CAUSAL MEDIATION ANALYSIS FOR NON-LINEAR MODELS

BY

WEI WANG

Submitted in partial fulfillment of the requirements

for the degree of Doctor of Philosophy

Thesis Adviser: Dr. Jeffrey M. Albert

Department of Epidemiology and Biostatistics

CASE WESTERN RESERVE UNIVERSITY

May 2012

CASE WESTERN RESERVE UNIVERSITY

SCHOOL OF GRADUATE STUDIES

We hereby approve the thesis/dissertation of

Wei Wang

candidate for the Doctor of Philosophy degree *

(signed) Jeffrey M. Albert
(chair of Committee)

Mark D. Schluchter

Pingfu Fu

Suchitra Nelson

(date) Mar. 9th, 2012

*We also certify that written approval has been obtained for any proprietary material contained therein.

Table of Contents

List of Tables

List of Figures

Acknowledgements

I would like to express my sincere gratitude to my thesis advisor, Dr. Jeffrey M. Albert, not only for his insightful guidance and endless support, but also for his patience and encouragement throughout my graduate study.

Thanks also go to Professors Mark D. Schluchter, Suchitra Nelson and Pingfu Fu. I am grateful to have you in my Ph.D. committee and for helping me improve my thesis with your superb suggestions and advice. Thanks to Dr. Fu for being my academic advisor and teaching me a lot during these years.

Thanks to Dr. Hiram G. Bezerra, Dr. Marco A. Costa and Dr. Henry Koon for their support during my graduate study.

I am also grateful to Ashok, Philip, Xiaowei and all my friends and classmates in the department of Epidemiology and Biostatistics for their friendship and support.

I would like to thank my parents and younger sister for their support of my goal and aspirations. Finally, my deepest gratitude goes to my wife, Yu and my son, David, for their love, care and support, for helping me become a better husband and father.

Causal Mediation Analysis for Non-linear Models

Abstract

by

WEI WANG

Mediators are intermediate variables in the causal pathway between an exposure and an outcome. Mediation analysis investigates the extent to which exposure effects occur through these variables, thus revealing causal mechanisms. One interesting question in causal inference area is mediation analysis for non-linear models.

In the first part of this dissertation, we consider the estimation of mediation effects in zero-inflated (ZI) models intended to accommodate ‘extra’ zeros in count data. Focusing on the ZI negative binomial (ZINB) models, we provide a mediation formula approach to estimate the (overall) mediation effect in the standard two-stage mediation framework under the key sequential ignorability assumption. We also consider a novel decomposition of the overall mediation effect for the ZI context using a three-stage mediation model. Simulation study results demonstrate low bias of mediation effect estimators and close-to-nominal coverage probability (CP) of confidence intervals. The method is applied to a retrospective cohort study of dental caries in very low birth weight adolescents. For overall mediation effect estimation, sensitivity analysis was conducted to

quantify the degree to which key assumption must be violated to reverse the original conclusion.

The second question we focus on is the mediation analysis for a dichotomous outcome in multiple-mediator models. We formulate a joint model (probit-normal) using continuous latent variables for any binary mediators to account for correlations among multiple mediators. A mediation formula approach is proposed to estimate the total mediation effect and decomposed mediation effects based on this parametric model. We conduct a simulation study that demonstrates low bias of mediation effect estimators for two-mediator models with various combinations of mediator types. The results also show that the power to detect a non-zero total mediation effect increases as the correlation coefficient between two mediators increases, while power for individual mediation effects reaches a maximum when the mediators are uncorrelated. We illustrate our approach by applying it to a retrospective cohort study of dental caries in adolescents with low and high socioeconomic status. Sensitivity analysis is performed to assess the robustness of conclusions regarding mediation effects when the assumption of no unmeasured mediator-outcome confounders is violated.

Estimation of Mediation effects for Zero-inflated Regression Models

1.1 Introduction

A common problem encountered in medical, public health and social studies is the presence of a high number of zeros, a problem known as zero inflation (ZI). For example, the distribution of decayed, missing, filled teeth index (DMFT), a count response variable from an ongoing longitudinal study which compares very low birth weight (VLBW) adolescents with full-term adolescents, contains a large number of zeros (Nelson et al., 2010). Excess zeroes can be categorized as either *sampling* or *structural zeros*. A sampling zero refers to a zero count or an unobserved event, due to a chance zero outcome for an individual with a non-zero probability of an event. In contrast, a structural zero refers to a zero count or unobserved event that is determined (the event being impossible or irrelevant) for an individual (Ridout et al., 1998). For example, to answer the question "How often did you drink alcohol during the last month?" there will be individuals who drink alcohol but chose not to drink during the last 30 days (sampling zeros) and individuals who never drink alcohol (structural zeros).

Zero-inflated models, two-component mixture models combining a point mass at zero with a proper count distribution are appropriate for data containing structural as well as possible sampling zeros. A well-known version, suitable for unbounded counts, is the zero-inflated Poisson regression (ZIP) model (Lambert, 1992); however, this model has the limitation that it may provide an inadequate fit to data when there is overdispersion. A popular model accounting for overdispersion is the negative binomial regression model

(NB) in which the Poisson mean has a gamma distribution. For bounded data, the comparable models include the zero-inflated binomial (ZIB) and the zero-inflated beta-binomial (ZIBB) model. The latter accommodates extra dispersion by assuming the success probability in the binomial model has a beta distribution. Recent medical applications of zero-inflated models are illustrated by Cheung (2002) who applied zero-inflated models to early growth and motor development, and Lewsey and Thomson (2004) who used ZIP and ZINB models to examine the effect of socio-economic status on DMFT. Cheung (2006) also used zero-inflated proportion models (ZIB and ZIBB) to model growth and cognitive function of Indonesian children.

Another type of model used to analyze data with excess structural zeros only is a two-part conditional model, known as the hurdle model (Mullahy, 1986). A hurdle model is a modified count model in which there are two processes, one generating the zeros and the other generating the positive values. For the first process, a model (typically, logistic regression) is used to determine the binary outcome of whether the count response is zero or greater than zero. If the value is positive, "the hurdle is crossed", and the conditional distribution of the positive values (the second process) is governed by a zero-truncated (for example, Poisson or NB) count model. A Swedish health care utilization study used a hurdle model (a logit model for the hurdle and a truncated NB model condition on a positive count) to model physician visit frequencies (Gerdtham, 1997). Similarly, another study evaluated the impact of managed care programs for Medicaid-eligible patients on utilization of health-care services using both parametric (Poisson and NB) and semi-parametric hurdle models (Gurmu, 1997).

Although both ZI models and hurdle models can provide a satisfactory fit for some ZI data (Rose and Martin, 2006), one model type may be more applicable based on the study objectives. ZI models involve a sub-model predicting membership into one of two latent populations, generally referred to the ‘susceptible’ (contributing sampling zeros) and ‘non-susceptible’ populations (contributing structural zeros). In hurdle models, the first stage is to induce an event, and once the hurdle to the first event has been cleared or crossed, the second stage determines the number of subsequent events. The approach for both types of models is similar in that they use two regression models, one (logit or probit regression) models the susceptible probability or probability the hurdle is crossed, and the other (log linear regression for Poisson/NB, logistic regression for binomial or beta-binomial) models the mean for the susceptible population or the count given crossing of the hurdle. Explanatory variables are allowed to have a different impact for the two processes (Moulton et al., 2002), and the two models are fit simultaneously using maximum likelihood estimation.

Extensions of ZI and hurdle regression models for correlated data have begun to appear to study subject-specific and marginal effects. Berk and Lachenbruch (2002) and Min and Agresti (2005) have proposed various random intercept zero-inflated models. Dalrymple et al. (2003) have developed finite mixture zero-inflated models that allow for clustering of subjects based on their latent response trajectories. Dobbie and Welsh (2001) and Hall and Zhang (2004) used generalized estimating equations (GEE) or its extension to estimate hurdle and zero-inflated marginal models, respectively. Several authors have recently proposed alternative Bayesian approaches to fitting zero-inflated models. Rodrigues (2003) developed Bayesian ZIP models for cross-sectional data and Neelon et al. (2010) proposed a Bayesian approach which incorporates prior information for repeated measures zero-inflated count data with application to outpatient psychiatric service use.

Generally, separate exposure effects were estimated for the susceptible probability and susceptible population mean respectively (Rose and Martin, 2006), though inference for the overall mean for the ZI model was also considered. Böhning et al. (1999) suggested two approaches to obtain confidence intervals for the overall mean of the ZIP model for a population. Yau and Lee (2001), considering a ZIP regression model with random effects, provided a confidence interval for the overall mean at a specified set of covariate values. Previously, we proposed two new methods for assessing an overall mean exposure effect in the context of ZI regression models. In the first approach, an 'average predicted value' (APV) method was developed to assess an overall mean exposure effect, which allows covariate-adjusted estimation of flexible functions of exposure group means. The second method presented uses log-linear models for both the binary and the count components of the ZINB model (Albert et al., 2011).

Though mediation analysis may be of interest for zero-inflated as well as others types of (e.g., normally-distributed) responses, this problem does not appear to have been previously addressed in the literature. The main goal of Part I is to develop the mediation analysis method for ZI models, particularly, the ZINB model. Different types of mediators will be considered. In order to accomplish this goal, three aims are proposed:

Aim 1: To develop a mediation analysis method to estimate overall mediation effects using the mediation formula within the standard two-stage framework.

Aim 2: To decompose the indirect effect into two paths: one through a latent variable representing the subpopulation ('susceptible' or 'non-susceptible'), the other going directly from the mediator to the final outcome variable.

Aim 3: To develop an approach for conducting sensitivity analyses that quantify the degree to which key assumptions must be violated to reverse the original conclusion for overall mediation effect estimation.

The rest for Part I is organized as follows. Section 1.2 presents a brief literature review of mediation analysis and also describes ZINB and hurdle models. In section 1.3, we define the natural indirect effect, natural direct effect, total causal effect and decomposition of the overall mediation effect in the ZINB model. Section 1.4 presents a difference in effects estimate of the mediation effect and also the mediation formula approach to estimate the overall mediation effect under given identifying assumptions. In this section we also discuss computation for the mediation formula using a Monte Carlo integration method as an alternative to 'exact' integration. Section 1.5 proposes the mediation formula approach to estimate the decomposition of the overall mediation effect as well as its Monte Carlo integration approximation. Simulation studies are used in Section 1.6 to examine the statistical properties of the proposed methods and to determine the number of samples needed for the Monte Carlo integration approach. Section 1.7 describes the application of the proposed method to the dental data and presents a sensitivity analysis. Discussion and suggestions for further research are presented in Section 1.8.

1.2 Background

1.2.1 Mediation analysis

The target of many empirical studies in the social, behavioral and health sciences is the causal mediation (or indirect) effect, which measures the extent to which the effect of the treatment on the outcome is mediated through some particular pathway. The examination of mediation is important because identification of the extent to which change in one or more mediating variables account for a treatment effect may shed light on the theoretical basis of the same effect, and therefore may help streamline and improve the treatment by focusing on effective components.

1.2.1.1 Definitions of mediators, moderators and confounders

Most research focuses on the relation between two variables, T and Y, in which T denotes treatment and Y denotes outcome variable, and T can be considered a possible cause of Y. Multiple relations may be present when a third variable is included in the analysis of a two-variable system.

Mediators. A mediator, or intervening causal variable, is on the causal pathway between the treatment (T in Figure 1.1) and the outcome (Y in Figure 1.1), and it must be a post-treatment variable that occurs before the outcome is realized. Beyond this minimal requirement, the causal model containing a specific mediator needs to be conceptually plausible (Bauman et al., 2002). There may be a single mediator (M_1) between the treatment and the outcome, or a series of cascading mediators (M_2, M_3) that are causally related in sequence, between the treatment and outcome. It is also possible that some of the effect may be direct (from T to Y, bypassing any mediator). For example, a continuous measure of job search self-efficacy may be the key mediator between the treatment (receiving a job training intervention on unemployed workers) and the binary outcome, whether or not the respondent had become employed (Vinokur and Schul, 1997).

Confounders. A confounder (C) is a predictor of the outcome, but is also associated with treatment, so that ignoring the confounder leads to incorrect inference about the relation between treatment and outcome. For example, cigarette smoking may be a confounder of the relationship between weight and long-term mortality in the Framingham Heart Study (Garrison et al., 1983). Cigarette smoking may be related to the outcome – smokers have higher mortality than nonsmokers and cigarette smoking also associated with the exposure – of men under desirable weight, more than 80% were smokers. The confounder may then account for the association between an exposure and an outcome and thus give a false impression about the causal effect if not adjusted for. Methods for dealing with confounders include matching, stratified analyses, and controlling for them using multivariable analytic techniques (Rothman, 2001).

Moderators. The third variable (B) may also modify the relation of treatment to outcome such that the relation of treatment to outcome differs at different values of B, and this is an example of an effect modifier (or moderator). This corresponds to the concept of statistical interaction, with the association $T \rightarrow Y$ varying across levels of the moderator, B

(Figure 1.1C). For example, Pediatric Traumatic Brain Injury (TBI) may generate less pronounced negative consequence for parents of black children than parents of white children at baseline, but become more pronounced at two years follow-up (Yeates et al., 2002). Moderators can be dealt with by including interaction terms in statistical models used to assess treatment effects, or by stratifying the data by the levels of the moderator, and re-examining effects.

In addition, the third variable (*W*) may be related to *Y* and/or *T*, so that information about *W* improves prediction of *Y* by *T*, but does not substantially alter the relation of *T* to *Y*; this is an example of a covariates (Figure 1.1D). Covariates may be confounding variables if they are related to both *T* and *Y*.

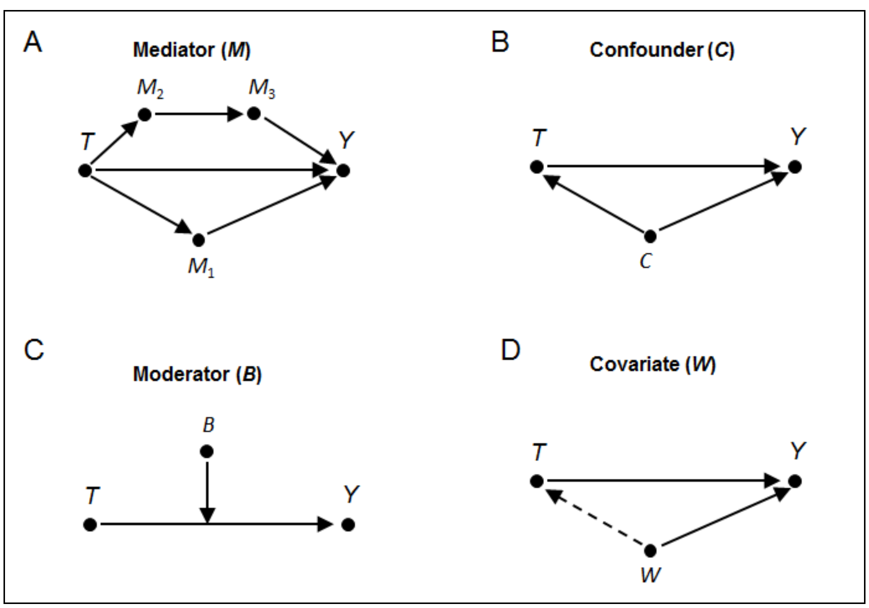

Figure 1.1. Definition of terms. Mediator – lies in the causal path from treatment to outcome. Confounder – affects the both the treatment and the outcome. Moderator – the relation of treatment to outcome differs with different values of moderator. Covariate – may affect the outcome and possibly be related to treatment.

1.2.1.2 Linear structural equation model

The two-stage mediation model is shown in Figure 1.2, where the variables T, M and Y are in rectangles and the arrows represent causal associations among variables. Figure 1.2 uses β_2 to denote the coefficient in the linear relationship between T to M, γ for the relation of M to Y adjusted for T, and β_3 for the relation of T to Y adjusted for M. The symbols ε_2 and ε_3 indicate errors for M and Y, respectively. The coefficients corresponding to Figure 1.2 are discussed below. Note that there is direct effect relating T to Y and a mediated effect by which T indirectly affects Y through M.

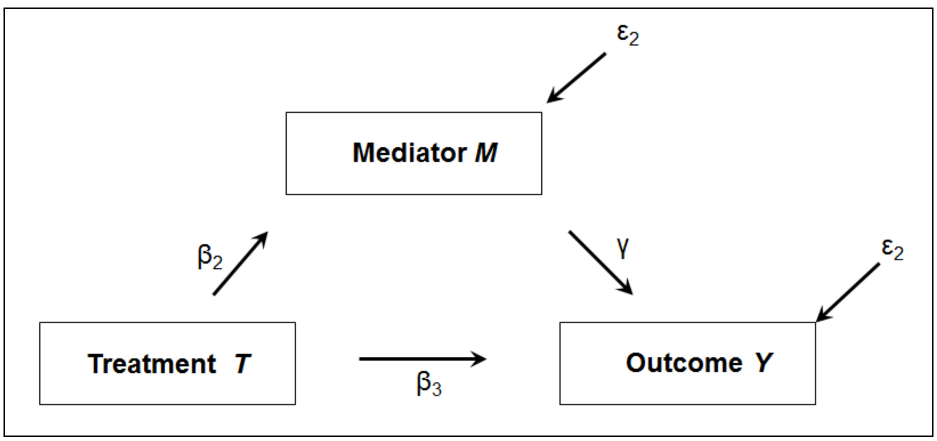

Figure 1.2. Two-stage mediation model

The statistical assessment of the mediation effect has been well developed under the normality assumption. Assuming both the mediator (M) and the outcome (Y) are normally distributed and are related according to the path diagram shown in Figure 1.2, the linear structural equation model (SEM) will be:

$$Y = \alpha_1 + \beta_1 T + \varepsilon_1 \quad (1)$$

$$M = \alpha_2 + \beta_2 T + \varepsilon_2 \quad (2)$$

$$Y = \alpha_3 + \beta_3 T + \gamma M + \varepsilon_3 \quad (3)$$

where $\varepsilon_i \sim N(0, \sigma_i^2)$, $i = 1, 2, 3$, and $\mathrm{Cov}(\varepsilon_2, \varepsilon_3) = 0$.

The most widely used method to assess mediation is the causal steps approach introduced by Baron and Kenny (1986). In general, under the assumptions that there is no unmeasured confounders between the mediator and the outcome, four conditions are involved in Baron and Kenny's approach to establishing mediation.

1. In the regression model of M regressed on T, T must be associated with M ($\beta_2 \neq 0$);
2. In the regression model of Y regressed on T, T must be associated with Y ($\beta_1 \neq 0$);
3. In the regression model of Y regressed on T and M, M must be associated with Y ($\gamma \neq 0$);
4. The regression coefficient of T from model (3) is less than that of model (1) ($\beta_3 < \beta_1$).

When all four of the above conditions hold, one may conclude that M is a mediator. It has been noted that the second condition is not necessary, since the mediation effect can be significant even when the total causal effect is zero. This happens when the mediation effect offsets the direct effect of the treatment. The mediation effect in the above two-stage mediation model (see Figure 1.2 as well as (1), (2) and (3)) may be estimated using two different approaches. The first one is called the difference in coefficients method, which evaluates the mediation effect as the difference in regression coefficients of the outcome on the treatment without and with adjustment for the mediator ($\beta_1 - \beta_3$) (MacKinnon, 1993). A second general method estimates mediation as the product of two coefficients, the coefficient of the mediator in its regression on the treatment, and the coefficient of the mediator in the regression of the outcome on the mediator adjusted for the treatment indicator. This method is called the product of coefficients method ($\beta_2\gamma$) (MacKinnon et al. 2002). The algebraic equivalence of both approaches was shown by Mackinnon et al. (1995) for normal theory ordinary least squares (or maximum likelihood estimation) under the above three mediation regression equations.

The asymptotic standard error of the indirect effect can be derived using the multivariate delta method (Sobel, 1982). Confidence intervals based on the normal distribution for the indirect effect are often inaccurate as found in simulation studies (MacKinnon et al., 1995, 2002; Stone and Sobel, 1990) and from bootstrap analysis of the mediated effect (Bollen and Stine 1990). These mediation effect confidence intervals tend to lie to the left of the true value of the mediation effect for positive mediation effect and to the right for negative mediation effect (Bollen and Stine 1990; MacKinnon et al. 1995; Stone and Sobel, 1990). Asymmetric confidence limits based on the distribution of the product and bootstrap estimation generally have better coverage (MacKinnon et al., 2004).

Often models are proposed that include more than a single mediator in the causal chain between treatment and outcome variables. An example can be found in Allen and Griffeth (2001) in which job performance positively affected employees' perceived employment alternatives, which in turn positively affected their intention to leave (turnover intention), which in turn affected actual turnover. A study of divorced mothers tested the hypothesis that the effect of negative life events on parenting behaviors would be mediated by two variables in turn: psychological distress and avoidant coping (Tein et al., 2000). In the linear case, the product of coefficients approach can be extended to estimate mediation effects in the context of multiple stages (Taylor et al., 2008). In this approach, the three-path mediation effects estimation requires that the following three regression equations be estimated:

$$M_1 = \alpha_1 + \beta_1 T + \varepsilon_1 \tag{4}$$

$$M_2 = \alpha_2 + \beta_2 T + \gamma_1 M_1 + \varepsilon_2 \tag{5}$$

$$Y = \alpha_3 + \beta_3 T + \gamma_2 M_1 + \gamma_3 M_2 + \varepsilon_3 \tag{6}$$

In these equations, Y is the outcome variable, T is the treatment indicator, and M_1 and M_2 are the two sequential mediators. There are a number of different effects of that might be defined using this model. It can be shown that the direct effect of T on Y is β_3 in equation (6), and the total mediation effect of T on Y, the effect passing through either mediator, is $\beta_1\gamma_1\gamma_3 + \beta_1\gamma_2 + \beta_2\gamma_3$. This effect can be broken down into the three-path mediation effect, which is the effect passing through both mediators ($\beta_1\gamma_1\gamma_3$), and the two-path mediation effects, the effects passing through only one of the mediator ($\beta_1\gamma_2$ and $\beta_2\gamma_3$). Another effect that may be studied is the mediation effect passing through one mediator, such as $\beta_1\gamma_1\gamma_3 + \beta_1\gamma_2$ for M_1 and $\beta_1\gamma_1\gamma_3 + \beta_2\gamma_3$ for M_2.

1.2.1.3 Mediation analysis with other types of outcomes and mediators

In some mediation analyses, the dependent variable is categorical, such as whether a new drug cures the patient or not after treatment. In this case, equation (1) and (3) must be rewritten for logistic or probit regression, where the dependent variable may be viewed as a latent continuous variable that has been dichotomized in its observed form. The mediation effects for a binary outcome may be defined using either the risk difference scale or the odds ratio scale (VanderWeele and Vansteelandt, 2010). For the odds ratio scale, there are two commonly used methods for computing mediation effects. Freedman et al. (1992) suggested the difference of coefficients method, while MacKinnon et al. (2007) advocated the product of coefficients method. Advantages of the latter are that it has less bias than the difference in coefficients method. In addition, the product of coefficients method was also shown to be robust against (logistic versus probit) model misspecification, as well as the normality assumption for the distribution of the mediator (MacKinnon et al., 2007). Schluchter (2008) illustrates how generalized estimating equations (GEE) modeling which applies to the class of generalized models, including linear, logistic, and Poisson regression, can be used to estimate the indirect effect, defined as the amount by which the regression coefficients of exposure on outcome changes after adjusting for mediator.

Li et al. (2007) consider the mediation model with a binary mediator and a continuous outcome. In their paper, they consider three types of mediation effect estimators and found that the estimators that account for the binary nature of the mediator (called the adjusted logit and probit estimators in the paper) that are consistent for the mediation effect defined in this paper while other estimators are inconsistent. They further concluded that the crude difference-in-coefficients estimator should be used with caution, and that the product of coefficients estimators should not be used. Huang et al. (2004) studied the situation where both the mediator and outcome are binary variables. Three approaches to the estimation of mediation effects – the delta method (relevant to product of coefficients approach), a bootstrap and a Bayesian modeling approach – were investigated. Their Monte Carlo simulations showed that Bayesian method using a non-informative prior outperformed both the bootstrap and delta methods, particularly for small sample sizes.

1.2.1.4 The potential outcomes framework

Recently, researchers (Robins and Greenland, 1992; Albert, 2008; Imai et al., 2010) have addressed mediation analysis using the potential outcomes framework. This framework allows clear definitions of mediation effects in causal terms and explication of assumptions required for causal inference. Generally, $M(t)$ denotes the potential value of the mediator under the treatment status t, $Y(t, m)$ represents the potential outcome of Y when $T = t$ and $M = m$. Then, $Y(t, M(t'))$ indicates the counterfactual value of Y that would be observed if T was set to t and M was set to value of M that would be observed if T was set to t'.

In the potential outcomes notation, the causal mediation effect in the standard two-stage mediation model under treatment status t can be defined as (Robins and Greenland, 1992; Pearl, 2001):

$$IE(t) \equiv E\{Y(t, M(1))\} - E\{Y(t, M(0))\} \tag{7}$$

for a binary treatment indicator t = 0, 1, Pearl (2001) called $IE(t)$ the *natural indirect effect*, while Robins (2003) used the term the *pure indirect effect* for $IE(0)$ and the *total indirect effect* for $IE(1)$. In words, $IE(t)$ represents the difference between two mean potential outcomes that would result under treatment indicator t, but where the mediator takes a value that would result under different treatment assignments.

Similarly, the *natural direct effect* and the *total causal effect* in the potential outcomes framework can be defined as:

$$DE(t) \equiv E\{Y(1, M(t))\} - E\{Y(0, M(t))\} \tag{8}$$

$$TE \equiv E\{Y(1, M(1))\} - E\{Y(0, M(0))\} \tag{9}$$

It is easy to derive the important relationship $TE = IE(t) + DE(1 - t)$; in words, the total causal effect is equal to the sum of the natural indirect effect under one treatment assignment and the natural direct effect under the other treatment assignment (Imai, 2010; Pearl, 2011).

The natural indirect effect and natural direct effect differ from the *controlled direct effect* of the mediator, that is $Y(t, m)$ - $Y(t, m')$ for t = 0, 1, and that of treatment, which is $Y(1, m)$ - $Y(0, m)$ (Pearl, 2001 and 2011). In contrast to the natural direct and indirect effects, the controlled direct effects of the mediator are defined in terms of specific values of the mediator, m and m', rather than its potential value, $M(t)$. In this sense, the natural indirect effect examines whether M mediates the causal relationship between T and Y, whereas the controlled direct effect of mediator investigates whether T moderates the causal effect of M on Y (Baron and Kenny, 1986; Imai, 2010).

1.2.1.5 Mediation analysis with mediation formula approach

There are necessary assumptions for the identifiability of the causal mediation effects, and the key assumption made is that is that of ‘sequential ignorability’ which consists of no unmeasured confounding as well as the requirement that no mediator-outcome confounder be affected by exposure (Peterson et al., 2006). The assumption sets proposed by Imai et al. (2010) are defined as:

Assumption 1 (Sequential Ignorability):

$$\{Y(t', m), M(t)\} \perp\!\!\!\perp T \mid W = w \tag{10}$$

$$Y(t', m) \perp\!\!\!\perp M(t) \mid W = w, T = t \tag{11}$$

Thus, the treatment is first assumed to be independent of potential outcomes and mediators given the baseline covariates, and then the mediator variable is assumed to be independent of potential outcomes given observed values of treatment assignment and the baseline covariates. Slight variants of this set of assumptions have been proposed by other authors (Pearl, 2001; Petersen et al. 2006; van der Laan and Petersen, 2008; VanderWeele, 2010). Imai et al. (2010) compared their identifying assumption with those proposed in the literature. Shpitser and VanderWeele (2011) show that Imai’s assumptions (2010) are equivalent to Pearl’s (2001) assumptions in the sense that if either set of assumptions holds for all models inducing a particular causal diagram, then the other set of assumptions will also hold for all models inducing that diagram.

The following theorem shows that the potential outcome means necessary for the estimation of mediation effect under the above assumptions are nonparametrically identified.

Theorem 1.Under Assumption 1, for continuous covariate *w* and binary treatment *t:*

$$E(Y(t, M(t'))) = \iint E\{Y \mid M = m, T = t, W = w\} dF_{M|T=t',W=w}(m) dF_W(w) \tag{12}$$

The proof is given in Appendix I. A similar proof is shown by Imai et al. (2010) and Shpitser and VanderWeele (2011). This theorem shows that under sequential ignorability the distribution of the required potential outcomes can be expressed as a function of the distribution of the observed data. Thus, the assumption lets us make inferences about the counterfactual quantities that are not observed. This approach to estimating causal effects is employed by numerous authors (Baron and Kenny, 1986; Li et al., 2007; Imai, 2010; Pearl, 2011), although its applicability in the context of the linear structural equation model is not always recognized. Pearl (2011) denotes the expression in Theorem 1 as the 'mediation formula'; this formula basically expresses a mean potential outcome of *Y* as an integral of the conditional mean of *Y* over the probability density distribution of the mediators under the identifiability assumption described above.

Since Theorem 1 is not based on any specific model, it can be used to develop a general estimation procedure for causal mediation effects under linear as well as various nonlinear conditions. Imai et al. (2010) proposed a general approach based on the mediation formula idea which can accommodate continuous and discrete mediators, and various types of outcome variables. The application of this approach to mediation effect estimation for a binary outcome using the risk difference scale was discussed in their paper. Imai et al. (2010) also developed a sensitivity analysis in the context of commonly used models, which enables researchers to formally assess the robustness of their empirical conclusion to violations of the key assumption.

Based on the mediation formula, Albert and Nelson (2011) presented a method applicable to multiple stages of mediation and mixed variable types using generalized linear models. They define pathway effects using a potential outcomes framework and present a general formula that provides the effect of exposure through any specified pathway. As shown by Avin et al. (2005), not all pathway effects will be identifiable without additional assumptions, in particular, the effect of a path from T to Y is nonidentifiable if and only if there is a path from some mediator M to Y that is activated (the response Y were set to the value it would have if M set to value it would have were the individual exposed) while another path from M to Y is deactivated (the response Y were set to the value it would have if M set to value it would have were the individual not exposed), and the path from T to M is activated (the mediator M were set to the value it would have were the individual exposed). A sensitivity analysis to assess the impact of the additional assumption was also proposed in this paper.

1.2.2 Statistical Models for Zero-inflated Response

We consider models for a count outcome y based on the zero-inflated negative binomial (ZINB) and negative binomial hurdle (NBH) models. In the ZINB model, two latent subpopulations can be defined, a susceptible population with responses distributed as negative binomial (mean λ) and a non-susceptible group with responses fixed at zero. The mixture probability, denoted as ψ, is the probability of being in the susceptible population. In the NBH model, the cross-hurdle probability is denoted as ψ, and conditional on crossing the hurdle, a positive outcome is distributed as truncated negative binomial (mean λ). The negative binomial distribution has mass function given by

$$nb(y;\lambda,\phi) = \frac{\Gamma(y+1/\phi)}{y!\Gamma(1/\phi)}\left(\frac{1}{1+\phi\lambda}\right)^{1/\phi} \times \left(\frac{\lambda}{1/\phi+\lambda}\right)^{y} \tag{13}$$

where ϕ is the dispersion parameter, and the mean and variance are λ and $\lambda(1 + \phi\lambda)$ respectively.

The ZINB model has a probability density function given by

$$P(y;\psi,\lambda,\phi) = \begin{cases} (1-\psi) + \psi\, nb(y;\lambda,\phi) & \text{for } y = 0 \\ \psi\, nb(y;\lambda,\phi) & \text{for } y > 0 \end{cases} \tag{14}$$

with mean and variance given by $\psi\lambda$ and $\psi\lambda(1 + \lambda(1 - \psi + \phi))$.

Similarly, the NBH model has a density given by

$$P(y;\psi,\lambda,\phi) = \begin{cases} 1-\psi & \text{for} \quad y = 0 \\ \dfrac{\psi\, nb(y;\lambda,\phi)}{1-p_0} & \text{for} \quad y > 0 \end{cases} \tag{15}$$

where p_0 is equal to $nb(0; \lambda, \phi)$. The mean and variance for the NBH are given by $\dfrac{\psi\lambda}{1-p_0}$ and $\dfrac{\psi\lambda}{1-p_0} \times \left\{ \dfrac{1+\phi\lambda+\lambda-p_0-\phi\lambda p_0-\lambda p_0-\psi\lambda-\psi^2\lambda}{1-p_0} + \dfrac{\psi^2\lambda}{(1-p_0)^2}\left[1+nb(0;\lambda,\phi)\right] \right\}$ respectively. The derivation of mean and variance of ZINB and NBH model is shown in Appendix II.

As ϕ approaches zero, the ZINB and NBH models approach the zero-inflated Poisson (ZIP) and Poisson hurdle (PH) models respectively. The non-zero (crossing hurdle) probability (ψ) and the (truncated) negative binomial mean (λ) may be modeled with a logistic regression model and a log-linear regression model, i.e. logit(ψ) = $\boldsymbol{\alpha}'\boldsymbol{x}$ and ln(λ) = $\boldsymbol{\beta}'\boldsymbol{x}$. Because the count responses for all subjects are assumed to be independent, the log-likelihood of the ZINB and NBH models are defined by summing the log of the expressions in equation (14) and (15) respectively over all the subjects. The resulting log-likelihood is then maximized simultaneously to estimate all parameters in the model. In our study, estimation was implemented using the adaptive Gaussian quadrature algorithm in the SAS (Version 9.2, SAS Institute Inc., Cary, NC, USA) PROC NLMIXED procedure.

1.3 Defining Overall and Decomposition of Meditation Effects

We now provide definitions and notation relevant for our proposed methods.

1.3.1 Defining causal mediation effects for zero-inflated regression models

Consider the general casual model including a binary exposure or treatment indicator (T), a mediator (M) and a zero-inflated distributed outcome (Y), where T may affect Y directly or T may affect M, which then affects Y. Figure 1.3 provides the path diagram. To define the causal mediation effects, we use the potential outcomes framework. Under the standard two-stage mediation model considering ZI count outcome, the natural indirect effect, natural direct effect and total causal effect in ZI regression model can also be defined as equation (7), (8) and (9) (listed above). Here, we consider t = 1 only and denote $IE(1)$ as IE and $DE(0)$ as DE; for t = 0, method will be similar.

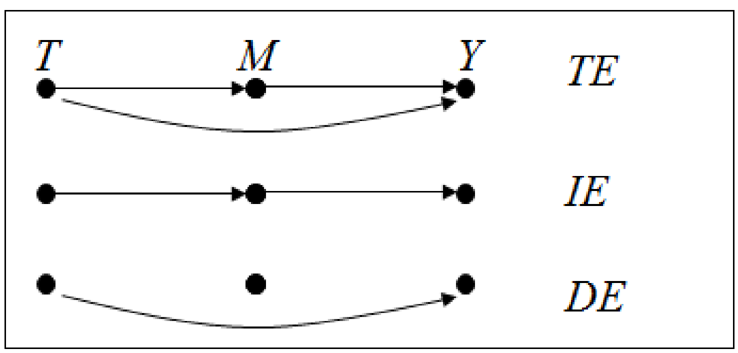

Figure 1.3. Path diagram in a two-stage mediation model. T = exposure or treatment, M = mediator, Y = outcome; TE = total casual effect, IE = natural indirect effect, DE = natural direct effect.

1.3.2 Defining decomposition of overall natural indirect effect

In ZI models, it is possible to further dissect the total natural indirect effect. We do this by introducing a latent variable Y_1 indicating whether the subject belongs to the 'susceptible'

(Y_1 = 1) or 'non-susceptible' populations (Y_1 = 0). Thus, our approach conceives of 'susceptibility' as manipulable, rather than a fixed trait. Incorporating this latent variable, which we refer to as the 'susceptibility indicator', we obtain an extended causal model involving, in causal order, the variables (T, M, Y_1, Y), and thus three stages of mediation (Figure 1.4). This model allows us to partition the overall total natural indirect effect into two paths, through or not through Y_1. We will follow the notation used by Albert and Nelson (2011) in their three-stage mediation framework. In our ZI model, the potential outcome of Y given manipulation D, with corresponding exposure indicator variables d = (d_0, d_1, d_2, $d_{1,2}$), can be written as $Y(d_0, M(d_1), Y_1(d_2, M(d_{1,2})))$, where $Y_1(d_2, M(d_{1,2}))$ is the potential outcome for Y_1, were exposure set to d_2, and M set to the value it would take were exposure set to $d_{1,2}$.

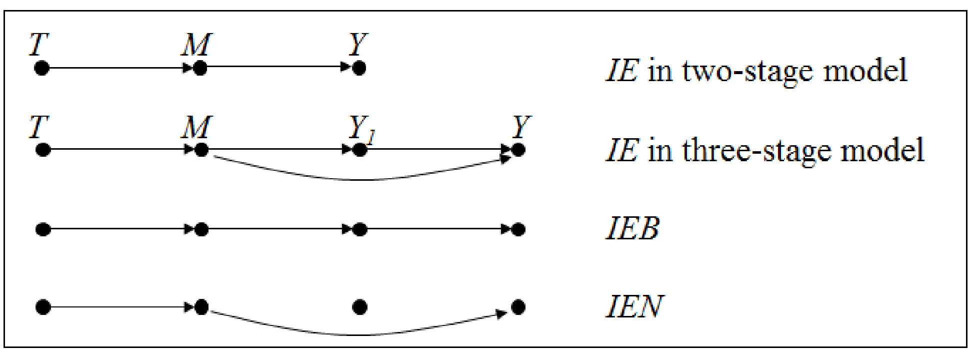

Figure 1.4. Path diagram for decomposition of overall mediation effect in a three-stage model. Y_1 = susceptible group indicator; *IEB* = natural indirect effect through susceptible group indicator (Y_1), *IEN* = natural indirect effect not through susceptible group indicator (Y_1). *T, M, Y,* and *IE* are shown as Figure 1.3.

The total natural indirect effect can be broken into two path-specific effects, one path effect (denoted '*IEB*', meaning **I**ndirect **E**ffect by affecting **B**inary component of

ZINB model) is through the susceptibility indicator Y_1; the other is the direct effect of M on Y (denoted as '*IEN*', meaning **I**ndirect **E**ffect by affecting **N**B component of ZINB model directly), that is, the part of the mediation effect not through Y_1 (Figure 1.4). *IEB* and *IEN* are each defined as the difference between two mean potential outcomes as follows,

$$IEB \equiv E\{Y(1, M(0), Y_1(1, M(1)))\} - E\{Y(1, M(0), Y_1(1, M(0)))\} \quad (16)$$

$$IEN \equiv E\{Y(1, M(1), Y_1(1, M(1)))\} - E\{Y(1, M(0), Y_1(1, M(1)))\} \quad (17)$$

Specifically, *IEB* represents the difference between two expected potential outcomes in which the exposure status (T = 1 indicating exposed) and the mediator potential outcome ($M(0)$) affecting the NB component of the ZI model is kept consistent, while the mediator affecting the binary component of the ZI model takes values that would result under different exposure statuses ($M(1)$ vs. $M(0)$). Note that the sum of *IEB* and *IEN* is equal to *IE*. An alternative decomposition of the overall mediation effect (where *IEB* is defined as difference in the means of $Y(1, M(1), Y_1(1, M(1)))$ and $Y(1, M(1), Y_1(1, M(0)))$, and *IEN* is defined as the difference in the means of $Y(1, M(1), Y_1(1, M(0)))$ and $Y(1, M(0), Y_1(1, M(0)))$) can be handled similarly and thus is not discussed here.

1.4 Analysis of Overall Mediation Effect for Zero-Inflated Models

We begin by presenting approaches for assessing the overall mediation effect. For comparison, we illustrate an ad hoc ‘difference in effects’ method. Following that we derive total and decomposition of mediation effect estimators utilizing the mediation formula.

1.4.1 Difference in effects approach

Before defining the mediation effect for the ZINB model, we present an ad hoc estimator of causal effects using an extension of the popular difference in coefficients idea, which we refer to as a ‘difference in effects’ approach. In this approach, two ZINB models, each involving two component models, are fit in order to estimate *IE* and *DE*. The first model does not adjust for the mediator and is written as:

$$\text{logit}(\psi) = \zeta_0 + \zeta_1 t + \boldsymbol{\zeta}'\boldsymbol{w} \qquad \ln(\lambda) = \xi_0 + \xi_1 t + \boldsymbol{\xi}'\boldsymbol{w} \tag{18}$$

where $\psi \equiv E(Y_1 \mid t, \boldsymbol{w})$, $\lambda \equiv E(Y \mid t, \boldsymbol{w}, Y_1 = 1)$, ζ_0, ζ_1, ξ_0 and ξ_1 are unknown parameters, $\boldsymbol{\zeta}$ and $\boldsymbol{\xi}$ are unknown parameter vectors, t is the exposure indicator (equal to 1 if exposed, 0 otherwise), and $\boldsymbol{w}$ is a vector of observed covariate values for an individual with response y. The overall mean of ZINB distributed response y for this individual (covariate $\boldsymbol{w}$) will be $\text{logit}^{-1}(\zeta_0 + \zeta_1 + \boldsymbol{\zeta}'\boldsymbol{w})\exp(\xi_0 + \xi_1 + \boldsymbol{\xi}'\boldsymbol{w})$ if this individual is exposed and $\text{logit}^{-1}(\zeta_0 + \boldsymbol{\zeta}'\boldsymbol{w})\exp(\xi_0 + \boldsymbol{\xi}'\boldsymbol{w})$ if not exposed based on formula provided in Section 1.2.2. From this model, we define the overall exposure effect as the total causal effect TE_i for an individual with covariate vector $\boldsymbol{w}$ (over the two components of the ZI model) (Albert et al., 2011)

$$TE_i = \text{logit}^{-1}(\zeta_0 + \zeta_1 + \boldsymbol{\zeta}'\boldsymbol{w})\exp(\xi_0 + \xi_1 + \boldsymbol{\xi}'\boldsymbol{w}) - \text{logit}^{-1}(\zeta_0 + \boldsymbol{\zeta}'\boldsymbol{w})\exp(\xi_0 + \boldsymbol{\xi}'\boldsymbol{w}) \tag{19}$$

The second model adjusts for the mediator and is written as:

$$\text{logit}(\psi) = \alpha_0 + \alpha_1 t + \alpha_2 m + \boldsymbol{\alpha}'\boldsymbol{w} \qquad \ln(\lambda) = \beta_0 + \beta_1 t + \beta_2 m + \boldsymbol{\beta}'\boldsymbol{w} \tag{20}$$

where the mediator m is adjusted and included in both components of models. The overall exposure effect from this model provides the direct effect DE_i (of exposure on response) for an individual with covariance vector $\boldsymbol{w}$:

$$\begin{aligned} DE_i = {} & \text{logit}^{-1}(\alpha_0 + \alpha_1 + \alpha_2 m + \boldsymbol{\alpha}'\boldsymbol{w})\exp(\beta_0 + \beta_1 + \beta_2 m + \boldsymbol{\beta}'\boldsymbol{w}) \\ & - \text{logit}^{-1}(\alpha_0 + \alpha_2 m + \boldsymbol{\alpha}'\boldsymbol{w})\exp(\beta_0 + \beta_2 m + \boldsymbol{\beta}'\boldsymbol{w}) \end{aligned} \tag{21}$$

The difference in the individual total and direct effects provides the individual indirect effect. Corresponding population effects are defined as averages over the subjects in the designated population (sample or subsample). Estimators for these effects are obtained by substituting the maximum likelihood estimators for the corresponding regression parameters obtained from the appropriate fitted model in the above expressions. This ad hoc method is an extension of the popular difference in coefficients approach (MacKinnon and Dwyer, 1993; MacKinnon et al., 2007). Note that different types of mediators (for example, continuous or discrete) are accommodated in this approach.

1.4.2 Estimation of overall mediation effect with mediation formula approach

We start by presenting identification results for the (overall) indirect and direct effects, defined by (7) and (8) using the potential outcomes framework described above. Under assumption 1 ((10) and (11)), the natural indirect effect, natural direct effect and total causal effect can be calculated respectively as:

$$IE = \iint E\left\{Y \mid M = m, T = 1, W = w\right\}\left(dF_{M|T=1,W=w}(m) - dF_{M|T=0,W=w}(m)\right) dF_W(w) \tag{22}$$

$$DE = \iint \left(E\left\{Y \mid M = m, T = 1, W = w\right\} - E\left\{Y \mid M = m, T = 0, W = w\right\}\right) dF_{M|T=0,W=w}(m) dF_W(w) \tag{23}$$

$$TE = \iint \left(E\left\{Y \mid M = m, T = 1, W = w\right\} dF_{M|T=1,W=w}(m) - E\left\{Y \mid M = m, T = 0, W = w\right\} dF_{M|T=0,W=w}(m)\right) dF_W(w) \tag{24}$$

Summation should be used in place of integration in the case of discrete mediators and/or discrete covariates.

Integrations can be approximated using Monte Carlo integration, which averages the integrand over randomly generated realizations of the assumed distribution of the mediator (James, 1980). Suppose that $m_1 \ldots m_S$ and $m_1' \ldots m_S'$ are randomly generated from the assumed distribution of $M(0)$ and $M(1)$ respectively. Formulae (22) and (23) can then be approximated by the following expressions:

$$IE \approx \frac{1}{N}\sum_{j=1}^{N}\left(\frac{1}{S}\sum_{i=1}^{S} E\{Y \mid M = m_i', T = 1, W = w_j\} - \frac{1}{S}\sum_{i=1}^{S} E\{Y \mid M = m_i, T = 1, W = w_j\}\right) \tag{25}$$

$$DE \approx \frac{1}{N}\sum_{j=1}^{N}\left(\frac{1}{S}\sum_{i=1}^{S} E\{Y \mid M = m_i, T = 1, W = w_j\} - \frac{1}{S}\sum_{i=1}^{S} E\{Y \mid M = m_i, T = 0, W = w_j\}\right) \tag{26}$$

To carry out estimation of the direct and indirect effects in (22) and (23) we need a model for the regression of Y on M, T, and W. For the present application of interest, we will consider the following zero-inflated model. The model for Y, allowing the mediator (M) to affect both the susceptibility indicator (Y_1) and the response Y given susceptible is that given in (20). For the mediator, we consider appropriate regression models for the binary and continuous mediator cases (allowing the mediator to depend on the exposure variable and baseline covariates ($\boldsymbol{W}$)) specifically, we use the following regression models for a binary and continuous mediator, respectively,

$$\text{logit}(m) = \gamma_0 + \gamma_1 t + \boldsymbol{\gamma}'\boldsymbol{w} \tag{27}$$

$$m = \gamma_0 + \gamma_1 t + \boldsymbol{\gamma}'\boldsymbol{w} + \varepsilon, \quad \textit{where } \varepsilon \sim N(0, \sigma^2) \tag{28}$$

1.4.3 Sensitivity analysis

For the identified overall mediation effect in our potential outcomes framework, the quantity cannot be given a causal interpretation without the particular assumption set (10) and (11). Assumption (10) is satisfied in a randomized treatment study or cohort study, where randomization probabilities may be a function of baseline covariates, $\boldsymbol{w}$. Assumption (11) may not hold if there exists unmeasured confounding for the mediator-outcome relationship. We examine the effect of violation of assumption (11) on estimation of the overall indirect and direct effects. To calculate the necessary joint distribution (possibly involving both continuous and discrete variables) we use the approach of Albert and Nelson (2011) which applies the Gaussian copula (Song et al., 2009).

We propose a sensitivity analysis based on a general model for the joint probability of $Y(t', m)$ and $M(t)$. The count outcome Y has $K + 1$ possible values (0, 1, 2, …K), where K is the assumed upper limit for Y. $P_{Y(t',m)}(y) = P\{Y(t', m) \leq y\}$ and $P_{M(t)}(m) = P\{M(t) \leq m\}$ will denote the cumulative distribution functions of Y and M. We suppose that there are, corresponding to $Y(t', m)$ and $M(t)$, latent variables, denoted as $Y^*(t', m)$ and $M^*(t)$ respectively, which are marginally distributed as standard normal and satisfy the relationship $Y^*(t', m) = \Phi^{-1}\{P_{Y(t',m)}(y)\}$ and $M^*(t) = \Phi^{-1}\{P_{M(t)}\}$, where Φ is the standard normal distribution function. In addition, $Y^*(t', m)$ and $M^*(t)$ are assumed bivariate normally distributed with correlation ρ, for all t, t' and m. In order to properly handle the discrete nature of the distributions of Y, we propose a Monte Carlo approach similar to that used by Albert and Nelson (2011) to compute the conditional distributions $P\{Y(t', m_i) = j, j = 1, 2, \ldots, K \mid M(t) = m_i\}$ for possible values of m_i. Specifically, we use the following algorithm:

(1) Sample a mediator $m_i \sim f_{M(t)}(m)$, $i = 1, 2, \ldots, n$.

(2) For a continuous m_i, get $U_1 = \Phi^{-1}\{P_{M(t)}(m_i)\}$. For a binary m_i, if $m_i = 0$ then draw a uniform variate u_1 from the interval $[0, P_{M(t)}(0)]$, if $m_i = 1$ then draw a uniform variate u_1 from the interval $(P_{M(t)}(0), 1]$, let $U_1 = \Phi^{-1}\{u_1\}$.

(3) Draw a variate $U_2 \sim N(\rho U_1, 1 - \rho^2)$, let $u_2 = \Phi(U_2)$.

(4) Let $C_{ji} = 1$ if $u_2 \in (P_{Y(t',mi)}(j-1), P_{Y(t',mi)}(j)]$, and $C_{ji} = 0$, otherwise, for $j = 1, 2, \ldots, K$ (with the subscript i indicating the conditioning on m_i).

(5) Repeat steps 2 – 4 with independent draws a large number of (say R) times obtaining C_{jir} in the rth replicate for $r = 1, 2, 3, \ldots, R$.

(6) Following the R replications for each m_i, we estimate the conditional probability $P\{Y(t', m_i) = j \mid M(t) = m_i\}$ as $\frac{\sum_r C_{jir}}{R}$ for j, $j = 1, 2, \ldots, K$.

(7) Repeat steps 1 – 6 n times, we can obtain an estimate of $E\{Y(t', M(t))\}$ as $\frac{1}{n}\sum_{i=1}^{n}\left(\sum_{j=1}^{K} jP\{Y(t', m_i) = j \mid M(t) = m_i\}\right)$.

Therefore, supposing that the correlation ρ between $Y^*(t', m)$ and $M^*(t)$ is given, the overall natural indirect effect and natural direct effect are identified and given by

$$IE = \frac{1}{n}\left\{\sum_{i=1}^{n}\left(\sum_{j=1}^{K} jP\left\{Y\left(1, m_i'\right) = j \mid M\left(1\right) = m_i'\right\}\right) - \sum_{i=1}^{n}\left(\sum_{j=1}^{K} jP\left\{Y\left(1, m_i\right) = j \mid M\left(0\right) = m_i\right\}\right)\right\} \tag{29}$$

$$DE = \frac{1}{n}\left\{\sum_{i=1}^{n}\left(\sum_{j=1}^{K} jP\left\{Y\left(1, m_i\right) = j \mid M\left(0\right) = m_i\right\}\right) - \sum_{i=1}^{n}\left(\sum_{j=1}^{K} jP\left\{Y\left(0, m_i\right) = j \mid M\left(0\right) = m_i\right\}\right)\right\} \tag{30}$$

where m_i and m_i' are randomly generated from the assumed distribution of $M(0)$ and $M(1)$ as shown in formula (25) and (26). These estimates can be recomputed using the above algorithm over varying values for ρ to provide a sensitivity analysis.

1.5 Decomposition of the Causal Effect for Zero-Inflated Models

1.5.1 Identifying decomposition path effects

To identify the path effects of the decomposed overall indirect effect ((16) and (17) given in Section 1.3.2), we extend the preceding sequential ignorability assumption to the three-stage causal model case. The three-stage sequential ignorability assumption is given as

Assumption 2:

$$\{Y(t'', m', y_1), Y_1(t', m), M(t)\} \perp\!\!\!\perp T \mid W = w \tag{31}$$

$$\{Y(t'', m', y_1), Y_1(t', m)\} \perp\!\!\!\perp M(t) \mid W = w, T = t \tag{32}$$

$$Y(t', m', y_1) \perp\!\!\!\perp Y_1(t, m) \mid W = w, T = t, M = m \tag{33}$$

Thus, each intermediate variable in the model is assumed to be ignorable (that is independent of subsequent potential outcomes) given all preceding variables in the causal model.

1.5.2 Estimation of decomposition of total mediation effect for the ZINB Model

To demonstrate identifiability of the decomposition path effects, we will examine identification of the relevant potential outcome means (of the general form $E(Y(d_0, M(d_1), Y_1(d_2, M(d_{1,2}))))$) given baseline covariate $\boldsymbol{w}$ under sequential ignorability (Assumption 2). Under Assumption 2 it can be shown that,

$$E\{Y(d_0, M(d_1), Y_1(d_2, M(d_{1,2})))\} = \iint E\{Y \mid T = d_0, M = m, Y_1 = 1\} E\{Y_1 \mid T = d_2, M = m'\} d^2 F_{M(d_1), M(d_{1,2})}(m, m') \tag{34}$$

As before, integration will be interpreted as summation in the case of a discrete mediator. Note that in general the expectations and distribution in (34) will be conditional on baseline covariates $\boldsymbol{w}$; however, this conditioning is left out of the expression, and those below, for brevity. When d_1 and $d_{1,2}$ are unequal (thus, equal to 1 and 0 or vice versa), expression (34) shows that we require the joint distribution of $M(1)$ and $M(0)$ in order to estimate $E(Y(d_0, M(d_1), Y_1(d_2, M(d_{1,2}))))$. As seen in equations (16) and (17), both *IEB* and *IEN* involve an expected potential outcome in which d_1 does not equal $d_{1,2}$ – namely, $(E\{Y(1, M(0), Y_1(1, M(1)))\})$; d_1 equals $d_{1,2}$ in the other potential outcomes ($E\{Y(1, M(0), Y_1(1, M(0)))\}$ and $E\{Y(1, M(1), Y_1(1, M(1)))\}$). Generally the joint distribution of $M(1)$ and $M(0)$ is unknown and cannot be estimated because both of these outcomes cannot be observed for the same subject. For identifiability we need to make an untestable assumption regarding the joint distribution of these two potential outcomes. For instance, in the continuous mediator case, we may assume that $M(1)$ and $M(0)$ are bivariate normally distributed with specified correlation coefficient τ. For a non-normally distributed mediator, we can specify joint distribution of these two counterfactuals ($M(1)$ and $M(0)$) using the copula method discussed above. Under this assumption in addition to Assumption 2, the expected potential outcome $E(Y(d_0, M(d_1), Y_1(d_2, M(d_{1,2}))))$ when $d_1 \neq d_{1,2}$ can be identified and computed using formula (34). The nonidentifiability of some pathway effects in three (or more) stage mediation is also mentioned by Avin et al. (2005) and Albert and Nelson (2011).

When d_1 equals $d_{1,2}$, $E(Y(d_0, M(d_1), Y_1(d_2, M(d_1))))$ can be estimated by,

$$E\{Y(d_0, M(d_1), Y_1(d_2, M(d_1)))\} = \int E\{Y \mid T = d_0, M = m, Y_1 = 1\} E\{Y_1 \mid T = d_2, M = m\} dF_{M(d_1)}(m) \qquad (35)$$

A proof of formulae (34) and (35) under assumptions (31) – (33) is given in Appendix III.

Similar expressions using Monte Carlo integration can be used to estimate the components of the overall mediation effect (that is, *IEB* and *IEN*),

$$IEB \approx \frac{1}{n}\left(\sum_{i=1}^{n} E\{Y \mid T=1, M=m_i, Y_1=1\} E\{Y_1 \mid T=1, M=m_i'\} - \sum_{i=1}^{n} E\{Y \mid T=1, M=m_i, Y_1=1\} E\{Y_1 \mid T=1, M=m_i\}\right) \qquad (36)$$

$$IEN \approx \frac{1}{n}\left(\sum_{i=1}^{n} E\{Y \mid T=1, M=m_i', Y_1=1\} E\{Y_1 \mid T=1, M=m_i'\} - \sum_{i=1}^{n} E\{Y \mid T=1, M=m_i, Y_1=1\} E\{Y_1 \mid T=1, M=m_i'\}\right) \qquad (37)$$

where (m_i, m_i') are drawn from the joint distribution $f_{M(0), M(1)}(m, m')$.

1.6 Simulation Study

In this section, we conduct a simulation study to compare natural indirect effect estimators from the difference in effects and mediation formula approaches for a binary as well as a continuous mediator. For the continuous mediator case, we also determine the number of samples needed for our Monte Carlo integration method to achieve acceptable mean squared error. In addition, we consider analogous scenarios in which data are simulated under the NBH model and investigate the robustness under this model of estimates assuming the ZINB model. Finally, we want to assess the effect of the correlation (τ) between the (assumed bivariate normal) counterfactuals $M(1)$ and $M(0)$ on the estimation of the mediation effects through susceptibility indicator Y_1 (IEB), and not through Y_1 (IEN).

1.6.1 Comparison of difference in effects and mediation formula approach

In our first simulation study, we studied the mediation effect (IE) assuming a ZINB model with different types of mediators. The logistic regression model for the susceptible probability and log-linear model for the susceptible population mean both include a binary exposure indicator (T = 1 if exposed, 0, otherwise), a common categorical covariate, W (constrained so that each exposure group had a 50% frequency of $w = 1$) and a mediator variable, M (either binary or continuous). The model is thus given as (20) above with $\boldsymbol{\alpha}' = \alpha_3$ and $\boldsymbol{\beta}' = \beta_3$ where α_3 and β_3 are unknown scalar coefficients for covariate w. Binary and continuous mediators were modeled as (27) and (28), and the same balanced covariate W was included in these two models. The other parameter that needs to be specified is the negative binomial dispersion parameter ϕ, and we chose 0.5 for all our simulation scenarios.

We considered nine scenarios which are distinguished in the magnitude of the natural indirect effect, IE, (corresponding to parameters α_2, β_2 and γ_1) and of the natural direct effect, DE, (corresponding to parameters α_1 and β_1). The scenarios were specified as: (1) large direct effect and zero indirect effect; (2) large direct effect and small indirect effect; (3) large direct effect and large indirect effect; (4) small direct effect and zero indirect effect; (5) small direct effect and small indirect effect; (6) small direct effect and large indirect effect; (7) zero direct effect and zero indirect effect; (8) zero direct effect and small indirect effect; (9) zero direct effect and large indirect effect. In the continuous mediator scenario, two standard deviation values for the mediator were considered, small σ (equal to 0.5) and large σ (equal to 2.5).

For each of the above scenarios and type of mediators, 1000 simulated datasets were generated. Sample sizes of 200 (100 per exposure group) and 1000 (500 per exposure group) were used. The exposure indicator and covariate were generated independently for individuals within each dataset using the pseudorandom number generator function 'RAND' in SAS/IML. For each given exposure and covariate, the mediator variable was generated using equation (27) and (28). The response variates were then generated independently according to the ZINB distribution with regression model (20) given the generated individual exposure, covariate and mediator variables.

The true natural indirect effect is defined by the function on the right hand side of (22), with true coefficients in place of the estimates. Summation should replace the integral calculation in equation (22) when dealing with discrete mediators or covariates. For each generated dataset, the difference in effects method and the mediation formula approach were used to calculate the estimated *IE*, and 95% confidence intervals for each approach were constructed with percentile estimates from 1000 bootstrap samples. From the simulations, we calculated the average estimate of *IE*; the average percent error (PE = 100 $\times$ (Average Estimated *IE* – True *IE*)/true *IE*) of *IE*, a measure of relative bias; the SD of estimated *IE*; the average estimated SE of estimated *IE*; the coverage probability (CP, percent of simulated datasets for which 95% confidence interval for *IE* covered the true value).

Table 1.1 gives results for the binary mediator case with total sample size 200 and 1000 respectively. With sample size 200, we see over all nine scenarios that the mediation formula approach produces a small bias in its estimation of *IE*, and the average PE is less than 2.1% for all scenarios. The coverage probabilities of 95% confidence intervals deviate from the nominal level for four of nine scenarios by around 3%. The difference in effects approach gives an *IE* estimator which is almost as good as that of the mediation formula approach when either true *DE* or *IE* is zero, but overestimates the *IE* (around 30%) when a large *DE* exists. When the sample size per group is increased to 500 (second part of Table 1.1), the mediation formula approach also shows very low bias (less than 2%) and the coverage probability gets closer to nominal rate compared with the smaller sample size scenarios. In contrast, the difference in effects approach still overestimates the *IE* in non-zero *IE* scenarios, and the overestimation increases as *DE* increases.

We also considered the continuous mediator variables case, and corresponding simulation results are shown in Table 1.2. Both methods can give good estimators in zero *IE* scenarios. In the non-zero *IE* scenarios, the mediation formula approach provides relative biases of less than 9% for all scenarios with n = 100 per group, and less than 6% for all scenarios with n = 500 per group for both small and large mediator SD scenarios. Bootstrap 95% confidence intervals are conservative (with greater than 99% coverage probability) for zero-*IE* scenarios with small mediator SD, but are good for all scenarios (within 2% of nominal level) when the mediator SD increases from 0.5 to 2.5. For non-zero *IE* scenarios, the performance of difference in effects approach is similar to that in the binary case; specifically, it overestimates the *IE*, and the overestimation increases as *DE* increases.

1.6.2 Determination of number of samples for Monte Carlo integration

In a second simulation study, we determined the number of Monte Carlo samples (S) needed for a good approximation of the integration in formula (22) when the mediator variable is continuous in the model shown in (20) and (28). The same $9\times2\times2$ scenarios (9 different magnitudes of *IE* and *DE*, 2 selected mediator standard deviations (σ), and 2 sample sizes: 200 and 1000) were considered as above. *IE* estimators from 1000 simulated datasets using the mediation formula approach with ‘exact’ integration (22) as

Table 1.1. Simulation statistics for the natural indirect effect of binary mediator using mediation formula approach and difference in effects approach on data generated from ZINB/logit-log model.

Sample Size	True *DE* Est	True *IE* Est	Mediation Formula approach: Ave Est *IE*	Ave PE (%)	SD of Est *IE*	AVE SE	CP (%)	Difference in Effects approach: Ave Est *IE*	Ave PE (%)	SD of Est *IE*	AVE SE	CP (%)
100	-0.766	0.000	-0.001	-	0.044	0.048	98.0	-0.003	-	0.083	0.095	98.1
	-0.888	-0.263	-0.265	0.61	0.117	0.121	94.3	-0.339	28.90	0.165	0.170	95.0
	-1.243	-0.753	-0.755	0.29	0.268	0.262	93.1	-0.978	29.94	0.323	0.321	90.2
	-0.265	0.000	-0.001	-	0.059	0.066	97.8	-0.005	-	0.093	0.103	98.5
	-0.308	-0.241	-0.242	0.34	0.126	0.131	95.1	-0.261	8.35	0.167	0.171	96.0
	-0.319	-0.773	-0.773	0.04	0.270	0.256	91.8	-0.833	7.71	0.270	0.268	95.0
	0.000	0.000	-0.002	-	0.065	0.075	97.7	-0.003	-	0.091	0.108	98.7
	0.000	-0.276	-0.277	0.14	0.145	0.149	95.3	-0.274	-0.92	0.172	0.182	95.7
	0.000	-0.710	-0.725	2.02	0.245	0.249	95.0	-0.723	1.76	0.257	0.265	95.4
500	-0.766	0.000	-0.001	-	0.017	0.018	94.2	-0.001	-	0.033	0.034	94.1
	-0.888	-0.263	-0.264	0.38	0.050	0.051	95.1	-0.340	29.02	0.071	0.072	82.9
	-1.243	-0.753	-0.753	0.02	0.111	0.111	95.2	-0.971	28.91	0.137	0.137	63.8
	-0.265	0.000	0.000	-	0.024	0.025	95.4	-0.001	-	0.038	0.038	94.3
	-0.308	-0.241	-0.240	-0.53	0.056	0.055	95.2	-0.256	6.03	0.074	0.073	95.0
	-0.319	-0.773	-0.771	-0.32	0.112	0.109	94.7	-0.827	6.99	0.117	0.115	92.4
	0.000	0.000	0.000	-	0.027	0.028	95.5	-0.001	-	0.038	0.040	95.9
	0.000	-0.276	-0.272	-1.49	0.062	0.062	93.7	-0.271	-1.99	0.078	0.077	95.0
	0.000	-0.710	-0.707	-0.55	0.105	0.105	95.1	-0.708	-0.33	0.112	0.114	95.2

Table 1.2. Simulation statistics for the natural indirect effect of continuous mediator using mediation formula approach and difference in effects approach on data generated from ZINB/logit-log model.

				Mediation Formula approach					Difference in Effects approach				
Sample Size	Sigma	True *DE* Est	True *IE* Est	Ave Est *IE*	Ave PE (%)	SD of Est *IE*	AVE SE	CP (%)	Ave Est *IE*	Ave PE (%)	SD of Est *IE*	AVE SE	CP (%)
100	0.5	-0.690	0.000	-0.001	-	0.023	0.036	99.6	-0.001	-	0.050	0.083	99.9
		-0.789	-0.197	-0.207	5.31	0.205	0.218	95.1	-0.257	30.68	0.237	0.254	95.6
		-1.194	-0.734	-0.775	5.47	0.394	0.384	92.7	-0.978	33.24	0.401	0.389	91.0
		-0.240	0.000	-0.001	-	0.032	0.048	99.9	-0.003	-	0.061	0.087	99.5
		-0.274	-0.280	-0.304	8.38	0.280	0.292	93.9	-0.311	10.86	0.275	0.283	94.9
		-0.340	-0.720	-0.748	3.93	0.398	0.410	93.2	-0.783	8.74	0.365	0.368	94.4
		0.000	0.000	-0.000	-	0.027	0.047	100.0	-0.002	-	0.045	0.080	99.9
		0.000	-0.148	-0.150	0.87	0.282	0.282	93.6	-0.136	-8.40	0.268	0.273	93.0
		0.000	-0.642	-0.668	4.07	0.365	0.389	95.3	-0.653	1.66	0.317	0.332	94.5
	2.5	-0.777	0.000	0.000	-	0.069	0.077	96.1	0.003	-	0.111	0.124	97.4
		-0.817	-0.234	-0.238	1.97	0.158	0.157	94.5	-0.294	25.67	0.240	0.234	94.9
		-1.074	-0.749	-0.721	-3.84	0.497	0.515	95.1	-0.803	7.10	0.962	0.868	95.6
		-0.269	0.000	0.001	-	0.095	0.100	95.6	0.003	-	0.133	0.137	96.5
		-0.280	-0.250	-0.245	-1.71	0.191	0.196	93.8	-0.256	2.43	0.274	0.258	92.4
		-0.383	-0.798	-0.788	-1.22	0.606	0.645	94.9	-0.812	1.78	1.092	0.980	93.6
		0.000	0.000	-0.001	-	0.111	0.111	94.6	-0.000	-	0.139	0.142	96.8
		0.000	-0.227	-0.209	-8.08	0.207	0.217	94.8	-0.213	-6.22	0.272	0.274	94.8
		0.000	-0.726	-0.753	3.70	0.579	0.595	95.0	-0.727	0.16	0.996	0.830	93.4
500	0.5	-0.690	0.000	0.000	-	0.006	0.008	99.3	0.000	-	0.013	0.017	99.3
		-0.789	-0.197	-0.201	2.05	0.093	0.089	92.9	-0.262	33.18	0.109	0.105	89.4
		-1.194	-0.734	-0.736	0.17	0.159	0.157	94.0	-0.965	31.39	0.167	0.166	70.6
		-0.240	0.000	0.000	-	0.010	0.011	98.3	0.000	-	0.015	0.019	98.7
		-0.274	-0.280	-0.278	-0.68	0.116	0.119	95.9	-0.298	6.45	0.114	0.118	96.2
		-0.340	-0.720	-0.720	0.01	0.161	0.164	96.3	-0.773	7.41	0.149	0.152	94.9
		0.000	0.000	0.000	-	0.006	0.009	99.7	0.000	-	0.009	0.015	99.9
		0.000	-0.148	-0.157	5.93	0.118	0.118	94.5	-0.154	3.83	0.113	0.113	94.5
		0.000	-0.642	-0.650	1.17	0.166	0.158	94.0	-0.649	1.06	0.145	0.139	93.9
	2.5	-0.777	0.000	-0.001	-	0.032	0.031	93.5	-0.001	-	0.047	0.049	95.5
		-0.817	-0.234	-0.235	0.50	0.063	0.067	96.0	-0.287	22.86	0.101	0.104	93.8
		-1.074	-0.749	-0.748	-0.19	0.225	0.222	94.5	-0.860	14.70	0.475	0.428	92.9
		-0.269	0.000	0.002	-	0.040	0.041	95.0	0.001	-	0.055	0.055	94.4
		-0.280	-0.250	-0.252	1.02	0.086	0.084	94.2	-0.265	6.16	0.116	0.114	94.5
		-0.383	-0.798	-0.810	1.55	0.267	0.277	94.6	-0.836	4.78	0.486	0.478	94.8
		0.000	0.000	-0.000	-	0.044	0.046	95.6	-0.002	-	0.057	0.058	94.7
		0.000	-0.227	-0.222	-2.03	0.095	0.093	95.1	-0.220	-3.04	0.122	0.121	94.0
		0.000	-0.726	-0.732	0.77	0.240	0.251	94.9	-0.715	-1.53	0.432	0.412	94.5

well as Monte Carlo integration (25) were calculated. The selected S used for Monte Carlo integration were 1, 2, 5, 10, 20, 50, 100, 200, 500, 1000, 2000, 5000, and the quality of the Monte Carlo approximation was assessed by two simulation statistics, CP and normalized mean squared error which divide the mean squared error (MSE) from Monte Carlo by that from 'exact' integration, where MSE = $\frac{1}{n}\sum_{i=1}^{1000}(Estimated\ IE_i - True\ IE)^2$.

Normalized MSE and CP were plotted against S in Figure 1.5 for 16 out of 36 scenarios considered above (scenarios with small *IE* or small *DE* were omitted for graphic presentation purpose). Panels A, B, C and D show the normalized MSE and panels E, F, G and H show the CP. Straight lines y = 1.0 and y = 95.0 are plotted in each set for reference. From these plots, we can see that both normalized MSE and CP decrease as the number of Monte Carlo samples increases, and large σ scenarios (B and D) generally produce a larger normalized MSE compared with the corresponding small σ scenarios (A and C). When S is equal to 100, the normalized MSE is less than 1.07 (less than 10% error relative to the 'exact' integration results) for all small σ scenarios, but as high as 1.45 for some large σ scenarios (Figure 1.5D, large *DE* and *IE* scenarios, marked with '■'). As the number of samples increases to 500, the normalized MSE is less than 1.02 for small σ scenarios and less than 1.09 for large σ scenarios. In addition, S = 100 for small σ and S = 500 for large σ also generate acceptable CPs in most scenarios. An exception is the zero-*IE* with small σ scenarios (Figure 1.5E '●' and '○' scenarios and 1.5G '○' scenario), for which SEs tend to be overestimated with the bootstrap method and

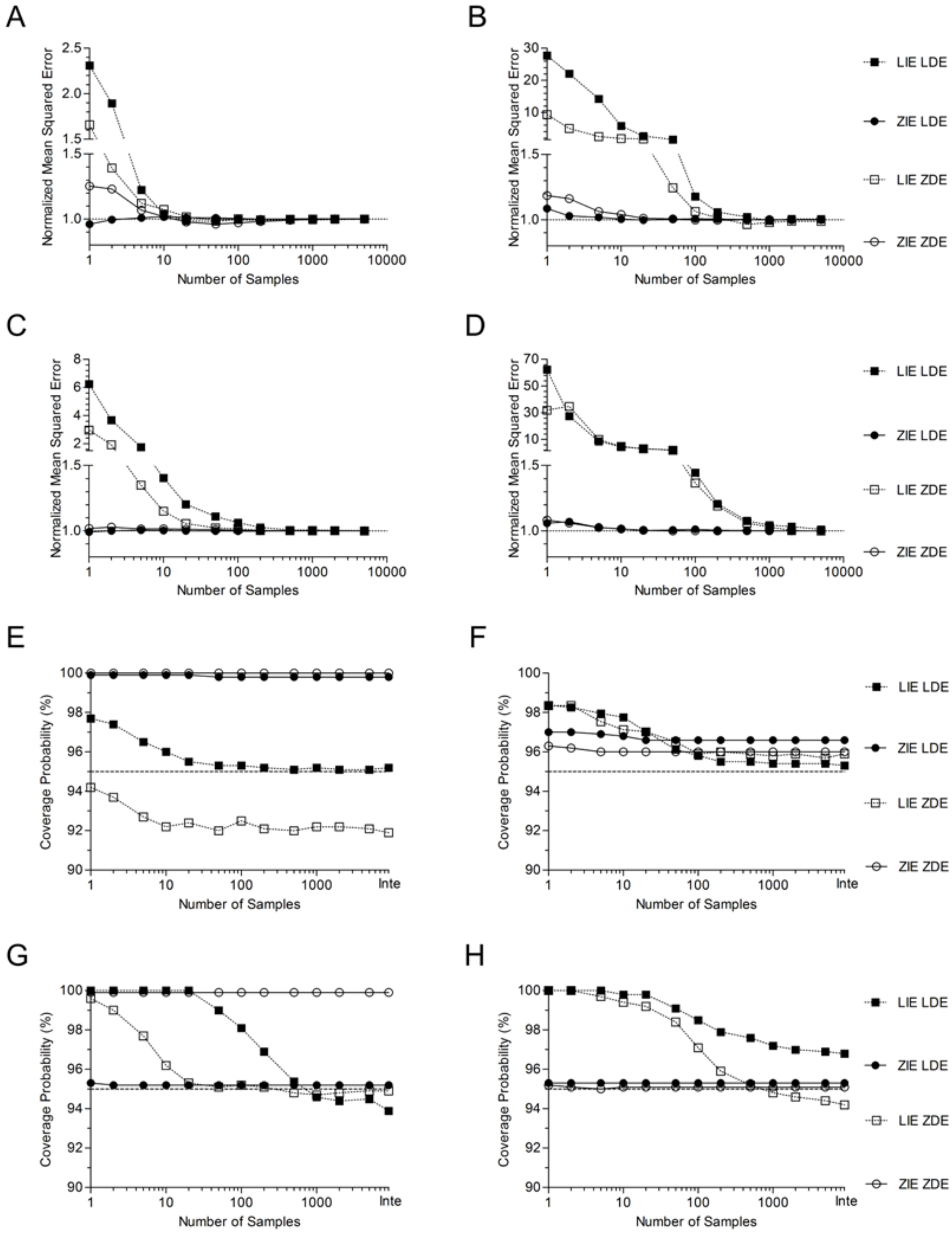

Figure 1.5. Normalized mean squared error (A, B, C, D) and 95% CI coverage probability (E, F, G, H) as function of number of Monte Carlo samples in simulation studies. A and E correspond to the first set of simulation scenarios in Table 1.2 (small σ 0.5, and small sample size n = 100 in each group), B and F correspond to the second set of simulation scenarios in Table 1.2 (large σ 2.5, and small sample size n = 100 in each group), C and G correspond to the third set of simulation scenarios in Table 1.2 (small σ 0.5,

and large sample size n = 500 in each group), D and H correspond to the last set of simulation scenarios in Table 1.2 (large σ 2.5, and large sample size n = 500 in each group). ZIE = **Z**ero **I**ndirect **E**ffect, LIE = **L**arge **I**ndirect **E**ffect. ZDE = **Z**ero **D**irect **E**ffect, LDE = **L**arge **D**irect **E**ffect.

conservative 95% confidence intervals obtained. This problem remains even for *S* as high as 5000 and also exists in the integration approach.

We also recorded the time required for integration calculation ('exact' integration using the 'QUAD' function in SAS/IML), and for Monte Carlo integration with different *S*'s (sampling the mediator using 'RAND' function in SAS/IML) when estimating *IE* with 200 simulation data sets and sample size 200. We found that the time used by the 'QUAD' function is comparable to Monte Carlo integration with *S* = 500 (16.0 seconds for 'QUAD' function vs. 16.2 seconds for Monte Carlo per dataset). When *S* is small, the time used for Monte Carlo does not decrease much (*S* = 1 used 14.8 seconds per dataset), and as *S* increases to 5000, the time consumption for Monte Carlo (32.7 seconds per dataset) doubled compared with 'exact' integration.

1.6.3 Robustness of ZINB model against NBH model in mediation analysis

In a third simulation study, we assumed a NBH model. Each NBH response was generated in two-step process: the first step used a logistic regression model to determine whether the outcome 'crosses the hurdle'; if yes, a truncated negative binomial distribution was used to model the final outcome. As before, we considered both binary and continuous mediators with the same simulation scenarios as in Section 1.6.1. In this study, two *IE* estimators – from the ZINB and NBH models – were compared on the

Table 1.3. Simulation statistics for the natural indirect effect of continuous mediator using NBH/logit-log and ZINB/logit-log model on data generated from NBH/logit-log model.

				NBH/logit-log Model					ZINB/logit-log Model				
Sample Size	Sigma	True *DE* Est	True *IE* Est	Ave Est *IE*	Ave PE (%)	SD of Est *IE*	AVE SE	CP (%)	Ave Est *IE*	Ave PE (%)	SD of Est *IE*	AVE SE	CP (%)
100	0.5	-0.712	0.000	0.001	-	0.024	0.038	99.9	0.000	-	0.024	0.038	99.8
		-0.809	-0.202	-0.214	6.15	0.217	0.228	94.9	-0.218	7.67	0.218	0.227	94.9
		-1.207	-0.740	-0.753	1.84	0.390	0.376	93.2	-0.758	2.49	0.397	0.383	92.9
		-0.243	0.000	-0.000	-	0.034	0.050	99.7	-0.000	-	0.034	0.050	99.7
		-0.276	-0.284	-0.290	2.40	0.292	0.291	94.1	-0.295	3.98	0.293	0.291	94.0
		-0.341	-0.721	-0.743	3.15	0.406	0.397	92.8	-0.753	4.51	0.415	0.402	92.5
		0.000	0.000	0.000	-	0.028	0.049	99.9	0.000	-	0.028	0.049	99.9
		0.000	-0.150	-0.148	-1.46	0.285	0.286	94.6	-0.154	2.57	0.287	0.286	94.2
		0.000	-0.647	-0.668	3.11	0.394	0.380	93.0	-0.679	4.87	0.403	0.385	92.8
	2.5	-0.792	0.000	0.001	-	0.073	0.077	97.3	0.001	-	0.074	0.078	97.1
		-0.831	-0.237	-0.238	0.05	0.154	0.159	95.0	-0.244	2.74	0.158	0.164	94.6
		-1.091	-0.761	-0.711	-6.57	0.451	0.475	95.5	-0.769	0.97	0.490	0.527	96.1
		-0.267	0.000	-0.003	-	0.094	0.099	96.2	-0.003	-	0.095	0.099	96.2
		-0.279	-0.250	-0.242	-3.34	0.184	0.189	94.3	-0.248	-0.75	0.190	0.194	94.1
		-0.383	-0.803	-0.804	0.16	0.581	0.599	95.5	-0.870	8.42	0.633	0.667	95.8
		0.000	0.000	0.004	-	0.104	0.110	96.8	0.004	-	0.105	0.112	96.5
		0.000	-0.226	-0.225	-0.20	0.208	0.210	95.0	-0.231	2.28	0.213	0.217	94.7
		0.000	-0.731	-0.648	-11.38	0.489	0.523	94.7	-0.708	-3.09	0.540	0.592	96.0
500	0.5	-0.712	0.000	-0.000	-	0.007	0.009	99.3	-0.000	-	0.007	0.009	99.3
		-0.809	-0.202	-0.205	1.49	0.097	0.093	94.2	-0.204	1.07	0.097	0.093	94.2
		-1.207	-0.740	-0.739	-0.06	0.156	0.157	94.8	-0.738	-0.21	0.159	0.160	94.6
		-0.243	0.000	-0.000	-	0.009	0.011	98.7	-0.000	-	0.009	0.011	98.6
		-0.276	-0.284	-0.287	1.25	0.118	0.122	94.6	-0.288	1.59	0.119	0.123	95.0
		-0.341	-0.721	-0.727	0.93	0.160	0.168	95.3	-0.732	1.59	0.163	0.171	95.3
		0.000	0.000	0.000	-	0.006	0.010	100.0	0.000	-	0.006	0.010	100.0
		0.000	-0.150	-0.145	-3.05	0.126	0.121	93.5	-0.147	-2.07	0.127	0.122	93.4
		0.000	-0.647	-0.648	0.06	0.157	0.159	94.6	-0.653	0.89	0.160	0.161	94.6
	2.5	-0.792	0.000	0.000	-	0.031	0.032	95.0	0.000	-	0.031	0.032	95.0
		-0.831	-0.237	-0.236	-0.58	0.068	0.067	94.7	-0.239	0.48	0.069	0.068	94.5
		-1.091	-0.761	-0.709	-6.94	0.204	0.208	93.9	-0.747	-1.86	0.217	0.221	94.4
		-0.267	0.000	-0.001	-	0.040	0.041	94.1	-0.001	-	0.040	0.041	94.1
		-0.279	-0.250	-0.244	-2.43	0.082	0.083	94.6	-0.247	-1.35	0.083	0.084	94.6
		-0.383	-0.803	-0.756	-5.76	0.254	0.257	95.6	-0.791	-1.43	0.268	0.272	95.4
		0.000	0.000	-0.000	-	0.045	0.046	96.0	-0.000	-	0.045	0.046	96.0
		0.000	-0.226	-0.222	-1.72	0.094	0.092	93.6	-0.225	-0.56	0.095	0.093	93.7
		0.000	-0.731	-0.682	-6.76	0.232	0.229	95.4	-0.726	-0.65	0.248	0.247	95.9

same simulation statistics listed above. We focus on results for the continuous mediator case, given in Table 1.3; results for the binary mediator case are similar and therefore not presented. Although the interpretation of ZI and hurdle models is different, when the real data came from NBH model, the *IE* estimators from both models are very similar and have low relative bias (less than 12%) for all scenarios considered. The bias of *IE* estimators from ZINB model in most scenarios is slightly higher than that from NBH when the latter is the true model, but the difference is trivial.

1.6.4 Estimation of decomposed mediation effects (*IEB* and *IEN*)

Finally, we studied the estimators of decomposed natural indirect effect that go through the susceptibility indicator (*IEB*), and directly (not through the susceptibility indicator, *IEN*). In this study we assumed that $M(1)$ and $M(0)$ are bivariate normally distributed with correlation coefficient τ. We considered 6 scenarios, all with large natural indirect effect (*IE*), but different magnitudes of the natural direct effect (*DE*) and natural indirect effects through susceptible probability (*IEB*). The scenarios were listed as: (1) small *DE* and small *IEB*; (2) small *DE* and medium *IEB*; (3) small *DE* and large *IEB*; (4) large *DE* and small *IEB*; (5) large *DE* and medium *IEB*; (6) large *DE* and large *IEB*. In each scenario, the potential outcome means for the true *IEB* and *IEN*, was computed using 'exact' integration formula (34) and (35) with true coefficients, and for the calculation of estimated *IEB* and *IEN*, the Monte Carlo integration formula (36) and (37) with corresponding coefficient estimates were used. Here we used 10,000 Monte Carlo samples (draws from the joint distribution of $M(1)$ and $M(0)$ with specified correlation coefficient τ). Three sets of *IEB* and *IEN* were calculated corresponding to three values for τ (0.0, 0.4 and 0.9). From the simulations, we computed the average estimate of *IEB* and *IE*, and then obtained the proportion *IE* accounted for by *IEB*, defined as ($IEB/IE \times 100$). The bias (mean estimate – true value) and the CP for two estimators, *IEB* and proportion of *IEB*, are given in Table 1.4. The true *IEB* magnitude decreases but the true *IE* stays constant as the specified correlation coefficient τ between $M(1)$ and $M(0)$ increases. Not surprisingly, the *IEB* estimator assuming the correct τ produces lower bias than *IEB* estimators assuming the wrong τs. When calculating the *IEB* estimator with mis-specified τ, scenarios with similar magnitude of *IEB* and *IEN* (scenario 2 and 5 in each set) always produce much higher bias compared with scenarios in which either *IEB* or *IEN* predominates. The average estimate of the *IEB* proportion is sometimes far off the true value when the sample size is small. This may be explained by the fact that the total indirect effect was relatively small in some data sets for these scenarios so that the estimate of the *IEB* proportion could be rather unstable. When sample size is increased to 500 per group, the *IEB* proportion estimators are more stable, and correct specification of correlation coefficient τ improves the estimation of *IEB* proportion. Furthermore, scenarios 2 and 5 (true *IEB* proportion around 50%) also produce higher bias than other scenarios for the *IEB* proportion estimation.

Table 1.4. Simulation statistics for the natural indirect effect decomposition with mediation formula approach for continuous mediator on data generated from ZINB/logit-log model, standard deviation of mediator and dispersion parameters are all 0.5.

Sample Size	τ (M(1), M(0))	True *DE*	True *IE*	True *IEB*	True Prop of *IEB*	Est *IE*	τ (M(1), M(0)) = 0.0				τ (M(1), M(0)) = 0.4				τ (M(1), M(0)) = 0.9			
							IEB		Prop of *IEB* (%)		*IEB*		Prop of *IEB* (%)		*IEB*		Prop of *IEB* (%)	
							Bias	CP (%)	Bias	CP (%)	Bias	CP (%)	Bias	CP (%)	Bias	CP (%)	Bias	CP (%)
100	0.0	-0.113	-0.780	-0.057	7.3	-0.843	-0.010	94.6	29.4	97.6	-0.007	94.6	21.7	97.7	-0.004	94.8	17.2	97.8
		-0.092	-0.828	-0.533	64.4	-0.826	0.009	94.3	-0.7	96.0	0.040	93.5	-4.0	96.7	0.079	91.8	-8.1	94.5
		-0.117	-0.775	-0.742	95.8	-0.768	0.007	94.6	51.7	97.4	0.010	95.1	51.6	97.3	0.014	95.3	39.7	97.4
		-1.371	-0.741	-0.053	7.1	-0.784	-0.007	94.0	-59.1	96.6	-0.005	93.9	-52.6	96.7	-0.002	93.9	-38.0	96.9
		-1.171	-0.817	-0.530	64.8	-0.839	-0.006	94.0	-1.5	96.7	0.028	93.6	-5.2	96.1	0.072	91.8	-9.8	93.7
		-0.857	-0.761	-0.735	96.5	-0.766	0.010	94.7	14.2	96.9	0.016	94.6	15.4	96.7	0.022	94.2	15.2	96.7
	0.4	-0.113	-0.780	-0.052	6.7	-0.830	-0.022	93.7	-36.2	96.8	-0.019	93.9	-33.8	97.0	-0.016	93.9	-29.8	97.6
		-0.092	-0.828	-0.501	60.5	-0.842	-0.034	95.2	2.1	95.8	-0.003	95.3	-1.5	96.6	0.038	94.6	-5.6	96.4
		-0.117	-0.775	-0.738	95.2	-0.789	0.003	95.3	5.8	97.4	0.008	95.3	6.6	97.3	0.014	95.3	7.5	97.5
		-1.371	-0.741	-0.047	6.4	-0.777	0.005	94.2	-36.7	97.2	0.005	94.4	-32.5	97.3	0.005	94.2	-27.9	97.3
		-1.171	-0.817	-0.495	60.6	-0.836	-0.041	95.1	2.8	94.6	-0.007	94.7	-1.0	95.6	0.037	94.1	-5.8	95.4
		-0.857	-0.761	-0.731	96.0	-0.752	0.006	93.9	11.7	96.3	0.009	94.1	12.6	96.4	0.012	93.5	14.0	96.2
	0.9	-0.113	-0.780	-0.046	5.9	-0.825	0.005	95.7	-17.8	98.2	0.005	95.7	-15.9	98.2	0.006	95.6	-13.7	98.4
		-0.092	-0.828	-0.460	55.5	-0.822	-0.057	95.3	8.3	92.6	-0.027	95.4	5.1	94.7	0.012	95.0	1.0	96.4
		-0.117	-0.775	-0.733	94.6	-0.789	-0.011	95.7	13.4	97.6	-0.006	95.6	14.3	97.4	-0.001	95.2	15.5	97.6
		-1.371	-0.741	-0.040	5.4	-0.800	-0.035	94.7	2.8	96.6	-0.031	94.6	1.7	96.6	-0.025	94.6	1.2	96.7
		-1.171	-0.817	-0.452	55.3	-0.837	-0.089	95.0	9.3	91.0	-0.054	95.6	5.4	94.8	-0.010	95.4	0.6	96.6
		-0.857	-0.761	-0.726	95.4	-0.760	-0.003	94.9	12.8	97.3	0.001	95.0	13.8	97.0	0.006	94.9	15.4	96.9
500	0.0	-0.113	-0.780	-0.057	7.3	-0.783	0.002	95.7	-3.7	95.4	0.006	95.7	-4.0	95.4	0.012	95.5	-4.3	95.1
		-0.092	-0.828	-0.533	64.4	-0.834	-0.005	96.4	0.2	95.7	0.028	95.0	-3.7	93.1	0.069	88.1	-8.6	79.7
		-0.117	-0.775	-0.742	95.8	-0.780	-0.000	95.4	0.8	95.8	0.004	95.3	0.4	96.0	0.009	94.7	-0.1	95.5
		-1.371	-0.741	-0.053	7.1	-0.736	0.004	94.3	-4.1	94.5	0.008	94.7	-4.3	94.3	0.014	94.9	-4.6	94.6
		-1.171	-0.817	-0.530	64.8	-0.825	-0.006	95.4	0.1	95.7	0.029	93.9	-4.1	92.2	0.073	87.3	-9.4	71.3
		-0.857	-0.761	-0.735	96.5	-0.762	0.003	95.8	0.7	94.8	0.007	95.7	0.3	94.8	0.012	96.1	-0.2	94.7
	0.4	-0.113	-0.780	-0.052	6.7	-0.795	-0.010	94.9	-2.5	95.4	-0.005	95.2	-2.8	95.1	0.002	95.2	-3.2	95.0
		-0.092	-0.828	-0.501	60.5	-0.839	-0.037	94.4	3.6	90.8	-0.004	94.4	-0.2	95.4	0.038	91.5	-5.1	90.1
		-0.117	-0.775	-0.738	95.2	-0.774	-0.002	95.2	1.8	95.5	0.003	94.9	1.5	95.8	0.008	94.5	1.1	95.8
		-1.371	-0.741	-0.047	6.4	-0.753	-0.011	94.9	-1.9	95.3	-0.006	95.2	-2.2	95.1	0.001	95.1	-2.7	95.1
		-1.171	-0.817	-0.495	60.6	-0.818	-0.033	94.3	3.9	90.1	0.001	94.7	-0.2	94.1	0.045	91.0	-5.5	88.8
		-0.857	-0.761	-0.731	96.0	-0.766	-0.005	95.3	1.1	95.7	-0.001	95.1	0.7	95.3	0.004	95.4	0.2	95.2
	0.9	-0.113	-0.780	-0.046	5.9	-0.789	-0.021	94.7	-0.8	94.7	-0.016	94.8	-1.1	94.7	-0.009	94.9	-1.6	94.8
		-0.092	-0.828	-0.460	55.5	-0.824	-0.072	92.1	9.0	76.2	-0.040	94.0	5.3	88.0	-0.000	94.9	0.5	94.7
		-0.117	-0.775	-0.733	94.6	-0.771	-0.005	94.3	2.9	93.3	-0.001	94.4	2.6	93.6	0.004	94.1	2.3	93.8
		-1.371	-0.741	-0.040	5.4	-0.759	-0.014	94.6	-1.9	94.5	-0.009	94.6	-2.2	94.8	-0.002	94.5	-2.6	94.5
		-1.171	-0.817	-0.452	55.3	-0.820	-0.077	90.3	9.1	76.1	-0.042	93.0	4.9	89.1	0.002	93.8	-0.3	95.3
		-0.857	-0.761	-0.726	95.4	-0.770	-0.012	95.5	1.6	94.9	-0.008	95.7	1.2	95.0	-0.002	95.4	0.7	95.2

1.7 Application

Our motivating example comes from a cohort study investigating the effect of VLBW (possibly with bronchopulmonary dysplasia) on dental caries in adolescence (Nelson et al., 2010). The subjects were previously recruited in a cohort study that followed them from birth and assessed various developmental outcomes (Singer, 1997). The study involved 224 infants (139 VLBW and 85 normal term), and the dental outcomes (including enamel defects, oral health behavior, and dental caries) were assessed at around age 14. The normal term group was selected in order to obtain similar distributions to the VLBW group for key baseline covariates, such as race, socioeconomic status (SES), and sex. The exposure variable is the binary variable referred to as "birth group", namely, VLBW (birth group = 1) and normal term (birth group = 0). The outcome considered in this example is the DMFT (decayed, missing, and filled teeth) count. We also considered, in separate models, the following potential mediators: "Sealant" (a binary variable indicating use of sealants), and average "Oral Hygiene Index" (AvgOHI), a bounded continuous variable. The primary study found that the VLBW group had a lower mean DMFT than the normal term group, which may be explained by the facts that VLBW children might be receiving more extensive dental care. We wanted to test this hypothesis and explore whether the effect of VLBW on DMFT is explained by its effect on sealant use or on AvgOHI using the proposed methods discussed above.

The dental dataset with complete model variables (exposure group, DMFT, Sealant, AvgOHI, baseline covariates SES, race and sex) gave us a final sample size of 203 subjects which are used for our analysis. We note that the observed average DMFT (SD) was 2.42 (2.94) for the normal term group (n = 78) and 1.67 (2.74) for VLBW group (n = 125). The models (20), (27) and (28) were fitted using maximum likelihood estimation. For the computations using the mediation formula, we assumed that AvgOHI follows a truncated normal distribution with values between 0 and 3, and mean depending on exposure status and covariate values for each individual. Three baseline covariates (SES, race and sex) were included in all three models, so that parameters were $\boldsymbol{\alpha}' = (\alpha_3, \alpha_4, \alpha_5)$, $\boldsymbol{\beta}' = (\beta_3, \beta_4, \beta_5)$, and $\boldsymbol{\gamma}' = (\gamma_3, \gamma_4, \gamma_5)$. Estimates of *IE*, *DE*, and *TE*, using both the difference in effects and mediation formula approaches are provided in Table 1.5. From this table, we see that both approaches provide similar causal effect estimates and nearly zero estimated mediation effect through either of the two potential mediators (Sealant and AvgOHI). The results are consistent with our simulation study for scenarios where *DE* predominates in the *TE*, in which case the difference in effects approach gave low bias, similarly to the mediation formula approach. From the analysis of the AvgOHI mediator by the mediation formula approach, we draw the following conclusions. When the birth group changes from normal term to VLBW, the mean DMFT decreases an estimated 0.67 (total causal effect). This decrease can be decomposed into an estimate change of 0.03 (-0.14, 0.19) attributable to the AvgOHI and an estimated change of -0.70 (-1.47, 0.06) attributable to other (unknown) pathways. Of note, in our example, *TE* is not statistically significant, which is consistent with previous findings indicating a non-significant exposure effect under alternative ZI models (Albert et al., 2011). With respect to the decomposition of the overall indirect effect, under the assumption of a moderate correlation ($\tau = 0.4$) between the two mediator counterfactuals, for the Sealant mediator, the natural indirect effect can be broken into path effects through susceptibility indicator Y_1 (estimated *IEB* = -0.01 (-0.07, 0.11)) and directly (estimated *IEN* = -0.01 (-0.17, 0.08)). As the correlation coefficient τ varies from 0 to 1, *IEB* and *IEN* estimates are contained in the bounded interval [-0.02, 0.00] and [-0.02, 0.01] respectively (0.00 denotes '< 0.005' in absolute value). For the AvgOHI mediator, the decomposed mediation effects estimates are 0.00 (-0.18, 0.27) for *IEB* and 0.04 (-0.19, 0.18) for *IEN*, and both estimates range between -0.01 and 0.07 when varying τ from 0 to 1. Note that 95% CIs for the decomposed indirect effect estimators above at any value of τ all cover 0 indicating no significant decomposed path effects.

In a sensitivity analysis, we examined the effect on *DE* and *IE* estimates of varying the assumed counterfactual correlation coefficient ρ (between $Y^*(t', m)$ and $M^*(t)$) from -1 to 1 in increments of 0.04. The plots of these relationships (Figure 1.6) show that the change in *IE* and *DE* estimates over ρ is in the opposite direction for Sealant and AvgOHI. For Sealant, the *IE* estimate is a maximum at $\rho = 0.44$ (Figure 1.6A) and *DE* increases as ρ increases (Figure 1.6B), while for AvgOHI, the *IE* estimate is a minimum at $\rho = 0.24$ (Figure 1.6C) and *DE* decreases as ρ increases (Figure 1.6D). However, the range for each estimator was consistent with the original conclusion indicating nonsignificant *DE* and *IE* effects (95% CIs cover 0). Thus, we find that mediation formula estimates of the direct and indirect effects in the dental data are robust to possible violations of the no M-Y confounding assumption (11).

Table 1.5. Estimated causal effects of interest based on the mediation formula approach and difference in effects approach from ZINB model using VLBW study data.

		Mediation Formula approach	Difference in Effects approach
Binary	Mediation *IE*	-0.01 (-0.12, 0.08)	-0.07 (-0.25, 0.11)
Mediator	Through Y_1 (*IEB*)*	-0.01 (-0.07, 0.11)	NA#
(Sealant)	Not through Y_1 (*IEN*)*	-0.01 (-0.17, 0.08)	NA#
	Direct *DE*	-0.62 (-1.49, 0.15)	-0.62 (-1.48, 0.14)
	Total *TE*	-0.64 (-1.51, 0.11)	-0.69 (-1.54, 0.06)
Continuous	Mediation *IE*	0.03 (-0.14, 0.19)	0.01 (-0.29, 0.23)
Mediator	Through Y_1 (*IEB*)*	0.00$ (-0.18, 0.27)	NA#
(AvgOHI)	Not through Y_1 (*IEN*)*	0.04 (-0.19, 0.18)	NA#
	Direct *DE*	-0.70 (-1.47, 0.06)	-0.70 (-1.49, 0.06)
	Total *TE*	-0.67 (-1.45, 0.10)	-0.69 (-1.54, 0.06)

Monte Carlo integration with 10,000 samples was used for estimation of *IEB* and *IEN*.

*: Requires assumed correlation (τ) between counterfactual $M(1)$ and $M(0)$ (or normally distributed latent variable for a binary mediator); here we assume $\tau = 0.4$ for both Sealant and AvgOHI.

$: 0.00 denotes < 0.005 in absolute value.

#: Not applicable.

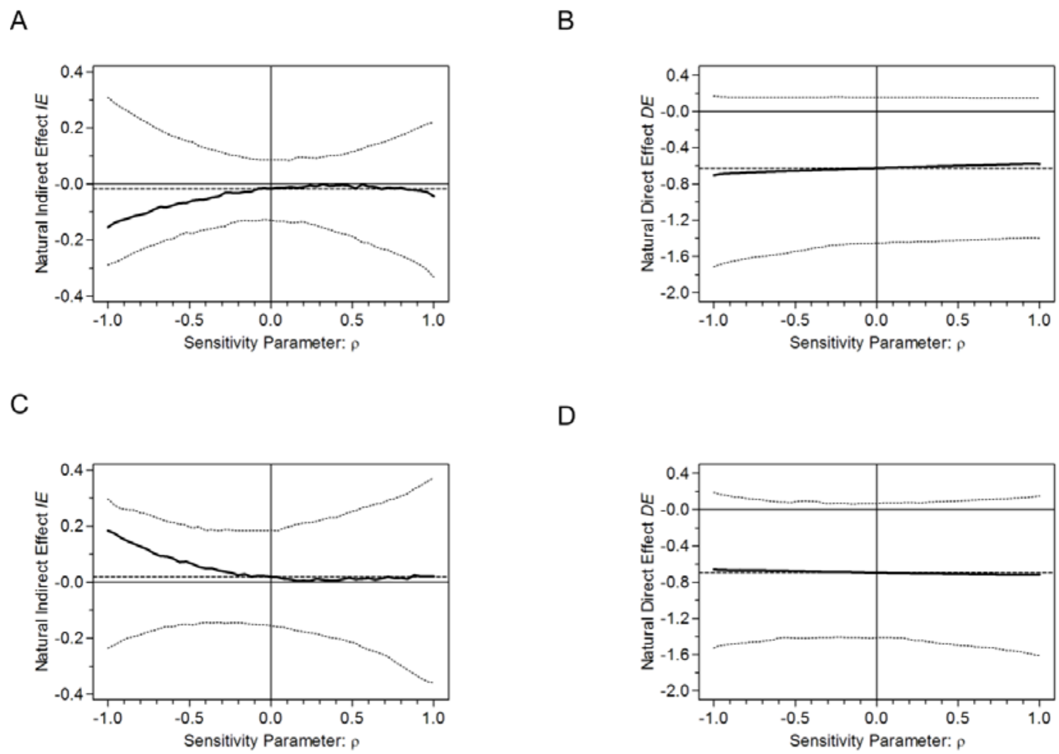

Figure 1.6. Sensitivity analysis for dental data involving the binary mediator ‘Sealant’ and the continuous mediator ‘AvgOHI’. Panels A and B show estimated *IE* and *DE* for varying correlation ρ between $Y^{*}(t', m)$ and $M^{*}(t)$ for Sealant. Panels C and D show same relationship for AvgOHI. The solid line represents the estimated indirect and direct effect. The areas between dotted lines represent the 95% confidence intervals for the natural indirect and natural direct effects at each value of ρ.

1.8 Discussion

In this part, we studied mediation analysis for ZI models using a mediation formula approach. A three-stage path framework was introduced due to the special two-component structure of the ZI model allowing us to further decompose the natural indirect effect, a decomposition which is not possible in mediation analysis based on standard linear regression models. An alternative ad hoc estimator based on a difference in effects approach was considered and compared to that of the mediation formula. The difference in effects method should be used with caution as it only gives *IE* estimator with low bias under small *DE* scenarios; an advantage of this approach is that it avoids the complex integration calculation or Monte Carlo sampling process used by the mediation formula method.

We addressed mediation analysis using the potential outcomes framework, where true causal effects are defined as differences in potential outcome means. The mediation formula calculates potential outcome means by integrating or summing over the mediator probability distribution space. Integration was carried out using the 'QUAD' function in SAS/IML, which uses an adaptive (Romberg-type) numerical integration technique. For some complex situations (for example, involving double integrals), this function may fail due to slow convergence or strong oscillation. We overcome this problem by approximating the integral calculation with Monte Carlo integration techniques, a version of which was used by Imai et al. (2010). Our simulation study, using standard linear regression to model the continuous mediator, shows that the required number of Monte Carlo samples increases as the mediator variance increases. This conclusion is understandable, as more sampling is needed in this case to achieve a good approximation of the mediator distribution, and thus less deviation from estimates using the mediation formula with ‘exact’ integration.

The three-stage path framework used for ZINB model mediation analysis allows us to further decompose the natural indirect effect under additional assumptions. Estimation of the decomposed effects (namely, through or not through the susceptibility indicator) involves the joint distribution of counterfactuals (namely, $M(1)$ and $M(0)$). This problem, which may occur when there are more than two stages of mediation, was also mentioned by Avin et al. (2005) and Albert and Nelson (2011). The simple identifying assumption that the counterfactuals are independent may not be scientifically plausible. In simulation studies, we examined the effect of varying a counterfactual correlation coefficient (under a bivariate normal model for $M(1)$ and $M(0)$) on the estimation of *IEB* and *IEN*. An interesting finding is that when either *IEB* or *IEN* predominates, estimation of the proportions corresponding to *IEB* or *IEB* is robust to mis-specification of the correlation coefficient (τ). However, in the case where *IEB* and *IEN* are of similar magnitude, the *IEB* estimator can be substantially biased if a wrong τ is chosen. Although the joint distribution for these two counterfactuals is nonidentifiable, it may be reasonable in many applications to assume a positive correlation between $M(1)$ and $M(0)$. For example, we assume a moderate correlation ($\tau = 0.4$) between $M(1)$ and $M(0)$ in the estimation of *IEB* and *IEN* for our VLBW study example. Alternatively, it may be sensible in a given application to present bounds for these estimates over a plausible range for τ.

ZI and hurdle models are two popular types of models for the analysis of zero-inflated count data, which are often found to provide a better fit than standard Poisson or negative binomial models for such data. Fitted ZI and hurdle models are often essentially indistinguishable using goodness of fit statistics (Gray, 2005). In our simulation studies, we found that mediation effect estimators based on the ZINB model retains validity for data generated from a NBH model. However, one model type may be more scientifically sensible depending on the context and study objectives. Zero-inflated models are indicated if there are two underlying processes, one which puts the subject at risk (susceptible population) and the other which determines the outcome in the at-risk population. In contrast, if all individuals are considered at risk, then the realization of the event represents a hurdle that has been crossed, and in this case the hurdle model may be more appropriate.

In conclusion, we have proposed a mediation formula approach for mediation analysis of ZI models which allows estimation of direct and indirect effects, as well as a further decomposition of the indirect effect through consideration of the latent susceptibility indicator. The difference in effects approach may be a quick and easy method for estimating indirect effect in the case of null direct effect, or for testing the hypothesis of a null direct effect. Further work is needed to study multiple mediators in ZI models.

Causal Mediation Analysis for a Dichotomous Outcome in Multiple-Mediator Models

2.1 Introduction

Randomized clinical trials and observation studies typically evaluate an overall treatment or exposure effect on a response. Of even greater scientific interest may be to explain by what means the treatment or exposure effect occurs, a goal that invokes the idea of mediation analysis. One way that a researcher can explain the mechanism by which one variable affects another is through the identification of mediating intermediate variables (or *mediators*). A well-known example of a mediated relationship in psychology is the effect of attitude on behavior which is mediated by intentions (Ajzen and Fishbein, 1980). Another example is the mediation of the effect of the apoliprotein E ε4 allele on cognitive impairment through an increase in the likelihood of chronic cerebral infarction (Schneider et al. 2005; Li et al., 2007).

At least a dozen methods have been proposed for testing the simple mediation hypothesis that the effect of an independent variable T on a dependent variable Y is mediated (at least in part) by an intermediate variable M (MacKinnon et al., 2002). Traditionally, causal mediation analysis has been dominated by linear regression paradigms, and the mediation effect assessed by difference in coefficient (MacKinnon and Dwyer, 1993) or product of coefficient approaches (MacKinnon et al., 2002). Pearl (2011) proposed a mediation formula approach which is applicable to nonlinear models with both discrete and continuous variables, and permits the evaluation of path-specific effects using the potential outcomes framework (Robins and Greenland, 1992; Albert, 2008). More generally, mediating processes may include multiple mediators. In school-based drug prevention, for example, primary prevention programs target multiple mediators such as resistance skills and social norms to reduce drug use (Donaldson et al. 1994). Reynolds et al. (2004) explored knowledge, availability of fruits and vegetables, and parental consumption as mediators of a school-based nutrition intervention to increase healthy food consumption in children.

Models with more than one mediator are straightforward extensions of single-mediator models in the linear case, and the product of coefficients approach can be used for the estimation of multiple mediation effects (MacKinnon, 2000; Preacher and Hayes, 2008). With multiple-mediator models, additional questions can be raised. For example, one may investigate the total mediation (or indirect) effect, that is, the extent to which a set of intermediate variables collectively mediate the effect of T to Y. Alternatively, one may wish to assess the extent of mediation through each individual mediator. For inference on mediation effects, Preacher and Hayes (2008) advocates the bootstrap – especially bias-corrected bootstrap – over the mulitivariate delta method (Bishop et al., 1975), since the former provides the most powerful and valid method of obtaining confidence intervals for specific indirect effect under most conditions.

Standard regression approaches generally assume normally distributed response and mediator variables. A possible way to handle multiple, and possibly non-normally distributed, mediators is through joint modeling. Methods that jointly analyze discrete and continuous outcomes have recently begun to appear, and include three main approaches. The first approach is based on a conditioning argument that allows the joint distribution to be factorized into a marginal and a conditional density (Little and Schluchter, 1985; Cox and Wermuth, 1992). A drawback of the conditioning approach for mixed outcome models is that it does not directly lead to marginal inferences (Geys et al., 2001). Also, conditional models do not easily extend to the settings of three or more outcomes, and the correlations between pairs of outcomes cannot be directly estimated. The second approach attempts to specify the joint density of two outcomes directly. Multivariate methods are well established for the modeling of multivariate normally distributed outcomes (Johnson and Wichern, 2002). To analyze mixed types of outcomes, bivariate (or multivariate) continuous variables were considered with components being either explicitly observed or latent continuous variables underlying the discrete outcomes (Catalano and Ryan, 1992). The level of an observed discrete outcome is determined according to whether or not the corresponding latent variable exceeds some threshold values. A common example is the probit-type model which assumes an unobservable normally distributed random variable underlying the binary outcome (Regan and Catalano, 1999). Instead of using latent variables, the third approach directly specifies the joint distribution via a mixed effects model, by specification of the marginal distribution for each outcome, conditional on a common or correlated random effect (Molenberghs and Verbeke, 2005). In this part, the second approach is considered. The bivariate or multivariate joint distribution of continuous/binary mediators is specified via a marginal joint distribution, components of which are continuous mediators or continuous latent variables underlying binary mediators. An advantage of this approach is that the correlation between the mediators is specified directly.

The multiple-mediator model is likely to provide a more accurate assessment of mediation effects in many research contexts. A limitation of the product of coefficients approach used for multiple mediator analysis is that it relies on linearity of the regression models, a restriction that may be difficult to justify unless the response and mediator are normally distributed. However, in medical research, the outcome or the mediator is often not normally distributed. Albert and Nelson (2011) handle sequential mediators, but not ‘contemporaneous’ mediators, for a discrete (count) outcome. Thus, the problem of mediation analysis in nonlinear (for example, probit) models for multiple (contemporaneous) mediators does not appear to have been previously addressed in the literature. In the second part, mediation formula approach is used to develop mediation analysis method for a dichotomous outcome in multiple-mediator models, and two aims are proposed:

Aim 1: To develop joint regression models to analyze binary and continuous mediators and adequately account for correlation structure in the data.

Aim 2: To estimate total and decomposed natural indirect effects for a dichotomous outcome with mixed types of mediators using mediation formula approach.

The remainder of this work proceeds as follows. Section 2.2 introduces the joint modeling approach used for multiple mediators. In Section 2.3, we define, for multiple-mediator models with a dichotomous outcome, the natural direct, (total) indirect effects, and decomposed mediation effects through individual or sets of mediators. Section 2.4 presents a mediation formula approach to estimate the total and decomposed mediation effects under given identifying assumptions. Section 2.5 proposes a sensitivity analysis that can be implemented by applied researchers to quantify the robustness of their conclusion to the potential violation of sequential ignorability assumptions. Simulation studies are used in Section 2.6 to examine the statistical properties of the proposed methods. Section 2.7 describes the application of the proposed method to a dental caries study and presents sensitivity analysis results. Discussion and suggestions for further research are presented in Section 2.8.

2.2 Background on Joint Models for Multiple Mediators

The mediation formula approach would appear to be applicable for mediation analysis of multiple-mediator models, though this extension has not yet been provided in the literature. Basically, the mediation formula expresses a mean potential outcome of the response as an integral of the conditional mean of the response over the probability density distribution of the mediators. Thus, the (estimated) joint distribution of multiple mediators is necessary for estimation of mediation effects in this model.

In this part, we only consider binary and normally distributed continuous mediators. We will illustrate the joint modeling approach with respect to one continuous mediator and one binary mediator, which is easily extended to cases with multiple continuous mediators or multiple binary mediators. Let M_{1i} and M_{2i} be two mediators measured on subject i (i = 1, 2, … , N), where M_1 denotes the binary mediator and M_2 the continuous mediator. Let M_{1i}^* be a latent variable underlying M_{1i}, such that $M_{1i} = I\{M_{1i}^* \geq 0\}$. For the bivariate response, we assume that the unobserved latent variable M_{1i}^* and the observed continuous mediator M_{2i} for individual i together follow a bivariate normal model of the form

$$M_{1i}^* = \boldsymbol{X_i}'\boldsymbol{\alpha_1} + \varepsilon_{1i} \qquad\qquad M_{2i} = \boldsymbol{X_i}'\boldsymbol{\alpha_2} + \varepsilon_{2i}$$

where $\boldsymbol{\alpha_1}$ and $\boldsymbol{\alpha_2}$ are fixed effect regression vectors of the covariate $\boldsymbol{X_i}$ on the binary and continuous endpoints, respectively, and the variance of the residual error is assumed to be

$$Var\begin{pmatrix} \varepsilon_{1i} \\ \varepsilon_{2i} \end{pmatrix} = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}.$$

As such, this model induces an unstructured correlation matrix for the two mediators. The parameter σ_1 is not identified, and standard practice assumes that $\sigma_1 = 1$ (Maddala, 1983; Edwards and Allenby, 2003). The presumption of a common design vector $\boldsymbol{X_i}$ for two mediators is not necessary, though a reasonable assumption in our application.

For the purpose of fitting, we note that the joint likelihood for the two mediators of subject i can be decomposed as

$$\begin{aligned} & f(M_{1i}, M_{2i} \mid \boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \sigma_2, \rho) = f(M_{2i} \mid \boldsymbol{\alpha}_2, \sigma_2) P(M_{1i} \mid M_{2i}, \boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \sigma_2, \rho) \\ & = f(M_{2i} \mid \boldsymbol{\alpha}_2, \sigma_2) \left\{ I(M_{1i} = 1) P(M_{1i}{}^* \geq 0 \mid M_{2i}, \boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \sigma_2, \rho) + I(M_{1i} = 0) \left[1 - P(M_{1i}{}^* \geq 0 \mid M_{2i}, \boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \sigma_2, \rho) \right] \right\} \end{aligned} \tag{38}$$

The first term above is the marginal normal distribution given previously. The conditional density function given by

$$M_{1i}{}^* \mid M_{2i} \sim N\left[\boldsymbol{X_i}'\boldsymbol{\alpha}_1 + (M_{2i} - \boldsymbol{X_i}'\boldsymbol{\alpha}_2)\, \rho / \sigma_2, (1 - \rho^2) \right]$$

Given this formulation, it is easy to then derive the contribution of M_{1i} to the likelihood. Specifically, we have

$$P(M_{1i}{}^* \geq 0 \mid M_{2i}, \boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \sigma_2, \rho) = \Phi \left\{ \frac{\boldsymbol{X_i}'\boldsymbol{\alpha}_1 + (M_{2i} - \boldsymbol{X_i}'\boldsymbol{\alpha}_2)\, \rho / \sigma_2}{\sqrt{(1 - \rho^2)}} \right\} \tag{39}$$

Therefore, we can write out the subject-specific contribution to the joint likelihood for (M_{1i}, M_{2i}) as the product of the marginal normal likelihood for M_{2i} times either the Bernoulli probability (39) or the probability of the complement, depending on whether M_{1i} was 1 or 0 respectively. Therefore, the log-likelihood is defined by summing the log of expressions in formula (38) over all subjects. Here, model estimation can be implemented using either SAS PROC NLMIXED procedure to specify the likelihood function manually or PROC QLIM procedure which constructs the likelihood automatically from specified regression models.

2.3 Defining Total and Decomposition of Mediation Effects

Consider the general causal model including a binary exposure or treatment indicator (T), J mediators (M_1, M_2, …, M_J) and a dichotomous outcome (Y), where T may affect Y directly and/or T may affect any of the M_j, j = 1 to J, which then affect Y. Figure 2.1 shows the path diagram. We can define the causal mediation effects of interest using nested potential outcomes (Albert, 2008) within this multiple-mediator model. The total mediation effect (through the set $\{M_j, j = 1, \dots, J\}$) under exposure t is defined as

$$IE(t) \equiv E\{Y(t, M_1(1), M_2(1), \dots, M_J(1))\} - E\{Y(t, M_1(0), M_2(0), \dots, M_J(0))\} \qquad (40)$$

Here, $Y(t, M_1(t'), M_2(t'), \dots, M_J(t'))$ denotes the potential outcome that Y would attain if T was set to t, and each M_j set to the counterfactual value that would be observed if T was set to t'. $IE(t)$ is called the *natural indirect effect*, and represents the difference between two mean potential outcomes that would result under exposure status t, but where all mediators takes values that would result under two different exposure statuses. Similarly, we can define the *natural direct effect* and the *total causal effect* in the potential outcomes framework as

$$DE(t) \equiv E\{Y(1, M_1(t), M_2(t), \dots, M_J(t))\} - E\{Y(0, M_1(t), M_2(t), \dots, M_J(t))\} \qquad (41)$$

$$TE \equiv E\{Y(1, M_1(1), M_2(1), \dots, M_J(1))\} - E\{Y(0, M_1(0), M_2(0), \dots, M_J(0))\} \qquad (42)$$

Thus, the natural indirect effect under one exposure status and the natural direct effect under the other exposure status sum to the total causal effect. Here, we consider the decomposition involving $IE(1)$ and $DE(0)$ denoted as IE and DE; for the other

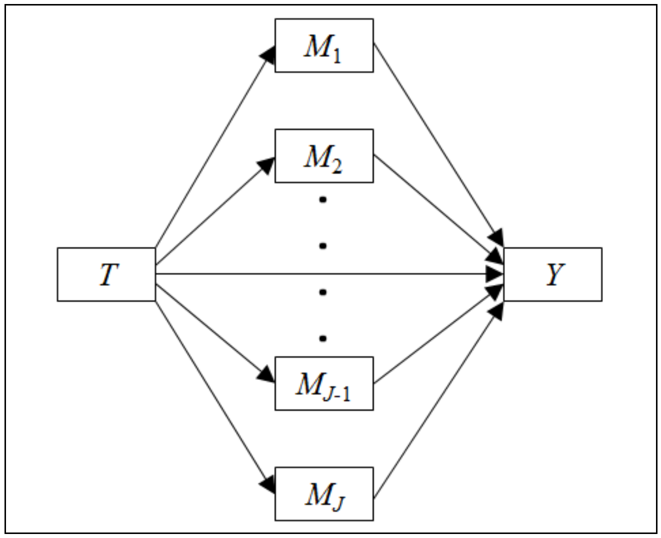

Figure 2.1. Path diagram for a multiple-mediator model with J mediators. T = exposure; M_j = mediator j, j = 1, … , J; Y = outcome. T may exert indirect effects on Y through M_1, M_2, … , M_J, or affect Y directly.

decomposition (involving $IE(0)$ and $DE(1)$) the method will be similar.

In multiple-mediator models, we can further consider mediation effects through an individual mediator or sets of mediators. The total natural indirect effect can be broken into J path effects within a J-mediator model; we use '$IE_j\,(t_0, t_1, \ldots, t_{j\text{-}1}, t_{j+1}, \ldots, t_J)$' to denote the path (or mediation) effect through the jth mediator (Figure 2.1) with exposure set to t_0 and other mediators except for M_j set to exposure levels $t_1, \ldots, t_{j\text{-}1}, t_{j+1}, \ldots, t_J$ respectively. $IE_j\,(t_0, t_1, \ldots, t_{j\text{-}1}, t_{j+1}, \ldots, t_J)$ can be defined as the difference between two mean potential outcomes as follows,

$$IE_j\,(t_0, t_1, \ldots, t_{j\text{-}1}, t_{j+1}, \ldots, t_J) \equiv E\{Y(t_0, M_1(t_1),\ldots, M_{j\text{-}1}(t_{j\text{-}1}), M_j(1), M_{j+1}(t_{j+1}),\ldots, M_J(t_J))\}$$
$$- E\{Y(t_0, M_1(t_1),\ldots, M_{j\text{-}1}(t_{j\text{-}1}), M_j(0), M_{j+1}(t_{j+1}),\ldots, M_J(t_J))\} \quad (43)$$

Mediation effects through sets of mediators can be defined as the sum of mediation effects through the component individual mediators. Note that 2^J versions of $IE_j\,(t_0, t_1, \ldots,$

$t_{j\text{-}1}$, t_{j+1}, … , $t_J = 0$ or 1), the indirect effect of M_j, can be formulated corresponding to 2^J combinations of exposure settings for the outcome and other mediators in a J-mediator model. We use only one of them, IE_j ($t_0 = 1$, $t_1 = 0$, … , $t_{j\text{-}1} = 0$, $t_{j+1} = 1$, … , $t_J = 1$), denoted as IE_j for simplicity; the other possible estimands can be handled similarly and thus are not discussed here. Of note, this defined IE_j´s provide a proper decomposition of a total (indirect) effect among individual mediators as follow,

$$IE = IE_1 + IE_2 + , \dots , + IE_{J\text{-}1} + IE_J \tag{44}$$

2.4 Analysis of Total and Decomposed Mediation Effects

2.4.1. Identification assumption for mediation effects in multiple-mediator models

We present identification results for the (total) indirect and direct effects (defined by (40) and (41)) as well as mediation effects through individual or sets of mediators (defined by (43)) using the potential outcomes framework described above. Under a particular version of the sequential ignorability assumption, the mediation effect estimators are identified nonparametrically in the multiple-mediator causal model. We first define our identifying assumption which extends Imai et al.'s version (Imai et al. 2010).

Assumption 3 (Sequential Ignorability)

$$\{Y(t, m_1, \dots, m_J), M_1(t_1), \dots, M_J(t_J)\} \perp\!\!\!\perp T \mid W = w \tag{45}$$

$$Y(t, m_1, \dots, m_J) \perp\!\!\!\perp M_1(t_1) \mid W = w, T = t_1, \dots, Y(t, m_1, \dots, m_J) \perp\!\!\!\perp M_J(t_J) \mid W = w, T = t_J \tag{46}$$

First, given the observed baseline covariates (W), the exposure or treatment is assumed to be statistically independent of potential outcomes and potential mediators. The second part of Assumption 3 states that all mediators are independent of potential outcomes given the observed exposure or treatment and pretreatment covariates.

2.4.2. Estimation of total and decomposed mediation effect for multiple-mediator models

To demonstrate identifiability of total and decomposed mediation effect for multiple-mediator models, we will examine the identification of the relevant potential outcome means (of the general form $E\{Y(t_0, M_1(t_1), \dots, M_J(t_J))\}$) given baseline covariate $\boldsymbol{W}$ under sequential ignorability (Assumption 3), and the conditioning on $\boldsymbol{W}$ is left out for brevity. Under Assumption 3 it can be shown that,

$$E\{Y(t_0, M_1(t_1),..., M_J(t_J))\} = \underbrace{\int ... \int}_{n=J} E\{Y \mid T = t_0, M_1 = m_1,..., M_J = m_J\} d^J F_{M_1(t_1),...,M_J(t_J)}(m_1,..., m_J) \quad (47)$$

A proof of formula (47) under assumptions (45) and (46) is given in Appendix IV. Integration will be replaced with summation in the case of a discrete mediator.

The total natural indirect effect, direct effect and total causal effect are identifiable, since the required joint distribution in formula (47) ($M_1(0)$, … , $M_J(0)$) and ($M_1(1)$, … , $M_J(1)$) can be estimated with the joint modeling approach described in Section 2.2. However, when assessing mediation effects through individual or sets of mediators we require a joint distribution ($M_1(t_1)$, … , $M_J(t_J)$) in which the t_j′s are not all equal; such a joint distribution is not estimable because this joint potential outcome cannot be observed for the same subject. For identifiability, we need to make additional assumptions regarding the joint distribution of ($M_1(t_1)$, … , $M_J(t_J)$). In fact, the marginal distribution of $M_j(t_j)$, $j = 1, \ldots, J$, is estimable, and if each correlation coefficient ρ_{kl}' between $M_k(t_k)$ and $M_l(t_l)$, $k < l$, $t_k \neq t_l$ (or underlying continuous latent variables) is pre-specified, then the mean potential outcome in formula (47) is identifiable and therefore so is the decomposed natural indirect effect through individual or sets of mediators. Of note, we may assume two different ρ_{kl}'s (one is for correlation coefficient between $M_k(0)$ and $M_l(1)$ and the other one is for that between $M_k(1)$ and $M_l(0)$), but for any specific mediation effect, the identification of individual mediation effect only needs one of them (e. g., the former one is needed for defined IE_j above).

To evaluate the joint integrals, we may also use Monte Carlo integration (James, 1980), a numerical integration technique that uses pseudo random numbers generated from a joint distribution of mediator counterfactuals. Suppose (m_{1n}, … , m_{Jn}) and (m_{1n}', … , m_{Jn}') are drawn from the joint distributions $f_{M_1(0),\, \ldots,\, M_J(0)}$ (m_1, … , m_J) and $f_{M_1(1),\, \ldots,\, M_J(1)}$ (m_1', … , m_J') respectively. Formula (40) can then be approximated by the following (using Monte Carlo approximations for formula (47)),

$$IE(t) \approx \frac{1}{N}\left(\sum_{n=1}^{N} E\{Y \mid M_1 = m_{1n}',\ldots,M_J = m_{Jn}',T=t\} - \sum_{n=1}^{N} E\{Y \mid M_1 = m_{1n},\ldots,M_J = m_{Jn},T=t\}\right) \tag{48}$$

The considered IE_j using formula (43) can be expressed in similar way:

$$\begin{aligned} IE_j \approx \frac{1}{N}\Bigg(&\sum_{n=1}^{N} E\{Y \mid M_1 = m_{1n},\ldots,M_{j-1} = m_{(j-1)n}, M_{jn} = m_{jn}', M_{(j+1)n} = m_{(j+1)n}',\ldots,M_J = m_{Jn}',T=1\} \\ &-\sum_{n=1}^{N} E\{Y \mid M_1 = m_{1n},\ldots,M_{j-1} = m_{(j-1)n}, M_{jn} = m_{jn}, M_{(j+1)n} = m_{(j+1)n}',\ldots,M_J = m_{Jn}',T=1\}\Bigg) \end{aligned} \tag{49}$$

where (m_{1n}, … , $m_{(j-1)n}$, m_{jn}', $m_{(j+1)n}'$,…, m_{Jn}'), (m_{1n}, … , $m_{(j-1)n}$, m_{jn}, $m_{(j+1)n}'$,…, m_{Jn}') are drawn from the joint distribution $f_{M_1(0),\, \ldots,\, M_{j-1}(0),\, M_j(1),\, M_{j+1}(1),\, \ldots,\, M_J(1)}$ (m_1, … m_{j-1}, m_j', m_{j+1}', … , m_J') and $f_{M_1(0),\, \ldots,\, M_{j-1}(0),\, M_j(0),\, M_{j+1}(1),\, \ldots,\, M_J(1)}$ (m_1, … m_{j-1}, m_j, m_{j+1}', … , m_J').

Estimation of mediation effects proceeds by estimating the regression function in the mediation formula. In the present context, we fit the data to the following regression models for the dichotomous outcome Y and the binary or continuous correlated mediators M_j, j = 1 , …, J,

$$\text{logit}(Y) = \beta_0 + \beta_1 M_1 + \ldots + \beta_J M_J + \beta_{J+1} T + \boldsymbol{\beta}'\boldsymbol{W} \tag{50}$$

$$M_j^{(*)} = \alpha_{j0} + \alpha_{j1}\, T + \boldsymbol{\alpha_j}'\boldsymbol{W} + \varepsilon_j,\ j = 1, \ldots, J\,. \tag{51}$$

If M_j is a binary mediator, $M_j^{(*)}$ denotes M_j^{*} and $Var(\varepsilon_j)$ = 1. Otherwise if M_j is a continuous mediator, $M_j^{(*)}$ denotes M_j and $Var(\varepsilon_j)$ is estimable. Further, we assume a constant unstructured correlation matrix for the J continuous (observed and latent) mediators regardless of the set exposure statuses. The β's and α_j's are regression parameters, and $\boldsymbol{W}$ is baseline covariate vector with corresponding coefficient vectors $\boldsymbol{\beta'}$ and $\boldsymbol{\alpha_j'}$.

2.5 Sensitivity Analysis

Sensitivity analysis explores the impact of departures from untestable assumptions and thus is an important component of our approach. Two sensitivity analyses are provided in this section. The first one assesses the effect of the assumed correlation coefficient ρ_{kl}' between $M_k^{(*)}(t_k)$ and $M_l^{(*)}(t_l)$, $k < l$, $t_k \neq t_l$ on the estimation of decomposed mediation effects, and the second one evaluates the effect of violation of the no unmeasured mediator-outcome confounder assumption (46) on mediation effect estimation. For ease of presentation, we use a two-mediator model with one binary mediator (M_1) and one continuous mediator (M_2) to illustrate our sensitivity analysis approach in this section. The approach described can be easily extended to models with more than two mediators.

2.5.1. Effect of assumed correlation coefficient on decomposed mediation effect estimation

As stated in section 2.4.2, the correlation coefficient ρ' between counterfactuals $M_1^*(0)$ and $M_2(1)$ needs to be pre-specified for the estimation of decomposed mediation effects IE_1 and IE_2. We may examine the robustness of the IE_1 and IE_2 estimators, thus providing a sensitivity analysis, by varying the specified counterfactual correlation ρ' over the interval -1 to 1.

2.5.2. Sensitivity analysis for violation of no unmeasured mediator-outcome confounder assumption

We propose a sensitivity analysis based on the approach of Albert and Nelson (2011) and Wang and Albert (2012), which applies the Gaussian copula (Song et al., 2009), to assess the joint probability of $Y(t_0, m_1, m_2)$, $M_1(t_1)$ and $M_2(t_2)$, where Y and M_1 are binary and M_2 is continuous (normally distributed). In order to handle binary variables, corresponding continuous latent variables are introduced, such that the observed binary event is realized if the latent variable exceeds some threshold. Specifically, for the probit-type model, the threshold is 0. Then $Y^*(t_0, m_1, m_2)$, $M_1^*(t_1)$ and $M_2(t_2)$ are assumed to have a trivariate normal distribution with covariance matrix Σ and mean μ. The marginal distributions of $Y^*(t_0, m_1, m_2)$, $M_1^*(t_1)$ and $M_2(t_2)$ as well as the correlation coefficient ρ between $M_1^*(t_1)$ and $M_2(t_2)$ are estimable; we assume that the correlation between $M_1^*(t_1)$ and $M_2(t_2)$ is constant over all t_1 and t_2. The correlation between $Y^*(t_0, m_1, m_2)$ and $M_1^*(t_1)$ is denoted as ρ_1 and the correlation between $Y^*(t_0, m_1, m_2)$ and $M_2(t_2)$ as ρ_2. Similarly, we also assume ρ_1 and ρ_2 keep consistent for all t_0, t_1, and t_2. Nonzero ρ_1 (or ρ_2) implies that there exists an omitted variable that affects both M_1 (or M_2) and Y. We propose a Monte Carlo approach to estimate the conditional probability $P\{Y(t_0, m_1, m_2) = 1\}$ under the assumed ρ_1 and ρ_2. This approach involves the following algorithm:

(1) Sample the pair of mediator values $(m_{1i}^*, m_{2i}) \sim f_{M_1^*(t_1), M_2(t_2)}(m_1^*, m_2)$, $i = 1, 2, \ldots, n$. If $m_{1i}^* > 0$ then $m_{1i} = 1$, otherwise, $m_{1i} = 0$.

(2) The conditional distribution $Y^*(t_0, m_{1i}, m_{2i})$ is obtained based on the trivariate normal distribution of $Y^*(t_0, m_{1i}, m_{2i})$, $M_1^*(t_1)$ and $M_2(t_2)$ with estimated ρ as well as pre-specified ρ_1 and ρ_2.

(3) Compute $P\{Y(t_0, m_{1i}, m_{2i}) = 1\}$ as $P\{Y^*(t_0, m_{1i}, m_{2i}) > 0\}$ based on the distribution in (2).

(4) Repeat steps 1-3 n times, we can obtain the potential outcome mean $E(Y(t_0, M_1(t_1),$

$M_2(t_2)))$ as $\frac{1}{n}\sum_{i=1}^{n} P\left\{Y(t_0, m_{1i}, m_{2i}) = 1\right\}$.

This algorithm provides an estimation method for the general expected value of potential outcome $Y(t_0, M_1(t_1), M_2(t_2))$ assuming specified correlations (ρ_1 and ρ_2) between each mediators and the outcome adjusted for the mediators. It thus provides a sensitivity analysis of the total and decomposed natural indirect effect for the sequential ignorability assumption (46).

2.6 Simulation Study

In this section, a simulation study was used to assess the empirical bias, coverage probability and power of total and decomposed natural indirect effect estimators from our mediation formula approach for a binary outcome in a two-mediator model. We use SAS/IML (Version 9.2) for all statistical simulation and analysis.

2.6.1. Empirical bias

To assess the empirical bias of mediation effect estimators, we conduct simulations using a 3 × 4 × 3 × 3 factorial design. First of all, we consider three two-mediator models with different combinations of mediator types, two binary, one binary and one continuous, or two continuous mediators. For each two-mediator model, four correlation coefficients values (-0.5, 0.0, 0.5 and 0.9) between the two mediators (or latent variables underlying any binary mediators) are assumed. The exposure effect is similar for all scenarios (around 0.2).Three cases are considered for the magnitudes of the total natural indirect effect, *IE*, (involving parameters β_1, β_2, α_{11} and α_{21}) and the natural direct effect, *DE*, (involving β_3): dominant *IE*, dominant *DE*, and similar magnitude of *IE* and *DE*. The last factor, related to the decomposition of the total natural indirect effect, consists of three scenarios: indirect effect exclusively through one mediator, similar mediation effect through each mediator, and substantial mediation effects through each mediator but with different directions. Sample sizes of 200 (100 per exposure group) and 500 (250 per exposure group) were used for each scenario.

The bivariate normal distributed error terms for two mediators (or underlying latent variables corresponding to binary mediators) were generated using the 'RANDNORMAL' function in SAS/IML. No baseline covariates were used in our simulation scenarios to simplify the calculations. For each given exposure, the mediator variables or corresponding latent variables were generated using equation (51) with above produced error terms. The response variates were then generated according to the logistic regression model (50) given the individual exposure and observed mediators (possibly from corresponding latent variables). For each dataset, we estimated the total and decomposed natural indirect effects using formula (40), (43) and (47). The true value is defined by the same function, with true coefficients in place of estimates. In addition, in the calculation of both true and estimated decomposed mediation effects, we assume that the correlation between $M_1^{(*)}(t_1)$ and $M_2^{(*)}(t_2)$ (which is estimable only for $t_1 = t_2$) is constant over all t_1 and t_2. We performed 1000 independent datasets and the following statistics for mediation effects estimators (total or individual) were provided: the average estimate; the average percent error (PE, estimate minus true value then divided by true value); the SD of the estimate.

We focus on results for the two-mediator model with one binary mediator and one continuous mediator; results for two-mediator model with either two binary or continuous mediators are similar and therefore not presented. The simulation results are given in Table 2.1 with total sample size 200. We see that the mediation formula approach generally produces a small bias in its estimation of IE, IE_1 and IE_2 in most scenarios.

Table 2.1. Simulation statistics of overall and decomposed mediation effect estimators for a dichotomous outcome model with two mediators (one binary (M_1) and one continuous (M_2)), n = 100 per group.

ρ (M_1, M_2)	True DE	True IE	True IE_1	True IE_2	Est DE	Overall IE Ave Est	Overall IE Ave PE (%)	Overall IE SD of Est	IE_1 Ave Est	IE_1 Ave PE (%)	IE_1 SD of Est	IE_2 Ave Est	IE_2 Ave PE (%)	IE_2 SD of Est
-0.5	0.185	0.024	0.024	0.000	0.184	0.025	3.66	0.034	0.025	1.66	0.017	0.000	-	0.026
	0.211	0.020	0.010	0.010	0.210	0.022	8.70	0.034	0.011	2.53	0.014	0.011	15.21	0.028
	0.183	0.020	0.198	-0.179	0.173	0.033	70.82	0.100	0.209	5.36	0.082	-0.176	-1.78	0.035
	0.097	0.107	0.107	0.000	0.098	0.107	0.63	0.069	0.107	0.44	0.056	0.000	-	0.024
	0.099	0.113	0.056	0.056	0.097	0.116	2.74	0.095	0.057	1.81	0.054	0.058	3.67	0.061
	0.092	0.101	0.309	-0.208	0.106	0.090	-10.89	0.124	0.297	-3.94	0.107	-0.207	-0.57	0.036
	0.023	0.179	0.179	0.000	0.022	0.178	-0.53	0.087	0.177	-1.58	0.056	0.002	-	0.047
	0.024	0.182	0.093	0.089	0.025	0.178	-2.07	0.078	0.090	-3.24	0.048	0.088	-0.85	0.048
	0.021	0.188	0.305	-0.117	0.020	0.193	2.26	0.122	0.308	0.88	0.109	-0.115	-1.33	0.030
0.0	0.185	0.024	0.024	0.000	0.184	0.025	2.99	0.031	0.025	3.82	0.017	-0.000	-	0.025
	0.210	0.020	0.010	0.010	0.209	0.021	0.74	0.029	0.011	8.27	0.013	0.009	-7.19	0.026
	0.192	-0.005	0.190	-0.195	0.187	0.001	-121.7	0.077	0.194	2.25	0.060	-0.193	-0.64	0.037
	0.097	0.107	0.107	0.000	0.090	0.111	3.93	0.057	0.109	2.10	0.053	0.002	-	0.021
	0.098	0.112	0.056	0.056	0.097	0.115	2.12	0.075	0.056	0.19	0.049	0.058	4.04	0.057
	0.095	0.083	0.303	-0.220	0.084	0.095	14.03	0.122	0.312	3.12	0.107	-0.217	-1.02	0.038
	0.023	0.179	0.179	0.000	0.022	0.180	0.27	0.071	0.178	-0.79	0.053	0.002	-	0.044
	0.024	0.180	0.093	0.088	0.025	0.177	-2.01	0.064	0.090	-2.38	0.048	0.086	-1.61	0.044
	0.021	0.180	0.302	-0.122	0.020	0.180	0.12	0.114	0.301	-0.15	0.107	-0.121	-0.56	0.031
0.5	0.185	0.024	0.024	0.000	0.188	0.024	1.01	0.028	0.024	-0.27	0.017	0.000	-	0.027
	0.210	0.020	0.010	0.010	0.206	0.021	3.72	0.026	0.011	1.66	0.015	0.010	5.89	0.027
	0.207	-0.019	0.185	-0.204	0.204	-0.012	-35.22	0.063	0.189	2.10	0.056	-0.202	-1.38	0.035
	0.097	0.107	0.107	0.000	0.096	0.108	1.28	0.053	0.108	1.07	0.055	0.000	-	0.023
	0.098	0.112	0.056	0.056	0.102	0.111	-0.75	0.065	0.053	-6.42	0.054	0.059	4.95	0.059
	0.100	0.076	0.305	-0.229	0.096	0.082	8.15	0.102	0.308	1.06	0.093	-0.226	-1.29	0.037
	0.023	0.179	0.179	-0.000	0.031	0.174	-2.88	0.062	0.176	-1.85	0.053	-0.002	-	0.047
	0.023	0.179	0.093	0.086	0.026	0.176	-1.66	0.058	0.093	-0.21	0.053	0.084	-3.22	0.049
	0.022	0.177	0.303	-0.126	0.013	0.184	3.85	0.097	0.307	1.53	0.096	-0.124	-1.73	0.031
0.9	0.185	0.024	0.024	0.000	0.185	0.024	-0.02	0.029	0.023	-4.46	0.019	0.001	-	0.034
	0.210	0.020	0.010	0.010	0.210	0.020	-1.26	0.028	0.009	-13.22	0.017	0.011	11.39	0.035
	0.224	-0.025	0.181	-0.206	0.228	-0.025	2.31	0.053	0.176	-2.93	0.054	-0.201	-2.31	0.036
	0.097	0.107	0.107	-0.000	0.096	0.108	1.40	0.055	0.107	0.09	0.068	0.001	-	0.028
	0.097	0.112	0.056	0.056	0.099	0.110	-1.70	0.060	0.051	-10.59	0.064	0.060	7.30	0.073
	0.106	0.080	0.313	-0.233	0.111	0.075	-6.04	0.090	0.308	-1.70	0.096	-0.233	-0.20	0.039
	0.023	0.179	0.179	0.000	0.024	0.179	-0.16	0.056	0.177	-1.53	0.067	0.002	-	0.059
	0.023	0.179	0.094	0.085	0.018	0.181	1.18	0.051	0.092	-1.64	0.060	0.089	4.28	0.059
	0.023	0.180	0.307	-0.128	0.024	0.180	0.18	0.096	0.306	-0.45	0.103	-0.126	-1.33	0.033

Table 2.2. Simulation statistics of overall and decomposed mediation effect estimators for a dichotomous outcome model with two mediators (one binary (M_1) and one continuous (M_2)), n = 250 per group.

ρ (M_1, M_2)	True *DE*	True *IE*	True IE_1	True IE_2	Est *DE*	Overall *IE*			IE_1			IE_2		
						Ave Est	Ave PE (%)	SD of Est	Ave Est	Ave PE (%)	SD of Est	Ave Est	Ave PE (%)	SD of Est
-0.5	0.185	0.024	0.024	0.000	0.185	0.024	-0.49	0.015	0.024	-0.66	0.007	0.000	-	0.011
	0.211	0.020	0.010	0.010	0.210	0.020	-1.68	0.014	0.010	-0.04	0.006	0.010	-3.40	0.011
	0.183	0.020	0.198	-0.179	0.183	0.021	6.63	0.043	0.200	0.71	0.034	-0.179	0.07	0.015
	0.097	0.107	0.107	0.000	0.098	0.107	0.03	0.030	0.107	0.44	0.025	-0.000	-	0.010
	0.099	0.113	0.056	0.056	0.098	0.114	1.29	0.042	0.057	1.40	0.023	0.057	1.18	0.027
	0.092	0.101	0.309	-0.208	0.083	0.110	8.83	0.070	0.318	2.72	0.063	-0.208	-0.24	0.017
	0.023	0.179	0.179	0.000	0.025	0.179	-0.11	0.038	0.179	-0.18	0.025	0.000	-	0.021
	0.024	0.182	0.093	0.089	0.022	0.185	1.43	0.036	0.094	0.99	0.022	0.091	1.89	0.022
	0.021	0.188	0.305	-0.117	0.016	0.194	2.78	0.063	0.311	1.68	0.058	-0.117	-0.09	0.014
0.0	0.185	0.024	0.024	0.000	0.187	0.024	0.24	0.014	0.024	-0.13	0.007	0.000	-	0.011
	0.210	0.020	0.010	0.010	0.211	0.021	0.58	0.013	0.010	-0.90	0.005	0.010	2.15	0.011
	0.192	-0.005	0.190	-0.195	0.193	-0.004	-17.56	0.033	0.190	0.03	0.027	-0.194	-0.38	0.015
	0.097	0.107	0.107	0.000	0.096	0.107	0.22	0.025	0.107	0.44	0.023	-0.000	-	0.009
	0.098	0.112	0.056	0.056	0.101	0.112	-0.45	0.034	0.056	-0.75	0.023	0.056	-0.14	0.025
	0.095	0.083	0.303	-0.220	0.091	0.086	3.73	0.054	0.306	1.14	0.048	-0.220	0.16	0.017
	0.023	0.179	0.179	0.000	0.022	0.180	0.60	0.031	0.180	0.53	0.022	0.000	-	0.020
	0.024	0.180	0.093	0.088	0.023	0.180	-0.36	0.029	0.093	-0.13	0.021	0.087	-0.60	0.020
	0.021	0.180	0.302	-0.122	0.019	0.182	1.33	0.051	0.304	0.78	0.046	-0.122	-0.04	0.014
0.5	0.185	0.024	0.024	0.000	0.188	0.023	-2.67	0.012	0.024	-1.43	0.007	-0.000	-	0.012
	0.210	0.020	0.010	0.010	0.212	0.020	-1.56	0.012	0.010	-0.53	0.006	0.010	-2.65	0.012
	0.207	-0.019	0.185	-0.204	0.206	-0.018	-3.62	0.027	0.185	-0.14	0.024	-0.203	-0.47	0.015
	0.097	0.107	0.107	0.000	0.095	0.108	0.92	0.023	0.108	0.72	0.024	0.000	-	0.010
	0.098	0.112	0.056	0.056	0.098	0.111	-0.79	0.029	0.056	-1.06	0.024	0.056	-0.53	0.026
	0.100	0.076	0.305	-0.229	0.100	0.077	1.75	0.041	0.306	0.31	0.039	-0.229	-0.17	0.017
	0.023	0.179	0.179	-0.000	0.023	0.180	0.24	0.026	0.179	0.06	0.025	0.000	-	0.021
	0.023	0.179	0.093	0.086	0.021	0.180	0.47	0.025	0.093	0.31	0.023	0.087	0.65	0.021
	0.022	0.177	0.303	-0.126	0.019	0.178	0.49	0.045	0.304	0.28	0.045	-0.126	-0.01	0.014
0.9	0.185	0.024	0.024	0.000	0.186	0.024	-1.84	0.012	0.024	-2.30	0.008	0.000	-	0.014
	0.210	0.020	0.010	0.010	0.210	0.021	2.72	0.012	0.010	-2.05	0.007	0.011	7.75	0.015
	0.224	-0.025	0.181	-0.206	0.223	-0.025	0.60	0.023	0.181	0.10	0.023	-0.206	0.16	0.016
	0.097	0.107	0.107	-0.000	0.098	0.106	-0.32	0.023	0.107	-0.19	0.028	-0.000	-	0.012
	0.097	0.112	0.056	0.056	0.095	0.114	1.18	0.027	0.056	-0.39	0.028	0.057	2.77	0.033
	0.106	0.080	0.313	-0.233	0.106	0.080	-0.27	0.040	0.313	0.01	0.042	-0.233	0.10	0.017
	0.023	0.179	0.179	0.000	0.023	0.180	0.36	0.024	0.179	-0.07	0.028	0.001	-	0.025
	0.023	0.179	0.094	0.085	0.024	0.179	-0.08	0.022	0.094	0.48	0.028	0.085	-0.69	0.027
	0.023	0.180	0.307	-0.128	0.022	0.182	1.09	0.042	0.309	0.42	0.045	-0.127	-0.51	0.015

However, the proposed approach occasionally provides some bias in two situations, scenarios in which the two mediators have mediation effects in different directions (the sixth scenario in each set) and scenarios with similar mediation effects magnitude through either mediator and where the correlation coefficient between the two mediators is 0.9 (the fifth scenario). In the first situation, our approach tends to produce high relative bias for the IE estimator, but low relative bias for the IE_1 and IE_2 estimators (less than 4%); on the contrary, high relative bias for individual and low relative bias (less than 3%) for overall mediation effect estimators can be detected in the second situation. When the sample size per group is increased to 250, the proposed approach shows low relative bias for all scenarios (less than 8.9%, Table 2.2).

2.6.2. Coverage probability and power

In a second simulation study, we compared the coverage probability (CP) of 95% CIs constructed by different methods. Four such methods were evaluated in this section: jackknife (Mosteller and Tukey, 1977), percentile bootstrap, bias-corrected bootstrap, and bootstrap-*t* (Efron and Tibshirani, 1993). The second goal was to assess the power to detect non-zero mediation effects under different correlation coefficients between two mediators. Four correlation coefficients were chosen as before, -0.5, 0.0, 0.5 and 0.9. For each correlation coefficient, we designed six different simulation scenarios: (1) small indirect effects with different signs for the two mediators; (2) substantial indirect effects with different signs and magnitudes for the two mediators; (3) zero indirect effect through each mediators; (4) zero indirect effect through the binary mediator and small indirect effect through the continuous mediator; (5) zero indirect effect through the continuous mediator and small indirect effect through the binary mediator; (6) small direct effects with the same sign for both mediators. Of note, for power comparison purpose, we adjusted the corresponding parameter β_1, β_2, α_{11} and α_{21} to make the true mediation estimand almost 'identical' (constant up to the fourth decimal) for different specified correlation coefficient values. The CP is calculated as the percent of simulated datasets for which the 95% CIs covered the true value, and the power is calculated as the proportion of these 1000 replicated CIs that do not include zero. The results are summarized in Table 2.3, Table 2.4 and Table 2.5 respectively.

Table 2.3 shows the CP of four different methods for total and individual indirect effect estimates with sample size 100 per group. The results indicate that the bootstrap percentile CIs, for which the CP is within 3% of the nominal level for all scenarios, performs best. The coverage for bias-corrected bootstrap and jackknife CIs is only around 85% for the second scenario with correlation coefficient -0.5, and the CP of bootstrap-*t* CIs for continuous mediator indirect effect, IE_2, is unstable in most scenarios. We repeated the simulations with a sample size of 250 per group (Table 2.4). With the larger sample size, the CP for bootstrap percentile, bias-corrected bootstrap and jackknife CIs gets closer to nominal level while problems for bootstrap-*t* CIs still exist.

We then compared the power of bootstrap percentile CIs to detect non-zero mediation effects under different values of the correlation between the two mediators.

The results, given in Table 2.5, show that the power to detect non-zero *IE* increases as the correlation coefficient between two mediators increases, and the power to detect non-zero IE_1 and IE_2 increases as the absolute value of the correlation coefficient between two mediators gets close to zero for most simulation scenarios. When the sample size per group is increased to 250, the power is larger than that in small sample size scenarios and similar results can be observed.

Table 2.3. Coverage probabilities of four confidence interval methods for overall and decomposed mediation effects in a dichotomous outcome model with two mediators (one binary (M_1) and one continuous (M_2)), n = 100 per group.

ρ (M_1, M_2)	True *IE*	True IE_1	True IE_2	Overall *IE*				IE_1				IE_2			
				Jack knife	Boot	Bc	Bt	Jack knife	Boot	Bc	Bt	Jack knife	Boot	Bc	Bt
-0.5	0.000	0.050	-0.050	93.2	94.6	94.5	93.8	95.5	94.4	92.9	94.3	92.8	93.4	92.7	82.1
	0.200	0.300	-0.100	83.8	97.3	83.1	90.4	81.5	97.7	84.4	87.5	95.2	94.9	94.3	99.9
	0.000	0.000	0.000	95.2	94.7	93.7	93.5	99.7	98.0	95.1	90.8	97.0	93.1	90.6	68.0
	0.051	0.000	0.051	96.3	94.8	94.7	94.9	99.5	97.3	94.6	92.0	94.9	94.3	94.9	64.2
	0.050	0.050	0.000	94.2	95.1	94.9	94.5	96.0	95.3	94.6	94.4	96.8	95.6	93.0	99.2
	0.100	0.050	0.050	93.6	94.7	95.1	93.2	95.4	95.0	93.4	94.3	95.0	94.9	94.0	81.8
0.0	0.000	0.050	-0.050	94.1	94.1	93.5	93.6	96.2	94.7	93.6	93.8	94.8	94.9	94.1	85.0
	0.200	0.300	-0.100	91.4	94.8	89.7	91.9	91.5	94.5	89.0	91.6	94.4	94.0	94.6	99.9
	0.000	0.000	0.000	94.7	95.2	93.1	95.0	99.7	97.9	95.2	92.3	95.8	93.7	91.1	67.5
	0.051	0.000	0.051	95.2	94.8	94.1	94.9	99.7	97.4	95.4	93.0	94.0	94.2	94.1	61.7
	0.050	0.050	0.000	96.8	96.2	95.6	94.8	96.8	96.5	95.5	95.3	96.8	94.5	92.9	98.7
	0.100	0.050	0.050	96.1	94.4	94.3	93.9	95.3	94.1	92.6	93.9	95.8	95.3	95.1	83.9
0.5	0.000	0.050	-0.050	95.2	94.6	94.6	93.9	95.8	94.4	93.7	95.0	93.2	94.2	93.9	85.8
	0.200	0.300	-0.100	92.6	95.0	94.0	93.5	92.8	94.8	93.6	93.3	94.2	93.5	94.5	100
	0.000	0.000	0.000	96.7	95.2	93.0	94.3	99.1	97.8	94.7	91.2	96.0	94.1	91.0	67.5
	0.051	0.000	0.051	95.5	94.7	94.5	93.9	99.4	97.9	95.3	92.4	94.9	93.1	93.1	62.5
	0.050	0.050	0.000	94.3	95.9	95.0	94.8	95.4	92.8	92.2	93.5	95.6	94.7	92.7	99.3
	0.100	0.050	0.050	96.2	94.4	94.1	93.6	95.5	93.9	93.0	94.7	95.2	94.0	93.4	84.2
0.9	0.000	0.050	-0.050	95.9	95.1	94.2	94.5	94.8	94.7	93.4	93.5	94.7	94.7	94.4	86.5
	0.200	0.300	-0.100	95.4	93.4	93.7	94.0	94.5	93.1	93.0	93.1	95.8	93.4	94.2	100
	0.000	0.000	0.000	96.4	95.3	94.5	93.9	99.2	97.6	94.4	93.6	97.1	94.7	93.0	69.2
	0.051	0.000	0.051	95.8	95.3	95.0	94.6	98.8	97.6	94.1	91.4	94.7	94.7	94.1	66.3
	0.050	0.050	0.000	95.6	95.2	94.8	95.0	95.1	94.8	94.3	95.3	96.7	94.6	93.8	99.1
	0.100	0.050	0.050	94.6	95.1	94.8	95.4	93.5	95.2	94.3	94.1	93.7	94.8	94.6	83.4

Jackknife: jackknife CIs; Boot: bootstrap percentile CIs; Bc: bias-corrected bootstrap CIs; Bt: bootstrap-*t* CIs.

Table 2.4. Coverage probabilities of four confidence interval methods for overall and decomposed mediation effects in a dichotomous outcome model with two mediators (one binary (M_1) and one continuous (M_2)), n = 250 per group.

ρ (M_1, M_2)	True IE	True IE_1	True IE_2	Overall IE				IE_1				IE_2			
				Jack knife	Boot	Bc	Bt	Jack knife	Boot	Bc	Bt	Jack knife	Boot	Bc	Bt
-0.5	0.000	0.050	-0.050	93.6	94.7	94.2	93.8	96.3	96.0	95.0	95.4	94.0	93.3	93.4	81.7
	0.200	0.300	-0.100	94.6	95.3	95.6	95.2	95.3	95.7	95.4	95.2	94.9	95.9	96.0	100
	0.000	0.000	0.000	96.8	94.9	94.2	95.3	97.8	96.0	93.6	93.8	96.9	94.1	93.0	64.1
	0.051	0.000	0.051	95.4	94.3	94.3	94.9	98.2	94.5	91.8	93.2	95.1	94.4	94.5	58.3
	0.050	0.050	0.000	95.1	94.3	93.8	94.7	95.7	94.8	94.3	95.3	95.8	94.0	93.2	98.9
	0.100	0.050	0.050	94.8	93.7	93.3	92.9	95.8	95.4	94.5	95.5	94.6	94.0	93.8	82.8
0.0	0.000	0.050	-0.050	92.8	94.2	94.2	94.6	94.7	95.3	94.5	95.0	93.2	94.2	93.7	83.5
	0.200	0.300	-0.100	95.3	94.3	94.8	93.2	95.4	94.2	94.6	93.5	95.2	94.7	95.1	100
	0.000	0.000	0.000	95.8	95.6	95.1	94.3	97.9	96.1	93.1	93.5	96.5	94.7	94.3	65.5
	0.051	0.000	0.051	93.9	95.8	96.4	95.2	98.5	95.6	93.4	94.1	94.0	94.9	95.4	59.6
	0.050	0.050	0.000	95.4	93.6	93.6	93.9	95.5	93.7	93.5	94.8	95.0	94.9	93.7	99.6
	0.100	0.050	0.050	94.8	94.2	94.0	94.3	95.7	94.9	94.0	93.6	94.0	94.4	94.1	84.3
0.5	0.000	0.050	-0.050	95.0	94.9	94.7	94.5	95.9	94.7	95.0	94.6	94.1	95.5	95.6	86.4
	0.200	0.300	-0.100	95.5	95.6	95.4	95.4	95.5	95.4	95.5	94.6	94.6	94.5	94.6	100
	0.000	0.000	0.000	96.1	95.3	94.8	95.5	98.7	96.7	93.1	93.4	96.2	95.3	95.0	65.0
	0.051	0.000	0.051	94.1	95.4	95.6	94.8	98.3	95.2	91.4	92.5	94.4	94.9	94.7	57.9
	0.050	0.050	0.000	94.5	94.3	94.2	94.2	95.4	95.1	95.0	94.4	95.8	94.9	94.2	99.6
	0.100	0.050	0.050	94.8	95.3	95.0	95.6	94.8	95.5	95.3	95.7	95.7	95.8	95.6	84.9
0.9	0.000	0.050	-0.050	94.8	96.5	96.0	96.1	94.9	94.4	94.2	94.2	95.2	93.7	94.1	85.0
	0.200	0.300	-0.100	95.0	96.3	96.7	95.4	94.6	96.7	96.3	95.2	95.5	95.8	95.5	100
	0.000	0.000	0.000	96.8	95.4	94.3	95.2	98.1	95.5	93.2	94.0	96.0	94.7	93.4	62.8
	0.051	0.000	0.051	95.1	94.6	94.5	94.8	98.6	95.1	92.7	93.3	95.4	94.4	94.0	64.0
	0.050	0.050	0.000	96.7	95.2	94.9	95.3	95.5	94.0	93.9	94.6	96.0	94.9	94.0	99.2
	0.100	0.050	0.050	95.6	95.0	94.4	94.0	95.1	94.8	94.9	93.9	95.4	92.6	91.9	83.4

Jackknife: jackknife CIs; Boot: bootstrap percentile CIs; Bc: bias-corrected bootstrap CIs; Bt: bootstrap-t CIs.

Table 2.5. Power of bootstrap percentile CIs to detect non-zero mediation effect for a dichotomous outcome model with two mediators (one binary (M_1) and one continuous (M_2)), n = 100 and 250 per group.

ρ (M_1, M_2)	True *IE*	True IE_1	True IE_2	Sample Size					
				100			250		
				Overall *IE*	IE_1	IE_2	Overall *IE*	IE_1	IE_2
-0.5	0.000	0.050	-0.050	-	24.8	18.6	-	54.8	35.0
	0.200	0.300	-0.100	36.6	91.6	95.6	74.0	100	100
	0.000	0.000	0.000	-	-	-	-	-	-
	0.051	0.000	0.051	33.2	-	53.0	70.2	-	89.1
	0.050	0.050	0.000	14.3	26.0	-	33.6	54.3	-
	0.100	0.050	0.050	26.1	24.4	16.0	55.0	53.0	33.8
0.0	0.000	0.050	-0.050	-	30.9	20.4	-	59.6	41.6
	0.200	0.300	-0.100	51.7	94.4	97.4	90.5	100	100
	0.000	0.000	0.000	-	-	-	-	-	-
	0.051	0.000	0.051	42.4	-	57.4	86.1	-	93.9
	0.050	0.050	0.000	18.6	27.3	-	46.0	61.8	-
	0.100	0.050	0.050	38.2	30.6	16.7	75.9	63.3	37.5
0.5	0.000	0.050	-0.050	-	25.6	17.3	-	53.5	36.4
	0.200	0.300	-0.100	62.2	94.1	96.2	94.4	100	100
	0.000	0.000	0.000	-	-	-	-	-	-
	0.051	0.000	0.051	53.7	-	53.9	93.0	-	88.1
	0.050	0.050	0.000	25.1	27.2	-	54.8	52.5	-
	0.100	0.050	0.050	45.3	26.7	16.1	85.6	52.0	34.0
0.9	0.000	0.050	-0.050	-	17.5	12.8	-	38.4	24.4
	0.200	0.300	-0.100	54.9	85.7	87.9	93.9	99.9	100
	0.000	0.000	0.000	-	-	-	-	-	-
	0.051	0.000	0.051	48.5	-	33.7	89.9	-	71.7
	0.050	0.050	0.000	30.3	19.3	-	62.3	39.2	-
	0.100	0.050	0.050	56.5	17.8	13.3	91.2	36.8	26.2

2.7 Study Example

The illustrative example considered here comes from a dental caries study measuring the number of decayed, filled, or missing teeth (#DMFT) at around age of 14 years for a cohort of very low birth group (VLBW) and a matched group of normal birth weight (NBW) children. The exposure variable of interest is the binary variable, socioeconomic status (SES), coded as SES = 1 for low SES (‘exposed’), SES = 0, for high SES (‘unexposed’). The outcome considered in this example is the dichotomous DMFT variable, DMFT = 0 (#DMFT = 0) versus DMFT = 1 (#DMFT > 0). Baseline covariates adjusted for in the model include birth group (VLBW vs. NBW), sex and race. We also considered the following potential mediators: “Sealant” (a binary variable indicating use of sealants), “AvgOHI” (the average oral hygiene index score with higher values indicating worse oral hygiene status), and “Visit” (a binary variable indicating whether the child received regular (at least once a year) checkups from the dentist or not). We wanted to assess the direct effect of SES on DMFT and its indirect effect through Sealant, AvgOHI and Visit.

The dental dataset we used for the analysis included 129 subjects in the exposed group (SES = 1) and 74 subjects in the unexposed group (SES = 0). 79/129 (61.2%) of subjects in the exposed group and 28/74 (37.8%) of subjects in the unexposed group had at least one DMFT indicating that children from families with low SES may have a higher risk of developing dental caries compared with those from families with high SES.

The system of models (including the outcome model (50) and the mediator model (51) for the selected mediators, each of which incorporated the three baseline covariates, birth group, sex and race), were fitted using maximum likelihood estimation. First, we considered the two-mediator model including Sealant and AvgOHI. Estimates of TE, DE, IE, IE_1 and IE_2 using the mediation formula approach with 'exact' integration (SAS, 'QUAD' function), and the Monte Carlo approximate integration approach (using 10,000 bivariate samples of $M_1^*(t_1)$ and $M_2(t_2)$) are provided in Table 2.6; 95% CIs for these estimators were computed using bootstrap percentile methodology. The 'exact' and Monte Carlo approaches provided almost identical mediation effect estimates as well as CIs. The results indicate that the total natural indirect effect accounts for approximate 25% of the total exposure effect and the mediation effect through AvgOHI predominates among the individual mediation effects. The effect estimates are interpreted as follows. Low (versus high) SES increases the probability of DMFT = 1 by 0.19 (95% confidence interval: 0.04, 0.34), by an estimated 0.04 (-0.00, 0.09) (0.00 denotes '< 0.005' in absolute value, same as below) attributable to the two mediators (Sealant and AvgOHI), and an estimated 0.15 (0.00, 0.30) due to the direct effect (or other unknown pathways). The overall indirect effect (attributable to the two mediators together) can be decomposed into mediation effects through Sealant (-0.00 (-0.02, 0.02)) and through AvgOHI (0.04 (0.01, 0.09)). In summary, we found, using the two-mediator model for the dental data, a significant total exposure effect (TE), direct effect (DE), and indirect effect through AvgOHI (IE_2), a marginally significant total indirect effect estimate (IE) and a non-significant indirect effect through Sealant (IE_1) (at the 0.05 level based on the confidence interval). We next included one additional potential mediator, Visit, in the model; the analysis results of this model are shown in Table 2.7. Similar conclusions were drawn as above. Low (versus high) SES increases the probability of DMFT = 1 by an estimated 0.18 (0.03, 0.34) with an estimated increase of 0.05 (0.00, 0.11) attributable to the set of mediators (with the following individual mediation effects: Sealant 0.00 (-0.02, 0.02), AvgOHI 0.04 (0.00, 0.08) and Visit 0.02 (-0.01, 0.05)) and an estimated increase of 0.13 (-0.03, 0.29) due to the direct effect (that is, any other pathways). Comparing with the results from Table 2.6, we may conclude that part of the direct effect in the two-mediator model was explained by an indirect effect through the added mediator Visit, although the indirect effect through Visit is not statistically significant.

We conducted two sensitivity analyses in the two-mediator model for our dental data. In the first sensitivity analysis, we examined the effect on the individual mediation effects, IE_1 and IE_2, of varying the correlation coefficient ρ' between $M_1^*(0)$ and $M_2(1)$ from -1 to 1 in increments of 0.01. The change in IE_1 and IE_2 estimates over ρ' is shown in Figure 2.2. The plots indicate that the IE_1 estimate increases (Figure 2.2A) and the IE_2 estimate decreases (Figure 2.2B) as ρ' increases; the sum of IE_1 and IE_2 is constant, since ρ' doesn't affect the total indirect effect estimate. Based on repeated computations of the confidence intervals, we found a nonsignificant indirect effect through Sealant and a significant indirect effect through AvgOHI over the whole range of ρ' indicating robustness of the decomposed mediation effect estimates to the correlation between

Table 2.6. Estimated causal effects (and 95% bootstrap percentile CIs) based on the mediation formula approach with ‘exact’ integration and Monte Carlo integration in the two-mediator model using dental data.

		Mediation Formula	
		‘exact’ Integration	Monte Carlo Integration#
Sealant (binary) and AvgOHI (continuous)	Total *TE*	0.19 (0.04, 0.34)	0.19 (0.04, 0.33)
	Direct *DE*	0.15 (0.00$, 0.30)	0.15 (0.00, 0.30)
	Mediation *IE*	0.04 (-0.00, 0.09)	0.04 (-0.00, 0.09)
	Through Sealant (IE_1)*	-0.00 (-0.02, 0.02)	-0.00 (-0.02, 0.02)
	Through AvgOHI (IE_2)*	0.04 (0.01, 0.09)	0.04 (0.01, 0.09)

#: Monte Carlo integration with 10,000 samples was used for estimation of causal effects.

*: Assumes correlation coefficient (ρ') between unobserved counterfactual values $M_1^*(0)$ and $M_2(1)$ the same as that between $M_1^*(0)$ and $M_2(0)$ and $M_1^*(1)$ and $M_2(1)$.

$: 0.00 denotes < 0.005 in absolute value.

Table 2.7. Estimated causal effects (and 95% bootstrap percentile CIs) based on the mediation formula approach with ‘exact’ integration and Monte Carlo integration in a three-mediator model using dental data.

		Mediation Formula	
		‘exact’ Integration	Monte Carlo Integration#
Sealant (binary), AvgOHI (continuous), and Visit (binary)	Total *TE*	0.18 (0.03, 0.34)	0.18 (0.03, 0.34)
	Direct *DE*	0.13 (-0.03, 0.29)	0.13 (-0.03, 0.29)
	Mediation *IE*	0.05 (0.00$, 0.11)	0.05 (0.00, 0.11)
	Through Sealant (IE_1)*	0.00 (-0.02, 0.02)	0.00 (-0.02, 0.02)
	Through AvgOHI and Visit (IE_{23})*	0.05 (0.01, 0.11)	0.05 (0.01, 0.11)
	Through AvgOHI (IE_2)*	0.04 (0.00, 0.08)	0.04 (0.00, 0.08)
	Through Sealant and Visit (IE_{13})*	0.02 (-0.02, 0.05)	0.02 (-0.02, 0.05)
	Through Visit (IE_3)*	0.02 (-0.01, 0.05)	0.02 (-0.01, 0.05)
	Through Sealant and AvgOHI (IE_{12})*	0.04 (-0.00, 0.08)	0.04 (-0.00, 0.08)

#: Monte Carlo integration with 1,000,000 samples was used for estimation of causal effects.

*: Assumes correlation between $M_i^{(*)}(t_i)$, $M_j^{(*)}(t_j)$ (i, j = 1, 2, 3 and $i \neq j$) is constant over all t_i and t_j.

$: 0.00 denotes < 0.005 in absolute value.

counterfactuals M_1*(0) and $M_2(1)$. In the second sensitivity analysis, we assessed the effect on mediation effect estimates (*IE*, IE_1 and IE_2) of possible violation of the no unmeasured mediator-outcome confounding assumption. Since there are two correlation parameters of interest, 3-D graphs (Figure 2.3) were plotted describing the change of mediation effect estimates over the ranges of the two correlations. Namely, the correlations are those between each of the two mediators (or underlying latent variables) and the latent variable underlying outcome variable adjusting for the mediators. We find that ρ_1 (the correlation between $Y^*(t_0, m_1, m_2)$ and $M_1^*(t_1)$) does not substantially affect estimate of overall indirect effect (*IE*), and as ρ_2 (the correlation between $Y^*(t_0, m_1, m_2)$ and $M_2(t_2)$) increases, *IE* estimate significantly decreases (Figure 2.3A). Furthermore, *IE* estimate is contained in the bounded interval [-0.14, 0.16], and when ρ_2 is greater than 0.56, we find significant negative *IE* estimate. For the IE_1 estimate, it is stable and trivial (bounded by [-0.01, 0.00]) over the range of ρ_1 and ρ_2 (Figure 2.3B). The fact that the 95% CIs recalculated over the full bivariate range for the two correlation coefficients always contain zero indicates the robustness of the original conclusion (that there is no evidence for a mediation effect through Sealant). The pattern of the change in the IE_2 estimate (over the ranges of ρ_1 and ρ_2) is similar with that of *IE* estimate (Figure 2.3C); IE_2 is close to *IE* in this example since the estimate of IE_1 is approximate zero. This result indicates that our original conclusion regarding the significant positive indirect effect through AvgOHI may not hold if unmeasured confounders (explaining a correlation of greater than 0.5) between the mediator and outcome exist.

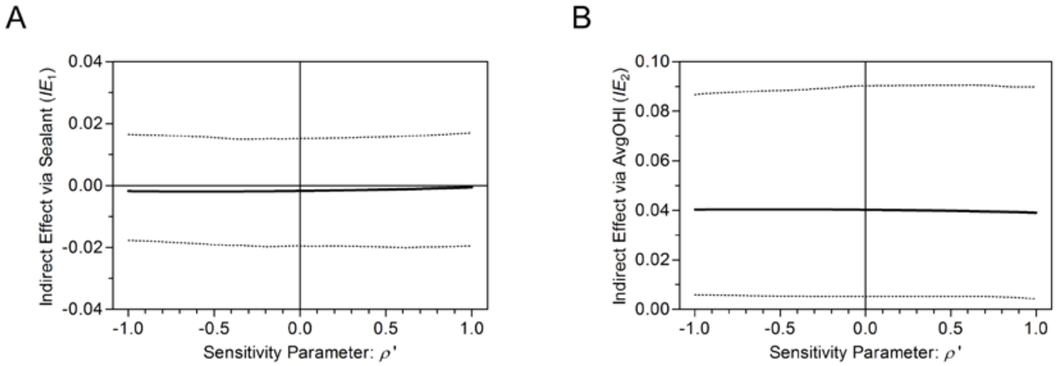

Figure 2.2. Sensitivity analysis for overall indirect effect decomposition of dental data involving binary ‘Sealant’ and continuous ‘AvgOHI’ mediators. Panels A and B show the estimated indirect effect through ‘Sealant’ and ‘AvgOHI’, respectively, for varying correlation ρ' between $M_1*(0)$ and $M_2(1)$. The areas between dotted lines represent the 95% CIs for the decomposed natural indirect effect estimator at each value of ρ'.

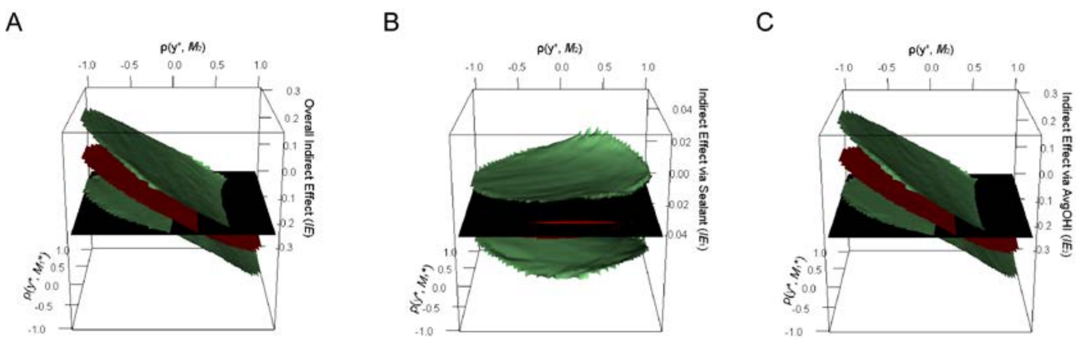

Figure 2.3. Sensitivity analysis for overall indirect effect estimation and indirect effect decomposition of dental data involving binary ‘Sealant’ (M_1) and continuous ‘AvgOHI’ (M_2) mediators. Panels A shows the estimated overall indirect effect for varying correlation ρ_1 (between $Y^*(t_0, m_1, m_2)$ and $M_1*(t_1)$) and ρ_2 (between $Y^*(t_0, m_1, m_2)$ and $M_2(t_2)$). Panel B shows indirect effect estimation through ‘Sealant’, and Panel C shows indirect effect estimation through ‘AvgOHI’. Red surface shows the mediation effect estimate, the space between the two green surfaces represents the 95% confidence intervals, and the black surface indicates zero mediation effect reference.

2.8 Discussion

In this part, we consider multiple-mediator models with binary or continuous mediators and a binary outcome. A marginal joint modeling approach with a probit-type model for binary mediator is applied to accommodate correlations among multiple mediators. The total and individual mediation effects are estimated with a mediation formula approach and sensitivity analyses are conducted to assess the robustness of the results.

Our simulation study showed good properties (low bias and close-to-nominal confidence interval coverage rates) for the proposed estimators under most scenarios. For the estimation of individual natural indirect effects, there appears to be some bias when there are substantial mediation effects through each mediator and the correlation coefficient between the two mediators is high. The reason for this bias may be due to the fact that our approach cannot dissect the total indirect effect accurately with the relatively small sample size of 200 when the two mediators are highly correlated. We note, however, that the total natural indirect effect can be estimated precisely in this situation. Comparing four methods of constructing CIs based on the coverage probability, we recommend that CIs for mediation effects in our multiple-mediator model context be obtained using the bootstrap percentile method. Preacher and Hayes (2008) advocate the bias-corrected bootstrap for CIs of mediation effects estimated with the product of coefficients approach in linear multiple-mediator models. However, we found that CIs from the bias-corrected bootstrap may be biased when there are substantial unbalanced indirect effects through both mediators with different signs and the sample size is small. An interesting finding is that the power for detecting a total indirect effect increases as the correlation between two mediators increases and the power for individual-mediator indirect effects is maximized when the two mediators are uncorrelated. Therefore it is relatively easy to detect a total mediation effect when the two mediators are highly correlated, whereas individual mediation effects are most easily detected with uncorrelated mediators.

Multiple-mediator models allow us to dissect the total mediation effect and estimate mediation effects through individual or sets of mediators under additional assumptions. Estimation of the decomposed mediation effects involves the joint distribution of counterfactuals that cannot be observed at the same time (namely, $M_1(0)$ and $M_2(1)$ in our data example). This problem, which may occur when there are more than two stages of mediation, was recognized previously in the literature (Avin, et al., 2007; Albert and Nelson, 2011; Wang and Albert, 2012). The simple identifying assumption of a constant correlation coefficient between counterfactuals $M_1^*(t_1)$ and $M_2(t_2)$ for all t_1, t_2 may not always be plausible. Therefore, it may be sensible to make use of bounds for the decomposed mediation effect estimate by considering ρ' over a plausible range, or even over the entire interval, [-1, 1]. To assess the effect on mediation effect estimates of violations of the no unmeasured mediator-outcome confounders assumption, we use the correlation coefficients between each mediator (or underlying latent variables) and the latent variable underlying the outcome variable (adjusting for the mediators), denoted as

ρ_1 and ρ_2, as our sensitivity analysis parameters. Alternatively, Imai et al. (2010) expressed the correlation coefficient between the mediator and the outcome as a function of the coefficients of determination (i.e. R^2, which represent the proportion of previously unexplained variance either in the mediator or outcome that is explained by the unobserved confounders), allowing for the sensitivity analysis to be based on the magnitudes of an effect of the omitted confounder. Imai et al. (2010) also extended this approach to the case of a binary mediator and/or binary outcome using the pseudo-R^2 of McKelvey and Zavoina (1975). This approach may possibly be applied to our model. Although R^2 as a sensitivity parameter has an advantage of interpretability, a disadvantage of this approach is that it requires two such R^2s instead of a single parameter (ρ). One consequence is that the graphical presentation of the sensitivity analysis results is more difficult, especially for our multiple-mediator models.

In conclusion, we have proposed a mediation formula approach for mediation analysis of multiple-mediator models with a dichotomous outcome that allows estimation of the total indirect effect, as well as further decomposition of the total indirect effect through individual mediators. This approach can be easily extended to other types of discrete outcomes (such as a count response). Future work is needed to explore possible dimension reduction strategies for mediation analysis with multiple mediators in non-linear models to avoid the extensive computation required by the present approach.

Appendix I: Proof of Theorem 1 (Section 1.2.1.5)

$$E(Y(t, M(t'))) = \int E(Y(t, M(t')) \mid W = w) dF_W(w) \qquad \textit{Definition}$$

$$= \iint E(Y(t, m) \mid M(t') = m, W = w) dF_{M(t')|W=w}(m) dF_W(w) \qquad \textit{Definition}$$

$$= \iint E(Y(t, m) \mid M(t') = m, T = t', W = w) dF_{M(t')|W=w}(m) dF_W(w) \qquad (10)$$

$$= \iint E(Y(t, m) \mid T = t', W = w) dF_{M(t')|W=w}(m) dF_W(w) \qquad (11)$$

$$= \iint E(Y(t, m) \mid T = t, W = w) dF_{M(t')|W=w}(m) dF_W(w) \qquad (10)$$

$$= \iint E(Y(t, m) \mid M(t) = m, T = t, W = w) dF_{M(t')|W=w}(m) dF_W(w) \qquad (11)$$

$$= \iint E(Y(t, m) \mid M = m, T = t, W = w) dF_{M(t')|W=w}(m) dF_W(w) \qquad \textit{Consisitency}$$

$$= \iint E(Y(t, m) \mid M = m, T = t, W = w) dF_{M(t')|T=t', W=w}(m) dF_W(w) \qquad (10)$$

$$= \iint E(Y \mid M = m, T = t, W = w) dF_{M|T=t', W=w}(m) dF_W(w) \qquad \textit{Consistency}$$

Appendix II: Derivation of ZINB and NBH mean and variance (Section 1.2.2)

For ZINB model:

$$E(y)=\sum_{y=1}^{\infty} y\times P(y;\psi,\lambda,\phi)=\sum_{y=0}^{\infty} y\times P(y;\psi,\lambda,\phi)$$
$$=\sum_{y=0}^{\infty} y\times\psi\times nb(y;\lambda,\phi)=\psi\times\sum_{y=0}^{\infty} y\times nb(y;\lambda,\phi)=\psi\times\lambda$$

$$Var(y)=\sum_{y=0}^{\infty}\left[y-E(y)\right]^2\times P(y;\psi,\lambda,\phi)=\left[0-\psi\times\lambda\right]^2\times\left[(1-\psi)+\psi\, nb(0;\lambda,\phi)\right]+\sum_{y=1}^{\infty}\left[y-\psi\times\lambda\right]^2\times\psi\times nb(y;\lambda,\phi)$$
$$=\sum_{y=0}^{\infty}\left[y-\psi\times\lambda\right]^2\times\psi\times nb(y;\lambda,\phi)+\psi^2\lambda^2(1-\psi)=\sum_{y=0}^{\infty}\left[y-\lambda+\lambda-\psi\times\lambda\right]^2\times\psi\times nb(y;\lambda,\phi)+\psi^2\lambda^2(1-\psi)$$
$$=\sum_{y=0}^{\infty}\left[y-\lambda\right]^2\times\psi\times nb(y;\lambda,\phi)+\sum_{y=0}^{\infty}\left\{\left[(1-\psi)\lambda\right]^2+2(1-\psi)\lambda(y-\lambda)\right\}\times\psi\times nb(y;\lambda,\phi)+\psi^2\lambda^2(1-\psi)$$
$$=\psi\times Var\left[nb(y;\lambda,\phi)\right]+\left[(1-\psi)\lambda\right]^2\times\psi+\psi^2\lambda^2(1-\psi)=\psi\lambda(1+\phi\lambda)+\left[(1-\psi)\lambda\right]^2\times\psi+\psi^2\lambda^2(1-\psi)$$
$$=\psi\lambda\left[1+\phi\lambda+\lambda+\psi^2\lambda-2\psi\lambda+\psi\lambda-\psi^2\lambda\right]=\psi\lambda\left[1+\lambda(1+\phi-\psi)\right]$$

For NBH model:

$$E(y)=\sum_{y=1}^{\infty} y\times P(y;\psi,\lambda,\phi)=\sum_{y=0}^{\infty} y\times P(y;\psi,\lambda,\phi)$$
$$=\sum_{y=0}^{\infty} y\times\frac{\psi}{1-p_0}\times nb(y;\lambda,\phi)=\frac{\psi}{1-p_0}\times\sum_{y=0}^{\infty} y\times nb(y;\lambda,\phi)=\frac{\psi\times\lambda}{1-p_0}$$

$$Var(y)=\sum_{y=0}^{\infty}\left[y-E(y)\right]^2\times P(y;\psi,\lambda,\phi)=\left[0-\frac{\psi\times\lambda}{1-p_0}\right]^2\times(1-\psi)+\sum_{y=1}^{\infty}\left[y-\frac{\psi\times\lambda}{1-p_0}\right]^2\times\frac{\psi}{1-p_0}\times nb(y;\lambda,\phi)$$
$$=\sum_{y=0}^{\infty}\left[y-\frac{\psi\times\lambda}{1-p_0}\right]^2\times\frac{\psi}{1-p_0}\times nb(y;\lambda,\phi)\;-\left[0-\frac{\psi\times\lambda}{1-p_0}\right]^2\times\frac{\psi}{1-p_0}\times nb(0;\lambda,\phi)+\left[0-\frac{\psi\times\lambda}{1-p_0}\right]^2\times(1-\psi)$$
$$=\sum_{y=0}^{\infty}\left[y-\lambda+\lambda-\frac{\psi\times\lambda}{1-p_0}\right]^2\times\frac{\psi}{1-p_0}\times nb(y;\lambda,\phi)\;+\frac{\psi^2\lambda^2}{(1-p_0)^2}\times\frac{\psi}{1-p_0}\times nb(0;\lambda,\phi)+\frac{\psi^2\lambda^2}{(1-p_0)^2}\times(1-\psi)$$
$$=\frac{\psi}{1-p_0}\times Var\left[nb(y;\lambda,\phi)\right]+\lambda^2\times\left(1-\frac{\psi}{1-p_0}\right)^2\times\frac{\psi}{1-p_0}+\frac{\psi^2\lambda^2}{(1-p_0)^2}\times\frac{\psi}{1-p_0}\times nb(0;\lambda,\phi)+\frac{\psi^2\lambda^2}{(1-p_0)^2}\times(1-\psi)$$
$$=\frac{\psi\lambda}{1-p_0}\times\left\{\frac{(1+\phi\lambda)\times(1-p_0)+\psi\lambda\times(1-\psi)+\lambda(1-p_0)-2\psi\lambda}{1-p_0}++\frac{\psi^2\lambda}{(1-p_0)^2}\left[1+nb(0;\lambda,\phi)\right]\right\}$$
$$=\frac{\psi\lambda}{1-p_0}\times\left\{\frac{1+\phi\lambda+\lambda-p_0-\phi\lambda p_0-\lambda p_0-\psi\lambda-\psi^2\lambda}{1-p_0}+\frac{\psi^2\lambda}{(1-p_0)^2}\left[1+nb(0;\lambda,\phi)\right]\right\}$$

Appendix III: Proof of Formula (34) and (35) (Section 1.5.2)

First, note that expressions for the mean of $Y(d_0, M(d_1), Y_1(d_2, M(d_{1,2})))$ in general involve the potential outcomes of M and Y_1, the latter being random variables (representing endogenous variables in the causal model (Albert and Nelson, 2011)). Therefore, obtaining the marginal mean of the potential outcome $Y(d_0, M(d_1), Y_1(d_2, M(d_{1,2})))$ requires integrating or summing over M and Y_1 as follows,

$$
\begin{aligned}
E\{Y(d_0, M(d_1), Y_1(d_2, M(d_{1,2})))\} &= \sum_{y_1=0,1} \int E\{Y(d_0, m, y_1) \mid M(d_1) = m, Y_1(d_2, M(d_{1,2})) = y_1\} \\
&\qquad P\{Y_1(d_2, M(d_{1,2})) = y_1 \mid M(d_1) = m\} dF_{M(d_1)}(m) \\
&= \int E\{Y(d_0, m, 1) \mid M(d_1) = m, Y_1(d_2, M(d_{1,2})) = 1\} P\{Y_1(d_2, M(d_{1,2})) = 1 \mid M(d_1) = m\} dF_{M(d_1)}(m)
\end{aligned}
$$

The second equation follows from the fact that when $y_1 = 0$ (the subject belongs to the non-susceptible group), $E\{Y(d_0, m, 0) \mid M(d_1) = m, Y_1(d_2, M(d_{1,2})) = 0\} = 0$. We can therefore express $E(Y(d_0, M(d_1), Y_1(d_2, M(d_{1,2}))))$, as a function of $E\{Y(d_0, m, 1) \mid M(d_1) = m, Y_1(d_2, M(d_{1,2})) = 1\}$ and $P\{Y_1(d_2, M(d_{1,2})) = 1 \mid M(d_1) = m\}$. The result that the former term is identifiable is shown as follows,

$$
\begin{aligned}
& E\{Y(d_0,m,1) \mid M(d_1)=m, Y_1(d_2,M(d_{1,2}))=1\} \\
&= E\{Y(d_0,m,1) \mid M(d_1)=m, (Y_1(d_2,m') \mid M(d_{1,2})=m')=1\} && \textit{Definition} \\
&= E\{Y(d_0,m,1) \mid M(d_1)=m, (Y_1(d_2,m') \mid T=d_{1,2}, M(d_{1,2})=m')=1\} && (31) \\
&= E\{Y(d_0,m,1) \mid M(d_1)=m, (Y_1(d_2,m') \mid T=d_{1,2})=1\} && (32) \\
&= E\{Y(d_0,m,1) \mid M(d_1)=m, (Y_1(d_2,m') \mid T=d_2)=1\} && (31) \\
&= E\{Y(d_0,m,1) \mid M(d_1)=m, (Y_1(d_2,m') \mid T=d_2, M(d_2)=m')=1\} && (32) \\
&= E\{Y(d_0,m,1) \mid M(d_1)=m, Y_1(d_2,M(d_2))=1\} && \textit{Consistency} \\
&= E\{Y(d_0,m,1) \mid T=d_1, M(d_1)=m, Y_1(d_2,M(d_2))=1\} && (31) \\
&= E\{Y(d_0,m,1) \mid T=d_1, Y_1(d_2,M(d_2))=1\} && (32) \\
&= E\{Y(d_0,m,1) \mid T=d_2, Y_1(d_2,M(d_2))=1\} && (31) \\
&= E\{Y(d_0,m,1) \mid T=d_2, M(d_2)=m, Y_1(d_2,M(d_2))=1\} && (32) \\
&= E\{Y(d_0,m,1) \mid T=d_2, M(d_2)=m\} && (33) \\
&= E\{Y(d_0,m,1) \mid T=d_2\} && (32) \\
&= E\{Y(d_0,m,1) \mid T=d_0\} && (31) \\
&= E\{Y(d_0,m,1) \mid T=d_0, M(d_0)=m\} && (32) \\
&= E\{Y(d_0,m,1) \mid T=d_0, M(d_0)=m, Y_1(d_0,M(d_0))=1\} && (33) \\
&= E\{Y \mid T=d_0, M=m, Y_1=1\} && \textit{Consistency}
\end{aligned}
$$

For formula (34), under the condition $d_1 \neq d_{1,2}$,

$$
\begin{aligned}
& P\{Y_1(d_2,M(d_{1,2}))=1 \mid M(d_1)=m\} = E\{Y_1(d_2,M(d_{1,2})) \mid M(d_1)=m\} && \textit{Definition} \\
&= \int E\{Y_1(d_2,m') \mid M(d_{1,2})=m', M(d_1)=m\} dF_{M(d_{1,2})|M(d_1)=m}(m') && \textit{Definition} \\
&= \int E\{Y_1(d_2,m') \mid T=d_1, M(d_{1,2})=m', M(d_1)=m\} dF_{M(d_{1,2})|M(d_1)=m}(m') && (31) \\
&= \int E\{Y_1(d_2,m') \mid T=d_1, M(d_{1,2})=m'\} dF_{M(d_{1,2})|M(d_1)=m}(m') && (32) \\
&= \int E\{Y_1(d_2,m') \mid T=d_{1,2}, M(d_{1,2})=m'\} dF_{M(d_{1,2})|M(d_1)=m}(m') && (31) \\
&= \int E\{Y_1(d_2,m') \mid T=d_{1,2}\} dF_{M(d_{1,2})|M(d_1)=m}(m') && (32) \\
&= \int E\{Y_1(d_2,m') \mid T=d_2\} dF_{M(d_{1,2})|M(d_1)=m}(m') && (31) \\
&= \int E\{Y_1(d_2,m') \mid T=d_2, M(d_2)=m'\} dF_{M(d_{1,2})|M(d_1)=m}(m') && (32) \\
&= \int E\{Y_1 \mid T=d_2, M=m'\} dF_{M(d_{1,2})|M(d_1)=m}(m') && \textit{Consistency}
\end{aligned}
$$

Then we have,

$$
\begin{aligned}
& E\{Y(d_0, M(d_1), Y_1(d_2, M(d_{1,2})))\} \\
&= \iint E\{Y \mid T=d_0, M=m, Y_1=1\} E\{Y_1 \mid T=d_2, M=m'\} dF_{M(d_{1,2})|M(d_1)=m}(m') dF_{M(d_1)}(m) \\
&= \iint E\{Y \mid T=d_0, M=m, Y_1=1\} E\{Y_1 \mid T=d_2, M=m'\} dF_{M(d_1),M(d_{1,2})}(m,m')
\end{aligned}
$$

Therefore when $d_1 \neq d_{1,2}$, formula (34) holds.

For formula (35), under the condition $d_1 = d_{1,2}$,

$$
\begin{aligned}
& P\{Y_1(d_2, M(d_{1,2})) = 1 \mid M(d_1) = m\} && \\
&= P\{Y_1(d_2, M(d_1)) = 1 \mid M(d_1) = m\} && \textit{Condition} \quad d_1 = d_{1,2} \\
&= E\{Y_1(d_2, m) \mid M(d_1) = m\} && \textit{Definition} \\
&= E\{Y_1(d_2, m) \mid T = d_1, M(d_1) = m\} && (31) \\
&= E\{Y_1(d_2, m) \mid T = d_1\} && (32) \\
&= E\{Y_1(d_2, m) \mid T = d_2\} && (31) \\
&= E\{Y_1(d_2, m) \mid T = d_2, M(d_2) = m\} && (32) \\
&= E\{Y_1 \mid T = d_2, M = m\} && \textit{Consistency}
\end{aligned}
$$

This gives:

$$
\begin{aligned}
& E\{Y(d_0, M(d_1), Y_1(d_2, M(d_1)))\} \\
&= \int E\{Y \mid T = d_0, M = m, Y_1 = 1\} E\{Y_1 \mid T = d_2, M = m\} dF_{M(d_1)}(m)
\end{aligned}
$$

So formula (35) also holds.

Appendix IV: Proof of Formula (47) (Section 2.4.2)

Note that, the potential outcomes for J mediators in $Y(t_0, M_1(t_1), \ldots, M_J(t_J))$ are random variables, therefore, to obtain the marginal expected value of $Y(t_0, M_1(t_1), \ldots, M_J(t_J))$ requires that we integrate over these variables, comprising the endogenous (random) explanatory variables in the model of Y (Albert and Nelson, 2011), as follows,

$$E\{Y(t_0, M_1(t_1), ..., M_J(t_J))\} =$$

$$= \underbrace{\int ... \int}_{n=J} E\{Y(t_0, m_1, ..., m_J) \mid M_1(t_1) = m_1, ..., M_J(t_J) = m_J\} d^J F_{M_1(t_1),...,M_J(t_J)}(m_1, ..., m_J) \quad \textit{Definition}$$

$$= \underbrace{\int ... \int}_{n=J} E\{Y(t_0, m_1, ..., m_J) \mid T = t_1, M_1(t_1) = m_1, ..., M_J(t_J) = m_J\} d^J F_{M_1(t_1),...,M_J(t_J)}(m_1, ..., m_J) \quad (45)$$

$$= \underbrace{\int ... \int}_{n=J} E\{Y(t_0, m_1, ..., m_J) \mid T = t_1, M_2(t_2) = m_2, ..., M_J(t_J) = m_J\} d^J F_{M_1(t_1),...,M_J(t_J)}(m_1, ..., m_J) \quad (46)$$

$$= \underbrace{\int ... \int}_{n=J} E\{Y(t_0, m_1, ..., m_J) \mid T = t_2, M_2(t_2) = m_2, ..., M_J(t_J) = m_J\} d^J F_{M_1(t_1),...,M_J(t_J)}(m_1, ..., m_J) \quad (45)$$

$$= \underbrace{\int ... \int}_{n=J} E\{Y(t_0, m_1, ..., m_J) \mid T = t_2, M_3(t_3) = m_3, ..., M_J(t_J) = m_J\} d^J F_{M_1(t_1),...,M_J(t_J)}(m_1, ..., m_J) \quad (46)$$

$$= ......$$

$$= \underbrace{\int ... \int}_{n=J} E\{Y(t_0, m_1, ..., m_J) \mid T = t_J\} d^J F_{M_1(t_1),...,M_J(t_J)}(m_1, ..., m_J) \quad (46)$$

$$= \underbrace{\int ... \int}_{n=J} E\{Y(t_0, m_1, ..., m_J) \mid T = t_0\} d^J F_{M_1(t_1),...,M_J(t_J)}(m_1, ..., m_J) \quad (45)$$

$$= \underbrace{\int ... \int}_{n=J} E\{Y(t_0, m_1, ..., m_J) \mid T = t_0, M_1(t_0) = m_1\} d^J F_{M_1(t_1),...,M_J(t_J)}(m_1, ..., m_J) \quad (46)$$

$$= \underbrace{\int ... \int}_{n=J} E\{Y(t_0, m_1, ..., m_J) \mid T = t_0, M_1(t_0) = m_1, M_2(t_0) = m_2\} d^J F_{M_1(t_1),...,M_J(t_J)}(m_1, ..., m_J) \quad (46)$$

$$= ......$$

$$= \underbrace{\int ... \int}_{n=J} E\{Y(t_0, m_1, ..., m_J) \mid T = t_0, M_1(t_0) = m_1, ..., M_J(t_0) = m_J\} d^J F_{M_1(t_1),...,M_J(t_J)}(m_1, ..., m_J) \quad (46)$$

$$= \underbrace{\int ... \int}_{n=J} E\{Y \mid T = t_0, M_1 = m_1, ..., M_J = m_J\} d^J F_{M_1(t_1),...,M_J(t_J)}(m_1, ..., m_J) \quad \textit{Consistency}$$

So formula (47) holds.

Bibliography

Ajzen I, Fishbein M. Understanding attitudes and predicting social behavior. Englewood Cliffs, 1980; NJ: Prentice Hall.

Albert JM. Mediation analysis via potential outcomes models. *Statistics in Medicine* 2008; **27**:1282-1304.

Albert JM, Nelson S. Generalized causal mediation analysis. *Biometrics* 2011; **67**:1028-1038.

Albert JM, Wang W, Nelson S. Estimating overall exposure effects for zero-inflated regression models with application to dental caries. *Statistical Methods in Medical Research* 2011; DOI: 10.1177/0962280211407800.

Allen, DG, Griffeth RW. Test of a mediated performance-turnover relationship highlighting the moderating roles of visibility and reward contingency. *Journal of Applied Psychology* 2001; **86**:1014-1021.

Avin C, Shpitser I, Pearl J. Identifiability of path-specific effects. *Proceedings of International Joint Conference on Artificial Intelligence* 2005; **19**:357-363.

Baron RM, Kenny DA. The moderator-mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology* 1986; **51**:1173-1182.

Bauman AE, Sallis JF, Dzewaltowski DA, Owen N. Toward a better understanding of the influences on physical activity: the role of determinants, correlates, causal variables, mediators, moderators, and confounders. *American Journal of Preventive Medicine* 2002; **23**:5-14.

Berk KN, Lachenbruch PA. Repeated measures with zeros. *Statistical Methods in Medical Research* 2002; **11**:303-316.

Bishop YM, Fienberg SE, Holland PW. Discrete multivariate analysis: Theory and practice. Cambridge, MA: MIT Press.

Böhning D, Dietz E, Schlattmann P, Mendonca L, Kirchner U. The zero-inflated Poisson Model and the Decayed, Missing and Filled Teeth Index in Dental Epidemiology. *Journal of the Royal Statistical Society, Series A* 1999; **162**:195-209.

Bollen KA, Stine RA. Direct and indirect effects: classical and bootstrap estimates of variability. *Sociological Methodology* 1990; **20**:115-140.

Catalano, PJ and Ryan LM. Bivariate latent variable models for clustered discrete and continuous outcomes. *Journal of the American Statistical Association* 1992; **87**:651-658.

Cheung YB. Zero-inflated models for regression analysis of count data: a study of growth and development. *Statistics in Medicine* 2002; **21**:1461-1469.

Cheung YB. Growth and cognitive function of Indonesian children: Zero-inflated proportion models. *Statistics in Medicine* 2006; **25**:3011-3022.

Cox DR, Wermuth N. Response models for mixed binary and quantitative variable. *Biometrika* 1992; **79**:441-461.

Dalrymle L, Hudson IL, Ford RPK. Finite mixture, zero-inflated Poisson and hurdle models with application to SIDS. *Computational Statistics and Data Analysis* 2003; **41**:491-504.

Dobbie MJ, Welsh AH. Modelling correlated zero-inflated count data. *Australian and New Zealand Journal of Statistics* 2001; **43**:431-444.

Donaldson SI, Graham JW, Hansen WB. Testing the generalizability of intervening mechanism theories: Understanding the effects of adolescent drug use prevention interventions. *Journal of Behavioral Medicine* 1994; **17**:195-216.

Edwards YD, Allenby GM. Multivariate analysis of multiple response data. *Journal of Marketing Research* 2003; **40**:321-334.

Efron, B, Tibshirani RJ. *An introduction to the bootstrap*. New York: Chapman & Hall.

Freedman LS, Graubard B, Schatzkin A. Statistical validation of intermediate endpoints for chronic diseases. *Statistics in Medicine* 1992; **11**:167-178.

Garrison RJ, Feinleib M, Castelli WP, MaNamara PM. Cigarette smoking as a confounder of the relationship between relative weight and long-term mortality. The Framingham Heart Study. *Journal of American Medical Association* 1983; **249**:2199-2203.

Gray BR. Selecting a distributional assumption for modeling relative densities of benthic macro-intvertebrates. *Ecological Modelling* 2005; **185**:1-12.

Gerdtham UG. Equality in health care utilization: further tests based on hurdle models and Swedish micro data. *Health Economics* 1997; **6**:303-319.

Geys H, Regan MM, Catalano PJ, Molenberghs G. Two latent variable risk assessment approaches for mixed continuous and discrete outcomes from developmental toxicity data. *Journal of Agricultural Biological and Environment Statistics* 2001; **6**:340-355.

Gurmu S. Semi-parametric estimation of hurdle regression models with an application to Medicaid utilization. *Journal of Applied Econometrics* 1997; **12**:225-242.

Hall, DB, Zhang Z. Marginal models for zero-inflated clustered data. *Statistical Modelling* 2004; **4**:161-180.

Huang B, Sivaganesan S, Succop P, Goodman E. Statistical assessment of meditational effects for logistic meditational models. *Statistics in Medicine* 2004; **23**:2713-2728.

Imai K, Keele L, Yamamoto T. Identification, inference and sensitivity analysis for causal mediation effects. *Statistical Science* 2010; **25**:51-71.

Imai K, Keele L, Tingley D. A general approach to causal mediation analysis. *Psychological Methods* 2010; **15**:309-334.

James F. Monte Carlo theory and practice. *Reports on Progress in Physics* 1980; **43**:1145-1189.

Johnson RA, Wichern DW. *Applied Multivariate Statistical Analysis* (5th edition). Prentice-Hall: Englewood Cliffs, NJ, 2002.

Lambert D. Zero-inflated Poisson regression, with an application to defects in manufacturing. *Technometrics* 1992; **34**:1-14.

Lewsey JD, Thomson WM. The utility of the zero-inflated Poisson and zero-inflated negative binomial models: a case study of cross-sectional and longitudinal DMF data examining the effect of socio-economic status. *Community Dental and Oral Epidemiology* 2004; **32**:183-189.

Li Y, Schneider JA Bennett DA. Estimation of mediation effect with a binary mediator. *Statistics in Medicine* 2007; **26**:3398-3414.

Little RJA, Schluchter MD. Maximum likelihood estimation for mixed continuous and categorical data with missing values. *Biometrika* 1985; **72**:497-512.

MacKinnon DP. Contrasts in multiple mediator models. *Multivariate Applications in Substance Use Research: New Methods for New Questions* 2000; 141-160.

MacKinnon DP, Dwyer JH. Estimating mediated effects in prevention studies. *Evaluation Review* 1993; **17**:144-158.

MacKinnon DP, Lockwood CM, Brown CH, Wang W and Hoffman JM. The intermediate endpoint effect in logistic and probit regression. *Clinical Trials* 2007; **4**:499-513.

MacKinnon DP, Lockwood CM, Hoffman JM, West SG Sheets V. A comparison of methods to test mediation and other intervening variable effects. *Psychological Methods* 2002; **7**:83-104.

MacKinnon DP, Lockwood CM, Williams J. Confidence limits for the indirect effect: distribution of the product and resampling methods. *Multivariate Behavioral Research* 2004; **39**:99-128.

MacKinnon DP, Warsi G, Dwyer JH. A simulation study of mediated effect measures. *Multivariate Behavioral Research* 1995; **30**:41-62.

Maddala GS. *Limited Dependent and Qualitative Variables in Econometrics.* Cambridge

University Press: Cambridge, UK, 1983.

McKelvey RD, Zavoina W. A statistical model for the analysis of ordinal level dependent variable. *Journal of Mathematical Sociology* 1975; **4**:103-120.

Min Y, Agresti A. Random effect models for repeated measures of zero-inflated count data. *Statistical Modelling* 2005; **5**:1-19.

Molenberghs G, Verbeke G. *Models for discrete longitudinal data*. Spring, New York, 2005.

Mosteller F, Tukey JW. *Data analysis and regression: A second course in statistics.* Reading, MA: Addison-Wesley.

Moulton LH, Curriero FC, Barroso PF. Mixture models for quantitative HIV RNA data. *Statistical Methods in Medical Research* 2002; **11**:317-325.

Mullahy, J. Specification and testing of some modified count data models. *Journal of Econometrics* 1986; **33**:341-365.

Neelon BH, O'Malley AJ, Normand SL. A Bayesian model for repeated measures zero-inflated count data with application to outpatient psychiatric service use. *Statistical Modelling* 2010; **10**:421-439.

Nelson S, Albert JM, Lombardi G, Wishnek S, Asaad G, Kirchner HL, Singer LT. Dental caries and enamel defects in very low birth weight adolescents. *Caries Research* 2010; **44**:509-518.

Pearl, J. Direct and indirect effects. In *Proceedings of the Seventeenth Conference on Uncertainty in Artificial Intelligence* (J.S. Breese and D. Koller, eds.) 2001; 411-420. Morgan Kaufman, San Francisco, CA.

Pearl J. The causal mediation formula – a guide to the assessment of pathways and mechanisms *Technical report* R-379, 2011: http://bayes.cs.ucla.edu/jp_home.html.

Petersen ML, Sinisi SE, van der Laan MJ. Estimation of direct causal effects. *Epidemiology* 2006; **17**:276-284.

Preacher KJ, Hayes AF. Asymptotic and resampling strategies for assessing and comparing indirect effects in multiple mediator models. *Behavior Research Methods* 2008; **40**:879-891.

Regan MM, Catalano PJ. Likelihood models for clustered binary and continuous outcomes: Application to developmental toxicology. *Biometrics* 1999; **55**:760-768.

Reynolds KD, Bishop DB, Chou CP, Xie B, Nebeling L, Perry CL. Contrasting mediating variables in two 5-a-day nutrition intervention programs. *Preventive Medicine* 2004; **39**:882-893.

Ridout MS, Demetrio CGB, Hinde JP. Models for count data with many zeros. *Proceedings of XIXth International Biometric Society Conference* 1998; IBC98:179-192.

Robins JM, Greenland S. Identifiability and exchangeability for direct and indirect effects. *Epidemiology* 1992; **3**:143-155.

Robins JM. Semantics of causal DAG models and the identification of direct and indirect effects. In highly Structured Stochastic Systems (P.J. Green, N. L. Hjort and S. Richardson, eds.) 2003; 70-81. Oxford Unv. Press, Oxford. MR2082403.

Rodrigues J. Bayesian analysis of zero-inflated distributions. *Communications in Statistics, Theory and Methods* 2003; **32**:281-289.

Rose CE, Martin SW, Wannemuehler KA, Plikaytis BD. On the use of zero-inflated and hurdle models for modeling vaccine adverse event count data. *Journal of Biopharmaceutical Statistics* 2006; **16**:463-481.

Rothman K, Greenland S. Modern epidemiology. Philadelphia: Lippincott-Raven Press, 2001.

Schneider JA, Bienias JL, Wilson RS, Berry-Kravis E, Evans DA, Bennett DA. The apolipoprotein E epsilon4 allele increases the odds of chronic cerebral infarction detected at autopsy in older persons. *Stroke* 2005; **36**:954-959.

Schluchter, MD. Flexible approaches to computing mediated effects in generalized linear models: generalized estimating equations and bootstrapping. Multivariate Behavioral Research 2008; **43**:268-288.

Shpitser I, VanderWeele TJ. A complete graphical criterion for the adjustment formula in mediation analysis. *International Journal of Biostatistics* 2011; **7**:Article 16.

Singer L, Yamashita T, Lilien L, Collin M, Baley J. A longitudinal study of developmental outcome of infants with bronchopulmonary dysplasia and very low birth weight. *Pediatrics* 1997; **100**:987-993.

Sobel ME. Asymptotic confidence intervals for indirect effects in structural equation models. *Sociological Methodology* 1982; **13**:290-312.

Song PXK, Li M, Yuan Y. Joint regression analysis of correlated data using Gaussian copulas. *Biometrics* 2009; **65**:60-68.

Stone CA, Sobel ME. The robustness of estimates of total indirect effects in covariance structural models estimated by maximum likelihood. *Psychometrika* 1990; **55**:337-352.

Taylor AB, MacKinnon D Tein JY. Test of the three-path mediated effect. *Organizational Research Methods* 2008; **11**:241-269.

Tein JY, Sandler IN, Zautra AJ. Stressful life events, psychological distress, coping, and parenting of divorced mothers: A longitudinal study. *Journal of Family Psychology*, 2000; **14**:27-41.

Van der Laan MJ, Petersen ML. Direct effect models. *International Journal of Biostatistics* 2008; **4**:Article 23.

VanderWeele TJ Bias formula for sensitivity analysis for direct and indirect effects. *Epidemiology* 2010; **21**:540-551.

VanderWeele TJ, Vansteelandt S. Odds ratios for mediation analysis for a dichotomous outcome. *American Journal of Epidemiology* 2010; **172**:1339-1348.

Vinokur, AD, Schul Y. Mastery and inoculation against setbacks as active ingredients in the JOBS intervention for the unemployed. *Journal of Consulting and Clinical Psychology* 1997; **65**:867-877.

Wang W, Albert JM. Estimation of mediation effects for zero-inflated regression models. Submitted to *Statistics in Medicine* 2012.

Yeates KO, Taylor HG, Woodrome SE, Wade SL, Stancin T, Drotar D. Race as a moderator of parent and family outcomes following pediatric traumatic brain injury. *Journal of Pediatric Psychology* 2002; **27**:393-403.

Yau KKW, Lee AH. Zero-inflated Poisson regression with random effects to evaluate an occupational injury prevention program. *Statistics in Medicine* 2001; **20**:2907-2920.