Biostatistics (2015), **16**, 2, *pp.* 339–351
doi:10.1093/biostatistics/kxu048
Advance Access publication on November 12, 2014

Sensitivity analyses for parametric causal mediation effect estimation

JEFFREY M. ALBERT*

Department of Epidemiology and Biostatistics, Case Western Reserve University, Cleveland, OH 44106, USA

jma13@case.edu

WEI WANG

Center for Biostatistics and Bioinformatics, University of Mississippi Medical Center, Jackson, MS 39216, USA

SUMMARY

Causal mediation analysis uses a potential outcomes framework to estimate the direct effect of an exposure on an outcome and its indirect effect through an intermediate variable (or mediator). Causal interpretations of these effects typically rely on sequential ignorability. Because this assumption is not empirically testable, it is important to conduct sensitivity analyses. Sensitivity analyses so far offered for this situation have either focused on the case where the outcome follows a linear model or involve nonparametric or semiparametric models. We propose alternative approaches that are suitable for responses following generalized linear models. The first approach uses a Gaussian copula model involving latent versions of the mediator and the final outcome. The second approach uses a so-called hybrid causal-observational model that extends the association model for the final outcome, providing a novel sensitivity parameter. These models, while still assuming a randomized exposure, allow for unobserved (as well as observed) mediator-outcome confounders that are not affected by exposure. The methods are applied to data from a study of the effect of mother education on dental caries in adolescence.

Keywords: Causal inference; Copula; Interaction; Mediation analysis; Mediation formula; Potential outcome; Structural equations model.

1. INTRODUCTION

Mediation analysis has become popular as an approach for studying the mechanisms through which a treatment or exposure affects an outcome. In practice, mediation analysis utilizes one or more measured intermediate variables (or mediators) hypothesized to lie on the causal pathway between the exposure and the outcome. Typically, the analysis involves the decomposition of the overall exposure effect into a direct effect and an indirect (mediation) effect occurring through one or more intermediate variables.

*To whom correspondence should be addressed.

A classical approach to mediation analysis involves the fit of linear regression models for the outcome and mediator (Baron and Kenny, 1986). Recently, causal model (potential outcomes) approaches have been developed that provide causally interpretable mediation effects under flexible situations, including generalized linear models and models with interactions (Robins and Greenland, 1992; Albert, 2008; Imai *and others*, 2010b).

One very general approach to causal mediation analysis is based on the mediation formula (Imai *and others*, 2010b; Albert and Nelson, 2011; Pearl, 2012). Although a nonparametric version of this method is possible (Imai *and others*, 2010b), a very general and practical use of the mediation formula is via parametric models. Particular applications of the parametric mediation formula are in several recent papers (Wang and Albert, 2012; Wang *and others*, 2013).

The mediation formula, which provides a key tool for the methods to be presented in this paper, is based on the potential outcomes framework. Some basic definitions and notation for this framework are as follows. We let $Y_i(x)$ denote the potential outcome for (response variable) Y if causal factor X were set to level x for individual i. Note that we will sometimes drop the subscript i when not needed for clarity. The central problem of causal inference is that a response, Y, for any given individual, is observed (at a given time) for only one level of the exposure (X). Potential outcomes for exposure levels not actually observed are referred to as counterfactuals. To define mediation effects, we use nested potential outcomes. For example, $Y(x, M(x'))$ denotes the potential outcome for Y if exposure, X, were set to x, but the mediator, M, set to its potential outcome if X were set to x'. We let $E\{(Y(x, M(x'))\}$ denote the average of this potential outcome over a specified population. Then we define the direct and indirect effects, respectively, as

$$D(x) = E\{Y(1, M(x))\} - E\{Y(0, M(x))\}, \quad I(x) = E\{Y(x, M(1))\} - E\{Y(x, M(0))\}. \tag{1.1}$$

Note that $D(x)$ and $I(x)$ represent "natural" direct and indirect effects whereby the mediator, M, is set to the value (potential outcome) that would "naturally" be observed under the specified exposure level. These estimands are in contrast to those of "controlled" direct and indirect effects whereby M is fixed at a particular common value. The total effect, $T = E\{Y(1, M(1))\} - E\{Y(0, M(0))\}$, can be written as $T = D(0) + I(1) = D(1) + I(0)$, demonstrating that the natural direct and indirect effects represent an exact decomposition of the total exposure effect.

A key assumption in the identification of the natural direct and indirect effects (as used, e.g., by Imai *and others*, 2010b) is sequential ignorability. This assumption (discussed in detail in Section 2 below) comprises two conditions loosely stated as follows: (i) there are no unobserved confounders of the X–M or X–Y relationships and (ii) there are no unobserved (baseline or post-exposure) M–Y confounders. Unfortunately, these assumptions are not testable, although the first condition can sometimes be justified by the design, namely, when the exposure or treatment is randomly assigned or a sufficient set of prognostic baseline variables measured. The second part of the assumption, however, is generally difficult to make with confidence due to the possibility of unobserved M–Y confounders. For this reason, a number of scholars (including Imai *and others*, 2010b) emphasize the importance of conducting a sensitivity analysis to accompany mediation effect estimates.

Approaches to sensitivity analysis offered in the past (Imai *and others*, 2010b) have often relied on linear or linearizable (e.g., probit) models. In these cases, sequential ignorability amounts to assuming a zero correlation between the error terms for the Y and M models. Imai *and others* (2010b) proposed a sensitivity analysis that examines the impact of varying this correlation, and Imai *and others* (2010a) extended this approach to a binary outcome using a probit regression model. VanderWeele (2010) presented a general nonparametric methodology that considered an unobserved confounder between the final outcome and mediator. Although simple bias formulae are provided under simplifying assumptions, more generally, this approach may require complex specifications of quantities involving the distribution of (or distributions that condition on) one or more confounders. Hafeman (2011) simplified the required specifications but

limited the sensitivity analysis to the case of binary variables (exposure, mediator, and outcome). Recent contributions include Tchetgen Tchetgen and Shpitser (2012) who proposed a double robust semiparametric approach, and VanderWeele and Chiba (2014) who provided a nonparametric approach allowing for unobserved M–Y confounders affected by X. A limitation of the latter two methods is that the number of sensitivity parameters increases with the dimensionality of M, making these approaches difficult when M is not binary.

In the present paper, we propose two new approaches to sensitivity analysis designed to accommodate different types of variables, each assumed to follow a generalized linear model. The first approach uses a Gaussian copula model for the joint distribution of potential outcomes for M and Y. Wang *and others* (2013) used a similar idea but made stronger assumptions than those of the present approach, resulting in a different derived model. The second approach considers a "hybrid" model that extends the association model for Y by incorporating causal as well as observational effects. The new approaches represent novel alternatives to the predominant approach that specifies effects of unobserved confounders, an approach that goes back to Cornfield *and others* (1959); see also Rosenbaum (2002).

Following some background in Section 2, we present the new methods in Section 3. In Section 4, we jointly apply the two new methods to data from a study of parental factors in the development of dental caries among adolescents. A concluding discussion, including advantages and limitations of each method, is provided in Section 5.

2. Background

To demonstrate identifiability of natural direct and indirect effects, Imai *and others* (2010b) assumed the following, with L denoting a vector of baseline covariates (unaffected by X):

$$\{Y_i(x', m),\, M_i(x)\} \amalg X_i \mid L_i = l_i \tag{2.1a}$$

$$Y_i(x', m) \amalg M_i(x) \mid X_i = x,\, L_i = l_i. \tag{2.1b}$$

Assumption (2.1a) implies that the exposure (X) is random (or that exposure groups are exchangeable) conditional on L, while Assumption (2.1b) requires that there are no unobserved confounders (whether baseline or post exposure) of the M–Y relationship conditioning on X and L. The two assumptions together are referred to as "sequential ignorability".

Imai *and others* (2010b) showed that under sequential ignorability (along with a standard consistency assumption),

$$E\{Y(x', M(x))\} = \int_l \int_m E(Y \mid M = m, X = x', L = l)\, \mathrm{d}F_{M|X=x,L=l}(m)\, \mathrm{d}F_L(l). \tag{2.2}$$

The right hand side of (2.2) contains only association parameters and is thus estimable under appropriate association models for Y and M. Rather than integrate over an assumed distribution for L as indicated in (2.2), a convenient alternative approach is to sum over the empirical distribution of L for the sample or other chosen reference group (e.g., subsample of exposed subjects). Equation (2.2), along with (1.1), yields estimable expressions for the natural direct and indirect effects. These expressions have been referred by Pearl (2012) as the mediation formula, although we tend to refer to Equation (2.2) itself as such.

The assumption (2.1b) can be relaxed slightly for natural direct and indirect effects defined (as in (1.1)) in terms of expected potential outcomes (2.2). In particular, it suffices to assume an "independence in expectation" version of (2.1b). In the present paper, we consider an alternative assumption which also allows identification of the natural direct and indirect effects in combination with randomization of X

(2.1a). This assumption, which we refer to as "mediator comparability", is written as,

$$E\{Y(x', m) \mid M(1) = m, X = 1, L = l\} = E\{(Y(x', m) \mid M(0) = m, X = 0, L = l\}, \tag{2.3}$$

for all x', m, l. This assumption says that the mean potential outcome of Y for a given exposure level x' and mediator value m is the same for subgroups of the two observed exposure groups ($X = 0, 1$) that are observed at a common level m of the mediator (and covariate values, l), this holding for each m, l. As with sequential ignorability, the mediator comparability assumption is not testable from the data. Identifiability of natural direct and indirect effects using the mediator comparability assumption is shown in Appendix A of supplementary material available at *Biostatistics* online.

In a parametric approach to causal mediation analysis, the above potential outcome assumptions are used in conjunction with parametric association models for the outcome variables. For example, we will consider the following generalized structural equations model (GSEM),

$$h_2\{E(Y \mid X, M, L)\} = \beta_0 + \beta_1 X + \beta_2 M + \beta_3 L, \tag{2.4a}$$

$$h_1\{E(M \mid X, L)\} = \gamma_0 + \gamma_1 X + \gamma_2 L, \tag{2.4b}$$

where h_1, h_2 are invertible link functions, and the $\beta's$ and $\gamma's$ are unknown regression parameters, with β_3, γ_2 representing parameter vectors compatible with L. In our approach to estimation, which is based on maximum likelihood, we need to further specify conditional distributions for each response variable (Y, M). We may also need to estimate non-regression parameters, for example, variances for normally distributed outcomes. In the present paper, we will assume response variables to follow exponential family distributions.

3. NEW SENSITIVITY ANALYSIS METHODS

Our proposed methods, like that of Imai *and others* (2010b), assume randomization (or conditional exchangeability) of X. The methods thus focus on the departure from the second sequential ignorability condition (i.e., (2.1b) or (2.3)), which is, roughly speaking, the assumption of no unobserved M–Y confounders. Further, also like Imai *and others* (2010b), these approaches assume that any M–Y confounders are not affected by exposure. Note that such an effect would imply a causal model with a three-link path, thus departing from the present (two-link) mediation set-up. The former situation is considered by Imai and Yamamoto (2013) focusing on the linear model case; see also Albert and Nelson (2011) and Lange *and others* (2014).

In addition, we note that while our focus is on the estimands $D(1)$ and $I(0)$, the approaches for $D(0)$ and $I(1)$ are completely analogous. The alternative versions of the direct and indirect effects are generally equal only in the case of a linear (identity link) model for Y, and in the absence of an $X \times M$ interaction. For other generalized linear models, they will tend to be very close (as in the case of our data example) when there is no interaction.

3.1 *Copula model approach*

The first approach uses a Gaussian copula model for the joint distribution of potential outcomes of Y and M. Specifically, we suppose the existence of continuous latent variables, $Y^*(x', m)$ and $M^*(x)$, corresponding to the (partially) observable potential outcomes, $Y(x', m)$ and $M(x)$, respectively, for each x', x, and m. We assume that $Y^*(x', m)$ and $M^*(x)$ are bivariate normally distributed with correlation $\rho_{x'x}$ (assumed constant over m) conditional on X and L.

We make the standard consistency assumption that $M_i(x) = M_i$ if $X_i = x$, and $Y_i(x, m) = Y_i$ if $X_i = x$ and $M_i = M_i(x) = m$, where X_i is the observed level of exposure (X) for individual i (similarly, for M_i,

Y_i). For discrete M (taking values $0, 1, \ldots, K$), we further say that $M^*(x) = m^*$ is "compatible" with $M(x) = m$ if $P(M(x) \leqslant m-1) < F_{M^*(x)}(m^*) \leqslant P(M(x) \leqslant m)$, $P(M(x) \leqslant -1) \equiv 0$, where $F_{M^*(x)}(\cdot)$ is the cumulative distribution function of $M^*(x)$. From here on in, using the same symbol, one with an asterisk (the latent version), one without (the discrete, observed version), such as m^* and m, will imply that the two values are compatible.

Given the assumed bivariate normal distribution for Y^* and M^*, we have the regression relationship,

$$E\{Y^*(x', m) \mid M^*(x) = m^*\} = E\{Y^*(x', m)\} + \frac{\sigma_{Y^*(x',m)}}{\sigma_{M^*(x)}} \rho_{x'x}[m^* - E\{M^*(x)\}], \tag{3.1}$$

where $\sigma^2_{Y^*(x',m)}$, $\sigma^2_{M^*(x)}$ denote the variances of $Y^*(x', m)$ and $M^*(x)$, respectively, all expectations (variances, and so on) in (3.1) being conditional on $L = l$. From here on in, we will tend to leave conditioning on X and L out of the notation unless needed for clarity; note that X can be dropped anyway due to the assumption of randomized X (2.1a). We will assume that $\sigma^2_{Y^*(x',m)}$ is constant over m and write $\sigma^2_{Y^*_{x'}} \equiv \sigma^2_{Y^*(x',m)}$. We also write $\rho_0 \equiv \rho_{00}$ and $\rho \equiv \rho_{01}$ for brevity.

Our goal is the estimation of $E\{Y(x', M(x))\}$ for all $x', x \in \{0, 1\}$, which will allow estimation of the natural direct and indirect effects defined earlier. The case of $x' \neq x$ requires particular attention, and in fact is where the sequential ignorability assumption (i.e., (2.1b) or (2.3)) is ordinarily needed. For concreteness, and applicability to our data example, we focus on the estimand $E\{Y(0, M(1))\}$, needed for $D(1)$ and $I(0)$. The approach for the estimand $E\{Y(1, M(0))\}$, which is involved in the expressions for $D(0)$ and $I(1)$, is derived in a parallel manner by simply switching the values for x' and x. We indicate maximum likelihood estimates (i.e., obtained under (2.4)) using the standard hat notation.

When Y and M are continuous and assumed to be bivariate normally distributed then (3.1) holds for Y, M in place of Y^*, M^*, and m in place of m^*. In the special case where Y and M follow additive linear models (i.e., (2.4a,b) with identity link functions), it can then be shown that $D(1) = \beta_1 + \eta_o\gamma_1$ and $I(0) = (\beta_2 - \eta_o)\gamma_1$ where $\eta_0 = \rho_0(\sigma_{Y_0}/\sigma_{M(0)})$. Under homogeneous variance assumptions, allowing estimation of σ_{Y_0} and $\sigma_{M(0)}$, the correlation parameter, ρ_0, may be used as the sensitivity parameter. The above and more general expressions for $D(1)$ and $I(0)$ in the continuous Y case are derived in Appendix B of supplementary material available at *Biostatistics* online.

For discrete Y, we assume the regression relationship (3.1) for the latent final response variable Y^* and mediator M^* (also latent if M is discrete). In this case, constraining the marginal distribution of $Y^*(x', m)$ to be standard normal, the right side of (3.1) can be reduced to give,

$$E\{Y^*(x', m) \mid M^*(x) = m^*\} = \rho_{x'x} m^*_{sx}, \tag{3.2}$$

where $m^*_{sx} \equiv [m^* - E\{M^*(x)\}]/\sigma_{M^*(x)}$. Given our assumptions, we have $Y^*(x', m) \mid M^*(x) = m^* \sim N(\rho_{x'x} m^*_{sx}, 1 - {\rho_{x'x}}^2)$.

We can then estimate $E\{Y(0, M(1))\}$ for given ρ_0 and ρ using Monte Carlo simulation. Here we present the algorithm for the case where M is continuous (and assumed normally distributed), in which case we use M in place of M^* (and m in place of m^*) in the above expressions. Further details, including the case where M is discrete, are given in Appendix C of supplementary material available at *Biostatistics* online. The algorithm is implemented for selected values for ρ_0 and ρ. For each person (i) in the chosen reference group, do the following for replications $r = 1, \ldots, R$,

(1) Draw m_{s1ir} from $N(0, 1)$. Let $m_{s0ir} = [m_{s1ir}\hat{\sigma}_{M(1)} + \hat{E}\{M_i(1)\} - \hat{E}\{M_i(0)\}]/\hat{\sigma}_{M(0)}$.
(2) Supposing the response for a given i is $Y_i \in \{0, 1, \ldots, J\}$, let $y_{ir} = \sum_{j=1}^{J} j(\hat{P}_{irj} - \hat{P}_{ir,j-1})$, where $\hat{P}_{irj} = \Phi\{[(1-\rho_0^2)^{1/2}\Phi^{-1}(\hat{P}_{0irj}) + (\rho_0 m_{s0ir} - \rho m_{s1ir})]/(1-\rho^2)^{1/2}\}$, $\hat{P}_{0irj} = \hat{P}(Y_i(0, m_{ir}) \leqslant j \mid M_i(0) = m_{ir}, L_i = l_i) = \hat{P}(Y_i \leqslant j \mid X_i = 0, M_i = m_{ir}, L_i = l_i)$, and $m_{ir} = m_{s1ir}\hat{\sigma}_{M(1)} + \hat{E}\{M_i(1)\}$.

Then, compute $\hat{E}\{Y_i(0, M(1)) \mid L_i = l_i\} = (1/R)\sum_{r=1}^{R} y_{ir}$ and $\hat{E}\{Y(0, M(1))\} = (1/n)\sum_{i=1}^{n} \hat{E}\{Y_i(0, M(1))\}$. Note that $\hat{P}_{irj}$ (in the second step above) represents the estimated conditional (cumulative) probability of $Y_i(0, m) \leqslant j$ given $M_i(1) = m_{ir}$, and y_{ir} represents the estimated conditional expected value of $Y_i(0, m)$ given $M_i(1) = m_{ir}$, both conditional on $L_i = l_i$.

The natural direct and indirect effects, using the above estimated expected value, may be recomputed for varying values for ρ_0 and ρ thus providing a sensitivity analysis. Confidence intervals for mediation effects for given ρ_0 and ρ can be computed using the percentile bootstrap resampling method.

3.2 *Hybrid model approach*

In our second approach, we consider a hybrid causal-observational model for Y,

$$E\{(Y(x', m) \mid M(x) = m, X = x, L = l\} = g(x', x, m, l), \tag{3.3}$$

where the right-hand side represents some specified parametric model. A key feature of this model is that it distinguishes different roles of the exposure, namely (i) a causal factor directly affecting Y (denoted as $X' = x'$) and (ii) the observed exposure group or cohort ($X = x$) representing selection bias. It is possible to further generalize this model by distinguishing the exposure term in $M(x)$, but this will not be pursued here.

For our application, we consider the following additive hybrid model

$$E\{Y(x', m) \mid M(x) = m, X = x, L = l\} = h_2^{-1}\{\beta_0 + \beta_{1c}x' + \beta_{1b}x + \beta_2 m + \beta_3 l\}, \tag{3.4}$$

where β_{1c} and β_{1b} represent causal and cohort (selection bias) effects, respectively. Mediator comparability is obtained when the coefficients of all terms involving x in the hybrid model (e.g., β_{1b} in (3.4)) are equal to 0. Note that the hybrid model (3.4) is distinct from previous causal models that may bear a resemblance to it, for example, Vansteelandt *and others* (2012) and Lange *and others* (2012). In the causal models in the latter papers, the dual exposure indicators both have a causal interpretation, as needed to define natural direct and indirect effects, whereas, in the hybrid model, they refer to different roles, i.e., causal versus observational, for the exposure variable.

We also consider the following useful reparameterization of the hybrid model (3.4):

$$E\{Y(x', m) \mid M(x) = m, X = x, L = l\} = h_2^{-1}\{\beta_0 + \beta_1(\phi x + (1 - \phi)x') + \beta_2 m + \beta_3 l\}, \tag{3.5}$$

where $\beta_1 \equiv \beta_{1c} + \beta_{1b}$ and $\phi \equiv \beta_{1b}/(\beta_{1c} + \beta_{1b})$. As in model (2.4a), β_1 in (3.5) represents an association parameter, namely the effect of X on Y, conditional on M and L, on the scale corresponding to the link function, h_2. The sensitivity parameter, ϕ, represents the (nonidentifiable) proportion of the association effect due to the cohort effect, and, consequently, $(1 - \phi)$ is the proportion of the association effect due to the causal effect of exposure. As the causal and cohort effects, β_{1c} and β_{1b} in (3.4), need not be in the same direction, it is possible for the "proportion", ϕ, to be negative or >1. Note that this reparameterization is only possible in the case of $\beta_1 \equiv \beta_{1b} + \beta_{1c} \neq 0$ (i.e., a non-zero observed effect of exposure).

The case of $\phi = 0$ in model (3.5) implies that $Y(x', m)$ is (mean) independent of X given M and L, and thus implies mediator comparability (2.3). In contrast, $\phi \neq 0$ represents a departure from mediator comparability, and thus from sequential ignorability. Note that, as seen in (3.5), ϕ does not affect $E\{Y(x', m)\}$ (nor, consequently, $E\{Y(x', M(x))\}$) when $x' = x$. This is as expected because then, under consistency, $E\{Y(x', M(x))\} = E\{Y(x')\}$, which is estimable under randomization of X (2.1a), that is, without having to assume (2.1b) or (2.3). In Appendix D of supplementary material available at *Biostatistics* online, we provide a demonstration of the identifiability of the natural direct and indirect effects under the hybrid

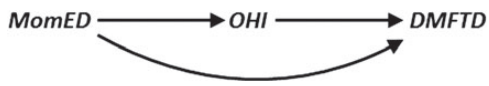

Fig. 1. Mediation model for dental data (confounding variables not shown).

model with specified ϕ. The effect of departures from sequential ignorability on estimates (and percentile bootstrap confidence intervals) of the natural direct and indirect effects can thus be studied by varying ϕ.

Some comments are in order on the specification of ϕ. The direct specification of ϕ may be difficult as ϕ represents a measure of "collider-stratification bias" (see, e.g., Greenland, 2003) whose magnitude, and even direction, may be non-intuitive to many researchers. Fortunately, this bias, or an approximation, may be derived as a function of the effects of (possibly multiple) unobserved M–Y confounders on M and Y. Detailed guidelines for the elicitation of plausible values for ϕ are provided in Appendix E of supplementary material available at *Biostatistics* online.

4. Application to dental data

We apply the two proposed sensitivity analysis methods to data from a dental caries study (Nelson *and others*, 2010). The study examined a cohort of very low birth weight (VLBW, with and without bronchopulmonary dysplasia, BPD) and a matched group of normal birth weight (NBW) children followed from birth through adolescence in an earlier study (Singer *and others*, 1997). This earlier study recorded parent information, including education, knowledge, and stress factors, from questionnaires given at the child's birth, ages 3 and 8. The children who continued on to the dental study underwent a dental clinical exam at around age 14. This exam provided, among other information, the number of decayed, missing, and filled teeth (DMFT) and the oral hygiene index score (OHI), an indicator of effective oral hygiene behavior.

Previous research considered the effect of mother education on child's dental outcomes at adolescence (Nelson *and others*, 2010). Nelson *and others* (2012) found that adolescents whose mothers had less than a high school education when the child was age 3 (compared with those whose mother had at least a high school education) had a higher mean DMFT. In the present paper, we seek to assess the extent to which this effect is mediated through OHI.

Our specific model variables included a binary exposure variable, "MomEd" ($X = 1$ for mother's education at or below high school level, "low", 0 otherwise, "high"), a binary final outcome, "DMFTD" ($Y = 1$ if the child had DMFT >0 at the age 14 exam, 0, otherwise), and a roughly continuous mediator, OHI (ranging from 0 to 3, where higher is worse). We assumed that the model variables are causally related as indicated in the graph in Figure 1. In an initial mediation analysis, we assumed sequential ignorability controlling for the (child) variables: sex, race (African American versus other), and birth status (two indicator variables for the three categories: VLBW plus BPD, VLBW alone, and NBW).

To describe the links in the mediation model, we assumed a special case of the GSEM (2.4a,b)) comprised of a logistic regression model for the binary outcome variable and a linear regression model for the continuous mediator as follows:

$$\text{logit}\{P(Y = 1 \mid X, M, L)\} = \beta_0 + \beta_1 X + \beta_2 M + \beta_3 L \tag{4.1a}$$

$$M = \gamma_0 + \gamma_1 X + \gamma_2 L + \varepsilon_M, \tag{4.1b}$$

where $\varepsilon_M \mid X, L \sim N(0, {\sigma_M}^2)$ and ${\sigma_M}^2$ is an unknown variance parameter. As an alternative, we considered the above set of models but where (4.1a) is extended by adding an exposure by mediator ($X \times M$)

Table 1. *Parameter estimates(standard errors) from ML fit of model (4.1) of DMFTD (without and with MomEd x OHI interaction) and OHI outcomes from dental data*

	DMFTD		
Covariates	Additive	Interaction	OHI
Intercept	−1.04 (0.45)	−1.24 (0.52)	0.83 (0.15)
MomEd	0.53 (0.33)	0.92 (0.61)	0.29 (0.12)
OHI	0.66 (0.22)	0.86 (0.35)	–
MomEd × OHI	–	−0.35 (0.44)	–
VLBW vs NBW	0.21 (0.42)	0.20 (0.43)	−0.093 (0.15)
BPD vs NBW	−0.46 (0.39)	−0.49 (0.39)	0.19 (0.14)
Race	0.57 (0.33)	0.55 (0.33)	0.29 (0.12)
Sex	−0.23 (0.33)	−0.24 (0.34)	0.04 (0.12)

interaction term. The component models ((4.1a), without and with interaction, and (4.2b)) may be fit separately using standard likelihood methods. The estimated regression parameters for both the additive and interaction GSEM (the model for M is the same in both cases) are given in Table 1.

For each GSEM, we then estimated the natural direct effect, $D(1)$, and the natural indirect effect, $I(0)$, using the mediation formula (2.2) plugging in parameter estimates obtained under model (4.1a,b) and using the VLBW subsample as the reference group. We present results from the additive model; results from the interaction model (not shown) are nearly the same, indicating that our mediation effect inferences are not changed substantially by inclusion of a MomEd by OHI ($X \times M$) interaction effect in the DMFTD (Y) model.

Using the mediation formula under the assumption of sequential ignorability, we obtained an estimated natural direct effect (bootstrap 95% confidence interval) of $\hat{D}(1) = 0.12$ $(-0.04, 0.27)$, indicating that the probability of a DMFT would increase by an estimated 0.12 if MomEd (X) were changed from "high" to "low" but the OHI (M) fixed as if the mother had "low" education. The estimated natural indirect effect was $\hat{I}(0) = 0.04$ (0.01, 0.09), indicating that the probability of a DMFT would increase by an estimated 0.04 if MomEd was changed from "high" to "low" in a way that naturally affected OHI without affecting DMFTD through any other path. We also note that the estimated total effect (95% confidence interval) was $\hat{T} = 0.16$ (0.003, 0.31) and the indirect effect (mediation) proportion an estimated 0.26. As indicated by the above confidence intervals, the indirect as well as the total (but not the direct) effect is statistically significant at the nominal 0.05 α-level.

The above estimates and confidence intervals rely on the assumption of sequential ignorability (or (2.1a) and mediator comparability (2.3)). In the context of the dental data, mediator comparability implies that if set to a common exposure and mediator level, subgroups of the two observed exposure groups (with high versus low mother education) that are observed at a common observed mediator (oral hygiene index) level (and the same baseline control variable values) would have the same probability of dental caries. This assumption is unlikely for the dental data because there are apt to be unobserved confounders of the relationship between the oral hygiene index and dental caries. It is of interest to examine how the estimates of the natural direct and indirect effects would change under violations of the mediator comparability assumption.

We conducted sensitivity analyses for the dental data using the two new approaches described in Section 3. The results of these analyses, including estimated (natural) direct and indirect effects and 95% confidence intervals, are presented in Figure 2. Scientific considerations may allow us to narrow the range of plausible values for the sensitivity parameters, and thus obtain sharper conclusions about mediation

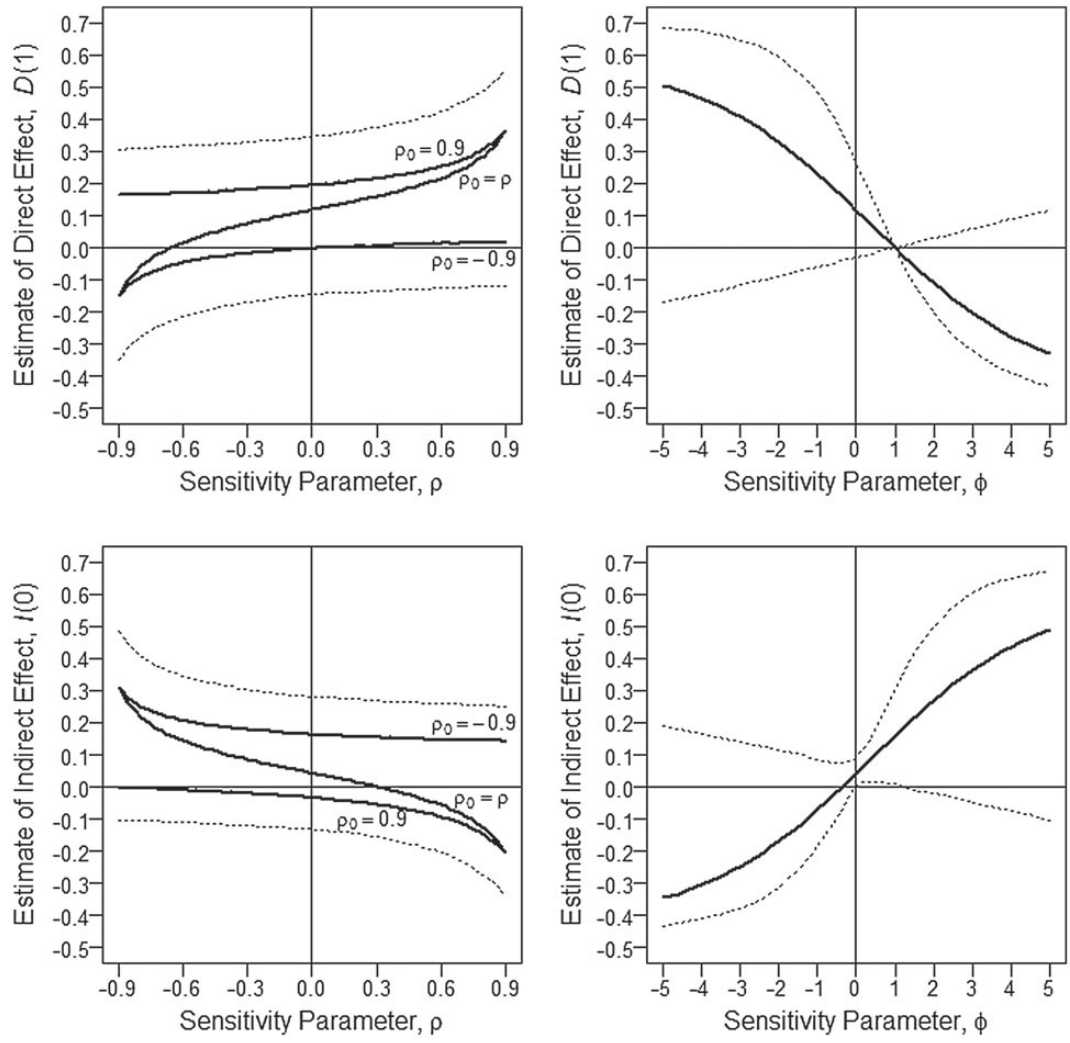

Fig. 2. Sensitivity analysis for dental data. Maximum likelihood estimates of direct (top) and indirect (bottom) effects versus sensitivity parameters from copula model (left) and hybrid model (right). Solid line = estimates, dotted lines = 95% confidence interval bounds. For copula method, the upper bounds are for $\rho_0 = 0.9$ (direct), $\rho_0 = -0.9$ (indirect); the lower bounds are for $\rho_0 = -0.9$ (direct), $\rho_0 = 0.9$ (indirect).

effects. The elicitation of plausible sensitivity parameter values for the present data example is described in Appendix E of supplementary material available at *Biostatistics* online.

In the context of our data example, the hybrid model sensitivity parameter, ϕ, is interpreted as the proportion of the association parameter, β_1, that is due to selection bias, the remaining proportion being due to a true causal effect of MomEd. In the hybrid model (3.5), as in the corresponding association model (4.1a), β_1 is interpreted as the log DMFTD odds ratio for low versus high MomEd groups conditional on OHI and the included baseline covariates.

For the dental data, β_1 was estimated to be 0.53, providing an estimated odds ratio for low versus high MomEd of $\exp(0.53) = 1.69$. Our elicitation suggested a range of plausible values for ϕ of -1 to 0, and

a best (pessimistic) guess value of -0.4. The value of $\phi = -0.4$ implies that the selection bias portion of the association effect is $-0.4(0.53) \simeq -0.21$, while the portion due to the causal effect of MomEd on DMFTD is $(1 - (-0.4))(0.53) \simeq 0.74$. Note that this decomposition can be expressed on the odds ratio scale: $\exp(0.53) \simeq \exp(-0.21)\exp(0.74)$ or $1.69 \simeq (0.81)\ (2.09)$. For the copula model, our elicitation (Appendix E of supplementary material available at *Biostatistics* online) provided plausible (pessimistic) values of $\rho_0 = 0.5$ and $\rho = 0.55$, and a plausible range for ρ_0 of [0.1, 0.7] and for ρ of [0.1, 0.75].

As indicated in the hybrid model curves in Figure 2, the elicited value of $\phi = -0.4$ provides an estimate (95% CI) for $D(1)$ of 0.16 (-0.04, 0.36) and an estimate (95% CI) for $I(0)$ of $-0.004(-0.07, 0.07)$. The plausible range for ϕ of $[-1, 0]$ provides a range for the estimate of $D(1)$ of 0.12 to 0.23 and for the estimate of $I(0)$ of -0.07 to 0.04. Figure 2 further shows that the 95% confidence intervals for $I(0)$ fail to exclude 0 for values of $\phi < 0$; also, for $\phi < -0.4$, the estimate for $I(0)$ is no longer positive. For $D(1)$, neither the sign nor lack of statistical significance changes over the plausible range for ϕ.

From the copula model, the pessimistic scenario of $\rho_0 = 0.5$, $\rho = 0.55$ results in an estimate (95% CI) for $D(1)$ of 0.2 (0.05, 0.37) and for $I(0)$ of $-0.04(-0.09, 0)$. As noted in Appendix E of supplementary material available at *Biostatistics* online, this scenario may be viewed as more pessimistic than the corresponding values obtained for the hybrid model. For the (lower bound) plausible values of $\rho_0 = \rho = 0.1$, we obtain an estimate (95% CI) for $D(1)$ of 0.13 (-0.023, 0.28) and for $I(0)$ of 0.028 (-0.03, 0.069). From these results, and as indicated in Figure 2, the estimate of $I(0)$ is no longer statistically significant (and goes from positive to negative) as ρ_0 and ρ increase over their plausible ranges. Estimates for $D(1)$ increase as the ρ's increase and are statistically significant in part of the plausible range (as seen for the "pessimistic" scenario discussed above). These results are similar (apart from some differences in confidence interval coverage) to those obtained from the hybrid model approach.

In regard to other features of the graphs, we note that the degenerate confidence interval at $\phi = 1$ in the hybrid model approach is due to the fact that this value implies no natural direct effect of X on Y. Another difference between the two methods is that the confidence intervals from the copula model are generally more symmetric, reflecting the assumption of an underlying bivariate normal distribution for M and Y.

5. DISCUSSION

We have presented two new approaches to sensitivity analysis for estimation of natural direct and indirect effects via the parametric mediation formula. Both approaches are flexible in allowing different variable types (i.e., discrete or continuous) for the outcome and mediator, each response variable following a generalized linear model. The approaches involve models (the copula and hybrid, respectively) providing particular departures from sequential ignorability, specifically (2.1b) or the mediator comparability assumption (2.3). Although our focus was on natural direct and indirect effects on a difference scale (as in (1.1)), because both sensitivity analysis methods provide estimates of expected potential outcomes, it would be straightforward to extend the method to inference on mediation effects defined on alternative scales, for example, relative risk or odds ratio for a binary Y. Of course, the relationship between the sensitivity analysis parameters presented here and the mediation effects may be different depending on the scale used for the latter.

The two new sensitivity analysis methods have been encoded in SAS macros, which are available for downloading from http://epbiwww.case.edu/index.php/people/faculty/53-albert. These macros include alternative model and distributional options then those discussed in this paper. Additional examples, assuming normal and negative binomial distributions for Y, are given in Appendix G of supplementary material available at *Biostatistics* online.

A difficulty of previous sensitivity analysis models that involve a hypothetical unobserved confounder is that the resulting model is generally incompatible with the model without the confounder (Lin *and others*,

1998). This problem is avoided in the proposed approaches as the corresponding sensitivity analysis models do not directly involve an unobserved confounder. Rather, in each of the new approaches the sensitivity analysis and association models are compatible by construction; for example, in the hybrid model the latter is obtained from the former when $x' = x$. However, unobserved confounders can be considered, and play a useful role, in the elicitation of plausible sensitivity parameter values for the proposed methods as described in Appendix E of supplementary material available at *Biostatistics* online.

Between our two proposed methods, the copula model approach has an advantage of involving bounded (correlation) parameters. In addition, due to its underlying bivariate normal assumption and consequent linear structure, it appears to have more favorable inference properties, in particular, relatively narrow and symmetric confidence intervals across the range of sensitivity parameter values. The hybrid model approach, on the other hand, has the advantage of greater computational simplicity and a single sensitivity parameter (though more complex hybrid models, for example, involving an X–M interaction, may be possible).

One way to reduce the sensitivity parameter dimensionality, and allow a simpler graphical presentation, for the copula model is to assume $\rho = \rho_0$ (analogously, $\rho_{10} = \rho_1$ for the estimation of $D(0)$ and $I(1)$). A more flexible approach would be to assume a functional relationship between the two correlation parameters, for example, $\text{logit}(\rho) = \text{logit}(\rho_0) + c$, where c is a specified constant.

As elaborated in Appendix E of supplementary material available at *Biostatistics* online, carrying out both methods has the advantage that each can be used to calibrate and check the other. Although more work is needed in systematic approaches to combining sensitivity analyses, the present paper suggests the great potential for multiple models to provide a more refined and complete sensitivity analysis.

Supplementary material

Supplementary material is available at http://biostatistics.oxfordjournals.org.

Acknowledgments

The authors thank the Associate Editor and referees for insightful and constructive comments that helped greatly in improving the paper. The authors also thank Cuiyu Geng for assistance in the construction of graphs and preparation of the paper, Dr Suchitra Nelson for helpful discussion and for providing data from her study of dental outcomes in VLBW and NBW children [NIDCR/NIH research grant number R21-DE16469], and Dr Lynn Singer for providing access to data from her cohort study of VLBW and NBW adolescents, supported by the Maternal and Child Health Program, Health Resources and Services Administration, Department of Health and Human Services (grant numbers MC-390592, MC-00127, and MC-00334). *Conflict of Interest*: None declared.

Funding

Support for this research was provided in part by the National Institute of Dental and Craniofacial Research, National Institutes of Health (R01DE022674 to J.A.).

References

Albert, J. M. (2008). Mediation analysis via potential outcomes models. *Statistics in Medicine* **27**, 1282–1304.

Albert, J. M. and Nelson, S. (2011). Generalized causal mediation analysis. *Biometrics* **67**, 1028–1038.

Baron, R. M. and Kenny, D. A. (1986). The moderator-mediator variable distinction in social psychological research: conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology* **51**, 1173–1181.

Cornfield, J., Haenszel, W., Hammond, E. C., Lilienfeld, A. M., Shimkin, M. B. and Wynder, E. L. (1959). Smoking and lung cancer: recent evidence and a discussion of some questions. *Journal of the National Cancer Institute* **22**, 173–203.

Greenland. (2003). Quantifying biases in causal models: classical confounding vs. collider-stratification bias. *Epidemiology* **14**, 300–306.

Hafeman, D. M. (2011). Confounding of indirect effects: a sensitivity analysis exploring the range of bias due to a cause common to both the mediator and the outcome. *American Journal of Epidemiology* **174**, 710–717.

Imai, K., Keele, L. and Tingley, J. (2010a). A general approach to causal mediation analysis. *Psychological Methods* **15**, 309–334.

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

Imai, K. and Yamamoto, T. (2013). Identification and sensitivity analysis for multiple causal mechanisms: revisiting evidence from framing experiments. *Political Analysis* **21**, 141–171.

Lange, T., Rasmussen, M. and Thygesen, L. C. (2014). Assessing natural direct and indirect effects through multiple pathways. *American Journal of Epidemiology* **179**, 513–518.

Lange, T., Vansteelandt, S. and Bekaert, M. (2012). A simple unified approach for estimating natural direct and indirect effects. *American Journal of Epidemiology* **176**, 190–195.

Lin, D. Y., Psaty, B. M. and Kronmal, R. A. (1998). Assessing the sensitivity of regression results to unmeasured confounders in observational studies. *Biometrics* **54**, 948–963.

Nelson, S., Albert, J. M., Lombardi, G., Wishnek, S., Asaad, G., Kirchner, H. L. and Singer, L. T. (2010). Dental caries and enamel defects in very low birth weight adolescents. *Caries Research* **44**, 509–518.

Nelson, S., Lee, W., Albert, J. M. and Singer, L. T. (2012). Early maternal psychosocial factors are predictors for adolescent caries. *Journal of Dental Research* **44**, 509–518.

Pearl, J. (2012). The causal mediation formula a guide to the assessment of pathways and mechanisms. *Prevention Science* **13**, 426–436.

Robins, J. M. and Greenland, S. (1992). Identifiability and exchangeability for direct and indirect effects. *Epidemiology* **3**, 143–155.

Rosenbaum, P. R. (2002). *Observational Studies*, 2nd edition. New York: Springer-Verlag.

Singer, L. T., Yamashita, T. S., Lilien, L., Collin, M. and Baley, J. (1997). A longitudinal study of infants with bronchopulmonary dysplasia and very low birthweight. *Pediatrics* **100**, 987–993.

Tchetgen Tchetgen, E. J. and Shpitser, I. (2012). Semiparametric theory for causal mediation anlaysis: efficiency bounds, multiple robustness, and sensitivity analysis. *Annals of Statistics* **40**, 1816–1845.

VanderWeele, T. J. (2010). Bias formulas for sensitivity analysis for direct and indirect effects. *Epidemiology* **21**, 540–551.

VanderWeele, T. J. and Chiba, Y. (2014). Sensitivity analysis for direct and indirect effects in the presence of exposure-induced mediator-outcome confounders. *Epidemiology, Biostatistics, and Public Health* e9027-1–e9027-15.

Vansteelandt, S., Bekaert, M. and Lange, T. (2012). Imputation strategies for the estimation of natural direct and indirect effects. *Epidemiologic Methods* **1**(1), Article 7.

Wang, W. and Albert, J. M. (2012). Estimation of mediation effects for zero-inflated regression models. *Statistics in Medicine* **31**, 3118–3132.

Wang, W., Nelson, S. and Albert, J. M. (2013). Estimation of causal mediation effects for a dichotomous outcome in multiple-mediator models using the mediation formula. *Statistics in Medicine* **32**, 4211–4228.

[*Received January 17, 2014; revised September 24, 2014; accepted for publication September 28, 2014*]