right]}{1+\exp\left(\beta_0+\beta_1 a^*+\boldsymbol{\beta}_2^{'}\boldsymbol{c}\right)}.
\end{aligned}$$

Ainsi, nous avons, conjointement avec (2.5), les expressions approximatives suivantes pour $OR^{NDE}_{a,a^*|\boldsymbol{c}}$

et $OR^{NIE}_{a,a^*|\boldsymbol{c}}$ (1.20) :

$$
\begin{aligned}
app.OR^{NDE}_{a,a^*|\boldsymbol{c}} &= \frac{app.P(Y(a,M(a^*))=1|\boldsymbol{C}=\boldsymbol{c})}{app.P(Y(a^*,M(a^*))=1|\boldsymbol{C}=\boldsymbol{c})} \\
&= \frac{\dfrac{\exp\left(\theta_0+\theta_1 a+\boldsymbol{\theta}_4^{'}\boldsymbol{c}\right)\left[\exp\left(\theta_2+\theta_3 a+\beta_0+\beta_1 a^*+\boldsymbol{\beta}_2^{'}\boldsymbol{c}\right)+1\right]}{1+\exp\left(\beta_0+\beta_1 a^*+\boldsymbol{\beta_2^{'}}\boldsymbol{c}\right)}}{\dfrac{\exp\left(\theta_0+\theta_1 a^*+\boldsymbol{\theta}_4^{'}\boldsymbol{c}\right)\left[\exp\left(\theta_2+\theta_3 a^*+\beta_0+\beta_1 a^*+\boldsymbol{\beta}_2^{'}\boldsymbol{c}\right)+1\right]}{1+\exp\left(\beta_0+\beta_1 a^*+\boldsymbol{\beta_2^{'}}\boldsymbol{c}\right)}} \\
&= \exp\left(\theta_1(a-a^*)\right)\frac{\exp\left(\theta_2+\theta_3 a+\beta_0+\beta_1 a^*+\boldsymbol{\beta}_2^{'}\boldsymbol{c}\right)+1}{\exp\left(\theta_2+\theta_3 a^*+\beta_0+\beta_1 a^*+\boldsymbol{\beta}_2^{'}\boldsymbol{c}\right)+1},
\end{aligned}
\tag{2.17}
$$

$$
\begin{aligned}
app.OR^{NIE}_{a,a^*|\boldsymbol{c}} &= \frac{app.P(Y(a,M(a))=1|\boldsymbol{C}=\boldsymbol{c})}{app.P(Y(a,M(a^*))=1|\boldsymbol{C}=\boldsymbol{c})} \\
&= \frac{\dfrac{\exp\left(\theta_0+\theta_1 a+\boldsymbol{\theta}_4^{'}\boldsymbol{c}\right)\left[\exp\left(\theta_2+\theta_3 a+\beta_0+\beta_1 a+\boldsymbol{\beta}_2^{'}\boldsymbol{c}\right)+1\right]}{1+\exp\left(\beta_0+\beta_1 a+\boldsymbol{\beta}_2^{'}\boldsymbol{c}\right)}}{\dfrac{\exp\left(\theta_0+\theta_1 a+\boldsymbol{\theta}_4^{'}\boldsymbol{c}\right)\left[\exp\left(\theta_2+\theta_3 a+\beta_0+\beta_1 a^*+\boldsymbol{\beta}_2^{'}\boldsymbol{c}\right)+1\right]}{1+\exp\left(\beta_0+\beta_1 a^*+\boldsymbol{\beta_2^{'}}\boldsymbol{c}\right)}} \\
&= \frac{\exp\left(\beta_0+\beta_1 a^*+\boldsymbol{\beta}_2^{'}\boldsymbol{c}\right)+1}{\exp\left(\beta_0+\beta_1 a+\boldsymbol{\beta}_2^{'}\boldsymbol{c}\right)+1}\cdot\frac{\exp\left(\theta_2+\theta_3 a+\beta_0+\beta_1 a+\boldsymbol{\beta}_2^{'}\boldsymbol{c}\right)+1}{\exp\left(\theta_2+\theta_3 a+\beta_0+\beta_1 a^*+\boldsymbol{\beta}_2^{'}\boldsymbol{c}\right)+1}.
\end{aligned}
\tag{2.18}
$$

Soient

$$
L_{NDE}=\ln\left(app.OR^{NDE}_{a,a^*|\boldsymbol{c}}\right),\quad L_{NIE}=\ln\left(app.OR^{NIE}_{a,a^*|\boldsymbol{c}}\right).
$$

Nous avons que

$$
\begin{aligned}
L_{NDE} &= \theta_1(a-a^*)+\ln\left[\exp\left(\theta_2+\theta_3 a+\beta_0+\beta_1 a^*+\boldsymbol{\beta}_2^{'}\boldsymbol{c}\right)+1\right] \\
&\quad -\ln\left[\exp\left(\theta_2+\theta_3 a^*+\beta_0+\beta_1 a^*+\boldsymbol{\beta}_2^{'}\boldsymbol{c}\right)+1\right],
\end{aligned}
$$

$$
\begin{aligned}
L_{NIE} &= \ln\left[\exp\left(\beta_0+\beta_1 a^*+\boldsymbol{\beta}_2^{'}\boldsymbol{c}\right)+1\right]-\ln\left[\exp\left(\beta_0+\beta_1 a+\boldsymbol{\beta}_2^{'}\boldsymbol{c}\right)+1\right] \\
&\quad +\ln\left[\exp\left(\theta_2+\theta_3 a+\beta_0+\beta_1 a+\boldsymbol{\beta}_2^{'}\boldsymbol{c}\right)+1\right] \\
&\quad -\ln\left[\exp\left(\theta_2+\theta_3 a+\beta_0+\beta_1 a^*+\boldsymbol{\beta}_2^{'}\boldsymbol{c}\right)+1\right],
\end{aligned}
$$

et, conséquemment,

$$
\begin{aligned}
&\frac{\partial L_{NDE}}{\partial \beta_0} = \operatorname{expit}\left(\theta_2 + \theta_3 a + \beta_0 + \beta_1 a^* + \boldsymbol{\beta}_2^{'}\boldsymbol{c}\right) - \operatorname{expit}\left(\theta_2 + \theta_3 a^* + \beta_0 + \beta_1 a^* + \boldsymbol{\beta}_2^{'}\boldsymbol{c}\right),\\
&\frac{\partial L_{NDE}}{\partial \beta_1} = a^* \, \frac{\partial L_{NDE}}{\partial \beta_0},\\
&\frac{\partial L_{NDE}}{\partial \beta_{2i}} = c_i \, \frac{\partial L_{NDE}}{\partial \beta_0}, \quad i = 1, 2, \ldots, k,\\
&\frac{\partial L_{NDE}}{\partial \theta_0} \equiv 0,\\
&\frac{\partial L_{NDE}}{\partial \theta_1} = a - a^*,\\
&\frac{\partial L_{NDE}}{\partial \theta_2} = \frac{\partial L_{NDE}}{\partial \beta_0},\\
&\frac{\partial L_{NDE}}{\partial \theta_3} = a \operatorname{expit}\left(\theta_2 + \theta_3 a + \beta_0 + \beta_1 a^* + \boldsymbol{\beta}_2^{'}\boldsymbol{c}\right) - a^* \operatorname{expit}\left(\theta_2 + \theta_3 a^* + \beta_0 + \beta_1 a^* + \boldsymbol{\beta}_2^{'}\boldsymbol{c}\right),\\
&\frac{\partial L_{NDE}}{\partial \theta_{4i}} \equiv 0, \quad i = 1, 2, \ldots, k,
\end{aligned}
$$

et

$$
\begin{aligned}
&\frac{\partial L_{NIE}}{\partial \beta_0} = \operatorname{expit}\left(\beta_0 + \beta_1 a^* + \boldsymbol{\beta}_2^{'}\boldsymbol{c}\right) - \operatorname{expit}\left(\beta_0 + \beta_1 a + \boldsymbol{\beta}_2^{'}\boldsymbol{c}\right)\\
&\qquad + \operatorname{expit}\left(\theta_2 + \theta_3 a + \beta_0 + \beta_1 a + \boldsymbol{\beta}_2^{'}\boldsymbol{c}\right) - \operatorname{expit}\left(\theta_2 + \theta_3 a + \beta_0 + \beta_1 a^* + \boldsymbol{\beta}_2^{'}\boldsymbol{c}\right),\\
&\frac{\partial L_{NIE}}{\partial \beta_1} = a^* \operatorname{expit}\left(\beta_0 + \beta_1 a^* + \boldsymbol{\beta}_2^{'}\boldsymbol{c}\right) - a \operatorname{expit}\left(\beta_0 + \beta_1 a + \boldsymbol{\beta}_2^{'}\boldsymbol{c}\right)\\
&\qquad + a \operatorname{expit}\left(\theta_2 + \theta_3 a + \beta_0 + \beta_1 a + \boldsymbol{\beta}_2^{'}\boldsymbol{c}\right) - a^* \operatorname{expit}\left(\theta_2 + \theta_3 a + \beta_0 + \beta_1 a^* + \boldsymbol{\beta}_2^{'}\boldsymbol{c}\right),\\
&\frac{\partial L_{NIE}}{\partial \beta_{2i}} = c_i \, \frac{\partial L_{NIE}}{\partial \beta_0}, \quad i = 1, 2, \ldots, k,\\
&\frac{\partial L_{NIE}}{\partial \theta_0} \equiv 0, \quad \frac{\partial L_{NIE}}{\partial \theta_1} \equiv 0,\\
&\frac{\partial L_{NIE}}{\partial \theta_2} = \operatorname{expit}\left(\theta_2 + \theta_3 a + \beta_0 + \beta_1 a + \boldsymbol{\beta}_2^{'}\boldsymbol{c}\right) - \operatorname{expit}\left(\theta_2 + \theta_3 a + \beta_0 + \beta_1 a^* + \boldsymbol{\beta}_2^{'}\boldsymbol{c}\right),\\
&\frac{\partial L_{NIE}}{\partial \theta_3} = a \, \frac{\partial L_{NIE}}{\partial \theta_2},\\
&\frac{\partial L_{NIE}}{\partial \theta_{4i}} \equiv 0, \quad i = 1, 2, \ldots, k.
\end{aligned}
$$

Les dérivées partielles présentées nous permettent de construire les gradients

$$
\nabla\left(L_{NDE}\right) = \left(\frac{\partial L_{NDE}}{\partial \boldsymbol{\beta}}, \frac{\partial L_{NDE}}{\partial \boldsymbol{\theta}}\right)^{'}, \qquad \nabla\left(L_{NIE}\right) = \left(\frac{\partial L_{NIE}}{\partial \boldsymbol{\beta}}, \frac{\partial L_{NIE}}{\partial \boldsymbol{\theta}}\right)^{'}
$$

des fonctions $L_{NDE}(a, a^*, \boldsymbol{c})$ et $L_{NIE}(a, a^*, \boldsymbol{c})$ par rapport au vecteur $(\boldsymbol{\beta}, \boldsymbol{\theta})$ et, par la suite, d'obtenir les intervalles de confiance à 95% approximés pour les expressions $app.OR^{NDE}_{a,a^*|\boldsymbol{c}}$ et $app.OR^{NIE}_{a,a^*|\boldsymbol{c}}$ (2.17-2.18) de la même manière que dans le cas d'un médiateur continu (voir la Sous-section 2.1.1).

La méthode bootstrap peut alternativement être appliquée pour construire des intervalles de confiance à 95% pour $app.OR^{NDE}_{a,a^*|\boldsymbol{c}}$ et $app.OR^{NIE}_{a,a^*|\boldsymbol{c}}$ (2.17-2.18).

2.3 Approche d'estimation d'Imai *et al.* (2010)

Imai *et al.* (2010) ont proposé une approche d'estimation des effets naturels applicable à une grande variété de modèles statistiques. Cette approche, implémentée par Tingley *et al.* (2014) dans le paquet R `mediation`, nécessite de spécifier un modèle pour le médiateur et un modèle pour la variable réponse :

$$M|A, \boldsymbol{C} \sim F_M(m|A, \boldsymbol{C}), \tag{2.19}$$

$$Y|A, M, \boldsymbol{C} \sim F_Y(y|A, M, \boldsymbol{C}). \tag{2.20}$$

Lorsque les modèles (2.19-2.20) sont paramétriques, nous les présentons comme suit :

$$M|A, \boldsymbol{C} \sim F_M(m|A, \boldsymbol{C}; \boldsymbol{\beta}), \tag{2.21}$$

$$Y|A, M, \boldsymbol{C} \sim F_Y(y|A, M, \boldsymbol{C}; \boldsymbol{\theta}), \tag{2.22}$$

où $\boldsymbol{\beta}$ et $\boldsymbol{\theta}$ sont des vecteurs de paramètres.

La procédure d'estimation repose sur l'idée que les modèles (2.19-2.20) nous permettent de simuler les réponses contrefactuelles $Y(a, M(a^*))$, $a, a^* \in \mathcal{A}$. Ensuite, nous pouvons obtenir les estimations ponctuelles des effets en question, ainsi que quantifier leur incertitude.

Deux algorithmes d'estimation sont implémentés dans le paquet R `mediation` (Tingley *et al.*, 2014). Le premier algorithme (*Parametric Inference*), se basant sur les approximations quasi bayésiennes de Monte-Carlo (King *et al.*, 2000), est très général puisqu'il peut être appliqué à n'importe quels modèles paramétriques (2.21-2.22).

Notons $\mathcal{E}$ l'échantillon observé ; soit n la taille de $\mathcal{E}$. L'algorithme *Parametric Inference* se décrit comme suit (Imai *et al.*, 2010; Starkopf *et al.*, 2017) :

Étape 1. En se basant sur l'échantillon observé $\mathcal{E}$, ajuster les modèles (2.21-2.22) afin d'obtenir les estimations $\widehat{\boldsymbol{\beta}}_{\mathcal{E}}$ et $\widehat{\boldsymbol{\theta}}_{\mathcal{E}}$ [2], ainsi que les estimations $\widehat{Var}_{\mathcal{E}}(\widehat{\boldsymbol{\beta}})$ et $\widehat{Var}_{\mathcal{E}}(\widehat{\boldsymbol{\theta}})$ [3].

Étape 2. Simuler J vecteurs $\tilde{\boldsymbol{\beta}}_1, \tilde{\boldsymbol{\beta}}_2, \ldots, \tilde{\boldsymbol{\beta}}_J$ de la loi normale multivariée $\mathcal{N}\left(\widehat{\boldsymbol{\beta}}_{\mathcal{E}}, \widehat{Var}_{\mathcal{E}}(\widehat{\boldsymbol{\beta}})\right)$; simuler J vecteurs $\tilde{\boldsymbol{\theta}}_1, \tilde{\boldsymbol{\theta}}_2, \ldots, \tilde{\boldsymbol{\theta}}_J$ de la loi normale multivariée $\mathcal{N}\left(\widehat{\boldsymbol{\theta}}_{\mathcal{E}}, \widehat{Var}_{\mathcal{E}}(\widehat{\boldsymbol{\theta}})\right)$.

Étape 3. Pour chaque j, $j = 1, 2, \ldots, J$, répéter les sous-étapes suivantes :

1. Pour chaque unité $i \in \mathcal{E}$, simuler K réalisations $M_i^{jk}(a) \sim F_M(m|A = a, \boldsymbol{C} = \boldsymbol{c}_i; \tilde{\boldsymbol{\beta}}_j)$, $k = 1, 2, \ldots, K$. Pour chaque $i \in \mathcal{E}$, simuler également K réalisations $M_i^{jk}(a^*) \sim F_M(m|A = a^*, \boldsymbol{C} = \boldsymbol{c}_i; \tilde{\boldsymbol{\beta}}_j)$, $k = 1, 2, \ldots, K$.
2. Pour chaque unité $i \in \mathcal{E}$, simuler une réalisation $Y_i^{jk}(a, M_i^{jk}(a)) \sim F_Y(y|A = a, M = M_i^{jk}(a), \boldsymbol{C} = \boldsymbol{c}_i; \tilde{\boldsymbol{\theta}}_j)$, une réalisation $Y_i^{jk}(a, M_i^{jk}(a^*)) \sim F_Y(y|A = a, M = M_i^{jk}(a^*), \boldsymbol{C} = \boldsymbol{c}_i; \tilde{\boldsymbol{\theta}}_j)$, ainsi qu'une réalisation $Y_i^{jk}(a^*, M_i^{jk}(a^*)) \sim F_Y(y|A = a^*, M = M_i^{jk}(a^*), \boldsymbol{C} = \boldsymbol{c}_i; \tilde{\boldsymbol{\theta}}_j)$, $k = 1, 2, \ldots, K$.
3. Calculer les estimations ponctuelles des effets naturels spécifiques au tirage j de Monte-Carlo comme suit :

$$\widehat{NDE}_j(a, a^*) = \frac{1}{nK} \sum_{i=1}^{n} \sum_{k=1}^{K} \left(Y_i^{jk}(a, M_i^{jk}(a^*)) - Y_i^{jk}(a^*, M_i^{jk}(a^*)) \right),$$
$$\widehat{NIE}_j(a, a^*) = \frac{1}{nK} \sum_{i=1}^{n} \sum_{k=1}^{K} \left(Y_i^{jk}(a, M_i^{jk}(a)) - Y_i^{jk}(a, M_i^{jk}(a^*)) \right).$$

Étape 4. Enfin, calculer les estimations ponctuelles de NDE et NIE :

$$\widehat{NDE}(a, a^*) = \frac{1}{J} \sum_{j=1}^{J} \widehat{NDE}_j(a, a^*),$$
$$\widehat{NIE}(a, a^*) = \frac{1}{J} \sum_{j=1}^{J} \widehat{NIE}_j(a, a^*).$$

Les percentiles des ensembles des estimations

$$\{\widehat{NDE}_1(a, a^*), \ldots, \widehat{NDE}_J(a, a^*)\}, \quad \{\widehat{NIE}_1(a, a^*), \ldots, \widehat{NIE}_J(a, a^*)\}$$

servent à construire des intervalles de confiance correspondants.

2. Les vecteurs $\widehat{\boldsymbol{\beta}}_{\mathcal{E}}$ et $\widehat{\boldsymbol{\theta}}_{\mathcal{E}}$ sont les valeurs des estimateurs $\widehat{\boldsymbol{\beta}}$ et $\widehat{\boldsymbol{\theta}}$ spécifiques à l'échantillon $\mathcal{E}$.

3. Les matrices $\widehat{Var}_{\mathcal{E}}(\widehat{\boldsymbol{\beta}})$ et $\widehat{Var}_{\mathcal{E}}(\widehat{\boldsymbol{\theta}})$ sont les valeurs des estimateurs $\widehat{Var}(\widehat{\boldsymbol{\beta}})$ et $\widehat{Var}(\widehat{\boldsymbol{\theta}})$ spécifiques à l'échantillon $\mathcal{E}$.

Le deuxième algorithme (*Nonparametric Inference*) est basé sur des techniques de bootstrap non paramétrique et permet d'utiliser des modèles complexes (2.19 -2.20), par exemple, des modèles non ou semi-paramétriques, des modèles de régression quantile, etc. L'algorithme *Nonparametric Inference* est aussi applicable aux modèles paramétriques (2.21-2.22) ; sans perte de généralité, nous présentons son déroulement pour ce type de modèles (Imai *et al.*, 2010; Starkopf *et al.*, 2017) :

Étape 1. Prélever J fois un échantillon aléatoire avec remise de taille n à partir de l'échantillon observé $\mathcal{E}$. Pour chaque échantillon bootstrap $\mathcal{E}_j$, $j = 1, 2, \ldots, J$, répéter les sous-étapes suivantes :

1. Ajuster les modèles (2.21-2.22) à l'échantillon bootstrap $\mathcal{E}_j$ afin d'obtenir les estimations $\widehat{\boldsymbol{\beta}}_j$ et $\widehat{\boldsymbol{\theta}}_j$.
2. Pour chaque unité i, $i = 1, 2, \ldots, n$, dans l'échantillon bootstrap $\mathcal{E}_j$, simuler K réalisations $M_i^{jk}(a) \sim F_M(m|A = a, \boldsymbol{C} = \boldsymbol{c}_i; \widehat{\boldsymbol{\beta}}_j)$ et K réalisations $M_i^{jk}(a^*) \sim F_M(m|A = a^*, \boldsymbol{C} = \boldsymbol{c}_i; \widehat{\boldsymbol{\beta}}_j)$, $k = 1, 2, \ldots, K$.
3. Pour chaque unité $i \in \mathcal{E}_j$, simuler une réalisation $Y_i^{jk}(a, M_i^{jk}(a)) \sim F_Y(y|A = a, M = M_i^{jk}(a), \boldsymbol{C} = \boldsymbol{c}_i; \widehat{\boldsymbol{\theta}}_j)$, une réalisation $Y_i^{jk}(a, M_i^{jk}(a^*)) \sim F_Y(y|A = a, M = M_i^{jk}(a^*), \boldsymbol{C} = \boldsymbol{c}_i; \widehat{\boldsymbol{\theta}}_j)$, ainsi qu'une réalisation $Y_i^{jk}(a^*, M_i^{jk}(a^*)) \sim F_Y(y|A = a^*, M = M_i^{jk}(a^*), \boldsymbol{C} = \boldsymbol{c}_i; \widehat{\boldsymbol{\theta}}_j)$, $k = 1, 2, \ldots, K$.
4. Calculer les estimations ponctuelles des effets naturels, spécifiques à l'échantillon bootstrap $\mathcal{E}_j$, comme suit :

$$\widehat{NDE}_j(a, a^*) = \frac{1}{nK} \sum_{i=1}^{n} \sum_{k=1}^{K} \Big(Y_i^{jk}(a, M_i^{jk}(a^*)) - Y_i^{jk}(a^*, M_i^{jk}(a^*)) \Big),$$
$$\widehat{NIE}_j(a, a^*) = \frac{1}{nK} \sum_{i=1}^{n} \sum_{k=1}^{K} \Big(Y_i^{jk}(a, M_i^{jk}(a)) - Y_i^{jk}(a, M_i^{jk}(a^*)) \Big).$$

Étape 2. Calculer les estimations ponctuelles de NDE et NIE :

$$\widehat{NDE}(a, a^*) = \frac{1}{J} \sum_{j=1}^{J} \widehat{NDE}_j(a, a^*),$$
$$\widehat{NIE}(a, a^*) = \frac{1}{J} \sum_{j=1}^{J} \widehat{NIE}_j(a, a^*).$$

Les percentiles des ensembles des estimations

$$\{\widehat{NDE}_1(a,a^*),\ldots,\widehat{NDE}_J(a,a^*)\},\quad \{\widehat{NIE}_1(a,a^*),\ldots,\widehat{NIE}_J(a,a^*)\}$$

servent à construire des intervalles de confiance correspondants.

Lorsque Y est binaire, l'approche générale d'Imai *et al.* (2010) n'invoque aucunement les hypothèses de la réponse binaire rare ou commune, contrairement aux approches décrites dans les Sous-sections (2.1.1) et (2.2). Le paquet R `mediation` ne fournit les résultats que sur l'échelle de la différence de risques.

2.4 Analyse de médiation causale par des modèles à effets naturels

Considérons le modèle pour la moyenne conditionnelle de la réponse contrefactuelle $Y(a,M(a^*))$ suivant :

$$E\left\{Y(a,M(a^*))|\boldsymbol{C}=\boldsymbol{c}\right\}=g^{-1}\left\{\boldsymbol{\gamma}'W(a,a^*,\boldsymbol{c})\right\}, \tag{2.23}$$

où $g(\cdot)$ est une fonction de lien connue et $W(\cdot)$ est une fonction vectorielle connue de a, a^* et $\boldsymbol{c}$; $\boldsymbol{\gamma}$ est un vecteur de paramètres à estimer. Nous appelons l'équation (2.23) un modèle à effets naturels (*natural effect model* ; Vansteelandt *et al.* (2012)) car le vecteur $\boldsymbol{\gamma}$ contient des éléments permettant de capturer les effets naturels.

Par exemple, comme dans les sections précédentes, considérons une variable réponse Y binaire et supposons que l'équation (2.23) s'écrit comme suit :

$$\text{logit}\left\{P(Y(a,M(a^*))=1|\boldsymbol{C}=\boldsymbol{c})\right\}=\gamma_0+\gamma_1 a+\gamma_2 a^*+\gamma_3 aa^*+\boldsymbol{\gamma}_4'\boldsymbol{c}, \tag{2.24}$$

où le terme aa^* est introduit pour permettre une interaction entre l'exposition et le médiateur dans

leur effet sur la réponse. Ainsi, nous avons pour les effets naturels (1.20) :

$$\begin{aligned} OR^{NDE}_{a,a^*|\boldsymbol{c}} &= \frac{\dfrac{\text{expit}(\gamma_0+\gamma_1 a+\gamma_2 a^*+\gamma_3 a a^*+\boldsymbol{\gamma}_4^{'}\boldsymbol{c})}{1-\text{expit}(\gamma_0+\gamma_1 a+\gamma_2 a^*+\gamma_3 a a^*+\boldsymbol{\gamma}_4^{'}\boldsymbol{c})}}{\dfrac{\text{expit}(\gamma_0+\gamma_1 a^*+\gamma_2 a^*+\gamma_3 a^* a^*+\boldsymbol{\gamma}_4^{'}\boldsymbol{c})}{1-\text{expit}(\gamma_0+\gamma_1 a^*+\gamma_2 a^*+\gamma_3 a^* a^*+\boldsymbol{\gamma}_4^{'}\boldsymbol{c})}} \\ &= \frac{\exp(\gamma_0+\gamma_1 a+\gamma_2 a^*+\gamma_3 a a^*+\boldsymbol{\gamma}_4^{'}\boldsymbol{c})}{\exp(\gamma_0+\gamma_1 a^*+\gamma_2 a^*+\gamma_3 a^* a^*+\boldsymbol{\gamma}_4^{'}\boldsymbol{c})} \\ &= \exp(\gamma_1(a-a^*)+\gamma_3(a-a^*)a^*), \end{aligned} \tag{2.25}$$

$$\begin{aligned} OR^{NIE}_{a,a|\boldsymbol{c}} &= \frac{\dfrac{\text{expit}(\gamma_0+\gamma_1 a+\gamma_2 a+\gamma_3 a a+\boldsymbol{\gamma}_4^{'}\boldsymbol{c})}{1-\text{expit}(\gamma_0+\gamma_1 a+\gamma_2 a+\gamma_3 a a+\boldsymbol{\gamma}_4^{'}\boldsymbol{c})}}{\dfrac{\text{expit}(\gamma_0+\gamma_1 a+\gamma_2 a^*+\gamma_3 a a^*+\boldsymbol{\gamma}_4^{'}\boldsymbol{c})}{1-\text{expit}(\gamma_0+\gamma_1 a+\gamma_2 a^*+\gamma_3 a a^*+\boldsymbol{\gamma}_4^{'}\boldsymbol{c})}} \\ &= \frac{\exp(\gamma_0+\gamma_1 a+\gamma_2 a+\gamma_3 a a+\boldsymbol{\gamma}_4^{'}\boldsymbol{c})}{\exp(\gamma_0+\gamma_1 a+\gamma_2 a^*+\gamma_3 a a^*+\boldsymbol{\gamma}_4^{'}\boldsymbol{c})} \\ &= \exp(\gamma_2(a-a^*)+\gamma_3(a-a^*)a). \end{aligned} \tag{2.26}$$

Cet exemple illustre que, contrairement aux autres approches présentées dans ce chapitre, le modèle (2.23) permet une paramétrisation directe des effets naturels en termes de ses coefficients. Ainsi, pour le cas spécial (2.24) du modèle général (2.23), chacun des effets naturels sur l'échelle du rapport de cotes est capturé en utilisant juste deux paramètres du modèle (2.24), notamment γ_1 et γ_3 pour l'effet naturel direct (2.25) et γ_2 et γ_3 pour l'effet naturel indirect (2.26). Nous pouvons aussi constater que les expressions (2.25 -2.26) ne dépendent pas des variables $\boldsymbol{c}$ (c'est-à-dire qu'elles sont les mêmes quelque soit $\boldsymbol{c}$) ; pour examiner si $OR^{NDE}_{a,a|\boldsymbol{c}}$ et/ou $OR^{NIE}_{a,a|\boldsymbol{c}}$ varient à travers différentes strates, les termes $a\boldsymbol{c}$ et/ou $a^*\boldsymbol{c}$ devraient être inclus dans le modèle (Steen *et al.*, 2017). Il est aussi important de noter que les expressions (2.25 -2.26) sont dérivées sans invoquer les hypothèses de la réponse rare ou commune.

Pour chaque unité des données étudiées, nous ne pouvons qu'observer la réponse contrefactuelle $Y(a, M(a^*))$ où a et a^* correspondent au niveau observé de l'exposition A ; par conséquent, $a = a^*$. Ainsi, pour ajuster un modèle à effets naturels (2.23), il suffit d'augmenter les données originales en introduisant pour chaque unité une observation additionnelle qui tient compte de la paire (a, a^*) où $a \neq a^*$. Deux techniques, notamment l'approche par pondération (*weighting-based approach*) et celle par imputation (*imputation-based approach*), ont été proposées pour construire des données augmentées (Lange *et al.*, 2012; Steen *et al.*, 2017; Vansteelandt *et al.*, 2012). Nous présentons la base

théorique justifiant les techniques d'augmentation des données précédemment citées en supposant, sans perte de généralité, que le médiateur M est discret.

L'équation (1.31) implique que

$$
\begin{aligned}
&E\left\{Y(a, M(a^*))|\boldsymbol{C}=\boldsymbol{c}\right\} \\
&\quad= \sum_{m\in\mathcal{M}} E\left\{Y|A=a, M=m, \boldsymbol{C}=\boldsymbol{c}\right\} P(M=m|A=a^*, \boldsymbol{C}=\boldsymbol{c}) \\
&\quad= \sum_{m\in\mathcal{M}} \left[\sum_{y\in\mathcal{Y}} y\, P(Y=y|A=a, M=m, \boldsymbol{C}=\boldsymbol{c}) P(M=m|A=a^*, \boldsymbol{C}=\boldsymbol{c})\right] \\
&\quad= \sum_{m\in\mathcal{M}} \left[\sum_{y\in\mathcal{Y}} y\, \frac{P(Y=y, M=m|A=a, \boldsymbol{C}=\boldsymbol{c})}{P(M=m|A=a, \boldsymbol{C}=\boldsymbol{c})} P(M=m|A=a^*, \boldsymbol{C}=\boldsymbol{c})\right] \\
&\quad= \sum_{y\in\mathcal{Y}} \left[\sum_{m\in\mathcal{M}} y\, \frac{P(M=m|A=a^*, \boldsymbol{C}=\boldsymbol{c})}{P(M=m|A=a, \boldsymbol{C}=\boldsymbol{c})} P(Y=y, M=m|A=a, \boldsymbol{C}=\boldsymbol{c})\right] \\
&\quad= E\left\{Y\, \frac{P(M|A=a^*, \boldsymbol{C}=\boldsymbol{c})}{P(M|A=a, \boldsymbol{C}=\boldsymbol{c})}\middle| A=a, \boldsymbol{C}=\boldsymbol{c}\right\} \\
&\quad= E\left\{YW\middle| A=a, \boldsymbol{C}=\boldsymbol{c}\right\}, \qquad (2.27)
\end{aligned}
$$

où

$$W = \frac{P(M|A=a^*, \boldsymbol{C}=\boldsymbol{c})}{P(M|A=a, \boldsymbol{C}=\boldsymbol{c})}.$$

Ainsi, l'équation (2.27) entraîne une procédure de pondération par laquelle les données observées sont augmentées par la construction d'une pseudo-population dans laquelle les mêmes unités sont évaluées aux différents niveaux du médiateur (correspondants à $A=a$ et $A=a^*$), mais au niveau observé de l'exposition, soit $A=a$, en utilisant les poids W dans l'équation (2.27) (Lange *et al.*, 2012; Steen *et al.*, 2017).

Également, nous pouvons réécrire l'équation (1.31) comme suit :

$$E\left\{Y(a, M(a^*))|\boldsymbol{C}=\boldsymbol{c}\right\} = E\left\{E\left\{Y|A=a, M, \boldsymbol{C}=\boldsymbol{c}\right\}\middle| A=a^*, \boldsymbol{C}=\boldsymbol{c}\right\}. \qquad (2.28)$$

L'équation (2.28) justifie la deuxième technique d'augmentation des données, celle par imputation. Lorsque a^* est la valeur observée de l'exposition A et $a=a^*$, les valeurs de Y dans les données observées coïncident avec les réponses contrefactuelles $Y(a, M(a^*))$. D'autre part, si la valeur ob-

servée de A est a^* et $a \neq a^*$, la réponse contrefactuelle $Y(a, M(a^*))$ n'est pas observée, tandis que son élément contrefactuel $M(a^*)$ coïncide avec la valeur observée de M (c'est-à-dire $M = M(a^*)$). Ainsi, nous pouvons augmenter les données originales en créant pour chaque unité une observation additionnelle pour laquelle la réponse contrefactuelle non observée $Y(a, M(a^*))$, $a \neq a^*$, est imputée par $E\{Y|A = a, M, \boldsymbol{C} = \boldsymbol{c}\}$ (Steen *et al.*, 2017; Vansteelandt *et al.*, 2012).

L'estimation des NDE et NIE en utilisant des modèles à effets naturels est implémentée dans le paquet R `medflex` (Steen *et al.*, 2017). Les procédures d'estimation correspondantes, incluant les algorithmes de l'augmentation des données par pondération et par imputation, sont décrites dans Vansteelandt *et al.* (2012) et Starkopf *et al.* (2017).

CHAPITRE III

RÉPONSE ET MÉDIATEUR BINAIRES : COMPARAISON DES APPROCHES PAR MODÈLES DE RÉGRESSION LOGISTIQUE ET LOG-BINOMIAL

Dans ce chapitre, nous présentons des expressions pour les effets naturels direct et indirect sur l'échelle du rapport de risques dans le contexte d'une réponse et d'un médiateur binaires. La dérivation de ces expressions se base sur un modèle de régression log-binomial pour la réponse, mais deux options sont considérées pour modéliser le médiateur : soit un modèle de régression logistique, soit un modèle de régression log-binomiale (comme, par exemple, dans l'article de Ananth et VanderWeele (2011)). Cependant, une erreur de la paramétrisation des effets naturels découlant du non-respect de l'hypothèse de la rareté du médiateur peut produire des estimations sévèrement biaisées de ces effets. Dans la dernière partie de ce chapitre, nous discutons plus généralement des difficultés de déterminer si une réponse binaire peut être classifiée comme rare dans le contexte de la médiation causale.

3.1 Expressions pour les effets naturels direct et indirect sur l'échelle du rapport de risques

Dans le contexte d'un médiateur M et d'une réponse Y binaires, l'approche d'estimation approximative des effets naturels sur l'échelle du rapport de cotes de Valeri et VanderWeele (2013) se base sur les modèles de régression logistique (2.15) pour M et (2.3) pour Y. Rappelons que Valeri et VanderWeele (2013) ont invoqué l'hypothèse de la réponse rare pour dériver les expressions $app.OR^{NDE}_{a,a^*|\boldsymbol{c}}$ et $app.OR^{NIE}_{a,a^*|\boldsymbol{c}}$ (2.17-2.18).

Lorsque la variable réponse Y n'est pas rare, l'approche de Valeri et VanderWeele (2013) considère le modèle logistique (2.15) pour M et le modèle log-binomial suivant pour Y :

$$\log\{P(Y=1|A=a,M=m,\boldsymbol{C}=\boldsymbol{c})\}=\theta_0+\theta_1 a+\theta_2 m+\theta_3 am+\boldsymbol{\theta}_4^{'}\boldsymbol{c}. \tag{3.1}$$

Sous les hypothèses de modélisation (2.15, 3.1), la probabilité emboîtée (2.14) s'exprime comme suit :

$$\begin{aligned}
&P(Y(a, M(a^*)) = 1 | \boldsymbol{C} = \boldsymbol{c}) \\
&\quad = \exp\left(\theta_0 + \theta_1 a + \theta_2 + \theta_3 a + \boldsymbol{\theta}_4^{'} \boldsymbol{c}\right) \operatorname{expit}\left(\beta_0 + \beta_1 a^* + \boldsymbol{\beta}_2^{'} \boldsymbol{c}\right) \\
&\qquad + \exp\left(\theta_0 + \theta_1 a + \boldsymbol{\theta}_4^{'} \boldsymbol{c}\right) \left(1 - \operatorname{expit}\left(\beta_0 + \beta_1 a^* + \boldsymbol{\beta}_2^{'} \boldsymbol{c}\right)\right) \\
&\quad = \frac{\exp\left(\theta_0 + \theta_1 a + \boldsymbol{\theta}_4^{'} \boldsymbol{c}\right) \left[\exp\left(\theta_2 + \theta_3 a + \beta_0 + \beta_1 a^* + \boldsymbol{\beta}_2^{'} \boldsymbol{c}\right) + 1\right]}{1 + \exp\left(\beta_0 + \beta_1 a^* + \boldsymbol{\beta}_2^{'} \boldsymbol{c}\right)}.
\end{aligned} \tag{3.2}$$

Nous avons de l'équation (3.2) les expressions suivantes pour les effets naturels direct et indirect sur l'échelle du rapport de risques (1.19) :

$$\begin{aligned}
RR^{NDE}_{a,a^*|\boldsymbol{c}} &= \frac{P(Y(a, M(a^*)) = 1 | \boldsymbol{C} = \boldsymbol{c})}{P(Y(a^*, M(a^*)) = 1 | \boldsymbol{C} = \boldsymbol{c})} \\
&= \frac{\dfrac{\exp\left(\theta_0 + \theta_1 a + \boldsymbol{\theta}_4^{'} \boldsymbol{c}\right) \left[\exp\left(\theta_2 + \theta_3 a + \beta_0 + \beta_1 a^* + \boldsymbol{\beta}_2^{'} \boldsymbol{c}\right) + 1\right]}{1 + \exp\left(\beta_0 + \beta_1 a^* + \boldsymbol{\beta}_2^{'} \boldsymbol{c}\right)}}{\dfrac{\exp\left(\theta_0 + \theta_1 a^* + \boldsymbol{\theta}_4^{'} \boldsymbol{c}\right) \left[\exp\left(\theta_2 + \theta_3 a^* + \beta_0 + \beta_1 a^* + \boldsymbol{\beta}_2^{'} \boldsymbol{c}\right) + 1\right]}{1 + \exp\left(\beta_0 + \beta_1 a^* + \boldsymbol{\beta}_2^{'} \boldsymbol{c}\right)}} \\
&= \exp\left(\theta_1 (a - a^*)\right) \cdot \frac{\exp\left(\theta_2 + \theta_3 a + \beta_0 + \beta_1 a^* + \boldsymbol{\beta}_2^{'} \boldsymbol{c}\right) + 1}{\exp\left(\theta_2 + \theta_3 a^* + \beta_0 + \beta_1 a^* + \boldsymbol{\beta}_2^{'} \boldsymbol{c}\right) + 1},
\end{aligned} \tag{3.3}$$

$$\begin{aligned}
RR^{NIE}_{a,a^*|\boldsymbol{c}} &= \frac{P(Y(a, M(a)) = 1 | \boldsymbol{C} = \boldsymbol{c})}{P(Y(a, M(a^*)) = 1 | \boldsymbol{C} = \boldsymbol{c})} \\
&= \frac{\dfrac{\exp\left(\theta_0 + \theta_1 a + \boldsymbol{\theta}_4^{'} \boldsymbol{c}\right) \left[\exp\left(\theta_2 + \theta_3 a + \beta_0 + \beta_1 a + \boldsymbol{\beta}_2^{'} \boldsymbol{c}\right) + 1\right]}{1 + \exp\left(\beta_0 + \beta_1 a + \boldsymbol{\beta}_2^{'} \boldsymbol{c}\right)}}{\dfrac{\exp\left(\theta_0 + \theta_1 a + \boldsymbol{\theta}_4^{'} \boldsymbol{c}\right) \left[\exp\left(\theta_2 + \theta_3 a + \beta_0 + \beta_1 a^* + \boldsymbol{\beta}_2^{'} \boldsymbol{c}\right) + 1\right]}{1 + \exp\left(\beta_0 + \beta_1 a^* + \boldsymbol{\beta}_2^{'} \boldsymbol{c}\right)}} \\
&= \frac{\exp\left(\beta_0 + \beta_1 a^* + \boldsymbol{\beta}_2^{'} \boldsymbol{c}\right) + 1}{\exp\left(\beta_0 + \beta_1 a + \boldsymbol{\beta}_2^{'} \boldsymbol{c}\right) + 1} \cdot \frac{\exp\left(\theta_2 + \theta_3 a + \beta_0 + \beta_1 a + \boldsymbol{\beta}_2^{'} \boldsymbol{c}\right) + 1}{\exp\left(\theta_2 + \theta_3 a + \beta_0 + \beta_1 a^* + \boldsymbol{\beta}_2^{'} \boldsymbol{c}\right) + 1}.
\end{aligned} \tag{3.4}$$

Contrairement aux expressions $app.OR^{NDE}_{a,a^*|\boldsymbol{c}}$ et $app.OR^{NIE}_{a,a^*|\boldsymbol{c}}$ (2.17-2.18) développées par Valeri et VanderWeele (2013) sous l'hypothèse de la réponse rare, aucune hypothèse simplificatrice n'a été invoquée pour dériver les formules (3.3 -3.4).

En comparant $RR^{NDE}_{a,a^*|\boldsymbol{c}}$ (3.3) avec $app.OR^{NDE}_{a,a^*|\boldsymbol{c}}$ (2.17), ainsi que $RR^{NIE}_{a,a^*|\boldsymbol{c}}$ (3.4) avec $app.OR^{NIE}_{a,a^*|\boldsymbol{c}}$ (2.18), nous pouvons constater la même forme analytique des expressions constituant chaque paire de comparaison, c'est-à-dire

$$\exp\left(\theta_1(a-a^*)\right)\cdot\frac{\exp\left(\theta_2+\theta_3 a+\beta_0+\beta_1 a^*+\boldsymbol{\beta}_2^{'}\boldsymbol{c}\right)+1}{\exp\left(\theta_2+\theta_3 a^*+\beta_0+\beta_1 a^*+\boldsymbol{\beta}_2^{'}\boldsymbol{c}\right)+1}$$

et

$$\frac{\exp\left(\beta_0+\beta_1 a^*+\boldsymbol{\beta}_2^{'}\boldsymbol{c}\right)+1}{\exp\left(\beta_0+\beta_1 a+\boldsymbol{\beta}_2^{'}\boldsymbol{c}\right)+1}\cdot\frac{\exp\left(\theta_2+\theta_3 a+\beta_0+\beta_1 a+\boldsymbol{\beta}_2^{'}\boldsymbol{c}\right)+1}{\exp\left(\theta_2+\theta_3 a+\beta_0+\beta_1 a^*+\boldsymbol{\beta}_2^{'}\boldsymbol{c}\right)+1},$$

respectivement. Toutefois, les coefficients θ_0, θ_1, θ_2, θ_3 et $\boldsymbol{\theta}_4$ obtenus du modèle log-binomial (3.1) pour construire les rapports de risques (3.3 -3.4) diffèrent de celles obtenus du modèle logistique (2.3) pour dériver les rapports de cotes (2.17-2.18). Les coefficients β_0, β_1 et $\boldsymbol{\beta}_2$ sont identiques pour chaque paire de comparaison, car les expressions (3.3 -3.4) et (2.17-2.18) sont dérivées sous le même modèle logistique (2.15) du médiateur M. Ainsi, quand les coefficients θ_0, θ_1, θ_2, θ_3 et $\boldsymbol{\theta}_4$ provenant du modèle logistique (2.3) de la réponse Y sont proches des coefficients correspondants du modèle log-binomial (3.1) de Y, nous pouvons interpréter les expressions approximatives $app.OR^{NDE}_{a,a^*|\boldsymbol{c}}$ (2.17) et $app.OR^{NIE}_{a,a^*|\boldsymbol{c}}$ (2.18) comme des rapports de risques, car elles mènent à des estimations des effets naturels proches de celles obtenues des expressions $RR^{NDE}_{a,a^*|\boldsymbol{c}}$ (3.3) et $RR^{NIE}_{a,a^*|\boldsymbol{c}}$ (3.4), respectivement (Samoilenko *et al.* (2018)).

Valeri et VanderWeele (2013) ont dérivé les expressions $app.OR^{NDE}_{a,a^*|\boldsymbol{c}}$, $app.OR^{NIE}_{a,a^*|\boldsymbol{c}}$ (2.17-2.18) et $RR^{NDE}_{a,a^*|\boldsymbol{c}}$, $RR^{NIE}_{a,a^*|\boldsymbol{c}}$ (3.3 -3.4) sous le modèle logistique (2.15) du médiateur. Toutefois, le modèle log-binomial

$$\log\{P(M=1|A=a,\boldsymbol{C}=\boldsymbol{c})\}=\beta_0+\beta_1 a+\boldsymbol{\beta}_2^{'}\boldsymbol{c}, \tag{3.5}$$

plutôt que le modèle logistique (2.15), peut être utilisé pour relier la probabilité conditionnelle $P(M=1|A=a,\boldsymbol{C}=\boldsymbol{c})$ au prédicteur linéaire $\beta_0+\beta_1 a+\boldsymbol{\beta}_2^{'}\boldsymbol{c}$, ce qui entraîne, conjointement avec l'hypothèse de modélisation (3.1), la paramétrisation suivante de la probabilité contrefactuelle emboîtée (2.14) :

$$
\begin{aligned}
&P(Y(a, M(a^*)) = 1|\boldsymbol{C} = \boldsymbol{c}) \\
&\quad = \exp\left(\theta_0 + \theta_1 a + \theta_2 + \theta_3 a + \boldsymbol{\theta_4'c}\right) \exp\left(\beta_0 + \beta_1 a^* + \boldsymbol{\beta_2'c}\right) \\
&\qquad + \exp\left(\theta_0 + \theta_1 a + \boldsymbol{\theta_4'c}\right) \left(1 - \exp\left(\beta_0 + \beta_1 a^* + \boldsymbol{\beta_2'c}\right)\right) \\
&\quad = \exp\left(\theta_0 + \theta_1 a + \boldsymbol{\theta_4'c}\right) \left\{\exp\left(\beta_0 + \beta_1 a^* + \boldsymbol{\beta_2'c}\right) \left[\exp\left(\theta_2 + \theta_3 a\right) - 1\right] + 1\right\}. \qquad (3.6)
\end{aligned}
$$

Ainsi, sous les hypothèses de modélisation (3.1, 3.5) impliquant l'équation (3.6), nous obtenons les expressions suivantes pour les effets naturels direct et indirect sur l'échelle du rapport de risques (1.19) :

$$
\begin{aligned}
RR^{NDE}_{a,a^*|\boldsymbol{c}} &= \frac{P(Y(a, M(a^*)) = 1|\boldsymbol{C} = \boldsymbol{c})}{P(Y(a^*, M(a^*)) = 1|\boldsymbol{C} = \boldsymbol{c})} \\
&= \frac{\exp\left(\theta_0 + \theta_1 a + \boldsymbol{\theta_4'c}\right) \left\{\exp\left(\beta_0 + \beta_1 a^* + \boldsymbol{\beta_2'c}\right) \left[\exp\left(\theta_2 + \theta_3 a\right) - 1\right] + 1\right\}}{\exp\left(\theta_0 + \theta_1 a^* + \boldsymbol{\theta_4'c}\right) \left\{\exp\left(\beta_0 + \beta_1 a^* + \boldsymbol{\beta_2'c}\right) \left[\exp\left(\theta_2 + \theta_3 a^*\right) - 1\right] + 1\right\}} \qquad (3.7) \\
&= \exp\left(\theta_1 (a - a^*)\right) \frac{\exp\left(\beta_0 + \beta_1 a^* + \boldsymbol{\beta_2'c}\right) \left[\exp\left(\theta_2 + \theta_3 a\right) - 1\right] + 1}{\exp\left(\beta_0 + \beta_1 a^* + \boldsymbol{\beta_2'c}\right) \left[\exp\left(\theta_2 + \theta_3 a^*\right) - 1\right] + 1}
\end{aligned}
$$

et

$$
\begin{aligned}
RR^{NIE}_{a,a^*|\boldsymbol{c}} &= \frac{P(Y(a, M(a)) = 1|\boldsymbol{C} = \boldsymbol{c})}{P(Y(a, M(a^*)) = 1|\boldsymbol{C} = \boldsymbol{c})} \\
&= \frac{\exp\left(\theta_0 + \theta_1 a + \boldsymbol{\theta_4'c}\right) \left\{\exp\left(\beta_0 + \beta_1 a + \boldsymbol{\beta_2'c}\right) \left[\exp\left(\theta_2 + \theta_3 a\right) - 1\right] + 1\right\}}{\exp\left(\theta_0 + \theta_1 a + \boldsymbol{\theta_4'c}\right) \left\{\exp\left(\beta_0 + \beta_1 a^* + \boldsymbol{\beta_2'c}\right) \left[\exp\left(\theta_2 + \theta_3 a\right) - 1\right] + 1\right\}} \qquad (3.8) \\
&= \frac{\exp\left(\beta_0 + \beta_1 a + \boldsymbol{\beta_2'c}\right) \left[\exp\left(\theta_2 + \theta_3 a\right) - 1\right] + 1}{\exp\left(\beta_0 + \beta_1 a^* + \boldsymbol{\beta_2'c}\right) \left[\exp\left(\theta_2 + \theta_3 a\right) - 1\right] + 1}.
\end{aligned}
$$

Il convient de noter que les paramétrisations (3.2) et (3.6) de la probabilité contrefactuelle emboîtée $P(Y(a, M(a^*)) = 1|\boldsymbol{C} = \boldsymbol{c})$ (2.14), basées respectivement sur le modèles logistique (2.15) et log-binomial (3.5) du médiateur, produisent généralement des estimations très proches des effets naturels sur l'échelle du rapport de risques en raison de la proximité entre les modèles logistiques et log-binomiaux pour l'estimation des probabilités. Toutefois, un remplacement des estimateurs de β_0, β_1 et $\boldsymbol{\beta}_2$ basés sur le modèle log-binomial (3.5) par leurs contreparties basées sur le modèle logistique (2.15) peut induire un biais non négligeable pour les estimateurs basés sur les expressions (3.3-3.4).

Ce phénomène a été étudié dans l'article de Samoilenko et Lefebvre (2019) publié dans l'*American Journal of Epidemiology*. Notre motivation principale pour la rédaction de cet article prend sa source dans l'identification de deux erreurs liées à la spécification du modèle du médiateur binaire dans la publication de Ananth et VanderWeele (2011). Dans notre article, nous explorons l'utilité des modèles log-binomiaux pour l'estimation des effets naturels direct et indirect pour une réponse et un médiateur binaires qui ne sont pas nécessairement rares. Nous évaluons aussi l'impact potentiel de l'erreur de la paramétrisation des effet naturels dans l'article de Ananth et VanderWeele (2011) en fonction de la rareté du médiateur. L'article de Samoilenko et Lefebvre (2019) est presenté dans la section suivante. Le lecteur peut aussi accéder à cet article en suivant le lien `https://doi.org/10.1093/aje/kwy275`.

3.2 ARTICLE 1. Point: Risk ratio equations for natural direct and indirect effects in causal mediation analysis of a binary mediator and a binary outcome – A fresh look at the formulas

Mariia Samoilenko and Geneviève Lefebvre

Abstract: In this article, we review the formulas for the natural direct and indirect effects' risk ratios introduced by Ananth and VanderWeele (*Am J Epidemiol.* 2011; 174(1): 99–108) for causal mediation analysis of a binary mediator and a binary outcome. In particular, we show that the closed-form equations Ananth and VanderWeele provided do not correspond to the log-binomial model specified by these authors for the mediator variable, but rather to a logistic model. We then provide risk ratio equations for natural direct and indirect effects that truly pertain to a log-binomial model. We conclude with a discussion on the practical implications of the binary mediator model's specification by analysts. The related impact can be negligible or not, depending on the rareness of the mediator.

Keywords: binary mediator; causal mediation analysis; log-binomial model; logistic model; risk ratio.

Abbreviations: NDE, natural direct effect; NIE, natural indirect effect; RR, risk ratio.

Editor's note: *A counterpoint to this article appears on page 1204*[1].

1. Voir VanderWeele *et al.* (2019) dans la bibliographie ou `https://doi.org/10.1093/aje/kwy281`.

We are concerned about theoretical results presented in the paper "Placental Abruption and Perinatal Mortality With Preterm Delivery as a Mediator: Disentangling Direct and Indirect Effects," by Ananth and VanderWeele (2011), published in the *American Journal of Epidemiology* in 2011. Therein, the authors proposed log-binomial models for conducting mediation analyses in the presence of a dichotomous outcome and mediator that are not necessarily rare. Specifically, a mediation methodology based on log-binomial models was described for estimation of natural direct and indirect (preterm-delivery–mediated) effects of placental abruption on mortality. While Ananth and VanderWeele (2011) did not provide guidelines for qualifying an outcome as rare in a mediation context, a (marginal) outcome probability of 10% has been proposed as a threshold (VanderWeele, 2015), although this convenient definition is not exempt from problems (Samoilenko *et al.*, 2018).

Our purpose in this article is to draw attention to the equations provided for the mediation effect risk ratios and point out two errors. One error is notably related to misspecification of the mediator model, and we explain herein that this may not be inconsequential in practice. Given the impact of Dr. VanderWeele's and his collaborators' body of work in many research areas worldwide and the absence of published errata for this paper, we believe it necessary to write the article to present correct information.

In the paper by Ananth and VanderWeele (2011), the natural direct and indirect effects of the placental abruption–mortality relationship were estimated by fitting a log-binomial model for mortality (Y), conditional on placental abruption (X), preterm delivery (M), an abruption–preterm delivery interaction variable ($X \times M$), and a set of confounders ($\boldsymbol{C}$):

$$\log\{P(Y=1|X=x, M=m, \boldsymbol{C}=\boldsymbol{c})\} = \theta_0 + \theta_1 x + \theta_2 m + \theta_3 xm + \boldsymbol{\theta}_4^{'}\boldsymbol{c}, \tag{3.9}$$

as well as a log-binomial model for preterm delivery, conditional on placental abruption and confounders:

$$\log\{P(M=1|X=x, \boldsymbol{C}=\boldsymbol{c})\} = \beta_0 + \beta_1 x + \boldsymbol{\beta}_2^{'}\boldsymbol{c}. \tag{3.10}$$

From models (3.9) and (3.10) (presented above in Equations (3.9) and (3.10)), Ananth and VanderWeele (2011) wrote that the risk ratios (RRs) for the natural direct effect (NDE), RR_{NDE}, and

natural indirect effect (NIE), RR_{NIE}, could be estimated for a dichotomous exposure (0/1) as

$$RR_{NDE}(\boldsymbol{c}) = \frac{\exp(\theta_1)\left(1+\exp\left(\theta_2+\theta_3+\beta_0+\boldsymbol{\beta}_2'\boldsymbol{c}\right)\right)}{1+\exp\left(\theta_2+\theta_3+\boldsymbol{\beta}_2'\boldsymbol{c}\right)} \tag{3.11}$$

and

$$RR_{NIE}(\boldsymbol{c}) = \frac{\left(1+\exp\left(\beta_0+\boldsymbol{\beta}_2'\boldsymbol{c}\right)\right)\left(1+\exp\left(\theta_2+\theta_3+\beta_0+\beta_1+\boldsymbol{\beta}_2'\boldsymbol{c}\right)\right)}{\left(1+\exp\left(\beta_0+\beta_1+\boldsymbol{\beta}_2'\boldsymbol{c}\right)\right)\left(1+\exp\left(\theta_2+\theta_3+\beta_0+\boldsymbol{\beta}_2'\boldsymbol{c}\right)\right)}. \tag{3.12}$$

The first error we identify is revealed when comparing Equations (3.11) and (3.12) with the RR_{NDE} and RR_{NIE} equations provided in subsequent published work by Valeri and VanderWeele (2013) for general exposure levels a and a^*. Indeed, the correct equation for the the RR_{NDE} should be

$$RR_{NDE}(\boldsymbol{c}) = \frac{\exp(\theta_1)\left(1+\exp\left(\theta_2+\theta_3+\beta_0+\boldsymbol{\beta}_2'\boldsymbol{c}\right)\right)}{1+\exp\left(\theta_2+\beta_0+\boldsymbol{\beta}_2'\boldsymbol{c}\right)}, \tag{3.13}$$

where θ_3 is replaced by β_0 in the denominator of Equation (3.11).

While the second error is probably also typographical, we argue that it has more profound implication. In the paper by Ananth and VanderWeele (2011), derivation of the equations for the RR_{NDE} and RR_{NIE} was deferred to a technical report (Valeri et VanderWeele, 2010) which does not seem to be readily available anymore. Valeri and VanderWeele (2013) also gave only the results; that is, they did not provide an explicit derivation of Equations (3.12) and (3.13) (see Valeri and VanderWeele's supplementary material, section 4, p. 18). However, despite the unavailability of proofs for these results, it is straightforward to show that Equations (3.12) and (3.13) are, strictly speaking, valid only when using a log-binomial model for the outcome but a logistic model for the mediator (see Appendix 3.2.1). In other words, the link function for model (3.10) (Equation (3.10)) should be *logit* instead of *log* to obtain Equations (3.12) and (3.13). Equations for the RR_{NDE} and RR_{NIE} that can be derived from a log-binomial model for the mediator do not simplify as in Equations (3.12) and (3.13) (and of course, neither as Equation (3.11)):

$$RR_{NDE}(\boldsymbol{c}) = \exp(\theta_1)\frac{\exp\left(\beta_0+\boldsymbol{\beta}_2'\boldsymbol{c}\right)\left[\exp(\theta_2+\theta_3)-1\right]+1}{\exp\left(\beta_0+\boldsymbol{\beta}_2'\boldsymbol{c}\right)\left[\exp(\theta_2)-1\right]+1}; \tag{3.14}$$

$$RR_{NIE}(\boldsymbol{c}) = \frac{\exp\left(\beta_0 + \beta_1 + \boldsymbol{\beta}_2'\boldsymbol{c}\right)\left[\exp\left(\theta_2 + \theta_3\right) - 1\right] + 1}{\exp\left(\beta_0 + \boldsymbol{\beta}_2'\boldsymbol{c}\right)\left[\exp\left(\theta_2 + \theta_3\right) - 1\right] + 1}. \tag{3.15}$$

(See Appendix 3.2.1 for proof.)

This second error required to be brought forward for the sake of transparency and mathematical correctness. However, it could also have nonnegligible impact from a practical standpoint. Imagine a practitioner fitting a log-binomial model for the mediator and using the risk ratio formulas truly pertaining to the logistic mediator model (Equations (3.12) and (3.13)) for estimation of the natural direct and indirect effects of exposure. This is the situation that we are directly concerned with: Ananth and VanderWeele (2011) referred to a log-binomial model for the mediator in their paper and suggested using Equations (3.12) and (3.13) for estimation of the mediation effects. Then, unless the mediator is sufficiently rare, marked differences could be seen between the RR_{NDE} and RR_{NIE} obtained (erroneously) with Equations (3.12) and (3.13) and those that should have been obtained with correct equations (Equations (3.14) and (3.15)). This is a consequence of the fact that when the mediator is not rare, mediator logistic probabilities

$$P(M = 1|X = x, \boldsymbol{C} = \boldsymbol{c}) = \frac{\exp\left(\beta_0 + \beta_1 x + \boldsymbol{\beta}_2'\boldsymbol{c}\right)}{1 + \exp\left(\beta_0 + \beta_1 x + \boldsymbol{\beta}_2'\boldsymbol{c}\right)} \quad (x = 0, 1)$$

computed with regression coefficients β_0, β_1, $\boldsymbol{\beta}_2$ obtained from a log-binomial model will not generally be close to the same probabilities computed with regression coefficients β_0, β_1, $\boldsymbol{\beta}_2$ obtained from a logistic model, since nonnegligible differences in the actual values of these coefficients will be seen from these two types of models. As such, *one should not plug in log-binomial parameter values in the mediator probabilities parametrized in terms of a logistic model when the mediator is not rare* (see Appendix 3.2.2 for an illustration of the impact on RR_{NDE} and RR_{NIE} estimates).

Now let us consider a practitioner who hesitates between a logistic model and a log-binomial model for the mediator and is aware of the two sets of equations presented here for estimation of RR_{NDE} and RR_{NIE}. In this case, the RR_{NDE} and RR_{NIE} obtained using Equations (3.12) and (3.13) for the logistic mediator model and Equations (3.14) and (3.15) for the log-binomial mediator model will typically be similar. This is a consequence of the fact that the mediator probabilities

$$P(M = 1|X = x, \boldsymbol{C} = \boldsymbol{c}), \quad x = 0, 1,$$

parametrized and obtained on the basis of a logistic regression will generally be close to those parametrized and obtained on the basis of a log-binomial regression even when the mediator is not rare (see Appendix 3.2.3 for an example).

In light of the previous comments, a common-sense advice is thus to use the formulas which directly pertain to the mediator model that is applied to the data: Use Equations (3.12) and (3.13) if the logistic model is fitted, and use Equations (3.14) and (3.15) if the log-binomial model is fitted. Which specific strategy to adopt will often be inconsequential because of the closeness between the logistic and log-binomial models for probability estimation, although one may want to take advantage of the greater numerical stability of the logistic model (Spiegelman et Hertzmark, 2005). In conclusion, it is noteworthy to recall that the SAS macro (SAS Institute, Inc., Cary, North Carolina) created by Valeri and VanderWeele (2013), a well-known resource for conducting regression-based mediation analyses, is not designed to handle a log-binomial model for both the outcome and the mediator. More precisely, with a binary outcome, the outcome model can be specified to be log-binomial (in addition to logistic), but only a logistic model is available for a binary mediator, thus being already in line with the above recommendation.

3.2.1 Appendix 1: Correct natural direct and indirect effects' risk ratio expressions

Log-binomial and logistic case

We begin by showing that expressions (3.12) and (3.13) (presented in the main text) are precisely obtained by using a log-binomial model for the outcome and a logistic model for the mediator. First, the logistic model is formulated as

$$\text{logit}\{P(M=1|X=x,\boldsymbol{C}=\boldsymbol{c})\}=\beta_0+\beta_1 x+\boldsymbol{\beta}_2^{'}\boldsymbol{c}, \tag{3.16}$$

Then we express $P\left(Y\left(1,M(1)\right)=1|\boldsymbol{C}=\boldsymbol{c}\right)$, $P\left(Y\left(1,M(0)\right)=1|\boldsymbol{C}=\boldsymbol{c}\right)$ and $P\left(Y\left(0,M(0)\right)=1|\boldsymbol{C}=\boldsymbol{c}\right)$ on the basis of the log-binomial outcome model (3.9) and the logistic mediator model (3.16). These counterfactual probabilities, which are involved in the numerator and denominator of the RR_{NDE} and RR_{NIE}, are given as:

$$
\begin{aligned}
&P\left(Y\left(1, M(1)\right)=1 | \boldsymbol{C}=\boldsymbol{c}\right) \\
&\quad=P\left(Y=1 | X=1, M=1, \boldsymbol{C}=\boldsymbol{c}\right) P\left(M=1 | X=1, \boldsymbol{C}=\boldsymbol{c}\right) \\
&\qquad+P\left(Y=1 | X=1, M=0, \boldsymbol{C}=\boldsymbol{c}\right) P\left(M=0 | X=1, \boldsymbol{C}=\boldsymbol{c}\right) \\
&\quad=\exp\left(\theta_0+\theta_1+\theta_2+\theta_3+\boldsymbol{\theta}_4^{'} \boldsymbol{c}\right) \cdot \frac{\exp\left(\beta_0+\beta_1+\boldsymbol{\beta}_2^{'} \boldsymbol{c}\right)}{1+\exp\left(\beta_0+\beta_1+\boldsymbol{\beta}_2^{'} \boldsymbol{c}\right)} \\
&\qquad+\exp\left(\theta_0+\theta_1+\boldsymbol{\theta}_4^{'} \boldsymbol{c}\right) \cdot \frac{1}{1+\exp\left(\beta_0+\beta_1+\boldsymbol{\beta}_2^{'} \boldsymbol{c}\right)} \\
&\quad=\frac{\exp\left(\theta_0+\theta_1+\boldsymbol{\theta}_4^{'} \boldsymbol{c}\right)\left\{\exp\left(\theta_2+\theta_3\right) \exp\left(\beta_0+\beta_1+\boldsymbol{\beta}_2^{'} \boldsymbol{c}\right)+1\right\}}{1+\exp\left(\beta_0+\beta_1+\boldsymbol{\beta}_2^{'} \boldsymbol{c}\right)};
\end{aligned}
$$

$$
\begin{aligned}
&P\left(Y\left(1, M(0)\right)=1 | \boldsymbol{C}=\boldsymbol{c}\right) \\
&\quad=P\left(Y=1 | X=1, M=1, \boldsymbol{C}=\boldsymbol{c}\right) P\left(M=1 | X=0, \boldsymbol{C}=\boldsymbol{c}\right) \\
&\qquad+P\left(Y=1 | X=1, M=0, \boldsymbol{C}=\boldsymbol{c}\right) P\left(M=0 | X=0, \boldsymbol{C}=\boldsymbol{c}\right) \\
&\quad=\exp\left(\theta_0+\theta_1+\theta_2+\theta_3+\boldsymbol{\theta}_4^{'} \boldsymbol{c}\right) \cdot \frac{\exp\left(\beta_0+\boldsymbol{\beta}_2^{'} \boldsymbol{c}\right)}{1+\exp\left(\beta_0+\boldsymbol{\beta}_2^{'} \boldsymbol{c}\right)} \\
&\qquad+\exp\left(\theta_0+\theta_1+\boldsymbol{\theta}_4^{'} \boldsymbol{c}\right) \cdot \frac{1}{1+\exp\left(\beta_0+\boldsymbol{\beta}_2^{'} \boldsymbol{c}\right)} \\
&\quad=\frac{\exp\left(\theta_0+\theta_1+\boldsymbol{\theta}_4^{'} \boldsymbol{c}\right)\left\{\exp\left(\theta_2+\theta_3\right) \exp\left(\beta_0+\boldsymbol{\beta}_2^{'} \boldsymbol{c}\right)+1\right\}}{1+\exp\left(\beta_0+\boldsymbol{\beta}_2^{'} \boldsymbol{c}\right)};
\end{aligned}
$$

$$
\begin{aligned}
&P\left(Y\left(0, M(0)\right)=1 | \boldsymbol{C}=\boldsymbol{c}\right) \\
&\quad=P\left(Y=1 | X=0, M=1, \boldsymbol{C}=\boldsymbol{c}\right) P\left(M=1 | X=0, \boldsymbol{C}=\boldsymbol{c}\right) \\
&\qquad+P\left(Y=1 | X=0, M=0, \boldsymbol{C}=\boldsymbol{c}\right) P\left(M=0 | X=0, \boldsymbol{C}=\boldsymbol{c}\right) \\
&\quad=\exp\left(\theta_0+\theta_2+\boldsymbol{\theta}_4^{'} \boldsymbol{c}\right) \cdot \frac{\exp\left(\beta_0+\boldsymbol{\beta}_2^{'} \boldsymbol{c}\right)}{1+\exp\left(\beta_0+\boldsymbol{\beta}_2^{'} \boldsymbol{c}\right)}+\exp\left(\theta_0+\boldsymbol{\theta}_4^{'} \boldsymbol{c}\right) \cdot \frac{1}{1+\exp\left(\beta_0+\boldsymbol{\beta}_2^{'} \boldsymbol{c}\right)} \\
&\quad=\frac{\exp\left(\theta_0+\boldsymbol{\theta}_4^{'} \boldsymbol{c}\right)\left\{\exp\left(\theta_2\right) \exp\left(\beta_0+\boldsymbol{\beta}_2^{'} \boldsymbol{c}\right)+1\right\}}{1+\exp\left(\beta_0+\boldsymbol{\beta}_2^{'} \boldsymbol{c}\right)}.
\end{aligned}
$$

Then we get, for a logistic mediator model,

$$\begin{aligned} RR_{NDE}(\boldsymbol{c}) &= \frac{P(Y(1,M(0))=1|\boldsymbol{C}=\boldsymbol{c})}{P(Y(0,M(0))=1|\boldsymbol{C}=\boldsymbol{c})} \\ &= \frac{\dfrac{\exp\left(\theta_0+\theta_1+\boldsymbol{\theta}_4'\boldsymbol{c}\right)}{1+\exp\left(\beta_0+\boldsymbol{\beta}_2'\boldsymbol{c}\right)}\cdot\left\{\exp(\theta_2+\theta_3)\exp\left(\beta_0+\boldsymbol{\beta}_2'\boldsymbol{c}\right)+1\right\}}{\dfrac{\exp\left(\theta_0+\boldsymbol{\theta}_4'\boldsymbol{c}\right)}{1+\exp\left(\beta_0+\boldsymbol{\beta}_2'\boldsymbol{c}\right)}\cdot\left\{\exp(\theta_2)\exp\left(\beta_0+\boldsymbol{\beta}_2'\boldsymbol{c}\right)+1\right\}} \\ &= \frac{\exp(\theta_1)\left\{\exp\left(\theta_2+\theta_3+\beta_0+\boldsymbol{\beta}_2'\boldsymbol{c}\right)+1\right\}}{\exp\left(\theta_2+\beta_0+\boldsymbol{\beta}_2'\boldsymbol{c}\right)+1} \end{aligned}$$

and

$$\begin{aligned} RR_{NIE}(\boldsymbol{c}) &= \frac{P(Y(1,M(1))=1|\boldsymbol{C}=\boldsymbol{c})}{P(Y(1,M(0))=1|\boldsymbol{C}=\boldsymbol{c})} \\ &= \frac{\dfrac{\exp\left(\theta_0+\theta_1+\boldsymbol{\theta}_4'\boldsymbol{c}\right)}{1+\exp\left(\beta_0+\beta_1+\boldsymbol{\beta}_2'\boldsymbol{c}\right)}\cdot\left\{\exp(\theta_2+\theta_3)\exp\left(\beta_0+\beta_1+\boldsymbol{\beta}_2'\boldsymbol{c}\right)+1\right\}}{\dfrac{\exp\left(\theta_0+\theta_1+\boldsymbol{\theta}_4'\boldsymbol{c}\right)}{1+\exp\left(\beta_0+\boldsymbol{\beta}_2'\boldsymbol{c}\right)}\cdot\left\{\exp(\theta_2+\theta_3)\exp\left(\beta_0+\boldsymbol{\beta}_2'\boldsymbol{c}\right)+1\right\}} \\ &= \frac{\left\{1+\exp\left(\beta_0+\boldsymbol{\beta}_2'\boldsymbol{c}\right)\right\}\cdot\left\{\exp\left(\theta_2+\theta_3+\beta_0+\beta_1+\boldsymbol{\beta}_2'\boldsymbol{c}\right)+1\right\}}{\left\{1+\exp\left(\beta_0+\beta_1+\boldsymbol{\beta}_2'\boldsymbol{c}\right)\right\}\cdot\left\{\exp\left(\theta_2+\theta_3+\beta_0+\boldsymbol{\beta}_2'\boldsymbol{c}\right)+1\right\}}. \end{aligned}$$

Log-binomial and log-binomial case

We now provide corresponding expressions for the RR_{NDE} and RR_{NIE} when using a log-binomial mediator model instead of a logistic. Similarly, we express $P(Y(1,M(1))=1|\boldsymbol{C}=\boldsymbol{c})$, $P(Y(1,M(0))=1|\boldsymbol{C}=\boldsymbol{c})$ and $P(Y(0,M(0))=1|\boldsymbol{C}=\boldsymbol{c})$ on the basis of the log-binomial outcome model (3.9) and the log-binomial mediator model (3.10):

$$
\begin{aligned}
&P\left(Y\left(1, M(1)\right)=1 | \boldsymbol{C}=\boldsymbol{c}\right) \\
&\quad= P\left(Y=1 | X=1, M=1, \boldsymbol{C}=\boldsymbol{c}\right) P\left(M=1 | X=1, \boldsymbol{C}=\boldsymbol{c}\right) \\
&\qquad+ P\left(Y=1 | X=1, M=0, \boldsymbol{C}=\boldsymbol{c}\right) P\left(M=0 | X=1, \boldsymbol{C}=\boldsymbol{c}\right) \\
&\quad= \exp\left(\theta_0+\theta_1+\theta_2+\theta_3+\boldsymbol{\theta}_4^{'} \boldsymbol{c}\right) \exp\left(\beta_0+\beta_1+\boldsymbol{\beta}_2^{'} \boldsymbol{c}\right) \\
&\qquad+ \exp\left(\theta_0+\theta_1+\boldsymbol{\theta}_4^{'} \boldsymbol{c}\right)\left(1-\exp\left(\beta_0+\beta_1+\boldsymbol{\beta}_2^{'} \boldsymbol{c}\right)\right) \\
&\quad= \exp\left(\theta_0+\theta_1+\boldsymbol{\theta}_4^{'} \boldsymbol{c}\right)\left\{\exp\left(\theta_2+\theta_3\right) \exp\left(\beta_0+\beta_1+\boldsymbol{\beta}_2^{'} \boldsymbol{c}\right)+1-\exp\left(\beta_0+\beta_1+\boldsymbol{\beta}_2^{'} \boldsymbol{c}\right)\right\} \\
&\quad= \exp\left(\theta_0+\theta_1+\boldsymbol{\theta}_4^{'} \boldsymbol{c}\right)\left\{\exp\left(\beta_0+\beta_1+\boldsymbol{\beta}_2^{'} \boldsymbol{c}\right)\left[\exp\left(\theta_2+\theta_3\right)-1\right]+1\right\};
\end{aligned}
$$

$$
\begin{aligned}
&P\left(Y\left(1, M(0)\right)=1 | \boldsymbol{C}=\boldsymbol{c}\right) \\
&\quad= P\left(Y=1 | X=1, M=1, \boldsymbol{C}=\boldsymbol{c}\right) P\left(M=1 | X=0, \boldsymbol{C}=\boldsymbol{c}\right) \\
&\qquad+ P\left(Y=1 | X=1, M=0, \boldsymbol{C}=\boldsymbol{c}\right) P\left(M=0 | X=0, \boldsymbol{C}=\boldsymbol{c}\right) \\
&\quad= \exp\left(\theta_0+\theta_1+\theta_2+\theta_3+\boldsymbol{\theta}_4^{'} \boldsymbol{c}\right) \exp\left(\beta_0+\boldsymbol{\beta}_2^{'} \boldsymbol{c}\right)+\exp\left(\theta_0+\theta_1+\boldsymbol{\theta}_4^{'} \boldsymbol{c}\right)\left(1-\exp\left(\beta_0+\boldsymbol{\beta}_2^{'} \boldsymbol{c}\right)\right) \\
&\quad= \exp\left(\theta_0+\theta_1+\boldsymbol{\theta}_4^{'} \boldsymbol{c}\right)\left\{\exp\left(\theta_2+\theta_3\right) \exp\left(\beta_0+\boldsymbol{\beta}_2^{'} \boldsymbol{c}\right)+1-\exp\left(\beta_0+\boldsymbol{\beta}_2^{'} \boldsymbol{c}\right)\right\} \\
&\quad= \exp\left(\theta_0+\theta_1+\boldsymbol{\theta}_4^{'} \boldsymbol{c}\right)\left\{\exp\left(\beta_0+\boldsymbol{\beta}_2^{'} \boldsymbol{c}\right)\left[\exp\left(\theta_2+\theta_3\right)-1\right]+1\right\};
\end{aligned}
$$

$$
\begin{aligned}
&P\left(Y\left(0, M(0)\right)=1 | \boldsymbol{C}=\boldsymbol{c}\right) \\
&\quad= P\left(Y=1 | X=0, M=1, \boldsymbol{C}=\boldsymbol{c}\right) P\left(M=1 | X=0, \boldsymbol{C}=\boldsymbol{c}\right) \\
&\qquad+ P\left(Y=1 | X=0, M=0, \boldsymbol{C}=\boldsymbol{c}\right) P\left(M=0 | X=0, \boldsymbol{C}=\boldsymbol{c}\right) \\
&\quad= \exp\left(\theta_0+\theta_2+\boldsymbol{\theta}_4^{'} \boldsymbol{c}\right) \exp\left(\beta_0+\boldsymbol{\beta}_2^{'} \boldsymbol{c}\right)+\exp\left(\theta_0+\boldsymbol{\theta}_4^{'} \boldsymbol{c}\right)\left(1-\exp\left(\beta_0+\boldsymbol{\beta}_2^{'} \boldsymbol{c}\right)\right) \\
&\quad= \exp\left(\theta_0+\boldsymbol{\theta}_4^{'} \boldsymbol{c}\right)\left\{\exp\left(\theta_2\right) \exp\left(\beta_0+\boldsymbol{\beta}_2^{'} \boldsymbol{c}\right)+1-\exp\left(\beta_0+\boldsymbol{\beta}_2^{'} \boldsymbol{c}\right)\right\} \\
&\quad= \exp\left(\theta_0+\boldsymbol{\theta}_4^{'} \boldsymbol{c}\right)\left\{\exp\left(\beta_0+\boldsymbol{\beta}_2^{'} \boldsymbol{c}\right)\left[\exp\left(\theta_2\right)-1\right]+1\right\}.
\end{aligned}
$$

Then we get, for a log-binomial mediator model,

$$\begin{aligned} RR_{NDE}(\boldsymbol{c}) &= \frac{P(Y(1,M(0))=1|\boldsymbol{C}=\boldsymbol{c})}{P(Y(0,M(0))=1|\boldsymbol{C}=\boldsymbol{c})} \\ &= \frac{\exp\left(\theta_0+\theta_1+\boldsymbol{\theta}_4'\boldsymbol{c}\right)\left\{\exp\left(\beta_0+\boldsymbol{\beta}_2'\boldsymbol{c}\right)\left[\exp(\theta_2+\theta_3)-1\right]+1\right\}}{\exp\left(\theta_0+\boldsymbol{\theta}_4'\boldsymbol{c}\right)\left\{\exp\left(\beta_0+\boldsymbol{\beta}_2'\boldsymbol{c}\right)\left[\exp(\theta_2)-1\right]+1\right\}} \\ &= \exp(\theta_1)\cdot\frac{\exp\left(\beta_0+\boldsymbol{\beta}_2'\boldsymbol{c}\right)\left[\exp(\theta_2+\theta_3)-1\right]+1}{\exp\left(\beta_0+\boldsymbol{\beta}_2'\boldsymbol{c}\right)\left[\exp(\theta_2)-1\right]+1} \end{aligned}$$

and

$$\begin{aligned} RR_{NIE}(\boldsymbol{c}) &= \frac{P(Y(1,M(1))=1|\boldsymbol{C}=\boldsymbol{c})}{P(Y(1,M(0))=1|\boldsymbol{C}=\boldsymbol{c})} \\ &= \frac{\exp\left(\theta_0+\theta_1+\boldsymbol{\theta}_4'\boldsymbol{c}\right)\left\{\exp\left(\beta_0+\beta_1+\boldsymbol{\beta}_2'\boldsymbol{c}\right)\left[\exp(\theta_2+\theta_3)-1\right]+1\right\}}{\exp\left(\theta_0+\theta_1+\boldsymbol{\theta}_4'\boldsymbol{c}\right)\left\{\exp\left(\beta_0+\boldsymbol{\beta}_2'\boldsymbol{c}\right)\left[\exp(\theta_2+\theta_3)-1\right]+1\right\}} \\ &= \frac{\exp\left(\beta_0+\beta_1+\boldsymbol{\beta}_2'\boldsymbol{c}\right)\left[\exp(\theta_2+\theta_3)-1\right]+1}{\exp\left(\beta_0+\boldsymbol{\beta}_2'\boldsymbol{c}\right)\left[\exp(\theta_2+\theta_3)-1\right]+1}. \end{aligned}$$

3.2.2 Appendix 2: Comparison between natural direct and indirect effects' risk ratio expressions with logistic and log-binomial mediator models

In order to demonstrate possible differences in the estimation of natural direct and indirect effects' risk ratios obtained by either Equations (3.12, 3.13) or Equations (3.14, 3.15), we have defined two relative ratios R_{NDE} and R_{NIE} contrasting these sets of expressions:

$$\begin{aligned} R_{NDE} &= \frac{\exp(\theta_1)\cdot\dfrac{\exp\left(\beta_0+\boldsymbol{\beta}_2'\boldsymbol{c}\right)\exp(\theta_2+\theta_3)+1}{\exp\left(\beta_0+\boldsymbol{\beta}_2'\boldsymbol{c}\right)\exp(\theta_2)+1}}{\exp(\theta_1)\cdot\dfrac{\exp\left(\beta_0+\boldsymbol{\beta}_2'\boldsymbol{c}\right)\left[\exp(\theta_2+\theta_3)-1\right]+1}{\exp\left(\beta_0+\boldsymbol{\beta}_2'\boldsymbol{c}\right)\left[\exp(\theta_2)-1\right]+1}} \\ &= \frac{\dfrac{\exp\left(\beta_0+\boldsymbol{\beta}_2'\boldsymbol{c}\right)\exp(\theta_2+\theta_3)+1}{\exp\left(\beta_0+\boldsymbol{\beta}_2'\boldsymbol{c}\right)\exp(\theta_2)+1}}{\dfrac{\exp\left(\beta_0+\boldsymbol{\beta}_2'\boldsymbol{c}\right)\left[\exp(\theta_2+\theta_3)-1\right]+1}{\exp\left(\beta_0+\boldsymbol{\beta}_2'\boldsymbol{c}\right)\left[\exp(\theta_2)-1\right]+1}}, \end{aligned}$$

$$R_{NIE} = \frac{\left\{\exp\left(\beta_0 + \boldsymbol{\beta}_2^{'}\boldsymbol{c}\right) + 1\right\} \cdot \left\{\exp\left(\beta_0 + \beta_1 + \boldsymbol{\beta}_2^{'}\boldsymbol{c}\right)\exp\left(\theta_2 + \theta_3\right) + 1\right\}}{\left\{\exp\left(\beta_0 + \beta_1 + \boldsymbol{\beta}_2^{'}\boldsymbol{c}\right) + 1\right\} \cdot \left\{\exp\left(\beta_0 + \boldsymbol{\beta}_2^{'}\boldsymbol{c}\right)\exp\left(\theta_2 + \theta_3\right) + 1\right\}}{\frac{\exp\left(\beta_0 + \beta_1 + \boldsymbol{\beta}_2^{'}\boldsymbol{c}\right)\left[\exp\left(\theta_2 + \theta_3\right) - 1\right] + 1}{\exp\left(\beta_0 + \boldsymbol{\beta}_2^{'}\boldsymbol{c}\right)\left[\exp\left(\theta_2 + \theta_3\right) - 1\right] + 1}}.$$

In R_{NDE}, the numerator represents expression (3.13), which corresponds to the NDE risk ratio associated with the logistic model for the mediator, while the denominator represents expression (3.14), corresponding to the same effect measure with a log-binomial mediator model. Similarly, the numerator and denominator of R_{NIE} pertain to expressions (3.12) and (3.15), respectively. To illustrate our purpose, we considered three scenarios with increasing degrees of commonness for the mediator. Specifically, we defined the baseline coefficient β_0 in Scenarios 1, 2 and 3 as $\beta_0 = -\log(20)$, $\beta_0 = -\log(3)$ and $\beta_0 = -\log(2)$, respectively. In each scenario, and without loss of generality, we considered a single binary covariate C (sex) with a prevalence of 0.5. We took $\beta_1 = \log(1.5)$ and $\beta_2 = \log(1.2)$ in all scenarios. Following the strategy in Valeri and VanderWeele (2013), the ratios R_{NDE} and R_{NIE} are reported for the mean level of the covariate C.

The results are presented as the contour lines of the R_{NDE} and R_{NIE} functions in the domain $D = \{(\theta_2, \theta_3) : \ |\theta_2 \mid \leq 3, \ |\theta_3 \mid \leq 2\}$ of the (θ_2, θ_3)-plane (see Figure 3.1). In the rare mediator case (Scenario 1), in which mediator probabilities ranged from $\exp(\beta_0) = 0.05$ to $\exp(\beta_0 + \beta_1 + \beta_2) = 0.09$ for all individuals, both ratios R_{NDE}, R_{NIE} were close to 1 for all (θ_2, θ_3) in D. To plug in the coefficients β_0, β_1, β_2 from a log-binomial mediator model in the mediator logistic expressions $(3.12, 3.13)$ thus leads to values of natural effects very close to what is obtained with mediator log-binomial expressions $(3.14, 3.15)$ when the mediator is rare. As expected however, departures of R_{NDE} and R_{NIE} from 1 were seen in the scenarios with a common mediator (with mediator probabilities ranging from 0.33 to 0.6 and 0.5 to 0.9 in Scenario 2 and Scenario 3, respectively). These departures were observed to be globally larger in the scenario in which the mediator was the most common (Scenario 3). In this scenario, the ratio of expression (3.12) to (3.15) (that is, R_{NIE}) could even reach 2 for some (θ_2, θ_3) (see Figure 3.1 (F)); put differently, in these cases the natural indirect effect risk ratio obtained from a logistic mediator model was twice the natural indirect effect risk ratio obtained from a log-binomial mediator model.

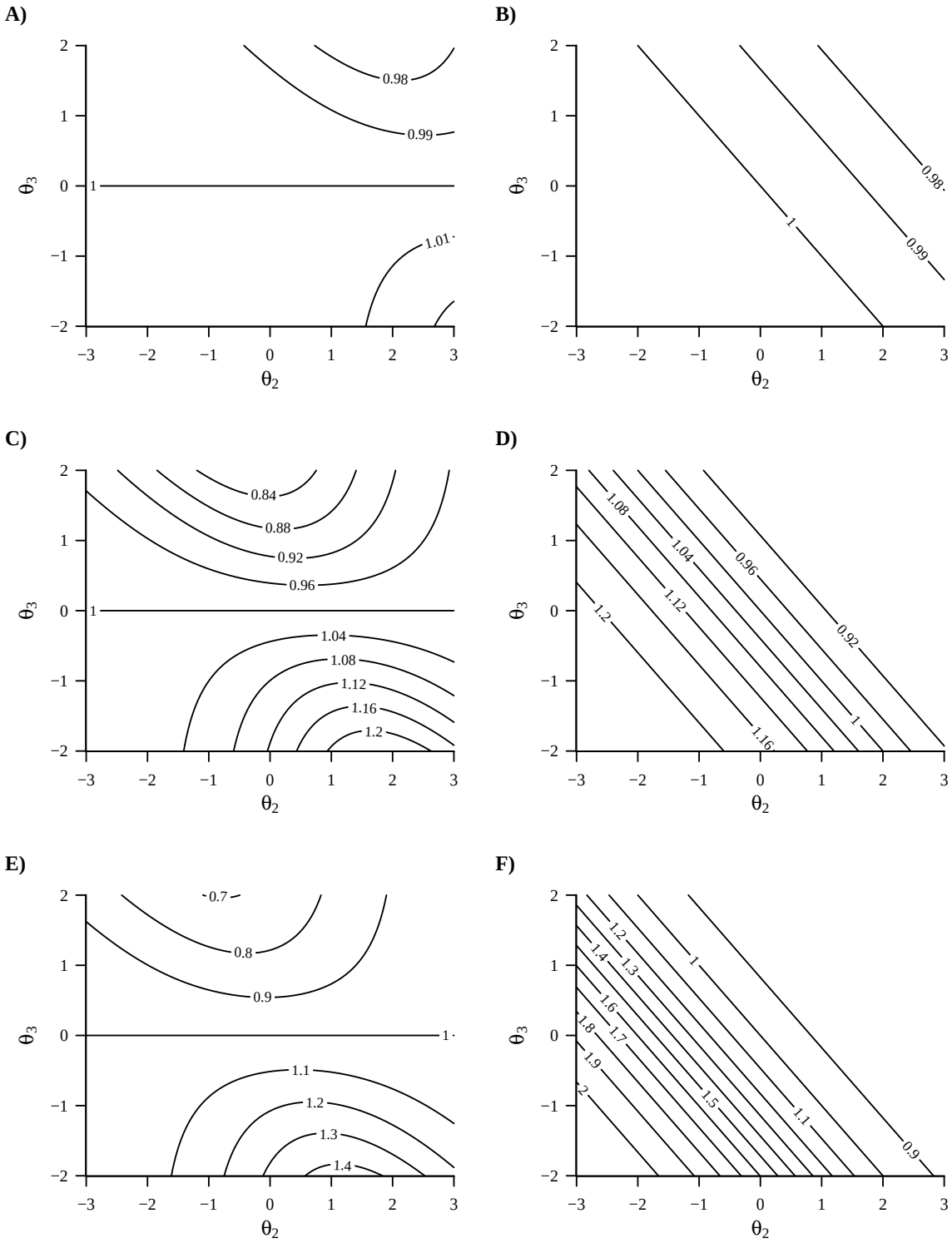
Figure 3.1 Comparing ratios of natural direct and indirect effects' risk ratio expressions derived from using a logistic model and a log-binomial model for the mediator (with a log-binomial outcome model) as a function of the outcome regression coefficients (θ_2, θ_3) for fixed scenarios of commonness of the mediator.

Legend: Contour lines of the function

A) $R_{NDE}(\theta_2, \theta_3)$ for Scenario 1: $\beta_0 = -\log(20)$, $\beta_2 = \log(1.2)$;
B) $R_{NIE}(\theta_2, \theta_3)$ for Scenario 1: $\beta_0 = -\log(20)$, $\beta_1 = \log(1.5)$, $\beta_2 = \log(1.2)$;
C) $R_{NDE}(\theta_2, \theta_3)$ for Scenario 2: $\beta_0 = -\log(3)$, $\beta_2 = \log(1.2)$;
D) $R_{NIE}(\theta_2, \theta_3)$ for Scenario 2: $\beta_0 = -\log(3)$, $\beta_1 = \log(1.5)$, $\beta_2 = \log(1.2)$;
E) $R_{NDE}(\theta_2, \theta_3)$ for Scenario 3: $\beta_0 = -\log(2)$, $\beta_2 = \log(1.2)$;
F) $R_{NIE}(\theta_2, \theta_3)$ for Scenario 3: $\beta_0 = -\log(2)$, $\beta_1 = \log(1.5)$, $\beta_2 = \log(1.2)$.

In a given scenario, a value of $R_{NDE}(\theta_2, \theta_3)$ greater than 1 indicates that the natural direct effect (NDE) risk ratio obtained under a logistic mediator model is greater than the natural direct effect risk ratio obtained under a log-binomial mediator model for a given set of values of the outcome regression coefficients (θ_2, θ_3). A similar interpretation prevails for $R_{NIE}(\theta_2, \theta_3)$, which applies to the natural indirect effect (NIE).

(The graph of Figure 3.1 appears on the following page.)


A)
B)
C)
D)
E)
F)
θ3
θ2
0.98
0.99
1
1.01
0.84
0.88
0.92
0.96
1.04
1.08
1.12
1.16
1.2
0.7
0.8
0.9
1.1
1.3
1.4
1.5
1.6
1.7
1.9
2

3.2.3 Appendix 3: Comparison between logistic and log-binomial probabilities

We present two examples showing the similarity or dissimilarity between probability estimates obtained from logistic and log-binomial models, and probability estimates based on a logistic formula and computed with estimated log-binomial coefficients.

The first example (Scenario 1; *Appendix 3.1*) concerns a case when the mediator is rare and the second example (Scenario 2; *Appendix 3.2*) concerns a case when the mediator is not rare. In both examples the mediator is generated from a log-binomial model. The R code to replicate the results is presented in *Appendix 3.3*.

Appendix 3.1. Example 1 – rare mediator

We generated a dataset of size n=100,000 with three covariates *Sex* (S), *Treatment* (X) and *Mediator* (M). The covariates *Sex* and *Treatment* were simulated independently from Bernoulli random variables with probabilities 0.5 and 0.3, respectively. Conditional on S and X, variables M were generated from Bernoulli random variables with probabilities

$$P\left(M=1|S=s,X=x\right)=\exp\left(\beta_0+\beta_s s+\beta_x x\right),$$

where $\beta_0=-\log(20)$, $\beta_s=\log(1.2)$, and $\beta_x=\log(1.5)$. For all individuals, the probability of the mediator thus ranged between 0.05 and 0.09 (with endpoint values occurring when $S=0,\ X=0$ and $S=1,\ X=1$, respectively).

Logistic and log-binomial models were then fitted on the data, where both specifications included *Sex* and *Treatment* as main effect terms. For an observation with $X=1,\ S=1$, the estimated probabilities $P\left(M=1|S=s,X=x\right)$ returned from the logistic and log-binomial models were 0.0935 and 0.0937, respectively (the true value is $\exp\left(-\log(20)+\log(1.2)\cdot 1+\log(1.5)\cdot 1\right)=0.09$). Using the estimated coefficients returned by the log-binomial model, the conditional probability of the outcome based on the logistic formula for such an observation was 0.0856.

This example illustrates the approximate equivalence between all three probabilities computed from these models when the mediator is "rare".

Appendix 3.2. Example 2 – common mediator

We replicated the experiment presented in Example 1, but this time the regression coefficient β_0

was modified to increase the prevalence of the mediator. Specifically, we used $\beta_0 = -\log(3)$, $\beta_s = \log(1.2)$, and $\beta_x = \log(1.5)$ to generate the mediator, yielding to mediator probabilities ranging between 0.33 and 0.6 for all individuals. We then fitted the same two mediator models as before on the data.

For an observation with $X = 1$, $S = 1$, the estimated probabilities $P(M = 1|S = s, X = x)$ returned from the logistic and log-binomial models were 0.5903 and 0.5998, respectively (the true value is $\exp(-\log(3) + \log(1.2) \cdot 1 + \log(1.5) \cdot 1) = 0.6$). Using the estimated coefficients returned by the log-binomial model, the conditional probability of the outcome based on the logistic formula for such an observation was 0.3749. The discrepancy between this probability and the two former probabilities arises from marked differences between the actual values of the regression coefficients estimated from the logistic and the log-binomial models in this case.

This illustrates the potential pitfall described in the main text when the mediator is not rare.

Appendix 3.3. R code and associated outputs

```
# Function to compute conditional probabilities for the binary
# mediator based on a log-binomial model (b0, bs, and bx are
# regression coefficients)
pcalcul<-function(b0,bs,bx,s,x){
  p<-exp(b0+bs*s+bx*x)
}

#####################################################################
# Example 1: Case when the mediator is rare                         #
#####################################################################

set.seed(1567)

# Sample size
n<-100000

# Prevalence of binary covariate (sex)
ps<-0.5

# Prevalence of binary treatment
px<-0.3

# Generate sex
s<-rbinom(n,1,ps)

# Generate treatment
x<-rbinom(n,1,px)

# Specify true values of regression coefficients
b0<- -log(20)
bs<-log(1.2)
bx<-log(1.5)

# Compute probability of the mediator given sex and treatment
pm<-pcalcul(b0,bs,bx,s,x)

# Generate mediator
m<-rbinom(n,1,pm)

# Fit a logistic model on the data
testlogit<-glm(m~s+x,family="binomial")
summary(testlogit)

# Fit a log-binomial model on the data
testlog<-glm(m~s+x,family=binomial(link = log))
summary(testlog)

## We observe that the regression coefficient estimates are similar
## in both models (with rare mediator)

# > summary(testlogit)
#
```

```
# Call:
# glm(formula = m ~ s + x, family = "binomial")
#
# Deviance Residuals:
#     Min       1Q   Median       3Q      Max
# -0.4430  -0.4012  -0.3544  -0.3205   2.4474
#
# Coefficients:
#             Estimate Std. Error  z value Pr(>|z|)
# (Intercept) -2.94343    0.02175 -135.338  < 2e-16 ***
# s            0.20741    0.02591    8.005 1.19e-15 ***
# x            0.46406    0.02644   17.553  < 2e-16 ***
# ---

# > summary(testlog)
#
# Call:
# glm(formula = m ~ s + x, family = binomial(link = log))
#
# Deviance Residuals:
#     Min       1Q   Median       3Q      Max
# -0.4435  -0.4007  -0.3543  -0.3206   2.4470
#
# Coefficients:
#             Estimate Std. Error  z value Pr(>|z|)
# (Intercept) -2.99382    0.02051 -145.948   <2e-16 ***
# s            0.19404    0.02418    8.026    1e-15 ***
# x            0.43181    0.02447   17.645   <2e-16 ***
# ---

# Data frame with individual observation with s = 1, x = 1
newdata<-data.frame(s = 1,x = 1)


## Compute predicted probability of the mediator when s = 1, x = 1...

# ...based on the logistic model
predict(testlogit,newdata,type="response")

# > predict(testlogit,newdata,type="response")
#         1
# 0.0934717

# ...based on the log-binomial model
predict(testlog, newdata, type="response")

# > predict(testlog, newdata, type="response")
#          1
# 0.09367106

## Both predicted values are close

# Compute logistic probability based on log-binomial regression
# coefficients when s = 1, x = 1
```

```
plogfromlogbin<-exp(-2.99382+0.19404+0.43181)/(1+exp(-2.99382+0.19404+0.43181))

# > plogfromlogbin
# [1] 0.08564798

## Similar to the previous two probabilities

#######################################################################
# Example 2: Case when the mediator is not rare                       #
#######################################################################

set.seed(1567)

# Sample size
n<-100000

# Prevalence of binary covariate (sex)
ps<-0.5

# Prevalence of binary treatment
px<-0.3


# Generate sex
s<-rbinom(n,1,ps)

# Generate treatment
x<-rbinom(n,1,px)

# Specify true values of regression coefficients
b0<- -log(3)
bs<-log(1.2)
bx<-log(1.5)

# Compute probability of the mediator given sex and treatment
pm<-pcalcul(b0,bs,bx,s,x)

# Generate mediator
m<-rbinom(n,1,pm)

# Fit a logistic model on the data
testlogit<-glm(m~s+x,family="binomial")
summary(testlogit)

# Fit a log-binomial model on the data
testlog<-glm(m~s+x,family=binomial(link = log))
summary(testlog)

## We observe that the regression coefficient estimates are now
## very different in both models (with not rare mediator)

# > summary(testlogit)
#
# Call:
```

```
# glm(formula = m ~ s + x, family = "binomial")
#
# Deviance Residuals:
#    Min      1Q  Median      3Q     Max
# -1.336  -1.023  -0.898   1.340   1.485
#
# Coefficients:
#             Estimate Std. Error z value Pr(>|z|)
# (Intercept) -0.69999    0.01038  -67.41   <2e-16 ***
# s            0.32451    0.01303   24.90   <2e-16 ***
# x            0.74064    0.01404   52.75   <2e-16 ***
# ---

# > summary(testlog)
#
# Call:
# glm(formula = m ~ s + x, family = binomial(link = log))
#
# Deviance Residuals:
#     Min       1Q   Median       3Q      Max
# -1.3534  -1.0156  -0.9052   1.3483   1.4767
#
# Coefficients:
#              Estimate Std. Error z value Pr(>|z|)
# (Intercept) -1.090286   0.006423 -169.74   <2e-16 ***
# s            0.181349   0.007230   25.08   <2e-16 ***
# x            0.397819   0.007129   55.80   <2e-16 ***
# ---

# Data frame with individual observation with s = 1, x = 1
newdata = data.frame(s = 1, x = 1)

## Compute predicted probability of the mediator when s = 1, x = 1...

# ...based on the logistic model
predict(testlogit, newdata, type="response")

# > predict(testlogit, newdata, type="response")
#         1
# 0.5902889

# ...based on the log-binomial model
predict(testlog, newdata, type="response")

# > predict(testlog, newdata, type="response")
#         1
# 0.5998247

## Both predicted values are again close

# Compute logistic probability based on log-binomial regression
# coefficients when s = 1, x = 1
plogfromlogbin<-exp(-1.090286+0.181349+0.397819)/(1+exp(-1.090286+0.181349+0.397819))
```

```
# > plogfromlogbin
# [1] 0.3749315

## This probability does not agree with the previous two
## probabilities, as opposed to when the mediator is rare
```

Acknowledgements

This publication was supported by grants from the Fonds de recherche Québec-Santé (FRQ-S) and the Natural Sciences and Engineering Research Council of Canada. Geneviève Lefebvre is a FRQ-S Research Scholar.

Conflict of interest: none declared.

3.3 Hypothèse de la réponse rare dans les analyses de médiation causale

Dans l'article de Samoilenko et Lefebvre (2019) présenté dans la Section 3.2, ainsi que dans la réponse de VanderWeele *et al.* (2019) sur cet article, la problématique d'absence de lignes directrices permettant de qualifier une réponse binaire comme rare dans le contexte de la médiation causale a été abordée. Bien que l'hypothèse de la réponse rare soit souvent référée dans les publications sur la médiation causale (voir, par exemple, les ouvrages de référence comme VanderWeele (2015), VanderWeele (2016) et Lash *et al.* (2021)), elle est rarement, voire jamais, formulée de manière explicite. En commentant l'article de Samoilenko et Lefebvre (2019) (Section 3.2), VanderWeele *et al.* (2019) ont souligné : « Greater care should certainly be taken in the articulation of this assumption. »

En pratique, l'hypothèse de la réponse rare est souvent examinée marginalement (c'est-à-dire en vérifiant que la prévalence observée de la réponse est $\leq 10\%$) afin d'appliquer les méthodes approximatives discutées dans la Section 2.2 et la Sous-section 2.1.1. Néanmoins, Samoilenko *et al.* (2018) ont montré dans leur étude de simulation que les estimateurs basés sur les expressions approximatives $app.OR^{NDE}_{a,a^*|\boldsymbol{c}}$ et $app.OR^{NIE}_{a,a^*|\boldsymbol{c}}$ (2.17-2.18) de Valeri et VanderWeele (2013) sont biaisés, si la réponse n'est pas rare dans au moins une strate définie par le traitement et le médiateur, même si la réponse est rare marginalement. Ainsi, VanderWeele *et al.* (2019) ont reconnu que l'hypothèse de la réponse rare devrait être satisfaite dans toutes les strates formées par le traitement, le médiateur et les covariables pour que leurs estimateurs approximatifs soient valides.

Dans les Chapitres 4 et 5, nous présenterons les estimateurs paramétriques des effets naturels direct et indirect pour une variable réponse binaire et un médiateur binaire ou continu qui ne reposent pas sur l'hypothèse de la réponse rare marginalement ou conditionnellement. De plus, dans le cas d'un médiateur continu, contrairement à l'approche de Gaynor *et al.* (2019), ces estimateurs ne s'appuient pas sur l'hypothèse de la réponse commune. De cette manière, nos estimateurs permettent d'éviter des difficultés liées à la vérification des hypothèses de la réponse rare ou commune dans un contexte de médiation causale.

CHAPITRE IV

ARTICLE 2. PARAMETRIC-REGRESSION—BASED CAUSAL MEDIATION ANALYSIS OF BINARY OUTCOMES AND BINARY MEDIATORS : MOVING BEYOND THE RARENESS OR COMMONNESS OF THE OUTCOME

Dans ce chapitre formé de l'article de Samoilenko et Lefebvre (2021) publié dans l'*American Journal of Epidemiology* (`https://doi.org/10.1093/aje/kwab055`), nous présentons les estimateurs paramétriques des effets naturels direct et indirect pour une variable réponse et un médiateur binaires qui ne reposent pas sur les hypothèses de la réponse rare (marginalement ou conditionnellement) ou commune. La présentation de cet article tient compte des corrections apportées dans Samoilenko et Lefebvre (2022a) (`https://doi.org/10.1093/aje/kwac078`). Notons aussi que cet article est présenté selon les exigences linguistiques du journal.

Mariia Samoilenko and Geneviève Lefebvre

Abstract: In the causal mediation framework, several parametric-regression–based approaches have been introduced in the last decade for estimating natural direct and indirect effects. For a binary outcome, a number of proposed estimators use a logistic model and rely on specific assumptions or approximations that may be delicate or not easy to verify in practice. To circumvent the challenges prompted by the rare outcome assumption in this context, an exact closed-form natural-effects estimator on the odds ratio scale was recently introduced for a binary mediator. In this work, we further push this exact approach and extend it for the estimation of natural effects on the risk ratio and risk difference scales. Explicit formulas for the delta method standard errors are provided. The performance of our proposed exact estimators is demonstrated in simulation scenarios featuring various levels of outcome rareness/commonness. The total effect decomposition property on the multiplicative scales is also examined. Using a SAS macro (SAS Institute, Inc., Cary, North

Carolina) we developed, our approach is illustrated to assess the separate effects of exposure to inhaled corticosteroids and placental abruption on low birth weight mediated by prematurity. Our exact natural-effects estimators are found to work properly in both simulations and the real data example.

Keywords: binary mediator; binary outcome; causal mediation regression-based analysis; exact natural-effects estimator; outcome rareness/commonness.

Abbreviations: 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.

4.1 Introduction

Mediation analysis approaches that rely on the specification of parametric models for the mediator and outcome variables are naturally appealing to practitioners because of their conceptual simplicity. However, notoriously, the development of such approaches is more challenging when the outcome is binary, as opposed to continuous, due to the consideration of nonlinear models (Loeys *et al.*, 2013). In this line of research, contributions made over the years in the causal inference framework have helped to increase the resources available for estimating direct and indirect effects with binary outcomes. However, a number of proposed approaches invoke specific assumptions or approximations, some of which may be delicate or not easy to verify in practice. VanderWeele and Vansteelandt (2010) and Valeri and VanderWeele (2013) relied on the rare outcome assumption (ROA) to propose regression-based estimators of the natural direct effect (NDE) and the natural indirect effect (NIE) on the odds ratio (OR) scale for continuous and binary mediators. For a normally distributed mediator, Gaynor *et al.* (2019) used a probit approximation to the logit function to provide an estimator of the NDE and NIE on the OR scale that can be used when the outcome is common. Previous work by Tchetgen Tchetgen (2014), which motivated the work of Gaynor *et al.* (2019), introduced an exact estimator for a nonrare outcome, but the approach assumed a bridge distribution for the continuous mediator.

For a binary outcome and a binary mediator, the logistic-regression–based causal mediation approach of Valeri and VanderWeele (2013) is popular among applied researchers, arguably because of its accessible implementation in standard statistical software (e.g., the SAS procedure PROC CAUSALMED (SAS Institute Inc., Cary, North Carolina) and the Stata module PARAMED (StataCorp LLC, College Station, Texas); Emsley et Liu (2013); SAS Institute Inc. (2017); Yung *et al.*

(2018)). First designed for cohort data, this approximate approach is based on the simplifying ROA, which is crucial in the development of the proposed closed-form natural-effects OR estimator. In practical contexts, the ROA is commonly verified by checking that the marginal outcome prevalence $P(Y = 1)$ is reasonably small (Feingold *et al.*, 2019; Rijnhart *et al.*, 2019; VanderWeele, 2015). However, as further expanded below, there is an increased awareness that this marginal definition is inadequate for the ROA in causal mediation settings.

For a binary mediator, both Samoilenko *et al.* (2018) and Gaynor *et al.* (2019) independently introduced a logistic-regression–based estimator for cohort data that uses the parametrized outcome and mediator probabilities to express the NDE and NIE on the OR scale. This estimator is qualified as exact since it does not rely on approximations and can be used regardless of the rareness or commonness of the outcome.

Samoilenko *et al.* (2018) presented a simulation scenario mimicking real perinatal data in which the outcome was rare marginally (i.e., with $P(Y = 1) < 0.1$) but not in the strata formed by the exposure and mediator. They compared the proposed exact OR estimator with the Valeri and VanderWeele (2013) approximate estimator and found that the former was unbiased for the NDE and NIE ORs ($\mathrm{OR}^{\mathrm{NDE}}$ and $\mathrm{OR}^{\mathrm{NIE}}$, respectively), unlike the latter. Commenting on Samoilenko *et al.* (2018), VanderWeele *et al.* (2019) acknowledged that the ROA needs to hold in strata formed by covariates, *including the mediator*, for their estimator to be valid. However, to require that the outcome be rare in strata of a mediator is questionable when the mediator is strongly associated with the outcome.

The recent parametric estimator proposed by Samoilenko *et al.* (2018) and Gaynor *et al.* (2019) for a binary mediator is attractive, since it overcomes the marginal or conditional verification of the ROA. However, more work is required to fully develop inference. In the paper by Samoilenko *et al.* (2018), the variance computation for the $\mathrm{OR}^{\mathrm{NDE}}$ and $\mathrm{OR}^{\mathrm{NIE}}$ estimators was done using bootstrapping only. In Gaynor *et al.* (2019), the standard error formulas were not provided in the paper but were implemented in a R code (R Foundation for Statistical Computing, Vienna, Austria) developed for scenarios based on specific data sets. In the paper by Doretti *et al.* (2019), the exact parametric formulas for the natural effects on the log OR scale were extended for all possible interactions in the outcome model (including exposure-mediator-confounding covariates' interactions); corresponding expressions for standard errors were derived using the delta method. However, the authors did not release computer code to provide easy implementation.

The purpose of our article is two-fold. Our first objective is to provide explicit and straightforward formulas for the delta method standard errors for the case of the mediator-exposure interaction and make this option available in the general SAS macro developed in the paper by Samoilenko *et al.* (2018). While the bootstrap is indicated for inference on the indirect effect (Hayes et Little, 2018), it is more computer-intensive and not assumption-free (Davison et Hinkley, 1997; Efron et Tibshirani, 1994). Therefore, providing both delta and percentile bootstrap confidence intervals allows for greater flexibility and increased confidence in mediation results. Our second objective is to go beyond the OR scale and provide analogous results for the NDE and NIE on the risk ratio (RR) and risk difference (RD) scales, with all three scales using the same logistic model for the outcome.

4.2 Methods

4.2.1 Models and nested counterfactual outcome probabilities

As in the papers by Samoilenko *et al.* (2018) and Gaynor *et al.* (2019), we assume the following logistic regression models for the binary mediator M and binary outcome Y, respectively:

$$\text{logit}\{P(M = 1|A = a, \boldsymbol{C} = \boldsymbol{c})\} = \beta_0 + \beta_1 x + \boldsymbol{\beta}_2^{'}\boldsymbol{c}, \tag{4.1}$$

$$\text{logit}\{P(Y = 1|A = a, M = m, \boldsymbol{C} = \boldsymbol{c})\} = \theta_0 + \theta_1 a + \theta_2 m + \theta_3 am + \boldsymbol{\theta}_4^{'}\boldsymbol{c}, \tag{4.2}$$

where A is the exposure (binary or continuous) and $\boldsymbol{C}$ is the set of covariates sufficient to control for exposure-outcome, mediator-outcome, and exposure-mediator confounding (VanderWeele, 2016).

Under identification assumptions (VanderWeele et Vansteelandt, 2009) and modelling assumptions (4.1-4.2), the nested counterfactual outcome $Y\left(a, M(a^*)\right)$ probability is expressed as

$$\begin{aligned} &P(Y(a, M(a^*)) = 1|\boldsymbol{C} = \boldsymbol{c}) \\ &\quad = \text{expit}\left(\theta_0 + \theta_1 a + \theta_2 + \theta_3 a + \boldsymbol{\theta}_4^{'}\boldsymbol{c}\right)\text{expit}\left(\beta_0 + \beta_1 a^* + \boldsymbol{\beta}_2^{'}\boldsymbol{c}\right) \\ &\qquad + \text{expit}\left(\theta_0 + \theta_1 a + \boldsymbol{\theta}_4^{'}\boldsymbol{c}\right)\left(1 - \text{expit}\left(\beta_0 + \beta_1 a^* + \boldsymbol{\beta}_2^{'}\boldsymbol{c}\right)\right), \end{aligned} \tag{4.3}$$

where

$$\text{expit}(\alpha) = \frac{\exp(\alpha)}{1 + \exp(\alpha)}, \quad 1 - \text{expit}(\alpha) = (1 + \exp(\alpha))^{-1}.$$

Generally, the NDE compares $Y\left(a, M(a^*)\right)$ to $Y\left(a^*, M(a^*)\right)$, while the NIE is defined as a contrast between $Y\left(a, M(a)\right)$ and $Y\left(a, M(a^*)\right)$. In the literature, the NDE and NIE are also referred to as the pure (natural) direct effect and the total (natural) indirect effect, respectively (De Stavola *et al.*, 2015; Robins et Greenland, 1992; Wang et Arah, 2015).

Equation (4.3) allows expression of the NDE and NIE ORs (OR^{NDE} and OR^{NIE}), as well as the NDE and NIE RRs (RR^{NDE} and RR^{NIE}) and the NDE and NIE RDs (RD^{NDE} and RD^{NIE}), in an exact manner.

4.2.2 Natural direct and indirect effects on the OR, RR and RD scales

Explicit expressions for the (conditional) NDE and NIE ORs, $OR^{NDE}_{a,a^*|\boldsymbol{c}}$ and $OR^{NIE}_{a,a^*|\boldsymbol{c}}$, corresponding to a change in the exposure level from $A = a^*$ to $A = a$ (also see Samoilenko *et al.* (2018) and Gaynor *et al.* (2019)) are derived using the nested counterfactual outcome probabilities defined in Equation (4.3) as follows:

$$\begin{aligned} OR^{NDE}_{a,a^*|\boldsymbol{c}} &= \frac{\dfrac{P(Y(a, M(a^*)) = 1|\boldsymbol{C} = \boldsymbol{c})}{1 - P(Y(a, M(a^*)) = 1|\boldsymbol{C} = \boldsymbol{c})}}{\dfrac{P(Y(a^*, M(a^*)) = 1|\boldsymbol{C} = \boldsymbol{c})}{1 - P(Y(a^*, M(a^*)) = 1|\boldsymbol{C} = \boldsymbol{c})}}, \\ OR^{NIE}_{a,a^*|\boldsymbol{c}} &= \frac{\dfrac{P(Y(a, M(a)) = 1|\boldsymbol{C} = \boldsymbol{c})}{1 - P(Y(a, M(a)) = 1|\boldsymbol{C} = \boldsymbol{c})}}{\dfrac{P(Y(a, M(a^*)) = 1|\boldsymbol{C} = \boldsymbol{c})}{1 - P(Y(a, M(a^*)) = 1|\boldsymbol{C} = \boldsymbol{c})}}. \end{aligned} \tag{4.4}$$

In an analogous manner, Equation (4.3) leads to exact NDE and NIE RR expressions, $RR^{NDE}_{a,a^*|\boldsymbol{c}}$ and $RR^{NIE}_{a,a^*|\boldsymbol{c}}$, respectively:

$$\begin{aligned} RR^{NDE}_{a,a^*|\boldsymbol{c}} &= \frac{P(Y(a, M(a^*)) = 1|\boldsymbol{C} = \boldsymbol{c})}{P(Y(a^*, M(a^*)) = 1|\boldsymbol{C} = \boldsymbol{c})}, \\ RR^{NIE}_{a,a^*|\boldsymbol{c}} &= \frac{P(Y(a, M(a)) = 1|\boldsymbol{C} = \boldsymbol{c})}{P(Y(a, M(a^*)) = 1|\boldsymbol{C} = \boldsymbol{c})}. \end{aligned} \tag{4.5}$$

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

$$OR^{TE}_{a,a^*|\boldsymbol{c}} = OR^{NDE}_{a,a^*|\boldsymbol{c}} \cdot OR^{NIE}_{a,a^*|\boldsymbol{c}}, \quad RR^{TE}_{a,a^*|\boldsymbol{c}} = RR^{NDE}_{a,a^*|\boldsymbol{c}} \cdot RR^{NIE}_{a,a^*|\boldsymbol{c}}. \tag{4.6}$$

From Equation (4.3), the NDE and NIE exact expressions on the RD scale are

$$\begin{aligned} RD^{NDE}_{a,a^*|\boldsymbol{c}} &= P(Y(a, M(a^*)) = 1|\boldsymbol{C} = \boldsymbol{c}) - P(Y(a^*, M(a^*)) = 1|\boldsymbol{C} = \boldsymbol{c}), \\ RD^{NIE}_{a,a^*|\boldsymbol{c}} &= P(Y(a, M(a)) = 1|\boldsymbol{C} = \boldsymbol{c}) - P(Y(a, M(a^*)) = 1|\boldsymbol{C} = \boldsymbol{c}). \end{aligned} \tag{4.7}$$

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

$$RD^{TE}_{a,a^*|\boldsymbol{c}} = RD^{NDE}_{a,a^*|\boldsymbol{c}} + RD^{NIE}_{a,a^*|\boldsymbol{c}}.$$

For each effect scale, the NDE and NIE estimators are induced by replacing the coefficients in Equations (4.1) and (4.2) with corresponding estimators. The formulas for calculating the natural-effects standard errors via the delta method are provided in Appendix 1 (Section 4.6).

4.2.3 Valeri and VanderWeele (2013) approximate NDE and NIE approach

As detailed by Samoilenko *et al.* (2018), the approximate expressions for the OR^{NDE} and OR^{NIE} provided in the paper by Valeri and VanderWeele (2013) are obtained by invoking the ROA multiple times. First, replace in Equation (4.3) the expit functions stemming from the outcome model with exponential functions; and second, approximate the OR by RR, that is, replace Equation (4.4) with Equation (4.5):

$$\begin{aligned} &P_{app}(Y(a, M(a^*)) = 1|\boldsymbol{C} = \boldsymbol{c}) \\ &\quad = \exp\left(\theta_0 + \theta_1 a + \theta_2 + \theta_3 a + \boldsymbol{\theta}_4' \boldsymbol{c}\right) \text{expit}\left(\beta_0 + \beta_1 a^* + \boldsymbol{\beta}_2' \boldsymbol{c}\right) \\ &\qquad + \exp\left(\theta_0 + \theta_1 a + \boldsymbol{\theta}_4' \boldsymbol{c}\right)\left(1 - \text{expit}\left(\beta_0 + \beta_1 a^* + \boldsymbol{\beta}_2' \boldsymbol{c}\right)\right); \end{aligned} \tag{4.8}$$

$$\begin{aligned} app\ OR^{NDE}_{a,a^*|\boldsymbol{c}} &= \frac{P_{app}(Y(a, M(a^*)) = 1|\boldsymbol{C} = \boldsymbol{c})}{P_{app}(Y(a^*, M(a^*)) = 1|\boldsymbol{C} = \boldsymbol{c})}, \\ app\ OR^{NIE}_{a,a^*|\boldsymbol{c}} &= \frac{P_{app}(Y(a, M(a)) = 1|\boldsymbol{C} = \boldsymbol{c})}{P_{app}(Y(a, M(a^*)) = 1|\boldsymbol{C} = \boldsymbol{c})}. \end{aligned} \tag{4.9}$$

The approximate expression for the TE is then given by

$$app\ OR^{TE}_{a,a^*|\boldsymbol{c}} = app\ OR^{NDE}_{a,a^*|\boldsymbol{c}} \cdot app\ OR^{NIE}_{a,a^*|\boldsymbol{c}}. \tag{4.10}$$

4.2.4 Simulation studies

We conducted two simulation studies to examine the behavior of proposed exact estimators. In the first simulation study, no covariates $\boldsymbol{C}$ were included for the sake of simplicity, while two covariates were included in the second study. Both studies considered four scenarios corresponding to different levels of outcome rareness/commonness:

Scenario 1. The outcome is rare in all of the strata defined by the binary exposure and binary mediator (conditional probabilities $P(Y = 1|A = i, M = j) \leq 10\%, \quad i, j = 0, 1$).

Scenario 2. The outcome is rare marginally ($P(Y = 1) \leq 10\%$), but it is not rare in one stratum defined by the binary exposure and binary mediator.

Scenario 3. This scenario is similar to *Scenario 2*, but it features two common strata and a slightly increased marginal outcome probability ($P(Y = 1) \approx 15\%$).

Scenario 4. The outcome is not rare marginally (is common) with $P(Y = 1) \approx 40\%$.

4.2.4.1 Simulation study without covariates

For each scenario, we generated 1000 independent samples of size $n = 5000$ nonparametrically using sequential Bernoulli sampling for A, M, and Y. The probability values used to generate the exposure, mediator and outcome variables are presented in Table 4.1.

Table 4.1 Data-generating mechanisms for a simulation study without covariates conducted to evaluate proposed exact estimators

Simulation parameters[a]	Scenario 1	Scenario 2	Scenario 3	Scenario 4
$P(A = 1)$	0.40	0.40	0.40	0.40
$P(M = 1\|A = 0)$	0.10	0.10	0.10	0.10
$P(M = 1\|A = 1)$	0.20	0.20	0.20	0.20
$P(Y = 1\|A = 0, M = 0)$	0.03	0.03	0.15	0.30
$P(Y = 1\|A = 0, M = 1)$	0.08	0.08	0.10	0.70
$P(Y = 1\|A = 1, M = 0)$	0.07	0.07	0.07	0.40
$P(Y = 1\|A = 1, M = 1)$	0.10	0.50	0.50	0.80
Marginal outcome probability	0.05	0.08	0.15	0.40

[a] A, binary exposure; M, binary mediator; P, probability; Y, binary outcome.

The true mediation OR, RR and RD effects were calculated as

$$\begin{aligned} true\ OR^{NDE} &= \frac{P_{10}/(1-P_{10})}{P_{00}/(1-P_{00})}, \quad true\ OR^{NIE} = \frac{P_{11}/(1-P_{11})}{P_{10}/(1-P_{10})}, \\ true\ RR^{NDE} &= \frac{P_{10}}{P_{00}}, \quad true\ RR^{NIE} = \frac{P_{11}}{P_{10}}, \\ true\ RD^{NDE} &= P_{10} - P_{00}, \quad true\ RD^{NIE} = P_{11} - P_{10}, \end{aligned} \tag{4.11}$$

with P_{11}, P_{10}, P_{00} computed using values from Table 4.1:

$$P_{ij} = P\left(Y=1|A=i,M=1\right)\cdot P\left(M=1|A=j\right) + P\left(Y=1|A=i,M=0\right)\cdot\left(1-P\left(M=1|A=j\right)\right).$$

The true total causal effects were calculated correspondingly as

$$\begin{aligned} true\ OR^{TE} &= true\ OR^{NDE} \cdot true\ OR^{NIE}, \\ true\ RR^{TE} &= true\ RR^{NDE} \cdot true\ RR^{NIE}, \\ true\ RD^{TE} &= true\ RD^{NDE} + true\ RD^{NIE}. \end{aligned}$$

For each sample, exact estimates of natural direct and indirect effects were calculated on the OR, RR and RD scales. The mean value, bias, relative bias, standard deviation, and root mean squared error of proposed exact estimators were then estimated over the 1000 samples generated [1]; the true ORs, RRs and RDs defined in Equation (4.11) were used as the gold standard. For each simulation scenario, the same statistics were also calculated for the approximate natural-effects estimator based on Equations (4.8-4.10). The approximate natural-effects OR estimator was evaluated in regard to both multiplicative scales (OR and RR). Indeed, because the approximate natural effects are generally reported as ORs (Oberg *et al.*, 2018), we first compared the approximate natural effect estimates to the true ORs. However, since the approximate ORs mimic RRs by construction (see correspondence between Equations (4.5) and (4.9)), we also evaluated the performance of the approximate estimator using the true RRs as the reference. The calculations described above were performed using SAS, Version 9.5.

1. Ici, l'écart-type (*standard deviation*, *SD*) est calculé selon la formule $SD = \sqrt{\frac{1}{1000-1}\sum_{i=1}^{1000}\left(\hat{\theta}_i - \overline{\hat{\theta}_i}\right)^2}$ et constitue une mesure de précision de l'estimateur étudié. Le terme *empirical standard error* est aussi utilisé dans la littérature (Morris *et al.*, 2019). La racine de l'erreur quadratique moyenne (*root mean squared error*, *RMSE*) est calculée selon la formule $RMSE = \sqrt{\frac{1}{1000}\sum_{i=1}^{1000}\left(\hat{\theta}_i - \theta\right)^2}$; cette mesure représente un moyen naturel d'intégrer le biais et la précision de l'estimateur dans une seule mesure de performance.

For each scenario and sample, we also considered two other existing approaches for comparison with the exact method being introduced here. For all three scales (OR, RR, and RD), we applied the natural effect model (NEM) approach (Lange *et al.*, 2012, 2017) using the R package `medflex` (Steen *et al.*, 2017). This approach is not based on the ROA and directly parameterizes the natural effects. Two procedures, weighting and imputation, are implemented in `medflex`; we used the weighting one which requires specifying a regression model for the mediator and a NEM for the counterfactual outcome. A logistic model was specified for the mediator for all scales. NEMs $g\left(E\left\{Y\left(a, M(a^*)\right)\right\}\right) = \gamma_0 + \gamma_1 a + \gamma_2 a^* + \gamma_3 a a^*$, where $g(\cdot)$ is a link function, were fitted using logistic, log-binomial and linear regressions for the OR, RR and RD scales, respectively. For the RD scale, we also applied Imai *et al.* (2010)'s *Parametric Inference Algorithm*, implemented in the R package `mediation` (Tingley *et al.*, 2014). This causal approach, which also does not rely on the ROA, is based on quasi-Bayesian Monte Carlo approximations and is provided as the default option in `mediation`. A logistic model was specified for the mediator as well as for the outcome, where the latter included a treatment-mediator interaction term as in the exact and approximate approaches; 1000 Monte Carlo draws were used for each sample generated. Note that `mediation` version 4.5.0 returns NDE and NIE estimates on the RD scale only.

We computed the coverage probabilities of 95% confidence interval estimators by calculating the proportion of times confidence intervals enclosed corresponding true values of the NDE, NIE and TE. For the exact and approximate approaches, 95% confidence intervals were constructed by percentile bootstrap based on 500 resamples with replacement (Chernick, 2011) and using the first-order delta method. For the NEM approach, 95% confidence intervals were obtained using robust standard errors based on the sandwich estimator (Liang et Zeger, 1986). For the quasi-Bayesian approach, 95% confidence intervals were based on the White's heteroskedasticity-consistent estimator for the covariance matrix (Tingley *et al.*, 2014).

4.2.4.2 Simulation study with covariates

In all scenarios, covariates C_1 and C_2 were generated independently as $Bernoulli(0.5)$ and $\mathcal{N}(0,1)$, respectively. The binary exposure A was generated according to the following model:

$$\text{logit}\left\{P\left(A = 1 | C_1 = c_1, C_2 = c_2\right)\right\} = -0.5 + 0.1c_1 - 0.15c_2.$$

Then, the binary mediator M and outcome Y were respectively generated under models

$$\text{logit}\left\{P\left(M=1|A=a,C_1=c_1,C_2=c_2\right)\right\}=\beta_0+\beta_1 a+\beta_{21}c_1+\beta_{22}c_2$$

and

$$\text{logit}\left\{P\left(Y=1|A=a,M=m,C_1=c_1,C_2=c_2\ \right)\right\}=\theta_0+\theta_1 a+\theta_2 m+\theta_3 am+\theta_{41}c_1+\theta_{42}c_2,$$

where $\beta_0=-2.3$, $\beta_1=0.8$, $\beta_{21}=0.2$, $\beta_{22}=0.25$. The outcome simulation parameters are presented in Table 4.2 for each simulation scenario. Under these parameter values, the stratum-specific outcome prevalences were similar to those from the simulations without covariates.

Table 4.2 Data-generating mechanisms for the simulation study with covariates: Outcome simulation parameters

Outcome simulation parameters	Scenario 1	Scenario 2	Scenario 3	Scenario 4
θ_0	-3.60	-3.60	-1.85	-0.90
θ_1	0.80	0.80	-0.85	0.40
θ_2	1.00	1.00	-0.50	1.50
θ_3	-0.60	1.50	3.00	0.40
θ_{41}	0.25	0.25	0.25	0.25
θ_{42}	0.20	0.20	0.20	0.20
Marginal outcome probability	0.049	0.077	0.148	0.405

The true mediation OR, RR and RD effects (gold standard) were calculated using simulation parameters according to Equation (4.11) , where

$$\begin{aligned}P_{ij}=&\ \text{expit}\left(\theta_0+\theta_1 i+\theta_2+\theta_3 i+\theta_{41}\overline{c_1}+\theta_{42}\overline{c_2}\right)\text{expit}\left(\beta_0+\beta_1 j+\beta_{21}\overline{c_1}+\beta_{22}\overline{c_2}\right)\\&+\text{expit}\left(\theta_0+\theta_1 i+\theta_{41}\overline{c_1}+\theta_{42}\overline{c_2}\right)\left(1-\text{expit}\left(\beta_0+\beta_1 j+\beta_{21}\overline{c_1}+\beta_{22}\overline{c_2}\right)\right)\end{aligned}$$

and $\overline{c_1}=0.5$, $\overline{c_2}=0$.

The simulation study with covariates was conducted the same way as the one without covariates regarding number of samples generated, sample size, and estimators investigated. However, for the RR scale in *Scenario 4*, the NEM was fitted using a Poisson regression model instead of log-binomial because of failed convergence of the latter model for 77.6% of samples generated. For all approaches, models included covariates as main-effect terms only, and mediation effects were estimated at the

sample-specific mean values for C_1 and C_2. Note that in absence of exposure-covariate interactions, the conditional mediation effects returned by `medflex` are the same for any level of adjustment covariates (Starkopf *et al.*, 2017).

The decomposition property of the exact and approximate TE estimators was examined in both simulation studies (see Appendix 1, Section 4.6). Further details on the estimation procedures are provided in Appendix 1 (Section 4.6).

4.3 Results

The performance of the proposed exact natural-effects estimators on the OR, RR and RD scales is summarized in Tables 4.3 - 4.5 and Tables 4.6 - 4.8 for the simulation studies without covariates and with covariates, respectively (type of estimator = *exact*).

For the multiplicative scales, the means values of exact NDE, NIE and TE estimates were very close to corresponding true values for each scenario and each type of simulation, with relative bias values ranging between -0.34% and 1.35%. All exact interval estimators (bootstrap and delta method) yielded coverage probability values close to 95%. For the simulations without covariates, the exact results were almost identical to those returned by the NEM approach (results omitted from tables), while they were very close in the simulations with covariates. The exact results were also very close to those obtained using the quasi-Bayesian approach (for RD scale; see Table 4.5 and Table 4.8).

The results for the approximate natural-effects estimator in the simulation studies without and with covariates under increasing degrees of the ROA violation are presented in Tables 4.3 and 4.4 and Tables 4.6 and 4.7, respectively (type of estimator = *approximate*). In *Scenario 1* (rare outcome in all strata defined by A and M), the approximate OR estimator demonstrated small relative bias values when either the true ORs or the true RRs were used as reference values (between 0.13% and 5.24%). Corresponding coverage probabilities by the delta method and bootstrap were close to the 95% nominal level. For *Scenario 2*, where the outcome Y is rare marginally, but not rare in the stratum defined by $A = 1$ and $M = 1$, we observed relative bias values ranging between 5.93% and 62.6% and a significant decrease in coverage probability values. The same tendencies for relative biases and coverage probabilities were seen for *Scenario 3*. For *Scenario 4*, which violated the ROA in all strata defined by A and M, we obtained relative bias values up to 69.62% and coverage probability values equal to 0% in some cases.

The TE estimates obtained from the exact approach by the multiplication of corresponding NDE and NIE estimates were closer to the nonmediated TE estimates as compared to the approximate approach (Tables 4.10 and 4.11).

Table 4.3 Exact and approximate natural-effects estimators on the odds ratio scale in scenarios with increasing levels of outcome commonness (simulation study without covariates based on 1000 independent samples of size $n = 5000$)

Effect	Estimator	True value	Mean	Bias	Relative bias (%)	SD	RMSE	Delta CP (%)	Bootstrap CP (%)
				Scenario 1					
NDE OR	exact[a]	2.171	2.197	0.026	1.20	0.297	0.299	95.7	94.9
	approx.[b]		2.184	0.013	0.59	0.297	0.297	95.6	95.1
NIE OR	exact	1.044	1.046	0.001	0.12	0.027	0.027	93.5	93.7
	approx.		1.047	0.003	0.24	0.028	0.028	93.4	93.7
TE OR	exact	2.268	2.296	0.028	1.25	0.304	0.305	95.3	95.2
	approx.		2.285	0.018	0.77	0.305	0.305	95.4	95.2
				Scenario 2					
NDE OR	exact	3.512	3.556	0.044	1.25	0.436	0.438	94.6	94.7
	approx.		4.663	1.151	32.77	0.600	1.298	40.6	38.1
NIE OR	exact	1.451	1.454	0.004	0.24	0.066	0.066	93.8	93.5
	approx.		1.555	0.104	7.18	0.080	0.131	74.2	72.7
TE OR	exact	5.096	5.165	0.069	1.35	0.616	0.620	95.5	95.1
	approx.		7.248	2.151	42.22	0.971	2.361	26.0	24.3
				Scenario 3					
NDE OR	exact	0.751	0.753	0.001	0.17	0.064	0.064	95.8	95.8
	approx.		0.992	0.241	32.11	0.094	0.259	17.7	18.3
NIE OR	exact	1.451	1.454	0.004	0.24	0.066	0.066	93.8	93.5
	approx.		1.555	0.104	7.18	0.080	0.131	74.2	72.7
TE OR	exact	1.090	1.093	0.003	0.27	0.087	0.087	96.3	95.9
	approx.		1.542	0.452	41.49	0.156	0.478	7.1	7.1
				Scenario 4					
NDE OR	exact	1.525	1.525	-0.001	-0.04	0.090	0.090	95.3	95.3
	approx.		1.616	0.091	5.97	0.129	0.158	90.8	90.1
NIE OR	exact	1.175	1.175	0.000	0.02	0.023	0.023	94.7	95.1
	approx.		1.335	0.160	13.61	0.052	0.168	6.3	5.2
TE OR	exact	1.791	1.792	-0.001	-0.04	0.105	0.105	95.4	95.5
	approx.		2.159	0.367	20.49	0.210	0.423	56.1	53.1

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] Exact estimator proposed.

[b] Approximate estimator of Valeri and VanderWeele (2013).

Table 4.4 Exact and approximate natural-effects estimators on the risk ratio scale in scenarios with increasing levels of outcome commonness (simulation study without covariates based on 1000 independent samples of size $n = 5000$)

Effect	Estimator	True value	Mean	Bias	Relative bias (%)	SD	RMSE	Delta CP (%)	Bootstrap CP (%)
				Scenario 1					
NDE RR	exact[a]	2.086	2.109	0.023	1.11	0.271	0.272	95.5	94.9
	approx.[b]		2.184	0.098	4.72	0.297	0.313	94.5	94.0
NIE RR	exact	1.041	1.042	0.001	0.11	0.025	0.025	93.5	93.8
	approx.		1.047	0.006	0.57	0.028	0.028	94.6	93.9
TE RR	exact	2.171	2.197	0.025	1.16	0.276	0.277	95.2	95.0
	approx.		2.285	0.114	5.24	0.305	0.325	94.9	94.2
				Scenario 2					
NDE RR	exact	3.229	3.266	0.037	1.16	0.377	0.379	94.9	94.5
	approx.		4.663	1.435	44.44	0.600	1.555	17.1	15.3
NIE RR	exact	1.381	1.383	0.003	0.20	0.055	0.055	94.0	93.7
	approx.		1.555	0.175	12.64	0.080	0.192	36.1	35.1
TE RR	exact	4.457	4.513	0.055	1.24	0.504	0.507	95.3	95.0
	approx.		7.248	2.790	62.60	0.971	2.955	3.6	3.0
				Scenario 3					
NDE RR	exact	0.779	0.780	0.001	0.10	0.058	0.058	95.8	95.7
	approx.		0.992	0.213	27.34	0.094	0.233	29.1	28.7
NIE RR	exact	1.381	1.383	0.003	0.20	0.055	0.055	94.0	93.7
	approx.		1.555	0.175	12.64	0.080	0.192	36.1	35.1
TE RR	exact	1.076	1.078	0.002	0.18	0.072	0.072	96.3	95.9
	approx.		1.542	0.466	43.33	0.156	0.491	5.6	5.6
				Scenario 4					
NDE RR	exact	1.294	1.293	-0.001	-0.08	0.046	0.046	95.2	95.2
	approx.		1.616	0.322	24.89	0.129	0.347	20.9	22.1
NIE RR	exact	1.091	1.091	0.000	0.01	0.012	0.012	94.6	95.1
	approx.		1.335	0.244	22.35	0.052	0.249	0.0	0.0
TE RR	exact	1.412	1.411	-0.001	-0.08	0.048	0.048	95.5	95.4
	approx.		2.159	0.747	52.93	0.210	0.776	0.7	0.7

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

[a] Exact estimator proposed.

[b] Approximate estimator of Valeri and VanderWeele (2013).

Table 4.5 Natural-effects estimators on the risk difference scale in scenarios with increasing levels of outcome commonness (simulation study without covariates based on 1000 independent samples of size $n = 5000$)

Effect	Estimator	True value	Mean	Bias	Relative bias (%)	SD	RMSE	Delta/ Robust SE CP (%)[a]	Boot-strap CP (%)
				Scenario 1					
NDE RD	exact[b]	0.038	0.038	0.000	0.04	0.007	0.007	95.8	95.4
	mediation[c]		0.038	0.000	0.16	0.007	0.007	95.8	-
NIE RD	exact	0.003	0.003	0.000	1.26	0.002	0.002	93.7	93.3
	mediation		0.003	0.000	3.81	0.002	0.002	94.4	-
TE RD	exact	0.041	0.041	0.000	0.13	0.007	0.007	96.1	95.8
	mediation		0.041	0.000	0.42	0.007	0.007	95.8	-
				Scenario 2					
NDE RD	exact	0.078	0.078	0.000	0.07	0.007	0.007	95.4	95.1
	mediation		0.078	0.000	0.06	0.007	0.007	95.4	-
NIE RD	exact	0.043	0.043	0.000	0.27	0.005	0.005	94.3	94.2
	mediation		0.043	0.000	0.25	0.005	0.005	97.4	-
TE RD	exact	0.121	0.121	0.000	0.14	0.009	0.009	95.9	95.8
	mediation		0.121	0.000	0.13	0.009	0.009	96.8	-
				Scenario 3					
NDE RD	exact	-0.032	-0.032	-0.000	0.39	0.009	0.009	95.8	95.4
	mediation		-0.032	-0.000	0.22	0.009	0.009	96.0	-
NIE RD	exact	0.043	0.043	0.000	0.27	0.005	0.005	94.3	94.2
	mediation		0.043	0.000	0.25	0.005	0.005	97.5	-
TE RD	exact	0.011	0.011	-0.000	-0.10	0.010	0.010	96.4	96.0
	mediation		0.011	0.000	0.33	0.010	0.010	96.8	-
				Scenario 4					
NDE RD	exact	0.10	0.099	-0.001	-0.50	0.014	0.014	95.0	95.4
	mediation		0.099	-0.000	-0.52	0.014	0.014	95.0	-
NIE RD	exact	0.04	0.040	-0.000	-0.06	0.005	0.005	94.6	95.2
	mediation		0.040	-0.000	-0.21	0.005	0.005	97.5	-
TE RD	exact	0.14	0.139	-0.001	-0.38	0.014	0.014	95.4	95.2
	mediation		0.139	-0.001	-0.43	0.014	0.014	96.0	-

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 for the exact estimator; for *mediation*, the 95% confidence intervals were based on the White's heteroskedasticity-consistent estimator for the covariance matrix (Tingley *et al.*, 2014).

[b] Exact estimator proposed.

[c] Qquasi-Bayesian approach of Imai *et al.* (2010) implemented in the R package `mediation` (Tingley *et al.*, 2014).

Table 4.6 Natural-effects estimators on the odds ratio scale in scenarios with increasing levels of outcome commonness (simulation study with covariates based on 1000 independent samples of size $n = 5000$)

Effect	Estimator[a]	True value	Mean	Bias	Relative bias (%)	SD	RMSE	CP (%)[b]
				Scenario 1				
NDE OR	exact	2.004	2.024	0.020	1.00	0.278	0.279	95.0
	approx.		2.013	0.009	0.45	0.277	0.278	95.2
	medflex		2.012	0.008	0.42	0.275	0.276	95.3
NIE OR	exact	1.045	1.045	0.000	0.01	0.027	0.027	93.6
	approx.		1.046	0.001	0.13	0.028	0.028	93.6
	medflex		1.046	0.001	0.13	0.028	0.028	93.8
TE OR	exact	2.094	2.114	0.020	0.95	0.286	0.287	95.8
	approx.		2.105	0.011	0.52	0.286	0.287	95.6
	medflex		2.104	0.010	0.49	0.284	0.285	95.7
				Scenario 2				
NDE OR	exact	3.206	3.239	0.033	1.02	0.398	0.399	95.4
	approx.		4.065	0.859	26.79	0.526	1.007	54.8
	medflex		3.265	0.060	1.86	0.401	0.405	95.3
NIE OR	exact	1.433	1.432	-0.000	-0.02	0.066	0.066	94.8
	approx.		1.517	0.085	5.93	0.078	0.115	80.5
	medflex		1.434	0.001	0.09	0.066	0.066	95.0
TE OR	exact	4.593	4.633	0.040	0.87	0.560	0.562	95.1
	approx.		6.166	1.573	34.25	0.840	1.783	41.9
	medflex		4.676	0.084	1.82	0.566	0.572	94.9
				Scenario 3				
NDE OR	exact	0.736	0.735	-0.001	-0.16	0.064	0.064	94.6
	approx.		0.941	0.205	27.81	0.089	0.223	27.6
	medflex		0.746	0.010	1.34	0.065	0.066	94.5
NIE OR	exact	1.425	1.424	-0.001	-0.04	0.063	0.063	94.7
	approx.		1.517	0.092	6.48	0.077	0.120	78.0
	medflex		1.424	-0.001	-0.04	0.063	0.063	94.8
TE OR	exact	1.049	1.046	-0.004	-0.34	0.086	0.086	94.2
	approx.		1.427	0.378	36.02	0.146	0.405	16.1
	medflex		1.062	0.012	1.17	0.088	0.089	94.0

Table 4.6 Natural-effects estimators on the odds ratio scale in scenarios with increasing levels of outcome commonness (simulation study with covariates based on 1000 independent samples of size $n = 5000$; continuation)

Effect	Estimator[a]	True value	Mean	Bias	Relative bias (%)	SD	RMSE	CP (%)[b]
				Scenario 4				
NDE OR	exact	1.504	1.500	-0.003	-0.23	0.090	0.090	95.0
	approx.		1.740	0.237	15.74	0.147	0.279	59.5
	medflex		1.503	-0.000	-0.03	0.091	0.091	95.4
NIE OR	exact	1.177	1.177	-0.001	-0.05	0.024	0.024	94.9
	approx.		1.356	0.179	15.23	0.058	0.188	4.9
	medflex		1.176	-0.001	-0.08	0.024	0.024	94.7
TE OR	exact	1.770	1.765	-0.005	-0.30	0.107	0.107	94.7
	approx.		2.363	0.593	33.52	0.250	0.644	20.9
	medflex		1.768	-0.002	-0.13	0.107	0.107	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] *exact*: exact estimator proposed; *approx.*: approximate estimator by Valeri and VanderWeele (2013); *medflex*: natural effect model approach (Lange *et al.*, 2012) using weighting method implemented in the R package `medflex` (Steen *et al.*, 2017).

[b] Delta method for exact and approximate estimators, robust standard error for *medflex*.

Table 4.7 Natural-effects estimators on the risk ratio scale in scenarios with increasing levels of outcome commonness (simulation study with covariates based on 1000 independent samples of size $n = 5000$)

Effect	Estimator[a]	True value	Mean	Bias	Relative bias (%)	SD	RMSE	CP (%)[b]
				Scenario 1				
NDE RR	exact	1.936	1.954	0.018	0.92	0.255	0.256	95.4
	approx.		2.013	0.077	3.96	0.277	0.288	94.2
	medflex		1.935	-0.001	-0.06	0.250	0.250	95.1
NIE RR	exact	1.042	1.042	0.000	0.01	0.025	0.025	93.5
	approx.		1.046	0.004	0.43	0.028	0.029	94.0
	medflex		1.043	0.001	0.10	0.026	0.026	93.8
TE RR	exact	2.017	2.035	0.018	0.88	0.262	0.262	96.0
	approx.		2.105	0.088	4.34	0.286	0.300	94.0
	medflex		2.017	-0.000	-0.01	0.258	0.258	95.4
				Scenario 2				
NDE RR	exact	2.977	3.006	0.029	0.96	0.349	0.350	95.4
	approx.		4.065	1.087	36.52	0.526	1.208	31.7
	medflex		2.999	0.022	0.74	0.344	0.345	95.0
NIE RR	exact	1.371	1.371	-0.000	-0.01	0.056	0.056	94.9
	approx.		1.517	0.146	10.68	0.078	0.166	48.6
	medflex		1.367	-0.004	-0.32	0.055	0.055	94.3
TE RR	exact	4.082	4.117	0.034	0.840	0.466	0.467	94.7
	approx.		6.166	2.083	51.03	0.84	2.246	14.3
	medflex		4.095	0.013	0.31	0.459	0.459	94.8
				Scenario 3				
NDE RR	exact	0.766	0.764	-0.001	-0.19	0.058	0.058	94.8
	approx.		0.941	0.175	22.90	0.089	0.197	41.7
	medflex		0.778	0.012	1.60	0.059	0.060	94.1
NIE RR	exact	1.360	1.360	-0.000	-0.03	0.053	0.053	94.7
	approx.		1.517	0.157	11.53	0.077	0.175	42.2
	medflex		1.357	-0.003	-0.25	0.052	0.052	94.3
TE RR	exact	1.042	1.038	-0.004	-0.34	0.073	0.073	94.4
	approx.		1.427	0.385	37.00	0.146	0.412	14.2
	medflex		1.054	0.013	1.23	0.073	0.075	93.6

Table 4.7 Natural-effects estimators on the risk ratio scale in scenarios with increasing levels of outcome commonness (simulation study with covariates based on 1000 independent samples of size $n = 5000$; continuation)

Effect	Estimator[a]	True value	Mean	Bias	Relative bias (%)	SD	RMSE	CP (%)[b]
				Scenario 4				
NDE RR	exact	1.278	1.275	-0.002	-0.17	0.046	0.046	95.1
	approx.		1.740	0.463	36.21	0.147	0.485	3.1
	medflex		1.269	-0.009	-0.68	0.044	0.045	93.9
NIE RR	exact	1.090	1.090	-0.000	-0.01	0.012	0.012	95.0
	approx.		1.356	0.266	24.39	0.058	0.272	0.0
	medflex		1.086	-0.004	-0.39	0.011	0.012	92.0
TE RR	exact	1.393	1.391	-0.003	-0.20	0.049	0.049	94.8
	approx.		2.363	0.970	69.62	0.250	1.002	0.0
	medflex		1.378	-0.015	-1.08	0.047	0.049	92.2

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

[a] *exact*: exact estimator proposed; *approx.*: approximate estimator by Valeri and VanderWeele (2013); *medflex*: natural effect model approach (Lange *et al.*, 2012) using weighting method implemented in the R package `medflex` (Steen *et al.*, 2017).

[b] Delta method for exact and approximate estimators, robust standard error for *medflex*.

Table 4.8 Natural-effects estimators on the risk difference scale in scenarios with increasing levels of outcome commonness (simulation study with covariates based on 1000 independent samples of size $n = 5000$)

Effect	Estimator[a]	True value	Mean	Bias	Relative bias (%)	SD	RMSE	CP (%)[b]
				Scenario 1				
	exact		0.032	-0.000	-0.46	0.007	0.007	95.8
NDE RD	medflex	0.033	0.033	0.000	1.20	0.007	0.007	95.2
	mediation		0.032	-0.000	-0.24	0.007	0.007	95.4
	exact		0.003	-0.000	-1.84	0.002	0.002	93.6
NIE RD	medflex	0.003	0.003	0.000	1.70	0.002	0.002	94.3
	mediation		0.003	0.000	0.72	0.002	0.002	95.3
	exact		0.035	-0.000	-0.57	0.007	0.007	94.8
TE RD	medflex	0.035	0.036	0.000	1.24	0.007	0.007	94.9
	mediation		0.035	-0.000	-0.16	0.007	0.007	95.2
				Scenario 2				
	exact		0.069	-0.000	-0.33	0.007	0.007	94.6
NDE RD	medflex	0.069	0.071	0.002	2.84	0.007	0.008	94.0
	mediation		0.069	-0.000	-0.29	0.007	0.007	95.0
	exact		0.038	-0.000	-0.81	0.005	0.005	94.2
NIE RD	medflex	0.038	0.038	-0.000	-0.19	0.005	0.005	94.2
	mediation		0.038	-0.000	-0.84	0.005	0.005	97.1
	exact		0.107	-0.001	-0.50	0.009	0.009	94.4
TE RD	medflex	0.107	0.109	0.002	1.76	0.009	0.009	94.7
	mediation		0.107	-0.001	-0.48	0.009	0.009	95.2
				Scenario 3				
	exact		-0.035	-0.000	1.42	0.009	0.009	94.9
NDE RD	medflex	-0.034	-0.033	0.001	-3.22	0.010	0.010	94.2
	mediation		-0.035	-0.000	1.19	0.009	0.009	95.4
	exact		0.040	-0.000	-0.78	0.005	0.005	94.3
NIE RD	medflex	0.040	0.040	-0.000	-0.37	0.005	0.005	94.6
	mediation		0.040	-0.000	-0.82	0.005	0.005	97.3
	exact		0.005	-0.001	-13.10	0.011	0.011	94.2
TE RD	medflex	0.006	0.007	0.001	15.59	0.011	0.011	94.0
	mediation		0.005	-0.001	-12.09	0.011	0.011	94.8

Table 4.8 Natural-effects estimators on the risk difference scale in scenarios with increasing levels of outcome commonness (simulation study with covariates based on 1000 independent samples of size $n = 5000$; continuation)

Effect	Estimator[a]	True value	Mean	Bias	Relative bias (%)	SD	RMSE	CP (%)[b]
				Scenario 4				
NDE RD	exact	0.097	0.096	-0.001	-1.06	0.014	0.014	95.1
	medflex		0.095	-0.002	-2.15	0.014	0.014	94.9
	mediation		0.096	-0.001	-1.08	0.014	0.014	94.9
NIE RD	exact	0.041	0.04	0.000	-0.51	0.005	0.005	94.8
	medflex		0.040	-0.001	-2.56	0.005	0.005	93.6
	mediation		0.040	-0.000	-0.71	0.005	0.005	97.0
TE RD	exact	0.138	0.137	-0.001	-0.90	0.015	0.015	94.7
	medflex		0.135	-0.003	-2.27	0.014	0.015	94.0
	mediation		0.137	-0.001	-0.98	0.015	0.015	94.7

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] *exact*: exact estimator proposed; *medflex*: natural effect model approach (Lange *et al.*, 2012) using weighting method implemented in the R package `medflex` (Steen *et al.*, 2017); *mediation*: quasi-Bayesian approach by Imai *et al.* (2010) implemented in the R package `mediation` (Tingley *et al.*, 2014).

[b] Delta method for exact and approximate estimators, robust standard error for *medflex*; for *mediation*, the 95% confidence intervals were based on the White's heteroskedasticity-consistent estimator for the covariance matrix (Tingley *et al.*, 2014).

4.4 Real-data example

We used cohort data presented in the paper by Samoilenko *et al.* (2018) to illustrate our exact mediation approach. Briefly, the data consisted of 6197 singleton pregnancies in asthmatic women who gave birth in the province of Quebec, Canada, between 1998 and 2008. Low birth weight and prematurity (preterm birth) were selected as the outcome and mediator, respectively, and two exposure variables were examined separately: 1) treatment with inhaled corticosteroids during pregnancy and 2) placental abruption. These data correspond to a scenario in which the outcome (low birth weight) is rare marginally, but not rare in some strata of mediator (preterm birth) and exposure.

We used our SAS macro `mediation_estimates` (see Appendices 2 and 3, Sections 4.7-4.8) to obtain exact NDE and NIE estimates on the OR, RR and RD scales for each exposure variable. Mediation analyses adjusted for *maternal age at the beginning of pregnancy* (<18 years, 18-34 years, or >34 years), *baby's sex*, *diabetes mellitus*, and *gestational diabetes*. We also applied the SAS CAUSALMED procedure to obtain natural effects on the multiplicative scales, implementing the approximate approach defined in Equations (4.8 -4.10) for the OR scale. Mediation effects on the OR and RR scales were also estimated using the NEM approach, as described in the simulation studies, and on the

RD scale using the quasi-Bayesian approach. For all approaches, exposure-mediator interaction was considered, and mediation effects were estimated at the sample-specific mean values of the covariates. However, since our SAS macro `mediation_estimates` allows for the estimation of conditional natural effects at user-specified values of the adjustment covariates (by default at the mean values of the covariates), we also obtained natural effects for placental abruption at more meaningful levels of the categorical covariates for purpose of illustration. More details on the real-data analyses are presented in Appendix 1 (Section 4.6).

The main results are presented in Table 4.9 and Figure 4.1. The exact and approximate OR estimates generally did not agree, the only exception being the NIE in the mediation analysis with inhaled corticosteroids as the exposure variable. For placental abruption, the observed discrepancies were quite remarkable. The RR point estimates computed by our SAS macro were close to those computed by PROC CAUSALMED with a log-binomial or Poisson outcome regression model. However, abnormally wide bootstrap 95% confidence intervals for RR^{NDE} and RR^{TE} were returned by PROC CAUSALMED for inhaled corticosteroids exposure.

For both exposures, the natural-effects OR and RR point and interval estimates obtained by our exact approach were similar to those obtained by the NEM approach. Exact point and interval estimates for the NDE and NIE on the RD scale were found close to corresponding effect estimates obtained using the quasi-Bayesian approach.

The exact TE point estimates were found to be close to the conventional TE estimates for both exposures and scales. However, the TE decomposition property was markedly not satisfied for the approximate OR estimates returned by PROC CAUSALMED; for example, the approximate TE was $2.24 \times 3.03 = 6.79$ for placental abruption, while the conventional TE was 5.13.

Finally, Figure 4.2 showcases our SAS macro by presenting natural effects on the OR and RD scales for placental abruption evaluated at two different levels of fetal sex, maternal age, and diabetes status.

The data that support the findings of this section are not publicly available because of privacy and ethical restrictions.

Table 4.9 Comparison between natural direct and indirect effect estimates on the odds ratio, risk ratio, and risk difference scales obtained from the exact estimator and existing estimators available in various software packages (real-data example)

Effect scale	Exact estimate[a]	Delta 95% CI	Bootstrap 95% CI[b]	Estimate by SAS PROC CAUSALMED[c]	Delta 95% CI	Bootstrap 95% CI[b]	Estimate by medflex[d]/ mediation[e] R package	95% CI[f]	Conventional TE	Bootstrap 95% CI[b]
Exposure: treatment with inhaled corticosteroids										
NDE OR	1.00	0.85, 1.16	0.85, 1.17	0.84	0.60, 1.07	0.63, 1.14	1.00	0.86, 1.17	-	-
NIE OR	0.95	0.86, 1.05	0.87, 1.05	0.94	0.83, 1.05	0.83, 1.07	0.95	0.86, 1.05	-	-
TE OR	0.94	0.78, 1.14	0.77, 1.15	0.79	0.54, 1.03	0.58, 1.07	0.95	0.79, 1.16	0.95	0.79, 1.16
NDE RR	1.00	0.86, 1.15	0.86, 1.16	0.98	0.84, 1.11	0.49, 217	1.00	0.87, 1.16	-	-
NIE RR	0.95	0.87, 1.05	0.87, 1.04	0.95	0.87, 1.04	0.84, 1.05	0.95	0.87, 1.05	-	-
TE RR	0.95	0.79, 1.13	0.79, 1.13	0.93	0.77, 1.09	0.45, 207	0.96	0.80, 1.15	0.96	0.80, 1.15
NDE RD	-0.00	-0.01, 0.01	-0.01, 0.01	NA	NA	NA	-0.00	-0.01, 0.01	-	-
NIE RD	-0.00	-0.01, 0.00	-0.01, 0.00	NA	NA	NA	-0.00	-0.01, 0.00	-	-
TE RD	-0.00	-0.02, 0.01	-0.02, 0.01	NA	NA	NA	-0.00	-0.02, 0.01	-	-
Exposure: placental abruption										
NDE OR	1.88	1.30, 2.72	1.23, 2.63	2.24	1.25, 3.24	1.44, 3.70	1.90	1.26, 2.67	-	-
NIE OR	2.70	1.99, 3.66	2.02, 3.86	3.03	2.29, 3.76	2.37, 3.81	2.70	2.03, 3.91	-	-
TE OR	5.07	3.70, 6.96	3.51, 6.90	6.79	3.12, 10.46	4.09, 12.04	5.14	3.66, 7.00	5.13	3.60, 6.92
NDE RR	1.78	1.28, 2.46	1.21, 2.38	1.76	1.12, 2.40	1.18, 2.32	1.78	1.29, 2.46	-	-
NIE RR	2.24	1.73, 2.91	1.76, 3.01	2.20	1.59 2.81	1.73, 2.97	2.21	1.71, 2.85	-	-
TE RR	3.99	3.14, 5.07	2.99, 5.02	3.86	2.66, 5.06	3.02, 4.80	3.94	3.12, 4.98	4.02	3.06, 5.03
NDE RD	0.05	0.01, 0.09	0.01, 0.09	NA	NA	NA	0.05	0.02, 0.10	-	-
NIE RD	0.15	0.10, 0.20	0.10, 0.20	NA	NA	NA	0.15	0.10, 0.20	-	-
TE RD	0.20	0.14, 0.26	0.14, 0.26	NA	NA	NA	0.20	0.14, 0.26	-	-

Abbreviations: CI, confidence interval; NA, not available; NDE, natural direct effect; NIE, natural indirect effect; OR odds ratio; RD, risk difference; RR, risk ratio; TE, total effect.

[a] Estimate returned by the SAS macro `mediation_estimates` (see Appendix 3, Section 4.8).

[b] Percentile bootstrap based on 1000 resamples with replacement.

[c] SAS procedure based on the approximate estimator by Valeri and VanderWeele (2013).

[d] Natural effect model approach (Lange *et al.*, 2012) using the weighting method implemented in the R package `medflex` (Steen *et al.*, 2017).

[e] Quasi-Bayesian approach by Imai *et al.* (2010) implemented in the R package `mediation` (Tingley *et al.*, 2014).

[f] See Appendix 1 (Section 4.6)) for details.

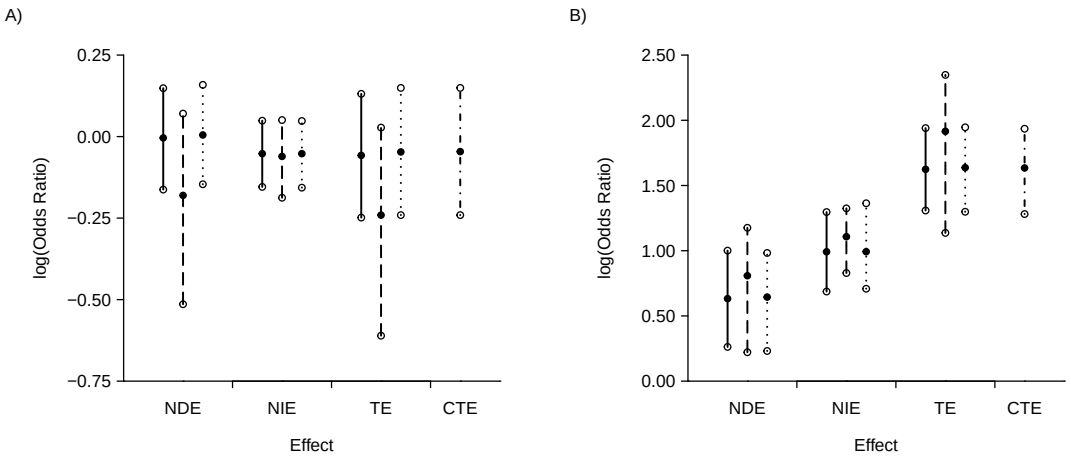


Figure 4.1 Comparison between natural direct effect (NDE), natural indirect effect (NIE), and total effect (TE) estimates on the odds ratio scale obtained from the exact estimator and existing estimators available in software (real-data example). A) Mediation analyses with use of inhaled corticosteroids as the exposure variable; B) mediation analyses with placental abruption as the exposure variable. The solid lines present 95% confidence intervals (CIs) obtained by the exact approach using the delta method. The dashed and dotted lines correspond to 95% CIs returned by the SAS PROC CAUSALMED procedure (via the delta method) and the R package `medflex` (via percentile bootstrap), respectively. The dotted-dashed line presents 95% CIs for the conventional (nonmediated) TE (CTE) by percentile bootstrap. The black circles show effect point estimates, and the white circles show the CI endpoints.

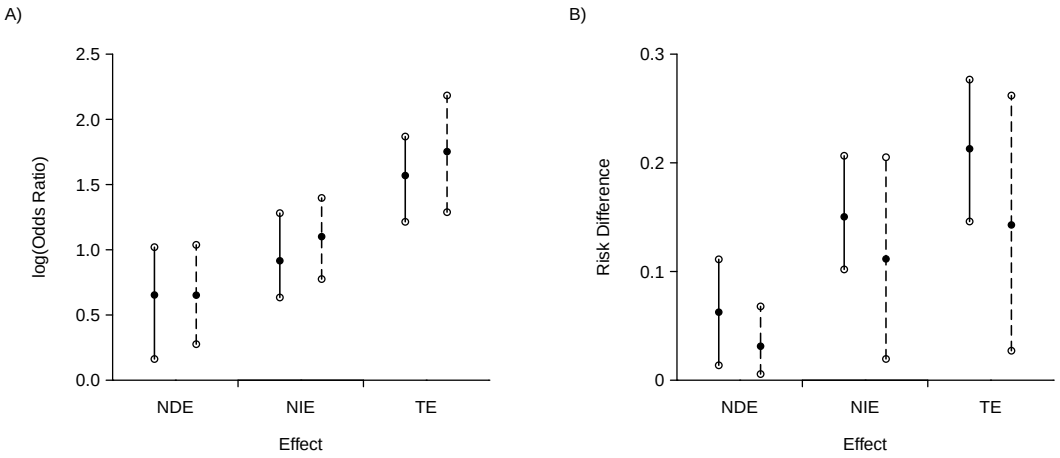


Figure 4.2 Exact natural direct effect (NDE), natural indirect effect (NIE), and total effect (TE) on the odds ratio (A) and risk difference (B) scales evaluated at particular levels of the adjustment covariates (real-data example with placental abruption as the exposure variable). Solid lines correspond to 95% confidence intervals (CIs) given the following set of covariate values: baby's sex = female, maternal age = 18–34 years, diabetes mellitus = no, and gestational diabetes = no. Dashed lines correspond to 95% CIs when the covariate values are specified as follows: baby's sex = male, maternal age <18 years, diabetes mellitus = no, and gestational diabetes = yes. The 95% CIs were constructed by percentile bootstrapping based on 1000 resamples with replacement. The black circles show effect point estimates, and the white circles show the CI endpoints.

4.5 Discussion

In this article, we introduced exact binary-binary regression-based estimators of the natural direct and indirect effects for the three most commonly used scales in epidemiology, namely the OR, RR and RD scales. Our work, which is based on the specification of a logistic outcome model, thus extends previous works that have proposed an exact binary-binary natural-effects estimator on the OR scale. Our exact estimators were observed to be virtually unbiased, regardless of the effect scale and the rareness or commonness of the outcome. Corresponding standard error formulas were derived for each scale using the first-order delta method, thereby providing an alternative approach for computing confidence intervals (in addition to the bootstrap). In our simulations, for which the sample size was relatively large, both the delta method and the bootstrap yielded coverage probabilities close to the nominal value. Unlike other mediation approaches implemented in the simulations and real-data analyses, our exact approach was observed to be numerically stable no matter the effect scale on which results were obtained.

Our investigations have produced additional evidence regarding the performance of the approximate natural-effects OR estimator proposed by Valeri and VanderWeele (2013) for binary mediators and outcomes. As expected, this estimator was found to behave adequately in the scenario where the outcome was rare in all strata defined by the mediator and the exposure (*Scenario* 1), while the exact estimator performed comparably or better. In other scenarios investigated (*Scenarios* 2-4), in which the outcome was either rare or common marginally but not rare conditionally, the bias and variance of the approximate estimator were found to be systematically larger than those of proposed exact estimator under both multiplicative scales, with large biases and poor coverage probabilities sometimes exhibited.

Our proposed exact approach can be implemented using the SAS macro accompanying Appendix 3 (Section 4.8). By default, the exact NDE and NIE are estimated at the sample-specific mean values of the adjustment covariates, but our macro also handles user-specified levels for the entire set of covariates or for some proper subset (in the latter case, our macro sets the other covariates to the sample mean values). Another functionality of our macro is that it allows for Firth penalization by calling the *Firth* option in PROC LOGISTIC. Firth penalization is a general method designed to reduce bias of the maximum likelihood parameter estimator (Firth, 1993). This penalization has been shown to be effective in dealing with separation problems in logistic regression models in presence of scarce or sparse data (Allison, 2012; Heinze et Schemper, 2002; Mansournia *et al.*, 2018).

Although the NDE and NIE are popular estimands in the applied literature, the controlled direct effect can also be of interest to practitioners (Imai *et al.*, 2013; VanderWeele, 2011). Valeri and VanderWeele (2013) provided an expression for the controlled direct effect on the OR scale derived from logistic regression models for the mediator and outcome. This expression is not obtained by invoking the ROA and thus is exact by construction. For completeness, our macro also returns the controlled direct effect on all scales considered (see Appendix 1, Section 4.6, for our extension to the RR and RD scales).

In conclusion, our exact estimator is indicated for those wanting to perform a conventional binary-binary regression-based mediation analysis on the effect scale of their choice without worrying about the rareness or commonness of the outcome. By using the same two fitted logistic models for all effect scales (OR, RR, and RD), our exact approach also simplifies applications and increases compatibility of mediation analyses results with binary mediators and outcomes. One limitation of our exact estimator is that it is currently only applicable to data from cohort studies; thus, more developments will thus be required to extend proposed approach to accommodate data from case-control study designs in which cases are overrepresented compared with controls. Moreover, since our work has thus far focused on the case of a single mediator, it will also be worthwhile to study the case of multiple mediators and expand our SAS macro further.

Acknowledgements

This work was funded by grants from the Fonds de recherche Québec-Santé (FRQ-S) and the Natural Sciences and Engineering Research Council of Canada. Geneviève Lefebvre is a FRQ-S Research Scholar.

Conflict of interest: none declared.

4.6 Appendix 1

4.6.1 General formulas used in the delta method for the exact regression-based mediation effects

The formulas for calculating the confidence intervals of the natural effects on the odd ratio (OR), risk ratio (RR), and risk difference (RD) scales depend upon the nested counterfactual outcome probability $P\left(Y\left(a, M\left(a^{*}\right)\right)=1 | \boldsymbol{C}=\boldsymbol{c}\right)$ and its gradient with respect to the coefficients of the mediator and outcome models ($\boldsymbol{\beta}$, $\boldsymbol{\theta}$). The following equations will be used to derive the expression for the aforementioned gradient as well as those for the standard errors of estimates.

For any scalar function $\varphi(\boldsymbol{\alpha})$ differentiable in $\mathbb{R}^n$, where $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, \ldots, \alpha_n)'$, the partial derivatives of $\text{expit}(\varphi(\alpha_1))$ and $1 - \text{expit}(\varphi(\alpha_1))$ are:

$$\begin{aligned}
\frac{\partial}{\partial \alpha_i}\left(\frac{\exp(\varphi(\boldsymbol{\alpha}))}{1+\exp(\varphi(\boldsymbol{\alpha}))}\right) &= \frac{\exp(\varphi(\boldsymbol{\alpha}))}{(1+\exp(\varphi(\boldsymbol{\alpha})))^2}\cdot\frac{\partial\varphi(\boldsymbol{\alpha})}{\partial\alpha_i}, \\
\frac{\partial}{\partial \alpha_i}\left(\frac{1}{1+\exp(\varphi(\boldsymbol{\alpha}))}\right) &= -\frac{\exp(\varphi(\boldsymbol{\alpha}))}{(1+\exp(\varphi(\boldsymbol{\alpha})))^2}\cdot\frac{\partial\varphi(\boldsymbol{\alpha})}{\partial\alpha_i},
\end{aligned} \tag{4.12}$$

where $i = 1, 2, \ldots, n$.

For any scalar functions $f(\boldsymbol{\alpha})$ and $g(\boldsymbol{\alpha})$, differentiable in $\mathbb{R}^n$ and taking values in $(0, 1)$:

$$\begin{aligned}
\frac{\partial}{\partial \alpha_i}\left(\ln\left(\frac{f(\boldsymbol{\alpha})}{1-f(\boldsymbol{\alpha})}\right)\right) &= \frac{\partial}{\partial\alpha_i}\left(\ln(f(\boldsymbol{\alpha})) - \ln(1-f(\boldsymbol{\alpha}))\right) \\
&= \frac{1}{f(\boldsymbol{\alpha})}\cdot\frac{\partial f(\boldsymbol{\alpha})}{\partial\alpha_i} + \frac{1}{1-f(\boldsymbol{\alpha})}\cdot\frac{\partial f(\boldsymbol{\alpha})}{\partial\alpha_i} \\
&= \frac{1}{f(\boldsymbol{\alpha})(1-f(\boldsymbol{\alpha}))}\cdot\frac{\partial f(\boldsymbol{\alpha})}{\partial\alpha_i},
\end{aligned}$$

$$\begin{aligned}
\frac{\partial}{\partial \alpha_i}\left(\ln\left(\frac{\frac{f(\boldsymbol{\alpha})}{1-f(\boldsymbol{\alpha})}}{\frac{g(\boldsymbol{\alpha})}{1-g(\boldsymbol{\alpha})}}\right)\right) &= \frac{\partial}{\partial\alpha_i}\left(\ln\left(\frac{f(\boldsymbol{\alpha})}{1-f(\boldsymbol{\alpha})}\right) - \ln\left(\frac{g(\boldsymbol{\alpha})}{1-g(\boldsymbol{\alpha})}\right)\right) \\
&= \frac{1}{f(\boldsymbol{\alpha})(1-f(\boldsymbol{\alpha}))}\cdot\frac{\partial f(\boldsymbol{\alpha})}{\partial\alpha_i} - \frac{1}{g(\boldsymbol{\alpha})(1-g(\boldsymbol{\alpha}))}\cdot\frac{\partial g(\boldsymbol{\alpha})}{\partial\alpha_i},
\end{aligned} \tag{4.13}$$

$$\frac{\partial}{\partial \alpha_i}\left(\ln\left(\frac{f(\boldsymbol{\alpha})}{g(\boldsymbol{\alpha})}\right)\right) = \frac{\partial}{\partial\alpha_i}\left(\ln f(\boldsymbol{\alpha}) - \ln g(\boldsymbol{\alpha})\right) = \frac{1}{f(\boldsymbol{\alpha})}\cdot\frac{\partial f(\boldsymbol{\alpha})}{\partial\alpha_i} - \frac{1}{g(\boldsymbol{\alpha})}\cdot\frac{\partial g(\boldsymbol{\alpha})}{\partial\alpha_i}, \tag{4.14}$$

where $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, \ldots, \alpha_n)' \in \mathbb{R}^n$, $i = 1, 2, \ldots, n$.

4.6.2 Nested probability formulas based on the mediator and outcome logistic models

Let $\boldsymbol{C}$ denote the set of adjustment variables. Under identifying and modeling assumptions (4.1 - 4.2) from the main text, the nested probability $P(Y(a, M(a^*)) = 1 | \boldsymbol{C} = \boldsymbol{c})$ for exposure levels a and a^*

is expressed as follows:

$$
\begin{aligned}
&P\left(Y\left(a, M\left(a^*\right)\right)=1 | \boldsymbol{C}=\boldsymbol{c}\right) \\
&\quad=P\left(Y=1 | A=a, M=1, \boldsymbol{C}=\boldsymbol{c}\right) P\left(M=1 | A=a^*, \boldsymbol{C}=\boldsymbol{c}\right) \\
&\qquad+P\left(Y=1 | A=a, M=0, \boldsymbol{C}=\boldsymbol{c}\right) P\left(M=0 | A=a^*, \boldsymbol{C}=\boldsymbol{c}\right) \\
&\quad=\frac{\exp\left(\theta_0+\theta_1 a+\theta_2+\theta_3 a+\boldsymbol{\theta}_4^{'} \boldsymbol{c}\right)}{1+\exp\left(\theta_0+\theta_1 a+\theta_2+\theta_3 a+\boldsymbol{\theta}_4^{'} \boldsymbol{c}_2\right)} \cdot \frac{\exp\left(\beta_0+\beta_1 a^*+\boldsymbol{\beta}_2^{'} \boldsymbol{c}\right)}{1+\exp\left(\beta_0+\beta_1 a^*+\boldsymbol{\beta}_2^{'} \boldsymbol{c}\right)} \\
&\qquad+\frac{\exp\left(\theta_0+\theta_1 a+\boldsymbol{\theta}_4^{'} \boldsymbol{c}\right)}{1+\exp\left(\theta_0+\theta_1 a+\boldsymbol{\theta}_4^{'} \boldsymbol{c}\right)} \cdot \frac{1}{1+\exp\left(\beta_0+\beta_1 a^*+\boldsymbol{\beta}_2^{'} \boldsymbol{c}\right)}.
\end{aligned}
$$

Let $g\left(a, a^*, \boldsymbol{c}\right)$ denote this last expression:

$$
\begin{aligned}
g\left(a, a^*, \boldsymbol{c}\right)=&\frac{\exp\left(\theta_0+\theta_1 a+\theta_2+\theta_3 a+\boldsymbol{\theta}_4^{'} \boldsymbol{c}\right)}{1+\exp\left(\theta_0+\theta_1 a+\theta_2+\theta_3 a+\boldsymbol{\theta}_4^{'} \boldsymbol{c}_2\right)} \cdot \frac{\exp\left(\beta_0+\beta_1 a^*+\boldsymbol{\beta}_2^{'} \boldsymbol{c}\right)}{1+\exp\left(\beta_0+\beta_1 a^*+\boldsymbol{\beta}_2^{'} \boldsymbol{c}\right)} \\
&+\frac{\exp\left(\theta_0+\theta_1 a+\boldsymbol{\theta}_4^{'} \boldsymbol{c}\right)}{1+\exp\left(\theta_0+\theta_1 a+\boldsymbol{\theta}_4^{'} \boldsymbol{c}\right)} \cdot \frac{1}{1+\exp\left(\beta_0+\beta_1 a^*+\boldsymbol{\beta}_2^{'} \boldsymbol{c}\right)}.
\end{aligned}
\tag{4.15}
$$

Then, using general formulas (4.12), the partial derivatives of $g\left(a, a^*, \boldsymbol{c}\right)$ with respect to each of the coefficients are:

$$
\begin{aligned}
\frac{\partial g\left(a, a^*, \boldsymbol{c}\right)}{\partial \beta_0}=&\frac{\exp(\theta_0+\theta_1 a+\theta_2+\theta_3 a+\boldsymbol{\theta}_4^{'} \boldsymbol{c})}{1+\exp(\theta_0+\theta_1 a+\theta_2+\theta_3 a+\boldsymbol{\theta}_4^{'} \boldsymbol{c})} \cdot \frac{\exp(\beta_0+\beta_1 a^*+\boldsymbol{\beta}_2^{'} \boldsymbol{c})}{\left(1+\exp(\beta_0+\beta_1 a^*+\boldsymbol{\beta}_2^{'} \boldsymbol{c})\right)^2} \\
&-\frac{\exp(\theta_0+\theta_1 a+\boldsymbol{\theta}_4^{'} \boldsymbol{c})}{1+\exp(\theta_0+\theta_1 a+\boldsymbol{\theta}_4^{'} \boldsymbol{c})} \cdot \frac{\exp(\beta_0+\beta_1 a^*+\boldsymbol{\beta}_2^{'} \boldsymbol{c})}{\left(1+\exp(\beta_0+\beta_1 a^*+\boldsymbol{\beta}_2^{'} \boldsymbol{c})\right)^2} \\
=&\frac{\exp\left(\beta_0+\beta_1 a^*+\boldsymbol{\beta}_2^{'} \boldsymbol{c}\right)}{\left(1+\exp\left(\beta_0+\beta_1 a^*+\boldsymbol{\beta}_2^{'} \boldsymbol{c}\right)\right)^2} \\
&\times\left(\frac{\exp(\theta_0+\theta_1 a+\theta_2+\theta_3 a+\boldsymbol{\theta}_4^{'} \boldsymbol{c})}{1+\exp(\theta_0+\theta_1 a+\theta_2+\theta_3 a+\boldsymbol{\theta}_4^{'} \boldsymbol{c})}-\frac{\exp(\theta_0+\theta_1 a+\boldsymbol{\theta}_4^{'} \boldsymbol{c})}{1+\exp(\theta_0+\theta_1 a+\boldsymbol{\theta}_4^{'} \boldsymbol{c})}\right),
\end{aligned}
$$

$$
\frac{\partial g\left(a, a^*, \boldsymbol{c}\right)}{\partial \beta_1}=a^* \cdot \frac{\partial g\left(a, a^*, \boldsymbol{c}\right)}{\partial \beta_0},
$$

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

$$\frac{\partial g\left(a,a^{*},\boldsymbol{c}\right)}{\partial\theta_{0}}=\frac{\exp(\theta_{0}+\theta_{1}a+\theta_{2}+\theta_{3}a+\boldsymbol{\theta}_{4}^{'}\boldsymbol{c})}{\left(1+\exp(\theta_{0}+\theta_{1}a+\theta_{2}+\theta_{3}a+\boldsymbol{\theta}_{4}^{'}\boldsymbol{c})\right)^{2}}\cdot\frac{\exp(\beta_{0}+\beta_{1}a^{*}+\boldsymbol{\beta}_{2}^{'}\boldsymbol{c})}{1+exp(\beta_{0}+\beta_{1}a^{*}+\boldsymbol{\beta}_{2}^{'}\boldsymbol{c})}$$
$$+\frac{\exp(\theta_{0}+\theta_{1}a+\boldsymbol{\theta}_{4}^{'}\boldsymbol{c})}{\left(1+\exp(\theta_{0}+\theta_{1}a+\boldsymbol{\theta}_{4}^{'}\boldsymbol{c})\right)^{2}}\cdot\frac{1}{1+\exp\left(\beta_{0}+\beta_{1}a^{*}+\boldsymbol{\beta}_{2}^{'}\boldsymbol{c}\right)},$$

$$\frac{\partial g\left(a,a^{*},\boldsymbol{c}\right)}{\partial\theta_{1}}=a\cdot\frac{\partial g\left(a,a^{*},\boldsymbol{c}\right)}{\partial\theta_{0}},$$

$$\frac{\partial g\left(a,a^{*},\boldsymbol{c}\right)}{\partial\theta_{2}}=\frac{\exp(\theta_{0}+\theta_{1}a+\theta_{2}+\theta_{3}a+\boldsymbol{\theta}_{4}^{'}\boldsymbol{c})}{\left(1+\exp(\theta_{0}+\theta_{1}a+\theta_{2}+\theta_{3}a+\boldsymbol{\theta}_{4}^{'}\boldsymbol{c})\right)^{2}}\cdot\frac{\exp\left(\beta_{0}+\beta_{1}a^{*}+\boldsymbol{\beta}_{2}^{'}\boldsymbol{c}\right)}{1+\exp\left(\beta_{0}+\beta_{1}a^{*}+\boldsymbol{\beta}_{2}^{'}\boldsymbol{c}\right)},$$

$$\frac{\partial g\left(a,a^{*},\boldsymbol{c}\right)}{\partial\theta_{3}}=a\cdot\frac{\partial g\left(a,a^{*},\boldsymbol{c}\right)}{\partial\theta_{2}},$$

$$\frac{\partial g\left(a,a^{*},\boldsymbol{c}\right)}{\partial\theta_{4i}}=c_{i}\cdot\frac{\partial g\left(a,a^{*},\boldsymbol{c}\right)}{\partial\theta_{0}},\quad i=1,\ldots,k,$$

where k is the cardinal number of $\boldsymbol{C}$.

Thus, the gradient of the scalar function $g\left(a,a^{*},\boldsymbol{c}\right)$ with respect to the vector

$$(\boldsymbol{\beta},\boldsymbol{\theta})=(\beta_{0},\beta_{1},\beta_{21},\ldots,\beta_{2k},\theta_{0},\theta_{1},\ \theta_{2},\theta_{3},\theta_{41},\ldots,\theta_{4k})^{'}$$

is

$$\nabla\left(g\left(a,a^{*},\boldsymbol{c}\right)\right)=\left(\frac{\partial g\left(a,a^{*},\boldsymbol{c}\right)}{\partial\boldsymbol{\beta}},\frac{\partial g\left(a,a^{*},\boldsymbol{c}\right)}{\partial\boldsymbol{\theta}}\right)^{'}\tag{4.16}$$

where

$$\frac{\partial g\left(a,a^{*},\boldsymbol{c}\right)}{\partial\boldsymbol{\beta}}=\left(\frac{\partial g\left(a,a^{*},\boldsymbol{c}\right)}{\partial\beta_{0}},\frac{\partial g\left(a,a^{*},\boldsymbol{c}\right)}{\beta_{1}},\frac{\partial g\left(a,a^{*},\boldsymbol{c}\right)}{\partial\beta_{21}},\ldots,\frac{\partial g\left(a,a^{*},\boldsymbol{c}\right)}{\partial\beta_{2k}}\right),$$
$$\frac{\partial g\left(a,a^{*},\boldsymbol{c}\right)}{\partial\boldsymbol{\theta}}=\left(\frac{\partial g\left(a,a^{*},\boldsymbol{c}\right)}{\partial\theta_{0}},\frac{\partial g\left(a,a^{*},\boldsymbol{c}\right)}{\partial\theta_{1}},\frac{\partial g\left(a,a^{*},\boldsymbol{c}\right)}{\partial\theta_{2}},\frac{\partial g\left(a,a^{*},\boldsymbol{c}\right)}{\partial\theta_{3}},\frac{\partial g\left(a,a^{*},\boldsymbol{c}\right)}{\partial\theta_{41}},\ldots,\frac{\partial g\left(a,a^{*},\boldsymbol{c}\right)}{\partial\theta_{4k}}\right).$$

4.6.3 Delta method for mediation exact regression-based odds ratios

We define the following notation:

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

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

where $\nabla\left(g\left(a,a^{*},\boldsymbol{c}\right)\right)$ is defined in Equation (4.16).

We can express the regression-based exact $OR^{NDE}_{a,a^{*}|\boldsymbol{c}}$ and $OR^{NIE}_{a,a^{*}|\boldsymbol{c}}$ in terms of $g\left(a,a^{*},\boldsymbol{c}\right)$ defined in Equation (4.15). as follows:

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

To construct the 95% confidence intervals (CIs) for the exact natural effect OR by first-order multivariate delta method (Xu et Long, 2005), we have applied the same approach as in VanderWeele and Vansteelandt (2010); that is we have expressed standard errors (se) for $\ln\left(\widehat{OR^{NDE}_{a,a^{*}|\boldsymbol{c}}}\right)$ and $\ln\left(\widehat{OR^{NIE}_{a,a^{*}|\boldsymbol{c}}}\right)$ according to the following approximate formulas:

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

where $\Sigma=diag\left\{\Sigma_{\widehat{\boldsymbol{\beta}}},\Sigma_{\widehat{\boldsymbol{\theta}}}\right\}$ is a block matrix, $\Sigma_{\widehat{\boldsymbol{\beta}}}$ and $\Sigma_{\widehat{\boldsymbol{\theta}}}$ are the covariance matrices for the vectors $\widehat{\boldsymbol{\beta}}$ and $\widehat{\boldsymbol{\theta}}$, respectively. The gradients of $\ln\left(\widehat{OR^{NDE}_{a,a^{*}|\boldsymbol{c}}}\right)$ and $\ln\left(\widehat{OR^{NIE}_{a,a^{*}|\boldsymbol{c}}}\right)$ are expressed using $\nabla\left(\widehat{g}\left(a,a^{*},\boldsymbol{c}\right)\right)$ as follows:

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

Thus,

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

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

mate 95% CIs for $OR^{NDE}_{a,a^*|\boldsymbol{c}}$ and $OR^{NIE}_{a,a^*|\boldsymbol{c}}$ are given by

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

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

$$\ln\left(\widehat{OR^{TE}_{a,a^*|\boldsymbol{c}}}\right) = \ln\left(\widehat{OR^{NDE}_{a,a^*|\boldsymbol{c}}}\right) + \ln\left(\widehat{OR^{NIE}_{a,a^*|\boldsymbol{c}}}\right),$$
$$\nabla\left(\ln\left(\widehat{OR^{TE}_{a,a^*|\boldsymbol{c}}}\right)\right) = \nabla\left(\ln\left(\widehat{OR^{NDE}_{a,a^*|\boldsymbol{c}}}\right)\right) + \nabla\left(\ln\left(\widehat{OR^{NIE}_{a,a^*|\boldsymbol{c}}}\right)\right)$$

and

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

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

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

and, consequently, a approximate 95% CI for $OR^{TE}_{a,a^*|\boldsymbol{c}}$ is given by

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

4.6.4 Delta method for exact mediation regression-based risk ratios

We have for the RR scale that

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

and, correspondingly,

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

The Equation (4.14) implies that

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

Thus,

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

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

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

Finally, we have for the total effect $RR^{TE}_{a,a^*|\boldsymbol{c}} = RR^{NDE}_{a,a^*|\boldsymbol{c}} \cdot RR^{NIE}_{a,a^*|\boldsymbol{c}}$:

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

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

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

4.6.5 Delta method for exact mediation regression-based risk differences

We have for the RD scale:

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

$$RD^{TE}_{a,a^*|\boldsymbol{c}} = RD^{NDE}_{a,a^*|\boldsymbol{c}} + RD^{NIE}_{a,a^*|\boldsymbol{c}},$$

$$\begin{aligned}
\nabla\left(\widehat{RD^{NDE}_{a,a^*|\boldsymbol{c}}}\right) &= \nabla\left(\widehat{g}\left(a, a^*, \boldsymbol{c}\right)\right) - \nabla\left(\widehat{g}\left(a^*, a^*, \boldsymbol{c}\right)\right), \\
\nabla\left(\widehat{RD^{NIE}_{a,a^*|\boldsymbol{c}}}\right) &= \nabla\left(\widehat{g}\left(a, a, \boldsymbol{c}\right)\right) - \nabla\left(\widehat{g}\left(a, a^*, \boldsymbol{c}\right)\right), \\
\nabla\left(\widehat{RD^{TE}_{a,a^*|\boldsymbol{c}}}\right) &= \nabla\left(\widehat{RD^{NDE}_{a,a^*|\boldsymbol{c}}}\right) + \nabla\left(\widehat{RD^{NIE}_{a,a^*|\boldsymbol{c}}}\right), \\
se\left(\widehat{RD^{NDE}_{a,a^*|\boldsymbol{c}}}\right) &\approx \sqrt{\nabla\left(\widehat{RD^{NDE}_{a,a^*|\boldsymbol{c}}}\right)' \cdot \widehat{\Sigma} \cdot \nabla\left(\widehat{RD^{NDE}_{a,a^*|\boldsymbol{c}}}\right)}, \\
se\left(\widehat{RD^{NIE}_{a,a^*|\boldsymbol{c}}}\right) &\approx \sqrt{\nabla\left(\widehat{RD^{NIE}_{a,a^*|\boldsymbol{c}}}\right)' \cdot \widehat{\Sigma} \cdot \nabla\left(\widehat{RD^{NIE}_{a,a^*|\boldsymbol{c}}}\right)}, \\
se\left(\widehat{RD^{TE}_{a,a^*|\boldsymbol{c}}}\right) &\approx \sqrt{\nabla\left(\widehat{RD^{TE}_{a,a^*|\boldsymbol{c}}}\right)' \cdot \widehat{\Sigma} \cdot \nabla\left(\widehat{RD^{TE}_{a,a^*|\boldsymbol{c}}}\right)}.
\end{aligned}$$

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

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

respectively.

4.6.6 Mediation controlled direct effects

Controlled direct effects reflect causal effect of the direct manipulation on the mediator and can be useful in policy evaluation (Imai *et al.*, 2013; Valeri et VanderWeele, 2013). We let $Y(a, m)$ be the potential outcome that would have been observed under the exposure and mediator levels $A = a$ and $M = m$, respectively. Under outcome model (Equation (4.2)) and independently of the mediator model, the corresponding counterfactual probability is expressed as

$$P\left(Y(a, m) = 1 | \boldsymbol{C} = \boldsymbol{c}\right) = \text{expit}\left(\theta_0 + \theta_1 a + \theta_2 m + \theta_3 am + \boldsymbol{\theta}_4' \boldsymbol{c}\right).$$

Thus, the logistic regression-based controlled direct effects (CDE) on the OR, RR and RD scales can be expressed as follows:

$$
\begin{aligned}
OR^{CDE}_{a,a^*|\boldsymbol{c}}(m) &= \exp\left(\theta_1\left(a-a^*\right)+\theta_3 m\left(a-a^*\right)\right),\\
RR^{CDE}_{a,a^*|\boldsymbol{c}}(m) &= \exp\left(\theta_1\left(a-a^*\right)+\theta_3 m\left(a-a^*\right)\right)\\
&\quad\times\frac{1+\exp\left(\theta_0+\theta_1 a^*+\theta_2 m+\theta_3 a^* m+\boldsymbol{\theta}_4^{'}\boldsymbol{c}\right)}{1+\exp\left(\theta_0+\theta_1 a+\theta_2 m+\theta_3 a m+\boldsymbol{\theta}_4^{'}\boldsymbol{c}\right)},\\
RD^{CDE}_{a,a^*|\boldsymbol{c}}(m) &= \operatorname{expit}\left(\theta_0+\theta_1 a+\theta_2 m+\theta_3 a m+\boldsymbol{\theta}_4^{'}\boldsymbol{c}\right)\\
&\quad-\operatorname{expit}\left(\theta_0+\theta_1 a^*+\theta_2 m+\theta_3 a^* m+\boldsymbol{\theta}_4^{'}\boldsymbol{c}\right).
\end{aligned}
$$

The derivatives of $\ln\left(RR^{CDE}_{a,a^*|\boldsymbol{c}}(m)\right)$ with respect to each of the coefficients

$$
\boldsymbol{\theta}=(\theta_1,\ \theta_2,\theta_3,\theta_{41},\ldots,\theta_{4k})
$$

of the outcome model are:

$$
\begin{aligned}
\frac{\partial}{\partial\theta_0}\ln\left(RR^{CDE}_{a,a^*|\boldsymbol{c}}(m)\right) &= \frac{\exp\left(\theta_0+\theta_1 a^*+\theta_2 m+\theta_3 a^* m+\boldsymbol{\theta}_4^{'}\boldsymbol{c}\right)}{1+\exp\left(\theta_0+\theta_1 a^*+\theta_2 m+\theta_3 a^* m+\boldsymbol{\theta}_4^{'}\boldsymbol{c}\right)}\\
&\quad-\frac{\exp\left(\theta_0+\theta_1 a+\theta_2 m+\theta_3 a m+\boldsymbol{\theta}_4^{'}\boldsymbol{c}\right)}{1+\exp\left(\theta_0+\theta_1 a+\theta_2 m+\theta_3 a m+\boldsymbol{\theta}_4^{'}\boldsymbol{c}\right)},\\
\frac{\partial}{\partial\theta_1}\ln\left(RR^{CDE}_{a,a^*|\boldsymbol{c}}(m)\right) &= \frac{a}{1+\exp\left(\theta_0+\theta_1 a+\theta_2 m+\theta_3 a m+\boldsymbol{\theta}_4^{'}\boldsymbol{c}\right)}\\
&\quad-\frac{a^*}{1+\exp\left(\theta_0+\theta_1 a^*+\theta_2 m+\theta_3 a^* m+\boldsymbol{\theta}_4^{'}\boldsymbol{c}\right)},\\
\frac{\partial}{\partial\theta_2}\ln\left(RR^{CDE}_{a,a^*|\boldsymbol{c}}(m)\right) &= m\cdot\frac{\partial}{\partial\theta_0}\ln\left(RR^{CDE}_{a,a^*|\boldsymbol{c}}(m)\right),\\
\frac{\partial}{\partial\theta_3}\ln\left(RR^{CDE}_{a,a^*|\boldsymbol{c}}(m)\right) &= m\cdot\frac{\partial}{\partial\theta_1}\ln\left(RR^{CDE}_{a,a^*|\boldsymbol{c}}(m)\right),\\
\frac{\partial}{\partial\theta_{4i}}\ln\left(RR^{CDE}_{a,a^*|\boldsymbol{c}}(m)\right) &= c_i\cdot\frac{\partial}{\partial\theta_0}\ln\left(RR^{CDE}_{a,a^*|\boldsymbol{c}}(m)\right),\quad i=1,2,\ldots,k.
\end{aligned}
$$

We define the following notation:

$$
\begin{aligned}
\ln\left(\widehat{RR^{CDE}_{a,a^*|\boldsymbol{c}}}(m)\right) &= \ln\left(RR^{CDE}_{a,a^*|\boldsymbol{c}}(m)\right)\Big|_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}},\\
\nabla\left(\ln\left(\widehat{RR^{CDE}_{a,a^*|\boldsymbol{c}}}(m)\right)\right) &= \nabla\left(\ln\left(RR^{CDE}_{a,a^*|\boldsymbol{c}}(m)\right)\right)\Big|_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}},
\end{aligned}
$$

where $\nabla\left(\ln\left(RR^{CDE}_{a,a^*|\boldsymbol{c}}(m)\right)\right)$ is the gradient of $\ln\left(RR^{CDE}_{a,a^*|\boldsymbol{c}}(m)\right)$ with respect to $\boldsymbol{\theta}$. Then

$$se\left(\ln\left(\widehat{RR^{CDE}_{a,a^*|\boldsymbol{c}}(m)}\right)\right) \approx \sqrt{\nabla\left(\ln\left(\widehat{RR^{CDE}_{a,a^*|\boldsymbol{c}}(m)}\right)\right)' \cdot \widehat{\Sigma}_{\widehat{\boldsymbol{\theta}}} \cdot \nabla\left(\ln\left(\widehat{RR^{CDE}_{a,a^*|\boldsymbol{c}}(m)}\right)\right)},$$

and

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

is the approximate 95% CI for $RR^{CDE}_{a,a^*|\boldsymbol{c}}(m)$ derived by the delta method.

Taking the derivative of $RD^{CDE}_{a,a^*|\boldsymbol{c}}(m)$ with respect to the outcome model coefficients $\boldsymbol{\theta}$, we get

$$\begin{aligned}
\frac{\partial}{\partial\theta_0} RD^{CDE}_{a,a^*|\boldsymbol{c}}(m) &= \frac{\exp\left(\theta_0+\theta_1 a+\theta_2 m+\theta_3 am+\boldsymbol{\theta}_4'\boldsymbol{c}\right)}{\left(1+\exp\left(\theta_0+\theta_1 a+\theta_2 m+\theta_3 am+\boldsymbol{\theta}_4'\boldsymbol{c}\right)\right)^2} \\
&\quad - \frac{\exp\left(\theta_0+\theta_1 a^*+\theta_2 m+\theta_3 a^* m+\boldsymbol{\theta}_4'\boldsymbol{c}\right)}{\left(1+\exp\left(\theta_0+\theta_1 a^*+\theta_2 m+\theta_3 a^* m+\boldsymbol{\theta}_4'\boldsymbol{c}\right)\right)^2}, \\
\frac{\partial}{\partial\theta_1} RD^{CDE}_{a,a^*|\boldsymbol{c}}(m) &= a \cdot \frac{\exp\left(\theta_0+\theta_1 a+\theta_2 m+\theta_3 am+\boldsymbol{\theta}_4'\boldsymbol{c}\right)}{\left(1+\exp\left(\theta_0+\theta_1 a+\theta_2 m+\theta_3 am+\boldsymbol{\theta}_4'\boldsymbol{c}\right)\right)^2} \\
&\quad - a^* \cdot \frac{\exp\left(\theta_0+\theta_1 a^*+\theta_2 m+\theta_3 a^* m+\boldsymbol{\theta}_4'\boldsymbol{c}\right)}{\left(1+\exp\left(\theta_0+\theta_1 a^*+\theta_2 m+\theta_3 a^* m+\boldsymbol{\theta}_4'\boldsymbol{c}\right)\right)^2}, \\
\frac{\partial}{\partial\theta_2} RD^{CDE}_{a,a^*|\boldsymbol{c}}(m) &= m \cdot \frac{\partial}{\partial\theta_0} RD^{CDE}_{a,a^*|\boldsymbol{c}}(m), \\
\frac{\partial}{\partial\theta_3} RD^{CDE}_{a,a^*|\boldsymbol{c}}(m) &= m \cdot \frac{\partial}{\partial\theta_1} RD^{CDE}_{a,a^*|\boldsymbol{c}}(m), \\
\frac{\partial}{\partial\theta_{4i}} RD^{CDE}_{a,a^*|\boldsymbol{c}}(m) &= c_i \cdot \frac{\partial}{\partial\theta_0} RD^{CDE}_{a,a^*|\boldsymbol{c}}(m), \quad i=1,2,\ldots,k.
\end{aligned}$$

Let us define

$$\widehat{RD^{CDE}_{a,a^*|\boldsymbol{c}}(m)} = RD^{CDE}_{a,a^*|\boldsymbol{c}}(m)\Big|_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}}, \quad \nabla\left(\widehat{RD^{CDE}_{a,a^*|\boldsymbol{c}}(m)}\right) = \nabla\left(RD^{CDE}_{a,a^*|\boldsymbol{c}}(m)\right)\Big|_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}}.$$

Thus, by applying the delta method,

$$se\left(\widehat{RD^{CDE}_{a,a^*|\boldsymbol{c}}(m)}\right) \approx \sqrt{\nabla\left(\widehat{RD^{CDE}_{a,a^*|\boldsymbol{c}}(m)}\right)' \cdot \widehat{\Sigma}_{\widehat{\boldsymbol{\theta}}} \cdot \nabla\left(\widehat{RD^{CDE}_{a,a^*|\boldsymbol{c}}(m)}\right)},$$

and

$$\widehat{RD^{CDE}_{a,a^*|\boldsymbol{c}}}(m) \pm \Phi^{-1}(0.975) \cdot se\left(\widehat{RD^{CDE}_{a,a^*|\boldsymbol{c}}}(m)\right)$$

is the approximate 95% CI for $RD^{CDE}_{a,a^*|\boldsymbol{c}}(m)$. See Valeri and VanderWeele (2013) for the delta method for $OR^{CDE}_{a,a^*|\boldsymbol{c}}(m)$.

4.6.7 Decomposition property of the exact total effect estimator

For binary outcomes, the total causal effect corresponding to a change in the binary exposure level from $A = 0$ to $A = 1$, conditional on $\boldsymbol{C} = \boldsymbol{c}$, is defined as a contrast between counterfactual probabilities $P(Y(1) = 1|\boldsymbol{C} = \boldsymbol{c})$ and $P(Y(0) = 1|\boldsymbol{C} = \boldsymbol{c})$, and can be expressed, under the no unmeasured confounders, positivity and consistency assumptions, as a contrast between $P(Y = 1|A = 1, \boldsymbol{C} = \boldsymbol{c})$ and $P(Y = 1|A = 0, \boldsymbol{C} = \boldsymbol{c})$ (Valeri et VanderWeele, 2013; VanderWeele et Vansteelandt, 2010). In order to examine the decomposition property of the exact and approximate TE estimators on the OR and RR scales (Equations (4.6) and (4.10)), we compared the estimate of the total effect calculated as the product of NIE and NDE estimates to the one obtained without consideration of the mediator. More precisely, for each sample generated, probabilities $P(Y = 1|A = 1, \boldsymbol{C} = \boldsymbol{c})$ and $P(Y = 1|A = 0, \boldsymbol{C} = \boldsymbol{c})$ were estimated from the outcome logistic regression model

$$\text{logit}\{P(Y = 1|A = a, \boldsymbol{C} = \boldsymbol{c})\} = \theta^*_0 + \theta^*_1 a + \boldsymbol{\theta}^{*'}_4 \boldsymbol{c}, \quad a = 0, 1,$$

as $\widehat{P}(Y = 1|A = a, \boldsymbol{C} = \boldsymbol{c}) = \text{expit}\left(\widehat{\theta}^*_0 + \widehat{\theta}^*_1 a + \boldsymbol{\theta}^{*'}_4 \boldsymbol{c}\right)$, $a = 0, 1$, and converted to the so-called conventional (that is, nonmediated) TE estimates on the OR and RR scales:

$$\widehat{conv.OR}^{TE}_{1,0|\boldsymbol{c}} = \frac{\dfrac{\widehat{P}(Y = 1|A = 1, \boldsymbol{C} = \boldsymbol{c})}{1 - \widehat{P}(Y = 1|A = 1, \boldsymbol{C} = \boldsymbol{c})}}{\dfrac{\widehat{P}(Y = 1|A = 0, \boldsymbol{C} = \boldsymbol{c})}{1 - \widehat{P}(Y = 1|A = 0, \boldsymbol{C} = \boldsymbol{c})}} = \exp\left(\widehat{\theta}^*_1\right),$$

$$\widehat{conv.RR}^{TE}_{1,0|\boldsymbol{c}} = \frac{\widehat{P}(Y = 1|A = 1, \boldsymbol{C} = \boldsymbol{c})}{\widehat{P}(Y = 1|A = 0, \boldsymbol{C} = \boldsymbol{c})}.$$

The conventional TE estimates were then compared to TE estimates calculated as the product of the NIE and NDE based on proposed exact estimators using absolute and relative differences:

$$\widehat{OR^{TE}_{1,0|\boldsymbol{c}}} - \widehat{conv.OR}^{TE}_{1,0|\boldsymbol{c}}, \quad \frac{\widehat{OR^{TE}_{1,0|\boldsymbol{c}}} - \widehat{conv.OR}^{TE}_{1,0|\boldsymbol{c}}}{\widehat{conv.OR}^{TE}_{1,0|\boldsymbol{c}}},$$

$$\widehat{RR^{TE}_{1,0|\boldsymbol{c}}} - \widehat{conv.RR}^{TE}_{1,0|\boldsymbol{c}}, \quad \frac{\widehat{RR^{TE}_{1,0|\boldsymbol{c}}} - \widehat{conv.RR}^{TE}_{1,0|\boldsymbol{c}}}{\widehat{conv.RR}^{TE}_{1,0|\boldsymbol{c}}}.$$

A similar approach was taken to examine the TE decomposition property of the approximate estimator $\widehat{app\,OR^{TE}_{1,0|\boldsymbol{c}}}$. For both the exact and approximate estimators, the mean and standard deviation of these differences were computed over all 1000 samples generated.

The TE decomposition property results for the exact and approximate estimators are reported for the simulation study without covariates in Table 4.10. We can see, for all scenarios studied and both multiplicative scales, that the products of the exact NDE and NIE estimates over all samples generated were almost identical to the corresponding conventional TE estimates, with mean relative differences ranging between 0.0001% and 0.0029%. This is to be expected given that the logistic outcome model is saturated in this case and that corresponding estimated model-based probabilities for Y given A and M and for M given A match those that would be obtained from the corresponding two- and one-way frequency tables. For *Scenario 1*, the differences between the products of the approximate NDE and NIE estimates and conventional TE estimates were small: the mean relative differences were equal to -0.48% and 3.94% for the OR and RR scales, respectively. These differences were much larger for *Scenarios 2 - 4* (between 20.5% and 60.4%). In the approximate approach, the expit of the linear predictor associated to the outcome model is replaced by an exponential after the regression coefficients have been estimated (compare Equation (4.3) to Equation (4.8)). Therefore, the corresponding model-based conditional probabilities do not match those from the frequency table anymore, hence producing a discrepancy between the two types of total effects. For the OR scale, this discrepancy is additionally explained by substitution of the OR by RR (see Equation (4.9)).

The TE decomposition property results for the simulation study with covariates are presented in Table 4.11. When covariates are included in the outcome logistic model, it is expected to observe some difference between mediated and nonmediated total effects (Wang et Arah, 2015), but the magnitude of the disagreement is of interest. The mean relative differences between the products of the exact NDE and NIE estimates and the corresponding conventional TE estimates over all samples

generated ranged between -1.59% and 0.50%. For *Scenario 1*, the differences between the products of the approximate NDE and NIE estimates and conventional TE estimates were equal to 0.06% and 3.83% for the OR and RR scales, respectively, while those were much larger for *Scenarios 2 - 4* (between 31.6% and 69.6%).

Table 4.10 Simulation study without covariates: Total effect multiplicative decomposition property of exact and approximate natural-effects estimators by scenarios with increasing levels of outcome commonness

	Exact TE estimate[a] vs. TE estimate by conventional approach				Approximate TE estimate[a] vs. TE estimate by conventional approach			
	Difference[b]		Relative difference[c] (%)		Difference[b]		Relative difference[c] (%)	
Scale	Mean	SD	Mean	SD	Mean	SD	Mean	SD
				Scenario 1				
OR	0.0000	0.0001	0.0002	0.0042	-0.0108	0.0127	-0.48	0.56
RR	0.0000	0.0001	0.0001	0.0040	0.0887	0.0315	3.94	0.99
				Scenario 2				
OR	0.0001	0.0004	0.0029	0.0084	2.0826	0.4383	40.20	6.13
RR	0.0001	0.0003	0.0028	0.0083	2.7350	0.5293	60.37	7.70
				Scenario 3				
OR	0.0000	0.0001	0.0003	0.0060	0.4493	0.0851	40.98	6.14
RR	0.0000	0.0001	0.0003	0.0051	0.4643	0.0950	42.85	6.91
				Scenario 4				
OR	0.0000	0.0000	0.0003	0.0020	0.3678	0.1622	20.50	8.83
RR	0.0000	0.0000	0.0002	0.0012	0.7484	0.1826	52.90	12.10

Abbreviations: OR, odds ratio; RR, risk ratio; SD, standard deviation; TE, total effect.
[a] TE estimate defined as a product of NDE and NIE estimates.
[b] TE estimate – TE estimate by conventional approach.
[c] (TE estimate – TE estimate by conventional approach)/TE estimate by conventional approach.

Table 4.11 Simulation study with covariates: Total effect multiplicative decomposition property of exact and approximate natural-effects estimators by scenarios with increasing levels of outcome commonness

	Exact TE estimate[a] vs. TE estimate by conventional approach				Approximate TE estimate[a] vs. TE estimate by conventional approach			
	Difference[b]		Relative difference[c] (%)		Difference[b]		Relative difference[c] (%)	
Scale	Mean	SD	Mean	SD	Mean	SD	Mean	SD
				Scenario 1				
OR	0.0104	0.0098	0.4955	0.4589	0.0014	0.0076	0.0576	0.3555
RR	0.0097	0.0090	0.4788	0.4380	0.0795	0.0274	3.8282	0.8760
				Scenario 2				
OR	-0.0485	0.0403	-1.0288	0.8297	1.4846	0.3400	31.572	5.2809
RR	-0.0410	0.0336	-0.9784	0.7802	2.0079	0.4178	48.026	6.6082
				Scenario 3				
OR	-0.0170	0.0070	-1.5937	0.6371	0.3645	0.0747	34.170	5.7028
RR	-0.0143	0.0059	-1.3601	0.5445	0.3746	0.0839	35.361	6.4171
				Scenario 4				
OR	-0.0024	0.0014	-0.1357	0.0757	0.5961	0.1957	33.656	10.571
RR	-0.0008	0.0005	-0.0543	0.0373	0.9719	0.2210	69.645	14.657

Abbreviations: OR, odds ratio; RR, risk ratio; SD, standard deviation; TE, total effect.
[a] TE estimate defined as a product of NDE and NIE estimates.
[b] TE estimate – TE estimate by conventional approach.
[c] (TE estimate – TE estimate by conventional approach)/TE estimate by conventional approach.

4.6.8 Comments on estimation procedures

Simulation studies

In order to reduce execution time when generating the exact and approximate results in the simulation study, we created a global macro (available upon request) based on our principal macro and the macro by Valeri and VanderWeele (2013) to compute results *simultaneously*. Consistency of results for the approximate estimates returned from this macro and Valeri and VanderWeele original macro was verified for a few randomly selected samples generated.

Real-data example

Logistic regression was used to model the mediator in all the mediation analyses performed. Logistic regression was used to model the outcome when estimating natural effects on the OR scale with the SAS CAUSALMED procedure (Yung *et al.*, 2018). For the RR scale, an outcome log-binomial

regression model was specified in PROC CAUSALMED. An outcome Poisson regression model was used instead of a log-binomial model when the latter did not converge (notably with placental abruption as exposure). When using the delta method for the effect E expressed on the OR and RR scales, our SAS macro `mediation_estimates` returns a confidence interval as $E \cdot \exp(\pm\Phi^{-1}(0.975) \cdot se(\ln(E)))$, as in the Valeri and VanderWeele SAS macro. Consequently, the symmetric confidence interval $E \pm \Phi^{-1}(0.975) \cdot se(E)$ returned by PROC CAUSALMED is not perfectly comparable to the one obtained by our SAS macro when using the delta method for multiplicative scales.

A weighting-based approach to estimate conditional (or stratum-specific) natural effects was taken for all the mediation analyses performed using the R package `medflex` (Lange *et al.*, 2012; Steen *et al.*, 2017); NEMs with *logit* and *log* link functions were used for OR and RR scales, respectively. The 95% confidence intervals were constructed by percentile bootstrap based on 1000 resamples for the mediation effects when the exposure variable was treatment with inhaled corticosteroids. Percentile bootstrap was also applied to obtain mediation effects on the OR scale with placental abruption as exposure; confidence intervals based on the robust standard errors were calculated for the RR scale since `medflex` failed to provide bootstrap confidence intervals for these specific exposure and scale.

The quasi-Bayesian approach was implemented using the R package `mediation` with a *logit* link for the outcome model. Corresponding results on the RD scale were based on 5000 Monte-Carlo draws. The 95% confidence intervals were based on the White's heteroskedasticity-consistent estimator for the covariance matrix (Tingley *et al.*, 2014).

4.7 Appendix 2

Comments on the SAS macro ***`mediation_estimates`*** *execution*

Use of the SAS macro `mediation_estimates` (see Appendix 3, Section 4.8) requires the specification of macro variables. We provide three examples showing how to specify values for these variables.

The following statement returns crude (unadjusted) estimates for OR^{NDE}, OR^{NDE} and OR^{TE} 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% confidence intervals. Conventional logistic regressions without Firth penalization are also used.

```
%mediation_estimates(mydata=data, A=exposure, M=mediator, Y=outcome, interaction=0,
adjusted=0, a1=t1, a0=t0, boot=0, scale="OR", Firth=0)
```

To perform an adjusted and penalized (with Firth) mediation analysis on the RR scale, allowing for an exposure-mediator interaction and using bootstrap based on 5000 samples with initial random seed = 1234 to construct 95% confidence intervals, our SAS macro `mediation_estimates` should be executed as follows:

```
%mediation_estimates(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, scale="RR", Firth=1)
```

where $\texttt{Mvar}_1$ $\texttt{Mvar}_2$ ... $\texttt{Mvar}_\texttt{p}$ and $\texttt{Yvar}_1$ $\texttt{Yvar}_2$ ... $\texttt{Yvar}_\texttt{s}$ are the set of adjustment covariates for the mediator and outcome models, correspondingly. Due to the non-collapsibility of the logistic regression model (Daniel *et al.*, 2021; Neuhaus et Jewell, 1993), we advise against using different sets of adjustment covariates in the outcome and mediator models unless it is known that excluded covariates are independent of the response being modeled given the rest of covariates.

By default, our SAS macro reports mediation effects evaluated at the sample-specific mean values of the 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 `mediation_estimates`. For example, in order to estimate mediation effects corresponding to $\texttt{Mvar}_1 = \texttt{Cm}_1$, $\texttt{Mvar}_2 = \texttt{Cm}_2$, $\texttt{Mvar}_3 = \texttt{Cm}_3$ (i.e., at user-defined values for the first three adjustment covariates in the mediator model), and $\texttt{Yvar}_3 = \texttt{Cy}_3$, $\texttt{Yvar}_4 = \texttt{Cy}_4$ (i.e., at user-defined values for the third and fourth covariates in the outcome model), datasets `DATA_M` and `DATA_Y` can be constructed using `datalines` statements as follows:

```
data DATA_M; input Mvar1 Mvar2 Mvar3; datalines;
Cm1 Cm2 Cm3
;
```

```
data DATA_Y; input Yvar3 Yvar4; datalines;
Cy3 Cy4
;
```

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.

For example, the user can estimate mediation effects on the RD scale that correspond to the covariate values specified in DATA_M and DATA_Y, assuming an exposure-mediator interaction and using the delta method to construct 95% confidence intervals, as follows:

```
%mediation_estimates(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=0, scale="RD", Firth=0, stratum=1,
cvar_M_data=DATA_M, cvar_Y_data=DATA_Y)
```

If the covariates specified in DATA_M (DATA_Y) constitute some proper subset of {$\texttt{Mvar}_1$, $\texttt{Mvar}_2$, ... $\texttt{Mvar}_p$} ({$\texttt{Yvar}_1$, $\texttt{Yvar}_2$, ... $\texttt{Yvar}_s$}), then the other covariates will be set to their sample-specific mean levels.

Categorical covariates

If, for example, $\texttt{Mvar}_1$ $\texttt{Mvar}_2$ are two dummy variables coding some categorical covariate $\texttt{V}_{\texttt{cat}}$ 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 $\texttt{V}_{\texttt{cat}}$, 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 $\texttt{V}_{\texttt{cat}}$, 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.

Missing values

Our SAS macro mediation_estimates performs all computations on complete cases only. Users can

apply multiple imputation techniques implemented in the R package *mice* (van Buuren et Groothuis-Oudshoorn, 2011) and the SAS MI procedure (Yuan, 2011) to handle missing data.

4.8 Appendix 3

```
/*************************************************************************************************/
/* The user needs to specify the values for the following macro variables in the macro        */
/* %mediation_estimates:                                                                      */
/*                                                                                           */
/* mydata:      input data that include the outcome, exposure and mediator variables as well  */
/*              as the covariates to be adjusted for in the model;                           */
/* A:           the name of the exposure variable;                                           */
/* M:           the name of the mediator variable;                                           */
/* Y:           the name of the outcome variable;                                            */
/* interaction: the user needs to specify INTERACTION=0 or INTERACTION=1 for the outcome      */
/*              model without or with interaction between the exposure and the mediator,     */
/*              respectively;                                                                */
/* adjusted:    the user needs to specify ADJUSTED=0 or ADJUSTED=1 to obtain unadjusted or    */
/*              adjusted NIE, NDE and TE, respectively;                                      */
/* cvar_M:      the list of adjustment variables (covariates) in the mediator model;         */
/*              categorical variables need to be coded as a series of dummy variables        */
/*              before being entered as covariates; use space to separate covariates' names; */
/* cvar_Y:      the list of adjustment variables (covariates) in the outcome model;          */
/*              categorical variables need to be coded as a series of dummy variables        */
/*              before being entered as covariates;                                          */
/* a0:          the exposure level corresponding to a*;                                      */
/* a1:          the exposure level corresponding to a;                                       */
/* boot:        the user needs to specify BOOT=0 or BOOT=1 to obtain 95% confidence intervals */
/*              by the delta method or bootstrapping, respectively;                          */
/* bootseed:    if BOOT=1, that is bootstrap 95% confidence intervals are required, then the */
/*              user needs to specify the initial seed (positive integer) for random         */
/*              number generation;                                                           */
/* nboot:       if BOOT=1, that is bootstrap 95% confidence intervals are required, then the */
/*              user needs to specify the number of bootstrap samples;                       */
/* scale:       the user needs to specify SCALE="OR", SCALE="RR"or SCALE="RD" to obtain      */
/*              estimated mediation effects on the odds ratio, risk ratio or risk            */
/*              difference scale, respectively; double quotation marks must be used,         */
/*              e.g., "OR";                                                                  */
/* Firth:       the user needs to specify FIRTH=1 in order to use the Firth method in        */
/*              logistic regressions; if FIRTH=0, conventional maximum likelihood estimates  */
/*              are returned by logistic regressions;                                        */
/* stratum:     the user needs to specify STRATUM=1 to estimate mediation effects at specific */
/*              values of some covariates (that is, stratum-specific effects);               */
/* cvar_M_data: if STRATUM=1, the user needs to provide a SAS dataset with a single row that */
/*              contains specific values for some or all of the adjustment covariates  cvar_M */
/*              in the mediator model; if the covariates specified in cvar_M_data constitute */
/*              some proper subset of cvar_M, then the other covariates will be set to their */
/*              sample-specific mean levels;                                                 */
/* cvar_Y_data: if STRATUM=1, the user needs to provide a SAS dataset with a single row that */
/*              contains specific values for some or all of the adjustment covariates cvar_Y */
/*              in the outcome model; if the covariates specified in cvar_Y_data constitute  */
/*              some proper subset of cvar_Y, then the other covariates will be set to their */
/*              sample-specific mean levels.                                                 */
/*                                                                                           */
/*************************************************************************************************/

%macro mediation_estimates(mydata,A,M,Y,interaction,adjusted,cvar_M,cvar_Y,a1,a0,
boot,bootseed,nboot,scale,Firth,stratum,cvar_M_data,cvar_Y_data);

%if &adjusted=0 %then %do;

  %if &Firth=0 %then %do;
```

```
  proc logistic data=&mydata noprint outest=coeffs_M (drop=_:) desc; model &M = &A; run;
  %if &interaction=1 %then %do;
  proc logistic data=&mydata noprint outest=coeffs_Y (drop=_:) desc; model &Y = &A|&M; run;
  %end;
  %if &interaction=0 %then %do;
  proc logistic data=&mydata noprint outest=coeffs_Y (drop=_:) desc; model &Y = &A &M; run;
  %end;
  %end;

  %if &Firth=1 %then %do;
  proc logistic data=&mydata noprint outest=coeffs_M (drop=_:) desc;
  model &M = &A / Firth; run;
  %if &interaction=1 %then %do;
  proc logistic data=&mydata noprint outest=coeffs_Y (drop=_:) desc;
  model &Y = &A|&M / Firth; run;
  %end;
  %if &interaction=0 %then %do;
  proc logistic data=&mydata noprint outest=coeffs_Y (drop=_:) desc;
  model &Y = &A &M / Firth; run;
  %end;
  %end;
  %end;

%if &adjusted=1 %then %do;

  %if &Firth=0 %then %do;
  proc logistic data=&mydata noprint outest=coeffs_M (drop=_:) desc; model &M = &A &cvar_M; run;
  %if &interaction=1 %then %do;
  proc logistic data=&mydata noprint outest=coeffs_Y (drop=_:) desc; model &Y = &A|&M &cvar_Y; run;
  %end;
  %if &interaction=0 %then %do;
  proc logistic data=&mydata noprint outest=coeffs_Y (drop=_:) desc; model &Y = &A &M &cvar_Y; run;
  %end;
  %end;
  %if &Firth=1 %then %do;
  proc logistic data=&mydata noprint outest=coeffs_M (drop=_:) desc;
  model &M = &A &cvar_M  / Firth; run;
  %if &interaction=1 %then %do;
  proc logistic data=&mydata noprint outest=coeffs_Y (drop=_:) desc;
  model &Y = &A|&M &cvar_Y  / Firth; run;
  %end;
  %if &interaction=0 %then %do;
  proc logistic data=&mydata noprint outest=coeffs_Y (drop=_:) desc;
  model &Y = &A &M &cvar_Y  / Firth; run;
  %end;
  %end;

  proc means data=&mydata mean noprint; var &cvar_M; output out=cvar_values_M (where=(_STAT_="MEAN"));
  run;
  proc means data=&mydata mean noprint; var &cvar_Y; output out=cvar_values_Y (where=(_STAT_="MEAN"));
  run;

  %if &stratum=1 %then %do;
  proc datasets lib=work nolist; delete cvar_names_M cvar_names_Y; quit; run;
  %if %length(&cvar_M_data) = 0 %then %do;
  %put ERROR: if STRATUM=1, then cvar_M_data and cvar_Y_data must be specified;
  %abort; %end;
  %if %length(&cvar_Y_data) = 0 %then %do;
  %put ERROR: if STRATUM=1, then cvar_M_data and cvar_Y_data must be specified;
```

```
  %abort; %end;
  %if %sysfunc(exist(&cvar_M_data)) %then %do;
  proc contents data=&cvar_M_data noprint out=cvar_names_M (keep=name); run;
  %end;
  %else %do;
  %put ERROR: if STRATUM=1, then cvar_M_data must be provided;
  %abort; %end;
  %if %sysfunc(exist(&&cvar_Y_data)) %then %do;
  proc contents data=&cvar_Y_data noprint out=cvar_names_Y (keep=name); run;
  %end;
  %else %do;
  %put ERROR: if STRATUM=1, then cvar_Y_data must be provided;
  %abort; %end;
  proc sort data=cvar_names_M; by name; run;
  proc sort data=cvar_names_Y; by name; run;
  data common_vars;  merge cvar_names_M (in=a) cvar_names_Y (in=b); by name;
  if a and b then output; run;
  proc sql noprint; select count(*) into :count1 from common_vars; quit;
  %if &count1 ne 0 %then %do;
  ods exclude CompareDatasets   CompareDifferences;
  proc compare base=&cvar_M_data compare=&cvar_Y_data nosummary out=check  outnoequal; run;
  proc sql noprint; select count(*) into :count2 from check; quit;
  %if &count2 ne 0 %then %do;
  %put WARNING: Some common variables in &cvar_M_data and &cvar_Y_data do not have the same values
                (are unequal);
  %put WARNING: See RESULTS for details;
  %abort;
  %end;
  %end;
  data cvar_values_M; merge cvar_values_M &cvar_M_data; run;
  data cvar_values_Y; merge cvar_values_Y &cvar_Y_data; run;
  %end;
  %end;

%if &boot=0 %then %do;

  %if &adjusted=0 %then %do;
  %if &Firth=0 %then %do;
  proc logistic data=&mydata noprint outest=sigma_M (drop=_:) covout desc;
  model &M = &A ; run;
  %if &interaction=1 %then %do;
  proc logistic data=&mydata noprint outest=sigma_Y (drop=_:) covout desc;
  model &Y = &A|&M ; run;
  %end;
  %if &interaction=0 %then %do;
  proc logistic data=&mydata noprint outest=sigma_Y (drop=_:) covout desc;
  model &Y = &A &M; run;
  %end;
  %end;
  %if &Firth=1 %then %do;
  proc logistic data=&mydata noprint outest=sigma_M (drop=_:) covout desc;
  model &M = &A / Firth; run;
  %if &interaction=1 %then %do;
  proc logistic data=&mydata noprint outest=sigma_Y (drop=_:) covout desc;
  model &Y = &A|&M / Firth; run;
  %end;
  %if &interaction=0 %then %do;
  proc logistic data=&mydata noprint outest=sigma_Y (drop=_:) covout desc;
  model &Y = &A &M / Firth; run;
```

```
  %end;
  %end;
  %end;

  %if &adjusted=1 %then %do;
  %if &Firth=0 %then %do;
  proc logistic data=&mydata noprint outest=sigma_M (drop=_:) covout desc;
  model &M = &A &cvar_M; run;
  %if &interaction=1 %then %do;
  proc logistic data=&mydata noprint outest=sigma_Y (drop=_:) covout desc;
  model &Y = &A|&M &cvar_Y; run;
  %end;
  %if &interaction=0 %then %do;
  proc logistic data=&mydata noprint outest=sigma_Y (drop=_:) covout desc;
  model &Y = &A &M &cvar_Y; run;
  %end;
  %end;
  %if &Firth=1 %then %do;
  proc logistic data=&mydata noprint outest=sigma_M (drop=_:) covout desc;
  model &M = &A &cvar_M / Firth; run;
  %if &interaction=1 %then %do;
  proc logistic data=&mydata noprint outest=sigma_Y (drop=_:) covout desc;
  model &Y = &A|&M &cvar_Y / Firth; run;
  %end;
  %if &interaction=0 %then %do;
  proc logistic data=&mydata noprint outest=sigma_Y (drop=_:) covout desc;
  model &Y = &A &M &cvar_Y / Firth; run;
  %end;
  %end;
  %end;
  data sigma_M; set sigma_M; if _n_=1 then delete; run;
  data sigma_Y; set sigma_Y; if _n_=1 then delete; run;

%end;

proc iml;
use coeffs_M;  read var {Intercept} into beta_0;
               read var {&A} into beta_1;
%if &adjusted=1 %then %do; read var {&cvar_M} into cov_coeffs_M; %end;
%if &adjusted=0 %then %do; cov_coeffs_M=0; %end;
use coeffs_Y;  read var {Intercept} into theta_0;
               read var {&A} into theta_1;
               read var {&M} into theta_2;
%if &interaction=1  %then %do;  read var {&A.&M} into theta_3; %end;
%if &interaction=0  %then %do;  theta_3=0; %end;
%if &adjusted=1  %then %do; read var {&cvar_Y} into cov_coeffs_Y; %end;
%if &adjusted=0  %then %do;   cov_coeffs_Y=0; %end;
%if &adjusted=1 %then %do;
use cvar_values_M; read var {&cvar_M} into cov_values_M;
use cvar_values_Y; read var {&cvar_Y} into cov_values_Y;
product_betas_c = cov_coeffs_M*t(cov_values_M);
product_thetas_c = cov_coeffs_Y*t(cov_values_Y);
%end;
%if &adjusted=0 %then %do;
product_betas_c  = 0; product_thetas_c = 0;
%end;
K_1 = exp(beta_0+beta_1*&a1+product_betas_c);
K_0 = exp(beta_0+beta_1*&a0+product_betas_c);
L_1 = exp(theta_0+theta_1*&a1+theta_2+theta_3*&a1+product_thetas_c);
```

```
L_0 = exp(theta_0+theta_1*&a0+theta_2+theta_3*&a0+product_thetas_c);
M_1 = exp(theta_0+theta_1*&a1+product_thetas_c);
M_0 = exp(theta_0+theta_1*&a0+product_thetas_c);
P11 = (L_1/(1+L_1))*(K_1/(1+K_1))+(M_1/(1+M_1))*(1/(1+K_1));
P10 = (L_1/(1+L_1))*(K_0/(1+K_0))+(M_1/(1+M_1))*(1/(1+K_0));
P00 = (L_0/(1+L_0))*(K_0/(1+K_0))+(M_0/(1+M_0))*(1/(1+K_0));
%if &scale = "OR" %then %do;
NDE = (P10/(1-P10))/(P00/(1-P00));
NIE = (P11/(1-P11))/(P10/(1-P10));
TE = NDE*NIE;
%end;
%if &scale = "RR" %then %do;
NDE = P10/P00;
NIE = P11/P10;
TE = NDE*NIE;
%end;
%if &scale = "RD" %then %do;
NDE = P10-P00;
NIE = P11-P10;
TE = NDE+NIE;
%end;
%if &scale = "OR" %then %do;
CDE_0 = exp(theta_1*(&a1-&a0));
CDE_1 = exp((theta_1+theta_3)*(&a1-&a0));
%end;
%if &scale = "RR" %then %do;
CDE_0 = exp(theta_1*(&a1-&a0))*(1+M_0)/(1+M_1);
CDE_1 = exp((theta_1+theta_3)*(&a1-&a0))*(1+L_0)/(1+L_1);
%end;
%if &scale = "RD" %then %do;
CDE_0 = M_1/(1+M_1) - M_0/(1+M_0);
CDE_1 = L_1/(1+L_1) - L_0/(1+L_0);
%end;
point_estimates  =  NDE // NIE // TE;
create  point_estimates from point_estimates[colname={"mediation_effects"}];
append from point_estimates; close point_estimates;
point_estimates_CDE  =  CDE_0 // CDE_1;
create  point_estimates_CDE from point_estimates_CDE[colname={"CDE"}];
append from point_estimates_CDE; close point_estimates_CDE;

/****************************************** delta ******************************************/

%if &boot=0 %then %do;

    q = quantile("NORMAL",.975);
    use sigma_Y; read all into sigma_Y;
    use sigma_M; read all into sigma_M;
    sigma = block(sigma_Y,sigma_M);
    sigma_CDE = sigma_Y;
    %if &interaction=0 %then %do;
    d=dimension(sigma); z = j(1,d[1],0); tmp = insert(sigma,z,4,0); z = j(1,d[1]+1,0); sigma=insert(tmp,t(z),0,4);
    d=dimension(sigma_CDE);
    z = j(1,d[1],0); tmp = insert(sigma_CDE,z,4,0); z = j(1,d[1]+1,0);
    sigma_CDE=insert(tmp,t(z),0,4);
    %end;

    /* DERIVATIVES for GRADIENTS */

    dP11_dtheta0=(L_1/((1+L_1)**2))*(K_1/(1+K_1))+(M_1/((1+M_1)**2))*(1/(1+K_1));
```

```
dP11_dtheta1=dP11_dtheta0*&a1;
dP11_dtheta2=(L_1/((1+L_1)**2))*(K_1/(1+K_1));
dP11_dtheta3=dP11_dtheta2*&a1*&interaction;
dP11_dbeta0=(K_1/((1+K_1)**2))*(L_1/(1+L_1)-M_1/(1+M_1));
dP11_dbeta1=dP11_dbeta0*&a1;

dP10_dtheta0=(L_1/((1+L_1)**2))*(K_0/(1+K_0))+(M_1/((1+M_1)**2))*(1/(1+K_0));
dP10_dtheta1=dP10_dtheta0*&a1;
dP10_dtheta2=(L_1/((1+L_1)**2))*(K_0/(1+K_0));
dP10_dtheta3=dP10_dtheta2*&a1*&interaction;
dP10_dbeta0=(K_0/((1+K_0)**2))*(L_1/(1+L_1)-M_1/(1+M_1));
dP10_dbeta1=dP10_dbeta0*&a0;

dP00_dtheta0=(L_0/((1+L_0)**2))*(K_0/(1+K_0))+(M_0/((1+M_0)**2))*(1/(1+K_0));
dP00_dtheta1=dP00_dtheta0*&a0;
dP00_dtheta2=(L_0/((1+L_0)**2))*(K_0/(1+K_0));
dP00_dtheta3=dP00_dtheta2*&a0*&interaction;
dP00_dbeta0=(K_0/((1+K_0)**2))*(L_0/(1+L_0)-M_0/(1+M_0));
dP00_dbeta1=dP00_dbeta0*&a0;

%if &adjusted=1 %then %do;
dP11_dtheta4=dP11_dtheta0*cov_values_Y;
dP11_dbeta2=dP11_dbeta0*cov_values_M;
dP10_dtheta4=dP10_dtheta0*cov_values_Y;
dP10_dbeta2=dP10_dbeta0*cov_values_M;
dP00_dtheta4=dP00_dtheta0*cov_values_Y;
dP00_dbeta2=dP00_dbeta0*cov_values_M;
Gamma_P11=dP11_dtheta0||dP11_dtheta1||dP11_dtheta2||dP11_dtheta3||dP11_dtheta4||
          dP11_dbeta0||dP11_dbeta1||dP11_dbeta2;
Gamma_P10=dP10_dtheta0||dP10_dtheta1||dP10_dtheta2||dP10_dtheta3||dP10_dtheta4||
          dP10_dbeta0||dP10_dbeta1||dP10_dbeta2;
Gamma_P00=dP00_dtheta0||dP00_dtheta1||dP00_dtheta2||dP00_dtheta3||dP00_dtheta4||
          dP00_dbeta0||dP00_dbeta1||dP00_dbeta2;
%end;

%if &adjusted=0 %then %do;
Gamma_P11=dP11_dtheta0||dP11_dtheta1||dP11_dtheta2||dP11_dtheta3||dP11_dbeta0||dP11_dbeta1;
Gamma_P10=dP10_dtheta0||dP10_dtheta1||dP10_dtheta2||dP10_dtheta3||dP10_dbeta0||dP10_dbeta1;
Gamma_P00=dP00_dtheta0||dP00_dtheta1||dP00_dtheta2||dP00_dtheta3||dP00_dbeta0||dP00_dbeta1;
%end;

%if &scale = "OR" %then %do;
Gamma_log_NDE = Gamma_P10/(P10*(1-P10)) -  Gamma_P00/(P00*(1-P00));
Gamma_log_NIE = Gamma_P11/(P11*(1-P11)) -  Gamma_P10/(P10*(1-P10));
Gamma_log_TE =  Gamma_log_NIE +  Gamma_log_NDE;
se_log_NDE = sqrt(Gamma_log_NDE*sigma*t(Gamma_log_NDE));
se_log_NIE = sqrt(Gamma_log_NIE*sigma*t(Gamma_log_NIE));
se_log_TE =  sqrt(Gamma_log_TE*sigma*t(Gamma_log_TE));
NDE_low = NDE*exp(-q*se_log_NDE);   NDE_upp = NDE*exp(q*se_log_NDE);
NIE_low = NIE*exp(-q*se_log_NIE);   NIE_upp = NIE*exp(q*se_log_NIE);
TE_low  = TE*exp(-q*se_log_TE);     TE_upp =  TE*exp(q*se_log_TE);
%end;
%if &scale = "RR" %then %do;
Gamma_log_NDE = Gamma_P10/P10 -  Gamma_P00/P00;
Gamma_log_NIE = Gamma_P11/P11 -  Gamma_P10/P10;
Gamma_log_TE =  Gamma_log_NIE +  Gamma_log_NDE;
se_log_NDE = sqrt(Gamma_log_NDE*sigma*t(Gamma_log_NDE));
se_log_NIE = sqrt(Gamma_log_NIE*sigma*t(Gamma_log_NIE));
se_log_TE =  sqrt(Gamma_log_TE*sigma*t(Gamma_log_TE));
```

```
NDE_low = NDE*exp(-q*se_log_NDE);  NDE_upp = NDE*exp(q*se_log_NDE);
NIE_low = NIE*exp(-q*se_log_NIE);  NIE_upp = NIE*exp(q*se_log_NIE);
TE_low  = TE*exp(-q*se_log_TE);     TE_upp =  TE*exp(q*se_log_TE);
%end;
%if &scale = "RD" %then %do;
Gamma_NDE = Gamma_P10 - Gamma_P00;
Gamma_NIE = Gamma_P11 - Gamma_P10;
Gamma_TE =  Gamma_NDE + Gamma_NIE;
se_NDE = sqrt(Gamma_NDE*sigma*t(Gamma_NDE));
se_NIE = sqrt(Gamma_NIE*sigma*t(Gamma_NIE));
se_TE =  sqrt(Gamma_TE*sigma*t(Gamma_TE));
NDE_low = NDE-q*se_NDE;   NDE_upp = NDE+q*se_NDE;
NIE_low = NIE-q*se_NIE;   NIE_upp = NIE+q*se_NIE;
TE_low  = TE -q*se_TE;    TE_upp =  TE +q*se_TE;
%end;
NDE_CI = NDE_low || NDE_upp;
NIE_CI = NIE_low || NIE_upp;
TE_CI  = TE_low  || TE_upp;
delta_CI  =  NDE_CI // NIE_CI // TE_CI;
create delta_CI from delta_CI[colname={"delta_CI_low" "delta_CI_upp"}];
append from delta_CI; close delta_CI;

/* CDE: DERIVATIVES for GRADIENTS */

%if &scale = "OR" %then %do;
dlog_CDE_0_dtheta0=0;
dlog_CDE_0_dtheta1=(&a1-&a0);
dlog_CDE_0_dtheta2=0;
dlog_CDE_0_dtheta3=0;
dlog_CDE_1_dtheta0=0;
dlog_CDE_1_dtheta1=(&a1-&a0);
dlog_CDE_1_dtheta2=0;
dlog_CDE_1_dtheta3=(&a1-&a0)*&interaction;
%if &adjusted=1 %then %do;
dlog_CDE_0_dtheta4=0*cov_values_Y;
dlog_CDE_1_dtheta4=0*cov_values_Y;
Gamma_log_CDE_0=
dlog_CDE_0_dtheta0||dlog_CDE_0_dtheta1||dlog_CDE_0_dtheta2||dlog_CDE_0_dtheta3||dlog_CDE_0_dtheta4;
Gamma_log_CDE_1=
dlog_CDE_1_dtheta0||dlog_CDE_1_dtheta1||dlog_CDE_1_dtheta2||dlog_CDE_1_dtheta3||dlog_CDE_1_dtheta4;
%end;
%if &adjusted=0 %then %do;
Gamma_log_CDE_0=dlog_CDE_0_dtheta0||dlog_CDE_0_dtheta1||dlog_CDE_0_dtheta2||dlog_CDE_0_dtheta3;
Gamma_log_CDE_1=dlog_CDE_1_dtheta0||dlog_CDE_1_dtheta1||dlog_CDE_1_dtheta2||dlog_CDE_1_dtheta3;
%end;
se_log_CDE_0 = sqrt(Gamma_log_CDE_0*sigma_CDE*t(Gamma_log_CDE_0));
se_log_CDE_1 = sqrt(Gamma_log_CDE_1*sigma_CDE*t(Gamma_log_CDE_1));
CDE_0_low = CDE_0*exp(-q*se_log_CDE_0); CDE_0_upp = CDE_0*exp(q*se_log_CDE_0);
CDE_1_low = CDE_1*exp(-q*se_log_CDE_1); CDE_1_upp = CDE_1*exp(q*se_log_CDE_1);
%end;

%if &scale = "RR" %then %do;
dlog_CDE_0_dtheta0=M_0/(1+M_0)-M_1/(1+M_1);
dlog_CDE_0_dtheta1=&a1/(1+M_1)-&a0/(1+M_0);
dlog_CDE_0_dtheta2=0;
dlog_CDE_0_dtheta3=0;
dlog_CDE_1_dtheta0=L_0/(1+L_0)-L_1/(1+L_1);
dlog_CDE_1_dtheta1=&a1/(1+L_1)-&a0/(1+L_0);
dlog_CDE_1_dtheta2=dlog_CDE_1_dtheta0;
```

```
    dlog_CDE_1_dtheta3=dlog_CDE_1_dtheta1*&interaction;
    %if &adjusted=1 %then %do;
    dlog_CDE_0_dtheta4=dlog_CDE_0_dtheta0*cov_values_Y;
    dlog_CDE_1_dtheta4=dlog_CDE_1_dtheta0*cov_values_Y;
    Gamma_log_CDE_0=
    dlog_CDE_0_dtheta0||dlog_CDE_0_dtheta1||dlog_CDE_0_dtheta2||dlog_CDE_0_dtheta3||dlog_CDE_0_dtheta4;
    Gamma_log_CDE_1=
    dlog_CDE_1_dtheta0||dlog_CDE_1_dtheta1||dlog_CDE_1_dtheta2||dlog_CDE_1_dtheta3||dlog_CDE_1_dtheta4;
    %end;
    %if &adjusted=0 %then %do;
    Gamma_log_CDE_0=dlog_CDE_0_dtheta0||dlog_CDE_0_dtheta1||dlog_CDE_0_dtheta2||dlog_CDE_0_dtheta3;
    Gamma_log_CDE_1=dlog_CDE_1_dtheta0||dlog_CDE_1_dtheta1||dlog_CDE_1_dtheta2||dlog_CDE_1_dtheta3;
    %end;
    se_log_CDE_0 = sqrt(Gamma_log_CDE_0*sigma_CDE*t(Gamma_log_CDE_0));
    se_log_CDE_1 = sqrt(Gamma_log_CDE_1*sigma_CDE*t(Gamma_log_CDE_1));
    CDE_0_low = CDE_0*exp(-q*se_log_CDE_0); CDE_0_upp = CDE_0*exp(q*se_log_CDE_0);
    CDE_1_low = CDE_1*exp(-q*se_log_CDE_1); CDE_1_upp = CDE_1*exp(q*se_log_CDE_1);
    %end;

    %if &scale = "RD" %then %do;
    dCDE_0_dtheta0=M_1/((1+M_1)**2) - M_0/((1+M_0)**2);
    dCDE_0_dtheta1=&a1*M_1/((1+M_1)**2) - &a0*M_0/((1+M_0)**2);
    dCDE_0_dtheta2=0;
    dCDE_0_dtheta3=0;
    dCDE_1_dtheta0=L_1/((1+L_1)**2) - L_0/((1+L_0)**2);
    dCDE_1_dtheta1=&a1*L_1/((1+L_1)**2) - &a0*L_0/((1+L_0)**2);
    dCDE_1_dtheta2=dCDE_1_dtheta0;
    dCDE_1_dtheta3=dCDE_1_dtheta1*&interaction;
    %if &adjusted=1 %then %do;
    dCDE_0_dtheta4=dCDE_0_dtheta0*cov_values_Y;
    dCDE_1_dtheta4=dCDE_1_dtheta0*cov_values_Y;
    Gamma_CDE_0=dCDE_0_dtheta0||dCDE_0_dtheta1||dCDE_0_dtheta2||dCDE_0_dtheta3||dCDE_0_dtheta4;
    Gamma_CDE_1=dCDE_1_dtheta0||dCDE_1_dtheta1||dCDE_1_dtheta2||dCDE_1_dtheta3||dCDE_1_dtheta4;
    %end;
    %if &adjusted=0 %then %do;
    Gamma_CDE_0=dCDE_0_dtheta0||dCDE_0_dtheta1||dCDE_0_dtheta2||dCDE_0_dtheta3;
    Gamma_CDE_1=dCDE_1_dtheta0||dCDE_1_dtheta1||dCDE_1_dtheta2||dCDE_1_dtheta3;
    %end;
    se_CDE_0 = sqrt(Gamma_CDE_0*sigma_CDE*t(Gamma_CDE_0));
    se_CDE_1 = sqrt(Gamma_CDE_1*sigma_CDE*t(Gamma_CDE_1));
    CDE_0_low = CDE_0-q*se_CDE_0;  CDE_0_upp = CDE_0+q*se_CDE_0;
    CDE_1_low = CDE_1-q*se_CDE_1;  CDE_1_upp = CDE_1+q*se_CDE_1;
    %end;
    CDE_0_CI = CDE_0_low || CDE_0_upp;
    CDE_1_CI = CDE_1_low || CDE_1_upp;
    delta_CDE_CI  = CDE_0_CI // CDE_1_CI;
    create delta_CDE_CI from delta_CDE_CI[colname={"delta_CI_low" "delta_CI_upp"}];
    append from delta_CDE_CI; close delta_CDE_CI;
%end; /* end boot=0 (i.e. delta=1) */
quit;

/****************************************** boot ******************************************/
%if &boot=1 %then %do;

    proc delete data=WORK.bootdata (gennum=all); run;
    %if %length(&nboot) = 0 %then %do;
    %put ERROR: if BOOT=1, then NBOOT and BOOTSEED must be specified;
    %abort; %end;
    %if %length(&bootseed) = 0 %then %do;
```

```
%put ERROR: if BOOT=1, then NBOOT and BOOTSEED must be specified;
%abort; %end;

options nonotes nosource nosource2 errors=0;

%if &adjusted=0 %then %do; data for_boot; set &mydata; keep &A &M &Y; run;
%end;
%if &adjusted=1 %then %do; data for_boot; set &mydata; keep &A &M &Y &cvar_M &cvar_Y; run;
%end;

proc surveyselect data= for_boot noprint
out=bootdata seed=&bootseed method=urs samprate=1 outhits rep=&nboot; run;

%if &adjusted=0 %then %do;
%if &Firth=0 %then %do;
proc logistic data=bootdata  noprint outest=coeffs_M (drop=_:) desc;
model &M  = &A; by Replicate; run;
%if &interaction=1 %then %do;
proc logistic data=bootdata noprint outest=coeffs_Y (drop=_:) desc;
model &Y  = &A|&M; by Replicate; run;
%end;
%if &interaction=0 %then %do;
proc logistic data=bootdata  noprint outest=coeffs_Y (drop=_:) desc;
model &Y  = &A &M; by Replicate; run;
%end;
%end;
%if &Firth=1 %then %do;
proc logistic data=bootdata  noprint outest=coeffs_M (drop=_:) desc;
model &M  = &A / Firth; by Replicate; run;
%if &interaction=1 %then %do;
proc logistic data=bootdata noprint outest=coeffs_Y (drop=_:) desc;
model &Y  = &A|&M / Firth; by Replicate; run;
%end;
%if &interaction=0 %then %do;
proc logistic data=bootdata  noprint outest=coeffs_Y (drop=_:) desc;
model &Y  = &A &M / Firth; by Replicate; run;
%end;
%end;
%end;
%if &adjusted=1  %then %do;
%if &Firth=0 %then %do;
proc logistic data=bootdata  noprint outest=coeffs_M (drop=_:) desc;
model &M  = &A &cvar_M; by Replicate; run;
%if &interaction=1 %then %do;
proc logistic data=bootdata noprint outest=coeffs_Y (drop=_:) desc;
model &Y  = &A|&M &cvar_Y; by Replicate; run;
%end;
%if &interaction=0 %then %do;
proc logistic data=bootdata  noprint outest=coeffs_Y (drop=_:) desc;
model &Y  = &A &M &cvar_Y; by Replicate; run;
%end;
%end;
%if &Firth=1 %then %do;
proc logistic data=bootdata  noprint outest=coeffs_M (drop=_:) desc;
model &M  = &A &cvar_M / Firth; by Replicate; run;
%if &interaction=1 %then %do;
proc logistic data=bootdata noprint outest=coeffs_Y (drop=_:) desc;
model &Y  = &A|&M &cvar_Y / Firth; by Replicate; run;
%end;
```

```
%if &interaction=0 %then %do;
proc logistic data=bootdata  noprint outest=coeffs_Y (drop=_:) desc;
model &Y  = &A &M &cvar_Y / Firth; by Replicate; run;
%end;
%end;
%end;
proc iml;
use coeffs_M;
read all var {Intercept} into beta_0; read all var {&A}  into beta_1;
%if &adjusted=1 %then %do; read all var {&cvar_M} into cov_coeffs_M; %end;
use coeffs_Y;
read all var {Intercept} into theta_0; read all var {&A} into theta_1;
read all var {&M} into theta_2;
%if &interaction=1 %then %do; read all var {&A.&M} into theta_3; %end;
%if &interaction=0 %then %do; theta_3=0; %end;
%if &adjusted=1 %then %do; read all var {&cvar_Y} into cov_coeffs_Y; %end;
%if &adjusted=1 %then %do;
use cvar_values_M;  read all var {&cvar_M}  into cov_values_M;
use cvar_values_Y;  read all var {&cvar_Y}  into cov_values_Y;
product_betas_c  = cov_coeffs_M*t(cov_values_M);
product_thetas_c = cov_coeffs_Y*t(cov_values_Y);
%end;
%if &adjusted=0 %then %do;  product_betas_c = 0; product_thetas_c = 0; %end;
K_1 = exp(beta_0+beta_1*&a1+product_betas_c);
K_0 = exp(beta_0+beta_1*&a0+product_betas_c);
L_1 = exp(theta_0+theta_1*&a1+theta_2+theta_3*&a1 + product_thetas_c);
L_0 = exp(theta_0+theta_1*&a0+theta_2+theta_3*&a0 + product_thetas_c);
M_1 = exp(theta_0+theta_1*&a1 + product_thetas_c);
M_0 = exp(theta_0+theta_1*&a0 + product_thetas_c);
P11 = (L_1/(1+L_1))#(K_1/(1+K_1))+(M_1/(1+M_1))#(1/(1+K_1));
P10 = (L_1/(1+L_1))#(K_0/(1+K_0))+(M_1/(1+M_1))#(1/(1+K_0));
P00 = (L_0/(1+L_0))#(K_0/(1+K_0))+(M_0/(1+M_0))#(1/(1+K_0));
%if &scale = "OR" %then %do;
CDE_0 = exp(theta_1*(&a1-&a0));
CDE_1 = exp((theta_1+theta_3)*(&a1-&a0));
%end;
%if &scale = "RR" %then %do;
CDE_0 = exp(theta_1*(&a1-&a0))#((1+M_0)/(1+M_1));
CDE_1 = exp((theta_1+theta_3)*(&a1-&a0))#((1+L_0)/(1+L_1));
%end;
%if &scale = "RD" %then %do;
CDE_0 = M_1/(1+M_1) - M_0/(1+M_0);
CDE_1 = L_1/(1+L_1) - L_0/(1+L_0);
%end;
probs = P11||P10||P00;
create probabilities from probs[colname={'P11' 'P10' 'P00'}];
append from probs; close probabilities;
CDE = CDE_0 || CDE_1;
create CDE from CDE[colname={'CDE_0' 'CDE_1'}]; append from CDE; close CDE;
quit;

%if &scale="OR" %then %do;
data boot_effects; set probabilities;
NDE = (P10/(1-P10))/(P00/(1-P00));
NIE = (P11/(1-P11))/(P10/(1-P10));
TE  = NDE*NIE;
run;
%end;
%if &scale="RR" %then %do;
```

```
    data boot_effects; set probabilities;
    NDE = P10/P00;
    NIE = P11/P10;
    TE  = NDE*NIE;
    run;
    %end;
    %if &scale="RD" %then %do;
    data boot_effects; set probabilities;
    NDE = P10-P00;
    NIE = P11-P10;
    TE  = NDE+NIE;
    run;
    %end;
    proc univariate data=boot_effects noprint;
    var NIE NDE TE;
    output out=boot_quantiles  pctlpre= NIE NDE TE  pctlpts=2.5 97.5  pctlname= pct_low pct_upp ;
    run;
    proc iml; use boot_quantiles;
    read var {NDEpct_low} into NDE_low; read var {NDEpct_upp} into NDE_upp; NDE_CI = NDE_low || NDE_upp;
    read var {NIEpct_low} into NIE_low; read var {NIEpct_upp} into NIE_upp;
NIE_CI = NIE_low || NIE_upp;
    read var {TEpct_low}  into TE_low;  read var {TEpct_upp}  into TE_upp;  TE_CI  = TE_low  || TE_upp;
    boot_CI  =  NDE_CI // NIE_CI // TE_CI;
    create boot_CI from boot_CI[colname={"boot_CI_low" "boot_CI_upp"}];
    append from boot_CI; close boot_CI; quit;
    proc univariate data=CDE noprint;
    var  CDE_0 CDE_1;
    output out=CDE_quantiles  pctlpre= CDE_0 CDE_1  pctlpts=2.5 97.5  pctlname= pct_low pct_upp ;
    run;
    proc iml; use CDE_quantiles;
    read var {CDE_0pct_low} into CDE_0_low; read var {CDE_0pct_upp} into CDE_0_upp;
    CDE_0_CI = CDE_0_low || CDE_0_upp;
    read var {CDE_1pct_low} into CDE_1_low; read var {CDE_1pct_upp} into CDE_1_upp;
    CDE_1_CI = CDE_1_low || CDE_1_upp;
    boot_CDE_CI  =  CDE_0_CI// CDE_1_CI;
    create boot_CDE_CI from boot_CDE_CI[colname={"boot_CI_low" "boot_CI_upp"}];
    append from boot_CDE_CI; close boot_CDE_CI; quit;
    options notes source source2 errors=20;
%end; /* end boot=1*/

/*****************************  Mediation point estimates with CI ******************************/

%if &boot=0 %then %do;
data all_estimates; retain effect;
if _n_=1 then effect="NDE";  if _n_=2 then effect="NIE"; if _n_=3 then effect="TE";
merge point_estimates delta_CI; run;
title  color=blue "Scale=%qsysfunc(compress(&scale ,%str(%'))),
adjusted=&adjusted,  A-M interaction=&interaction,  Firth=&Firth,";
title2 color=blue "95% CI: delta method";
proc print data=all_estimates; run; title;

data CDE_estimates; retain effect;
if _n_=1 then effect="CDE: M=0";  if _n_=2 then effect="CDE: M=1";
merge point_estimates_CDE delta_CDE_CI; run;
title  color=blue "Scale=%qsysfunc(compress(&scale ,%str(%'))),
adjusted=&adjusted,  A-M interaction=&interaction,  Firth=&Firth,";
title2 color=blue "CDE 95% CI: delta method";
proc print data=CDE_estimates; run; title;
%end;
```

```
%if &boot=1 %then %do;
data all_estimates; retain effect;
if _n_=1 then effect="NDE";  if _n_=2 then effect="NIE"; if _n_=3 then effect="TE";
merge point_estimates boot_CI; run;
title  color=blue "Scale=%qsysfunc(compress(&scale ,%str(%'))),
adjusted=&adjusted,  A-M interaction=&interaction,  Firth=&Firth,";
title2 color=blue "95% CI: percentile bootstrap based on &nboot samples";
proc print data=all_estimates; run; title;

data CDE_estimates; retain effect;
if _n_=1 then effect="CDE: M=0";  if _n_=2 then effect="CDE: M=1";
merge point_estimates_CDE boot_CDE_CI; run;
title  color=blue "Scale=%qsysfunc(compress(&scale ,%str(%'))),
adjusted=&adjusted,  A-M interaction=&interaction,  Firth=&Firth,";
title2 color=blue "CDE 95% CI: percentile bootstrap based on &nboot samples";
proc print data=CDE_estimates; run; title;
%end;

%mend mediation_estimates;
```

CHAPITRE V

ARTICLE 3. AN EXACT REGRESSION-BASED APPROACH FOR THE ESTIMATION OF NATURAL DIRECT AND INDIRECT EFFECTS WITH A BINARY OUTCOME AND A CONTINUOUS MEDIATOR

Dans ce chapitre formé de l'article de Samoilenko et Lefebvre (2023) publié dans *Statistics in Medicine*, nous présentons les estimateurs paramétriques des effets naturels direct et indirect pour une variable réponse binaire et un médiateur continu qui ne reposent pas sur les hypothèses de la réponse rare (marginalement ou conditionnellement) ou commune. Le lecteur peut aussi accéder à cet article en suivant le lien `https://doi.org/10.1002/sim.9621`.

Mariia Samoilenko and Geneviève Lefebvre

Abstract: 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.

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.

5.1 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 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* (Imai *et al.*, 2010; Pearl, 2001, 2012). 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 (Loeys *et al.*, 2013).

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 (Gaynor *et al.*, 2019). To circumvent this inconvenience, VanderWeele and Vansteelandt (2010) 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 (2013), the SAS procedure PROC CAUSALMED (Yung *et al.*, 2018) and in the novel R package

`CMAverse` (Shi *et al.*, 2021).

Under the above standard specification of the binary outcome and continuous mediator models, Gaynor *et al.* (2019) 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 (e.g., minimax solution (Camilli, 1994)) or using some statistical criteria (e.g., equality of variances (Cox, 1970) or Kullback-Leibler information criterion (Savalei, 2006)). In the literature, proposed values for this scaling constant range between 0.551 and 0.625 (Savalei, 2006). In Gaynor *et al.* (2019), 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.* (2019) 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 (e.g., 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 (2018) 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 (2010)'s approach with a continuous mediator.

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 (Cheng *et al.*, 2021; Doretti *et al.*, 2021; Samoilenko *et al.*, 2018; Samoilenko et Lefebvre, 2021). Notably, Samoilenko and Lefebvre (2021) 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 (2021)'s exact approach 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 (2021) compared exact and approximate natural effects estimators (Gaynor *et al.*, 2019; Valeri et VanderWeele, 2013; VanderWeele et Vansteelandt, 2010) 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.* (2021) 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 (2010) were observed in the simulation settings with a rare marginal outcome they considered.

The present work is a follow up of Samoilenko and Lefebvre (2021) and Cheng *et al.* (2021). As in Cheng *et al.* (2021), 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.* (2021) – 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 (2010) and Gaynor *et al.* (2019). 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.* (2012) and Imai *et al.* (2010)'s 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 (Allison, 2012; Mansournia *et al.*, 2018). In Cheng *et al.* (2021), numerical problems were observed for small sample sizes and low outcome prevalences. Therefore, an additional objective of this paper was to investigate the impact of Firth's penalization, a popular approach widely implemented in statistical software to deal with aforementioned estimation problems (Greenland et Mansournia, 2015), on the exact natural effects estimators proposed.

5.2 Methods

5.2.1 Models and nested counterfactual outcome probabilities

Let us note A the exposure (binary or continuous). As in VanderWeele and Vansteelandt (2010), Valeri and VanderWeele (2013) and Gaynor *et al.* (2019), we assume the following linear and logistic regression models for the continuous mediator M and binary outcome Y, respectively:

$$E\{M|A=a,\boldsymbol{C}=\boldsymbol{c}\}=\beta_0+\beta_1 a+\boldsymbol{\beta}_2^{'}\boldsymbol{c},\quad \epsilon\sim\mathcal{N}(0,\sigma^2), \tag{5.1}$$

$$\text{logit}\{P(Y=1|A=a,M=m,\boldsymbol{C}=\boldsymbol{c})\}=\theta_0+\theta_1 a+\theta_2 m+\theta_3 am+\boldsymbol{\theta}_4^{'}\boldsymbol{c}, \tag{5.2}$$

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 5.6.1) and modeling assumptions in Equations (5.1) and (5.2), 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:

$$
\begin{aligned}
P(Y(a, M(a^*)) = 1|\boldsymbol{C} = \boldsymbol{c}) &= \int_{-\infty}^{\infty} P(Y = 1|A = a, M = m, \boldsymbol{C} = \boldsymbol{c})\, d\Phi_{M|A=a^*,\boldsymbol{C}=\boldsymbol{c}}(m) \\
&= \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} \text{expit}\left(\theta_0 + \theta_1 a + \theta_2 m + \theta_3 am + \boldsymbol{\theta}_4^{'}\boldsymbol{c}\right) \\
&\quad \times \exp\left(-\frac{\left(m - \left(\beta_0 + \beta_1 a^* + \boldsymbol{\beta}_2^{'}\boldsymbol{c}\right)\right)^2}{2\sigma^2}\right) dm,
\end{aligned} \tag{5.3}
$$

where $\Phi_{M|A=a^*,\boldsymbol{C}=\boldsymbol{c}}$ is the cumulative distribution function of the normal distribution $\mathcal{N}\big(\beta_0 + \beta_1 a^* + \boldsymbol{\beta}_2^{'}\boldsymbol{c}, \sigma^2\big)$ and $\text{expit}(\alpha) = \exp(\alpha)/(1 + \exp(\alpha))$.

The mediation formula (Imai *et al.*, 2010; Pearl, 2001, 2012) 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 (5.3) in an exact manner; for a change in the exposure level from $A = a^*$ to $A = a$, these effects are:

$$
OR^{NDE}_{a,a^*|\boldsymbol{c}} = \frac{g(a, a^*, \boldsymbol{c})/(1 - g(a, a^*, \boldsymbol{c}))}{g(a^*, a^*, \boldsymbol{c})/(1 - g(a^*, a^*, \boldsymbol{c}))}, \quad OR^{NIE}_{a,a^*|\boldsymbol{c}} = \frac{g(a, a, \boldsymbol{c})/(1 - g(a, a, \boldsymbol{c}))}{g(a, a^*, \boldsymbol{c})/(1 - g(a, a^*, \boldsymbol{c}))}, \tag{5.4}
$$

$$
RR^{NDE}_{a,a^*|\boldsymbol{c}} = \frac{g(a, a^*, \boldsymbol{c})}{g(a^*, a^*, \boldsymbol{c})}, \quad RR^{NIE}_{a,a^*|\boldsymbol{c}} = \frac{g(a, a, \boldsymbol{c})}{g(a, a^*, \boldsymbol{c})}, \tag{5.5}
$$

$$
RD^{NDE}_{a,a^*|\boldsymbol{c}} = g(a, a^*, \boldsymbol{c}) - g(a^*, a^*, \boldsymbol{c}), \quad RD^{NIE}_{a,a^*|\boldsymbol{c}} = g(a, a, \boldsymbol{c}) - g(a, a^*, \boldsymbol{c}), \tag{5.6}
$$

where

$$
\begin{aligned}
g(a, a^*, \boldsymbol{c}) &= \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} \text{expit}\left(\theta_0 + \theta_1 a + \theta_2 m + \theta_3 am + \boldsymbol{\theta}_4^{'}\boldsymbol{c}\right) \\
&\quad \times \exp\left(-\frac{\left(m - \left(\beta_0 + \beta_1 a^* + \boldsymbol{\beta}_2^{'}\boldsymbol{c}\right)\right)^2}{2\sigma^2}\right) dm.
\end{aligned} \tag{5.7}
$$

The improper integral in Equation (5.7) 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, $OR^{TE}_{a,a^*|\boldsymbol{c}}$ and $RR^{TE}_{a,a^*|\boldsymbol{c}}$, are defined as the product of the

NDE and NIE on their respective scale:

$$OR^{TE}_{a,a^*|\boldsymbol{c}} = OR^{NDE}_{a,a^*|\boldsymbol{c}} \times OR^{NIE}_{a,a^*|\boldsymbol{c}}, \quad RR^{TE}_{a,a^*|\boldsymbol{c}} = RR^{NDE}_{a,a^*|\boldsymbol{c}} \times RR^{NIE}_{a,a^*|\boldsymbol{c}}. \tag{5.8}$$

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

$$RR^{TE}_{a,a^*|\boldsymbol{c}} = RR^{NDE}_{a,a^*|\boldsymbol{c}} + RR^{NIE}_{a,a^*|\boldsymbol{c}}. \tag{5.9}$$

For each scale, the exact NDE and NIE estimators are obtained from Equations (5.4)-(5.7) by replacing the parameters involved in Equation (5.7) 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 (5.7) 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 (SAS Institute Inc., 2010) that implements adaptive Romberg-type integration techniques and which is devised to deal with singularities, functions with large derivatives, and infinite domains (SAS Institute Inc., 2010; Wicklin, 2011). Adaptive Romberg-type integration techniques are notably known to have advantages over Gauss-Hermite and Gauss-Laguerre quadratures for infinite intervals (such as in Equation (5.7))(SAS Institute Inc., 2010).

The formulas for calculating the standard errors via the first-order multivariate delta method (Casella et Berger, 2002) are provided in Appendix 5.6.2. The theoretical convergence of the improper integrals involved in the delta method is established in Appendix 5.6.3.

5.2.2 Simulation studies

5.2.2.1 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 $\mathcal{N}\big(\beta_0 + \beta_1 a,\, \sigma^2\big)$ 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 $Bernoulli\,(p_{C_1})$ and a $\mathcal{N}(\mu_{C_2}, \sigma^2_{C_2})$ distributions, respectively, where $p_{C_1} = 0.5$, $\mu_{C_2} = 0$ and $\sigma^2_{C_2} = 0.75^2$. Secondly, 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 $\mathcal{N}\big(\beta_0 + \beta_1 a + \beta_{21} c_1 + \beta_{22} c_2, \sigma^2\big)$ 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 (5.4)-(5.9), 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 (SAS Institute Inc., 2010).

Description of analyses

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 (2010) and Gaynor *et al.* (2019). For the approach by Gaynor *et al.* (2019), 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.* (2012) and Imai *et al.* (2010)'s 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` (Steen *et al.*, 2017) 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.* (2017) 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. Imai *et al.* (2010)'s approach was implemented with the R package `mediation` (Tingley *et al.*, 2014), 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 (Steen *et al.*, 2017). 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 (2010), 95% CIs were constructed by percentile bootstrap (Chernick, 2011) based on 500 resamples with replacement and using the first-order delta method (Casella et Berger, 2002). For

the approximate approach by Gaynor *et al.* (2019), 95% CIs were obtained by percentile bootstrap only. For the NEM approach, 95% CIs were constructed using robust standard errors based on the sandwich estimator (Liang et Zeger, 1986). For the parametric inference algorithm by Imai *et al.* (2010), 95% CIs were based on White's heteroskedasticity-consistent estimator for the covariance matrix (White, 1980).

5.2.2.2 Simulation study with a marginal but not conditional rare outcome

As in Samoilenko and Lefebvre (2021), 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 $S_1 = \{m\text{: } m \leq Q_1\}$, $S_2 = \{m\text{: } Q_1 < m \leq Q_2\}$, $S_3 = \{m\text{: } Q_2 < m \leq Q_3\}$, $S_4 = \{m\text{: } m > Q_3\}$, 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 $\hat{P}(Y = 1|A = 1, M \in S_4) = 14.84\%$.

5.2.2.3 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 (i.e., cases when $Y = 1$) for some

combinations of regressors levels (Cole *et al.*, 2014; Greenland *et al.*, 2016). In logistic regression analyses 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) (Gelman *et al.*, 2008; Šinkovec *et al.*, 2019). Separation problems are likely to be associated with a low number of events per variable (EPV) (van Smeden *et al.*, 2016). Furthermore, low EPV is one of the data features contributing to sparse-data bias (Greenland *et al.*, 2016; van Smeden *et al.*, 2016). Firth's penalization is generally considered as a effective tool to deal with small-sample or sparse-data biases (Cole *et al.*, 2014; Heinze et Schemper, 2002). 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 (Heinze et Schemper, 2002). Firth's penalization is also a default choice to handle separation issues in logistic regression analyses as this method yields finite and consistent estimates when separation occurs (Allison, 2012; Heinze et Schemper, 2002; Šinkovec *et al.*, 2019).

To explore the impact of Firth's penalization (Firth, 1993; Heinze et Schemper, 2002) 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.

5.2.2.4 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 (5.2)). As in the main simulation study, 1000

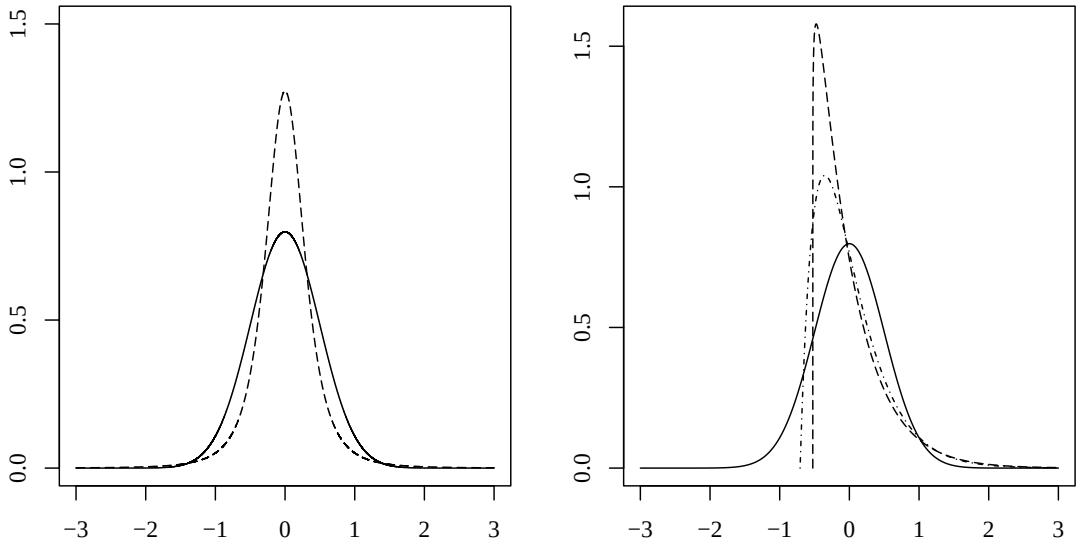


Figure 5.1 Left graph: density function of the generalized t-distribution with $\nu = 3$ degrees of freedom, position parameter $\mu = 0$ and scale parameter $\sigma = \frac{1}{2\sqrt{3}}$ (dashed line). Right graph: density functions of the gamma distributions with $shape = 1.1025$, $scale = 0.4762$ (dashed line; $skewness = 1.9048$) and $shape = 2$, $scale = 0.3636$ (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 $variance = 0.5^2$

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 (5.4) and (5.7)) 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: $\theta_3 = 0.30$.

5.2.2.5 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 5.1) affects the performance of the exact estimators proposed. Two cases were considered.

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 5.1, 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 5.1, 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 $shape = 2$, $scale = 0.3636$, 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 (5.7) by the corresponding true density function (generalized t-distribution or gamma) and using the true (simulation) parameters values.

5.3 Results

5.3.1 Results of the main simulation studies

Tables 5.1 - 5.3 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 5.5 - 5.7 (Appendix 5.6.4). 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 (Lange *et al.*, 2012), while they were very close in the adjusted simulation study (see Tables 5.5 - 5.6 and Tables 5.1 - 5.2). For the RD scale, the exact results were close to those obtained using Imai *et al.* (2010)'s quasi-Bayesian approach.

Compared to the exact OR estimators, the approximate OR estimators by Gaynor *et al.* (2019)

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 Table 5.5 and Table 5.1).

It is noteworthy to emphasize the key role of the parameter s in the Gaynor *et al.* (2019) approach. 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 5.8 the distribution of the s estimates in each crude and adjusted simulation scenario (see Appendix 5.6.5). 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 (1981). An increase in relative biases for *Crude* and *Adjusted scenarios 4*, corresponding to estimated marginal probabilities of $\approx 80\%$, could be explained by the fact that the observed s estimates were closer to the minimax solution (Camilli, 1994), which performs better at the middle abscissas (i.e., argmax) of the logistic density (Savalei, 2006). Therefore, the procedure to estimate s proposed by Gaynor *et al.* (2019) can produce suboptimal values for this parameter.

The approximate OR estimators by VanderWeele and Vansteelandt (2010) demonstrated small relative bias values (between 0.66% and 1.74%) for *Crude* and *Adjusted scenarios 1*, corresponding to the marginally rare outcome (see Table 5.5 and Table 5.1). 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.

5.3.2 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 5.9 (see Appendix 5.6.6). 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 (Lange *et al.*, 2012) for the multiplicative scales and were almost identical to those returned by Imai *et al.* (2010)'s quasi-Bayesian approach for the RD scale.

Compared to the approximate OR estimators by Gaynor *et al.* (2019), 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 (2010) 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.

Table 5.1 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	-

Table 5.1 Adjusted simulation study: odds ratio scale (1000 samples of size $n = 5000$; continuation)

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 (Casella et Berger, 2002) for exact and VanderWeele and Vansteelandt's estimators; robust standard errors based on the sandwich estimator (Liang et Zeger, 1986) for NEMs.

[b] Percentile bootstrap (Chernick, 2011) based on 500 resamples with replacement.

Table 5.2 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 (Casella et Berger, 2002) for exact estimators; robust standard errors based on the sandwich estimator (Liang et Zeger, 1986) for NEMs.

[b] Percentile bootstrap (Chernick, 2011) based on 500 resamples with replacement.

Table 5.3 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 (%)[a]	Bootstrap CP (%)[b]
				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 (Casella et Berger, 2002) for exact estimators; White's heteroskedasticity-consistent estimator for the covariance matrix White (1980) for approach by Imai et al.

[b] Percentile bootstrap (Chernick, 2011) based on 500 resamples with replacement.

5.3.3 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 5.10 - 5.12 (see Appendix 5.6.7).

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 5.10 - 5.11) 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 (e.g., mean NDE OR=72.98, corresponding relative bias=4608.6%). Quasi-separation was not observed in the other analyses.

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.

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

In Table 5.13 of Appendix 5.6.8, 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 (5.2), 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 5.1 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 5.1).

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

We present in Tables 5.14 - 5.15 (see Appendix 5.6.9) 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.

5.4 Real data example

We applied the proposed exact approach to the cohort data studied in our previous works (Samoilenko *et al.*, 2018; Samoilenko et Lefebvre, 2021). 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 (i.e., the observed prevalence of low birth weight was 7.7%). Both unadjusted (crude) and adjusted analyses were performed; in the latter, we examined the same set of adjustment variables as in Samoilenko and Lefebvre (2021): 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 Appendix 5.6.11) 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 (2010) and Gaynor *et al.* (2019).

Natural effects were also estimated using NEMs (Lange *et al.*, 2012; Steen *et al.*, 2017). 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 (Chernick, 2011) based on 5000 resamples with replacement to calculate 95% CIs.

The results are presented in Table 5.4. In both crude and adjusted analyses, 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.* (2019). Interestingly, the results obtained using Gaynor *et al.* (2019) 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.* (2019)'s approach 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 (Savalei, 2006). 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 versus 9). Lastly, results from the approximate approach by VanderWeele and Vansteelandt (2010) generally stood apart from the results returned by the other approaches in both crude and adjusted analyses.

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 birth) instead of a continuous variable (see our previous analysis in Samoilenko and Lefebvre (2021)). This is a reassuring finding since, conceptually, the total effect should not be affected by the mediator and its type.

5.5 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.* (2021) for estimating natural direct and indirect effects on the (log) OR scale. As in Cheng *et al.* (2021), 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

Table 5.4 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 analyses				
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 analyses				
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 (Chernick, 2011) based on 5000 resamples with replacement.

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 5.2). 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 Appendix 5.6.11).

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 Vanstelandt (2010) 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.* (2010) and the NEM approach by Lange *et al.* (2012). These benchmark approaches also do not rely on the outcome rareness or commonness. The R package `mediation` (Tingley *et al.*, 2014), which implements the approach by Imai et al. examined (and its non-parametric bootstrap version; Imai *et al.* (2010)), 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` (Steen *et al.*, 2017), implementing the NEM approach by Lange *et al.* (2012), which requires different NEM specifications in order to obtain natural effects estimates on these scales (e.g., logistic model for OR, log-binomial or Poisson model for RR).

As Tchetgen Tchetgen (2014) pointed out, the modeling assumption encoded in Equation (5.1) 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 (e.g., 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 (5.2) for reference) is reasonably satisfied.

Our exact approach is conceptually straightforward and a logical choice when both the mediator and outcome models (5.1), (5.2) 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` (Steen *et al.*, 2017) may be more adequate since it does not require modelling 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, (Bartlett et Hughes, 2020; Steen *et al.*, 2017) which is prone to occur with nonlinear outcome models (Steen *et al.*, 2017). While the approach by Imai *et al.* (2010) used the same logistic and linear models as the exact approach in our simulations, the R package `mediation` (Tingley *et al.*, 2014) – which implements this approach – allows going beyond standard regression models for the mediator and outcome models (e.g., 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 analyses of small and/or sparse data because of complete separation or quasi-complete separation (Allison, 2012; Heinze, 2006). 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 (Allison, 2012; Heinze, 2006; Mansournia *et al.*, 2018), some authors have mentioned that this penalization can introduce bias in both average and individual predicted probabilities (Puhr *et al.*, 2017) (of note, the expit function in the integrand of Equation (5.3) 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.* (2021) 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.* (2021) advised to avoid adjusting for such predictors. We recommend for applying Diop *et al.* (2021) 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 (e.g., outcome-adaptive LASSO; Ye *et al.* (2021)) could be used to select covariates to control for confounding.

The SAS macro `bin_cont_exactmed` was firstly 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 (Naimi *et al.*, 2014; Valeri et VanderWeele, 2013). 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 (2021). As the VanderWeele and Vansteelandt (2010), Gaynor *et al.* (2019), and Cheng *et al.* (2021) 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 minutes were required to obtain results on all three binary scales considered with a machine having a CPU speed of 3.40GHz 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 (i.e., 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 paper, it will be worthwhile to extend it to case-control designs in order to increase its practical applicability.

Supporting Information

A SAS macro `bin_cont_exactmed` may be found online in the Supporting Information section (see `https://onlinelibrary.wiley.com/doi/full/10.1002/sim.9621`) and in Appendix 5.6.11 at the end of this article.

Author contributions

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

Acknowledgments

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.

5.6 Appendices

5.6.1 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* (Lange *et al.*, 2017; VanderWeele et Vansteelandt,

2009). 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 (Lange *et al.*, 2017; Andrews et Didelez, 2021).

5.6.2 Delta method for exact mediation approach

Let us note:

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

In Equation (5.7), $g(a, a^*, \boldsymbol{c})$ is a function of the vector

$$(\boldsymbol{\beta}, \boldsymbol{\theta}, \sigma^2) = (\beta_0, \beta_1, \beta_{21}, \beta_{22}, \ldots, \beta_{2k}, \theta_0, \theta_1, \theta_2, \theta_3, \theta_{41}, \theta_{42}, \ldots, \theta_{4k}, \sigma^2)^{'}; \tag{5.10}$$

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

Thus

$$\begin{aligned}
\frac{\partial g(a, a^*, \boldsymbol{c})}{\partial \beta_0} &= \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} \operatorname{expit}\left(\theta_0 + \theta_1 a + \theta_2 m + \theta_3 a m + \boldsymbol{\theta}_4^{'}\boldsymbol{c}\right) \exp\left(-\frac{\left(m - \left(\beta_0 + \beta_1 a^* + \boldsymbol{\beta}_2^{'}\boldsymbol{c}\right)\right)^2}{2\sigma^2}\right) \\
&\quad \times \frac{m - (\beta_0 + \beta_1 a^* + \boldsymbol{\beta}_2^{'}\boldsymbol{c})}{\sigma^2} dm \\
&= \frac{1}{\sigma^3\sqrt{2\pi}} \int_{-\infty}^{\infty} m \operatorname{expit}\left(\theta_0 + \theta_1 a + \theta_2 m + \theta_3 a m + \boldsymbol{\theta}_4^{'}\boldsymbol{c}\right) \\
&\quad \times \exp\left(-\frac{\left(m - \left(\beta_0 + \beta_1 a^* + \boldsymbol{\beta}_2^{'}\boldsymbol{c}\right)\right)^2}{2\sigma^2}\right) dm - \frac{\beta_0 + \beta_1 a^* + \boldsymbol{\beta}_2^{'}\boldsymbol{c}}{\sigma^2}\, g(a, a^*, \boldsymbol{c}),
\end{aligned} \tag{5.11}$$

$$\frac{\partial g(a,a^*,\boldsymbol{c})}{\partial \beta_1} = a^* \, \frac{\partial g(a,a^*,\boldsymbol{c})}{\partial \beta_0}, \tag{5.12}$$

$$\frac{\partial g(a,a^*,\boldsymbol{c})}{\partial \beta_{2i}} = c_i \, \frac{\partial}{\partial \beta_0} g(a,a^*,\boldsymbol{c}), \quad i = 1,2,\ldots,k, \tag{5.13}$$

$$\begin{aligned}\frac{\partial g(a,a^*,\boldsymbol{c})}{\partial \theta_0} = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} & \frac{\exp\left(\theta_0 + \theta_1 a + \theta_2 m + \theta_3 am + \boldsymbol{\theta}_4^{'}\boldsymbol{c}\right)}{\left(1 + \exp\left(\theta_0 + \theta_1 a + \theta_2 m + \theta_3 am + \boldsymbol{\theta}_4^{'}\boldsymbol{c}\right)\right)^2} \\ & \times \exp\left(-\frac{\left(m - \left(\beta_0 + \beta_1 a^* + \boldsymbol{\beta}_2^{'}\boldsymbol{c}\right)\right)^2}{2\sigma^2}\right) dm,\end{aligned} \tag{5.14}$$

$$\frac{\partial g(a,a^*,\boldsymbol{c})}{\partial \theta_1} = a \, \frac{\partial g(a,a^*,\boldsymbol{c})}{\partial \theta_0}, \tag{5.15}$$

$$\begin{aligned}&\frac{\partial g(a,a^*,\boldsymbol{c})}{\partial \theta_2} \\ &= \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} m \, \frac{\exp\left(\theta_0 + \theta_1 a + \theta_2 m + \theta_3 am + \boldsymbol{\theta}_4^{'}\boldsymbol{c}\right)}{\left(1 + \exp\left(\theta_0 + \theta_1 a + \theta_2 m + \theta_3 am + \boldsymbol{\theta}_4^{'}\boldsymbol{c}\right)\right)^2} \exp\left(-\frac{\left(m - \left(\beta_0 + \beta_1 a^* + \boldsymbol{\beta}_2^{'}\boldsymbol{c}\right)\right)^2}{2\sigma^2}\right) dm,\end{aligned} \tag{5.16}$$

$$\frac{\partial g(a,a^*,\boldsymbol{c})}{\partial \theta_3} = a \, \frac{\partial g(a,a^*,\boldsymbol{c})}{\partial \theta_2}, \tag{5.17}$$

$$\frac{\partial g(a,a^*,\boldsymbol{c})}{\partial \theta_{4i}} = c_i \, \frac{\partial g(a,a^*,\boldsymbol{c})}{\partial \theta_0}, \quad i = 1,2,\ldots,k, \tag{5.18}$$

$$
\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^{'}\boldsymbol{c}\right)\exp\left(-\frac{\left(m-\left(\beta_0+\beta_1 a^*+\boldsymbol{\beta}_2^{'}\boldsymbol{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^{'}\boldsymbol{c}\right)\exp\left(-\frac{\left(m-\left(\beta_0+\beta_1 a^*+\boldsymbol{\beta}_2^{'}\boldsymbol{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^{'}\boldsymbol{c}\right)\exp\left(-\frac{\left(m-(\beta_0+\beta_1 a^*+\boldsymbol{\beta}_2^{'}\boldsymbol{c})\right)^2}{2t}\right) \\
&\quad \times \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^{'}\boldsymbol{c}\right)\exp\left(-\frac{\left(m-\left(\beta_0+\beta_1 a^*+\boldsymbol{\beta}_2^{'}\boldsymbol{c}\right)\right)^2}{2t}\right) \\
&\quad \times \left(m-\left(\beta_0+\beta_1 a^*+\boldsymbol{\beta}_2^{'}\boldsymbol{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^{'}\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) \\
&\quad \times \left(m-\left(\beta_0+\beta_1 a^*+\boldsymbol{\beta_2^{'}c}\right)\right)^2 dm.
\end{aligned}
\tag{5.19}
$$

The gradient of the scalar function $g(a,a^*,\boldsymbol{c})$ with respect to the vector $(\boldsymbol{\beta},\boldsymbol{\theta},\sigma^2)$ (see Equation (5.10)) is

$$
\nabla\left(g(a,a^*,\boldsymbol{c})\right) = \left(\frac{\partial g(a,a^*,\boldsymbol{c})}{\partial \boldsymbol{\beta}}, \frac{\partial g(a,a^*,\boldsymbol{c})}{\partial \boldsymbol{\theta}}, \frac{\partial g(a,a^*,\boldsymbol{c})}{\partial \sigma^2}\right)^{'}, \tag{5.20}
$$

where

$$
\begin{aligned}
&\frac{\partial g(a,a^*,\boldsymbol{c})}{\partial \boldsymbol{\beta}} = \left(\frac{\partial g(a,a^*,\boldsymbol{c})}{\partial \beta_0}, \frac{\partial g(a,a^*,\boldsymbol{c})}{\partial \beta_1}, \frac{\partial g(a,a^*,\boldsymbol{c})}{\partial \beta_{21}}, \frac{\partial g(a,a^*,\boldsymbol{c})}{\partial \beta_{22}}, \ldots, \frac{\partial g(a,a^*,\boldsymbol{c})}{\partial \beta_{2k}}\right), \\
&\frac{\partial g(a,a^*,\boldsymbol{c})}{\partial \boldsymbol{\theta}} \\
&\quad = \left(\frac{\partial g(a,a^*,\boldsymbol{c})}{\partial \theta_0}, \frac{\partial g(a,a^*,\boldsymbol{c})}{\partial \theta_1}, \frac{\partial g(a,a^*,\boldsymbol{c})}{\partial \theta_2}, \frac{\partial g(a,a^*,\boldsymbol{c})}{\partial \theta_3}, \frac{\partial g(a,a^*,\boldsymbol{c})}{\partial \theta_{41}}, \frac{\partial g(a,a^*,\boldsymbol{c})}{\partial \theta_{42}}, \ldots, \frac{\partial g(a,a^*,\boldsymbol{c})}{\partial \theta_{4k}}\right).
\end{aligned}
$$

Let us note

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

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

where $\nabla\left(g(a, a^*, \boldsymbol{c})\right)$ is defined in Equation (5.20).

5.6.2.1 Delta method for exact natural effects odds ratios

We can express the exact $OR^{NDE}_{a,a^*|\boldsymbol{c}}$ and $OR^{NIE}_{a,a^*|\boldsymbol{c}}$ in terms of $g(a, a^*, \boldsymbol{c})$ defined in Equation (5.7) as follows:

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

To construct the 95% confidence intervals (CIs) for $OR^{NDE}_{a,a^*|\boldsymbol{c}}$ and $OR^{NIE}_{a,a^*|\boldsymbol{c}}$ by the first-order multivariate delta method (Casella et Berger, 2002), we expressed standard errors (se) for $\ln\left(\widehat{OR^{NDE}_{a,a^*|\boldsymbol{c}}}\right)$ and $\ln\left(\widehat{OR^{NIE}_{a,a^*|\boldsymbol{c}}}\right)$ according to the following approximate formulas:

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

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

where $\Sigma = diag\left\{\Sigma_{\widehat{\boldsymbol{\beta}}}, \Sigma_{\widehat{\boldsymbol{\theta}}}, \Sigma_{\widehat{\sigma^2}}\right\}$ is a block matrix; $\Sigma_{\widehat{\boldsymbol{\beta}}}$, $\Sigma_{\widehat{\boldsymbol{\theta}}}$, and $\Sigma_{\widehat{\sigma^2}}$ are the covariance matrices for $\widehat{\boldsymbol{\beta}}$, $\widehat{\boldsymbol{\theta}}$, and $\widehat{\sigma^2}$, respectively. To estimate $\Sigma_{\widehat{\sigma^2}} = \mathrm{Var}\left(\widehat{\sigma^2}\right)$, we used the following unbiased estimator proposed by Gao and Luo Gao et Luo (2019):

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

where MSE is the residual mean squared error from the mediator model (Equation (5.1)), $l = n - p$, and n and p are the number of observations and the number of regression coefficients (including intercept) in the mediator model (Equation (5.1)), respectively; see Gao et Luo (2019) for more information on the properties of W_l [1].

1. Notamment, l'estimateur W_l possède la variance minimale parmi tous les estimateurs sans biais de $\mathrm{Var}\left(\widehat{\sigma^2}\right)$.

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

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

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

Thus,

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

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

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

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

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

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

$$\ln\left(\widehat{OR^{TE}_{a,a^*|\boldsymbol{c}}}\right) = \ln\left(\widehat{OR^{NDE}_{a,a^*|\boldsymbol{c}}}\right) + \ln\left(\widehat{OR^{NIE}_{a,a^*|\boldsymbol{c}}}\right),$$

$$\nabla\left(\ln\left(\widehat{OR^{TE}_{a,a^*|\boldsymbol{c}}}\right)\right) = \nabla\left(\ln\left(\widehat{OR^{NDE}_{a,a^*|\boldsymbol{c}}}\right)\right) + \nabla\left(\ln\left(\widehat{OR^{NIE}_{a,a^*|\boldsymbol{c}}}\right)\right),$$

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

Thus,

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

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

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

5.6.2.2 Delta method for exact natural effects risk ratios

For the RR scale, we have that

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

and, correspondingly,

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

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

Thus,

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

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

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

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

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

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

$$\ln\left(\widehat{RR^{TE}_{a,a^*|\boldsymbol{c}}}\right) = \ln\left(\widehat{RR^{NDE}_{a,a^*|\boldsymbol{c}}}\right) + \ln\left(\widehat{RR^{NIE}_{a,a^*|\boldsymbol{c}}}\right),$$

$$\nabla\left(\ln\left(\widehat{RR^{TE}_{a,a^*|\boldsymbol{c}}}\right)\right) = \nabla\left(\ln\left(\widehat{RR^{NDE}_{a,a^*|\boldsymbol{c}}}\right)\right) + \nabla\left(\ln\left(\widehat{RR^{NIE}_{a,a^*|\boldsymbol{c}}}\right)\right),$$

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

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

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

5.6.2.3 Delta method for exact natural effects risk differences

For the RD scale, we have:

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

$$RD^{TE}_{a,a^*|\boldsymbol{c}} = RD^{NDE}_{a,a^*|\boldsymbol{c}} + RD^{NIE}_{a,a^*|\boldsymbol{c}}$$

and, correspondingly,

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

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

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

Thus,

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

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

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

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

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

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

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

5.6.3 Convergence of improper integrals in the exact approach

All improper integrals used to construct the exact point and interval estimators (see Equations (5.7), (5.11 - 5.19)) 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 (Khuri, 2002):

$$0 < \frac{1}{\sqrt{2\pi\sigma^2}} \operatorname{expit}(a+bm) \exp\left(-\frac{(m-c)^2}{2\sigma^2}\right) < \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(m-c)^2}{2\sigma^2}\right),$$

$$0 < \frac{1}{\sqrt{2\pi\sigma^2}} \frac{\exp(a+bm)}{(1+\exp(a+bm))^2} \exp\left(-\frac{(m-c)^2}{2\sigma^2}\right) < \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(m-c)^2}{2\sigma^2}\right),$$

$$\left|\frac{m}{\sqrt{2\pi\sigma^2}} \operatorname{expit}(a+bm) \exp\left(-\frac{(m-c)^2}{2\sigma^2}\right)\right| < \frac{|m|}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(m-c)^2}{2\sigma^2}\right) < \frac{m^2+1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(m-c)^2}{2\sigma^2}\right),$$

$$\left|\frac{m}{\sqrt{2\pi\sigma^2}} \frac{\exp(a+bm)}{(1+\exp(a+bm))^2} \exp\left(-\frac{(m-c)^2}{2\sigma^2}\right)\right| < \frac{|m|}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(m-c)^2}{2\sigma^2}\right)$$

$$< \frac{m^2+1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(m-c)^2}{2\sigma^2}\right),$$

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

5.6.4 Results of the crude simulation study

The results of the crude simulation study are presented in Tables 5.5 - 5.7.

Table 5.5 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 CP (%)[a]	Bootstrap CP (%)[b]
				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	-

Table 5.5 Crude simulation study: odds ratio scale (1000 samples of size $n = 5000$; continuation)

Effect	Estimator	True value	Mean	Bias	Relative bias (%)	SD	RMSE	Delta / robust SE CP (%)[a]	Bootstrap CP (%)[b]
			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 (Casella et Berger, 2002) for exact and VanderWeele and Vansteelandt's estimators; robust standard errors based on the sandwich estimator (Liang et Zeger, 1986) for NEMs.

[b] Percentile bootstrap (Chernick, 2011) based on 500 resamples with replacement.

Table 5.6 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 (Casella et Berger, 2002) for exact estimators; robust standard errors based on the sandwich estimator (Liang et Zeger, 1986) for NEMs.

[b] Percentile bootstrap (Chernick, 2011) based on 500 resamples with replacement.

Table 5.7 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 (Casella et Berger, 2002) for exact estimators; White's heteroskedasticity-consistent estimator for the covariance matrix White (1980) for approach by Imai et al.

[b] Percentile bootstrap (Chernick, 2011) based on 500 resamples with replacement.

5.6.5 Estimation of the parameter s in the approach by Gaynor et al.

Table 5.8 presents the distribution of the parameter s involved in the approximate approach by Gaynor *et al.* (2019) for each crude and adjusted simulation scenario.

Table 5.8 Distribution of the parameter s estimates in the approach by Gaynor *et al.* (2019)

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

Abbreviations: SD, standard deviation.
[a] Values of the scaling parameter s according to the literature: $s \approx 0.551$ by Cox (1970) based on the equality of variances; $s \approx 0.572$ using Kullback-Leibler information criterion (Savalei, 2006); $s \approx 0.588$ from the minimax solution (Camilli, 1994); $s \approx 0.625$ using comparative approach by Amemiya (Amemiya, 1981; Savalei, 2006).

5.6.6 Results of the simulation study with marginally but not conditionally rare outcome

Table 5.9 reports the results of the adjusted simulation study with a marginally but not conditionally rare outcome.

5.6.7 Results of the simulation study with Firth's penalization

Tables 5.10 - 5.12 present the results of the simulation study performed to examine the impact of Firth's penalization on the exact estimators.

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

Table 5.13 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.

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

Tables 5.14 - 5.15 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 5.9 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 (Casella et Berger, 2002) for exact and VanderWeele and Vansteelandt's estimators; percentile bootstrap (Chernick, 2011) based on 500 resamples with replacement for approach by Gaynor et al.; robust standard errors based on the sandwich estimator (Liang et Zeger, 1986) for NEMs; White's heteroskedasticity-consistent estimator for the covariance matrix (White, 1980) for approach by Imai et al.

Table 5.10 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 5.11 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 5.12 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.

Table 5.13 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 (Chernick, 2011) based on 500 resamples with replacement.

Table 5.14 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 (Chernick, 2011) based on 500 resamples with replacement.

Table 5.15 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 (Chernick, 2011) based on 500 resamples with replacement.

5.6.10 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 Appendix 5.6.11) 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 $\texttt{t}_0$ to level $\texttt{t}_1$ 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 (i.e., 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`, `mcontrol=`$\texttt{m}_c$, our SAS macro `bin_cont_exactmed` additionally returns the controlled direct effect estimated at the level $\texttt{m}_c$ 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 $\texttt{Mvar}_1$=$\texttt{Cm}_1$, $\texttt{Mvar}_2$=$\texttt{Cm}_2$, $\texttt{Mvar}_3$=$\texttt{Cm}_3$ (i.e., at user-defined values for the first three adjustment covariates in the mediator model), and

Yvar$_3$=Cy$_3$, Yvar$_4$=Cy$_4$ (i.e., 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 $\mathtt{V}_{cat}$, `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 (e.g., when the integrand values in Equation (5.7) and/or Equations (5.11-5.19) are close to zero on $(-\infty, \infty)$, possibly except on some small intervals). To overcome these problems, the user can change the default value (that is, 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 $\mathtt{t}_0$ to $\mathtt{t}_1$, 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 (5.7) 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 5.11-5.19), 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, etc.

5.6.11 SAS macro bin_cont_exactmed

```
/*********************************************************************************************/
/* The user needs to specify the values for the following macro variables                    */
/* in the macro %bin_cont_exactmed:                                                          */
/*                                                                                           */
/* mydata:      input data that include the outcome, exposure and mediator variables as      */
/*              well as the covariates to be adjusted for in the model;                      */
/* A:           the name of the exposure variable (binary or continuous);                    */
/* M:           the name of the mediator variable (continuous);                              */
/* Y:           the name of the outcome variable (binary);                                   */
/* interaction: the user needs to specify INTERACTION=0 or INTERACTION=1 for the outcome     */
/*              model without or with interaction between the exposure and the mediator,     */
/*              respectively;                                                                */
/* adjusted:    the user needs to specify ADJUSTED=0 or ADJUSTED=1 to obtain                 */
/*              unadjusted or adjusted NIE, NDE and TE, respectively;                        */
/* cvar_M:      the list of adjustment variables (covariates) in the mediator model;         */
/*              categorical variables need to be coded as a series of dummy variables       */
/*              before being entered as covariates; use space to separate covariates'       */
/*              names;                                                                       */
/* cvar_Y:      the list of adjustment variables (covariates) in the outcome model;          */
/*              categorical variables need to be coded as a series of dummy variables       */
/*              before being entered as covariates;                                          */
/* a0:          the exposure level corresponding to a*;                                      */
/* a1:          the exposure level corresponding to a;                                       */
/* boot:        the user needs to specify BOOT=1 to obtain 95% confidence intervals by       */
/*              percentile bootstrap (in addition to the default 95% confidence              */
/*              intervals based on the delta method);                                        */
/* bootseed:    if BOOT=1, that is bootstrap 95% confidence intervals are required,          */
/*              then the user needs to specify the initial seed (positive integer) for       */
/*              random number generation;                                                    */
/* nboot:       if BOOT=1, that is bootstrap 95% confidence intervals are required, then     */
/*              the user needs to specify the number of bootstrap resamples;                 */
/* Firth:       the user needs to specify FIRTH=1 in order to use Firth's  penalization      */
/*              in the outcome logistic regression model; if FIRTH=0, conventional           */
/*              maximum likelihood estimates are returned for the outcome logistic           */
/*              regression model;                                                            */
/* pscale:      the user needs to specify PSCALE=1 in order to replace the default value     */
/*              (that is, 1) of the scale parameter in the QUAD subroutine to calculate      */
/*              the nested counterfactual outcome probabilities;                             */
/* pscalevalue: if PSCALE=1, the user needs to provide a value for the scale parameter       */
/*              in the QUAD subroutine to calculate the nested counterfactual outcome        */
/*              probabilities;                                                               */
/* dscale:      the user needs to specify DSCALE=1 in order to replace the default value     */
/*              (that is, 1) of the scale parameter in the QUAD subroutine to calculate      */
/*              the integrals needed for the delta method;                                   */
/* dscalevalue: if DSCALE=1, the user needs to provide a value for the scale parameter       */
/*              in the QUAD subroutine to calculate the integrals needed for the delta       */
/*              method;                                                                      */
/* stratum:     the user needs to specify STRATUM=1 to estimate mediation effects at         */
/*              specific values of some covariates (that is, stratum-specific effects);      */
/* cvar_M_data: if STRATUM=1, the user needs to provide a SAS dataset with a single row      */
/*              that contains specific values for some or all of the adjustment              */
/*              covariates cvar_M in the mediator model; if the covariates specified in      */
/*              cvar_M_data constitute some proper (strict) subset of cvar_M, then the       */
/*              other covariates will be set to their sample-specific mean levels;           */
/* cvar_Y_data: if STRATUM=1, the user needs to provide a SAS dataset with a single row      */
/*              that contains specific values for some or all of the adjustment covariates   */
/*              cvar_Y in the outcome model; if the covariates specified in cvar_Y_data      */
/*              constitute some proper (strict) subset of cvar_Y, then the other             */
```

```
/*              covariates will be set to their sample-specific mean levels;               */
/* cde:         the user needs to specify CDE=1 to estimate controlled direct effects at   */
/*              some specific level of the mediator;                                       */
/* mcontrol:    if CDE=1, the user needs to provide the level of the mediator used  to     */
/*              estimate controlled direct effects.                                        */
/*                                                                                         */
/*******************************************************************************************/


%macro bin_cont_exactmed(mydata,exposure,mediator,outcome,a1,a0,adjusted,cvar_M,cvar_Y,
interaction,boot,bootseed,nboot,Firth,pscale,pscalevalue,dscale,dscalevalue,
stratum,cvar_M_data,cvar_Y_data,cde,mcontrol);

data mydata; set &mydata; keep &exposure &mediator &outcome &cvar_M &cvar_Y; run;

proc sql noprint; select count(*) into :nobs
from mydata; quit;
%put Count=&nobs.;

%if &adjusted = 1 %then %do;

  proc means data=mydata mean noprint; var &cvar_M;
  output out=cvar_values_M (where=(_STAT_="MEAN")); run;
  proc means data=mydata mean noprint; var &cvar_Y;
  output out=cvar_values_Y (where=(_STAT_="MEAN")); run;

    %if &stratum=1 %then %do;
      proc datasets lib=work nolist;
      delete cvar_names_M cvar_names_Y; quit; run;

      %if %length(&cvar_M_data) = 0 %then %do;
      %put ERROR: if STRATUM=1, then cvar_M_data and cvar_Y_data must be specified;
      %abort; %end;

      %if %length(&cvar_Y_data) = 0 %then %do;
      %put ERROR: if STRATUM=1, then cvar_M_data and cvar_Y_data must be specified;
      %abort; %end;

      %if %sysfunc(exist(&cvar_M_data)) %then %do;
      proc contents data=&cvar_M_data noprint out=cvar_names_M (keep=name); run;
      %end; %else %do;
      %put ERROR: if STRATUM=1, then cvar_M_data must be provided;
      %abort; %end;

      %if %sysfunc(exist(&cvar_Y_data)) %then %do;
      proc contents data=&cvar_Y_data noprint out=cvar_names_Y (keep=name); run;
      %end; %else %do;
      %put ERROR: if STRATUM=1, then cvar_Y_data must be provided;
      %abort; %end;

      proc sort data=cvar_names_M; by name; run;
      proc sort data=cvar_names_Y; by name; run;

      data common_vars;  merge cvar_names_M (in=a) cvar_names_Y (in=b); by name;
      if a and b then output; run;

      proc sql noprint; select count(*) into :count1 from common_vars; quit;

      %if &count1 ne 0 %then %do;
```

```
      ods exclude CompareDatasets CompareDifferences;
      proc compare base=&cvar_M_data compare=&cvar_Y_data
      nosummary out=check  outnoequal; run;
      proc sql noprint; select count(*) into :count2 from check; quit;
      %if &count2 ne 0 %then %do;
      %put WARNING: Some common variables  &cvar_M_data and &cvar_Y_data
                    do not have the same values (are unequal);
      %put WARNING: See RESULTS for details; %abort; %end;
      %end;

     data cvar_values_M; merge cvar_values_M &cvar_M_data; run;
     data cvar_values_Y; merge cvar_values_Y &cvar_Y_data; run;

  %end;

  proc reg data=mydata outest=coeffs_mediator (keep=_RMSE_ Intercept &exposure &cvar_M) noprint;
  model &mediator = &exposure &cvar_M; run; quit;
  proc reg data=mydata covout outest=sigma_mediator (drop=_: &mediator) noprint;
  model &mediator = &exposure &cvar_M; run; quit;

  %if &Firth=1 %then %do;

    %if &interaction=0 %then %do;
    proc logistic data= mydata descending outest = coeffs_outcome (drop=_:) noprint;
    model &outcome = &exposure &mediator &cvar_Y/Firth; run;
    proc logistic data= mydata descending covout outest = sigma_outcome (drop=_:) noprint;
    model &outcome = &exposure &mediator &cvar_Y/Firth; run;
    %end;

    %if &interaction=1 %then %do;
    proc logistic data= mydata descending outest = coeffs_outcome (drop=_:) noprint;
    model &outcome = &exposure|&mediator &cvar_Y/Firth; run;
    proc logistic data= mydata descending covout outest = sigma_outcome (drop=_:) noprint;
    model &outcome = &exposure|&mediator &cvar_Y/Firth; run;
    %end;

  %end; %else %do;

    %if &interaction=0 %then %do;
    proc logistic data= mydata descending outest = coeffs_outcome (drop=_:)  noprint;
    model &outcome = &exposure &mediator &cvar_Y; run;
    proc logistic data= mydata descending covout outest = sigma_outcome (drop=_:) noprint;
    model &outcome = &exposure &mediator &cvar_Y; run;
    %end;

    %if &interaction=1 %then %do;
    proc logistic data= mydata descending outest = coeffs_outcome (drop=_:) noprint;
    model &outcome = &exposure|&mediator &cvar_Y; run;
    proc logistic data= mydata descending covout outest = sigma_outcome (drop=_:) noprint;
    model &outcome = &exposure|&mediator &cvar_Y; run;
    %end;

  %end;

%end;

%if &adjusted = 0 %then %do;

  proc reg data=mydata outest=coeffs_mediator(keep=_RMSE_ Intercept &exposure) noprint;
```

```
  model &mediator = &exposure; run; quit;
  proc reg data=mydata covout outest=sigma_mediator(drop=_: &mediator) noprint;
  model &mediator = &exposure; run; quit;

  %if &Firth=1 %then %do;

    %if &interaction=0 %then %do;
    proc logistic data= mydata descending outest = coeffs_outcome (drop=_:) noprint;
    model &outcome = &exposure &mediator/Firth; run;
    proc logistic data= mydata descending covout outest = sigma_outcome (drop=_:) noprint;
    model &outcome = &exposure &mediator/Firth; run;
    %end;

    %if &interaction=1 %then %do;
    proc logistic data= mydata descending outest = coeffs_outcome (drop=_:) noprint;
    model &outcome = &exposure|&mediator/Firth; run;
    proc logistic data= mydata descending covout outest = sigma_outcome (drop=_:) noprint;
    model &outcome = &exposure|&mediator/Firth; run;
    %end;

  %end; %else %do;

    %if &interaction=0 %then %do;
    proc logistic data= mydata descending outest = coeffs_outcome (drop=_:) noprint;
    model &outcome = &exposure &mediator; run;
    proc logistic data= mydata descending covout outest = sigma_outcome (drop=_:) noprint;
    model &outcome = &exposure &mediator; run;
     %end;

    %if &interaction=1 %then %do;
    proc logistic data= mydata descending outest = coeffs_outcome (drop=_:) noprint;
    model &outcome = &exposure|&mediator; run;
    proc logistic data= mydata descending covout outest = sigma_outcome (drop=_:) noprint;
    model &outcome = &exposure|&mediator; run;
    %end;

  %end;

%end;

data sigma_mediator; set sigma_mediator; if _n_=1 then delete; run;
data sigma_outcome;  set sigma_outcome;  if _n_=1 then delete; run;

%macro nest_probs(a1,a0);
B=&beta_0+&beta_1*&a0+&product_betas_c; call symputx("B", B);
/* P: nested couterfactual probs */
start fun(m);
v_0=exp(&theta_0+&theta_1*&a1+&theta_2*m+&theta_3*m*&a1+&product_thetas_c);
v_1=v_0/(1+v_0);
v_2=exp(-(m-&B)**2/(2*&sigma**2)); v =v_1*v_2; return(v); finish;
%if &pscale=1 %then %do;
call quad(P_0,"fun",{.M .P}) peak=&B scale=&pscalevalue;
%end;
%else %do; call quad(P_0,"fun",{.M .P}) peak=&B; %end;
P = P_0/(sigma*sqrt(2*pi));
%mend  nest_probs;

%macro derivs(a1,a0);
```

```
B=&beta_0+&beta_1*&a0+&product_betas_c; call symputx("B", B);
P_right = &B+0.5*&sigma**2; call symputx("P_right", P_right);
P_left = &B-0.5*&sigma**2; call symputx("P_left",  P_left);

C = 1/((&sigma**3)*sqrt(2*pi));
D = &B/(&sigma**2);
F = 1/(&sigma*sqrt(2*pi));
E = 1/(2*&sigma**5*sqrt(2*pi));

/* P: nested couterfactual probs */
start fun(m);
v_0=exp(&theta_0+&theta_1*&a1+&theta_2*m+&theta_3*m*&a1+&product_thetas_c);
v_1=v_0/(1+v_0);
v_2=exp(-(m-&B)**2/(2*&sigma**2)); v =v_1*v_2; return(v); finish;
%if &pscale=1 %then %do;
call quad(P_0,"fun",{.M .P}) peak=&B scale=&pscalevalue;
%end;
%else %do; call quad(P_0,"fun",{.M .P}) peak=&B; %end;
P = P_0/(sigma*sqrt(2*pi));

/* derivative of P by beta0 */
start fun(m);
v_0=exp(&theta_0+&theta_1*&a1+&theta_2*m+&theta_3*m*&a1+&product_thetas_c);
v_1=exp(-(m-&B)**2/(2*&sigma**2));
v =m*v_0*v_1/(1+v_0); return(v); finish;
%if &dscale=1 %then %do;
call quad(P_beta0_0,"fun",{.M .P}) peak=&B scale=&dscalevalue;
%end;
%else %do; call quad(P_beta0_0,"fun",{.M .P}) peak=&B; %end;
P_beta0 = C*P_beta0_0 - D*P;

/* derivative of P by beta1 */
P_beta1=P_beta0*&a0;

/* derivative of P by theta0 */
start fun(m);
v_0=exp(&theta_0+&theta_1*&a1+&theta_2*m+&theta_3*m*&a1+&product_thetas_c);
v_1=exp(-(m-&B)**2/(2*&sigma**2));
v =v_0*v_1/((1+v_0)**2); return(v); finish;
%if &dscale=1 %then %do;
call quad(P_theta0_0,"fun",{.M .P}) peak=&B scale=&dscalevalue;
%end;
%else %do; call quad(P_theta0_0,"fun",{.M .P}) peak=&B; %end;
P_theta0 = F*P_theta0_0;

/* derivative of P by theta1 */
P_theta1=P_theta0*&a1;

/* derivative of P by theta2  */
start fun(m);
v_0=exp(&theta_0+&theta_1*&a1+&theta_2*m+&theta_3*m*&a1+&product_thetas_c);
v_1=exp(-(m-&B)**2/(2*&sigma**2));
v =m*v_0*v_1/((1+v_0)**2); return(v); finish;
%if &dscale=1 %then %do;
call quad(P_theta2_0,"fun",{.M .P}) peak=&B scale=&dscalevalue;
%end;
%else %do; call quad(P_theta2_0,"fun",{.M .P}) peak=&B; %end;
P_theta2 = F*P_theta2_0;
```

```
/*  derivative of P by theta3  */
P_theta3=P_theta2*&a1*&interaction;

/* derivative of P by squared_sigma */
start fun(m);
v_0=exp(&theta_0+&theta_1*&a1+&theta_2*m+&theta_3*m*&a1+&product_thetas_c);
v_1=v_0/(1+v_0);
v_2=exp(-(m-&B)**2/(2*&sigma**2));
v_3=(m-&B)**2;
v =v_1*v_2*v_3; return(v); finish;
%if &dscale=1 %then %do;
call quad(P_squared_sigma_left, "fun",{.M &B}) peak=&P_left  scale=&dscalevalue;
call quad(P_squared_sigma_right,"fun",{&B .P}) peak=&P_right scale=&dscalevalue;
%end;
%else %do;
call quad(P_squared_sigma_left,"fun",{ .M &B}) peak=&P_left;
call quad(P_squared_sigma_right,"fun",{&B .P}) peak=&P_right;
%end;
P_squared_sigma_0 = P_squared_sigma_left + P_squared_sigma_right;
P_squared_sigma = -1/(2*&sigma**2)*P+E*P_squared_sigma_0;

/* Gradient: crude analysis */
%if &adjusted=0 %then %do;
P_all_derivs
= P_theta0||P_theta1||P_theta2||P_theta3||P_beta0||P_beta1||P_squared_sigma;
%end;
/* Gradient: adjusted analysis */
%if &adjusted=1 %then %do;
P_beta2=P_beta0*cov_values_M; P_theta4=P_theta0*cov_values_Y;
P_all_derivs
= P_theta0||P_theta1||P_theta2||P_theta3||P_theta4||P_beta0||P_beta1||P_beta2||P_squared_sigma;
%end;
%mend  derivs;

proc iml;

pi = constant("pi");
use coeffs_mediator;
read var {Intercept} into beta_0; call symputx("beta_0", beta_0);
read var {&exposure} into beta_1; call symputx("beta_1", beta_1);
read var {_RMSE_}    into sigma;  call symputx("sigma", sigma);
use coeffs_outcome;
read var {Intercept} into theta_0; call symputx("theta_0", theta_0);
read var {&exposure} into theta_1; call symputx("theta_1", theta_1);
read var {&mediator} into theta_2; call symputx("theta_2", theta_2);

%if &interaction=1  %then %do;
read var {&&exposure.&mediator} into theta_3;
%end;
%if &interaction=0  %then %do;  theta_3=0; %end;

call symputx("theta_3", theta_3);

%if &adjusted=1 %then %do;
use cvar_values_M;   read var {&cvar_M} into cov_values_M;
use cvar_values_Y;   read var {&cvar_Y} into cov_values_Y;
use coeffs_mediator; read var {&cvar_M} into cov_coeffs_M;
use coeffs_outcome;  read var {&cvar_Y} into cov_coeffs_Y;
product_betas_c = cov_coeffs_M*t(cov_values_M);
```

```
product_thetas_c = cov_coeffs_Y*t(cov_values_Y);
%end;
%if &adjusted=0 %then %do; product_betas_c=0; product_thetas_c=0; %end;

call symputx("product_betas_c",  product_betas_c);
call symputx("product_thetas_c", product_thetas_c);

squared_sigma=sigma**2;  call symputx("squared_sigma", squared_sigma);

%nest_probs(a1=&a1,a0=&a1); P11=P;
%nest_probs(a1=&a1,a0=&a0); P10=P;
%nest_probs(a1=&a0,a0=&a0); P00=P;

NIE_OR=P11/(1-P11)*(1-P10)/P10;
NDE_OR=P10/(1-P10)*(1-P00)/P00;
TE_OR=NIE_OR*NDE_OR;

NIE_RR=P11/P10;
NDE_RR=P10/P00;
TE_RR=NIE_RR*NDE_RR;

NIE_RD=P11-P10;
NDE_RD=P10-P00;
TE_RD=NIE_RD+NDE_RD;

q = quantile("NORMAL",.975);
use sigma_mediator; read all  into sigma_mediator;
use sigma_outcome;  read all  into sigma_outcome;
sigma_for_sigma_est = 2*(sigma**4)/(&nobs-1); /*p=3, n-p+2=n-3+2=n-1*/
big_sigma = block(sigma_outcome,sigma_mediator,sigma_for_sigma_est);
%if &interaction=0 %then %do;
d=dimension(big_sigma); z = j(1,d[1],0);
tmp = insert(big_sigma,z,4,0); z = j(1,d[1]+1,0);
big_sigma=insert(tmp,t(z),0,4);
%end;

%derivs(a1=&a1,a0=&a1); Gamma_P11=P_all_derivs;
%derivs(a1=&a1,a0=&a0); Gamma_P10=P_all_derivs;
%derivs(a1=&a0,a0=&a0); Gamma_P00=P_all_derivs;

Gamma_log_NDE = Gamma_P10/(P10*(1-P10)) - Gamma_P00/(P00*(1-P00));
Gamma_log_NIE = Gamma_P11/(P11*(1-P11)) - Gamma_P10/(P10*(1-P10));
Gamma_log_TE =  Gamma_log_NIE +  Gamma_log_NDE;

se_log_NDE = sqrt(Gamma_log_NDE*big_sigma*t(Gamma_log_NDE));
se_log_NIE = sqrt(Gamma_log_NIE*big_sigma*t(Gamma_log_NIE));
se_log_TE =  sqrt(Gamma_log_TE*big_sigma*t(Gamma_log_TE));

NDE_low_OR = NDE_OR*exp(-q*se_log_NDE);
NDE_upp_OR = NDE_OR*exp(q*se_log_NDE);
NIE_low_OR = NIE_OR*exp(-q*se_log_NIE);
NIE_upp_OR = NIE_OR*exp(q*se_log_NIE);
TE_low_OR  = TE_OR*exp(-q*se_log_TE);
TE_upp_OR =  TE_OR*exp(q*se_log_TE);

NDE_CI_OR = NDE_OR || NDE_low_OR || NDE_upp_OR;
NIE_CI_OR = NIE_OR || NIE_low_OR || NIE_upp_OR;
TE_CI_OR  = TE_OR  || TE_low_OR  || TE_upp_OR;
delta_CI_OR  =  NDE_CI_OR // NIE_CI_OR // TE_CI_OR;
```

```
create results_OR
from delta_CI_OR[colname={"OR" "delta_CI_OR_low" "delta_CI_OR_upp"}];
append from delta_CI_OR; close results_OR;


Gamma_log_NDE = Gamma_P10/P10 - Gamma_P00/P00;
Gamma_log_NIE = Gamma_P11/P11 - Gamma_P10/P10;
Gamma_log_TE =  Gamma_log_NIE + Gamma_log_NDE;

se_log_NDE = sqrt(Gamma_log_NDE*big_sigma*t(Gamma_log_NDE));
se_log_NIE = sqrt(Gamma_log_NIE*big_sigma*t(Gamma_log_NIE));
se_log_TE =  sqrt(Gamma_log_TE*big_sigma*t(Gamma_log_TE));

NDE_low_RR = NDE_RR*exp(-q*se_log_NDE);
NDE_upp_RR = NDE_RR*exp(q*se_log_NDE);
NIE_low_RR = NIE_RR*exp(-q*se_log_NIE);
NIE_upp_RR = NIE_RR*exp(q*se_log_NIE);
TE_low_RR  = TE_RR*exp(-q*se_log_TE);
TE_upp_RR =  TE_RR*exp(q*se_log_TE);

NDE_CI_RR = NDE_RR || NDE_low_RR || NDE_upp_RR;
NIE_CI_RR = NIE_RR || NIE_low_RR || NIE_upp_RR;
TE_CI_RR  = TE_RR  || TE_low_RR  || TE_upp_RR;
delta_CI_RR  =  NDE_CI_RR // NIE_CI_RR // TE_CI_RR;
create results_RR
from delta_CI_RR[colname={"RR" "delta_CI_RR_low" "delta_CI_RR_upp"}];
append from delta_CI_RR; close results_RR;

Gamma_NDE = Gamma_P10 - Gamma_P00;
Gamma_NIE = Gamma_P11 - Gamma_P10;
Gamma_TE =  Gamma_NDE + Gamma_NIE;

se_NDE = sqrt(Gamma_NDE*big_sigma*t(Gamma_NDE));
se_NIE = sqrt(Gamma_NIE*big_sigma*t(Gamma_NIE));
se_TE =  sqrt(Gamma_TE*big_sigma*t(Gamma_TE));

NDE_low_RD = NDE_RD-q*se_NDE;  NDE_upp_RD = NDE_RD+q*se_NDE;
NIE_low_RD = NIE_RD-q*se_NIE;  NIE_upp_RD = NIE_RD+q*se_NIE;
TE_low_RD  = TE_RD -q*se_TE;   TE_upp_RD =  TE_RD+q*se_TE;

NDE_CI_RD = NDE_RD || NDE_low_RD || NDE_upp_RD;
NIE_CI_RD = NIE_RD || NIE_low_RD || NIE_upp_RD;
TE_CI_RD  = TE_RD  || TE_low_RD  || TE_upp_RD;
delta_CI_RD  =  NDE_CI_RD // NIE_CI_RD // TE_CI_RD;
create results_RD
from delta_CI_RD[colname={"RD" "delta_CI_RD_low" "delta_CI_RD_upp"}];
append from delta_CI_RD; close results_RD;

%if &cde=1 %then %do;

  sigma_CDE = sigma_outcome;
  %if &interaction=0 %then %do;
  d=dimension(sigma_CDE); z = j(1,d[1],0);
  tmp = insert(sigma_CDE,z,4,0); z = j(1,d[1]+1,0);
  sigma_CDE=insert(tmp,t(z),0,4);
  %end;

  L_1 = exp(theta_0+theta_1*&a1+(theta_2+theta_3*&a1)*&mcontrol+product_thetas_c);
  L_0 = exp(theta_0+theta_1*&a0+(theta_2+theta_3*&a0)*&mcontrol+product_thetas_c);
```

```
CDE_OR = exp((theta_1+theta_3*&mcontrol)*(&a1-&a0));
CDE_RR = CDE_OR*(1+L_0)/(1+L_1);
CDE_RD = L_1/(1+L_1)-L_0/(1+L_0);

/* CDE_OR: DERIVATIVES for GRADIENT and delta CI*/

dOR_dtheta0=0;
dOR_dtheta1=(&a1-&a0);
dOR_dtheta2=0;
dOR_dtheta3=(&a1-&a0)*&interaction*&mcontrol;

%if &adjusted=1 %then %do;
dOR_dtheta4=0*cov_values_Y;
Gamma_CDE_OR=dOR_dtheta0||dOR_dtheta1||dOR_dtheta2||dOR_dtheta3||dOR_dtheta4;
%end;
%if &adjusted=0 %then %do;
Gamma_CDE_OR=dOR_dtheta0||dOR_dtheta1||dOR_dtheta2||dOR_dtheta3;
%end;

se_CDE_OR = sqrt(Gamma_CDE_OR*sigma_CDE*t(Gamma_CDE_OR));
CDE_OR_low = CDE_OR*exp(-q*se_CDE_OR);  CDE_OR_upp = CDE_OR*exp(q*se_CDE_OR);

/* CDE_RR: DERIVATIVES for GRADIENT and delta CI*/

dRR_dtheta0=L_0/(1+L_0)-L_1/(1+L_1);
dRR_dtheta1=&a1/(1+L_1)-&a0/(1+L_0);
dRR_dtheta2=dRR_dtheta0*&mcontrol;
dRR_dtheta3=dRR_dtheta1*&mcontrol*&interaction;

%if &adjusted=1 %then %do;
dRR_dtheta4=dRR_dtheta0*cov_values_Y;
Gamma_CDE_RR=dRR_dtheta0||dRR_dtheta1||dRR_dtheta2||dRR_dtheta3||dRR_dtheta4;
%end;
%if &adjusted=0 %then %do;
Gamma_CDE_RR=dRR_dtheta0||dRR_dtheta1||dRR_dtheta2||dRR_dtheta3;
%end;

se_CDE_RR = sqrt(Gamma_CDE_RR*sigma_CDE*t(Gamma_CDE_RR));
CDE_RR_low = CDE_RR*exp(-q*se_CDE_RR); CDE_RR_upp = CDE_RR*exp(q*se_CDE_RR);

/* CDE_RD: DERIVATIVES for GRADIENT and delta CI*/

dRD_dtheta0=L_1/((1+L_1)**2)-L_0/((1+L_0)**2);
dRD_dtheta1=&a1*L_1/((1+L_1)**2)-&a0*L_0/((1+L_0)**2);
dRD_dtheta2=dRD_dtheta0*&mcontrol;
dRD_dtheta3=dRD_dtheta1*&mcontrol*&interaction;

%if &adjusted=1 %then %do;
dRD_dtheta4=dRD_dtheta0*cov_values_Y;
Gamma_CDE_RD=dRD_dtheta0||dRD_dtheta1||dRD_dtheta2||dRD_dtheta3||dRD_dtheta4;
%end;
%if &adjusted=0 %then %do;
Gamma_CDE_RD=dRD_dtheta0||dRD_dtheta1||dRD_dtheta2||dRD_dtheta3;
%end;

se_CDE_RD = sqrt(Gamma_CDE_RD*sigma_CDE*t(Gamma_CDE_RD));
CDE_RD_low = CDE_RD-q*se_CDE_RD; CDE_RD_upp = CDE_RD+q*se_CDE_RD;
```

```
  /* CDE results*/
  results_CDE_OR = CDE_OR||CDE_OR_low||CDE_OR_upp;
  results_CDE_RR = CDE_RR||CDE_RR_low||CDE_RR_upp;
  results_CDE_RD = CDE_RD||CDE_RD_low||CDE_RD_upp;

  delta_CI_CDE=results_CDE_OR//results_CDE_RR//results_CDE_RD;
  create results_CDE from delta_CI_CDE[colname={"CDE" "delta_CI_CDE_low" "delta_CI_CDE_upp"}];
  append from delta_CI_CDE; close results_CDE;

%end;

quit;

data results_OR; retain effect;
if _n_=1 then effect="NDE"; if _n_=2 then effect="NIE"; if _n_=3 then effect="TE";
set results_OR; run;
data results_RR; retain effect;
if _n_=1 then effect="NDE"; if _n_=2 then effect="NIE"; if _n_=3 then effect="TE";
set results_RR; run;
data results_RD; retain effect;
if _n_=1 then effect="NDE"; if _n_=2 then effect="NIE"; if _n_=3 then effect="TE";
set results_RD; run;

%if &cde=1 %then %do;
data results_CDE; retain scale;
if _n_=1 then scale="CDE OR"; if _n_=2 then scale="CDE RR"; if _n_=3 then scale="CDE RD";
set results_CDE; run;
%end;

%if &boot=1 %then %do;

  proc datasets library = WORK noprint; delete boot_estimates boot_CDE_estimates; run; quit;
  proc surveyselect data= mydata out=bootdata seed=&bootseed
  method=urs samprate=1 outhits rep=&nboot noprint; run;

  options nonotes nosource nosource2 errors=0;

  %if &adjusted = 1 %then %do;

    proc reg data=bootdata outest=coeffs_mediator
    (keep=Replicate _RMSE_ Intercept &exposure &cvar_M)  noprint;
    model &mediator = &exposure &cvar_M; by Replicate; run; quit;

    %if &Firth=1 %then %do;
    %if &interaction=0 %then %do;
    proc logistic data= bootdata descending
    outest = coeffs_outcome (drop=_:)  noprint;
    model &outcome = &exposure &mediator &cvar_Y/Firth; by Replicate; run;
    %end;
    %if &interaction=1 %then %do;
    proc logistic data= bootdata descending outest = coeffs_outcome
    (drop=_:)  noprint;
    model &outcome = &exposure|&mediator &cvar_Y/Firth;
    by Replicate; run;
    %end;
    %end; %else %do;
    %if &interaction=0 %then %do;
    proc logistic data= bootdata descending outest = coeffs_outcome
    (drop=_:)  noprint;
```

```
  model &outcome = &exposure &mediator &cvar_Y; by Replicate; run;
  %end;
  %if &interaction=1 %then %do;
  proc logistic data= bootdata descending outest = coeffs_outcome
  (drop=_:)  noprint;
  model &outcome = &exposure|&mediator &cvar_Y; by Replicate; run;
  %end;
  %end;

%end;

%if &adjusted = 0 %then %do;

  proc reg data=bootdata outest=coeffs_mediator
  (keep=Replicate _RMSE_ Intercept &exposure)  noprint;
  model &mediator = &exposure; by Replicate; run; quit;

  %if &Firth=1 %then %do;
  %if &interaction=0 %then %do;
  proc logistic data= bootdata descending outest = coeffs_outcome
  (drop=_:) noprint;
  model &outcome = &exposure &mediator/Firth; by Replicate; run;
  %end;
  %if &interaction=1 %then %do;
  proc logistic data= bootdata descending outest = coeffs_outcome
  (drop=_:)  noprint;
  model &outcome = &exposure|&mediator/Firth; by Replicate; run;
  %end;
  %end; %else %do;
  %if &interaction=0 %then %do;
  proc logistic data= bootdata descending outest = coeffs_outcome
  (drop=_:) noprint;
  model &outcome = &exposure &mediator; by Replicate; run;
  %end;
  %if &interaction=1 %then %do;
  proc logistic data= bootdata descending outest = coeffs_outcome
  (drop=_:)  noprint;
  model &outcome = &exposure|&mediator; by Replicate; run;
  %end;
  %end;

%end;

%do i=1 %to &nboot; %put iter = &i;

data betas_temp;  set coeffs_mediator;  where Replicate=&i; run;
data thetas_temp; set coeffs_outcome;  where Replicate=&i; run;

proc iml;
pi = constant("pi");
use betas_temp;
read var {Intercept} into beta_0; call symputx("beta_0", beta_0);
read var {&exposure} into beta_1; call symputx("beta_1", beta_1);
read var {_RMSE_}   into sigma;  call symputx("sigma", sigma);
use thetas_temp;
read var {Intercept} into theta_0; call symputx("theta_0", theta_0);
read var {&exposure} into theta_1; call symputx("theta_1", theta_1);
read var {&mediator} into theta_2; call symputx("theta_2", theta_2);
%if &interaction=1  %then %do;
```

```
read var {&&exposure.&mediator} into theta_3;
%end;
%if &interaction=0  %then %do;  theta_3=0; %end;
call symputx("theta_3", theta_3);
%if &adjusted=0 %then %do;
product_betas_c=0; product_thetas_c=0;
%end;
%if &adjusted=1 %then %do;
use betas_temp;    read var {&cvar_M} into cov_coeffs_M;
use thetas_temp;   read var {&cvar_Y} into cov_coeffs_Y;
use cvar_values_M; read var {&cvar_M} into cov_values_M;
use cvar_values_Y; read var {&cvar_Y} into cov_values_Y;
product_betas_c  = cov_coeffs_M*t(cov_values_M);
product_thetas_c = cov_coeffs_Y*t(cov_values_Y);
%end;
call symputx("product_betas_c",  product_betas_c);
call symputx("product_thetas_c", product_thetas_c);
squared_sigma=sigma**2; call symputx("squared_sigma", squared_sigma);

%nest_probs(a1=&a1,a0=&a1); P11=P;
%nest_probs(a1=&a1,a0=&a0); P10=P;
%nest_probs(a1=&a0,a0=&a0); P00=P;

NIE_OR=P11/(1-P11)*(1-P10)/P10;
NDE_OR=P10/(1-P10)*(1-P00)/P00;
TE_OR=P11/(1-P11)*(1-P00)/P00;
point_OR  = NIE_OR || NDE_OR || TE_OR;

NIE_RR=P11/P10;
NDE_RR=P10/P00;
TE_RR=P11/P00;
point_RR  = NIE_RR || NDE_RR || TE_RR;

NIE_RD=P11-P10;
NDE_RD=P10-P00;
TE_RD=P11-P00;
point_RD  = NIE_RD || NDE_RD || TE_RD;

point_estimates = point_OR || point_RR || point_RD;

create point_estimates from
point_estimates[colname
={'NIE_OR' 'NDE_OR' 'TE_OR' 'NIE_RR' 'NDE_RR' 'TE_RR' 'NIE_RD' 'NDE_RD' 'TE_RD'}];
append from point_estimates; close point_estimates;

%if &cde=1 %then %do;
L_1 = exp(theta_0+theta_1*&a1+(theta_2+theta_3*&a1)*&mcontrol+product_thetas_c);
L_0 = exp(theta_0+theta_1*&a0+(theta_2+theta_3*&a0)*&mcontrol+product_thetas_c);
CDE_OR = exp((theta_1+theta_3*&mcontrol)*(&a1-&a0));
CDE_RR = CDE_OR*(1+L_0)/(1+L_1);
CDE_RD = L_1/(1+L_1)-L_0/(1+L_0);
CDE_estimates = CDE_OR || CDE_RR || CDE_RD;
create CDE_estimates from CDE_estimates[colname={'CDE_OR' 'CDE_RR' 'CDE_RD'}];
append from CDE_estimates; close CDE_estimates;
%end;

quit;  /*end proc iml*/

proc append base=boot_estimates  data=point_estimates; run;
```

```
%if &cde=1 %then %do;
proc append base=boot_CDE_estimates data=CDE_estimates;   run;
%end;

%end;

options notes source source2 errors=20;

proc univariate data=boot_estimates noprint; var NIE_OR NDE_OR TE_OR;
output out=boot_CI_OR pctlpre=NIE_OR NDE_OR TE_OR
pctlpts=2.5 97.5  pctlname= _low _upp; run;
proc univariate data=boot_estimates noprint; var NIE_RR NDE_RR TE_RR;
output out=boot_CI_RR pctlpre=NIE_RR NDE_RR TE_RR pctlpts=2.5 97.5
pctlname= _low _upp; run;
proc univariate data=boot_estimates noprint; var NIE_RD NDE_RD TE_RD;
output out=boot_CI_RD pctlpre=NIE_RD NDE_RD TE_RD pctlpts=2.5 97.5
pctlname= _low _upp; run;

%if &cde=1 %then %do;
proc univariate data=boot_CDE_estimates noprint;
var CDE_OR CDE_RR CDE_RD;
output out=boot_CI_CDE  pctlpre= CDE_OR CDE_RR CDE_RD pctlpts=2.5 97.5
pctlname= _low _upp ;
run;
%end;

proc iml;

use boot_CI_OR;
read var {NDE_OR_low} into NDE_OR_low; read var {NDE_OR_upp} into NDE_OR_upp;
read var {NIE_OR_low} into NIE_OR_low; read var {NIE_OR_upp} into NIE_OR_upp;
read var {TE_OR_low}  into TE_OR_low;  read var {TE_OR_upp}  into TE_OR_upp;
NDE_CI_OR = NDE_OR_low || NDE_OR_upp;
NIE_CI_OR = NIE_OR_low || NIE_OR_upp;
TE_CI_OR =  TE_OR_low  || TE_OR_upp;
boot_CI_OR  =  NDE_CI_OR // NIE_CI_OR // TE_CI_OR;
create boot_OR from boot_CI_OR[colname={"boot_CI_OR_low" "boot_CI_OR_upp"}];
append from boot_CI_OR; close boot_OR;

use boot_CI_RR;
read var {NDE_RR_low} into NDE_RR_low; read var {NDE_RR_upp} into NDE_RR_upp;
read var {NIE_RR_low} into NIE_RR_low; read var {NIE_RR_upp} into NIE_RR_upp;
read var {TE_RR_low}  into TE_RR_low;  read var {TE_RR_upp}  into TE_RR_upp;
NDE_CI_RR = NDE_RR_low || NDE_RR_upp;
NIE_CI_RR = NIE_RR_low || NIE_RR_upp;
TE_CI_RR =  TE_RR_low  || TE_RR_upp;
boot_CI_RR  =  NDE_CI_RR // NIE_CI_RR // TE_CI_RR;
create boot_RR from boot_CI_RR[colname={"boot_CI_RR_low" "boot_CI_RR_upp"}];
append from boot_CI_RR; close boot_RR;

use boot_CI_RD;
read var {NDE_RD_low} into NDE_RD_low; read var {NDE_RD_upp} into NDE_RD_upp;
read var {NIE_RD_low} into NIE_RD_low; read var {NIE_RD_upp} into NIE_RD_upp;
read var {TE_RD_low}  into TE_RD_low;  read var {TE_RD_upp}  into TE_RD_upp;
NDE_CI_RD = NDE_RD_low || NDE_RD_upp;
NIE_CI_RD = NIE_RD_low || NIE_RD_upp;
TE_CI_RD =  TE_RD_low  || TE_RD_upp;
boot_CI_RD  =  NDE_CI_RD // NIE_CI_RD // TE_CI_RD;
```

```
  create boot_RD from boot_CI_RD[colname={"boot_CI_RD_low" "boot_CI_RD_upp"}];
  append from boot_CI_RD; close boot_RD;

  %if &cde=1 %then %do;
  use boot_CI_CDE;
  read var {CDE_OR_low} into CDE_OR_low; read var {CDE_OR_upp} into CDE_OR_upp;
  read var {CDE_RR_low} into CDE_RR_low; read var {CDE_RR_upp} into CDE_RR_upp;
  read var {CDE_RD_low} into CDE_RD_low; read var {CDE_RD_upp} into CDE_RD_upp; ;
  CDE_CI_OR = CDE_OR_low || CDE_OR_upp;
  CDE_CI_RR = CDE_RR_low || CDE_RR_upp;
  CDE_CI_RD = CDE_RD_low || CDE_RD_upp;
  boot_CI_CDE  =  CDE_CI_OR // CDE_CI_RR // CDE_CI_RD;
  create boot_CDE from boot_CI_CDE[colname={"boot_CI_CDE_low" "boot_CI_CDE_upp"}];
  append from boot_CI_CDE; close boot_CDE;
  %end;

  quit;

  data results_OR; merge results_OR boot_OR; run;
  data results_RR; merge results_RR boot_RR; run;
  data results_RD; merge results_RD boot_RD; run;

  %if &cde=1 %then %do;
  data results_CDE; merge results_CDE boot_CDE; run;
  %end;

%end;

title "Odds ratio scale";      proc print data=results_OR; run; title;
title "Risk ratio scale";      proc print data=results_RR; run; title;
title "Risk difference scale"; proc print data=results_RD; run; title;

%if &cde=1 %then %do;
title "Controlled direct effects";  proc print data=results_CDE; run; title;
%end;

%mend bin_cont_exactmed;
```

CHAPITRE VI

IMPLÉMENTATION DE L'APPROCHE EXACTE D'ANALYSE DE MÉDIATION CAUSALE POUR UNE RÉPONSE BINAIRE EN SAS ET R

Dans ce chapitre, nous présentons les fonctionnalités principales des macros SAS `mediation_estimates` et `bin_cont_exactmed` et du paquet R `ExactMed`, conçus pour implémenter l'approche exacte de Samoilenko et Lefebvre (2021, 2023) pour une réponse binaire.

Dans les Chapitres 4 et 5, nous avons présenté les macros SAS `mediation_estimates` et `bin_cont_exactmed` permettant d'implémenter notre approche exacte d'analyse de médiation causale pour une réponse binaire basée sur la régression (Samoilenko et Lefebvre, 2021, 2023). La première macro a été développée pour un médiateur binaire. Lorsque le médiateur est continu, la macro `bin_cont_exactmed` est applicable.

Récemment, le paquet R `ExactMed` (Caubet *et al.*, 2023) a aussi été mis à la disposition des chercheurs appliqués souhaitant effectuer des analyses de médiation causale basées sur l'approche exacte de Samoilenko et Lefebvre (2021, 2023).

Nos macros SAS et paquet R susmentionnés permettent d'estimer des effets naturels et directs contrôlés sur les échelles du rapport de cotes, du rapport de risques et de la différence de risques. Ils sont conçus pour l'estimation des effets non ajustés (bruts), ainsi que pour estimer des effets ajustés aux covariables (c'est-à-dire des effets conditionnels). Par défaut, comme dans la macro SAS de Valeri et VanderWeele (2013), les estimations des effets conditionnels sont obtenues en fixant les covariables à leurs valeurs moyennes spécifiques à l'échantillon pour les covariables numériques ; pour les covariables catégorielles, la procédure est effectuée par le biais des variables indicatrices (*dummy variables*) associées. Il est également possible d'obtenir les estimations des effets ajustés pour des valeurs spécifiques des covariables fournies par l'utilisateur.

Nos outils SAS et R permettent aussi d'incorporer un terme d'interaction entre l'exposition et le médiateur dans le modèle de la réponse. En ce qui concerne l'estimation par intervalle, nos outils génèrent des intervalles de confiance basés sur la méthode delta (Casella et Berger, 2002) ou par le bootstrap basé sur les percentiles (Efron et Tibshirani, 1994). Toutefois, contrairement aux macros SAS `mediation_estimates` et `bin_cont_exactmed` qui ne construisent que des intervalles de confiance à 95%, le paquet R `ExactMed` est plus flexible en permettant de différentes valeurs du coefficient de confiance.

La pénalisation de Firth sert à réduire le biais des estimateurs des coefficients de régression logistique dans le cas où les données sont séparées ou quasi séparées (Greenland et Mansournia, 2015; Mansournia *et al.*, 2018). Nos macros SAS et paquet R possèdent une fonctionnalité optionnelle permettant d'appliquer cette méthode de réduction du biais. Si cette fonctionnalité est activée, la pénalisation de Firth est appliquée au modèle de la réponse ; elle est aussi appliquée au modèle du médiateur lorsque le médiateur est binaire.

Les macros SAS `mediation_estimates` et `bin_cont_exactmed` ne sont applicables qu'aux données provenant des études de cohorte. Afin de surmonter cette limitation, les fonctionnalités permettant d'implémenter l'approche exacte de Samoilenko et Lefebvre (2021, 2023) aux données issues des études cas-témoins ont été incluses dans notre paquet R `ExactMed`. Ces fonctionnalités se basent sur l'extension théorique de l'approche exacte de Samoilenko et Lefebvre (2021, 2023) aux dévis cas-témoins (Lefebvre et Caubet, 2022).

Les détails techniques relatifs à l'utilisation des macros SAS `mediation_estimates` et `bin_cont_exactmed` sont discutés dans les Sous-sections 4.7 et 5.6.10, respectivement. Nous référons le lecteur à la vignette *The ExactMed functions* du paquet R `ExactMed` (Caubet *et al.*, 2023) pour une illustration de l'utilisation de ses fonctions.

CONCLUSION

Les approches d'analyse de médiation causale qui reposent sur la spécification des modèles paramétriques pour le médiateur et la réponse sont naturellement attrayantes pour les chercheurs appliqués en raison de leur lien naturel avec l'approche traditionnelle intrinsèquement paramétrique, ainsi que de leur simplicité conceptuelle. Dans le contexte de l'analyse de médiation simple, un nombre d'approches causales basées sur la régression logistique pour une réponse binaire ont été proposées dans la littérature pour estimer les effets naturels direct et indirect (Gaynor *et al.*, 2019; Valeri et VanderWeele, 2013; VanderWeele et Vansteelandt, 2010). Ces approches ont invoqué lesdites hypothèses de la réponse rare ou commune (non rare) afin d'obtenir des expressions *approximatives* simples et fermées pour les effets naturels sur l'échelle du rapport de cotes. Plus précisément, les techniques d'estimation de VanderWeele et Vansteelandt (2010) et de Valeri et VanderWeele (2013) ont été développées sous l'hypothèse de la réponse rare respectivement pour un médiateur binaire ou continu, tandis que l'approche de Gaynor *et al.* (2019) a été proposée pour une réponse commune (ayant une prévalence entre 20% et 60%) dans le cas d'un médiateur continu. Cependant, dans les études appliquées, l'évaluation de l'hypothèse de la réponse rare présente une grande difficulté due à l'absence des lignes directrices explicites permettant de qualifier une réponse binaire comme rare dans le dans le contexte de la médiation causale. Quant à l'approche de Gaynor *et al.* (2019), la question qui se pose naturellement est quelle stratégie qu'on devrait adopter lorsque la prévalence de la réponse se situe hors de l'intervalle entre 20% et 60%.

Il est important de noter que, lorsqu'on parle de l'hypothèse de la réponse rare dans le contexte de la médiation causale, le terme « réponse » doit être compris au sens large, y incluant le médiateur. Ainsi, dans notre article *« Point : Risk ratio equations for natural direct and indirect effects in causal mediation analysis of a binary mediator and a binary outcome — A fresh look at the formulas »* publié dans *American Journal of Epidemiology* et présenté dans le Chapitre 3 de cette thèse (voir Samoilenko et Lefebvre (2019) dans la bibliographie), nous avons illustré que la violation de l'hypothèse de la rareté du médiateur binaire peut engendrer des biais non négligeables des estimateurs des effets naturels dérivés sous cette hypothèse. De plus, cet article et notre autre article (Samoilenko *et al.*, 2018) ont suscité une réponse de VanderWeele *et al.* (2019) où ces auteurs ont

reconnu que l'hypothèse de la réponse rare se doit d'être satisfaite dans toutes les strates formées par le traitement, le médiateur et les covariables pour que leurs estimateurs approximatifs dérivés sous cette hypothèse soient valides. Ainsi, une évaluation naïve de l'hypothèse de la réponse rare de façon marginale n'est pas suffisante pour garantir une performance adéquate des estimateurs des effets naturels proposés par VanderWeele et Vansteelandt (2010) et Valeri et VanderWeele (2013).

Une prise de conscience des problèmes liés à l'évaluation de l'hypothèse de la réponse rare dans le contexte de la médiation causale d'une réponse binaire, ainsi que l'impact potentiel non négligeable du non-respect de cette hypothèse sur la qualité des estimateurs dérivés sous cette hypothèse, nous a incitées à développer des estimateurs des effets naturels basés sur la régression qui n'invoquent aucune hypothèse théorique simplificatrice dans leur dérivation et qui, conséquemment, permettent d'éviter des difficultés liées à la vérification des hypothèses basées sur la prévalence de la réponse.

Ainsi, dans l'article *« Parametric-regression–based causal mediation analysis of binary outcomes and binary mediators : Moving beyond the rareness or commonness of the outcome »* de Samoilenko et Lefebvre (2021) publié dans *American Journal of Epidemiology* et présenté dans le Chapitre 4 de cette thèse, nous avons proposé des estimateurs des effets naturels dans le cas d'une réponse Y et d'un médiateur M binaires développés sous la spécification des modèles de régression logistique pour Y et M. Ces estimateurs sont *exacts* dans le sens qu'ils sont dérivés sans aucune hypothèse théorique simplificatrice et, conséquemment, leur précision est définie, au-delà de la taille échantillonnale, par la précision numérique des logiciels utilisés et par la tolérance par défaut ou spécifiée par l'utilisateur des routines appliquées. Contrairement à l'approche approximative de Valeri et VanderWeele (2010), notre approche exacte a introduit des estimateurs des effets naturels sur trois échelles binaires standards, à savoir le rapport de risques, le rapport de cotes et la différence de risques. Pour chaque échelle considérée, des formules exactes pour les erreurs standard ont également été dérivées en utilisant la méthode delta multivariée du premier ordre. Nous avons montré une performance adéquate de nos estimateurs exacts ponctuels et par intervalle (bootstrap ou basés sur la méthode delta) dans des études de simulation où la réponse était rare ou commune, y compris des scénarios où la réponse était rare marginalement, mais pas conditionnellement. De cette manière, nos estimateurs proposés permettent de contourner des difficultés liées à l'évaluation des hypothèses de la réponse rare ou commune dans le contexte de médiation causale. De plus, dans certaines études appliquées, les chercheurs peuvent être intéressés à estimer l'effet direct contrôlé (VanderWeele, 2011; Imai *et al.*, 2013). Valeri et VanderWeele (2013) ont proposé un estimateur de cet effet sur l'échelle du rapport de cotes

qui a été dérivé sous le modèle de régression logistique pour la réponse. Nous avons généralisé leur estimateur aux échelles du rapport de risques et de la différence de risques. Il faut noter que les dérivations correspondantes n'invoquent pas l'hypothèse de la réponse rare et donc ces estimateurs de l'effet direct contrôlé sont exacts par construction.

L'article *« An exact regression-based approach for the estimation of natural direct and indirect effects with a binary outcome and a continuous mediator »* de Samoilenko et Lefebvre (2023) publié dans *Statistics in Medicine* et présenté dans le Chapitre 5 de cette thèse est une extension naturelle de notre approche exacte pour une réponse et un médiateur binaires (Samoilenko et Lefebvre, 2021) au cas d'un médiateur continu. Ainsi, dans cet article, nous avons proposé les estimateurs exacts des effets naturels sur trois échelles binaires standards basés sur les régressions logistique et linéaire pour la réponse et le médiateur, respectivement. Les formules correspondantes pour l'estimation des erreurs par la méthode delta ont été dérivées. Les études de simulation effectuées ont montré une performance adéquate des estimateurs exacts proposés pour les réponses rares ou communes marginalement. Nous avons aussi constaté une bonne performance des nos estimateurs dans une étude de simulation dans laquelle la réponse était rare marginalement alors qu'il y avait une grande déviation de l'hypothèse de la réponse rare dans certaines strates définies par l'exposition et le médiateur. Nous avons aussi étudié par simulation l'impact de l'omission du terme d'interaction dans le modèle de la réponse lorsqu'une telle interaction existe ; une réduction de performance a été observée en fonction de la magnitude de la valeur du coefficient d'interaction dans le mécanisme de génération des données. Dans notre étude de simulation évaluant l'impact d'une déviation de la normalité de l'erreur du médiateur, nous avons observé que nos estimateurs exacts étaient robustes à la non-normalité dans les scénarios et les distributions de médiateurs considérés (t généralisée et *Gamma*).

Notre contribution pratique consiste au développent des macros SAS `mediation_estimates` et `bin_cont_exactmed` conçues pour implémenter l'approche exacte de Samoilenko et Lefebvre (2021, 2023) pour une réponse binaire. Ces macros fournissent les estimations des effets naturels et directs contrôlés sur les échelles du rapport de risques, du rapport de cotes et de la différence de risques facilitant ainsi une comparaison directe avec les résultats obtenus par d'autres approches de médiation causale pour la réponse binaire. Récemment, nous avons aussi mis à la disposition des chercheurs appliqués le paquet R `ExactMed` (Caubet *et al.* (2023) ; créateur : Miguel Caubet) pour effectuer des analyses de médiation causale basées sur l'approche exacte. Ma contribution principale à la création

de ce paquet a consisté dans l'exécution des tests exhaustifs de ses fonctionnalités ainsi qu'à une revue de la documentation associée.

Il est bien documenté dans la littérature que, lorsqu'un modèle de régression logistique est appliqué à de petits échantillons et/ou à des données éparses, les méthodes conventionnelles d'estimation du maximum de vraisemblance sont susceptibles d'être biaisées ou de produire des estimations infinies. La pénalisation de Firth (1993) est une approche populaire pour traiter ce type de problèmes (Greenland et Mansournia, 2015) et nos macros SAS permettent de l'appliquer dans les modèles logistiques utilisés dans les analyses. Cependant, les valeurs du biais significatives obtenues pour les échelles multiplicatives et/ou la différence de risques dans certains de nos scénarios de l'étude de simulation effectuée pour évaluer l'impact de cette pénalisation dans le contexte d'une réponse binaire et d'un médiateur continu (Samoilenko et Lefebvre, 2023) suggèrent que des études supplémentaires sont nécessaires pour examiner l'impact de la pénalisation de Firth dans le contexte de l'approche exacte proposée. Alors que cette méthode est généralement considérée comme un outil efficace pour traiter les problèmes d'estimation susmentionnés, il est documenté par certains auteurs (Puhr *et al.*, 2017; Rahman et Sultana, 2017) que la pénalisation de Firth peut introduire un biais dans l'estimation des probabilités prédites moyennes ou individuelles (à noter que les fonctions *expit* dans les équations des probabilités contrefactuelles exactes (4.3) et (5.3) peuvent être considérées comme des probabilités prédites). Deux modifications de la méthode de Firth proposées par Puhr *et al.* (2017), FLIC (*Firth's logistic regression with intercept-correction*) et FLAC (*Firth's logistic regression with added covariate*), améliorent efficacement la performance prédictive de la pénalisation de Firth (en termes des probabilités prédites moyennes et individuelles). Ainsi, l'utilisation des modifications FLIC et FLAC de cette pénalisation est une solution potentielle afin d'améliorer la performance des estimateurs exacts proposés dans le contexte de petits échantillons et/ou de données éparses. Par ailleurs, dans le contexte générale de l'analyse de ce type de données, lorsque la précision de la prédiction est une question centrale, des méthodes qui se concentrent sur l'erreur de prédiction (par exemple, LASSO) sont recommandées (Mansournia *et al.*, 2018). Ainsi, des méthodes basées sur le LASSO (par exemple, la méthode *outcome-adaptive LASSO* de Ye *et al.* (2021) développée pour la sélection de variables dans le cadre de l'analyse de la médiation causale) pourraient être aussi investiguées.

L'une des limites de notre approche exacte présentée dans les Chapitres 4 et 5 est qu'elle n'est actuellement applicable qu'aux données provenant d'études de cohortes. Effectivement, notre approche nécessite des estimateurs convergents pour les paramètres populationnels des régressions du média-

teur sur l'exposition et les covariables et de la réponse sur l'exposition, le médiateur et les covariables, ce qui peut être réalisé dans une étude de cohorte vu que, sous ce devis, la sélection des unités est effectuée sur la base de leur statut d'exposition. Ainsi, des développements supplémentaires sont nécessaires pour généraliser l'approche proposée aux données provenant d'études cas-témoins, car ce devis repose sur une sélection des unités sur la base de leur statut de réponse (Stürmer et Brookhart, 2013).

Notre approche est actuellement limitée aux expositions et médiateurs binaires ou continues. Cependant, des expositions et des médiateurs catégoriels ne sont pas rares dans les recherches appliquées. Par conséquent, il serait naturel de généraliser notre approche exacte pour ce type d'exposition et/ou de médiateur en faisant aussi une extension correspondante de nos macros SAS. Dans le contexte d'un médiateur binaire ou continu, une généralisation de notre approche exacte au cas d'une exposition catégorielle, ainsi que son implémentation en SAS, sont directes. Dans le cas d'un médiateur catégoriel et d'une exposition binaire ou continue, certains développements théoriques basés sur le modèle multinomial logistique du médiateur ont déjà été effectués et implémentés en SAS (Samoilenko et Lefebvre, 2022b). La combinaison d'une exposition *et* d'un médiateur catégoriels a un coût en termes de la complexité des dérivations des expressions correspondants et de leur implémention informatique, mais cet objectif est tout à fait réalisable.

Notre approche exacte est un choix naturel lorsque les modèles du médiateur et de la réponse sont correctement spécifiés. Notamment, il est à noter que le modèle de la réponse (2.3), une pierre angulaire de notre approche exacte, présume que l'effet du médiateur M sur la réponse Y est linéaire sur l'échelle logit. En pratique, nous recommandons à l'utilisateur de vérifier empiriquement la plausibilité de cette hypothèse de linéarité. Lorsqu'elle ne semble pas être raisonnable, d'autres techniques d'estimation peuvent être utilisées. Dans ce contexte, nous mentionnons l'approche d'Imai *et al.* (2010) dont l'implémentation dans le paquet R `mediation` permet des modèles plus flexibles (par exemple, des modèles additifs généralisés) ainsi que l'approche par des modèles à effets naturels basée sur la pondération (Lange *et al.*, 2012) qui ne requiert pas la modélisation de Y en fonction de M et qui est implémentée dans le paquet R `medflex` (voir aussi les Sous-sections 2.3 et 2.4 pour les approches susmentionnées).

Enfin, notre approche exacte a été développée dans le cadre d'un modèle de médiation simple. Dans de futures études, il serait intéressant de considérer le cas de médiateurs multiples et développer une extension correspondante de nos macros SAS. Cependant, même dans le cas où il n'y a que

deux médiateurs continus, le calcul des estimations des probabilités contrefactuelles emboîtées peut nécessiter une intégration numérique non-triviale (sur $\mathbb{R}^2$ dans ce cas). Toutefois, l'utilisation de la sous-routine QUAD dans SAS IML (SAS Institute Inc., 2010) peut aider à la résolution des problèmes. Notons que la complexité des dérivations des expressions pour les erreurs standards basées sur la méthode delta augmente considérablement dans le contexte des médiateurs multiples.

ANNEXE A

DISTRIBUTION LOGIT NORMALE

Nous disons qu'une variable aléatoire X suit la loi *logit-normale*(μ, σ^2) sur l'intervalle $(0, 1)$ si sa transfomation logit, $\ln(X/(1-X))$, suit une loi normale $\mathcal{N}(\mu, \sigma^2)$ (Frederic et Lad, 2008). Selon la formule pour la fonction de densité d'une transformation d'une variable aléatoire continue (Casella et Berger, 2002, p. 51), nous avons que la densité de X s'exprime comme

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}x(1-x)} \exp\left\{-\frac{1}{2}\left(\frac{\ln\left(\frac{x}{1-x}\right)-\mu}{\sigma}\right)^2\right\}, \quad x \in (0,1), \tag{A.1}$$

Dans le cas général, il n'existe pas d'expressions de forme analytique fermée pour l'espérance et la variance de X (Frederic et Lad, 2008).

Nous avons de l'équation (A.1) pour l'espérance de X :

$$\begin{aligned} E\{X\} &= \frac{1}{\sqrt{2\pi\sigma^2}} \int_0^1 \frac{x}{x(1-x)} \exp\left\{-\frac{1}{2}\left(\frac{\ln\left(\frac{x}{1-x}\right)-\mu}{\sigma}\right)^2\right\} dx \\ &= \frac{1}{\sqrt{2\pi\sigma^2}} \int_0^1 \frac{1}{1-x} \exp\left\{-\frac{1}{2}\left(\frac{\ln\left(\frac{x}{1-x}\right)-\mu}{\sigma}\right)^2\right\} dx. \end{aligned} \tag{A.2}$$

Le changement de la variable d'intégration

$$t = \ln\left(\frac{x}{1-x}\right), \quad x \in (0,1), \tag{A.3}$$

implique que

$$x = \frac{\exp(t)}{1+\exp(t)}, \quad 1-x = \frac{1}{1+\exp(t)}, \quad t \in (-\infty, \infty),$$

et

$$dx = \frac{\exp(t)}{(1+\exp(t))^2} dt.$$

Ainsi, nous pouvons réécrire l'espérance (A.2) sous le changement de la variable d'intégration (A.3)

$$\begin{aligned} E\{X\} &= \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} (1+\exp(t)) \exp\left\{-\frac{1}{2}\left(\frac{t-\mu}{\sigma}\right)^2\right\} \frac{\exp(t)}{(1+\exp(t))^2} dt \\ &= \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} \frac{\exp(t)}{1+\exp(t)} \exp\left\{-\frac{1}{2}\left(\frac{t-\mu}{\sigma}\right)^2\right\} dt \\ &= \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} \text{expit}(t) \exp\left\{-\frac{1}{2}\left(\frac{t-\mu}{\sigma}\right)^2\right\} dt. \end{aligned} \tag{A.4}$$

Holmes et Schofield (2022) ont présenté les résultats permettant d'éviter l'intégration numérique pour calculer l'espérance (A.4) :

$$E\{X\} = 0.5 + \frac{a + \frac{2\pi}{\sigma^2} \cdot b}{1+2c}$$

où

$$\begin{aligned} a &= \sum_{k=1}^{\infty} \exp\left(-k^2\sigma^2/2\right) \sinh(k\mu) \tanh(k\sigma^2/2), \\ b &= \sum_{k=1}^{\infty} \frac{\exp\left(-(2k-1)^2\pi^2/2\sigma^2\right) \sin\left((2k-1)\pi\mu/\sigma^2\right)}{\sinh\left((2k-1)\pi^2/\sigma^2\right)}, \\ c &= \sum_{k=1}^{\infty} \exp\left(-k^2\sigma^2/2\right) \cosh(k\mu) \end{aligned}$$

et

$$\sinh(x) = \frac{\exp(x)-\exp(-x)}{2}, \quad \cosh(x) = \frac{\exp(x)+\exp(-x)}{2}, \quad \tanh(x) = \frac{\exp(x)-\exp(-x)}{\exp(x)+\exp(-x)}.$$

Toutefois, Holmes et Schofield (2022) ne proposent pas une alternative à l'intégration numérique des expressions (5.11-5.19).

BIBLIOGRAPHIE

Albert, J. M. et Wang, W. (2015). Sensitivity analyses for parametric causal mediation effect estimation. *Biostatistics*, *16*(2), 339–351. https://dx.doi.org/10.1093/biostatistics/kxu048

Allison, P. D. (2012). *Logistic regression using SAS : Theory and application* (2 éd.). SAS Institute Inc.

Amemiya, T. (1981). Qualitive response models : A survey. *Journal of Economic Literature*, *19*, 1483–1536.

Ananth, C. V. et VanderWeele, T. J. (2011). Placental abruption and perinatal mortality with preterm delivery as a mediator : Disentangling direct and indirect effects. *American Journal of Epidemiology*, *174*(1), 99–108. https://dx.doi.org/10.1093/aje/kwr045

Andrews, R. M. et Didelez, V. (2021). Insights into the cross-world independence assumption of causal mediation analysis. *Epidemiology*, *32*(2), 209–219. https://dx.doi.org/10.1097/EDE.0000000000001313

Baron, R. M. et Kenny, D. A. (1986). The moderator-mediator variable distinction in social psychological research : Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology*, *51*(6), 1173–1182. https://dx.doi.org/10.1037//0022-3514.51.6.1173

Bartlett, J. W. et Hughes, R. A. (2020). Bootstrap inference for multiple imputation under uncongeniality and misspecification. *Statistical Methods in Medical Research*, *29*(12), 3533–3546. https://dx.doi.org/10.1177/0962280220932189

Breen, R., Choi, S. et Holm, A. (2015). Heterogeneous causal effects and sample selection bias. *Sociological Science*, *2*, 351–369. https://dx.doi.org/110.15195/v2.a17

Camilli, G. (1994). Origin of the scaling Constant d = 1.7 in item response theory. *Journal of Educational and Behavioral Statistics*, *19*(3), 293–295. https://dx.doi.org/10.2307/1165298

Casella, G. et Berger, R. L. (2002). *Statistical inference* (2 éd.). Duxbury/Thomson Learning.

Caubet, M., Samoilenko, M. et Lefebvre, G. (2023). *ExactMed : Exact mediation analysis for binary outcomes*. R package version 0.2.0. https://CRAN.R-project.org/package=ExactMed

Cheng, C., Spiegelman, D. et Li, F. (2021). Estimating the natural indirect effect and the mediation proportion via the product method. *BMC Medical Research Methodology*, *21*, 253. https://dx.doi.org/10.1186/s12874-021-01425-4

Chernick, M. R. (2011). *Bootstrap methods : A guide for practitioners and researchers* (2 éd.). John Wiley & Sons.

Coffman, D. L., Schuler, M. S., Nguyen, T. Q. et McCaffrey, D. F. (2023). *Weighting estimators for causal mediation*, Dans *Handbook of matching and weighting adjustments for causal inference*, (p. 373–412). Chapman and Hall/CRC.

Cole, S. R., Chu, H. et Greenland, S. (2014). Maximum likelihood, profile likelihood, and penalized likelihood : A primer. *American Journal of Epidemiology*, *179*(2), 252–260. https://dx.doi.org/10.1093/aje/kwt245

Cox, D. R. (1970). *The analysis of binary data*. Methuen.

Daniel, R., Zhang, J. et Farewell, D. (2021). Making apples from oranges : Comparing noncollapsible effect estimators and their standard errors after adjustment for different covariate sets. *Biometrical Journal*, *63*(3), 528–557. https://dx.doi.org/10.1002/bimj.201900297

Davison, A. C. et Hinkley, D. V. (1997). *Bootstrap methods and their application*. Cambridge University Press. https://dx.doi.org/10.1017/CBO9780511802843

De Stavola, B. L., Daniel, R. M., Ploubidis, G. B. et Micali, N. (2015). Mediation analysis with intermediate confounding : structural equation modeling viewed through the causal inference lens. *American Journal of Epidemiology*, *181*(1), 64–80. https://dx.doi.org/10.1093/aje/kwu239

Diop, A., Lefebvre, G., Duchaine, C. S., Laurin, D. et Talbot, D. (2021). The impact of adjusting for pure predictors of exposure, mediator, and outcome on the variance of natural direct and indirect effect estimators. *Statistics in Medicine*, *40*(10), 2339–2354. https://dx.doi.org/10.1002/sim.8906

Doretti, M., Raggi, M. et Stanghellini, E. (2019). *Exact parametric causal mediation analysis for a binary outcome with a binary mediator*. arXiv. https://dx.doi.org/10.48550/arXiv.1811.00439

Doretti, M., Raggi, M. et Stanghellini, E. (2021). Exact parametric causal mediation analysis for a binary outcome with a binary mediator. *Statistical Methods & Applications*, *31*(1), 87–108. https://dx.doi.org/10.1007/s10260-021-00562-w

Efron, B. et Tibshirani, R. J. (1994). *An Introduction to the bootstrap*. Chapman and Hall/CRC Press. https://dx.doi.org/10.1201/9780429246593

Emsley, R. et Liu, H. (2013). PARAMED : Stata module to perform causal mediation analysis using parametric regression models. *Statistical Software Components*.

Fay, M. P. et Brittain, E. H. (2022). *Statistical hypothesis testing in context : Reproducibility, inference, and science*. Cambridge University Press. https://dx.doi.org/10.1017/9781108528825

Feingold, A., MacKinnon, D. P. et Capaldi, D. M. (2019). Mediation analysis with binary outcomes : Direct and indirect effects of pro-alcohol influences on alcohol use disorders. *Addictive Behaviors*, *94*, 26–35. https://dx.doi.org/10.1016/j.addbeh.2018.12.018

Firth, D. (1993). Bias reduction of maximum likelihood estimates. *Biometrika*, *80*(1), 27–38. https://dx.doi.org/10.1093/biomet/80.1.27

Frederic, P. et Lad, F. (2008). Two moments of the logitnormal distribution. *Communications in Statistics - Simulation and Computation*, *37*(7), 1263–1269. https://dx.doi.org/10.1080/03610910801983178

Gao, X. et Luo, L. (2019). An improvement in estimation of the standard error for the natural direct effect in causal mediation analysis. *Epidemiology*, *30*(4), e25–e26. https://dx.doi.org/10.1097/EDE.0000000000001005

Gaynor, S. M., Schwartz, J. et Lin, X. (2019). Mediation analysis for common binary outcomes. *Statistics in Medicine*, *38*(4), 512–529. https://dx.doi.org/10.1002/sim.7945

Gelman, A., Jakulin, A., Pittau, M. G. et Su, Y.-S. (2008). A weakly informative default prior distribution for logistic and other regression models. *Annals of Applied Statistics*, *2*(4), 1360–1383. https://dx.doi.org/10.1214/08-AOAS191

Greenland, S. et Brumback, B. (2003). An overview of relations among causal modelling methods. *International Journal of Epidemiology*, *31*(5), 1030–1037. https://dx.doi.org/10.1093/ije/31.5.1030

Greenland, S. et Mansournia, M. A. (2015). Penaization, bias reduction, and default priors in logistic and related categorical and survival regressions. *Statistics in medicine*, *34*(23), 3133–3143. https://dx.doi.org/10.1002/sim.6537

Greenland, S., Mansournia, M. A. et Altman, D. G. (2016). Sparse data bias : A problem hiding in plain sight. *British Medical Journal*, *352*, i1981. https://dx.doi.org/10.1136/bmj.i1981

Greenland, S. et Robins, J. M. (2009). Identifiability, exchangeability and confounding revisited. *Epidemiologic Perspectives & Innovations*, *6*(1), 1–9. https://dx.doi.org/10.1186/1742-5573-6-4

Hayes, A. et Little, T. (2018). *Introduction to mediation, moderation, and conditional process analysis : A Regression-based approach* (2 éd.). Guilford publications.

Heinze, G. (2006). A comparative investigation of methods for logistic regression with separated or nearly separated data. *Statistics in Medicine*, *25*(24), 4216–4226. https://dx.doi.org/10.1002/sim.2687

Heinze, G. et Schemper, M. (2002). A solution to the problem of separation in logistic regression. *Statistics in Medicine*, *21*(16), 2409–2419. https://dx.doi.org/https://doi.org/10.1002/sim.1047

Hernán, M. et Robins, J. (2023). *Causal inference : What if*. Chapman & Hall/CRC. https://dx.doi.org/10.1201/9781315374932

Hernán, M. A. (2004). A definition of causal effect for epidemiological research. *Journal of Epidemiology & Community Health*, *58*(4), 265–271. https://dx.doi.org/10.1136/jech.2002.006361

Hernán, M. A., Hernández-Díaz, S., Werler, M. M. et Mitchell, A. A. (2002). Causal knowledge as a prerequisite for confounding evaluation : An application to birth defects epidemiology. *American journal of epidemiology*, *155*(2), 176–184. https://dx.doi.org/10.1093/aje/155.2.176

Holland, P. W. (1986). Statistics and causal inference. *Journal of the American Statistical Association*, *81*(396), 945–960. https://dx.doi.org/10.1080/01621459.1986.10478354

Holmes, J. B. et Schofield, M. R. (2022). Moments of the logit-normal distribution. *Communications in Statistics - Theory and Methods*, *51*(3), 610–623. https://dx.doi.org/10.1080/03610926.2020.1752723

Iacobucci, D. (2012). Mediation analysis and categorical variables : The final frontier. *Journal of Consumer Psychology*, *22*(4), 582–594. https://dx.doi.org/10.1016/j.jcps.2012.03.006

Imai, K., Keele, L. et Tingley, D. (2010). A general approach to causal mediation analysis. *Psychological Methods*, *15*(4), 309–334. https://dx.doi.org/10.1037/a0020761

Imai, K., Tingley, D. et Yamamoto, T. (2013). Experimental designs for identifying causal mechanisms. *Journal of the Royal Statistical Society : Series A (Statistics in Society)*, *176*(1), 5–51. https://dx.doi.org/https://doi.org/10.1111/j.1467-985X.2012.01032.x

Jewell, N. (2004). *Statistics for epidemiology*. Chapman and Hall/CRC. https://dx.doi.org/10.1201/9781482286014

Khuri, A. D. (2002). *Advanced Calculus with Applications in Statistics* (2 éd.). John Wiley & Sons.

King, G., Tomz, M. et Wittenberg, J. (2000). Making the most of statistical analyses : Improving interpretation and presentation. *American Journal of Political Science*, *44*, 341–355. https://dx.doi.org/10.2307/2669316

Lange, T., Hansen, K. W., Sørensen, R. et Galatius, S. (2017). Applied mediation analyses : a review and tutorial. *Epidemiology and Health*, *39*, e2017035. https://dx.doi.org/10.4178/epih.e2017035

Lange, T., Vansteelandt, S. et Bekaert, M. (2012). A simple unified approach for estimating natural direct and indirect effects. *American Journal of Epidemiology*, *176*(3), 190–195. https://dx.doi.org/10.1093/aje/kwr525

Lash, T. L., VanderWeele, T. J., Haneause, S. et Rothman, K. J. (2021). *Modern epidemiology* (4 éd.). Lippincott Williams & Wilkins.

Lefebvre, G. et Caubet, M. (2022). *Investigating the performance of the exact estimator for causal mediation analysis of binary outcomes and binary mediators in case-control designs* [Conference presentation]. 49th Annual Meeting of the Statistical Society of Canada, May 29 – June 5, 2022.

Liang, K.-Y. et Zeger, S. L. (1986). Longitudinal data analysis using generalized linear models. *Biometrika*, *73*(1), 13–22. https://dx.doi.org/10.1093/biomet/73.1.13

Lilienfeld, A. M. (1957). Epidemiological methods and inferences in studies of noninfectious diseases. *Public Health Reports*, *72*(1), 51–60.

Loeys, T., Moerkerke, B., Smet, O. D., Buysse, A., Steen, J. et Vansteelandt, S. (2013). Flexible mediation analysis in the presence of nonlinear relations : Beyond the mediation formula. *Multivariate Behavioral Research*, *48*(6), 871–894. https://dx.doi.org/10.1080/00273171.2013.832132

Lundberg, I., Johnson, R. et Stewart, B. M. (2021). What is your estimand ? Defining the target quantity connects statistical evidence to theory. *American Sociological Review*, *86*(3), 532–565. https://dx.doi.org/10.1177/00031224211004187

MacMahon, B. et Pugh, T. F. (1970). *Epidemiology : Principles and methods*. Little Brown & Co.

Mansournia, M. A., Geroldinger, A., Greenland, S. et Heinze, G. (2018). Separation in logistic regression : Causes, consequences, and control. *American Journal of Epidemiology*, *187*(4), 864–870. https://dx.doi.org/10.1093/aje/kwx299

Morgan, S. L. et Winship, C. (2014). *Counterfactuals and causal inference : Methods and principles for social research* (2 éd.). Cambridge University Press. https://dx.doi.org/10.1017/CBO9781107587991

Morris, T. P., White, I. R. et Crowther, M. J. (2019). Using simulation studies to evaluate statistical methods. *Statistics in Medicine*, *38*(11), 2074–2102. https://dx.doi.org/https://doi.org/10.1002/sim.8086

Naimi, A. I., Kaufman, J. S. et MacLehose, R. F. (2014). Mediation misgivings : Ambiguous clinical and public health interpretations of natural direct and indirect effects. *International Journal of Epidemiology*, *43*(5), 1656–1661. https://dx.doi.org/10.1093/ije/dyu107

Neuhaus, J. M. et Jewell, N. P. (1993). A geometric approach to assess bias due to omitted covariates in generalized linear models. *Biometrika*, *80*(4), 807–815. https://dx.doi.org/10.1093/biomet/80.4.807

Nguyen, T. Q., Schmid, I., Ogburn, E. L. et Stuart, E. A. (2022). Clarifying causal mediation analysis : Effect identification via three assumptions and five potential outcomes. *Journal of Causal Inference*, *10*(1), 246–279. https://dx.doi.org/10.1515/jci-2021-0049

Nguyen, T. Q., Schmid, I. et Stuart, E. A. (2021). Clarifying causal mediation analysis for the applied researcher : Defining effects based on what we want to learn. *Psychological Methods*, *26*(2), 255–271. https://dx.doi.org/10.1037/met0000299

Oberg, A. S., VanderWeele, T. J., Almqvist, C. et Hernandez-Diaz, S. (2018). Pregnancy complications following fertility treatment—disentangling the role of multiple gestation. *International Journal of Epidemiology*, *47*(4), 1333–1342. https://dx.doi.org/10.1093/ije/dyy103

Patel, J. K. et Read, C. B. (1996). *Handbook of the normal distribution* (2 éd.). CRC Press.

Pearl, J. (2001). Direct and indirect effects. Proceedings of the Seventeenth Conference on Uncertainty in Artificial Intelligence, 411–420. Morgan Kaufmann Publishers Inc.

Pearl, J. (2012). The mediation formula : A guide to the assessment of causal pathways in nonlinear models. In C. Berzuini, P. Dawid, et L. Bernardinell (dir.), *Causality : Statistical Perspectives and Applications* 151–179. John Wiley & Sons, Ltd.

Pirlott, A. G. et MacKinnon, D. P. (2016). Design approaches to experimental mediation. *Journal of Experimental Social Psychology*, *66*, 29–38. https://dx.doi.org/10.1016/j.jesp.2015.09.012

Preacher, K. J. et Hayes, A. F. (2004). SPSS and SAS procedures for estimating indirect effects in simple mediation models. *Behavior Research Methods, Instruments, & Computers*, *36*, 717–731. https://dx.doi.org/10.3758/BF03206553

Puhr, R., Heinze, G., Nold, M., Lusa, L. et Geroldinger, A. (2017). Firth's logistic regression with rare events : Accurate effect estimates and predictions ? *Statistics in Medicine*, *36*(14), 2302–2317. https://dx.doi.org/10.1002/sim.7273

Rahman, M. S. et Sultana, M. (2017). Performance of Firth-and logF-type penalized methods in risk prediction for small or sparse binary data. *BMC Medical Research Methodology*, *17*. https://dx.doi.org/10.1186/s12874-017-0313-9

Richiardi, L., Bellocco, R. et Zugna, D. (2013). Mediation analysis in epidemiology : Methods, interpretation and bias. *International Journal of Epidemiology*, *42*(5), 1511–1519. https://dx.doi.org/10.1093/ije/dyt127

Rijnhart, J. J., Twisk, J. W., Eekhout, I. et Heymans, M. W. (2019). Comparison of logistic-regression based methods for simple mediation analysis with a dichotomous outcome variable. *BMC Medical Research Methodology*, *19*, 1–10. https://dx.doi.org/10.1186/s12874-018-0654-z

Robins, J. M. et Greenland, S. (1992). Identifiability and exchangeability for direct and indirect effects. *Epidemiology*, *3*(2), 143–155. https://dx.doi.org/10.1097/00001648-199203000-00013

Ross, S. (2009). *Initiation aux probabilités*. Lausanne : Presses polythechniques et universitaires romandes.

Rubin, D. B. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology*, *66*(5), 688–701. https://dx.doi.org/10.1037/h0037350

Rubin, D. B. (2005). Causal inference using potential outcomes : Design, modeling, decisions. *Journal of the American Statistical Association*, *100*(469), 322–331. https://dx.doi.org/10.1198/016214504000001880

Samoilenko, M., Blais, L. et Lefebvre, G. (2018). 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. *Observational Studies*, *4*(1), 193–216. https://dx.doi.org/10.1353/obs.2018.0013

Samoilenko, M. et Lefebvre, G. (2019). Point : Risk ratio equations for natural direct and indirect effects in causal mediation analysis of a binary mediator and a binary outcome — A fresh look at the formulas. *American Journal of Epidemiology*, *188*(7), 1201–1203. https://dx.doi.org/10.1093/aje/kwy275

Samoilenko, M. et Lefebvre, G. (2021). Parametric-regression–based causal mediation analysis of binary outcomes and binary mediators : Moving beyond the rareness or commonness of the outcome. *American Journal of Epidemiology*, *190*(9), 1846–1858. https://dx.doi.org/10.1093/aje/kwab055

Samoilenko, M. et Lefebvre, G. (2022a). Correction to : "Parametric-regression–based causal mediation analysis of binary outcomes and binary mediators : Moving beyond the rareness or commonness of the outcome". *American Journal of Epidemiology*, *191*(9), 1670. https://dx.doi.org/10.1093/aje/kwac078

Samoilenko, M. et Lefebvre, G. (2022b). *On the power to detect a natural indirect effect in causal mediation analysis with a categorical mediator and a minary Outcome* [Conference presentation]. 49th Annual Meeting of the Statistical Society of Canada, May 29 – June 5, 2022.

Samoilenko, M. et Lefebvre, G. (2023). 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*, *42*(3), 353–387. https://dx.doi.org/https://doi.org/10.1002/sim.9621

SAS Institute Inc. (2010). *SAS/IML®9.22 User's Guide*. SAS Institute Inc.

SAS Institute Inc. (2017). *SAS/STAT®14.3 User's Guide*. SAS Institute Inc.

Savalei, V. (2006). Logistic approximation to the normal : The KL rationale. *Psychometrika*, *71*(4), 763–767. https://dx.doi.org/110.1007/s11336-004-1237-y

Shi, B., Choirat, C., Coull, B. A., VanderWeele, T. J. et Valeri, L. (2021). CMAverse : A Suite of functions for reproducible causal mediation analyses. *Epidemiology*, *32*(5), e20–e22. https://dx.doi.org/10.1097/EDE.0000000000001378

Šinkovec, H., Geroldinger, A. et Heinze, G. (2019). Bring more data ! — a good advice ? Removing separation in logistic regression by increasing sample size. *International Journal of Environmental Research and Public Health*, *16*(23), 4658. https://dx.doi.org/10.3390/ijerph16234658

Spiegelman, D. et Hertzmark, E. (2005). Easy SAS calculations for risk or prevalence ratios and differences. *American Journal of Epidemiology*, *162*(3), 199–200. https://dx.doi.org/10.1093/aje/kwi188

Splawa-Neyman, J. (1923). Próba uzasadnienia zastosowań rachunku prawdopodobieństwa do doświadczeń polowych. *Roczniki Nauk Rolniczych i Leśnych*, *10*, 1–51.

Splawa-Neyman, J., Dabrowska, D. M. et Speed, T. P. (1990). On the application of probability theory to agricultural experiments. Essay on principles. *Statistical Science*, *5*(4), 465–472. https://dx.doi.org/10.1214/ss/1177012031

Starkopf, L., Andersen, M. P., Gerds, T. A., Torp-Pedersen, C. et Lange, T. (2017). *Comparison of five software solutions to mediation analysis*. Rapport technique, University of Copenhagen, Department of Biostatistics.

StataCorp. (2023). *STATA Causal inference and treatment-effects estimation reference reference manual : Release 18*. Stata Press.

Steen, J., Loeys, T., Moerkerke, B. et Vansteelandt, S. (2017). medflex : An R package for flexible mediation analysis using natural effect models. *Journal of Statistical Software*, *76*, 1–46. https://dx.doi.org/10.18637/jss.v076.i11

Stürmer, T. et Brookhart, M. A. (2013). *Study design considerations*, Dans P. Velentgas, N. A. Dreyer, P. Nourjah, S. R. Smith, et M. M. Torchia (dir.). *Developing a protocol for observational comparative effectiveness research : A user's guide*. Agency for Healthcare Research and Quality (US).

Tchetgen Tchetgen, E. (2014). A note on formulae for causal mediation analysis in an odds ratio context. *Epidemiologic Methods*, *2*(1), 21–31. https://dx.doi.org/10.1515/em-2012-0005

Tchetgen Tchetgen, E. J. et Shpitser, I. (2012). Semiparametric theory for causal mediation analysis : Efficiency bounds, multiple robustness, and sensitivity analysis. *The Annals of Statistics*, *40*(3), 1816–1845. https://dx.doi.org/10.2307/41713695

Tingley, D., Yamamoto, T., Hirose, K., Keele, L. et Imai, K. (2014). mediation : R package for causal mediation analysis. *Journal of Statistical Software*, *59*(5), 1–38. https://dx.doi.org/10.18637/jss.v059.i05

Valeri, L. et VanderWeele, T. J. (2010). *Extending the Baron and Kenny mediation analysis to allow for exposure-mediator interactions : SAS and SPSS macros*. Rapport technique, Harvard School of Public Health.

Valeri, L. et VanderWeele, T. J. (2013). Mediation analysis allowing for exposure-mediator interactions and causal interpretation : Theoretical assumptions and implementation with SAS and SPSS macros. *Psychological Methods*, *18*(2), 137–150. https://dx.doi.org/10.1037/a0031034

van Buuren, S. et Groothuis-Oudshoorn, K. (2011). mice : Multivariate imputation by chained equations in R. *Journal of Statistical Software*, *45*(3), 1–67. https://dx.doi.org/10.18637/jss.v045.i03

van Smeden, V. M., de Groot, J. A., Moons, K. G., Collins, G. S., Altman, D. G., Eijkemans, M. J. et Reitsma, J. B. (2016). No rationale for 1 variable per 10 events criterion for binary logistic regression analysis. *BMC Medical Research Methodology*, *16*(1), 1–12. https://dx.doi.org/10.1186/s12874-016-0267-3

VanderWeele, T. J. (2011). Controlled direct and mediated effects : definition, identification and bounds. *Scandinavian Journal of Statistics*, *38*(3), 551–563. https://dx.doi.org/10.1111/j.1467-9469.2010.00722.x

VanderWeele, T. J. (2013). Policy-relevant proportions for direct effects. *Epidemiology*, *24*(1), 175–176. https://dx.doi.org/10.1097/EDE.0b013e3182781410

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

VanderWeele, T. J. (2016). Mediation analysis : a practitioner's guide. *Annual review of public health*, *37*, 17–32. https://dx.doi.org/10.1146/annurev-publhealth-032315-021402

VanderWeele, T. J. et Hernan, M. A. (2013). Causal inference under multiple versions of treatment. *Journal of Causal Inference*, *1*(1), 1–20. https://dx.doi.org/10.1515/jci-2012-0002

VanderWeele, T. J., Valeri, L. et Ananth, C. V. (2019). Counterpoint : Mediation formulas with binary mediators and outcomes and the "Rare outcome assumption". *American Journal of Epidemiology*, *188*(7), 1204–1205. https://dx.doi.org/10.1093/aje/kwy281

VanderWeele, T. J. et Vansteelandt, S. (2009). Conceptual issues concerning mediation, interventions and composition. *Statistics and Its Interface*, *2*, 457–468. https://dx.doi.org/10.4310/SII.2009.v2.n4.a7

VanderWeele, T. J. et Vansteelandt, S. (2010). Odds ratios for mediation analysis for a dichotomous outcome. *American Journal of Epidemiology*, *172*(12), 1339–1348. https://dx.doi.org/10.1093/aje/kwq332

VanderWeele, T. J. et Vansteelandt, S. (2014). Mediation analysis with multiple mediators. *Epidemiologic Methods*, *2*(1), 95–115. https://dx.doi.org/doi:10.1515/em-2012-0010

Vansteelandt, S., Bekaert, M. et Lange, T. (2012). Imputation strategies for the estimation of natural direct and indirect effects. *Epidemiologic Methods*, *1*(1), 131–158. https://dx.doi.org/10.1515/2161-962X.1014

Vo, T.-T., Superchi, C., Boutron, I. et Vansteelandt, S. (2020). The conduct and reporting of mediation analysis in recently published randomized controlled trials : Results from a methodological systematic review. *Journal of Clinical Epidemiology, Volume 117, 78 - 88*, *117*(3), 78–88. https://dx.doi.org/10.1016/j.jclinepi.2019.10.001

Wang, A. et Arah, O. A. (2015). G-computation demonstration in causal mediation analysis. *European Journal of Epidemiology*, *30*(10), 1119–1127. https://dx.doi.org/10.1007/s10654-015-0100-z

Westreich, D. et Cole, S. R. (2010). Invited commentary : Positivity in practice. *American Journal of Epidemiology*, *171*(6), 674–677. https://dx.doi.org/10.1146/annurev.soc.25.1.659

White, H. (1980). A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica*, 817–838. https://dx.doi.org/10.2307/1912934

Wicklin, R. (2011). How to numerically integrate a function in SAS. The Do Loop SAS blog. https://blogs.sas.com/content/iml/2011/05/06/how-to-numerically-integrate-a-function-in-sas.html. Accessed September 7, 2021.

Xu, J. et Long, J. S. (2005). Confidence intervals for predicted outcomes in regression models for categorical outcomes. *The Stata Journal*, *5*(4), 537–559. https://dx.doi.org/10.1177/1536867X0500500405

Ye, Z., Zhu, Y. et Coffman, D. L. (2021). Variable selection for causal mediation analysis using LASSO-based methods. *Statistical Methods in Medical Research*, *30*(6), 1413–1427. https://dx.doi.org/10.1177/0962280221997505

Yuan, Y. (2011). Multiple imputation using SAS software. *Journal of Statistical Software*, *45*(6), 1–25. https://dx.doi.org/10.18637/jss.v045.i06

Yung, Y.-F., Lamm, M. et Wei, Z. (2018). Causal mediation analysis with the causalmed procedure. Dans *Proceedings of the SAS Global Forum 2018 Conference*. SAS Institute Inc. https://www.sas.com/content/dam/SAS/support/en/sas-global-forum-proceedings/2018/1991-2018.pdf

Zugna, D., Popovic, M., Fasanelli, F., Heude, B., Scelo, G. et Richiardi, L. (2022). Applied causal inference methods for sequential mediator. *BMC Medical Research Methodology*, *22*, 301. https://dx.doi.org/10.1186/s12874-022-01764-w