Bias Analysis for The Principal Stratum Direct Effect in The Presence of Confounded Intermediate Variables

Yasutaka Chiba

Department of Environmental Medicine and Behavioral Science, Kinki University School of Medicine, Japan

Abstract

In epidemiological and clinical research, investigators often want to estimate the direct effect of a treatment on an outcome, which is not relayed by intermediate variables. Even if the total effect is unconfounded, the direct effect is not identified when unmeasured variables affect the intermediate and outcome variables. This article focuses on the principal stratum direct effect (PSDE) of a randomized treatment, which is the difference between expectations of potential outcomes within latent subgroups of subjects for whom the intermediate variable would be constant, regardless of the randomized treatment assignment. Unfortunately, the PSDE will not generally be estimated in an unbiased manner without untestable conditions, even if monotonicity is assumed. Thus, we propose bounds and a simple method of sensitivity analysis for the PSDE under a monotonicity assumption. To develop them, we introduce sensitivity parameters that are defined as the difference in potential outcomes with the same value of the intermediate variable between subjects who are assigned to the treatment and those who are assigned to the control group. Investigators can use the proposed method without complex computer programming. The method is illustrated using a randomized trial for coronary heart disease.

Keywords: Bounds; Causal inference; Intention-to-treat; Monte Carlo sensitivity analysis; Potential outcome

Introduction

Adjusting for an intermediate variable is a common analytic strategy in estimating a direct effect [1-4]. Even if the total effect is unconfounded, the direct effect is not identified when unmeasured variables affect the intermediate (mediator) and outcome variables. The total and direct effects can be formalized most readily by representing the problem nonparametrically in terms of directed acyclic graphs and counterfactual notation [5,6]. For example, in the context of randomized trials (Figure 1), the total effect of binary randomized treatment R on outcome Y is obtained without regard to intermediate D as simply the contrast between E[$Y \mid R = 1$] and E[$Y \mid R = 0$]: i.e., the intention-to-treat (ITT) effect. However, the salient scientific question of interest often involves not the total effect of R on Y, but rather only the portion of that effect that is not transmitted through the influence of R on intermediate D: i.e., the direct effect.

In many epidemiological and clinical studies in which investigators are interested in the direct effect, some factors that confound the relationship between the intermediate and outcome variables are present. Such factors are often unmeasured or not controlled for. If no control is made, the direct effect will not generally be estimated in an unbiased manner [7]. Thus, it is important to conduct a bias analysis for the direct effect, in the presence of unmeasured confounding between the intermediate and outcome variables.

Here, we focus on the application of the principal stratification approach for estimating the direct effect of a randomized treatment. Using this approach, we develop the bounds and a simple method of sensitivity analysis for the principal stratum direct effect (PSDE), which is the difference between expectations of potential outcomes within latent subgroups of subjects for whom the intermediate variable would be constant, regardless of the randomized treatment assignment. For example, the PSDE is closely related to issue of inference with a surrogate marker, where a good surrogate outcome serves as a mediator of treatment effect, leaving little effect of the treatment to directly impact the true outcome of interest though other channels [8]. Although bounds on the PSDE have been presented [9,10], we develop the bounds with narrower width by adding a plausible assumption in some situations. The methods of sensitivity

analysis have been also presented [11-13], but the methods require some functional model, and use somewhat complex formulae and calculations. Here, we develop a simple method that is much easier to use formulae.

We require the monotonicity assumption, a standard assumption often used in the literature of causal inference [14,15], and introduce sensitivity parameters that are defined as the difference in potential outcomes with the same value of the intermediate variable between subjects who are assigned to the treatment group and those who are assigned to the control group. The remainder of this manuscript is organized as follows. We review the PSDE in the next section. In the third section, we introduce sensitivity parameters, and propose the bounds and a method of sensitivity analysis on these bases. The developed bounds and sensitivity analysis are applied to a randomized trial for coronary heart disease (CHD) in the fourth section. The last section discusses some implications of the developed approach.

Definitions and Assumptions

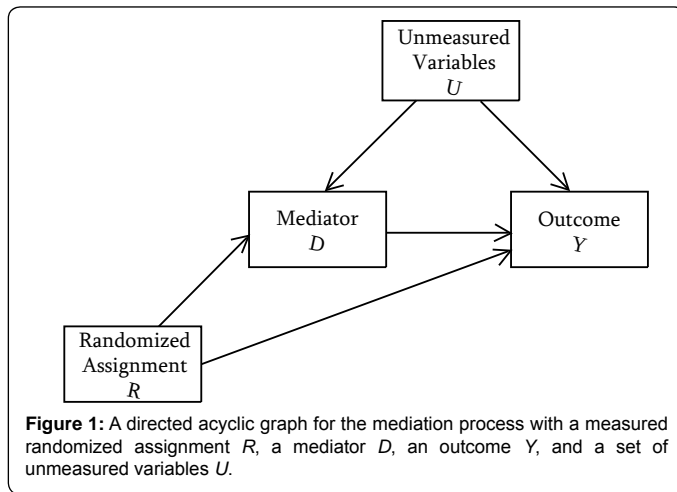### Potential outcome and principal stratification

We assume a deterministic potential outcomes framework [16-18]. Let $Y_{R=r}$ and $D_{R=r}$ denote the respective values of the potential outcome and mediator that would have been observed if the treatment R had been set. We require the consistency assumption; this assumption is that $Y_{R=r} = Y$, i.e., that the value of the potential outcome that would have been observed if the treatment r had been equal to the value of the observed outcome when actually

Figure 1: A directed acyclic graph for the mediation process with a measured randomized assignment R, a mediator D, an outcome Y, and a set of unmeasured variables U.

assigned to the treatment r. We further assume the independence of treatment assignment. This assumes that $Y_{R=r}$ is independent of R, and means that the treatment assignment gives no information about the distribution of potential outcomes. Note that the independency between the potential outcome $Y_{R=r}$ and R does not mean that the observed outcome Y is independent of R.

Using the principal stratification approach [2,8], four principal strata are formulated when the randomized treatment assignment and intermediate variable are dichotomous. These four principal strata are constructed of the following compliant-mediators, always-mediators, never-mediators, and defiant-mediators. Compliant-mediators exhibit positive intermediate behavior when assigned to the treatment, but do not exhibit positive intermediate behavior when assigned to the control. Therefore, $D_{R=1} = 1$ and $D_{R=0} = 0$. Always-mediators always exhibit positive intermediate behavior, regardless of the treatment assignment. Therefore, $D_{R=1} = 1$ and $D_{R=0} = 1$. Never-mediators never exhibit positive intermediate behavior, regardless of the treatment assignment. Therefore, $D_{R=1} = 0$ and $D_{R=0} = 0$. Defiant-mediators do not exhibit positive intermediate behavior when assigned to the treatment, but do exhibit such behavior when assigned to the control. Therefore, $D_{R=1} = 0$ and $D_{R=0} = 1$.

The principal stratum direct effect

Under the principal stratification approach, we focus on ITT effects in two of the four principal strata formed by the potential behavior. In Figure 1, the pathway between R and Y does not include D for the always- and never-mediating principal strata because the potential level of the mediator is constant within each of these two strata. Thus, the separate ITT effect of treatment within the always- and never-mediating principal strata is the PSDE [12].

We denote that t takes on the values 1, 2, 3, and 4, corresponding to the compliant-mediating, always-mediating, never-mediating, and defiant-mediating principal strata, respectively, and $C = t$ corresponds to the tth principal stratum. Then, the ITT effect for the tth principal stratum is

$$\theta_{\text{ITT}t} = \text{E}[Y_{R=1} \mid C = t] - \text{E}[Y_{R=0} \mid C = t].$$

The standard ITT effect over the whole population equals the weighted sum of the stratum-specific ITT effects across the four strata, with weights corresponding to the probabilities of membership in each principal stratum $\pi_t = \Pr(C = t)$ such that $\sum_{t=1}^{4} \pi_t = 1$. Therefore,

$$\theta_{\text{ITT}} = \sum_{t=1}^{4} \text{E}[Y_{R=1} - Y_{R=0} \mid C = t] \pi_t = \sum_{t=1}^{4} \pi_t \theta_{\text{ITT}t}. \tag{1}$$

The PSDE corresponds to the weighted sum of the ITT effect across the always- and never-mediating principal strata and is computed as

$$\text{PSDE} = \frac{\sum_{t=2}^{3} \text{E}[Y_{R=1} - Y_{R=0} \mid C = t] \pi_t}{\pi_2 + \pi_3} = \frac{\pi_2 \theta_{\text{ITT2}} + \pi_3 \theta_{\text{ITT3}}}{\pi_2 + \pi_3}. \tag{2}$$

The relationships between π_t and $p_r = \Pr(D = 1 \mid R = r)$ are as follows:

$$\pi_1 + \pi_2 = p_1, \pi_3 + \pi_4 = 1 - p_1, \pi_2 + \pi_4 = p_0, \text{ and } \pi_1 + \pi_3 = 1 - p_0,$$

because the subjects within each principal stratum should be homogeneously assigned to $R = 0$ and $R = 1$ due to randomization. Furthermore, $\text{E}[Y_{R=r} \mid D = d, R = r]$ is translated into the weighted sum of $\text{E}[Y_{R=r} \mid C = t]$ as follows [12,19]:

$$\text{E}[Y_{R=r} \mid D = 1, R = 1] = \frac{\pi_1}{\pi_1 + \pi_2} \text{E}[Y_{R=r} \mid C = 1] + \frac{\pi_2}{\pi_1 + \pi_2} \text{E}[Y_{R=r} \mid C = 2],$$

$$\text{E}[Y_{R=r} \mid D = 1, R = 0] = \frac{\pi_2}{\pi_2 + \pi_4} \text{E}[Y_{R=r} \mid C = 2] + \frac{\pi_4}{\pi_2 + \pi_4} \text{E}[Y_{R=r} \mid C = 4],$$

$$\text{E}[Y_{R=r} \mid D = 0, R = 1] = \frac{\pi_3}{\pi_3 + \pi_4} \text{E}[Y_{R=r} \mid C = 3] + \frac{\pi_4}{\pi_3 + \pi_4} \text{E}[Y_{R=r} \mid C = 4],$$

$$\text{E}[Y_{R=r} \mid D = 0, R = 0] = \frac{\pi_1}{\pi_1 + \pi_3} \text{E}[Y_{R=r} \mid C = 1] + \frac{\pi_3}{\pi_1 + \pi_3} \text{E}[Y_{R=r} \mid C = 3].$$

Unfortunately, neither $\text{E}[Y_{R=r} \mid C = t]$ nor π_t can be identified from the observed data.

We assume monotonicity, a standard assumption often used in the literature of causal inference [14,15]. This assumption is that no defiant-mediator exists (i.e., $\pi_4 = 0$). Then, $\pi_1 = p_1 - p_0$, $\pi_2 = p_0$, and $\pi_3 = 1 - p_1$. $p_0 = p_1$ ($\pi_1 = 0$) indicates that no compliant-mediator exists, and the monotonicity assumption does not hold if $p_0 > p_1$ ($\pi_1 < 0$). From these relationships between π_t and p_r, the following relationships between $\text{E}[Y_{R=r} \mid C = t]$ and $\text{E}[Y_{R=r} \mid D = d, R = r]$ are obtained:

$$\frac{p_1 - p_0}{p_1} \text{E}[Y_{R=r} \mid C = 1] + \frac{p_0}{p_1} \text{E}[Y_{R=r} \mid C = 2] = \text{E}[Y_{R=r} \mid D = 1, R = 1],$$

$$\text{E}[Y_{R=r} \mid C = 2] = \text{E}[Y_{R=r} \mid D = 1, R = 0],$$

$$\text{E}[Y_{R=r} \mid C = 3] = \text{E}[Y_{R=r} \mid D = 0, R = 1],$$

$$\frac{p_1 - p_0}{1 - p_0} \text{E}[Y_{R=r} \mid C = 1] + \frac{1 - p_1}{1 - p_0} \text{E}[Y_{R=r} \mid C = 3] = \text{E}[Y_{R=r} \mid D = 0, R = 0].$$

Bias Analysis

We define the D-specific sensitivity parameters as follows:

$$\alpha_d \equiv \text{E}[Y_{R=1} \mid D = d, R = 1] - \text{E}[Y_{R=1} \mid D = d, R = 0],$$

$$\beta_d \equiv \text{E}[Y_{R=0} \mid D = d, R = 1] - \text{E}[Y_{R=0} \mid D = d, R = 0],$$

where $d = 0, 1$. This definition of sensitivity parameters is an extension of an idea of the bias factors introduced in the context of randomized trials with noncompliance [20-22]. In these reports, it was assumed that the treatment affects the outcome only through the mediator: i.e., no direct pathway from R to Y exists in Figure 1, and the causal effect of interest was $\text{E}[Y_{D=1}] - \text{E}[Y_{D=0}]$. The bias factors were defined as $\gamma_r \equiv \text{E}[Y_{D=1} \mid D = 1, R = r] - \text{E}[Y_{D=1} \mid D = 0, R = r]$, and $\delta_r \equiv \text{E}[Y_{D=0} \mid D = 1, R = r] - \text{E}[Y_{D=0} \mid D = 0, R = r]$.

α_d and β_d are the difference in potential outcomes with the same mediator value between subjects who are assigned to the treatment group and those who are assigned to the control group. $\text{E}[Y_{R=r} \mid R = 1] = \text{E}[Y_{R=r} \mid R = 0]$ holds because it is assumed that $Y_{R=r}$ is independent of R, but $\text{E}[Y_{R=r} \mid D = d, R = 1] = \text{E}[Y_{R=r} \mid D = d, R =$

0] does not hold in general [1]. Then, these sensitivity parameters are interpreted as biases caused by conditioning on the mediator. $\alpha_d = 0$ and $\beta_d = 0$ hold conditional on some covariates, if these covariates include all of the confounders of the relationships between D and Y.

Using these sensitivity parameters, $E[Y_{R=1} \mid C = t]$ are represented by

$$E[Y_{R=1} \mid C = 1] = E_{01} - (1-p_0)\alpha_0 / (p_1 - p_0) = E_{11} + p_0\alpha_1 / (p_1 - p_0), \quad (3)$$

$$E[Y_{R=1} \mid C = 2] = E_{11} - \alpha_1, \quad (4)$$

$$E[Y_{R=1} \mid C = 3] = E_{01}, \quad (5)$$

and $E[Y_{R=0} \mid C = t]$ are represented by

$$E[Y_{R=0} \mid C = 1] = E_{00} - (1-p_1)\beta_0 / (p_1 - p_0) = E_{10} + p_1\beta_1 / (p_1 - p_0), \quad (6)$$

$$E[Y_{R=0} \mid C = 2] = E_{10}, \quad (7)$$

$$E[Y_{R=0} \mid C = 3] = E_{00} + \beta_0, \quad (8)$$

where $E_{dr} = E[Y \mid D = d, R = r]$. By substituting equations (4), (5), (7), and (8) into equation (2), the PSDE has the following formula:

$$PSDE = \frac{p_0(E_{11} - E_{10} - \alpha_1) + (1-p_1)(E_{01} - E_{00} - \beta_0)}{1 - p_1 + p_0} \quad (9)$$

We propose bounds on the PSDE using equations (3)–(8) and a sensitivity analysis using equation (9).

Bounds

Let (K_0, K_1) be the finite range of y. Then, $-(K_1 - K_0) \leq E[Y_{R=1} - Y_{R=0} \mid C = 1] \leq K_1 - K_0$ because $K_0 \leq E[Y_{R=r} \mid C = 1] \leq K_1$. Substituting $-(K_1 - K_0) \leq E[Y_{R=1} - Y_{R=0} \mid C = 1] \leq K_1 - K_0$ into equation (1) gives

$$\theta_{ITT} - (K_1 - K_0)\pi_1 \leq \sum_{t=2}^{3} E[Y_{R=1} - Y_{R=0} \mid C = t]\pi_t \leq \theta_{ITT} + (K_1 - K_0)\pi_1$$

[9,10]. From equation (2), bounds on the PSDE become

$$\frac{\theta_{ITT} - (K_1 - K_0)(p_1 - p_0)}{1 - p_1 + p_0} \leq PSDE \leq \frac{\theta_{ITT} + (K_1 - K_0)(p_1 - p_0)}{1 - p_1 + p_0} \quad (10)$$

In cases of a binary outcome, $K_0 = 0$ and $K_1 = 1$ are applied.

By making two assumptions, we improve inequality (10) to bounds with narrower width. Note that the following Assumption 1 is newly proposed, but Assumption 2 has been presented.

ASSUMPTION 1. *The expectation of potential outcomes for the compliant-mediators is between the expectations of potential outcomes for the always-mediators and never-mediators, which can be formalized as*

$$E[Y_{R=r} \mid C = 2] \leq E[Y_{R=r} \mid C = 1] \leq E[Y_{R=r} \mid C = 3]$$

or

$$E[Y_{R=r} \mid C = 3] \leq E[Y_{R=r} \mid C = 1] \leq E[Y_{R=r} \mid C = 2].$$

Under $E[Y_{R=1} \mid C = 2] \leq E[Y_{R=1} \mid C = 1] \leq E[Y_{R=1} \mid C = 3]$, equations (3)–(5) yield $\alpha_0 \geq 0$ and $\alpha_1 \geq 0$, and similarly, equations (6)–(8) yield $\alpha_0 \leq 0$ and $\alpha_1 \leq 0$ under $E[Y_{R=1} \mid C = 3] \leq E[Y_{R=1} \mid C = 1] \leq E[Y_{R=1} \mid C = 2]$. Therefore, the signs of α_0 and α_1 must be the same under Assumption 1. It is readily verified that the converse holds, i.e.,

$$E[Y_{R=1} \mid C = 2] \leq E[Y_{R=1} \mid C = 1] \leq E[Y_{R=1} \mid C = 3] \text{ if } \alpha_0 \geq 0 \text{ and } \alpha_1 \geq 0,$$

$$E[Y_{R=1} \mid C = 3] \leq E[Y_{R=1} \mid C = 1] \leq E[Y_{R=1} \mid C = 2] \text{ if } \alpha_0 \leq 0 \text{ and } \alpha_1 \leq 0.$$

From equation (3), bounds on $E[Y_{R=1} \mid C = 1]$ become $E_{11} \leq E[Y_{R=1} \mid C = 1] \leq E_{01}$ under $\alpha_0 \geq 0$ and $\alpha_1 \geq 0$, and $E_{01} \leq E[Y_{R=1} \mid C = 1] \leq E_{11}$ under $\alpha_0 \leq 0$ and $\alpha_1 \leq 0$. If the observed data show that $E_{01} \leq E_{11}$, α_0

≥ 0 and $\alpha_1 \geq 0$ cannot hold and then $E[Y_{R=1} \mid C = 2] \leq E[Y_{R=1} \mid C = 1] \leq E[Y_{R=1} \mid C = 3]$ cannot hold. Consequently, under Assumption 1,

$$E_{01} \leq E_{11} \Rightarrow \alpha_0 \leq 0 \text{ and } \alpha_1 \leq 0, \text{ and } E_{01} \leq E[Y_{R=1} \mid C = 1] \leq E_{11},$$

and conversely

$$E_{11} \leq E_{01} \Rightarrow \alpha_0 \geq 0 \text{ and } \alpha_1 \geq 0, \text{ and } E_{11} \leq E[Y_{R=1} \mid C = 1] \leq E_{01}.$$

Likewise, bounds on $E[Y_{R=0} \mid C = 1]$ under Assumption 1 are as follows:

$$E_{00} \leq E_{10} \Rightarrow \beta_0 \leq 0 \text{ and } \beta_1 \leq 0, \text{ and } E_{00} \leq E[Y_{R=0} \mid C = 1] \leq E_{10},$$

$$E_{10} \leq E_{00} \Rightarrow \beta_0 \geq 0 \text{ and } \beta_1 \geq 0, \text{ and } E_{10} \leq E[Y_{R=0} \mid C = 1] \leq E_{00}.$$

These bounds provide bounds on the PSDE. If the observed data show that $E_{01} \leq E_{11}$ and $E_{00} \leq E_{10}$, bounds on the PSDE become

$$\frac{\theta_{ITT} - (p_1 - p_0)(E_{11} - E_{00})}{1 - p_1 + p_0} \leq PSDE \leq \frac{\theta_{ITT} + (p_1 - p_0)(E_{10} - E_{01})}{1 - p_1 + p_0},$$

and conversely if the observed data show that $E_{11} \leq E_{01}$ and $E_{10} \leq E_{00}$, bounds on the PSDE become

$$\frac{\theta_{ITT} - (p_1 - p_0)(E_{01} - E_{10})}{1 - p_1 + p_0} \leq PSDE \leq \frac{\theta_{ITT} + (p_1 - p_0)(E_{00} - E_{11})}{1 - p_1 + p_0}.$$

Note that the other bounds on sensitivity parameters can be also derived from equations (3) and (6) under Assumption 1. When $E_{01} \leq E_{11}$, bounds on α_0 and α_1 are

$$-\frac{p_1 - p_0}{1 - p_0}(E_{11} - E_{01}) \leq \alpha_0 \leq 0 \text{ and } -\frac{p_1 - p_0}{p_0}(E_{11} - E_{01}) \leq \alpha_1 \leq 0$$

and when $E_{00} \leq E_{10}$, bounds on β_0 and β_1 are

$$-\frac{p_1 - p_0}{1 - p_1}(E_{10} - E_{00}) \leq \beta_0 \leq 0 \text{ and } -\frac{p_1 - p_0}{p_1}(E_{10} - E_{00}) \leq \beta_1 \leq 0$$

Conversely, these bounds are

$$0 \leq \alpha_0 \leq \frac{p_1 - p_0}{1 - p_0}(E_{01} - E_{11}) \text{ and } 0 \leq \alpha_1 \leq \frac{p_1 - p_0}{p_0}(E_{01} - E_{11}) \text{ when } E_{11} \leq E_{01},$$

$$0 \leq \beta_0 \leq \frac{p_1 - p_0}{1 - p_1}(E_{00} - E_{10}) \text{ and } 0 \leq \beta_1 \leq \frac{p_1 - p_0}{p_1}(E_{00} - E_{10}) \text{ when } E_{10} \leq E_{00}.$$

Chiba [19] presented the following assumption to derive an estimator of the PSDE.

ASSUMPTION 2. *The causal effects are the same between subpopulations with $(D, R) = (d, 1)$ and $(D, R) = (d, 0)$, which can be formalized as*

$$E[Y_{R=1} - Y_{R=0} \mid D = d, R = 1] = E[Y_{R=1} - Y_{R=0} \mid D = d, R = 0] \text{ for } d = 0, 1.$$

Assumption 2 is equivalent to $\alpha_d = \beta_d$ for $d = 0, 1$, and holds under the null hypothesis $Y_{R=1} = Y_{R=0}$. Under this assumption, we can obtain $\theta_{ITT1} = \theta_{ITT2} = E_{11} - E_{10} - \alpha_1$ and $\theta_{ITT1} = \theta_{ITT3} = E_{01} - E_{00} - \beta_0$ from equations (2) and (4). Therefore, PSDE $= \theta_{ITTt} = \theta_{ITT}$ for $t = 1, 2, 3$. $\alpha_0 = \beta_0$ and $\alpha_1 = \beta_1$ are estimated by

$$\alpha_0 = \beta_0 = p_1(E_{01} - E_{11}) + p_0(E_{10} - E_{00}),$$

$$\alpha_1 = \beta_1 = (1 - p_1)(E_{11} - E_{01}) + (1 - p_0)(E_{00} - E_{10}).$$

Note that PSDE $= \theta_{ITTt} = \theta_{ITT}$ for $t = 1, 2, 3$ is also derived under the assumption of

$$E[Y_{R=1} - Y_{R=0} \mid D = 1, R = r] = E[Y_{R=1} - Y_{R=0} \mid D = 0, R = r] \text{ for } r = 0, 1,$$

which holds under the null hypothesis $Y_{R=1} = Y_{R=0}$ [19].

Sensitivity analysis

Several approaches may be considered for sensitivity analyses.

Sjölander et al. [11] presented a method that was conducted by assuming structural regression models for $E[Y_{R=r} \mid C = c]$ and estimating the parameters using the expectation-maximization (EM) algorithm. Other researchers used Markov Chain Monte Carlo techniques to estimate the parameters in the framework of Bayesian inference [12,13]. Here, we propose a method without functional models or complex calculations. We use equation (9) only for the sensitivity analysis. The simplest approach is to vary the values of α_1 and β_0 within the relevant ranges of these values. Then, our approach can be regarded as a re-parameterization of their approaches, and its advantage is that it is much easier to use the formula in the sensitivity analysis.

We can also apply the Monte Carlo sensitivity analysis (MCSA) [23-25] using equation (9). For the MCSA, investigators assume prior distributions of the sensitivity parameters, and generate a large number (L) of estimates of the PSDE by drawing L sets of random values from their distributions. Then, a frequency distribution of L PSDE is generated, and we obtain the result without incorporating the random error of the estimate. To incorporate the random error, the distributions of E_{dr} and p_r based on the observed data are applied.

If investigators do not have reasonable information about prior distributions of the sensitivity parameters, they can use the bounds on α_1 and β_0 introduced here, once their bounds are obtained, uniform distributions within the ranges can be applied.

Application

Data

We illustrate the proposed bounds and sensitivity analysis using data from the Lipid Research Clinics Coronary Primary Prevention Trial (LRC-CPPT) [26]. The purpose of that study was to evaluate the efficacy of the cholesterol-lowering drug cholestyramine in the prevention of CHD in 3806 asymptomatic middle-aged men with hypercholesterolemia. In this study, 1888 subjects were randomly assigned to receive cholestyramine treatment ($R = 0$) and 1918 subjects were randomly assigned to receive a placebo ($R = 1$). During a follow-up period of 1 year, each CHD event was recorded ($Y = 0$ for no event and $Y = 1$ for an event). At the end of follow-up, cholesterol levels were recorded for each subject. We dichotomize cholesterol levels as $D = 0$ for < 280 mg/dL and $D = 1$ for ≥ 280 mg/dL, as in previous studies [27–30]. Data from the LRC-CPPT are displayed in Table 1 [27]. Note that this example is for illustrative purposes only, as the mediator D has been dichotomized and this can give rise to misleading influences.

In the LRC-CPPT, the four principal strata are as follows. Compliant-mediators are subjects whose cholesterol levels were higher than 280 mg/dL when assigned to the placebo group, but lower than 280 mg/dL when assigned to cholestyramine treatment. For always-mediators, regardless of treatment assignment, cholesterol levels were always higher than 280 mg/dL. Conversely, for never-mediators, regardless of treatment assignment, cholesterol levels were always

lower than 280 mg/dL (and never higher than 280 mg/dL). In contrast to compliant-mediators, defiant-mediators are subjects whose cholesterol levels were higher than 280 mg/dL when assigned to cholestyramine treatment, but lower than 280 mg/dL when assigned to the placebo group.

Bounds

Inequality (10) yielded bounds of $-22.39\% \leq \text{PSDE} \leq 27.06\%$. The width of the bounds is 49.45%, which is very wide and thus rather uninformative.

Although whether Assumptions 1 and 2 hold cannot be confirmed from the observed data, it is important to discuss it. In the LRC-CPPT, health-minded individuals may tend not to experience CHD and to have lower cholesterol levels than people who are not as health-conscious. Then, always-mediators, who are individuals with high cholesterol level regardless of treatment assignment, may mostly tend to experience CHD. Conversely, never-mediators, who are individuals with low cholesterol level regardless of treatment assignment, may mostly tend not to experience CHD. The probability of experiencing CHD in compliant-mediators, whose cholesterol levels depend on treatment assignment, may be between the probabilities for always- and never-mediators. This observation shows that $E[Y_{R=r} \mid C = 3] \leq E[Y_{R=r} \mid C = 1] \leq E[Y_{R=r} \mid C = 2]$. Therefore, Assumption 1 may hold. Investigators may not be able to insist that Assumption 2 holds, until the estimate of θ_{ITT} is 0 or at least close to 0. Even though the estimate of θ_{ITT} is close to 0, it may be difficult to insist on that.

Under Assumption 1, $\alpha_d \leq 0$, $\beta_d \leq 0$, and $E[Y_{R=r} \mid C = 3] \leq E[Y_{R=r} \mid C = 1] \leq E[Y_{R=r} \mid C = 2]$, because $E_{11} = 10.92\% (= 82/751) \geq E_{01} = 7.37\% (= 86/1167)$ and $E_{10} = 9.04\% (= 33/365) \geq E_{00} = 6.37\% (= 97/1523)$. Then, bounds on the PSDE were $1.21\% \leq \text{PSDE} \leq 2.75\%$. The width of the bounds has been improved to 1.54%, which yields significant information about the PSDE. The result shows that the PSDE is positive. Thus, it is concluded that cholestyramine treatment prevents CHD for subjects within the always- and never-mediating principal strata. Note that the lower bounds on α_d and β_d were -0.87%, -3.64%, -0.87%, and -1.35% for α_0, α_1, β_0, and β_1, respectively.

Under Assumption 2, $\text{PSDE} = \theta_{ITTt} = \theta_{ITT} = 1.87\%$ (95% confidence interval: 0.17%, 3.58%). Note that $\alpha_0 = \beta_0 = -0.87\%$ and $\alpha_1 = \beta_1 = 0.00\%$.

Sensitivity analysis

While we did not know about the distributions of the sensitivity parameters, we assumed that the sensitivity parameters followed uniform distributions with ranges obtained under Assumption 2, i.e., $-0.0364 \leq \alpha_1 \leq 0$ and $-0.0087 \leq \beta_0 \leq 0$. It was assumed that E_{dr} and p_r followed binomial distributions, with observed numbers and proportions estimated from the observed data.
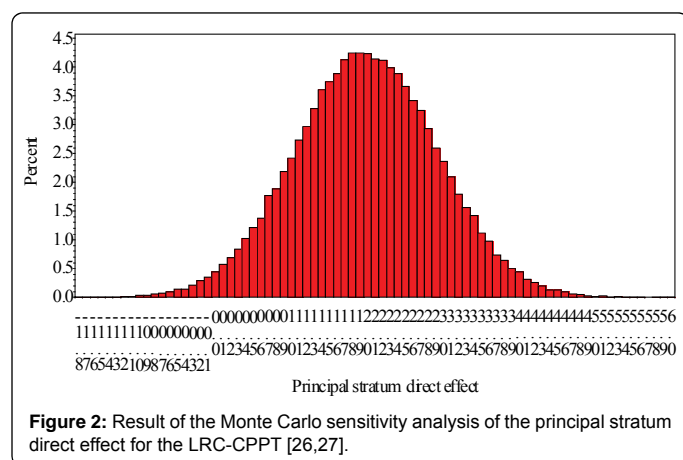
We drew 100,000 sets of random values from these distributions, and generated a frequency distribution of 100,000 PSDE. The result is shown in Figure 2. The 50th percentile of the resulting PSDE distribution was 1.98%, (2.5th percentile: 0.15%, 97.5th percentile: 3.81%), which was larger than θ_{ITT}. Again, the result shows that the PSDE is positive.

Discussion

We have proposed the bounds and a simple method of sensitivity analysis for the PSDE. To introduce bounds with narrower width, we made Assumption 1. The advantages of the proposed bounds are that their formulae are simple and the width is narrow. Although the

	Placebo ($R = 1$)			Cholestyramine Treatment ($R = 0$)		
	Cholesterol ≥ 280 mg/dL ($D = 1$)	Cholesterol < 280 mg/dL ($D = 0$)	Total Placebo	Cholesterol ≥ 280 mg/dL ($D = 1$)	Cholesterol < 280 mg/dL ($D = 0$)	Total Treatment
$Y = 1$	82	86	168	33	97	130
$Y = 0$	669	1081	1750	332	1426	1758
Total	751	1167	1918	365	1523	1888

Table 1: Definite CHD mortality or myocardial infarction events (Y) in the LRC-CPPT according to randomized cholestyramine treatment group (R) and serum cholesterol (mg/dL) at 1 year (D) [27].

Figure 2: Result of the Monte Carlo sensitivity analysis of the principal stratum direct effect for the LRC-CPPT [26,27].

bounds have a weakness in requiring some untestable assumptions, Assumption 1 is a reasonable assumption in some situations, when the observed data shows that $E_{01} \leq E_{11}$ and $E_{00} \leq E_{10}$, or $E_{11} \leq E_{01}$ and $E_{10} \leq E_{00}$.

In this paper, we have discussed randomized trials, where treatment is unconfounded. Some researchers may be interested in an extension to non-randomized trials, where treatment R is confounded. Such an extension can be achieved as follows, if all baseline covariates X are measured. In the presence of measured covariates X, all formulae in this paper hold by applying the expectations and probabilities conditional on X. Then,

$$\theta_{ITT2,x} = E[Y \mid D = 1, R = 1, X = x] - E[Y \mid D = 1, R = 0, X = x] - \alpha_{1,x},$$

$$\theta_{ITT3,x} = E[Y \mid D = 0, R = 1, X = x] - E[Y \mid D = 0, R = 0, X = x] - \beta_{0,x},$$

where

$$\alpha_{d,x} \equiv E[Y_{R=1} \mid D = d, R = 1, X = x] - E[Y_{R=1} \mid D = d, R = 0, X = x],$$

$$\beta_{d,x} \equiv E[Y_{R=0} \mid D = d, R = 1, X = x] - E[Y_{R=0} \mid D = d, R = 0, X = x].$$

With fixed values of $\alpha_{d,x}$ and $\beta_{d,x}$, we can estimate θ_{ITTt} after adjusting for x, for example, using regression analysis. Thus, the covariates-adjusted version of equation (9) can be obtained because equations (1) and (2) is also defined with adjusted θ_{ITTt} and π_t. This shows that our method can be used in the presence of baseline covariates that should be adjusted for. In practice, it is very hard that we assume the values or distributions of $\alpha_{d,x}$ and $\beta_{d,x}$ for all x. To reduce the number of sensitivity parameters, common $\alpha_d (= \alpha_{d,x})$ and $\beta_d (= \beta_{d,x})$ for all x will be applied.

A sensitivity analysis technique for the PSDE has previously been developed [11–13]. The technique requires some functional models, and use somewhat complex formulae and calculations in the sensitivity analysis. An advantage of our approach is that it is much easier to use formulae. Applying the MCSA, investigators can use our approach without complex computer programming. However, our approach has a disadvantage that it assumes monotonicity. In the LRC-CPPT, investigators can insist that the monotonicity assumption holds, if cholestyramine had beneficial effects for all subjects. In fact, however, cholestyramine may be beneficial on average but may be harmful for particular individuals. A logical next step in this research program would therefore be that the monotonicity assumption is relaxed without using complex formulae and calculations.

In this paper, we have discussed the PSDE, which is a causal effect that is not affected by intermediate variables. For example, such an

effect is closely related to issue of inference with a surrogate marker, where a good surrogate outcome serves as a mediator of treatment effect, leaving little effect of the treatment to directly impact the true outcome of interest though other channels. The developed bounds and sensitivity analysis for the PSDE will be used in such situations, and will further be extended to issue of inference with non-compliance and truncation by death.

Acknowledgements

References

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

2. Rubin DB (2004) Direct and indirect causal effects via potential outcomes. Scand J Stat 31: 161–170.

3. Kaufman S, Kaufman JS, MacLehose RF, Greenland S, Poole C (2005) Improved estimation of controlled direct effects in the presence of unmeasured confounding of intermediate variables. Stat Med 24: 1683–1702.

4. Petersen ML, Sinisi SE, van der Laan MJ (2006) Estimation of direct causal effects. Epidemiology 17: 276–284.

5. Pearl J (1995) Causal diagrams for empirical research. Biometrika 82: 669–688.

6. Greenland S, Pearl J, Robins JM (1999) Causal diagrams for epidemiologic research. Epidemiology 10: 37–48.

7. VanderWeele TJ (2009) Marginal structural models for the estimation of direct and indirect effects. Epidemiology 20: 18–26.

8. Frangakis CE, Rubin DB (2002) Principal stratification in causal inference. Biometrics 58: 21–29.

9. Zhang JL, Rubin DB (2003) Estimation of causal effects via principal stratification when some outcomes are truncated by "death". J Educ Behav Stat 28: 353–368.

10. Imai K (2008) Sharp bounds on causal effects in randomized experiments with "truncation-by-death." Stat Probab Lett 78: 144–149.

11. Sjölander A, Humphreys K, Vansteelandt S, Bellocco R, Palmgren J (2009) Sensitivity analysis for principal stratum direct effects, with an application to a study of physical activity and coronary heart disease. Biometrics 65: 514–520.

12. Gallop R, Small DS, Lin JY, Elliott MR, Joffe M, et al. (2009) Mediation analysis with principal stratification. Stat Med 28: 1108–1130.

13. Elliott MR, Raghunathan TE, Li Y (2010) Bayesian inference for causal mediation effects using principal stratification with dichotomous mediators and outcomes. Biostatistics 11: 353–372.

14. Angrist JD, Imbens GW, Rubin DB (1996) Identification of causal effects using instrumental variables. J Am Stat Assoc 91: 444–455.

15. Hudgens MG, Halloran ME (2006) Causal vaccine effects on binary postinfection outcomes. J Am Stat Assoc 101: 51–64.

16. Rubin DB (1974) Estimating causal effects of treatments in randomized and nonrandomized studies. J Educ Psychol 66: 688–701.

17. Rubin DB (1978) Bayesian inference for causal effects: The role of randomization. Ann Stat 6: 34–58.

18. Rubin DB (1990) Formal models of statistical inference for causal effects. J Stat Plann Infer 25: 279–292.

19. Chiba Y (2010) Estimating the principal stratum direct effect when the total effects are consistent between two standard populations. Stat Probab Lett 80: 958–961.

20. Chiba Y, Sato T, Greenland S (2007) Bounds on potential risks and causal risk differences under assumptions about confounding parameters. Stat Med 26: 5125–5135.

21. Chiba Y (2010) Bias analysis of the instrumental variable estimator as an estimator of the average causal effect. Contemp Clin Trials 31: 12–17.

22. Chiba Y (2010) An approach for estimating causal effects in randomized trials with noncompliance. Commun Stat Theory Methods 39: 2146–2156.

23. Greenland S (2003) The impact of prior distributions for uncontrolled confounding and response bias: A case study of the relation of wire codes and magnetic fields to childhood leukemia. J Am Stat Assoc 98: 47–54.

24. Steenland K, Greenland S (2004) Monte Carlo sensitivity analysis and Bayesian analysis of smoking as an unmeasured confounder in a study of silica and lung cancer. Am J Epidemiol 160: 384-392.

25. Greenland S (2005) Multiple-bias modeling for analysis of observational data. J R Stat Soc Ser A 168: 267–306.

26. The Lipid Research Clinics Coronary Primary Prevention Trial results.I. Reduction in incidence of coronary heart disease (1984) JAMA 251: 351-364.

27. Freedman LS, Graubard BI, Schatzkin A (1992) Statistical validation of intermediate endpoints for chronic diseases. Stat Med 11: 167–178.

28. Cai Z, Kuroki M, Pearl J, Tian J (2008) Bounds on direct effects in the presence of confounded intermediate variables. Biometrics 64: 695–701.

29. Sjölander A (2009) Bounds on natural direct effects in the presence of confounded intermediate variables. Stat Med 28: 558–571.

30. Chiba Y (2010) Bounds on controlled direct effects under monotonic assumptions about mediators and confounders. Biom J (in press) doi:10.1002/bimj.201000051.