Published in final edited form as:
Biometrics. 2011 September ; 67(3): 1028–1038. doi:10.1111/j.1541-0420.2010.01547.x.

Generalized Causal Mediation Analysis

Jeffrey M. Albert[1,*] and **Suchitra Nelson**[2,**]

[1] Department of Epidemiology and Biostatistics, School of Medicine, WG-43, Case Western Reserve University, 10900 Euclid Avenue, Cleveland, Ohio 44106, U.S.A

[2] Department of Community Dentistry, Case School of Dental Medicine, 10900 Euclid Avenue, Cleveland, Ohio 44106, U.S.A

Summary

The goal of mediation analysis is to assess direct and indirect effects of a treatment or exposure on an outcome. More generally, we may be interested in the context of a causal model as characterized by a directed acyclic graph (DAG), where mediation via a specific path from exposure to outcome may involve an arbitrary number of links (or 'stages'). Methods for estimating mediation (or pathway) effects are available for a continuous outcome and a continuous mediator related via a linear model, while for a categorical outcome or categorical mediator, methods are usually limited to two-stage mediation. We present a method applicable to multiple stages of mediation and mixed variable types using generalized linear models. We define pathway effects using a potential outcomes framework and present a general formula that provides the effect of exposure through any specified pathway. Some pathway effects are nonidentifiable and their estimation requires an assumption regarding the correlation between counterfactuals. We provide a sensitivity analysis to assess of the impact of this assumption. Confidence intervals for pathway effect estimates are obtained via a bootstrap method. The method is applied to a cohort study of dental caries in very low birth weight adolescents. A simulation study demonstrates low bias of pathway effect estimators and close-to-nominal coverage rates of confidence intervals. We also find low sensitivity to the counterfactual correlation in most scenarios.

Keywords

Copula; Generalized linear model; G-computation algorithm; Path analysis; Potential outcome; Sensitivity analysis

1. Introduction

Mediation analysis has been of increasing interest in health and medical research as a tool for illuminating the mechanisms by which a treatment or exposure leads to a disease or health outcome. The starting point of a mediation analysis is a causal or path model. In applications, such models often include multiple stages (that is, where a path between the exposure and the final outcome, also referred to as a 'pathway', has more than two links) and involve variables of multiple types - for example, continuous, count, and dichotomous.

*jma13@case.edu
**sxn15@case.edu

Supplementary Materials
Web Table 1 and Web Figure 1, referenced in Section 5, are available under the Paper Information link at the Biometrics website http://www.biometrics.tibs.org.

The objective in this more general context is to determine the relative contribution of different paths to the overall exposure effect.

A motivating example is provided by a cohort study investigating very low birth weight (VLBW, possibly with bronchopulmonary dysplasia) to dental caries in adolescence (Nelson et al., 2010). The study involved 224 infants (139 VLBW, 85 normal term) who were part of an ongoing longitudinal study of family and psychosocial measures assessed from birth until 14 years of age (Singer et al., 1997). The normal term (control) group was selected in order to obtain similar distributions to the VLBW group for race, SES, and sex. The dental outcomes (including enamel defects, oral health behavior, and dental caries) were assessed at around age 14 years. The exposure variable in this example is the binary variable referred to as 'birth group', namely, VLBW (birth group = 1) versus normal term (birth group = 0). Originally, it was expected that VLBW would result in more adolescent dental caries, as measured by the number of decayed, filled or missing teeth (DMFT), and that this effect would be mediated by multiple pathways, including a biological mechanism (via the effect of birth group on enamel defects) and psychosocial mechanisms (for example, via the effect of birth group on oral health behavior or access to dental care). The primary study found, however, that the VLBW group had a lower mean DMFT than the term group. This result led to a revised hypothesis that while some (in particular, biological) pathways might favor the normal term children, other pathways might favor the VLBW group. In particular, it was supposed that the VLBW children might be receiving more extensive dental care, as indicated, for example, by the use of dental sealants. To provide an initial assessment of the contribution of such pathways, as well as a starting point for the development of our new method, we proposed the relatively simple path model presented in Figure 1. This model, described further in Section 4, includes a biological pathway (through enamel defects) and a dental access pathway (through use of sealants). As the occurrence of enamel defects would temporally precede decisions regarding use of sealants, we allowed for the possibility of a three-stage pathway (birth group → enamel defects → sealants → DMFT) in which enamel defects caused by VLBW induce an increased use of sealants which in turns reduces DMFT. The problem presented by this study thus involves the assessment of multiple mediators of different types, possibly occurring in multiple stages.

A classical method of mediation analysis for two stages (Baron and Kenny, 1986) involves the fitting of a succession of linear regression models. Structural equations model (SEM) based methods have also been proposed for mediation analysis (Ditlevsen et al., 2005). These methods have focused on the case of a linear model, typically leading to an estimator of the mediation (or indirect) effect in the form of a 'product of coefficients' (MacKinnon et al., 2002). In the linear case, it is straightforward to estimate mediation effects in the context of multiple stages via the product of coefficients approach (Taylor et al., 2008).

Some recent research has focused on mediation for binary or mixed types of variables. Huang et al. (2004) and Schluchter (2008) addressed two-stage mediation for a binary outcome based on a logistic regression model. Li et al. (2007) studied the two-stage case with a binary mediator. Eskima et al. (2001) addressed a general path model involving all binary variables, although their definition of a mediation effect was still confined to two stages.

Recently, researchers (Robins and Greenland, 1992; Rubin, 2004; Ten Have et al., 2007; Albert, 2008; Imai, Keele, and Yamamoto, 2010) have addressed mediation analysis using the potential outcomes framework. This framework allows clear definitions of mediation effects in causal terms and the explication of assumptions required for causal inference. These papers, however, were confined to two-stage mediation. Pearl (2001) presented definitions that extended direct and indirect effects ('path effects') to the case of a

nonparametric directed acyclic graph (DAG, see Pearl, 2000). In a further extension, Avin, Shpitser, and Pearl (2005) provided criteria for identifiability of path effects, as well as expressions for identifiable effects, in the nonparametric context. However, the latter papers did not consider possible regression relationships and did not provide methods for inference. In addition, there is a need for a practical approach for assessing path effects that are not nonparametrically identified. In sum, there has been little previous work on methods for mediation analysis in complex causal models involving mixed types of variables and multiple stages. The present paper seeks to extend the potential outcome approach to allow the assessment of mediation or path effects for such general causal models.

The remainder of the paper proceeds as follows. Section 2 provides the potential outcome framework and notation, Section 3 presents the general method, Section 4 describes the application to the dental data, Section 5 presents a simulation study, and Section 6 provides a summary and discussion.

2. Framework and Notation

2.1 A Causal (Manipulation) Approach to Path Models

This section and the next provide the framework and notation for deriving specific path effects for general causal models. In particular, we develop a potential outcomes-based approach to mediation analysis in the context of a DAG (or path) model. The approach thus extends the standard two-stage mediation analysis to allow for multiple stages of mediation. To simplify the situation somewhat, we focus on a causally ordered set of variables for which the exposure or treatment variable is the first and the outcome variable is the last in the causal order. All variables in between are referred to as intermediate variables or mediators. Note also that all of these variables, aside from the exposure, are endogenous variables (as they are caused by preceding variables) as well as response variables.

Besides the multiple stages, a generalized aspect of our model is to allow a link function, as in the generalized linear model, to relate the mean of each response variable to its directly causally preceding variables (‘parents’). A primary goal will then be to partition the overall exposure effect on the final outcome among the specific paths through which the overall effect is hypothesized to occur. Note that we use the term ‘pathway’ for a path going from exposure to the final response. For simplicity, we focus on the case of *two* levels of the exposure variable, which we will refer to as ‘exposed’ and ‘not exposed’.

Our main focus will be the path model shown at the top of Figure 1. This is a saturated model in that all possible pathways are allowed; that is, the variables may be ordered in a way so that all preceding variables are considered as parents. In this case of four variables (thus, involving as many as three stages) there are four possible pathways. More generally, in a saturated model with m mediators, there are $p = 2^m$ possible pathways. To see this, note that a specific pathway is obtained by selecting a subset of the m mediators (of which there are 2^m possibilities) through which the exposure effect will occur. The pathway involving none of the mediators represents the direct effect of exposure on the final response variable. The possible specific pathways in the saturated three-stage model are displayed at the bottom of Figure 1.

Although we focus on the saturated case, our method is applicable as well to unsaturated models. Thus the method can accommodate ‘contemporaneous’ mediators and/or multiple endpoints, in which case the set of model variables will be partially, as opposed to strictly, causally ordered. An example would be obtained from the model displayed in Figure 1 if the direct link from Z_1 to Z_2 were removed. In this case, Z_1 and Z_2 are not causally ordered (thus

contemporaneous), and Z_1 and Z_2 can be written in either order in the (partially) ordered vector of model variables.

In either the structural model (Pearl, 2000) or potential outcomes framework (for example, Albert, 2008), causal effects are conceived as resulting from possible (or imagined) experimental manipulations. Mediation or path effects may be defined as a function of a type of manipulation referred to as a 'path-deactivation process' (Pearl, 2001). This process was defined in terms of structural models by Pearl (2001), and is elucidated using a potential outcome framework below. Informally, a path-deactivation process can be described as a sequence of actions applied to the endogenous variables in a causal model (in causal order) in which each variable is set to the value it would have if it and each of its parents (and their parents, and so on) were subject to a specified combination of exposure levels. For example, starting with the simplest case, if the response variable Z_1 were set to the value it would have if the individual were exposed ($X = 1$), we say that the path (or 'link' in this case) from X to Z_1 is 'activated'; alternatively, if Z_1 were set to the value it would have if the individual were not exposed ($X = 0$) we say that the path from X to Z_1 is 'deactivated'. Next, suppose that response variable Z_2 (causally subsequent to Z_1) were set to the value it would have were the individual exposed (affecting Z_2 directly) but Z_1 set to the value it would have were the individual not exposed. In this case, the path $X \rightarrow Z_2$ is activated but the path $X \rightarrow Z_1 \rightarrow Z_2$ is deactivated. Note that the latter manipulation effectively isolates the direct effect of X on Z_2. The process becomes cumbersome to describe when more than two links are involved, but can be characterized precisely using the nested potential outcomes notation (Albert, 2008) as illustrated below. Note that the above definition of path activation/deactivation corresponds to *natural* (as opposed to *controlled*) direct and indirect effects, whereby manipulations dictate the *exposure* rather than the specific level of each mediator.

2.2 Notation and Potential Outcomes Framework

We next set up notation leading to a potential outcomes characterization of specific path effects. We let $X \equiv Z_0$ denote the exposure (with $X = 1$ if exposed, $X = 0$ if non-exposed); $Z_1, \ldots, Z_m$ are the causally ordered mediators, and $Y \equiv Z_{m+1}$ is the outcome variable. We further let $\mathcal{D}$ represent the set of manipulations (path activations) corresponding to the possible specific pathways. A particular realization D of $\mathcal{D}$ will be represented by listing involved mediators as subscripts; for example, D_1 represents the manipulation activating the pathway through Z_1 alone, and D_0 represents the direct pathway. For the three-stage model of Figure 1, we thus have $\mathcal{D} = \{D_0, D_1, D_2, D_{1,2}\}$. We will augment $\mathcal{D}$ by including $D = 0$ and $D = 1$ for the no exposure and exposure 'manipulations' (that is, the two naturally observed conditions), respectively. We also define corresponding manipulation indicator variables, for example $d_0 = 1$ if the manipulation activates the direct pathway only ($d_0 = 0$, otherwise), $d_1 = 1$ if the pathway through Z_1 alone is activated ($d_1 = 0$, otherwise), and so on. Thus, for each specific pathway exactly one of the d's is equal to 1 and the others are equal to 0. Other manipulations are possible that activate multiple pathways simultaneously. Such manipulations (corresponding to setting more than one of the d's equal to 1) are easily handled by our approach but are not the present focus.

We can consider the value of some response variable Y for individual i, if subject to manipulation D, as a potential outcome, written as $Y_i(D)$. (Hereon in we drop the subscript i, indexing the subjects, in the notation.) Note that the standard notation for a potential outcome (for example, $Y(0)$) can be considered as consistent with the above notation (where the argument is a manipulation) if we consider the number in the argument as representing a manipulation where all the d's are equal to that number. For example $Y(0)$ represents the potential outcome with all the d's equal to 0.

The potential outcome of response Y for a given manipulation can be written using nested potential outcomes as in Albert (2008). In the case of a single mediator (Z_1), the (nested) potential outcome of Y under manipulation D is written as $Y(D) \equiv Y(X=d_0, Z_1(X=d_1))$. A nested potential outcome can be read in a sequential manner to define the manipulation (path activation/deactivation). For example, $Y(X=1, Z_1(X=0))$ indicates the potential outcome for Y were X set to 1 (that is, all paths activated) and then Z_1 set to the value it would have were X set to 0. The latter action deactivates the path $X \rightarrow Z_1 \rightarrow Y$, thus leaving only the path $X \rightarrow Y$ (the direct effect) activated. Similarly, $Y(X=0, Z_1(X=1))$ represents the potential outcome of Y were X set to 0 (all paths deactivated) and then Z_1 set to the value it would have were X set to 1. This manipulation corresponds to the activation of the indirect path $X \rightarrow Z_1 \rightarrow Y$.

In the case of four variables (two mediators), as in the model shown in Figure 1, the potential outcome of Y given manipulation D, with corresponding indicator variables $d=(d_0, d_1, d_2, d_{1,2})$, can be written as $Y(D) \equiv Y(X=d_0, Z_1(X=d_1), Z_2(X=d_2, Z_1(X=d_{1,2})))$ or $Y(d_0, Z_1(d_1), Z_2(d_2, Z_1(d_{12})))$ for brief. Thus, the potential outcomes for the four possible pathways may be written as follows: $Y(D_0) = Y(1, Z_1(0), Z_2(0, Z_1(0)))$; $Y(D_1) = Y(0, Z_1(1), Z_2(0, Z_1(0)))$; $Y(D_2) = Y(0, Z_1(0), Z_2(1, Z_1(0)))$; and $Y(D_{12}) = Y(0, Z_1(0), Z_2(0, Z_1(1)))$. For example, the manipulation D_1 can be described (from the expression above for $Y(D_1)$) as the following three step process: (1) X is set to 0 (all paths from X to Y are deactivated); (2) Z_1 is set to the value it would take were X set to 1 (that is, the paths through Z_1, namely, $X \rightarrow Z_1 \rightarrow Y$ and $X \rightarrow Z_1 \rightarrow Z_2 \rightarrow Y$, are activated); and (3) Z_2 is set to the value it would take were X set to 0 and Z_1 set to the value it would take were X set to 0 (that is, the paths $X \rightarrow Z_2 \rightarrow Y$ and $X \rightarrow Z_1 \rightarrow Z_2 \rightarrow Y$ are deactivated). This manipulation leaves only the path $X \rightarrow Z_1 \rightarrow Y$ as activated. The above explication of a nested potential outcome reveals that the activation of a given pathway may involve a rather elaborate manipulation. Indicating the pathway activated thus provides a useful shorthand for the manipulation (Avin et al., 2005).

It will be convenient to write potential outcomes of intermediate variables as a function of the manipulation (D) even though not all the actions in the manipulation may be relevant to those variables. Such potential outcomes are easily found according to the nested values in the expression for Y. For example, for manipulation D_{12} with $Y(D_{12}) = Y(0, Z_1(0), Z_2(0, Z_1(1)))$, we have $Z_2(D_{12}) = Z_2(0, Z_1(1))$ and $Z_1(D_{12}) = Z_1(0)$. Note that the latter is with respect to response Y ; however, $Z_1(D_{12}) = Z_1(1)$ with respect to response Z_2. We will also use the expression $X(D)$ (where X is a control variable rather than response variable) to indicate the initial set value for X in a given manipulation; for example, $X(D_{12}) = 0$ with respect to response Y. Furthermore, for a vector $\boldsymbol{Z} = (Z_1, \ldots, Z_M)$ we define $\boldsymbol{Z}(D) \equiv (Z_1(D), \ldots, Z_M(D))$. We denote the expected value of $Y(D)$ as $E\{Y(D)\}$. Other estimands of interest will be the pathway effects relative to no exposure, that is, $R(D) \equiv E\{Y(D)\} - E\{Y(0)\}$, and the proportions of exposure effect due to (that is, mediated by the intermediate variables in) each pathway, defined as $P(D) \equiv R(D)/T$ where $T \equiv \sum_{D \in \mathcal{D}} R(D)$.

3. Inference for Pathway Effects

3.1 A General Formula for Pathway Effects

We next derive specific pathway effects in the context of a system of generalized linear structural models. This system of models corresponding to our DAG may be written as

$$h_j(\mu_j) = \beta_j \boldsymbol{Z}_{pj}(D),\ j = 1, \ldots, m+1 \equiv M, \tag{1}$$

where $\mu_j \equiv E\{Z_j(D)|\boldsymbol{Z}_{pj}(D)\}$; h_j is an invertible link function relating the expected value of Z_j to its parents; $\boldsymbol{Z}_{pj} = (1, Z_{pj1}, \ldots, Z_{pja_j})'$ is a vector containing the a_j parents of response Z_j; and $\boldsymbol{\beta}_j$ is the corresponding vector of unknown regression parameters. We also let $g_j \equiv h_j^{-1}$. The complete model also specifies the conditional probability distribution (or density) function for $Z_j(D)$ given $\boldsymbol{Z}_{pj}(D)$, denoted as $C_j\{Z_j(D), \boldsymbol{\theta}_j\} \equiv f_{Z_j(D)|\boldsymbol{Z}_{pj}(D)}(z_j|\boldsymbol{z}_{pj}; \boldsymbol{\theta}_j)$, for each j and D. The vector $\boldsymbol{\theta}_j$ contains the fixed unknown parameters involved in the conditional model for $Z_j(D)$; it will be dropped from the notation hereon in.

Note that the mediators (that is, the endogenous variables) in $Z_{pj}(D)$ are random variables, whereas the control variable (X) given D is fixed for each individual. Therefore, to obtain the marginal expected value of $Z_M(D) \equiv Y(D)$, namely, $E\{Y(D)\}$, will require that we integrate over the endogenous (random) parents of Y. Let $\boldsymbol{Z}_{RM} = (Z_{RM1}, \ldots, Z_{RMb})$ represent the subvector of $\boldsymbol{Z}_{pM}$ containing these b non-control variables ($b \leq a_M$) in causal order. This leads to the formula

$$E\{Y(D)\} = \int_{Z_{RM1}} \cdots \int_{Z_{RMb}} g_M(\beta_M \boldsymbol{z}_{pM}) f_{\boldsymbol{Z}_{RM}(D)}(\boldsymbol{z}_{RM}) d\boldsymbol{z}_{RM} \tag{2}$$

where $f_{\boldsymbol{Z}_{RM}(D)}(\boldsymbol{z}_{RM})$ is the joint density function for the random vector $Z_{RM}(D)$. Note that the joint density of $\boldsymbol{Z}_{RM}(D)$ may be factored as

$$f_{\boldsymbol{Z}_{RM}(D)}(\boldsymbol{z}_{RM}) = \prod_{k=1}^{b} f_{Z_{RMk}(D)|\overline{Z}_{RM,k-1}(D)}(Z_{RMk}|\overline{Z}_{RM,k-1}),$$

where $\overline{\boldsymbol{Z}}_{RM,k-1} \equiv (Z_{RM1}, \ldots, Z_{RM,k-1})$, and $\overline{\boldsymbol{Z}}_{RM,0} \equiv 0$. The integration in (2) will be interpreted as summation in the case of discrete mediators. Expression (2) may be seen as a variation of the G-computation algorithm (Robins, 1986). While the latter was proposed for repeated measures where the manipulation involves sequential treatments, the present formula is for a causal path model where the manipulation represents a path-deactivation process as described above. The present context has special implications and will require different methods of implementation. In causal inference terms, the above model and notation imply the stable unit-treatment value assumption (SUTVA, Rubin, 1990) and sequential ignorability (Ten Have et al., 2007).

One consequence of moving from the two-stage to the three-stage mediation problem is that not all pathway effects will be identifiable without additional assumptions. In particular, as shown by Avin et al. (2005), the effect of a path from X to Y is nonidentifiable if and only if and there is a path from some mediator Z to Y that is activated while another path from Z to Y is deactivated, and the path from X to Z is activated. Applying this criterion to the model in Figure 1, we find that (the effect of) pathway D_1 is nonidentifiable because the path $X \rightarrow Z_1 \rightarrow Y$ is activated but the path $X \rightarrow Z_1 \rightarrow Z_2 \rightarrow Y$ is deactivated. The reverse is true for pathway D_{12}, so it is also nonidentifiable. However, the effects of the other two pathways (D_0 and D_2) are identifiable. Below, we introduce an assumption to allow identification of the otherwise nonidentifiable pathways and also propose a sensitivity analysis.

We next illustrate the use of formula (2) and show the correspondence of our method with the standard approach in the case of two-stage mediation with continuous response variables. With a single mediator (Z) and the link function h for each equation as the identity function, the system of models in (1) would be written as: $E\{Y(D)|Z(D)\} = \beta_{Y\,I} + \beta_{Y\,0}X(D) + \beta_{Y\,1}Z(D)$ and $E\{Z(D)\} = \beta_{ZI} + \beta_{Z0}X(D)$. In this case, there are two possible pathways: 1) $D =$

D_0 (the direct pathway) giving $X(D_0) = 1$ and $Z(D_0) = Z(0)$ for response Y ; and 2) $D = D_1$ (the indirect pathway) giving $X(D_1) = 0$ and $Z(D_1) = Z(1)$. Then by formula (2) (in this case, simply integrating out $Z(D)$) we have, $E\{Y(D_0)\} = \int_z (\beta_{Y\,I} + \beta_{Y\,0} X(D_0) + \beta_{Y\,1} z) f_{Z(0)}(z) dz = \beta_{Y\,I} + \beta_{Y\,0} + \beta_{Y\,1}\beta_{ZI}$, since $E\{Z(D_0)\} = \beta_{ZI}$ with $X(D_0) = 0$ for response Z; similarly, $E\{Y(D_1)\} = \int_z (\beta_{Y\,I} + \beta_{Y\,0} X(D_1) + \beta_{Y\,1} z) f_{Z(1)}(z) dz = \beta_{Y\,I} + \beta_{Y\,1}(\beta_{ZI} + \beta_{Z0})$, since $E\{Z(D_1)\} = \beta_{ZI} + \beta_{Z0}$ with $X(D_1) = 1$ for response Z. Subtracting the expected value for no exposure ($E\{Y(0)\} = \beta_{Y\,I} + \beta_{Y\,1}\beta_{ZI}$), we obtain the direct effect as $E\{Y(D_0)\} - E\{Y(0)\} = \beta_{Y\,0}$ and the indirect effect as $E\{Y(D_1)\} - E\{Y(0)\} = \beta_{Y\,1}\beta_{Z0}$. These are the well-known expressions for these respective effects as discussed, for example, by MacKinnon et al. (2002).

3.2 Inference for Pathway Effects

Estimation of pathway expected values, $E\{Y(D)\}$, can be conducted using (2), plugging in estimates for regression parameters and any other parameters involved in the density functions. Regression parameter estimates can be obtained by fitting the generalized linear models corresponding to the structural models in (1). In our data application, we fit each model separately using maximum likelihood estimation.

In addition to the regression coefficients, the pathway effects also involve density (or probability) functions of mediators, thus, the parameters in the $\boldsymbol{\theta}_j$'s. For single-parameter distributions (such as binomial or Poisson) the probability function, $P(Z_j|\boldsymbol{Z}_{pj})$, follows immediately from the corresponding estimated mean for Z_j. For multiple-parameter distributions (such as negative binomial or normal with unknown variance) additional parameters may have to be estimated. Further details on estimation of the probability functions are discussed below. Confidence intervals for the expected values, as well as the relative pathway effects and pathway (mediation) proportions, can be obtained using a bootstrap approach (Efron and Tibshirani, 1993).

3.3 Evaluation of Probability Functions

For more than two stages of mediation, as discussed above, not all pathway effects are identifiable in general. Practically, this means that for a nonidentifiable pathway, the probability function in (2), and therefore the expected response, cannot be evaluated without further assumptions. We further consider this issue here, focusing, for simplicity and due to its relevance to our data example, on the context of a saturated three-stage model (as in Figure 1) with discrete mediators. In this case applying formula (2) yields

$$E\{Y(D)\} = \sum_{z_1} \sum_{z_2} g_3(\beta_{YI} + \beta_{Y0} d_0 + \beta_{Y1} z_1 + \beta_{Y2} z_2) p_2 p_1 \tag{3}$$

where $p_2 = P\{Z_2(d_2, Z_1(d_{12})) = z_2 | Z_1(d_1) = z_1\}$, $p_1 = P\{Z_1(d_1) = z_1\}$ and the summations are over the possible values of Z_1 and Z_2, respectively.

While estimation of p_1 is straightforward, the approach to evaluating p_2 depends on the case. When $d_1 = d_{12}(= 0)$, that is, the pathway being considered is D_0 or D_2, we have $P\{Z_2(d_2, Z_1(d_{12}))|Z_1(d_1)\} = P\{Z_2(d_2, Z_1(d_{12}))|Z_1(d_{12})\}$. The right hand side of this equality is readily evaluable and thus p_2 and $E\{Y(D)\}$ are estimable. In the case where $d_1 \neq d_{12}$ (which occurs when the pathway is D_1 or D_{12}), the above equality does not hold in general. To evaluate p_2 in this case we use the identity

$$P\{Z_2(d_2, Z_1(d_{12}))=z_2|Z_1(d_1)=z_1\}=\sum_{z_{12}} P\{Z_2(d_2, Z_1(d_{12}))=z_2|Z_1(d_{12})=z_{12}\}\times P\{Z_1(d_{12})=z_{12}|Z_1(d_1)=z_1\} \quad (4)$$

where the summation is over the possible values of Z_1. Note that when $d_1 \neq d_{12}$, the term $P\{Z_1(d_{12})|Z_1(d_1)\}$ in (4) involves the relationship between two counterfactuals, namely, $Z_1(1)$ and $Z_1(0)$, and thus is not estimable (resulting in the nonidentifiability of $E\{Y(D)\}$ for pathways D_1 and D_{12}) without further assumptions.

One way to allow estimation to proceed is to supply a value for the conditional probability $P\{Z_1(d_{12})|Z_1(d_1)\}$. If we assume that $Z_1(1)$ and $Z_1(0)$ are independent then (4) can be written as $p_2 = \Sigma_{z12} P\{Z_2(d_2, Z_1(d_{12})) = z_2|Z_1(d_{12}) = z_{12}\}P\{Z_1(d_{12}) = z_{12}\}$ which is estimable. Below, we discuss the possibility of alternative assumptions regarding $P\{Z_1(d_{12})|Z_1(d_1)\}$ which may be considered as part of a sensitivity analysis.

3.4 Sensitivity Analysis

We may wish to check the sensitivity of the pathway effect estimates to assumptions regarding the conditional probability $P\{Z_1(d_{12})|Z_1(d_1)\}$. As noted above, this is only an issue in the three-stage model for pathways for which $d_1 \neq d_{12}$. We propose a sensitivity analysis based on a general model for the joint probability of $Z_1(d_1)$ and $Z_1(d_{12})$. In this analysis we suppose that each of these variables is ordinal with (the same) K possible values, and let $z_{1,j}$ denote the jth ordered value. We let $P_1(z_1) \equiv P\{Z_1(d_1) \leq z_1\}$ denote the probability distribution function of $Z_1(d_1)$. Similarly, $P_{12}(z_{12}) \equiv P\{Z_1(d_{12}) \leq z_{12}\}$ will denote the probability distribution function of $Z_1(d_{12})$. These will generally be based on an assumed model, for example, Poisson. Of course, for the pathways under consideration, one variate ($Z_1(d_1)$ or $Z_1(d_{12})$) will equal $Z_1(0)$ and the other will equal $Z_1(1)$. We suppose that there are, corresponding to $Z_1(d_1)$ and $Z_1(d_{12})$, latent variables, denoted as $Z_1^*(d_1)$ and $Z_1^*(d_{12})$, respectively, which are marginally distributed as standard normal and satisfy the relationships, $Z_1^*(d_1)=\Phi^{-1}[P_1\{Z_1(d_1)\}]$ and $Z_1^*(d_{12})=\Phi^{-1}[P_{12}\{Z_1(d_{12})\}]$ where Φ is the standard normal distribution function. In addition, $Z_1^*(d_1)$ and $Z_1^*(d_{12})$ are assumed to have a bivariate normal distribution with correlation ρ. Due to the assumed bivariate normal distribution we have the relationship $E\{Z_1^*(d_1)|Z_1^*(d_{12})\}=\rho Z_1^*(d_{12})$. This model may be seen as a special case of the Gaussian copula (Song, Li, and Yuan 2009).

In order to properly handle the discrete nature of the distributions of the original variables, we propose a Monte Carlo approach to compute the conditional distributions $P\{Z_1(d_{12})|Z_1(d_1) = z_1\}$ for possible values of z_1. The method uses the following algorithm which has as inputs, the estimated or empirical marginal probability distribution functions, denoted $\hat{P}_1$ and $\hat{P}_{12}$ for $Z_1(d_1)$ and $Z_1(d_{12})$, respectively, and a specified value for correlation parameter ρ; we also let $\hat{P}_{1,j} \equiv \hat{P}_1(z_{1,j})$, $\hat{P}_{12,j} \equiv \hat{P}_{12}(z_{1,j})$, and $\hat{P}_{1,0} = \hat{P}_{12,0} \equiv 0$:

1. For a given z_1, say $z_{1,k}$, draw a uniform variate (u_1, say) from the interval $(\hat{P}_{1,k-1}, \hat{P}_{1,k}]$. Let $U_1 = \Phi^{-1}(u_1)$.
2. Draw a variate (U_{12}, say) from $N(0, 1)$ independent of U_1.
3. Calculate $U_{12C} = \rho U_1 + (1 - {}^2)^{1/2}U_{12}$. Let $u_{12C} = \Phi(U_{12C})$.
4. Let $C_{12jk} = 1$ if $u_{12C} \in (\hat{P}_{12,j-1}, \hat{P}_{12,j}]$, $C_{12jk} = 0$, otherwise, for $j = 1, \ldots, K$ (with the subscript k indicating the conditioning on $z_{1,k}$).
5. Repeat Steps 1–4 with independent draws a large number of (say, R) times obtaining $C_{12jkr} = C_{12jk}$ in the rth replication for $r = 1, \ldots, R$.

Following the R replications for each $z_{1,k}$, we estimate the conditional probability $P\{Z_1(d_{12}) = z_{1,j}|Z_1(d_1) = z_{1,k}\}$ as $\Sigma_r C_{12jkr}/R$ for $j, k = 1, \ldots, K$. With this substitution we can obtain an estimate of $E\{Y(D)\}$ for each specific pathway D using expressions (3) and (4); consequently we can also estimate $R(D)$ and $P(D)$. These estimates can be recomputed using the above algorithm over varying values for ρ to provide a sensitivity analysis.

3.5 Adjustment for Covariates

So far, we have addressed the situation where the vector of model variables ($\boldsymbol{Z}$), consists only of the (endogenous) mediators and the exogenous exposure variable. However, there may be covariates (that is, other pre-exposure or otherwise exogenous variables) that affect two or more endogenous variables in the model. Such variables represent potential confounders that may be important to control for in the assessment of mediation/pathway effects. We now show how the above mediation analysis may be extended to allow covariate adjustment.

The general model (1) expanded to include covariates can be written as

$$h_j(\mu_j)=\beta_j \boldsymbol{Z}_{pj}(D)+\gamma_j \boldsymbol{W},\ j=1,\ldots,m+1 \equiv M,$$

where $\boldsymbol{W}$ is the vector of (exogenous) covariates and $\boldsymbol{\gamma}_j$ is the corresponding vector of regression parameters for the jth outcome variable. The covariates, being exogenous, are unaffected by the manipulation, D. For simplicity, the above model implies the use of the same set of covariates for all equations, but this is not necessary.

Our approach is to estimate the marginal expected value of response Y under manipulation D (that is, $E\{Y(D)\}$) as the weighted average (over the joint distribution of the covariates, $\boldsymbol{W}$) of the conditional expected responses, $E\{Y(D)|\boldsymbol{W}\}$. We thus use the basic identity, $E\{Y(D)\} = \int_{\boldsymbol{w}} E\{Y(D)|\boldsymbol{W}=\boldsymbol{w}\} f_{\boldsymbol{W}}(\boldsymbol{w}) d\boldsymbol{w}$ where $f_{\boldsymbol{W}}$ is the joint density function of $\boldsymbol{W}$, and the integral is interpreted as summation in the case of categorical covariates. The conditional expectation, $E\{Y(D)|\boldsymbol{W}=\boldsymbol{w}\}$ is computed as in formula (2) but with $g_M(\boldsymbol{\beta}_M \boldsymbol{z}_{pM} + \boldsymbol{\gamma}_M \boldsymbol{W})$ in place of $g_M(\boldsymbol{\beta}_M \boldsymbol{z}_{pM})$, and the conditional joint density $f_{\boldsymbol{Z}_{RM}(D)|\boldsymbol{W}}(\boldsymbol{z}_{RM}|\boldsymbol{W})$ in place of $f_{\boldsymbol{Z}_{RM}(D)}(\boldsymbol{z}_{RM})$. With these substitutions, the conditional expected value is estimated in the same manner as the marginal expectation as described above. To estimate the joint distribution of the covariates, we use the empirical distribution (which will easily accommodate continuous as well as categorical covariates). For our application, we use the empirical distribution for the whole sample, although other applications could differ, for example, by using the distribution for the exposed group. Essentially, we are then obtaining the estimate of $E\{Y(D)\}$ as the average of the estimate $E\{Y(D)|\boldsymbol{W} = \boldsymbol{w}_i\}$ for individual covariate values, $\boldsymbol{w}_i$, over the sample. Covariate-adjusted estimates of $R(D)$ and $P(D)$ are obtained (based, on the definitions at the end of Section 2) using the covariate-adjusted estimates of $E\{Y(D)\}$ and $E\{Y(0)\}$. The bootstrap method for estimating standard errors and confidence intervals and the sensitivity analysis are applied, in a parallel manner to that described above, to the covariate-adjusted estimates.

4. Application to Dental Data

The dental study that we consider was briefly described in the introduction. The proposed causal model (represented in Figure 1) is a saturated model involving, in order, the exposure variable, ‘birth group’ (X), coded as 1 for VLBW, 0 for normal term; ‘enamel defects’, specifically, the number of teeth with any demarcation (Z_E); ‘sealants’, a binary indicator of the use of sealants (Z_S), coded as 1 if the individual received any sealant, 0 otherwise; and

DMFT (*Y*), the number of decayed, missing, and filled permanent teeth. The enamel defect and DMFT counts used in the present analysis were confined to incisors and first molars as these were anticipated to be more vulnerable than other teeth to problems due to prematurity. We begin by specifying the model and deriving mediation effects. Following the methodological setup we present the results for the data.

4.1 Model and Formulae for Pathway Effects

To begin, we assumed that the count variables (DMFT and enamel defects) are distributed as Poisson and that sealant use is Bernoulli. We used the following system of generalized structural linear models with canonical links (log for Poisson and logit for Bernoulli); here we use subscripts "*E*" (for enamel defects), "*S*" (for sealant), and "*Y*" (for the final response, DMFT) in place of subscripts 1, 2, and 3 (or *M*), respectively:

$$\begin{aligned} ln[\,E\{Y(D)|Z_{pY}(D)\}] &= \beta_{Y1}+\beta_{Y0}X(D)+\beta_{Y1}Z_E(D)+\beta_{Y2}Z_S(D) \\ logit[\,E\{Z_S(D)|Z_{pS}(D)\}] &= \beta_{S1}+\beta_{S0}X(D)+\beta_{S1}Z_E(D) \\ ln[\,E\{Z_E(D)\}] &= \beta_{E1}+\beta_{E0}X(D). \end{aligned} \tag{5}$$

In this two-mediator model there are four possible pathways from the exposure (birth group) to the outcome (dental caries). These pathways (corresponding to alternative explanations of the effect of birth group on dental caries) are represented graphically in Figure 1 and described as follows: D_0 - effect is due to the direct effect of birth group on dental caries; D_E - effect is through enamel defects only; D_S - effect is through sealants only; D_{ES} - effect is through the chain of enamel defects followed by sealants.

Our goal was to determine the contribution of each of the four possible pathways in explaining the association between birth group and DMFT. For computational ease we used an upper bound of 9 for enamel defects counts. In fact, 12 is the biological upper bound (corresponding to the 8 permanent incisors and 4 permanent first molar teeth), but the probability of a count higher than 9 (which in fact did not occur in our data) was considered as negligible.

For the present model, we apply formula (3) with the appropriate change in notation. We use $P\{Z_E(d_E)=z_E\}=\{(\mu_E^{z_E}e^{-\mu_E})/z_E!\}/P_{EM}$ where $\mu_E \equiv E\{Z_E(D_E)\} = exp(\beta_{EI} + \beta_{E0}d_E)$ and $P_{EM}=\sum_{j=0}^{9}(\mu_E^j e^{-\mu_E})/j!$. Similarly, $P\{Z_E(d_{ES})=z_{ES}\}=\{(\mu_{ES}^{z_{ES}}e^{-\mu_{ES}})/z_{ES}!\}/P_{ESM}$ where $\mu_{ES} = exp(\beta_{EI} + \beta_{E0}d_{ES})$ and $P_{ESM}=\sum_{j=0}^{9}(\mu_{ES}^j e^{-\mu_{ES}})/j!$. For pathways D_0 and D_S we have $p_S = z_S\mu_S + (1 - z_S)(1 - \mu_S)$ where $\mu_S = [1 + exp\{-(\beta_{SI} + \beta_{S0}d_S + \beta_{S1}z_E)\}]^{-1}$. For the pathways D_E and D_{ES} we made the working assumption of independence of $Z_E(0)$ and $Z_E(1)$, equivalently $\rho = 0$ in the latent variable model described in Section 3.4. This assumption yields $p_S = \Sigma_{z_{ES}}\{z_S\mu_{SC} + (1 - z_S)(1 - \mu_{SC})\}P\{Z_E(d_{ES})=z_{ES}\}$ where $\mu_{SC} = [1 + exp\{-(\beta_{SI} + \beta_{S0}d_S + \beta_{S1}z_{ES})\}]^{-1}$.

In addition, we fit the above model controlling for covariates, namely, sex, race (African American or white), and socioeconomic status (low or high). All three of these variables were added to each model in (5) and the method of Section 3.5 used to estimate pathway effects. Finally, we considered the negative binomial distribution as an alternative to the Poisson for the two count variables. The model is the same as above (that is, (5) with added covariates) but here the mediator probabilities in (3) are determined using the negative binomial rather than Poisson distribution (truncating above a count of 9 as before). Model parameters were estimated via maximum likelihood estimation and the Akaike information

criterion (*AIC*) was calculated for each model to allow comparisons of model fit to the dental data.

For each of the above models we obtained the estimated pathway effects, that is, $E\{Y(D)\}$, $R(D)$, and $P(D)$, for each specified D. For each estimand, upper and lower bounds for a 95% confidence interval were obtained, respectively, as the 2.5th and 97.5th percentiles of the distribution of the corresponding estimate over 1000 bootstrap samples. In addition, we used the method of Section 3.4 to assess the sensitivity of these estimates to the choice of ρ.

4.2 Results of Dental Data Analysis

The dental data provided a sample size of 203 using complete cases on the model variables (and one other variable considered but not included in the present analysis, namely, the oral hygiene score, which provides an alternative candidate as a mediator). For later reference we note that the observed average DMFT (and standard deviation) was 1.54 (1.75) for the normal birth weight group (n=78) and 1.04 (1.58) for the VLBW group (n=125). Table 1 provides maximum likelihood estimates of the regression parameters (along with standard errors) for the three models examined: 1) Model (5) assuming Poisson-distributed counts, 2) the same model but including three covariates (sex, race, and SES) in each response model, and 3) Model (5) with covariates and assuming negative binomial distributed counts. Table 1 also has *AIC* values for each response model. The lower *AIC* for each response variable in Model 2 relative to the same response variable in Model 1 indicates that adding the three covariates improves model fit. In addition, comparing Models 2 and 3, we find lower *AIC*s using a negative binomial distribution for the count variables (Z_E and Y) relative to using a Poisson distribution. The model with negative binomial counts generally provides larger estimated standard errors, presumably reflecting the fact that this distribution is able to account for extra-Poisson variation. An examination of the regression parameter estimates (focusing on Model 3 results) reveals marginally statistically significant direct effects of birth group (X) and sealant use (Z_S) on DMFT (Y). However, birth group (X) does not appear to significantly affect enamel defects (Z_E) or use of sealants (Z_S) directly. Also, there do not appear to be significant effects of enamel defects on sealant use or on DMFT.

Estimates of expected DMFT ($E\{Y(D)\}$), the pathway effect relative to no exposure ($R(D)$), and the mediation proportion ($P(D)$), for each pathway, are provided in Table 2. From this table, we see that estimated pathway effects are similar in the three models and the conclusions are substantially the same. The estimated expected response for the direct pathway between birth group and DMFT is around 1.0 for each model. This value is interpreted as the expected DMFT for the VLBW group were the birth group effect through enamel defects and through sealant use to be blocked. We also see that the expected DMFT for the other three pathways is around 1.5. The estimated mediation proportions (from Models 2 and 3) are around 1.04 for the direct effect, −0.07 for the pathway through enamel defects alone, 0.03 for the pathway though sealants alone, and close to 0 for the pathway though enamel defects and sealants. It thus appears that the effect of birth group on DMFT is primarily due to the direct effect. Note, of course, that this effect will include effects of pathways through any unobserved mediators. The fact that the estimated proportion for the direct effect is greater than 1 indicates that the assessed mediating factors (enamel defects and sealant use) together actually slightly favor the normal birth group; that is, were the birth group effects through these mediators to be blocked, the expected birth group effect (favoring the VLBW group) would be even greater (by an estimated 4 percent) than that observed. The positive mediation proportion for the pathway through sealants alone (in the models that adjust for covariates) indicates that this pathway explains some (3 percent) of the observed association between birth group and DMFT, while the negative proportion for the pathway through enamel defects alone indicates that this pathway actually favors the normal birth group. However, none of these mediation proportions (aside from that of the

direct effect) are significantly different from 0. Note that negative values for $P(D)$, as well as values greater than 1, are interpretable due to the possibility of 'negative' mediation, as discussed in Albert (2008).

The expected effects for the D_E and D_{ES} pathways were estimated under the assumption of independence between $Z_E(0)$ and $Z_E(1)$. We checked the impact of this assumption by applying the sensitivity analysis described in Section 3.4. For pathways D_E and D_{ES}, we re-estimated $E\{Y(D)\}$ over varying values of the correlation parameter ρ, namely, from -1 to 1 in increments of 0.1. The number of replicates used in the Monte Carlo simulations to estimate each conditional probability was R=1000. We found that the re-estimates, both for $E\{Y(D_E)\}$ and $E\{Y(D_{ES})\}$, estimated under $\rho = 0$ as 1.54 and 1.50 respectively (for Model 3), were within 0.001 over the range of ρ. The results thus appear to be quite insensitive to assumptions regarding ρ.

5. Simulation Study

We conducted a simulation study to examine our proposed estimators in terms of bias and efficiency. Seven different scenarios, each representing a set of selected values of the regression parameters (β's) in Model (5) possibly including covariates, were considered. The scenarios may be characterized as follows: (1) 'Dental Data' - parameter values approximately equal to the estimates obtained from the analysis of the dental data in Section 4; (2) 'All D_1' -entire exposure effect due to pathway $X \rightarrow Z_1 \rightarrow Y$;(3) 'All D_2' - entire exposure effect though pathway $X \rightarrow Z_2 \rightarrow Y$; (4) 'All D_{12}' - entire exposure effect though pathway $X \rightarrow Z_1 \rightarrow Z_2 \rightarrow Y$; (5) 'All D_0' - entire exposure effect though direct pathway ($X \rightarrow Y$);(6) 'Equal' - approximately equal mediation proportions though the four possible pathways; (7) 'Equal + Covariate' - approximately equal mediation proportions as in Scenario (6) but where a covariate (W) is included that affects two of the responses (namely, Z_1 and Y). For Scenario 7 we fit two models, one with and one without the covariate; the latter was intended to allow us to study the impact on estimates of an omitted confounder. The regression coefficient values used for each scenario are given in Table 3.

Simulated data sets for each scenario were obtained with a sample size of 200, similar to the dental dataset. For each individual (observational unit), X, Z_1, Z_2, and Y were generated sequentially, with the mean for each variable calculated as a function of its predictors (parents) as given by model (5) (with the subscripts "E" and "S" exchanged with "1" and "2", respectively). The exposure group, X, coded as 0 (non-exposed) or 1 (exposed) was assigned via constrained randomization to provide 100 individuals per group; Z_2 was generated as a Bernoulli random variable; and Z_1 and Y were generated as truncated Poisson with an upper bound of 9. Poisson was chosen rather than negative binomial - although the latter provided a better fit to the dental data - to allow greater computational speed and because the estimated pathway effects in the dental data were not greatly affected by the choice of model. Realizations of this set of four variates were generated independently across individuals. For each dataset, the method of Section 3, under the specific implementation of Section 4.1, was used to estimate $E\{Y(D)\}$, $R(D)$, and $P(D)$ for each pathway (D). One thousand bootstrap samples were drawn for each dataset to obtain confidence intervals. For each scenario 500 replications were performed.

The main statistics of interest for each set of replications were the mean of the estimates, the coverage (percent of replications for which the 95% confidence interval covered the true value of the estimand), and, for $R(D)$ and $P(D)$, the power (percent of replications for which the 95% confidence interval excluded zero).

The true values for the estimands were obtained by applying the formula for the expected values (3) using the true values for the regression parameters (β's) in model (5). As in the analysis of actual data, this approach requires a specification of $P\{Z_1(d_1)|Z_1(d_{12})\}$ for otherwise nonidentifiable pathways. For the simulated data, we assumed independence, equivalent to using $\rho = 0$ in the model describe in Section 3.4.

The simulation results are shown in Table 4. We see that for all scenarios average estimates for $E\{Y(D)\}$ and $R(D)$ were close to the true values for each of the four pathways (D's). This was also the case for $P(D)$ for Scenario 1, but not always for the other scenarios. In particular, in Scenario 4, and to a lesser extent in the other scenarios, the average estimate of $P(D)$ was sometimes rather far off from the true value. This may be explained by the fact that that the total exposure effect was relatively small in these scenarios, so that that the estimate of $P(D)$ could be rather unstable. If we include only datasets where the estimated total exposure effect is at least 0.005 in absolute value, then the estimates of $P(D)$ for Scenario 4 become more reasonable (0.022, 0.057, −0.069, and 0.99, for pathways D_0, D_1, D_2, and D_{12}, respectively). The coverage of 95% confidence intervals for $R(D)$ was close to the nominal rate (plus or minus two percent) in most cases, though sometime conservative or anti-conservative. The 95% confidence intervals for $P(D)$ tended to be conservative, presumably due to the frequent instability of this measure as noted above. The results for power reflect the effect sizes and indicate that rather large effects (corresponding to mediation proportions near 1) are needed for high power for our studied sample size. The results for Scenario 7 (Web Table 1) show that the method for covariate adjustment could provide reasonable estimates of pathway effects in the presence of a confounder. We find that estimated pathway effects are somewhat but not very far off when the covariate is omitted.

In addition, some of the scenarios had a lack of convergence or questionable convergence due to a non-positive definite negative Hessian for some of the simulated datasets. This occurred in a few of the bootstrap-generated datasets for Scenarios 4, 6 and 7, and also in a small number of the originally-generated datasets in Scenario 4. These datasets were excluded from the results. Such computational difficulties appeared to be more likely in cases, such as Scenario 4, where there is a substantial effect of Z_1 on Z_2, presumably because this situation will tend to induce multicollinearity when both Z_1 and Z_2 are included in the model for Y. Nevertheless even in these cases, our simulations showed low bias in estimated pathway effects.

Additionally, we examined the effect on true estimand values of varying the counterfactual correlation ρ from −1 to 1 in increments of 0.1. As expected, there was no effect of ρ on the estimands for pathways D_0 or D_2 (for which the derived formulae do not involve ρ.) For D_1 and D_{12} there was no effect of ρ on the estimands in Scenario 2, 3, and 5. This result could be expected because there is no effect of Z_1 on Z_2 in these scenarios. In Scenario 1 ('Dental Data'), which has a small effect of Z_1 on Z_2, the expected values of Y (that is, $E\{Y(D)\}$) varied by less than 0.005 over the range of ρ. The other scenarios showed a more appreciable impact of varying ρ, with a range of around 0.3 for Scenario 4, and up to around 0.5 for Scenarios 6 and 7. The effect of ρ on $E\{Y(D)\}$ for pathways D_1 and D_{12} is displayed graphically in Web Figure 1 for Scenario 6 ('equal'). The more pronounced impact of ρ on estimand values in these scenarios can be explained by the relatively large effect of Z_1 on Z_2. Note, however, that in practice it may often be reasonable to assume a positive correlation. This would reducing the ranges mentioned above for $E\{Y(D)\}$ by around half.

6. Discussion

This paper presents a general method for assessing pathway-specific (or mediation) effects in the relationship of a treatment or exposure and an outcome. The basic formula presented (2) is applicable to any DAG where the expected value of the outcome and each mediator can be expressed as a function of its parents via a generalized linear model. In particular, this general approach allows for multiple stages of mediation, different types of model variables, and adjustment for pre-exposure covariates. For concreteness, we focused on the special case of a saturated three-stage path model with bounded discrete mediators. Our approach should be readily extendable to unsaturated models (that is, models that assume the absence of some direct effects) as well as models with more than three stages of mediation. Because our proposed approach is based on likelihood, it has the advantages of the availability of likelihood-based methods for model testing and selection, and of allowing valid inference when data are missing at random. Estimation of model parameters may be performed using standard statistical packages. We used *SAS* (Version 9.2), including PROC GENMOD and *SAS/IML*, to conduct the data analysis and simulation studies presented in this paper.

Our approach was successfully applied to data from a study of dental caries in VLBW versus normal term adolescents. The analysis showed that the effect of VLBW on DMFT was largely explained by its direct effect, with little contribution of hypothesized pathways through enamel defects or through use of sealants. Although this data example was intended to be illustrative, we provide some precautions for interpreting the results. First, we note that the dental responses (including both the final outcome and the mediators in our model) were obtained at around the same time, thus placing heavy reliance on assumptions of their causal order. In addition, using birth weight as an exposure variable implies that it is manipulable; while arguable, this is a conceptualization that needs further explication. Finally, the model presented for the dental data represents a considerable simplification of a more comprehensive model for dental caries which would include additional mediators and potential confounders.

The implementation of our method may be more difficult if unbounded or continuous mediators are used; in the latter case, formula (2) involves integration rather than summation. In cases where a continuous mediator is of interest, numerical or Monte Carlo methods can be incorporated in the pathway effect estimators.

Our simulation studies showed little if any bias of estimated pathway effects under the models and good coverage properties of bootstrap confidence intervals. More caution may be needed for inference regarding mediation proportions which we found to be potentially unstable for small exposure effects. This observation echoes previous work on the proportion of treatment effect explained by a surrogate endpoint (Freedman, Graubaud, and Shatzkin, 1992). An additional concern is the possibility of lack of convergence when fitting the generalized linear models. The usual precaution of avoiding multicollinearity among baseline covariates should be taken.

We indicated that some pathways may be nonidentifiable, namely, where their expected value involves the joint probability among counterfactuals. This problem, which only occurs when there are more than two stages of mediation, appears to have received little attention aside from Avin et al. (2005). The assumption of independence of the counterfactuals allows a straightforward solution, but may not be scientifically plausible. Consequently, we devised a sensitivity analysis to allow an assessment of the impact of this assumption on the pathway effect estimates. Our proposed sensitivity parameter is based on a copula model and is interpreted as the correlation between latent normally distributed variables underlying the

counterfactuals. This approach has the advantage of being applicable to different types of variables. A limitation is that, being a correlation of counterfactuals, some may find this parameter difficult to interpret. Fortunately, our simulation studies, as well as our data analysis, showed that pathway effect estimates have low sensitivity to the counterfactual correlation under most scenarios.

A basic limitation of our method is that it assumes a correct specification of the causal model, implying the assumption of no unmeasured confounders, or equivalently, sequential ignorability. Specifying adequately elaborate models may allow a better approximation to this assumption. A sensitivity analysis based on this assumption (Imai et al., 2010) might also be considered.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

The authors would like to thank the Associate Editor and three referees for their insightful comments which helped greatly to improve the paper. We are also grateful to Wei Wang and Yuan-Chiao Lu for assistance in conducting the simulation studies and in preparation of the paper, and Dr. Lynn Singer for providing access to the longitudinal cohort of VLBW and NBW adolescents. Support for this research was provided in part by Research Grants R03-DE018391 (Dr. Albert) and R21-DE16469 (Dr. Nelson) from NIDCR/NIH and MC-390592, MC-00127, and MC-00334 (Dr. Singer) from HRSA/DHHS.

References

Albert JM. Mediation analysis via potential outcomes models. Statistics in Medicine. 2008; 27:1282–1304. [PubMed: 17691077]

Avin C, Shpitser I, Pearl J. Identifiability of path-specific effects. Proceedings of the 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. [PubMed: 3806354]

Ditlevsen S, Christensen U, Lynch J, Damsgaard MT, Keiding N. The mediation proportion: a structural equation approach for estimating the proportion of exposure effect on outcome explained by an intermediate variable. Epidemiology. 2005; 16:114–120. [PubMed: 15613954]

Efron, B.; Tibshirani, R. An Introduction to the Bootstrap. Chapman & Hall Ltd; London: 1993.

Eskima N, Tabata M, Zhi G. Path analysis with logistic regression models: effect analysis of fully recursive causal systems of categorical variables. Journal of the Japan Statistical Society. 2001; 31:1–14.

Freedman LS, Graubard BI, Schatzkin A. Statistical validation of intermediate endpoints for chronic diseases. Statistics in Medicine. 1992; 11:167–178. [PubMed: 1579756]

Huang B, Sivaganesan S, Succop P, Goodman E. Statistical assessment of mediational effects for logistic mediational models. Statistics in Medicine. 2004; 23:2713–2728. [PubMed: 15316954]

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

Li Y, Schneider JA, Bennett DA. Estimation of the mediation effect with a binary mediator. Statistics in Medicine. 2007; 26:3398–3414. [PubMed: 17066450]

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. [PubMed: 11928892]

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. [PubMed: 20975268]

Pearl, J. Models, Reasoning, and Inference. Cambridge University Press; Cambridge, UK: 2000.

Pearl, J. Direct and indirect effects. In Proceedings of the 17th Conference on Uncertainty in Artificial Intelligence; San Francisco: Morgan Kaufmann; 2001. p. 411-420.

Robins J. A new approach to causal inference in mortality studies with sustained exposure periods - applications to control of the healthy worker survivor effect. Mathematical Modeling. 1986; 7:1393–1512.

Robins JM, Greenland S. Identifiability and exchangeability for direct and indirect effects. Epidemiology. 1992; 3:143–155. [PubMed: 1576220]

Rubin DB. Neyman (1923) and causal inference in experiments and observational studies. Statistical Science. 1990; 5:472–480.

Rubin DB. Direct and indirect causal effects via potential outcomes. Scandinavian Journal of Statistics. 2004; 31:161–170.

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

Singer LT, Yamashita TS, Lilien L, Collin M, Baley JA. Longitudinal study of infants with bronchopulmonary dysplasia and very low birth weight. Pediatrics. 1997; 100:987–993. [PubMed: 9374570]

Song PXK, Li M, Yuan Y. Joint regression analysis of correlated data using Gaussian copulas. Biometrics. 2009; 65:60–68. [PubMed: 18510653]

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

Ten Have TR, Joffe MM, Lynch KG, Brown GK, Maisto SA, Beck AT. Causal mediation analyses with rank preserving models. Biometrics. 2007; 63:926–934. [PubMed: 17825022]

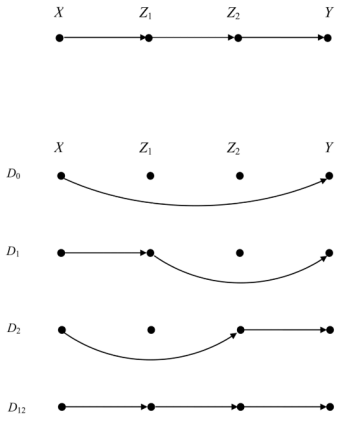

Figure 1.
Three-stage path model. On top is the overall path model; below are the possible specific pathways. For the dental data, X = birth group, Z_1 = enamel defects, Z_2 = sealant use, and Y = DMFT.

Table 1

Maximum likelihood estimates (and standard errors) of regression parameters for alternative models applied to dental data; also AIC values (smaller is better) for each response submodel. (X = birth group, Z_E = enamel defects, Z_S = sealant use, and Y = DMFT)

			Predictor				
Model	**Response**	**Dist**	**Intercept**	***X***	Z_E	Z_S	***AIC***
	Z_E	Poisson	−0.32 (0.13)	0.32 (0.15)	-	-	683.0
1 (No cov adj)	Z_S	Binomial	−0.92 (0.25)	−0.04 (0.31)	−0.018 (0.10)	-	270.2
	Y	Poisson	0.44 (0.10)	−0.39 (0.12)	0.092 (0.03)	−0.50 (0.16)	765.2
	Z_E	Poisson	−0.34 (0.19)	0.29 (0.15)	-	-	631.4
2 (Cov adj)	Z_S	Binomial	−0.63 (0.40)	0.10 (0.33)	0.004 (0.10)	-	245.4
	Y	Poisson	0.18 (0.17)	−0.41 (0.13)	0.076 (0.036)	−0.46 (0.16)	673.9
	Z_E	Neg Bin	−0.30 (0.30)	0.27 (0.24)	-	-	547.9
3 (Cov adj)	Z_S	Binomial	−0.63 (0.40)	0.10 (0.33)	0.004(0.10)	-	245.4
	Y	Neg Bin	0.15(0.27)	−0.39(0.21)	0.080(0.06)	−0.45(0.24)	614.5

Table 2

Estimates (and 95% bootstrap confidence intervals) for expected values, E{Y (D)}, relative effects, R(D) = E{Y (D)} − E{Y (0) }, and proportion of total exposure effect explained, P(D), for each pathway and for each model fit to the dental data. Absolute values less than 0.001 are indicated by “0.000” (for positive values) or “−0.000” (for negative values).

Model	Pathway(D)	$E\{Y(D)\}$	$R(D)$	$P(D)$
1 (Poisson Counts, no covs)	D_0	1.01 (0.76, 1.30)	−0.48 (−0.92, −0.004)	1.11 (0.08, 2.2)
	D_E	1.53 (1.14, 1.92)	0.040 (−0.022, 0.13)	−0.093 (−0.84, 0.54)
	D_S	1.49 (1.12, 1.86)	0.005 (−0.089, 0.11)	−0.012 (−0.59, 0.51)
	D_{ES}	1.49 (1.11, 1.86)	0.000 (−0.008, 0.010)	−0.001 (−0.044, 0.059)
2 (Poisson Counts, with covs)	D_0	1.02 (0.77, 1.30)	−0.52 (−0.95, −0.029)	1.04 (0.70, 1.69)
	D_E	1.57 (1.17, 1.95)	0.034 (−0.015, 0.12)	−0.068 (−0.49, 0.085)
	D_S	1.52 (1.16, 1.89)	−0.012 (−0.11, 0.075)	0.025 (−0.33, 0.38)
	D_{ES}	1.54 (1.16, 1.91)	−0.000 (−0.010, 0.008)	0.000 (−0.034, 0.035)
3 (Neg Bin Counts, with covs)	D_0	1.03 (0.77, 1.32)	−0.48 (−1.02, 0.049)	1.04 (0.53, 1.91)
	D_E	1.54 (1.12, 1.99)	0.030 (−0.022, 0.13)	−0.065 (−0.62, 0.27)
	D_S	1.50 (1.10, 1.94)	−0.012 (−0.10, 0.076)	0.026 (−0.42, 0.43)
	D_{ES}	1.51 (1.10, 1.96)	−0.000 (−0.007, 0.007)	0.000 (−0.029, 0.035)

Table 3

Specified values of regression coefficients in each of seven scenarios (based on the three-stage path model (5), possibly with covariate W) used in the simulation study.

Scenario Outcome		Intercept	X	Z_1	Z_2	W
1 Dental Data	Z_1	−0.3	0.3	-	-	-
	Z_2	−0.4	0.1	0.01	-	-
	Y	0.4	−0.4	0.1	−0.5	-
2 All D_1	Z_1	−0.3	0.6	-	-	-
	Z_2	0.1	0	0	-	-
	Y	−0.1	0	0.3	0	-
3 All D_2	Z_1	−0.3	0	-	-	-
	Z_2	0.1	1.2	0	-	-
	Y	0.3	0	0	0.7	-
4 All D_{12}	Z_1	−0.6	1.0	-	-	-
	Z_2	0.3	0	2.3	-	-
	Y	−2.2	0	0	2.6	-
5 All D_0	Z_1	−0.1	0	-	-	-
	Z_2	0.6	0	0	-	-
	Y	0.1	0.2	0	0	-
6 Equal D_{12}	Z_1	−0.3	0.6	-	-	-
	Z_2	−0.7	0.6	1.8	-	-
	Y	−1.2	0.1	0.21	1.5	-
7 Equal + Covariate	Z_1	−0.5	0.6	-	-	0.5
	Z_2	−0.7	0.6	1.8	-	-
	Y	−1.8	0.1	0.21	1.5	0.7

Table 4

Simulation study results for six scenarios with n=200. For each estimand - the expected response, E{Y (D)}, the effect relative to no exposure, R(D), and the percent exposure effect due to pathway, P(D) - the table gives the true value, and the average estimate and percent coverage for the 95% confidence interval based on 500 replications. 'Power' (percent of 95% CIs that exclude 0) is also given for R(D) and P(D). Absolute values less than 0.01 are indicated by "0.00" (for positive values) or "−0.00" (for negative values).

		E{Y (D)}			*R (D)*				*P (D)*			
Scenario	**Pathway**	**True**	**Ave Est**	**% Cover**	**True**	**Ave Est**	**% Cover**	**Power**	**True**	**Ave Est**	**% Cover**	**Power**
1 Dental Data	D_0	0.91	0.92	0.94	−0.45	−0.44	0.93	0.81	1.05	1.10	0.97	0.82
	D_1	1.39	1.40	0.94	0.04	0.04	0.94	0.11	−0.09	−0.11	0.96	0.07
	D_2	1.34	1.35	0.94	−0.02	−0.02	0.95	0.06	0.04	0.01	0.97	0.05
	D_{12}	1.36	1.36	0.93	−0.00	−0.00	0.99	0.01	0.00	0.00	1.00	0.00
2 All D_1	D_0	1.17	1.17	0.93	0.00	0.00	0.94	0.06	0.00	−0.63	0.97	0.03
	D_1	1.45	1.45	0.93	0.28	0.28	0.93	0.99	1.00	1.67	0.97	0.47
	D_2	1.17	1.16	0.96	0.00	0.00	1.00	0.00	0.00	−0.04	1.00	0.00
	D_{12}	1.17	1.16	0.95	0.00	0.00	1.00	0.00	0.00	0.00	1.00	0.00
3 All D_2	D_0	1.14	1.13	0.95	0.00	−0.00	0.95	0.05	0.00	−0.18	0.97	0.03
	D_1	1.14	1.14	0.93	0.00	0.00	1.00	0.00	0.00	0.05	1.00	0.00
	D_2	1.33	1.33	0.93	0.20	0.19	0.93	0.95	1.00	1.20	0.97	0.29
	D_{12}	1.14	1.14	0.93	0.00	0.00	1.00	0.00	0.00	−0.07	1.00	0.00
4 All D_{12}	D_0	1.12	1.12	0.95	0.00	−0.01	0.97	0.03	0.00	1.11	0.98	0.02
	D_1	1.12	1.13	0.96	−0.00	0.01	0.96	0.02	−0.00	0.15	0.99	0.01
	D_2	1.12	1.12	0.95	0.00	−0.00	0.97	0.03	0.00	−0.27	0.99	0.01
	D_{12}	1.32	1.33	0.96	0.20	0.20	0.94	0.72	1.00	0.01	0.96	0.12
5 All D_0	D_0	1.35	1.35	0.95	0.24	0.24	0.96	0.33	1.00	0.96	1.00	0.62
	D_1	1.11	1.11	0.95	0.00	−0.00	1.00	0.00	0.00	−0.06	1.00	0.00
	D_2	1.11	1.11	0.94	0.00	−0.00	0.99	0.01	0.00	0.09	1.00	0.00
	D_{12}	1.11	1.11	0.94	0.00	−0.00	1.00	0.00	0.00	0.01	1.00	0.00
6 Equal	D_0	1.27	1.27	0.95	0.12	0.12	0.96	0.12	0.25	0.18	0.97	0.12

Scenario	Pathway	$E\{Y(D)\}$			$R(D)$				$P(D)$			
		True	Ave Est	% Cover	True	Ave Est	% Cover	Power	True	Ave Est	% Cover	Power
	D_1	1.26	1.26	0.93	0.12	0.11	0.90	0.67	0.25	0.28	0.95	0.54
	D_2	1.27	1.27	0.95	0.12	0.12	0.93	0.37	0.25	0.27	0.95	0.30
	D_{12}	1.26	1.27	0.94	0.12	0.12	0.95	0.72	0.25	0.28	0.97	0.61