Mediation Analysis for Count and Zero-inflated Count Data

Jing Cheng*1, Nancy F. Cheng*, Zijian Guo**, Steve Gregorich***, Amid I. Ismail**** and Stuart A. Gansky*

* Division of Dental Public Health & Epidemiology, University of California at San Francisco
** Department of Statistics, Wharton School, University of Pennsylvania
*** Department of Medicine, School of Medicine, University of California at San Francisco
**** Department of Restorative Dentistry, Maurice H. Kornberg School of Dentistry, Temple University

Abstract

Different conventional and causal approaches have been proposed for mediation analysis to better understand the mechanism of a treatment. Count and zero-inflated count data occur in biomedicine, economics and social sciences. However, not much attention has been paid to the mediation analysis for such data. This paper considers mediation analysis for count and zero-inflated count data under the potential outcome framework with nonlinear models. When there are post-treatment confounders which are independent of, or affected by, the treatment, we first define the direct, indirect and total effects of our interest and then discuss various conditions under which the effects of interest can be identified. Proofs are provided for the sensitivity analysis proposed in the paper. Simulation studies show that the methods work well. We apply the methods to the Detroit Dental Health Project’s Motivational Interviewing DVD (DDHP MI-DVD) trial for the direct and indirect effects of motivational interviewing on count and zero-inflated count dental caries outcomes.

Keywords: Direct effect, indirect effect, post-treatment confounder, sensitivity analysis, sequential ignorability.

1 Introduction

In many health studies, the intervention is designed to change some post-randomization (intermediate) variable, such as knowledge, attitudes, behavior, biomarkers or social factors, so that the change in the intermediate variable will lead to improvement in the final health outcomes of interest ([1]). In these studies, researchers are not only interested if the intervention works but also if and how much the intervention affects the outcome through and around the intermediate variable. Such an intermediate variable is usually called a mediator; the effect of the treatment through the mediator is called indirect or mediation effect, while the effect around the mediator is called the direct effect. An indirect or mediation effect shows that

---

1 Address for correspondence: Jing Cheng, Associate Professor, Division of Dental Public Health & Epidemiology, University of California at San Francisco, 3333 California Street, Ste 495, San Francisco, CA 94143-1361, USA. Email: jing.cheng@ucsf.edu. This study was supported by grant U54DE019285 from the National Institute of Dental and Craniofacial Research.

1
the intervention affects the outcome through the intermediate variables as designed, while a
direct effect indicates that the intervention changes the outcome directly or involving some
other intermediate variables in a heretofore undiscovered mechanism. Knowing those effects
helps us to better understand the working mechanism of an intervention such that in future
research and applications in specific populations, we can tailor specific intervention compo-
ments to target important mediators and consequently lead to bigger improvement in health
outcomes.

Conventional mediation approaches since Baron and Kenny ([2, 3, 4]) (e.g., regression,
path and structural equation model (SEM)) and recently developed causal methods ([5] -
[21]) make different assumptions on the intervention and mediator to achieve a causal inter-
pretation on the indirect (mediation) effect and direct effect of the intervention through and
around a mediator. Conventional approaches model observed treatment and mediator values
and may not provide a general definition/interpretation of causal effects independent of spe-
cific statistical models. Different from conventional approaches, causal mediation approaches
first conceptually define causal direct and indirect effects under the potential outcome frame-
work ([22, 23]) without reference to a specific statistical model and then different statistical
models can be used to identify and estimate causal direct and indirect effects under different
assumptions.

Most conventional and causal approaches focus on continuous or binary mediators and/or
outcomes. However, the outcome variable in many studies is often a count following a Pois-
son or Negative Binomial distribution, or a zero-inflated count that has a higher probability
of being zero than expected under a Poisson or Negative Binomial, such as number of doc-
tor visits, number of admissions and readmissions to a hospital, number of complications,
and number of decayed, missing and filled (dmf/DMF) teeth or tooth surfaces. Assuming a
Poisson or negative binomial (NB) distribution on dental outcomes such as dmf/DMF teeth
or tooth surfaces, Albert and Nelson ([24]) developed a nice approach for mediation analysis
based on the potential outcome framework ([7]) in the context of a directed acyclic graph
(DAG) using generalized linear models, and Albert ([25]) considered an inverse-probability
weighted estimator for the mediation effect on count outcomes. Assuming a zero-inflated neg-
ative binomial (ZINB) model for the outcome, Wang and Albert ([26]) provided a mediation
formula for the mediation effect estimation in a two-stage model and considered a decompo-
sition of the mediation effect in a three-stage model with application to zero-inflated count
outcomes. Instead of decomposing the mediation (indirect) effect into different components
for ZINB data as Wang and Albert ([26]), in this paper, we are interested in the overall
direct, mediation (indirect) and total effects for Poisson, NB, zero-inflated Poisson (ZIP)
and ZINB data. Additionally, we consider cases when there are post-treatment confounders
in a study with count and zero-inflated count data, and discuss some conditions to identify
the effects of our interest and propose a sensitivity analysis under those cases.

Our study is motivated by the Detroit Dental Health Project’s Motivational Interviewing
DVD (DDHP MI-DVD) trial. It was a randomized dental trial of a Motivational Interviewing
(MI) intervention to prevent early childhood caries (ECC) in low income African-American
In the study, caregivers in both intervention and control groups watched a 15-minute education video on children’s oral health. For the intervention group, a MI interviewer reviewed the child’s dental examination with caregivers, and discussed caregivers’ personal thoughts and concerns about specific goals for their child’s oral health. A brochure with caregivers’ specific goals and a general list of 10 recommendations on diet, oral hygiene and dental visits was given to caregivers in the intervention and control groups, respectively. The outcome of interest is the number of new cavitated, and new untreated lesions 2 years later ([27]). Since the majority of the children did not have any new cavitated and new untreated lesions at the end of the study, the distribution of the outcome contains a lot of zeros (Figure 2). We are interested in whether and how much the MI intervention prevented new cavities in children through its effect in changing subjects’ oral health behavior.

The rest of the paper is organized as follows. In Section 2, we set up the causal framework, introduce notation and assumptions, and define the indirect and direct effects. In Section 3, we extend the method to estimate the indirect and direct effect in randomized trials with count or zero-inflated count outcomes. In Section 4, we present some simulation studies. An application of our method to the DDHP MI-DVD study is shown in Section 5. Finally, we provide conclusions and discussion in Section 6.

All the programming used and analyses conducted in this paper were written in R (https://cran.r-project.org/) and are available from the authors.

2 The Framework

In this study, we will use the potential (counterfactual) outcome framework ([22, 23]) to specify the direct, indirect (mediation) and overall effects of the treatment. We will make the Stable Unit Treatment Value Assumption (SUTVA) in the paper. SUTVA says that a subject’s potential outcome is not related to the randomization or mediation value of other subjects or the method of administration of randomization or the mediator. Under SUTVA, we use $Z_i$ to denote the treatment variable, $M_i$ for the observed mediator level, $X_i$ for the observed baseline covariates and $Y_i$ for observed outcome for subject $i$. In a two-arm trial, $Z_i = 1$ if subject $i$ is randomized to the intervention group and $Z_i = 0$ if randomized to the control group. We let $M_i^z$ denote the potential value of a mediator under treatment $Z_i = z$ for subject $i$, which has two versions $M_i^1$ under intervention and $M_i^0$ under control. However, in practice we are not able to observe both potential mediator values but only one of $M_i^1$ and $M_i^0$ depending on which treatment group subject $i$ was actually assigned to. We use $Y_i^{z,m}$ to denote the potential outcome subject $i$ would have under the treatment $Z_i = z$ and mediator $M_i = m$, and $Y_i^{z,M_i^z}$ for potential outcome under $Z_i = z$, where $Y_i^{z,m}$ will be used below to define controlled effects and $Y_i^{z,M_i^z}$ for natural effects. Again, we can only observe one version of multiple potential outcomes for a subject depending on the actual treatment and mediator value subject $i$ had.
The total effect (TE) or intent-to-treat (ITT) effect of the intervention and its average are

\[ TE = Y_{i}^{1,M_{i}^{1}} - Y_{i}^{0,M_{i}^{0}}, \quad \bar{TE} = E(Y_{i}^{1,M_{i}^{1}} - Y_{i}^{0,M_{i}^{0}}), \]

which is the total effect of the intervention \((Z = 1)\) on outcome \(Y\) compared to control \((Z = 0)\) no matter whether the effect is through or around mediator \(M\). The total or ITT effect of the intervention has two components: the effect of the intervention around the mediator, called the direct effect, and the effect of the intervention through the mediator, called the indirect or mediation effect. Two sets of definitions on these effects have been proposed in the literature ([6, 8, 16, 28, 29]): controlled and natural effects.

The controlled direct effect (CDE) of the intervention and its average while fixing the mediator at \(m\) are

\[ CDE_{m} = Y_{i}^{1m} - Y_{i}^{0m}, \quad \bar{CDE}_{m} = E(Y_{i}^{1m} - Y_{i}^{0m}), \]

which is the effect of intervention compared to control while fixing the mediator at \(m\); and the controlled mediation effect (CME) of \(m\) vs. \(m'\) when fixing \(z\) and its average are

\[ CME_{z} = Y_{i}^{zm} - Y_{i}^{zm'}, \quad \bar{CME}_{z} = E(Y_{i}^{zm} - Y_{i}^{zm'}), \]

for \(z = 0, 1\) and all \(m \neq m'\),

which is the effect of mediator (at \(m\) vs. \(m'\)) on the outcome under treatment \(z\).

Alternatively, instead of setting the mediator at a fixed level \(m\) in the controlled effects, the natural effects set the mediator at its “natural” level that would be achieved under treatment assignment \(z\). The natural direct effect (NDE) of intervention and its average when the mediator is set at its level under treatment assignment \(z\) are

\[ NDE_{z} = Y_{i}^{1,M_{i}^{z}} - Y_{i}^{0,M_{i}^{z}}, \quad \bar{NDE}_{z} = E(Y_{i}^{1,M_{i}^{z}} - Y_{i}^{0,M_{i}^{z}}), \]

which is the effect of intervention on outcome compared to control while having the mediator at its potential level \(M_{i}^{z}\); and the natural mediation (indirect) effect (NME) and its average when fixing treatment \(z\) are

\[ NME_{z} = Y_{i}^{z,M_{i}^{1}} - Y_{i}^{z,M_{i}^{0}}, \quad \bar{NME}_{z} = E(Y_{i}^{z,M_{i}^{1}} - Y_{i}^{z,M_{i}^{0}}), \]

which is the outcome change under treatment \(z\) that would be observed if the mediator would change from the value under control \(M_{i}^{0}\) to the value under treatment \(M_{i}^{1}\). In some studies, natural effects are probably preferred since we may not be able to set the mediator at a specific level. However, stronger assumptions are often needed to identify natural effects than controlled effects since the potential outcome corresponding to both levels of \(Z, Y_{i}^{z,M_{i}^{z'}}\) \((z \neq z')\), is involved in natural effects. In this paper, we will focus on the natural effects while the controlled effects will be mentioned in the discussion of existing approaches.
3 Mediation Analysis for Count and Zero-inflated Count Data

As discussed above, the counterfactual potential outcome involved in the natural effects $Y_{iz,M}^{z',M'} (z \neq z')$ is not observed. To identify the effects, we assume sequential ignorability as per Imai et al ([16, 30]):

$$\{Y_{iz,M}^{z',m}, M_i^z\} \perp Z_i|X_i = x; \; Y_{iz,M}^{z',m} \perp M_i^z|Z_i = z, X_i = x, \; \text{for all } z, z', m. \quad (1)$$

This assumption says that (a) given the baseline covariates, the treatment is independent of potential mediators and potential outcomes; and (b) given the treatment and baseline covariates, the mediators are independent of the potential outcomes. In the DVD-MI study, the first ignorability assumption is reasonable because participants were randomized to the MI intervention. The random assignment of the intervention does not guarantee the second ignorability assumption because the oral health behavior after randomization was not randomly assigned. However, the second ignorability assumption may hold after conditioning on baseline covariates and treatment; that is, the oral health behavior was as if randomized among subjects in the same treatment group who have the same baseline characteristics.

Under sequential ignorability, Imai, Keele and Tingley ([30]) showed that the distribution of the potential outcome is nonparametrically identified, i.e., the distribution of the potential outcome on the left hand side can be expressed as a function of the distribution of observed data on the right hand side:

$$f(Y_{iz,M}^{z',m}|X_i = x) = \int_{M} f(Y_i|m, Z_i = z, X_i = x)dF_M(m|Z_i = z', X_i = x), \; x \in X; z, z' = 0, 1. \quad (2)$$

This result allows us to estimate the potential outcome and mediators we do not observe. Based on this result, we further assume the following mediator and outcome models:

$$M_i^{Z_i} \sim f_M(\theta_M = h^{-1}(\alpha_M + \beta_M Z_i + \eta_T X_i)) \quad (3)$$

$$Y_i^{Z_i,M_i^{Z_i}} \sim f_Y(\theta_Y = g^{-1}(\alpha_Y + \beta_Y Z_i + \gamma_Y M_i^{Z_i} + \eta_Y Z_i M_i^{Z_i} + \eta_T X_i)) \quad (4)$$

where the link functions $h$ and $g$ are monotonic and differentiable functions; e.g., identity link for normally distributed $M_i$ or $Y_i$, and probit link for binary $M_i$ or $Y_i$. For a count outcome or mediator following a Poisson or Negative Binomial distribution, a loglinear model can be used as per Albert and Nelson ([24]). For zero-inflated outcomes, different approaches ([31]) have been proposed outside the mediation context. In this paper, we will adopt the zero-inflated Poisson (ZIP) ([32]) or zero-inflated Negative Binomial (ZINB) ([33]) model for zero-inflated counts in the mediation context. The basic idea of these models is that the outcome is a mixture of zeros and Poisson (or Negative Binomial) random variables with the mixture proportion $p(Z_i, M_i^{Z_i}, X_i)$ and Poisson (or Negative Binomial) mean $\lambda(Z_i, M_i^{Z_i}, X_i)$.
depending on the covariates $X_i$. When an interpretation only relies on the second part (positive outcome) of the ZIP or ZINB model, the conclusion could be misleading because the two groups with the positive outcome are not ensured to be comparable by randomization ([34]). In this paper, our estimates of direct, mediation and total effects and their comparisons between groups will use information from all the randomized subjects with both parts of the model so that the ignorability of randomization holds. The outcome distribution under ZIP is:

$$P(Y_i | Z_i, M_i = 0) = \omega_i + (1 - \omega_i)e^{-\lambda_i};$$
$$P(Y_i | Z_i, M_i = j) = (1 - \omega_i)\frac{e^{-\lambda_i \lambda_i^j}}{j!}; \quad j > 0 \tag{5}$$

while the outcome distribution under ZINB is:

$$P(Y_i | Z_i, M_i = 0) = \omega_i + (1 - \omega_i)(1 + \sigma \lambda_i)^{-\frac{1}{\sigma}};$$
$$P(Y_i | Z_i, M_i = j) = (1 - \omega_i)\frac{\Gamma(j + \frac{1}{\sigma})}{j!\Gamma(\frac{1}{\sigma})}(\sigma \lambda_i)^j(1 + \sigma \lambda_i)^{-j - \frac{1}{\sigma}}; \quad j > 0 \tag{6}$$

where

$$\log \frac{\omega_i}{1 - \omega_i} = \alpha Y_1 + \beta Y_1 Z_i + \gamma Y_1 M_i Z_i + \xi Y_1 Z_i M_i Z_i + \eta_{Y_1}^T X_i,$$
$$\log \lambda_i = \alpha Y_2 + \beta Y_2 Z_i + \gamma Y_2 M_i Z_i + \xi Y_2 Z_i M_i Z_i + \eta_{Y_2}^T X_i,$$
$$\sigma (\geq 0)$$ is a dispersion parameter that does not depend on covariates.

Then as in Imai et al. ([16, 30]), the procedure based on the quasi-Bayesian Monte Carlo approximation of King, Tomz, and Wittenberg ([35]) will be used to make inference on the direct and indirect effects of treatment:

(I) Fit the mediator and outcome models with observed mediator and outcome, and obtain estimated parameters (coefficients) and their estimated asymptotic covariance matrix.

(II) Simulate model parameters (coefficients) from their sampling distribution based on the approximate multivariate normal distribution with mean and variance equal to the estimated parameters (coefficients) and their estimated asymptotic covariance matrix obtained in (I), and sample $J$ copies of the mediator and outcome model coefficients from their sampling distributions: $\theta^j_M$ and $\theta^j_Y$.

(III) For each copy $j = 1, \ldots, J$, repeat the following steps:

(a) simulate potential values of the mediator under each $z = 0, 1$ for each subject based on the mediator model (3) with simulated parameters (coefficients) obtained in (II);
(b) simulate potential outcomes under each \( z = 0, 1 \) for each subject based on the outcome model (4) with simulated potential mediator values obtained in (a) and simulated parameters (coefficients) obtained in (II);

(c) compute the direct, mediation and total treatment effects by averaging the difference between the corresponding two predicted potential outcomes discussed in Section 2.

(IV) Compute the point estimates of direct, mediation and total effects, confidence intervals and p values based on the results from J repetitions. We use the sample median, standard deviation, and percentiles of the corresponding distributions from the J repetitions as the point estimate, standard error and confidence interval for the direct, indirect (mediation) and total effects.

4 Mediation Analysis with Post Treatment Confounders

In Section 3, we only consider situations with measured baseline confounders \( X_i \). In this section, we will consider mediation analysis for cases with some confounding after randomization. For example, to evaluate the direct and indirect effects of an educational intervention on parents around and through children’s salivary bacterial challenge at one time point on caries prevention, parents behavior of making sure children for dental visit after randomization could be associated with both the mediator salivary bacterial challenge and dental outcomes.

We let \( U_i \) denote post-treatment confounders. Figure 1 shows the treatment mechanism through and around the mediator when the treatment (a) does not affect and (b) does affect the post-treatment confounder, respectively.

4.1 Post Treatment Confounders not Affected by the Treatment

When the post-treatment confounder \( U_i \) is not affected by treatment \( Z_i \) (Figure 1(a)), average natural effects are identified ([29]) under the sequential ignorability (8):

\[
(Y_{i}^{z', m}, M_{i}^{z}) \perp Z_i | X_i = x; \quad \text{and} \quad Y_{i}^{z', m} \perp M_{i}^{z} | Z_i = z, X_i = x, U_i = u, \text{ for all } z, z', m, u. \tag{8}
\]

The first part of (8) is the same as the first part of (1), which says that the treatment is randomly assigned conditional on \( X_i \). The second part of (8) is similar to the second part of (1) except that now the ignorability of the mediator holds given not only the treatment assignment and baseline covariates but also post-treatment confounders. That is, the mediator is effectively random (independent of confounding) among subjects in the same treatment group who have the same values of baseline characteristics and post-treatment confounders.
To estimate the direct and indirect natural effects of the treatment when the post-treatment confounder $U_i$ is not affected by treatment $Z_i$, we can modify the outcome model by including the post-treatment confounder in the model:

$$Y_{i,Z_i,M_i} \sim f_Y(\theta_Y = g^{-1}(\alpha_Y + \beta_Y Z_i + \gamma_Y M_i^{Z_i} + \eta_Y^T X_i + \phi^T U_i))$$

Then the same procedure discussed in Section 3 can be used for the estimation of direct and indirect natural effects. For zero-inflated count data, (7) changes to

$$\log \frac{\omega_i}{1-\omega_i} = \alpha_{Y1} + \beta_{Y1} Z_i + \gamma_{Y1} M_i^{Z_i} + \eta_{Y1}^T X_i + \phi^T U_i,$$
$$\log \lambda_i = \alpha_{Y2} + \beta_{Y2} Z_i + \gamma_{Y2} M_i^{Z_i} + \eta_{Y2}^T X_i + \phi^T U_i$$

### 4.2 Post Treatment Confounders Affected by the Treatment

When treatment $Z_i$ affects the post-treatment confounder $U_i$ (Figure 1(b)), average natural effects are not identified under assumption (8) without additional information. Instead, average controlled effects can be estimated under sequential ignorability (8) and the extended outcome model (10):

$$Y_{i,Z_i,M_i} \sim f_Y(\theta_Y = g^{-1}(\alpha_Y + \beta_Y Z_i + \gamma_Y M_i + \eta_Y^T X_i + \phi^T U_i)).$$

The average controlled mediation effect can be estimated by a function of $\hat{\gamma}_Y$, but the estimate of the average controlled direct effect by $\hat{\beta}_Y$ could be biased ([29]) because $U_i$ is also affected by $Z_i$ and the effect through $U_i$ is not incorporated in the estimation of the controlled direct effect. For continuous outcomes with an identity link function in (10), Vansteelandt ([36]) and Joffe and Greene ([37]) used a two-stage ordinary least squares (OLS) procedure to estimate the average controlled direct effect by correcting the bias in the second stage. Some researchers considered the derivation of bounds for the natural direct and indirect effects ([38, 39, 40]). Tchetgen Tchetgen and Shpitser ([41]) and VanderWeele and Chiba ([42]) considered various contrasts of the outcome between two subpopulations as sensitivity parameters and then corrected the bias with specified values of sensitivity parameters. Tchetgen Tchetgen and VanderWeele ([43]) assumed monotonicity about the effect of the treatment (exposure) on the confounder and showed the nonparametrical identifiability of the natural direct effect. For binary mediators, Taguri and Chiba ([44]) classified subjects into four principal M-response strata and estimated the natural direct and indirect effects under additional monotonicity assumption on treatment-mediator effect and assumption of common average mediator effects between compliant and never intermediates.

In this section, we will consider a sensitivity analysis for the direct and indirect effects on count and zero-inflated count outcomes when the treatment affects the post-treatment
confounder. We consider the average natural mediation, direct and total effects as:

\[
\bar{NME}_z = E(\bar{Y}_z, U_z, M_1; U_1) - E(\bar{Y}_z, U_z, M_0; U_0), \quad \text{for } z = 0, 1
\]

\[
\bar{NDE}_z = E(Y_1, U_1, M_z; U_z) - E(Y_0, U_0, M_z; U_0), \quad \text{for } z = 0, 1
\]

\[
\bar{NTE} = E(Y_1, U_1, M_1; U_1) - E(Y_0, U_0, M_0; U_0) = \bar{NDE}_1 + \bar{NME}_0.
\]

That is, the mediation effect is the causal effect of the treatment on the outcome through the mediator \(M\) under treatment \(z\); and the direct effect is all other causal effects of the treatment on the outcome around \(M\), including the effect through the post-treatment confounder \(U\). That is, the confounding effect is included in the direct effect when it is not the interest. Please see Daniel et al. ([45]) for discussion on various approaches when more than one intermediate variables exist in a study. When effects through the post-treatment confounder \(U\) is the interest of investigators, Imai and Yamamoto ([46]) assumed a linear structural equation model for the outcome and mediators and estimated the effects, Daniel et al. ([45]) considered the finest possible decomposition of the total effect, and VanderWeele and Vansteelandt ([47]) considered the mediators one at a time as joint mediators and proposed decomposition of the total effect with regression-based and weighting approaches.

We assume sequential ignorability (12) and (13) and mediator and outcome models:

\[
\left(Y^{z,u,m}_{i,z,u}, M^{z,u'}_i, U^z_i\right) \perp Z_i \mid X_i = x, Z_i \sim f_M(\theta_M = h^{-1}(\alpha_M + \beta_M Z_i + \phi_M U^Z_i + \eta^T_M X_i))
\]

\[
Y^{z,u,m}_{i,z,u} \perp M^{z,u'}_i \mid X_i = x, Z_i = z, U^z_i = u
\]

\[
M^{Z_i}_i \sim f_M(\theta_M = h^{-1}(\alpha_M + \beta_M Z_i + \phi_M U^Z_i + \eta^T_M X_i))
\]

\[
Y^{Z_i,m}_{i,z} \sim f_Y(\theta_Y = g^{-1}(\alpha_Y + \beta_Y Z_i + \gamma_Y M^{Z_i}_i + \phi_Y U^{Z_i}_i + \eta^T_Y X_i))
\]

Additionally, we assume various models below for the post-treatment confounder \(U^{Z_i}_i\). Then we can show that the effects (11) are identified under (12) - (15) and one of (16) - (18).

\[
U^1_i = U^0_i + \beta_U + \tau^T_U X_i, \quad \text{where } \delta_i \perp (Z_i, X_i, U^0_i, Y^{z,u,m}_{i,z,u}, M^{z,u'}_i)
\]

Models (16) - (18) are good for continuous post-treatment confounders, where Model (18) allows the heterogeneity treatment effect on \(U\) for individuals. For a binary confounder \(U\), one can also assume an underlying continuous variable following one of Models (16) -
For general post-treatment confounders, we assume the following set of assumptions to identify the effects (11),

\[
\begin{align*}
\left( Y_i^{z,u,m}, M_i^{z',u',1}, U_i^0, U_i^0 \right) \perp Z_i \mid X_i = x \\
Y_i^{z,u,m} \perp M_i^{z',u'} \mid X_i = x, Z_i = z, U_i^0 = u, U_i^1 = u'
\end{align*}
\]  \hspace{0.5cm} (19)

and

\[
U_i^{Z_i} \sim f_U(\theta_U = o^{-1}(\alpha_U + \beta_U Z_i + \tau_U^T X_i)),
\]

\[
U_i^1 \perp U_i^0 \mid X_i = x \quad \text{and} \quad \left( Y_i^{z,u,m}, M_i^{z',u'} \right) \perp (U_i^0, U_i^1) \mid X_i = x, Z_i = z
\]  \hspace{0.5cm} (20)

Note that (19) and (20) are slightly different from the assumptions (12) and (13) since (19) and (20) are involved with joint distribution of \((U_i^0, U_i^1)\) while (12) and (13) are only involved with marginal distribution \(U_i^z\). In practice, if the ignorability holds for marginal distribution \(U_i^z\), it is reasonable to assume that the ignorability also holds for the joint distribution \(U_i^1\) and \(U_i^0\).

**Result 1** Given sequential ignorability (12) and (13), mediator model (14) and outcome model (15), and one of confounder models (16) - (18), then the average effects \(\bar{NME}_z\), \(\bar{NDE}_z\) and \(\bar{NTE}_z\) are identified. Given sequential ignorability (13), (19) and (20), mediator model (14) and outcome model (15), and the confounder model (21), then the average effects \(\bar{NME}_z\), \(\bar{NDE}_z\) and \(\bar{NTE}_z\) are identified.

Please see the Appendix for the proof. Note that Model (21) works for general post-treatment confounders while The procedure based on the quasi-Bayesian Monte Carlo approximation ([35]) discussed in Section 3 can then be used for inference on the direct, mediation and total treatment effects but with one additional confounder model (16), (17), (18) or (21).

In a real study, we can conduct a sensitivity analysis by varying the values of parameters \(\beta_U\) and \(\tau_U^T\) one or two at a time and see how the estimates of effects (11) will change. Although we are not able to know the values of those parameters for sure, information from the study is helpful for specifying values of those parameters under sequential ignorability (12) and (13) or (19) and (20). Estimates from a regression of observed \(U_i\) on treatment \(Z_i\), covariates \(X_i\) and their interaction can provide reasonable starting points for the choice of values for \(\beta_U\) and \(\tau_U^T\) in the sensitivity analysis. For example, in \(U_i = \alpha_U + \delta_U Z_i + \nu_U X_i + \epsilon_i\), \(\delta_U\) would be a reasonable starting value for \(\beta_U\) in (16). We suggest to use the estimated value based on observables \(\pm c\%\) (say \(\pm 5\%, 50\%\) or \(100\%\)) of the estimated value as a range for the parameters, where the choice of \(c\%\) will be based on expert knowledge in a study such that the range will represent the possible treatment effect on the confounder. Then equally divided 10-20 values in the range can be used for the sensitivity analysis.
5 Simulation Studies

In this section, we will present simulation studies to examine the finite sample performance of the methods discussed in Sections 3 and 4.

The treatment $Z_i$ was assigned randomly with a probability of 0.5 to either treatment or control group. The baseline covariates were drawn independently from $N(0, 1)$, Bernoulli(0.5), and/or multinomial((1, 2, 3, 4), (0.25, 0.25, 0.25, 0.25)). The results are similar with different types of covariates and only results with the normal and binary covariates are reported. We consider both continuous and binary mediators:

\[
M_i^{Z_i} \sim N(\alpha_M + \beta_M Z_i + \eta_M^T X_i, 1); \quad M_i^{Z_i} \sim \text{Binary}(\frac{\exp(\alpha_M + \beta_M Z_i + \eta_M^T X_i)}{1 + \exp(\alpha_M + \beta_M Z_i + \eta_M^T X_i)})
\]

Four families of outcome distributions were considered in the simulation studies: Poisson (Poi), Negative Binomial (NB), Zero-inflated Poisson (ZIP) and Zero-inflated Negative Binomial (ZINB).

\[
Y_i^{Z_i,M_i^{Z_i}} \sim \text{Poi}(\exp[\alpha_Y + \beta_Y Z_i + (\gamma_Y + \delta Z_i)M_i^{Z_i} + \eta_Y^T X_i])
\]
\[
Y_i^{Z_i,M_i^{Z_i}} \sim \text{NB}(\exp[\alpha_Y + \beta_Y Z_i + (\gamma_Y + \delta Z_i)M_i^{Z_i} + \eta_Y^T X_i], \text{size} = c)
\]
\[
Y_i^{Z_i,M_i^{Z_i}} = 0 \text{ with } p_i = \expit\{\alpha_Y + \beta_Y Z_i + (\gamma_Y + \delta Z_i)M_i^{Z_i} + \eta_Y^T X_i\}; \quad Y_i^{Z_i,M_i^{Z_i}} \sim \text{Poi}(\exp[\alpha_Y + \beta_Y Z_i + (\gamma_Y + \delta Z_i)M_i^{Z_i} + \eta_Y^T X_i]) \text{ with } (1 - p_i).
\]
\[
Y_i^{Z_i,M_i^{Z_i}} = 0 \text{ with } p_i = \expit\{\alpha_Y + \beta_Y Z_i + (\gamma_Y + \delta Z_i)M_i^{Z_i} + \eta_Y^T X_i\}; \quad Y_i^{Z_i,M_i^{Z_i}} \sim \text{NB}(\exp[\alpha_Y + \beta_Y Z_i + (\gamma_Y + \delta Z_i)M_i^{Z_i} + \eta_Y^T X_i], \text{size} = c) \text{ with } (1 - p_i).
\]

There are about 20% zeroes for Poisson and negative binomial data and about 50% zeroes for ZIP and ZINB data. For each distribution family, we simulated one setting where the treatment affected the outcome and about 30% of its effect was through the mediator, and another setting corresponding to the null hypothesis of no direct and indirect effects. For each setting, we performed 1,000 Monte Carlo replications, generating data for 100 and 500 subjects respectively on each replication.

In the simulation for cases with a post-treatment confounder affected by the treatment, we consider and present results from one normal $U$ model with a normal covariate but other $U$ models work similarly.

\[
U_i^0 \sim N(\alpha_U + \eta_U^T X_i, \sigma_U^2), \quad U_i^1 = U_i^0 + \beta_U;
\]
The corresponding mediator and outcome models are:

\[
M_i^{Z_i} \sim N(\alpha_M + \beta_M Z_i + \phi_M U_i^{Z_i} + \eta_M^T X_i, 1);
\]

\[
M_i^{Z_i} \sim \text{Binary} \left( \frac{\exp(\alpha_M + \beta_M Z_i + \phi_M U_i^{Z_i} + \eta_M^T X_i)}{1 + \exp(\alpha_M + \beta_M Z_i + \phi_M U_i^{Z_i} + \eta_M^T X_i)} \right)
\]

\[
Y_i^{Z_i,M_i^{Z_i}} \sim \text{Poi}(\exp(\alpha_Y + \beta_Y Z_i + \gamma_Y M_i^{Z_i} + \phi_Y U_i^{Z_i} + \eta_Y^T X_i))
\]

\[
Y_i^{Z_i,M_i^{Z_i}} \sim \text{NB}(\exp(\alpha_Y + \beta_Y Z_i + \gamma_Y M_i^{Z_i} + \phi_Y U_i^{Z_i} + \eta_Y^T X_i), \text{size} = c)
\]

\[
y_i^{Z_i,M_i^{Z_i}} = 0 \text{ with } p_i = \text{expit}(\alpha_Y + \beta_Y Z_i + \gamma_Y M_i^{Z_i} + \phi_Y U_i^{Z_i} + \eta_Y^T X_i);
\]

\[
y_i^{Z_i,M_i^{Z_i}} \sim \text{Poi}(\exp(\alpha_Y + \beta_Y Z_i + \gamma_Y M_i^{Z_i} + \phi_Y U_i^{Z_i} + \eta_Y^T X_i)) \text{ with } (1 - p_i).
\]

\[
y_i^{Z_i,M_i^{Z_i}} = 0 \text{ with } p_i = \text{expit}(\alpha_Y + \beta_Y Z_i + \gamma_Y M_i^{Z_i} + \phi_Y U_i^{Z_i} + \eta_Y^T X_i);
\]

\[
y_i^{Z_i,M_i^{Z_i}} \sim \text{NB}(\exp(\alpha_Y + \beta_Y Z_i + \gamma_Y M_i^{Z_i} + \phi_Y U_i^{Z_i} + \eta_Y^T X_i), \text{size} = c) \text{ with } (1 - p_i).
\]

The true values of natural direct, indirect (mediation) and total effects were computed as the average difference between two corresponding potential outcomes with the true values of the parameters (coefficients). The average estimated values, root mean squared errors (RMSE), confidence interval coverages, and empirical rejection rates for a level of 0.05 are shown in Tables 1 and 2 without and with the post-treatment confounder (from (16)) respectively when there are direct and mediation effects (the alternative hypothesis is true). We can see that the bias and RMSEs are small under all outcome distributions with and without post treatment confounders. The 95% confidence interval coverage is good for most cases, but when there is a post-treatment confounder affected by the treatment, the coverage is less than 95% for the mediation effect for ZIP and ZINB, where around 40% – 70% observations are zero. The test has higher power to detect direct and total effects than the power to detect mediation effects, and the power to detect the mediation effect is increased when the sample size is increased from 100 to 500. A more detailed investigation on the power will be performed in a future study. When there are no direct and indirect effects (the null hypothesis is true), the pattern of results is similar with Type I error < 0.05 for all cases (The results are not shown to save space).

6 Application

In this section, we will apply the method discussed in this paper to the DDHP MI-DVD trial ([27]). In the study, 790 families (0-5 years old children and their caregivers) were randomly assigned to one of two education groups (DVD only or MI+DVD). Both groups of families received a copy of a special 15-minute DVD on how the caregivers could help their children stay free from tooth decay. Additionally the families in the intervention group (MI+DVD) met a MI interviewer, developed their own preventive goals, and received booster calls within 6 months of the intervention. Table 3 shows the baseline characteristics of participants by randomization assignment.
The primary analyses of the study showed that caregivers in the MI+DVD group were more likely to make sure their child brushed at bedtime at 6 months and 2 years, but the intervention did not have a significant effect on children’s dental outcomes at 2 years ([27]). In this study, we are interested in if there was a direct effect of the intervention on children’s dental outcomes that cancelled out a mediation effect in an opposite direction so that no significant total effect of the intervention was found. The dental outcomes of interest include the number of new untreated lesions, number of decayed, missing and filled surfaces (dmfs) and number of decayed, missing and filled teeth (dmft) at 2 years. The number of new untreated lesions, dmfs and dmft took values of integers and had around 60%, 26% and 47% zeros respectively. Figure 2 shows the dental outcome histograms by group at two years.

Because the motivating interview had a significant effect on parents’ behavior to make sure their child brushed at bedtime at 6 months, we used it as a mediator and would like to see if the intervention had an effect on dental outcomes through parents’ behavior to make sure their child brushed at bedtime. A logistic regressions was used to model the binary mediator whether or not parents made sure their child brushed at bedtime at 6 months on intervention while controlling for baseline covariates such as soda consumption, household income, caregivers’ education, number of times child brushed, whether or not caregivers made sure their child brushed, whether or not caregivers provided child healthy meals, and dental visits at baseline. Poisson, Negative Binomial, zero-inflated Poisson and zero-inflated Negative Binomial outcome models were fitted for the number of new untreated cavities, new dmfs and new dmft at two years with intervention, mediator, and baseline covariates included in the models. Baseline covariates, such as soda consumption, household income, caregivers’ education, number of times child brushed, whether or not caregivers made sure their child brushed, whether or not caregivers provided child healthy meals, and dental visits at baseline, are known as important confounders in the relations between the MI-DVD intervention and dental outcomes and between the mediator (whether or not parents made sure their child brushed at bedtime) and dental outcome. Controlling for those covariates, the sequential ignorability seems a reasonable assumption in this study. The Vuong test ([48]) was used to compare different outcome models and showed that the zero-inflated Negative Binomial outcome models were preferred. Table 4 shows the estimated direct, mediation and total effects for the three dental outcomes. None of the direct, mediation and total effects were significant, indicating no significant evidence that the effect of the intervention on parents made sure children brushed at bedtime translated to an improvement of dental outcomes at 2 years.

The sensitivity analyses considered a potential post-treatment confounder: Parents made sure the children see a dentist every six months. We modeled the post-treatment confounder on treatment and baseline covariates. The estimated treatment effect on the confounder was -0.15, and we used $-0.15 \pm \frac{1}{2}(-0.15)$, i.e., $(-0.20, -0.10)$ as the reasonable range for $\beta_U$ in terms of possible treatment effect on the confounder. Figure 3 shows that with various values of $\beta_U$, the mediation effects stayed around 0 while the direct and total effects vary within a range from 0.03 for untreated cavities to 0.15 for dmfs.
7 Discussion

This paper considers mediation analysis for count and zero-inflated count outcomes – common outcomes in dental studies and other fields. Sequential ignorability is assumed in the methods discussed in this paper. Although the mediator is not randomly assigned such that the ignorability of the mediator is not guaranteed, the assumption is more likely satisfied after controlling for relevant baseline covariates. When we evaluate the direct and mediation effect of the treatment through a mediator of interest, it is common that there are some other intermediate variables, which are affected by treatment and also associated with the outcome and mediator of interest, so called post-treatment confounders. Those post-treatment confounders make the evaluation of natural direct and mediation effect difficult. In this paper, we consider mediation sensitivity analysis with the presence of post-treatment confounders by modeling the post-treatment confounders on treatment and baseline covariates along with quasi-Bayesian Monte Carlo approximation based g-computation. This method allows us to evaluate the natural direct and mediation effects with sensitivity parameters easily specified.

In addition to the dental outcomes discussed in this paper, healthcare utilizations such as the number of doctor visits or emergency visits and number of admissions and readmissions to a hospital, and medical outcomes such as the number of complications, are often count or zero-inflated count data. The methods discussed in this paper can be applied to those data. Important baseline confounders should be controlled in the mediator and outcome models such that the sequential ignorability is a reasonable assumption. When there is a concern of a post-treatment confounder which is affected by the treatment, sensitivity analysis proposed in this paper should be considered to see how the results will change while the sensitivity parameters vary in a realistic range in the study.

Acknowledgements

This study was made possible by Grant Number U54 DE 019285 from the National Institute for Dental and Craniofacial Research (NIDCR), a component of the National Institute of Health (NIH). The authors thank Dylan S. Small for valuable discussions on the paper. The authors are grateful to the referees and Editors for insightful comments that improved the paper.

References


Appendix

Proof of Result 1

To estimate the natural direct and indirect effect, it is essential to estimate 

\[ E \left( Y_{i}^{0,U_{i}^{0},M_{i}^{1,U_{i}^{1}}} \right), \]

\[ E \left( Y_{i}^{1,U_{i}^{1},M_{i}^{1,U_{i}^{1}}} \right) \text{ and } E \left( Y_{i}^{0,U_{i}^{0},M_{i}^{0,U_{i}^{0}}} \right). \]

Let \( F_{Z} (\cdot) \) and \( F_{Z|W} (\cdot) \) represent the distribution function of a random variable \( Z \) and the conditional distribution function of \( Z \) given \( W \).
Note that
\[
E \left( Y_{i}^{0,0,0,0} \right) \\
= \int E \left( Y_{i}^{0,0,0,0} \mid X_i = x \right) dF_{X_i} (x) \\
= \int E \left( Y_{i}^{0,0,0,0} \mid X_i = x, U_i^0 = u \right) dF_{U_i^0 | X_i = x} (u) dF_{X_i} (x) \\
= \int E \left( Y_{i}^{0,0,0,0} \mid X_i = x, U_i^0 = u, M_i^0 = m \right) dF_{M_i^0 | X_i = x, U_i^0 = u} (m) dF_{U_i^0 | X_i = x} (u) dF_{X_i} (x) .
\]

By the ignorability assumption (12), we have
\[
E \left( Y_{i}^{0,0,0,0} \mid X_i = x, U_i^0 = u, M_i^0 = m \right) = E \left( Y_i^0 \mid X_i = x, Z_i = 0, U_i = u, M_i = m \right) = E \left( Y_i \mid X_i = x, Z_i = 0, U_i = u, M_i = m \right) ,
\]
and
\[
dF_{M_i^0 | X_i = x, U_i^0 = u} (m) = dF_{M_i^0 | X_i = x, Z_i = 0, U_i = u} (m) = dF_{M_i | X_i = x, Z_i = 0, U_i = u} (m) ,
\]
\[
dF_{U_i^0 | X_i = x} (u) = dF_{U_i^0 | X_i = x, Z_i = 0} (u) = dF_{U_i | X_i = x, Z_i = 0} (u) .
\]
By combining (22), (23) and (24), we have
\[
E \left( Y_{i}^{0,0,0,0} \right) = \int E \left( Y_i \mid X_i = x, Z_i = 0, U_i = u, M_i = m \right) dF_{M_i | X_i = x, Z_i = 0, U_i = u} (m) dF_{U_i | X_i = x, Z_i = 0} (u) dF_{X_i} (x) .
\]

Similarly, we can also obtain
\[
E \left( Y_{i}^{1,1,1,1} \right) = \int E \left( Y_i \mid X_i = x, Z_i = 1, U_i = u, M_i = m \right) dF_{M_i | X_i = x, Z_i = 1, U_i = u} (m) dF_{U_i | X_i = x, Z_i = 1} (u) dF_{X_i} (x) .
\]
By (12), we have
\( E \left( Y_{i,1}^{0,0, \mathcal{M}_i^{1,1}} \right) \)
\[= \int E \left( Y_{i,1}^{0,0, \mathcal{M}_i^{1,1}} | X_i = x \right) dF_{X_i}(x) \]
\[= \int E \left( Y_{i,1}^{0,0, \mathcal{M}_i^{1,1}} | X_i = x, U_i^0 = u \right) dF_{U_i^0 | X_i = x}(u) dF_{X_i}(x) \]
\[= \int E \left( Y_{i,1}^{0,0, \mathcal{M}_i^{1,1}} | X_i = x, U_i^0 = u, U_i^1 = u' \right) dF_{U_i^1 | X_i = x, U_i^0 = u}(u') dF_{U_i^0 | X_i = x}(u) dF_{X_i}(x) \]
\[= \int E \left( Y_{i,1}^{0,0, \mathcal{M}_i^{1,1}} | X_i = x, U_i^0 = u, U_i^1 = u', M_i^{1,1} = m \right) \]
\[dF_{M_i^{1,1} | U_i^0 = u', X_i = x, U_i^0 = u}(u') dF_{U_i^0 | X_i = x}(u) dF_{X_i}(x). \]  
\[(27)\]

Proof under Model (16)

By (16), we have
\[dF_{U_i^1 | X_i = x, U_i^0 = u}(u') = 1_{u'=u+\beta_U}, \]  
\[(28)\]
where \(1_{u'=u+\beta_U}\) is the indicator function taking value 1 when \(u' = u + \beta_U\) and value 0 on all other places. Hence, (27) can be expressed as
\[E \left( Y_{i,1}^{0,0, \mathcal{M}_i^{1,1}} \right) = \int E \left( Y_{i,1}^{0,0, \mathcal{M}_i^{1,1}} | X_i = x, U_i^0 = u, M_i^{1,1} = m \right) \]
\[dF_{M_i^{1,1} | U_i^0 = u+\beta_U, X_i = x}(u') dF_{U_i^0 | X_i = x}(u) dF_{X_i}(x). \]
\[(29)\]

Note that
\[dF_{U_i^0 | X_i = x}(u) = dF_{U_i^0 | X_i = x, Z_i = 0}(u). \]

The remaining goal is to identify the following quantities,
\[E \left( Y_{i,1}^{0,0, \mathcal{M}_i^{1,1}} | X_i = x, U_i^0 = u, M_i^{1,1} = m \right), \]
\[(30)\]
and
\[dF_{M_i^{1,1} | U_i^0 = u+\beta_U, X_i = x}. \]
\[(31)\]

By (12), we have
\[dF_{M_i^{1,1} | U_i^0 = u+\beta_U, X_i = x} = dF_{M_i^{1,1} | X_i = x, Z_i = 1, U_i^1 = u+\beta_U}(m) \]
\[= dF_{M_i | X_i = x, Z_i = 1, U_i = u+\beta_U}(m). \]
\[(32)\]
For the conditional expectation part, we have
\[
E \left( Y_{i}^{0,u,m} \mid X_i = x, U_i^0 = u, M_i^{1,u+\beta_U} = m \right)
= E \left( Y_{i}^{0,u,m} \mid X_i = x, Z_i = 0, U_i^0 = u, M_i^{1,u+\beta_U} = m \right)
= E \left( Y_{i}^{0,u,m} \mid X_i = x, Z_i = 0, U_i^0 = u \right)
= E \left( Y_{i}^{0,u,m} \mid X_i = x, Z_i = 0, U_i^0 = u, M_i^{0,u} = m \right)
= E \left( Y_{i} \mid X_i = x, Z_i = 0, U_i = u, M_i = m \right),
\]
where the first equality follows from (12) and the second and third equalities follow from (13). Combing (28), (32) and (33), (27) can be expressed as
\[
E \left( Y_{i}^{0,U_0,M_i^{1,u^T}} \right) = \int E \left( Y_i \mid X_i = x, Z_i = 0, U_i = u, M_i = m \right) dF_{M_i \mid X_i=x,Z_i=0,U_i=\beta_U}(m) dF_{U_i \mid X_i=x,Z_i=0}(u) dF_X(x).
\]
(34)

Proof under Model (17)
All the results under Model (16) hold by replacing \( \beta_U \) with \( \beta_U + \tau^T_0 x \).

Proof under Model (18)
By (27), we have
\[
E \left( Y_{i}^{0,U_0,M_i^{1,u^T}} \right) = \int E \left( Y_{i}^{0,u,M_i^{1,u'}} \mid X_i = x, U_i^0 = u, U_i^1 = u' \right) dF_{U_i^1 \mid X_i=x,U_i^0=\beta_U}(u') dF_{U_i^0 \mid X_i=x}(u) dF_X(x)
= \int E \left( Y_{i}^{0,u,M_i^{1,u+\beta_U+\tau^T_0 x + \delta}} \mid X_i = x, U_i^0 = u, U_i^1 = u + \beta_U + \tau^T_0 x + \delta \right) dF_{\delta_i}(\delta) dF_{U_i^0 \mid X_i=x}(u) dF_X(x)
= \int E \left( Y_{i}^{0,u,m} \mid X_i = x, U_i^0 = u, U_i^1 = u + \beta_U + \tau^T_0 x + \delta, M_i^{1,u+\beta_U+\tau^T_0 x + \delta} = m \right)
\]
\[
\int dF_{M_i^{1,u+\beta_U+\tau^T_0 x + \delta}}(u') dF_{\delta_i}(\delta) dF_{U_i^0 \mid X_i=x}(u) dF_X(x).
\]
(35)

By the assumption \( \delta_i \perp (Z_i, X_i, U_i^0, Y_i^{z,a,m}, M_i^{x,a'}) \), we have
\[
E \left( Y_{i}^{0,U_0,M_i^{1,u^T}} \right) = \int E \left( Y_i \mid X_i = x, Z_i = 0, U_i = u, M_i = m \right) dF_{M_i \mid X_i=x,Z_i=1,U_i=\beta_U+\tau^T_0 x + \delta}(m) dF_{\delta_i}(\delta) dF_{U_i \mid X_i=x,Z_i=0}(u) dF_X(x).
\]
(36)

Proof under Model (21)

21
By (21), we have
\[ dF_{U_i^1 \mid X_i = x, U_i^0 = u, U_i^1 = u'} = dF_{U_i^1 \mid X_i = x, U_i^0 = u, U_i^1 = u'} (u'). \]

By (20), we have
\[ dF_{M_i^{1,u'} \mid X_i = x, U_i^0 = u, U_i^1 = u'} (m) = dF_{M_i^{1,u'} \mid X_i = x, Z_i = 1} (m) \]
\[ = dF_{M_i^{1,u'} \mid X_i = x, Z_i = 1, U_i^1 = u} \]

Note that
\[
E \left( Y_{i,u,m}^0 \mid X_i = x, U_i^0 = u, U_i^1 = u', M_i^{1,u'} = m \right) \\
= E \left( Y_{i,u,m}^0 \mid X_i = x, Z_i = 0, U_i^0 = u, U_i^1 = u', M_i^{1,u'} = m \right) \\
= E \left( Y_{i,u,m}^0 \mid X_i = x, Z_i = 0, U_i^0 = u, U_i^1 = u \right) \\
= E \left( Y_{i,u,m}^0 \mid X_i = x, Z_i = 0 \right) \\
= E \left( Y_{i,u,m}^0 \mid X_i = x, Z_i = 0, U_i^0 = u \right) \\
= E \left( Y_{i,u,m}^0 \mid X_i = x, Z_i = 0, U_i = u, M_i = m \right),
\]

where the first equality follows from (19), the second and the forth equality follow from (20) and the third equality follows from (13). Then,
\[
E \left( Y_{i,u,m}^0 \mid M_i^{1,u'} \right) = \int E \left( Y_i \mid X_i = x, Z_i = 0, U_i = u, M_i = m \right) \\
\quad \quad \quad \quad \quad \quad \quad dF_{M_i \mid X_i = x, Z_i = 1} \left( u' \right) dF_{U_i \mid X_i = x, U_i^1 = 1} \left( u' \right) dF_{X_i \mid x} (x)
\]
Figure 1: Treatment mechanism when $Z$ does not affect $U$ (a) and when $Z$ affects $U$ (b)
Figure 2: Histograms of the numbers of new untreated cavities, new dmfs and new dmft in participants at 2 years in DDHP MI-DVD study.
Figure 3: Sensitivity analysis for direct, mediation and total effects on the numbers of new untreated cavities, new dmfs and new dmft with varying treatment effects on the post treatment confounder $\beta_U$
<table>
<thead>
<tr>
<th>D'n</th>
<th>M</th>
<th>N</th>
<th>Z</th>
<th>NDE</th>
<th>NDE</th>
<th>RMSE</th>
<th>Cov.</th>
<th>rr</th>
<th>NTE</th>
<th>NTE</th>
<th>RMSE</th>
<th>Cov.</th>
<th>rr</th>
</tr>
</thead>
<tbody>
<tr>
<td>Poi</td>
<td>N</td>
<td>100</td>
<td>1</td>
<td>1.641</td>
<td>1.656</td>
<td>0.019</td>
<td>98.3</td>
<td>0.979</td>
<td>0.647</td>
<td>0.662</td>
<td>0.018</td>
<td>94.3</td>
<td>0.690</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>1.268</td>
<td>1.286</td>
<td>0.020</td>
<td>98.0</td>
<td>0.978</td>
<td>0.273</td>
<td>0.291</td>
<td>0.019</td>
<td>94.1</td>
<td>0.485</td>
<td></td>
<td></td>
</tr>
<tr>
<td>B</td>
<td>100</td>
<td>1</td>
<td>1.517</td>
<td>1.504</td>
<td>0.018</td>
<td>98.8</td>
<td>0.959</td>
<td>0.464</td>
<td>0.451</td>
<td>0.018</td>
<td>99.8</td>
<td>0.104</td>
<td>1.813</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>1.349</td>
<td>1.341</td>
<td>0.013</td>
<td>97.9</td>
<td>0.946</td>
<td>0.297</td>
<td>0.289</td>
<td>0.011</td>
<td>99.1</td>
<td>0.104</td>
<td></td>
<td></td>
</tr>
<tr>
<td>NB</td>
<td>N</td>
<td>100</td>
<td>1</td>
<td>1.513</td>
<td>1.551</td>
<td>0.043</td>
<td>95.0</td>
<td>0.725</td>
<td>0.724</td>
<td>0.028</td>
<td>95.6</td>
<td>0.508</td>
<td>1.935</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>1.236</td>
<td>1.278</td>
<td>0.045</td>
<td>94.5</td>
<td>0.725</td>
<td>0.422</td>
<td>0.451</td>
<td>0.030</td>
<td>94.6</td>
<td>0.508</td>
<td></td>
<td></td>
</tr>
<tr>
<td>B</td>
<td>100</td>
<td>1</td>
<td>1.513</td>
<td>1.554</td>
<td>0.047</td>
<td>96.1</td>
<td>0.547</td>
<td>0.471</td>
<td>0.471</td>
<td>0.013</td>
<td>98.4</td>
<td>0.094</td>
<td>1.804</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>1.333</td>
<td>1.382</td>
<td>0.054</td>
<td>95.5</td>
<td>0.532</td>
<td>0.290</td>
<td>0.299</td>
<td>0.012</td>
<td>95.9</td>
<td>0.092</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ZINB</td>
<td>N</td>
<td>100</td>
<td>1</td>
<td>1.614</td>
<td>1.706</td>
<td>0.099</td>
<td>99.9</td>
<td>0.544</td>
<td>0.675</td>
<td>0.675</td>
<td>0.014</td>
<td>96.2</td>
<td>0.295</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>1.262</td>
<td>1.626</td>
<td>0.032</td>
<td>99.7</td>
<td>0.984</td>
<td>0.538</td>
<td>0.522</td>
<td>0.020</td>
<td>93.3</td>
<td>0.963</td>
<td></td>
<td></td>
</tr>
<tr>
<td>B</td>
<td>100</td>
<td>1</td>
<td>1.539</td>
<td>1.565</td>
<td>0.037</td>
<td>99.2</td>
<td>0.501</td>
<td>0.498</td>
<td>0.016</td>
<td>91.4</td>
<td>0.129</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>1.338</td>
<td>1.381</td>
<td>0.048</td>
<td>98.6</td>
<td>0.498</td>
<td>0.309</td>
<td>0.313</td>
<td>0.007</td>
<td>94.0</td>
<td>0.118</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 1: Simulation results without post-treatment confounders
<table>
<thead>
<tr>
<th>D'n</th>
<th>M</th>
<th>N</th>
<th>Z</th>
<th>NDE</th>
<th>NME</th>
<th>NTE</th>
<th>RMSE</th>
<th>Cov.</th>
<th>r²</th>
<th>NDE</th>
<th>NME</th>
<th>NTE</th>
<th>RMSE</th>
<th>Cov.</th>
<th>r²</th>
</tr>
</thead>
<tbody>
<tr>
<td>Poi</td>
<td>0</td>
<td>0.568</td>
<td>0.582</td>
<td>0.016</td>
<td>97.2</td>
<td>0.571</td>
<td>0.122</td>
<td>0.142</td>
<td>0.020</td>
<td>95.3</td>
<td>0.199</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>500</td>
<td>0.709</td>
<td>0.704</td>
<td>0.006</td>
<td>95.1</td>
<td>1.000</td>
<td>0.264</td>
<td>0.258</td>
<td>0.007</td>
<td>98.7</td>
<td>0.950</td>
<td>0.833</td>
<td>0.834</td>
<td>0.004</td>
<td>96.7</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>0.569</td>
<td>0.576</td>
<td>0.008</td>
<td>95.5</td>
<td>1.000</td>
<td>0.124</td>
<td>0.130</td>
<td>0.006</td>
<td>93.9</td>
<td>0.950</td>
<td>0.833</td>
<td>0.834</td>
<td>0.004</td>
<td>96.7</td>
</tr>
<tr>
<td>B</td>
<td>100</td>
<td>0.721</td>
<td>0.710</td>
<td>0.016</td>
<td>96.8</td>
<td>0.430</td>
<td>0.307</td>
<td>0.279</td>
<td>0.029</td>
<td>93.8</td>
<td>0.412</td>
<td>0.907</td>
<td>0.900</td>
<td>0.013</td>
<td>97.3</td>
</tr>
<tr>
<td></td>
<td>500</td>
<td>0.721</td>
<td>0.714</td>
<td>0.009</td>
<td>97.8</td>
<td>0.991</td>
<td>0.313</td>
<td>0.300</td>
<td>0.013</td>
<td>95.2</td>
<td>1.000</td>
<td>0.910</td>
<td>0.909</td>
<td>0.005</td>
<td>98.1</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>0.598</td>
<td>0.609</td>
<td>0.013</td>
<td>96.9</td>
<td>0.982</td>
<td>0.189</td>
<td>0.195</td>
<td>0.007</td>
<td>92.8</td>
<td>1.000</td>
<td>0.982</td>
<td>0.828</td>
<td>0.844</td>
<td>0.006</td>
</tr>
<tr>
<td>NB</td>
<td>100</td>
<td>0.696</td>
<td>0.709</td>
<td>0.018</td>
<td>97.1</td>
<td>0.363</td>
<td>0.252</td>
<td>0.267</td>
<td>0.017</td>
<td>96.3</td>
<td>0.246</td>
<td>0.817</td>
<td>0.854</td>
<td>0.039</td>
<td>97.6</td>
</tr>
<tr>
<td></td>
<td>500</td>
<td>0.565</td>
<td>0.587</td>
<td>0.024</td>
<td>96.2</td>
<td>0.336</td>
<td>0.120</td>
<td>0.145</td>
<td>0.025</td>
<td>93.6</td>
<td>0.113</td>
<td>0.396</td>
<td>0.415</td>
<td>0.030</td>
<td>95.6</td>
</tr>
<tr>
<td>B</td>
<td>100</td>
<td>0.698</td>
<td>0.699</td>
<td>0.005</td>
<td>98.0</td>
<td>0.982</td>
<td>0.253</td>
<td>0.251</td>
<td>0.004</td>
<td>95.6</td>
<td>0.934</td>
<td>0.820</td>
<td>0.829</td>
<td>0.010</td>
<td>98.3</td>
</tr>
<tr>
<td></td>
<td>500</td>
<td>0.566</td>
<td>0.578</td>
<td>0.013</td>
<td>98.1</td>
<td>0.983</td>
<td>0.122</td>
<td>0.130</td>
<td>0.008</td>
<td>91.7</td>
<td>0.920</td>
<td>0.153</td>
<td>0.156</td>
<td>0.009</td>
<td>96.5</td>
</tr>
<tr>
<td>ZIP</td>
<td>100</td>
<td>0.751</td>
<td>0.743</td>
<td>0.011</td>
<td>99.0</td>
<td>0.949</td>
<td>0.252</td>
<td>0.239</td>
<td>0.013</td>
<td>87.7</td>
<td>0.938</td>
<td>0.899</td>
<td>0.892</td>
<td>0.009</td>
<td>99.1</td>
</tr>
<tr>
<td></td>
<td>500</td>
<td>0.647</td>
<td>0.653</td>
<td>0.008</td>
<td>98.4</td>
<td>0.933</td>
<td>0.147</td>
<td>0.150</td>
<td>0.003</td>
<td>84.6</td>
<td>0.929</td>
<td>0.841</td>
<td>0.843</td>
<td>0.010</td>
<td>98.1</td>
</tr>
<tr>
<td>ZINB</td>
<td>100</td>
<td>0.760</td>
<td>0.700</td>
<td>0.061</td>
<td>99.0</td>
<td>0.129</td>
<td>0.309</td>
<td>0.252</td>
<td>0.057</td>
<td>80.5</td>
<td>0.144</td>
<td>0.897</td>
<td>0.866</td>
<td>0.035</td>
<td>99.0</td>
</tr>
<tr>
<td></td>
<td>500</td>
<td>0.589</td>
<td>0.614</td>
<td>0.028</td>
<td>97.9</td>
<td>0.125</td>
<td>0.138</td>
<td>0.167</td>
<td>0.028</td>
<td>89.3</td>
<td>0.088</td>
<td>0.922</td>
<td>0.917</td>
<td>0.049</td>
<td>98.9</td>
</tr>
<tr>
<td>B</td>
<td>100</td>
<td>0.765</td>
<td>0.656</td>
<td>0.109</td>
<td>95.4</td>
<td>0.790</td>
<td>0.295</td>
<td>0.220</td>
<td>0.076</td>
<td>70.5</td>
<td>0.706</td>
<td>0.902</td>
<td>0.798</td>
<td>0.104</td>
<td>95.7</td>
</tr>
<tr>
<td></td>
<td>500</td>
<td>0.607</td>
<td>0.579</td>
<td>0.029</td>
<td>95.7</td>
<td>0.846</td>
<td>0.138</td>
<td>0.142</td>
<td>0.005</td>
<td>81.2</td>
<td>0.704</td>
<td>0.838</td>
<td>0.833</td>
<td>0.020</td>
<td>98.0</td>
</tr>
<tr>
<td>ZIP</td>
<td>100</td>
<td>0.889</td>
<td>0.826</td>
<td>0.065</td>
<td>99.6</td>
<td>0.210</td>
<td>0.317</td>
<td>0.295</td>
<td>0.023</td>
<td>87.0</td>
<td>0.092</td>
<td>1.084</td>
<td>1.030</td>
<td>0.056</td>
<td>99.6</td>
</tr>
<tr>
<td></td>
<td>500</td>
<td>0.767</td>
<td>0.736</td>
<td>0.034</td>
<td>99.3</td>
<td>0.192</td>
<td>0.195</td>
<td>0.204</td>
<td>0.010</td>
<td>93.3</td>
<td>0.201</td>
<td>0.806</td>
<td>0.811</td>
<td>0.010</td>
<td>98.4</td>
</tr>
<tr>
<td>ZINB</td>
<td>100</td>
<td>0.915</td>
<td>0.809</td>
<td>0.106</td>
<td>97.8</td>
<td>0.902</td>
<td>0.331</td>
<td>0.291</td>
<td>0.040</td>
<td>85.2</td>
<td>0.985</td>
<td>1.106</td>
<td>1.011</td>
<td>0.096</td>
<td>98.2</td>
</tr>
<tr>
<td></td>
<td>500</td>
<td>0.776</td>
<td>0.720</td>
<td>0.056</td>
<td>98.9</td>
<td>0.905</td>
<td>0.192</td>
<td>0.202</td>
<td>0.010</td>
<td>87.5</td>
<td>0.991</td>
<td>0.969</td>
<td>0.971</td>
<td>0.011</td>
<td>97.6</td>
</tr>
</tbody>
</table>
Table 3: Baseline characteristics by randomization assignment

<table>
<thead>
<tr>
<th>Child characteristics</th>
<th>MI + DVD</th>
<th>DVD only</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td>4.6 ± 1.6</td>
<td>4.5 ± 1.7</td>
</tr>
<tr>
<td>Gender</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>197(53.2%)</td>
<td>194(53.3%)</td>
</tr>
<tr>
<td>Soda consumption</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Never</td>
<td>117(36.2%)</td>
<td>112(36.6%)</td>
</tr>
<tr>
<td>1 day/week</td>
<td>28(8.7%)</td>
<td>35(11.4%)</td>
</tr>
<tr>
<td>2-6 days/week</td>
<td>127(39.3%)</td>
<td>124(40.5%)</td>
</tr>
<tr>
<td>Every day</td>
<td>51(15.8%)</td>
<td>35(11.4%)</td>
</tr>
<tr>
<td>Dental visit in the past 2 years</td>
<td>249(67.3%)</td>
<td>236(64.8%)</td>
</tr>
<tr>
<td>Untreated cavities</td>
<td>3.0 ± 5.9</td>
<td>2.9 ± 5.7</td>
</tr>
<tr>
<td>dmfs</td>
<td>9.2 ± 10.5</td>
<td>8.8 ± 10.2</td>
</tr>
<tr>
<td>dmft</td>
<td>5.3 ± 8.8</td>
<td>5.0 ± 8.3</td>
</tr>
<tr>
<td>Caregiver/family characteristics</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age</td>
<td>31.6 ± 8.8</td>
<td>31.0 ± 9.2</td>
</tr>
<tr>
<td>Gender</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>355(95.9%)</td>
<td>344(94.5%)</td>
</tr>
<tr>
<td>Education</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Less than high school</td>
<td>179(48.4%)</td>
<td>151(41.5%)</td>
</tr>
<tr>
<td>High school/GED</td>
<td>114(30.8%)</td>
<td>126(34.6%)</td>
</tr>
<tr>
<td>Some college or more</td>
<td>77(20.8%)</td>
<td>87(23.9%)</td>
</tr>
<tr>
<td>Household income</td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt; $10K</td>
<td>156(42.2%)</td>
<td>139(38.2%)</td>
</tr>
<tr>
<td>$10K ~</td>
<td>105(28.4%)</td>
<td>97(26.7%)</td>
</tr>
<tr>
<td>$20K ~</td>
<td>63(17.0%)</td>
<td>71(19.5%)</td>
</tr>
<tr>
<td>$30K ~</td>
<td>46(12.4%)</td>
<td>57(15.7%)</td>
</tr>
</tbody>
</table>

Table 4: Effects of MI on preventing early childhood caries

<table>
<thead>
<tr>
<th>Dental Outcome</th>
<th>Direct Effect</th>
<th>Mediation Effect</th>
<th>Total Effect</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Estimate (95% CI)</td>
<td>P</td>
<td>Estimate (95% CI)</td>
</tr>
<tr>
<td>New untreated lesion</td>
<td>0.007 (-0.666, 0.696)</td>
<td>0.99</td>
<td>-0.001 (-0.053, 0.050)</td>
</tr>
<tr>
<td>dmfs</td>
<td>0.381 (-1.076, 1.770)</td>
<td>0.63</td>
<td>-0.045 (-0.219, 0.058)</td>
</tr>
<tr>
<td>dmft</td>
<td>0.114 (-1.070, 1.234)</td>
<td>0.83</td>
<td>0.020 (-0.057, 0.134)</td>
</tr>
</tbody>
</table>