General Multiple Mediation Analysis With an Application to Explore Racial Disparities in Breast Cancer Survival

Qingzhao Yu[1]*, Ying Fan[2] and Xiaocheng Wu[3]

[1]*Associate Professor of Biostatistics, Louisiana State University Health Sciences Center, USA*
[2]*Biostatistician, Covance at Otsuka Pharmaceutical Development & Commercialization, Inc., USA*
[3]*Associate Professor of Epidemiology, Director of Louisiana Tumor Registry, USA*

Abstract

Mediation refers to the effect transmitted by mediators that intervenes in the relationship between an exposure and a response variable. Mediation analysis has been broadly studied in many fields. However, it remains a challenge for researchers to differentiate individual effect from multiple mediators. This paper proposes general definitions of mediation effects that are consistent for all different types (categorical or continuous) of response, exposure, or mediation variables. With these definitions, multiple mediators can be considered simultaneously, and the indirect effects carried by individual mediators can be separated from the total effect. Moreover, the derived mediation analysis can be performed with general predictive models. For linear predictive models with continuous mediators, we show that the proposed method is equivalent to the conventional coefficients product method. We also establish the relationship between the proposed definitions of direct or indirect effect and the natural direct or indirect effect for binary exposure variables. The proposed method is demonstrated by both simulations and a real example examining racial disparities in three-year survival rates for female breast cancer patients in Louisiana.

Keywords: Causal inference; Confounder; Health disparity; (In) Direct effect; Multiple mediators

Introduction

Mediation effect refers to the effect conveyed by an intervening variable to an observed relationship between an exposure and a response variable of interest. The concept of mediation originated in psychological research. In 1929, Robert S. Woodworth introduced the expression Stimulus-Organism-Response to describe the pathway between stimulus and response. This concept has then been expanded to many fields such as social science, prevention study, behavior research, and epidemiology. Investigators are interested in discovering not only the relationship between the exposure variable and response variable but also the mechanism of risk factors mediating the relationship. For example, it has been well established that low socioeconomic status is associated with poor health status. To reduce this health disparity, investigators would need to quantify the mediation effects from different risk factors so that efficient interventions can be carried out [1].

There are generally two settings for mediation analysis. One is based on linear models. According to Baron and Kenny [2], three conditions are required to establish a mediation: (a) the exposure variable (X) is significantly associated with the response variable (Y); (b) the presumed mediator (M) is significantly related to X; and (c) M significantly relates to Y controlling for X. When the relationships are represented by linear regression models, the indirect effect is typically measured by two methods: the difference in the coefficients of the exposure variable when it is regressed on Y with or without controlling for the mediator (abbreviated as `CD' thereafter. [3-5]; or the product of the coefficient of $X$ when regressed on $M$, and the coefficient of $M$ in explaining $Y$ controlling for $X$ (`CP' for abbreviation [6]). MacKinnon et al. [7] showed that when the relationship among mediators, exposure and response variables are fitted with linear regression models, the mediation effects measured by CD or CP are equivalent. However, neither CD nor CP is easily adaptable to separate multiple mediation effects when Y or M is not continuous or when the relationships cannot be fitted with linear regressions [8].

Counterfactual framework is the other popular setting to implement mediation analysis [9-12]. Let $Y_i(X)$ denote the post-treatment potential outcome if subject i is exposed to X. To compare the change in outcome when the exposure changes from x to $x^\star$ (e.g., 0 or 1 for binary X), only one of the responses, $Y_i(x)$ or $Y_i(x^\star)$, is observed. The causal effect of treatment on the response variable for subject i is defined as $Y_i(x)-Y_i(x^\star)$. It is impossible to estimate the individual causal effect since the estimation depends on a non-observable response. Holland [13] proposed, instead of estimating causal effect on a specific subject, to estimate the average causal effect over a pool of subjects $-E(Y_i(x)-Y_i(x^\star))$. If the subjects are randomly assigned to control or treatment groups, the average causal effect equals the expected conditional causal effect, $E(Y_i|X=x)-E(Y_i|X=x^\star)$.

Denote $M_i(X)$ as the potential value of M when subject i is exposed to treatment X. Let $Y_i(x,m)$ be the potential outcome of subject i for a given x and m. It has been established that the total effect of X on Y when X changes from x to $x^\star$ is $Y_i(x,M_i(x))-Y_i(x^\star,M_i(x^\star))$. The conventional mediation analysis decomposes the total effect into direct effect and indirect effect. Namely, direct effect is the effect of X directly on Y, while indirect effect is the effect of X on Y through M. Robins and Greenland [9] introduced the concepts of controlled direct effect, defined as $Y_i(x,m)-Y_i(x^\star,m)$ and of natural direct effect, defined as $\varsigma_i(x) = Y_i(x,M_i(x^\star))-Y_i(x^\star,M_i(x^\star))$. The difference between controlled and natural direct effects is that the controlled direct effect is measured when M is fixed at m, whereas for the natural direct effect, M is random as if the actual exposure were $x^\star$. The difference between total effect

*Corresponding author: Qingzhao Yu, Associate Professor of Biostatistics, Louisiana State University Health Sciences Center, USA, Tel: (504) 568-6086; E-mail: qyu@lsuhsc.edu

and natural direct effect is defined as the natural indirect effect, $\delta_i(x) \equiv Y_i(x, M_i(x)) - Y_i(x^\star, M_i(x^\star))$. In comparison, the difference between a total effect and a controlled direct effect cannot in general be interpreted as an indirect effect [13,14]. A common restriction for definitions of controlled and natural direct effects is that the exposure levels x and $x^\star$ have to be preset. When the relationship among variables cannot be assumed linear, it is hard to choose representative exposure levels especially if the exposure variable is multi-categorical or continuous.

In this paper, we propose general definitions of mediation effects. The derived mediation analysis is promising in that the indirect effects contributed by different mediators are separable, which enables the comparison of relative mediation effects carried by different third variables. Furthermore, the mediation analysis is generalized so that we can deal with binary, multicategorical or continuous exposure, mediator and response variables. The method allows general predictive models, in addition to general linear models, to be used to fit variable relationships. We show that the proposed method is equivalent to the conventional CD and CP methods in single continuous mediator cases. We also establish the relationship between the proposed definitions of direct or indirect effect, and the natural direct or indirect effects in single binary mediator cases.

The paper is organized as follows: in Section 2, we present a motivating example that explores racial disparities of female breast cancer three-year survival rates in Louisiana. Section 3 proposes general definitions of mediation effects. We demonstrate the mediation analysis with linear regression and logistic regression, which deals with different types of mediators in Section 4 and 5. In Section 6, we propose algorithms for mediation analysis with non/semi-parametric predictive models and binary exposures. Statistical inference on indirect effect by the Delta and bootstrap methods is discussed in Section 7. In Section 8, the proposed method is adapted to the motivating example. Also a simulation study is used to demonstrate its identifiability and sensitivity. Finally, conclusions and future research are discussed in Section 9.

Motivating Example

Breast cancer is the most common cancer and the second leading cause of cancer death among American women of all races. Owing to advanced screening technologies for early stage breast cancer detection, as well as improved treatment modalities, the overall death rate of breast cancer in US has decreased in recent years. However, when compared with white women, African-American women have a higher death rate from breast cancer despite the fact their incidence rate is lower. Previous studies have found that more advanced and aggressive tumors and less than optimal treatment may explain the lower survival rates among black women [15-17]. There are many other contributing factors, such as patient demographic information and health care provider information. However, the relative effect contributed by each factor to racial disparities in breast cancer survival cannot be differentiated due to the lack of comprehensive data and the limitations of current analysis methods (discussed in Section 1).

The Institute of Medicine (IOM) reported in 1999 that cancer patients did not consistently receive the care known to be effective for their conditions [18]. In response to the IOM reports, the National Program of Cancer Registries (NPCR) of the Centers of Disease Control and Prevention (CDC) established a series of Pattern of Care (PoC) studies. The PoC-Breast and Prostate (BP) study collected information from breast and prostate cancer patients. The routinely collected registry data were supplemented by re-abstracting hospital records and obtaining information about adjuvant treatment and comorbidity from

physicians and outpatient facilities. We use the data set collected by the Louisiana Tumor Registry on about 1453 non-Hispanic white and black women diagnosed with malignant breast cancer in 2004 in Louisiana. All patients were followed up for five years or until death, whichever is shorter. We found that the odds of dying of breast cancer within three years for black women was significantly higher than that for white patients (OR=2.03; CI:(1.468,2.809)). To identify and differentiate attributable risk factors for the racial disparities, we developed a novel mediation analysis method.

A third variable can intervene in the relationship between an explanatory variable and a response variable through many forms. In this paper, we focus on the two forms: mediation or confounding. Although they are conceptually distinctive, MacKinnon et al. [19] claimed that these effects are statistically similar in the sense that all of them measure the change of association between the explanatory and response variables when considering a third variable. Therefore the statistical methods developed for mediation framework can be used for confounding effect analysis, although the scientific interpretations of the analysis might be different. In this paper, we call the effect carried by a third variable indirect effect. We demonstrate the proposed mediation analysis method to differentiate the indirect effects from a wide range of potential mediators/confounders that account for racial disparities in breast cancer survival. Multiple potential third variables including patient residence census tract level and individual level variables are considered. Details of the analysis are in Section 8.1.

General Mediation Analysis

We would like to measure the direct effect of a variable X on the response variable Y and the indirect effect of X on Y through the third variable M. What we call a mediator in this paper should be equivalent to a confounder in confounding analysis. When there is more than one mediator, we should be able to differentiate the indirect effect from each mediator. For these objectives, we first propose general definitions of mediation effects in the counterfactual framework. Figure 1 is the graphical model showing the relationship among $X, Y,$ and M (M_i can be a vector), illustrating the motivating example. The figure could be explained as that the distributions of risk factors (M) are different for different race groups (X), and for different levels of risk factors, the survival rates (Y) vary.

Notations

Let $\mathbf{M} = (M_1, \ldots, M_p)^T$ where M_j be the jth mediator/confounder. $\mathbf{Z}$ is the vector of other independent explanatory variables that directly relate to Y, but does not interact with X. We are interested in exploring the mechanism of the changes in the response, Y, when X is altered. Let $Y(x, \mathbf{m})$ be the potential outcome if the exposure X is at x, and M at m. Denote the domain of X as dom_X. Let $u^\star$ (>0) be the infimum positive unit such that there is a biggest subset of dom_X, denoted as $\text{dom}_X^\star$, in which any x also satisfies $x + u^\star \in \text{dom}_X$, and that $\text{dom}_X = \text{dom}_X^\star \cup \{x + u^\star \mid x \in \text{dom}_X^\star\}$. If $u^\star$ exists, it is unique. Note that if X is continuous,
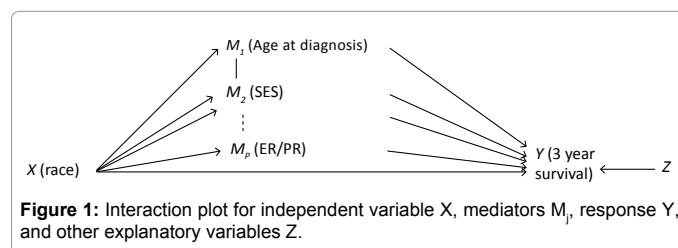

Figure 1: Interaction plot for independent variable X, mediators M_j, response Y, and other explanatory variables Z.

$u^*=0+$ and $\text{dom}_X=\text{dom}_X{}^*$; if X is an ordered binary variable, taking the values 0 or 1, then $u^*=1$ and $\text{dom}_x{}^*=\{0\}$. For random variables A,B and $C, A \perp\!\!\!\perp B \mid C$ denotes that given C, A is conditionally independent of B.

For any function $g(x)$, denote $E_x g = \int_{x \in \text{dom} x} g \cdot f(x) dx$ and $E_{x \mid y} g = \int_{x \in \text{dom} x} g \cdot f(x \mid y) dx$ when x is continuous; and $E_x g = \Sigma_{x \in \text{dom} x} g \cdot f(x)$ and $E_{x \mid y} g = \Sigma_{x \in \text{dom} x} g \cdot f(x \mid y)$ when x is discrete, where $f(x)$ is the density of x and $f(x \mid y)$ is the conditional density of x given y.

Total effect

We define the total effect in terms of the average changing rate of Y with the treatment/exposure, X.

Definition 3.1 Given **Z**, the **total effect** (TE) of X on Y at $X=x^*$ is defined as the change in $E(Y(X,Z))$ when X changes by a u^* unit:

$$TE_{|z}\left(x^*\right) = lim_{u \to u^*} \frac{E(Y \mid \mathbf{Z},(X = x^* + u) - E(Y \mid \mathbf{Z},(X = x^*)}{u}.$$

The average total effect is defined as $ATE_{|z} = E_{x^*}\left[TE_{|z}(x^*)\right]$, where the density of x^* is $f^*(x^*) = \dfrac{f(x^*)}{\int_{x \in \text{dom}_x} f(x) dx}$.

For the identification of total effects, we need two assumptions, which were also stated in VanderWeele and Vansteelandt [14]:

A1 *No-unmeasured-confounder for the exposure-outcome relationship* This assumption can be expressed as $Y(x,\mathbf{m}) \perp\!\!\!\perp X \mid \mathbf{Z}$ for all levels of x and $\mathbf{m}$.

A2 *No-unmeasured-confounder for the mediator-outcome relationship* This assumption can be expressed as $Y(x,\mathbf{m}) \perp\!\!\!\perp \mathbf{M}X \mid \mathbf{Z}$ for all levels of x and $\mathbf{m}$.

With the assumptions, it is easy to see that $TE_{|z}\left(x^*\right) = lim_{u \to u^*} \frac{E(Y \mid \mathbf{Z},(X = x^* + u) - E(Y \mid \mathbf{Z},(X = x^*)}{u}.$ Note that if X is binary, $\text{dom}_x{}^*=\{0\}$, so the (average) total effect is $E(Y|Z,X=1)-E(Y|Z,X=0)$, which is equivalent to the commonly used definition of total effect in literature for binary outcomes.

Compared with conventional definitions of the average total effect that look at the differences in expected Y when X changes from x to x^*, we define total effect based on the rate of change. The motivation is that the effect will not change with either the unit or the changing unit (x^*-x) of X, thus generalizing the definitions of mediation effects, which will be consistent for exposure variables measured at any scale (binary, multi-categorical, or continuous). The benefits of the modification will be further discussed in Section 4 after the average direct and indirect effects are similarly defined.

Direct and indirect effects

We define the direct effect by fixing M at its marginal distribution, $f(M|Z)$, while X shifts. The direct effect of X at x^* is defined as

$$DE_{|z}\left(x^*\right) = \mathrm{E}_m\left[lim_{u \to u^*} \frac{E(Y(x^* + u, m) - Y(x^*, \mathrm{m}) \mid Z)}{u}\right].$$ When A1 and A2 are satisfied, $DE_{|z}\left(x^*\right) = \mathrm{E}_m\left[lim_{u \to u^*} \frac{E(Y \mid X = x^* + u, \mathrm{M} = m, \mathbf{Z}) - E(Y \mid X = x^*, \mathrm{M} = m, \mathbf{Z})}{u}\right]$,

where m takes the values in the support set $\{m \mid prob(m \mid X=x^*)>0$ & $prob(m \mid X=x^* + u^*)>0\}$. The average direct effect is defined as $\mathrm{E}_{x^*}[DE_{|Z}(x^*)]$. In the same vein, we extend the definition to multiple mediators.

Definition 3.2 For given Z, the direct effect (DE) of X on Y not from M_j is defined as

$H_2SO_4 \leftrightarrow 2H^+ + SO_4$ and the average direct effect of X on Y not from M_j is $ADE_{\backslash Mj|\mathbf{Z}}=E_{x^*}DE_{\backslash Mj|\mathbf{Z}}(x^*)$, where $\mathbf{M}_{-j}$ denotes the vector $\mathbf{M}$ without M_j.

For the identifiability of the DE, an additional assumption is needed:

A3 M_j is not causally prior to $\mathbf{M}_{-j}$.

Note that if there are overlapping pathways through multiple mediators, A3 is violated. We recommend using the sensitivity analysis proposed by Imai and Yamanoto [20] to assess the robustness of empirical results. If causal relationship among mediators exists, it is necessary to combine those mediators. For example, if the real relationship is $X \to M_{t1} \to M_{t2} \to Y$, we should combine the effects from M_{t1} and M_{t2}, and consider the pathway $X \to \begin{pmatrix} M_{t1} \\ M_{t2} \end{pmatrix} \to Y$ instead, where the mediator is a vector. The within $\begin{pmatrix} M_{t1} \\ M_{t2} \end{pmatrix}$ effects can be explored separately. With A3, we formally calculate the average direct effect from exposure to outcome not through M_j as

$$\mathrm{E}_{x^*}\mathrm{E}_{mj}\left[\lim_{u \to u^*} \frac{E_{\mathrm{M}_{-j}|X=x^*+u}[E(Y \mid \mathbf{Z}, M_j = m_j, \mathrm{M}_{-j}(X = x^* + u)] - E_{\mathrm{M}_{-j}|X=x^*}[E(Y \mid \mathbf{Z}, M_j = m_j, \mathrm{M}_{-j}(X = x^*)]}{u}\right].$$

It is easy to show that when X is binary, $ADE_{\backslash Mj|\mathbf{Z}}=DE_{\backslash Mj|\mathbf{Z}}(0)=$

$$E_{mj}\left\{E_{\mathrm{M}_{-j}|X=1}[E(Y \mid \mathbf{Z}, M_j = m_j, \mathrm{M}_{-j} X = 1)] - E_{\mathrm{M}_{-j}|X=0}[E(Y \mid \mathbf{Z}, M_j = m_j, \mathrm{M}_{-j}(X = 0)]\right\}$$

Mathematically, the definitions of TE and DE differ in that for given Z, the former takes the conditional expectation over M_j for given x^*while the latter takes the marginal expectation. The direct effect measures the changing rate in the potential outcome with X, where Mj is fixed at its marginal level, while all the other mediators can change with X. As in Figure 1, DE not from Mj is calculated as TE but the line between X and M_j is broken. Therefore, we call the changing rate the direct effect of X on Y not from M_j, which is the summation of the direct effect of X on Y and the indirect effects through mediators other than M_j. With definitions 3.1 and 3.2, the definition of indirect effect is straightforward.

Definition 3.3 Given Z, the indirect effect of X on Y through M_j is defined as $IE_{M_j|\mathbf{Z}}(x^*) = TE_{|\mathbf{Z}}(x^*) - DE_{\backslash M_j|\mathbf{Z}}(x^*)$. Similarly, the average indirect effect through M_j is $AIE_{M_j|\mathbf{Z}} = ATE_{|\mathbf{Z}} - DE_{\backslash M_j|\mathbf{Z}}$.

Finally, the average direct effect of X not from **M** is $ADE_{|\mathbf{Z}} = ADE_{|\mathbf{Z}} - \sum_{j=1}^{p} AIE_{m_j|z}$. If we can assume that M includes all potential mediators, ADE|Z is the average direct effect of X on Y. Note that sometimes the direct effect $DE_{|\mathbf{Z}}(x) = TE_{|\mathbf{Z}}(x) - \sum_{j=1}^{p} IE_{m_j|z}(x)$ is more relevant than ADE when the direct effects from exposure are not constant at different exposure levels.

Based on the definitions, mediation analyses can be generalized to different types of exposure, outcome or mediator variables, whose relationships can be modeled by different predictive models. Moreover, indirect effect from individual mediator can be differentiated from multiple mediators.

Multiple Mediation Analysis in Linear Regressions

In this section, we show that for continuous mediators and outcomes that are modeled with linear regressions, the proposed average indirect effects are identical to those measured by the CD and CP methods. We also show that the assumption of none exposure-mediator interaction

for controlled direct effect is not required for the measurements of mediation effects defined in this paper. For simplicity, we ignore the independent variable(s) Z for the following discussion. Without loss of generality, assume that there are two mediators and that the true relationships among variables are:

$$M1_i = a_{01} + a_1 X_i + \varepsilon_{1i}; \tag{1}$$

$$M_{2i} = a_{02} + a_2 X_i + \varepsilon_{2i}; \tag{2}$$

$$Y_i = b_0 + b_1 M_{1i} + b_2 M_{2i} + cX_i + \varepsilon_{3i}; \tag{3}$$

where $(\varepsilon_{1i}, \varepsilon_{2i})$ have identical and independent (iid) bivariate normal distributions, $\begin{pmatrix} \varepsilon_{1i} \\ \varepsilon_{2i} \end{pmatrix} \sim N_2 \left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix} \right)$, and are independent with ε_3, which are iid $N(0.\sigma_3^2)$, for $i=1,\ldots,n$. Following Definitions 3.1 and 3.3, we have Lemma 4.1 (proof in supplementary materials) for the measurements of direct and indirect effects.

Lemma 4.1 The average total effect of X on Y is $a_1b_1+a_2b_2+c$; the average indirect effect through M_1 is a_1b_1; and that through M_2 is a_2b_2. Of all the effect from X to Y, a $\dfrac{a_1b_1}{a_1b_1+a_2b_2+c}$ fraction is indirectly from M_1; $\dfrac{a_2b_2}{a_1b_1+a_2b_2+c}$ fraction is from M_2; and $\dfrac{c}{a_1b_1+a_2b_2+c}$ fraction is directly from X.

These measurements of average indirect effects are identical to those from the CD and CP methods in linear regressions. Note that correlations among mediators are allowed here. Moreover, compared with the controlled direct effects, we do not require the none exposure mediator interaction assumption. To illustrate this, assume that there is an interaction effect of M_1 and X on Y, so that equation (3) should be $Y_i = b_0 + b_1 M_{1i} + b_2 M_{2i} + cX_i + dX_i M_{1i} + \varepsilon_{3i}$. Based on the models, we have the following lemma for the mediation effects.

Lemma 4.2 The total effect of X on Y at X=x is $a_1b_1 + a_2b_2 + c + a_{01}d + 2a_1dx$, among which the indirect effect through M_1 is $a_1b_1 + a_1dx$; and that through M_2 is a_2b_2. The direct effect from X is $c + a_{01}d + a_1dx$.

When X has no effect on $M_1(a_1=0)$ and consequently there is no indirect effect from M1, the indirect effect from M1 defined by Definition 3.3 will be 0 by Lemma 4.2. However, the total effect minus the controlled direct effect may be non-zero because of the exposure mediator interactions [14]. Similar to our results, the natural direct effect from X and the natural indirect effect of M_1 when X changes from x to $x^\star$ defined by VanderWeele and Vansteelandt [14] are $(c+a_{01}d+a_1da^\star)$ $(x-x^\star)$ and $(a_1b_1+a_1da)(x-x^\star)$, both of which depend on the changing unit of X, $x-x^\star$. However, the natural direct effect depends further on the end value of X, $x^\star$, while the natural indirect effect depends further on the start value x.

Multiple Mediation Analysis in Logistic Regression

In this section, we illustrate the mediation analysis when the mediator variable is not continuous and when the response variable is binary and a logistic regression is adapted to model the relationship with X and M. In this case, the response is the log-odds of the outcome variable, e.g., $Y_i = \log \dfrac{\text{Pr(patient } i \text{ dies within 3 years)}}{Pr(\text{patient } i \text{ is alive at the end of the 3rd year})}$ in the motivating example. The independent variable X can be continuous or categorical. It is the patient race (white or black) in the motivating example. In the following, we assume the exposure variable X to be binary.

When M is binary

We first consider mediator M_1 to be binary and assume the underlying true models:

$$logit\left(Pr\left(M_{1i} = 1\right)\right) = a_{01} + a_1 X_i;$$
$$\vdots$$
$$Y_i = b_0 + b_1 M_{1i} + \sum_{j=2}^{p} b_j M_{ji} + cX_i$$

In this situation, the CP method cannot be used directly since a_1 denotes the change of $logit(M_{1i}=1)$ with X_i, while b_1 is the change of Y_i with M_{1i}. The CD method is also not readily adaptable, since first we have to assume two true models for Y, one fitted with X and mediators and the other with X only; second, the coefficients of X in two models fitted with different subsets of explanatory variables have different scales, and consequently are not comparable; third, the indirect effects from different mediators are not separable. With the definitions in Section 3, we derive the neat results in Lemma 5.1 to calculate the indirect effect of M_1.

Lemma 5.1 *Using the above notations, the indirect effect from X to Y through M_1 is*

$$\mu_{jkl1} = average\{f(X=0, M_j = k, \bar{M}_{-j,i,l}), \text{for } i = 1,\ldots, \left[\frac{n}{2}\right] \text{ and } \tilde{M}_{j,i,l} = K\}$$

Note that the indirect effect from M_1 can be separated into two parts: b_1, representing the effect of M_1 on Y, and $\left(\dfrac{e^{a_{01+a_1}}}{1+e^{a_{01+a_1}}} - \dfrac{e^{a_{01}}}{1+e^{a_{01}}} \right)$, denoting the difference in the prevalence of $M_1=1$ when X changes from 0 to 1.

The counterfactual framework is popular in dealing with the special case when X is binary (0 or 1 denoting control or treatment), for example, the natural indirect effect, $\delta_i(x)$, and the natural direct effect, $\varsigma_i(x)$, discussed in Section 1. Based on the definition, Imai et al. [20,21] defined the average causal indirect effect (ACIE) as $\bar{\delta}(x) = E(\delta_i(x))$ and the direct effect as $\bar{\zeta}(x) = E(\zeta_i(x))$. Their methods have to make the assumption that there is no interaction between X and the indirect effect, i.e., $\bar{\delta}(0) = \bar{\delta}(1)$ or $\bar{\zeta}(0) = \bar{\zeta}(1)$. Otherwise, there could be two measurements of indirect effect or direct effect, which brings in challenges to generalizing the mediation analysis to multi-categorical or continuous exposures. Our method relaxes this assumption. One can easily show that the average direct effect we defined for binary X in single mediator scenario is $P(X=0) \cdot \bar{\zeta}(0) + P(X=1) \cdot \bar{\zeta}(1)$.

When M is multi-categorical

When the mediator is multi-categorical with $K+1$ distinct categories, i.e., M takes one of the values $0,1,\cdots,K$, multinomial logit regression model can be adapted to fit the relationship between M and X. The true models are assumed to be:

$$log \frac{Pr(M_{1i}=1)}{Pr(M_{1i}=0)} = a_{01} + a_1 X_i;$$
$$\vdots$$
$$log \frac{Pr(M_{1i}=K)}{Pr(M_{1i}=0)} = a_{0K} + a_K X_i;$$
$$\vdots$$
$$Y_i = b_0 + \sum_{k=1}^{K} b_{1k} I(M_{1i}=k) + \sum_{j=2}^{p} b_j M_{ji} + cX_i;$$

where $I(M=m)$ is 1 if M=m, and 0 otherwise.

Lemma 5.2 *With the above assumptions and notations, the indirect effect from X to Y through M_1 is*

$$\sum_{m=1}^{k} b_{1m} \left(\frac{e^{a_{0m}+a_m}}{1+\sum_{k=1}^{K} e^{a_{0k}+a_k}} - \frac{e^{a_{0m}}}{1+\sum_{k=1}^{K} e^{a_{0k}}} \right).$$

Multiple Mediation Analysis for Non/Semi-Parametric Predictive Models with Binary Exposure

When (generalized) linear regression is insufficient in describing variable relationships, mediation analysis can be very difficult. The following algorithms that derived directly from the definitions of mediation effects provide a non/semi-parametric method to calculate mediation effects when the exposure variable is binary and the sample size at each exposure level is large. More general mediation analysis with any types of exposures and nonlinear relationships is discussed in a separate paper [22].

Algorithm 6.1 Estimate the total effect: the total effect for binary X is $E(Y|X=1)-E(Y|X=0)$. Under certain conditions, it can be directly obtained by averaging the response variable Y in subgroups of $X=0$ and 1 separately and taking the difference.

Algorithm 6.2 Estimate the direct effect not through M_j, which is defined as

$$E_{m_j}\{E_{M_{-j}|X=1}[E(Y|\mathbf{Z},M_j=m_j,M_{-j},X=1)]-E_{M_{-j}|X=0}[E(Y|\mathbf{Z},M_j=m_j,M_{-j},X=0)]\}:$$

$$\log \frac{Pr(M_{1i}=1)}{Pr(M_{1i}=0)} = a_{01} + a_1 X_i;$$
$$\vdots$$

1 Divide the original data sets
$$\log \frac{Pr(M_{1i}=K)}{Pr(M_{1i}=0)} = a_{0K} + a_K X_i;$$
$$\vdots$$

$$Y_i = b_0 + \sum_{k=1}^{K} b_{1k} I(M_{1i}=k) + \sum_{j=2}^{p} b_j M_{ji} + c X_i;$$

and $D_j = \{M_{j,i}\}_{i=1}^{n}$ according to $X=0$ *or* 1, denote the separated data sets as $D_{0,-j}, D_{1,-j}, D_{0,j}, D_{1,j}$.

2 Fit predictive model of Y on X and M on the whole observation set, denote the pre-dictive model as $E(Y)=f(X,M_j,M_{-j})$.

3 For $l=1,…,N$, where N is the number of iteration:

(a) Sample $\left[\frac{n}{2}\right] M_j s$ with replacement from $D_{0,j}$ and $D_{1,j}$ separately, randomly mix the two sets of samples to form a sample of M_j from its marginal distribution, which is denoted as $\{\tilde{M}_{jil}\}_{i=1}^{n}$.

(b) Take $\left[\frac{n}{2}\right]$ samples with replacement from $D_{0,-j}$ and denote it as $\breve{D}_{0,-j,l} = \{\breve{M}_{-j,i,l}\}_{i=1}^{\left[\frac{n}{2}\right]}$. Take $\left[\frac{n}{2}\right]$ samples with replacement from $D_{1,-j}$ and denote it as $\breve{D}_{1,-j,l} = \{\breve{M}_{-j,i,l}\}_{i=\left[\frac{n}{2}\right]+1}^{n}$. $\breve{D}_{0,-j,l}$ and $\breve{D}_{1,-j,l}$ are joint samples of M_{-j} from their conditional distributions on $X=0$ or 1 respectively.

(c) If M_j is categorical, taking $K+1$ potential groups, let

$$\mu_{jkl1} = average\{f(X=1,M_j=k,\breve{M}_{-j,i,l}), \text{for } i=\left[\frac{n}{2}\right]+1,...,n, \text{and } \tilde{M}_{j,i,l}=k\}$$

and $\mu_{jkl1} = average\{f(X=1,M_j=k,\breve{M}_{-j,i,l}), \text{for } i=\left[\frac{n}{2}\right]+1,...,n, \text{and } \tilde{M}_{j,i,l}=k\}$, where $k=0,…,K$. Let $DE_{\backslash M_j,l} = \sum_{k=0}^{K} Pr(\tilde{M}_{jl}=k) \times (\mu_{jkl1} - \mu_{jkl0})$, where $Pr(\tilde{M}_{jl}=k)$ is the probability of $M_j=$k in the sample $\{\tilde{M}_{jil}\}_{i=1}^{n}$.

(d) If M_j is continuous, let

$$DE_{\backslash M_j,l} = \frac{1}{n-\left[\frac{n}{2}\right]} \sum_{i=\left[\frac{n}{2}\right]+1}^{n} f(1,\tilde{M}_{jil},\breve{M}_{-j,i,l}) - \frac{1}{\left[\frac{n}{2}\right]} \sum_{i=1}^{\left[\frac{n}{2}\right]} f(0,\tilde{M}_{jil},\breve{M}_{-j,i,l})$$

4 The direct effect not from M_j is estimated as $\frac{1}{N} \sum_{l=1}^{N} DE_{\backslash M_j,l}$.

The model f fitted in step 2 can be parametric or nonparametric. In Section 8.1, we illustrate the method with both logistic model and multivariate additive regression trees (MART).

Inferences on the Estimated Mediation Effects

Delta and bootstrap are two popular methods to measure the uncertainties of mediation estimators [23,24]. When the variable relationships are modeled with (generalized) linear models. Delta method can be used to estimate the variances of the mediation effect estimators. In the supplementary materials, we provide the Delta method estimated variances of the mediation effects estimated in Lemmas 4.1, 5.1 and 5.2. The variances of mediation effect estimators can also be obtained by bootstrap method, especially for nonlinear predictive models. First randomly draw n observations from the original data set with replacements, and then conduct mediation analysis on the bootstrap sample and obtain the direct, indirect and total effect estimates. The resampling and analysis are repeated B times. The variances of the mediation effects are calculated based on the B sets of mediation effect estimators from the bootstrap samples. The $OR = \frac{Pr(M_1=1,M_2=1)Pr(M_1=0,M_2=0)}{Pr(M_1=0,M_2=1)Pr(M_1=1,M_2=0)}$ and $\left(1-\frac{\alpha}{2}\right)$th percentiles of the bootstrap estimates construct the 95% confidence interval of the mediation effects.

Real Example and Simulation Study

The motivating example

In this example the explanatory variable is the race of patient (0 for white and 1 for black). The binary response variable is patient's vital status at the end of the third year of diagnosis (death from breast cancer or not). The potential mediators (confounders) are listed in Table 1. Among these variables, age is continuous; poverty, education, employment, radiation, chemotherapy and hormonal therapy variables are binary; and all other variables are polytomous.

We used the criteria proposed by Baron and Kenny [2] to check the qualification of potential mediators. Those variables significantly associated with race, and with vital status controlling for race were candidate mediators. A variable not related with race but significantly related with vital status adjusted for race was included as a covariate. The variables not significantly related to vital status controlling for race were excluded for further analysis. As a result, radiation was included as a covariate, and chemotherapy and all census tract level variables were left out for further analysis. The remaining variables were analyzed as potential mediators in multiple mediation analysis. We used both the logistic model and MART [25] to explore the mediation effects. Note that the IEs were measured in terms of log odds of death in logistic model, but in terms of probability of death in MART. To compare results, we define the relative indirect effect: $RE = \frac{IE}{TE}$. Table 2

Variables	Variable Description
1. Census tract variables	
Poverty	≥ 20% vs < 20% of households with income below the federal poverty level
Education	≥ 25% vs < 25% of adults with less than a high school education
Residence area	grouped using Beale codes: 100% rural; urban-rural mix; 100% urban
Employment	≥ 66% versus < 66% of population aged 16 and over who are unemployed
2. Demographic variables	
Insurance	no insurance; Medicaid; Medicare and public; private insurance
Marital status	single-never married; married; separated; widowed; divorced; unknown
Age	age at diagnosis
3. Tumor characteristics	
Stage	regional; distant; localized
Grade	moderately differentiate; poorly/un-differentiate; well differentiate; unknown
Tumor size	< 1 cm; 1:1 ~ 2 cm; 2:1 ~ 3 cm; > 3 cm; unknown
Comorbidity	mild; moderate; severe; none; unknown
4. Treatment indicators	
Surgery	mastectomy, lumpectomy, no surgery
Radiation	not administered; administered
Chemotherapy	not administered, administered
Hormonal therapy	not administered; administered
ER/PR receptor	either is positive; both are negative; unknown

Table 1: Potential Mediators in the Motivating Example.

Mediator	Logistic Regression		Nonparametric Method	
	IE (95% CI)	RE (95% CI)	IE (95% CI)	RE (95% CI)
Stage	0.276 (0.127,0.488)	28.14 (12.2,76.7)	0.023 (0.008,0.038)	28.78 (9.6,48.0)
Insurance	0.275 (0.074,0.430)	28.05 (5.8,70.8)	0.006 (0.004,0.008)	7.84 (5.6,10.1)
ER/PR	0.181 (0.068,0.332)	18.46 (5.5,56.0)	0.015 (0.009,0.020)	18.53 (11.8,25.3)
Grade	0.158 (0.027,0.379)	16.09 (2.4,47.6)	0.008 (0.005,0.011)	10.34 (6.23,14.5)
Surgery	0.145 (0.067,0.333)	14.75 (6.3,50.4)	0.016 (0.008,0.023)	19.87 (10.4,29.3)
Tumor Size	0.135 (0.025,0.417)	13.77 (2.2,56.1)	0.016 (0.011,0.022)	20.88 (14.3,27.5)
Hormonal Therapy	0.114 (0.02,0.253)	11.57 (1.6,40.2)	0.008 (0.004,0.011)	9.75 (5.6,13.9)
Age	-0.113 (-0.301,-0.034)	-11.5 (-43.2,-3.1)	-0.005 (-0.011,0.001)	-6.04 (-13.6,1.6)
Marital Status	0.001 (-0.001,0.004)	1.65 (-1.27,4.57)	0.001 (-0.001,0.004)	1.65 (-1.3,4.6)
Comorbidity	-0.002 (-0.150,0.109)	-0.17 (-20.5,13.6)	0.002 (0.001,0.003)	3.06 (1.9,4.2)

[1] After considering all indirect effects through mediators, direct effect of race on mortality was -0.115 with 95% CI(-0.713,0.526) in logistic model and was -0.014(0.034,0.007).
[2] 95% confidence interval is 0.025 and 0.975 percentiles of the distribution of statistics obtained by bootstrap with 1000 repetitions.

Table 2: Indirect Effects (IE) and Relative Effects (RE).

presents the indirect effect, relative indirect effect estimates, and their 95% confidence intervals from the bootstrap method. There were some differences in the results from logistic regression and from MART: 1. the orders of relative effects were slightly different; 2. comorbidity was a significant mediator by MART but not by logistic regression; and 3. the confidence intervals were narrower using MART. We recommend using MART in this case, since it is more sensitive in finding significant mediators especially if the assumed relationship in logistic regression is inappropriate. From the results, stage explained 29% of racial disparities in the three-year survival breast cancer patients. Compared with localized breast cancer, patients diagnosed with regional or distant cancer were significantly more likely to die from breast cancer. Also, the larger tumor size and/or worse grade of breast tumor contributed to the higher risk of breast cancer death among black women than white women; the relative indirect effects from tumor size and grade were 21% and 10%, respectively.

Compared with patients with negative ER/PR receptors, patients with positive ER/PR receptors developed less aggressive breast cancer and had better three-year survival rates. Black women were 54:7% less likely to be diagnosed with positive ER/PR receptors than whites. Compared with patients with lumpectomy surgery or hormonal therapy after surgery, those without surgery or hormonal therapy had worse three-year survival. Black women were less likely to undergo hormonal therapy or lumpectomy surgery. As results, the relative indirect effects for hormonal receptor, surgery, and hormonal therapy were 10%, 20%, and 10%, respectively.

Black breast cancer patients were more likely to be diagnosed at younger age at which the patients had longer survival time. This fact indicated age at diagnosis was a suppression factor for racial disparities in mortality.

Insurance had a significant indirect effect (8%) on racial disparities in the risk of death from breast cancer. Compared with patients with private insurance, those having no insurance or having Medicaid, Medicare or other public insurance had higher three-year mortality, which might relate to the more restricted accessibility to necessary breast cancer treatment. A higher proportion of blacks than whites in this study had Medicaid, Medicare, public or no insurance.

The estimated average direct effect of race on mortality after considering all mediators was -.014 with 95% confidence interval (-.034; .007), which was statistically insignificant. This suggests that racial disparities of breast cancer survival could be satisfactorily explained by all mediators in the model.

Simulation study

To evaluate the sensitivity and identifiability of the proposed method, we conducted simulations to check the bias, type I error, and power in estimating mediation effects. We checked the method comprehensively with all types of exposures, mediators and response variables, and with many types of relationships. In this section, we show a special case where all variables were binary. For more scenarios, readers are referred to Fan [22]. We used two potential mediators, which were independent or correlated with each other. Data were generated from the following true models:

$$logit(Pr(M_1=1))=a_1X$$

$$logit(Pr(M_2=1))=a_2X$$

$$logit(Pr(Y=1))=cX+b_1M_1+b_2M_2$$

The correlation between M_1 and M_2 was controlled by the odds ratio

$$OR=\frac{Pr(M_1=1,M_2=1)Pr(M_1=0,M_2=0)}{Pr(M_1=0,M_2=1)Pr(M_1=1,M_2=0)}$$ [26]. Four different sample sizes of

the simulation (200, 500, 1000 and 1500) were applied. Parameters a_1 and a_2 were chosen from 0, 0.518 or 2.150, which indicated different proportions of variances in M that were explained by X. Without loss of generality, we made $a_1 \leq a_2$. Thus there were six different combinations of (a_1, a_2). Parameters c and OR took values from the sets (0, 0.518, 2.150) and (0.2, 1, 5), respectively. The values of bi (i=1,2) depended on a_i (i=1, 2). When a_i=0, a_i=0:518 or a_i=2:150, b_i took values from the sets (0, 0.518, 2.150), (0, 0.522, 2.168) or (0, 0.564, 2.341), respectively. Larger a_j and b_j indicated greater indirect effect from M_j. When either of them was zero, there was no mediation effect from the corresponding variable. A total of 6×3⁴=486 parameter combinations were used for each of the four sample sizes, yielding 1944 simulation scenarios. We present the simulation results for a_1= 0:518; a_2=0:518 and sample sizes 500 and 1000 in this paper. The complete simulation results will be provided upon request.

Empirical bias

For each simulation scenario, 500 replications were conducted. We estimated the IEs of M_1 and M_2 from these 500 replicates. The empirical bias is the difference between the averaged IE and true IE. The simulation results are summarized in Table 3. For all the scenarios, we found no empirical bias that was significantly different from 0.

Type I error rate and power

For each replication described in the previous section, the variance of the estimated IE was estimated from both the Delta method and the bootstrap method (B=500). We tested the hypothesis: H_0: IE=0 via the

test statistics $\frac{IE}{SE(I|E)}$, which was assumed to be normally distributed.

The type I error rate or power was the proportion of times rejecting the null hypothesis at 5% significance level. The simulation results were summarized in Table 4. If the true indirect effect is 0, the values listed in the table represent the type I error rates, otherwise the powers. Figure 2 presents the power curves obtained by the Delta and bootstrap methods (OR=0.2) for mediator M_1 at different b_1 and sample sizes when b_2 and c were fixed at 0.522 and 0.518, respectively. The Delta and bootstrap methods showed similar patterns. Figure 2 suggests that for all sample sizes, the statistical power increased with b_1 and then reached a plateau when b_1 hit a certain point. At fixed b_1, larger sample size had greater power. For a given sample size, when b_2 and c were fixed, the statistical

Parameters			TRUE	N=500			N=1000		
b1	b2	c	AIE	OR=0.2	1	5	OR=0.2	1	5
0	0	0	0	.008 (.006)	.016 (.012)	.008 (.006)	.036 (.028)	.028 (.022)	.022 (.018)
0	0	0.518	0	.012 (.01)	.016 (.012)	.012 (.008)	.026 (.024)	.02 (.018)	.014 (.01)
0	0	2.15	0	.006 (.008)	.02 (.01)	.01 (.008)	.026 (.022)	.026 (.024)	.028 (.022)
0	0.522	0	0	.004 (.002)	.01 (.01)	.014 (.006)	.028 (.026)	.026 (.022)	.02 (.012)
0	0.522	0.518	0	.01 (.006)	.008 (.008)	.014 (.006)	.034 (.03)	.028 (.03)	.026 (.022)
0	0.522	2.15	0	.012 (.006)	.012 (.006)	.01 (.0014)	.034 (.028)	.028 (.028)	.03 (.032)
0	2.168	0	0	.01 (.006)	.01 (.004)	.004 (.004)	.024 (.02)	.024 (.022)	.026 (.018)
0	2.168	0.518	0	.01 (.006)	.01 (.002)	.004 (.002)	.036 (.03)	.028 (.022)	.034 (.028)
0	2.168	2.15	0	.004 (.004)	.012 (.008)	.006 (.006)	.026 (.02)	.02 (.024)	.03 (.026)
0.522	0	0	0.0662	.416 (.356)	.444 (.388)	.374 (.31)	.87 (.85)	.904 (.876)	.888 (.87)
0.522	0	0.518	0.0662	.388 (.334)	.432 (.36)	.378 (.33)	.848 (.828)	.89 (.75)	.868 (.854)
0.522	0	2.15	0.0662	.274 (.214)	.29 (.252)	.272 (.22)	.718 (.696)	.776 (.874)	.712 (.678)
0.522	0.522	0	0.0662	.358 (.304)	.39 (.34)	.366 (.302)	.846 (.792)	.896 (.846)	.876 (.862)
0.522	0.522	0.518	0.0662	.32 (.256)	.368 (.332)	.336 (.284)	.818 (.656)	.866 (.72)	.85 (.814)
0.522	0.522	2.15	0.0662	.208 (.17)	.264 (.218)	.242 (.19)	.69 (.656)	.742 (.72)	.686 (.648)
0.522	2.168	0	0.0662	.204 (.164)	.26 (.204)	.232 (.194)	.668 (.618)	.766 (.722)	.718 (.676)
0.522	2.168	0.518	0.0662	.196 (.168)	.252 (.204)	.224 (.176)	.6 (.556)	.704 (.656)	.634 (.59)
0.522	2.168	2.15	0.0662	.136 (.106)	.174 (.138)	.162 (.106)	.48 (.458)	.558 (.512)	.516 (.456)
2.168	0	0	0.2747	.804 (.794)	.818 (.812)	.796 (.788)	.972 (.974)	.98 (.974)	.978 (.978)
2.168	0	0.518	0.2747	.802 (.794)	.816 (.804)	.796 (.78)	.97 (.974)	.98 (.974)	.978 (.978)
2.168	0	2.15	0.2747	.8 (.772)	.812 (.786)	.788 (.748)	.97 (.968)	.98 (.97)	.978 (.976)
2.168	0.522	0	0.2747	.802 (.794)	.816 (.806)	.796 (.778)	.97 (.972)	.98 (.974)	.978 (.978)
2.168	0.522	0.518	0.2747	.8 (.786)	.814 (.8)	.796 (.764)	.97 (.974)	.98 (.972)	.978 (.978)
2.168	0.522	2.15	0.2747	.796 (.742)	.81 (.738)	.782 (.66)	.968 (.968)	.978 (.97)	.978 (.974)
2.168	2.168	0	0.2747	.798 (.772)	.812 (.78)	.786 (.672)	.97 (.97)	.98 (.97)	.978 (.976)
2.168	2.168	0.518	0.2747	.792 (.756)	.806 (.736)	.766 (.524)	.968 (.968)	.978 (.97)	.976 (.972)
2.168	2.168	2.15	0.2747	.774 (.642)	.786 (.496)	.744 (.242)	.966 (.966)	.974 (.968)	.976 (.93)

Table 3: Empirical Bias (Standard Error) of Estimated AIE$_{M1}$.

Parameters			TRUE	N=500			N=1000		
b1	b2	c	AIE	OR=0.2	1	5	OR=0.2	1	5
0	0	0	0	.008 (.006)	.016 (.012)	.008 (.006)	.036 (.028)	.028 (.022)	.022 (.018)
0	0	0.518	0	.012 (.01)	.016 (.012)	.012 (.008)	.026 (.024)	.02 (.018)	.014 (.01)
0	0	2.15	0	.006 (.008)	.02 (.01)	.01 (.008)	.026 (.022)	.026 (.024)	.028 (.022)
0	0.522	0	0	.004 (.002)	.01 (.01)	.014 (.006)	.028 (.026)	.026 (.022)	.02 (.012)
0	0.522	0.518	0	.01 (.006)	.008 (.008)	.014 (.006)	.034 (.03)	.028 (.03)	.026 (.022)
0	0.522	2.15	0	.012 (.006)	.012 (.006)	.01 (.0014)	.034 (.028)	.028 (.028)	.03 (.032)
0	2.168	0	0	.01 (.006)	.01 (.004)	.004 (.004)	.024 (.02)	.024 (.022)	.026 (.018)
0	2.168	0.518	0	.01 (.006)	.01 (.002)	.004 (.002)	.036 (.03)	.028 (.022)	.034 (.028)
0	2.168	2.15	0	.004 (.004)	.012 (.008)	.006 (.006)	.026 (.02)	.02 (.024)	.03 (.026)
0.522	0	0	0.0662	.416 (.356)	.444 (.388)	.374 (.31)	.87 (.85)	.904 (.876)	.888 (.87)
0.522	0	0.518	0.0662	.388 (.334)	.432 (.36)	.378 (.33)	.848 (.828)	.89 (.75)	.868 (.854)
0.522	0	2.15	0.0662	.274 (.214)	.29 (.252)	.272 (.22)	.718 (.696)	.776 (.874)	.712 (.678)
0.522	0.522	0	0.0662	.358 (.304)	.39 (.34)	.366 (.302)	.846 (.792)	.896 (.846)	.876 (.862)
0.522	0.522	0.518	0.0662	.32 (.256)	.368 (.332)	.336 (.284)	.818 (.656)	.866 (.72)	.85 (.814)
0.522	0.522	2.15	0.0662	.208 (.17)	.264 (.218)	.242 (.19)	.69 (.656)	.742 (.72)	.686 (.648)
0.522	2.168	0	0.0662	.204 (.164)	.26 (.204)	.232 (.194)	.668 (.618)	.766 (.722)	.718 (.676)
0.522	2.168	0.518	0.0662	.196 (.168)	.252 (.204)	.224 (.176)	.6 (.556)	.704 (.656)	.634 (.59)
0.522	2.168	2.15	0.0662	.136 (.106)	.174 (.138)	.162 (.106)	.48 (.458)	.558 (.512)	.516 (.456)
2.168	0	0	0.2747	.804 (.794)	.818 (.812)	.796 (.788)	.972 (.974)	.98 (.974)	.978 (.978)
2.168	0	0.518	0.2747	.802 (.794)	.816 (.804)	.796 (.78)	.97 (.974)	.98 (.974)	.978 (.978)
2.168	0	2.15	0.2747	.8 (.772)	.812 (.786)	.788 (.748)	.97 (.968)	.98 (.97)	.978 (.976)
2.168	0.522	0	0.2747	.802 (.794)	.816 (.806)	.796 (.778)	.97 (.972)	.98 (.974)	.978 (.978)
2.168	0.522	0.518	0.2747	.8 (.786)	.814 (.8)	.796 (.764)	.97 (.974)	.98 (.972)	.978 (.978)
2.168	0.522	2.15	0.2747	.796 (.742)	.81 (.738)	.782 (.66)	.968 (.968)	.978 (.97)	.978 (.974)
2.168	2.168	0	0.2747	.798 (.772)	.812 (.78)	.786 (.672)	.97 (.97)	.98 (.97)	.978 (.976)
2.168	2.168	0.518	0.2747	.792 (.756)	.806 (.736)	.766 (.524)	.968 (.968)	.978 (.97)	.976 (.972)
2.168	2.168	2.15	0.2747	.774 (.642)	.786 (.496)	.744 (.242)	.966 (.966)	.974 (.968)	.976 (.93)

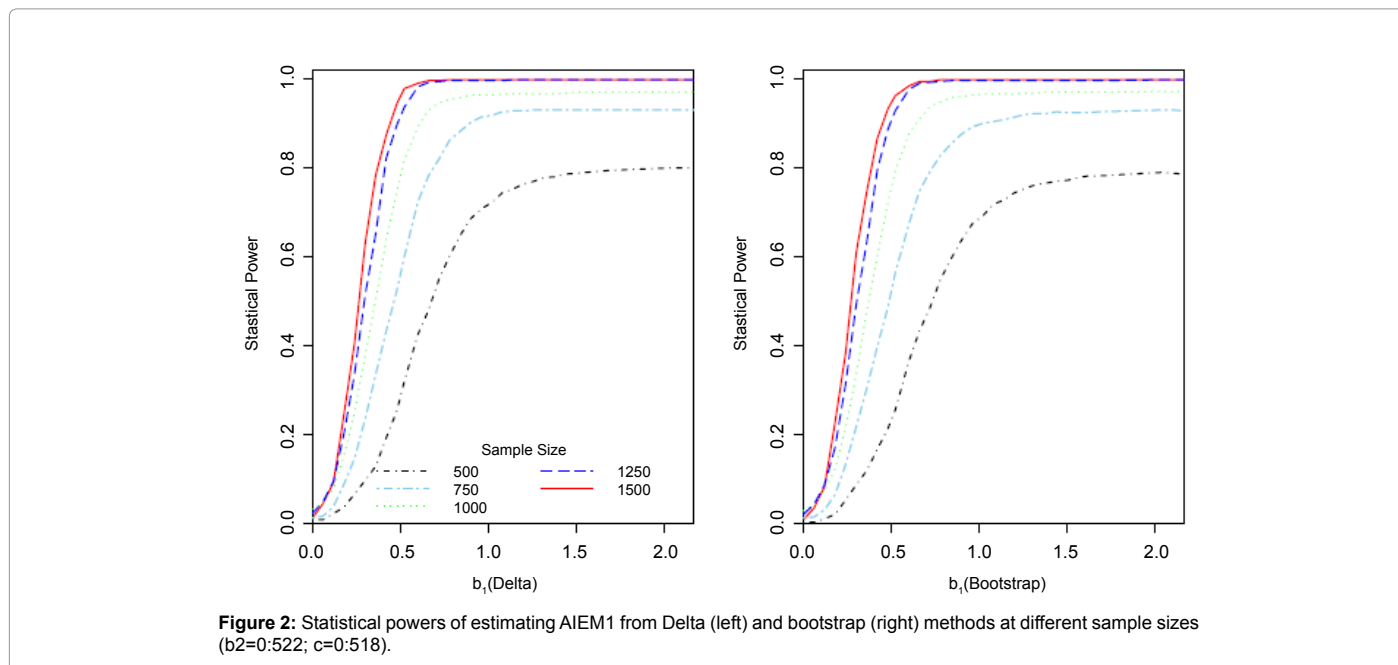**Table 4:** Type I Error Rates and Power of Estimated AIEM1 by Delta (Bootstrap) Method.

Figure 2: Statistical powers of estimating AIEM1 from Delta (left) and bootstrap (right) methods at different sample sizes ($b2=0:522$; $c=0:518$).

power increased as b_1 increased; when b_1 and $b_2=c$ were fixed, the statistical power decreased as $c=b_2$ increased. At the same parameter configuration, indirect effect inference showed slightly greater power if two mediators were generated independently.

Discussion and Future Work

In this paper we propose a mediation analysis through general definitions of total effect, direct effect and indirect effect. We demonstrate the method with various predictive models. The proposed method is

a general extension of mediation analysis under the counterfactual framework. It is also related to the traditional CD and CP methods. The method generalizes and improves the existing mediation analysis methodologies in many ways. First, responses, exposure variables and mediators can be measured at any scale: continuous, bi nary or multi-categorical. Second, multiple mediators of different types are allowed in the pathway analysis simultaneously. Indirect effect transmitted by an individual mediator can be differentiated from the total effect, which enables the comparison of the importance of the mediators. This property is especially useful for developing policies that aim at altering the relationship between a specific exposure variable and a response variable through controlling the intervention from third variables. With the knowledge of the indirect effect carried by each mediator/confounder, a policymaker is able to focus limited resources on changing the most important factors. Third, the mediation study allows correlations among mediators. Fourth, the concepts of mediation analysis can be applied in general predictive models. Finally, we provide two approaches to estimate the variance of indirect effect in parametric or nonparametric models. Our methods are demonstrated through a real example and a simulation study.

Several aspects of the proposed method will be explored in future work. We provide a non-parametric procedure of mediation analysis for binary exposures. The procedure will be extended to more general predictive models and other types of exposures. Only limited work has focused on mediation analysis in survival model contexts. The idea of mediation analysis proposed in this paper will also be extended to the additive models in survival analysis. With the proposed definitions, we can also take advantage of previous knowledge and information in the analysis. Therefore, we propose further research on implementing multiple mediation analysis in Bayesian settings.

Supplementary Material

1. The proofs of Lemmas 4.1, 5.1, 5.2. 2. Delta method to measure the variances of mediation effects.

Acknowledgments

References

1. Lynch JW, Kaplan GA, Cohen RD, Tuomilehto J, Salonen JT (1996) Do cardiovascular risk factors explain the relation between socioeconomic status, risk of all -cause mortality, cardiovascular mortality, and acute myocardial infarction?" Am J Epidemiol 144: 934-942.

2. Baron RM, Kenny DA (1986) The moderator-mediator variable distinction in social psychological research: conceptual, strategic, and statistical considerations. J Pers Soc Psychol 51: 1173-1182.

3. Hyman HH (1955) Survey design and analysis: Principles, cases and procedures. Glencoe, IL: Free Press.

4. Alwin DF, Hauser RM (1975) The decomposition of effects in path analysis. American Sociological Review 40: 37-47.

5. Judd CM, Kenny DA (1981) Process Analysis: Estimating mediation in treatment evaluations. Evaluation Review 5: 602-619.

6. MacKinnon DP, Dwyer JH (1993) Estimating mediated effects in prevention studies. Evaluation review 17: 144-158.

7. Mackinnon DP, Warsi G, Dwyer JH (1995) A Simulation Study of Mediated Effect Measures. Multivariate Behav Res 30: 41.

8. Kim J, Ferree GD (1981) Standardization in Causal Analysis. Sociological Method & Research 10, 187-210.

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

10. Pearl J (2001) Direct and indirect effects. Proceedings of the Seventeenth Conference on Uncertainty and Artificial Intelligence. San Francisco, CA: Morgan Kaufmann 411-420.

11. Have TR, Joffe MM, Lynch KG, Brown GK, Maisto SA, et al. (2007) Causal mediation analyses with rank preserving models. Biometrics 63: 926-934.

12. Albert JM (2008) Mediation analysis via potential outcomes models. Stat Med 27: 1282-1304.

13. Holland PW (1986) Statistics and Causal inference. J of the American Statistical Association 81: 945-960.

14. VanderWeele TJ, Vansteelandt S (2009) Conceptual issues concerning mediation, interventions and composition. Statistics and Its Interface 2: 457-468.

15. Rosenberg J, Chia YL, Plevritis S (2005) The effect of age, race, tumor size, tumor grade, and disease stage on invasive ductal breast cancer survival in the U.S. SEER database. Breast Cancer Res Treat 89: 47-54.

16. Sassi F, Luft HS, Guadagnoli E (2006) Reducing racial/ethnic disparities in female breast cancer: screening rates and stage at diagnosis. Am J Public Health 96: 2165-2172.

17. Gerend MA, Pai M (2008) Social determinants of Black-White disparities in breast cancer mortality: a review. Cancer Epidemiol Biomarkers Prev 17: 2913-2923.

18. Hewitt M, Somone JV (1999) Ensuring quality cancer care, National Academy press, Washington, DC.

19. MacKinnon DP, Krull JL, Lockwood CM (2000) Equivalence of the mediation, confounding and suppression effect. Prev Sci 1: 173-181.

20. Imai K, Yamamoto T (2010a) Identification and Sensitivity Analysis for Multiple Causal Mechanisms: Revisiting Evidence from Framing Experiments. Political Analysis 21: 141-171.

21. Imai K, Keele L, Tingley D (2010) Identification, Inference and sensitivity analysis of causal mediation effects. Statistical Science 25: 1-144.

22. Fan Y (2012) Multiple mediation analysis with general predictive models. Ph.D. dissertation, Louisiana State University Health Sciences Center.

23. Huang B, Sivaganesan S, Succop P, Goodman E (2004) Statistical assessment of mediational effects for logistic mediational models. Stat Med 23: 2713-2728.

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

25. Friedman JH (2001) Greedy function approximation: a gradient boosting machine. The Annals of Statistics 29: 1189-1232.

26. Palmgren J (1989) Regression models for bivariate binary response. University of Washington Biostatistics Working Paper Series.