Bounds on Direct Effects in the Presence of Confounded Intermediate Variables

Zhihong Cai,[1,*] **Manabu Kuroki,**[2] **Judea Pearl**[3] **and Jin Tian**[4]

[1] Department of Biostatistics, Kyoto University

Yoshida-Konoe-cho, Sakyo-ku, Kyoto, 606-8501, Japan

[2] Department of Systems Innovation, Osaka University

1-3, Machikaneyama-cho, Toyonaka, Osaka, 560-8531, Japan

[3] Department of Computer Science, UCLA

4532 Boelter Hall, Los Angeles, CA, 90024-1596

[4] Department of Computer Science, Iowa State University

226 Atanasoff Hall, Iowa State University, Ames, IA 50011

September 10, 2007

* *email:* cai@pbh.med.kyoto-u.ac.jp

Summary. This paper considers the problem of estimating the average controlled direct effect (ACDE) of a treatment on an outcome, in the presence of unmeasured confounders between an intermediate variable and the outcome. Such confounders render the direct effect unidentifiable even in cases where the total effect is unconfounded (hence identifiable). Kaufman et al. (2005) applied a linear programming software to find the minimum and maximum possible values of the ACDE for specific numerical data. In this paper, we apply the symbolic Balke-Pearl (1997) linear programming method to derive closed-form formulas for the upper and lower bounds on the ACDE under various assumptions of monotonicity. These universal bounds enable clinical experimenters to assess the direct effect of treatment from observed data with minimum computational effort, and they further shed light on the sign of the direct effect and the accuracy of the assessments.

Key words: Causal effect; Midpoint estimator; Potential response type; Stratified analysis

1. Introduction

Estimation of the direct effect of a treatment on an outcome is a central concern in epidemiological and clinical research (Buyse and Molenberghs, 1998; Kaufman et al., 2005; Petersen et al., 2006; Robins and Greenland, 1992; Rubin, 2004; Taylor et al., 2005; Wang and Taylor, 2002). Pearl (2001) gave a formal definition of the total effect decomposition into direct and indirect effects, and distinguished between the controlled direct effect and the natural direct effect, the former obtains when intermediate variables are held constant at specific values. Kaufman et al. (2005) considered the problem of estimating the average controlled direct effect (ACDE) of a treatment on an outcome, in the presence of unmeasured confounders between an intermediate variable and the outcome. Such confounders render the direct effect unidentifiable even in cases where the total effect is unconfounded (hence identifiable). Kaufman et al. (2005) applied a linear programming software to find the minimum and maximum possible values of the ACDE for specific numerical data. They further proposed the midpoint between the minimum and maximum values as an estimator of the ACDE. However, they did not provide exact formulas of the bounds on the ACDE.

In this paper, we apply the symbolic Balke-Pearl linear programming method (Balke, 1995; Balke and Pearl, 1997) to derive closed-form formulas of the upper and lower bounds on the ACDE under various assumptions of monotonicity. In contrast to the numerical method of Kaufman et al. (2005), these symbolic formulas enable clinical experimenters to assess the direct effect of a treatment on an outcome from observed data with minimum computational effort, and they further shed lights on the accuracy of the assessment. In addition, we derive bounds on the ACDE when covariate information is available. Moreover, we provide a formal formula for the midpoint estimator chosen by Kaufman et al. (2005), and propose a new stratified midpoint estimator that is more accurate when covariate measurements are available. In addition to the binary case, we further propose bounds on the ACDE in the case where observed variables are multi-categorical. Finally, we illustrate our results through an empirical example in both binary and multi-categorical cases.

2. Bounding formulas

2.1 *Problem description*

To motivate our problem, we examine the data from the Lipid Research Clinics Coronary Primary Prevention Trial (LRC-CPPT), shown in Table 1 (LRT-CPPT group, 1984; Kaufman et al, 2005). The purpose of this study is to evaluate the efficacy of the cholesterol lowering drug cholestyramine for the prevention of coronary heart disease (CHD) in 3806 hypercholesterolemia men. Our interest is to examine whether serum cholesterol level 1 year after initiation of cholestyramine was an adequate surrogate endpoint (i.e., explanation) for the outcome of CHD.

According to Freedman et al. (1992), a good surrogate endpoint is one that explains a large proportion of the total effect. A conventional approach to validate a surrogate endpoint is to estimate the relative contributions of the direct and indirect effects to the total effect. However, if there exist unmeasured confounding factors, for example, if there exist unmeasured genetic or life style factors that affect both cholesterol and CHD, estimating the direct effect requires careful causal analysis.

[Table 1 about here.]

To model presence of unmeasured confounding, we consider the directed acyclic graph shown in Fig. 1, where a treatment X, an intermediate Z and an outcome Y are binary variables taken values as x, z and y respectively ($x \in \{x_0, x_1\}$, $y \in \{y_0, y_1\}$, $z \in \{z_0, z_1\}$), and U is a set of unmeasured variables, which is independent of X. In this figure, the treatment is assumed randomized, hence there is no confounder between X and Y, and the total effect of X on Y is identifiable. However, the set of unmeasured confounders U between Z and Y renders the direct effect of X on Y unidentifiable. In other words, it is impossible to estimate this direct effect without making further assumptions. The central aim of this paper is to derive formulas of the bounds on the direct effect of X on Y in this causal structure.

[Figure 1 about here.]

The average controlled direct effects (ACDEs) are defined as

$$ACDE(z) = \text{pr}\{y_1|\text{do}(x_1), \text{do}(z)\} - \text{pr}\{y_1|\text{do}(x_0), \text{do}(z)\} \qquad (1)$$

for $z \in \{z_0, z_1\}$, where 'do($\cdot$)' denotes an imposed intervention (Kaufman et al., 2005; Pearl, 2000). $\text{pr}\{y|\text{do}(x), \text{do}(z)\}$ indicates the probability of $Y = y$ when we set X and Z to specific values x and z respectively by a joint intervention (Pearl, 2000).

Using the counterfactual notation of Neyman (1923) and Rubin (1974), Eqn (1) can be also written as

$$ACDE(z) = \text{pr}(Y_{x_1,z} = y_1) - \text{pr}(Y_{x_0,z} = y_1). \qquad (2)$$

A formal translation from graphs to counterfactual models is given by Pearl (2000).

Eqn (1) represents the average causal effect of X on Y when the causal path through Z is blocked by holding Z fixed at z_0 or z_1. Note that Eqn (1) is different from the crude stratum-specific risk difference $\mathrm{pr}(y_1|x_1, z) - \mathrm{pr}(y_1|x_0, z)$. The latter stands for the observed conditional risk difference in stratum z (in our example, the subgroup with cholesterol<280mg/dl or cholesterol≥280mg/dl), which represents the direct effect of X on Y, plus the spurious correlation between X and Y through the path $X \rightarrow Z \leftarrow U \rightarrow Y$. On the other hand, Eqn (1) represents the direct effect only, since the path $X \rightarrow Z \leftarrow U \rightarrow Y$ had been blocked by an intervention on Z. If the ACDE equals 0 in our example, then we can judge that cholesterol (Z) is a perfect surrogate endpoint for CHD (Y), which suggests that lowering cholesterol level constitutes an adequate explanation for how the drug prevents the occurrence of CHD.

In order to derive bounds on the ACDE, we follow Kaufman et al. (2005) and define 64 potential response types. First, we consider X (cholestyramine) as a treatment and Z (cholesterol) as an outcome. Since X and Z are binary variables, there are four possible potential response types at the unit-level: (1) a subject whose cholesterol increases regardless of taking cholestyramine or placebo (doomed), (2) a subject whose cholesterol decreases only by taking cholestyramine (causative), (3) a subject whose cholesterol decreases only by not taking cholestyramine (preventive), and (4) a subject whose cholesterol decreases regardless of taking cholestyramine or placebo (immune) (Greenland and Robins, 1986). We index these four types by a mapping variable

$r_z = 1, 2, 3, 4$. Similarly, when we consider X (cholestyramine) as a treatment and Y (CHD) as an outcome with Z (cholesterol) fixed to z_0 or z_1, there still exist doomed, causative, preventive and immune potential response types. We denote these four types by a mapping variable $r_{y|z_0} = 1, 2, 3, 4$ when Z is fixed to z_0, and another mapping variable $r_{y|z_1} = 1, 2, 3, 4$ when Z is fixed to z_1. Therefore, any of the $4 \times 4 \times 4$ index triples, $(r_z,\, r_{y|z_0}, r_{y|z_1})$, represents a potential response type. The joint probability distribution of $(r_z,\, r_{y|z_0},\, r_{y|z_1})$ is defined by

$$q_{ijk} = \mathrm{pr}(r_z = i, r_{y|z_0} = j, r_{y|z_1} = k)$$

for $i, j, k = 1, 2, 3, 4$, and the $\{q_{ijk}\}$ represent the proportion of the 64 potential response types among the population. Thus, the population of interest is fully characterized by $\{q_{ijk}\}$, and we can rewrite Eqn (2) as

$$ACDE(z_1) = \sum_{i=1}^{4}\sum_{j=1}^{4}\left(\sum_{k\in\{1,2\}} q_{ijk} - \sum_{k\in\{1,3\}} q_{ijk}\right)$$

$$ACDE(z_0) = \sum_{i=1}^{4}\sum_{k=1}^{4}\left(\sum_{j\in\{1,2\}} q_{ijk} - \sum_{j\in\{1,3\}} q_{ijk}\right).$$

See Appendix A for a detail derivation.

Kaufman et al. (2005) applied a linear program software package to find the minimum and maximum possible values of the ACDE for specific numerical data. However, they did not provide exact formulas for the ACDE. Balke (1995) and Balke and Pearl (1997) describe a computer program that takes symbolic description of linear programming problems and returns symbolic expressions for the desired bounds. In this paper, we apply this symbolic method to derive closed-form formulas for the ACDEs under three sets of assumptions. Details of this method are included in Appendix A.

2.2 *No assumption case*

When no assumption is made, there are 64 potential response types, while there are only 8 observed conditional probabilities $\mathrm{pr}(y, z|x)$. Using the Balke-Pearl method (Balke, 1995; Balke and Pearl, 1997), the formulas for the tightest upper and lower bounds on the ACDEs, are given by

$$\mathrm{pr}(y_0, z|x_0) + \mathrm{pr}(y_1, z|x_1) - 1 \leq ACDE(z) \leq 1 - \mathrm{pr}(y_1, z|x_0) - \mathrm{pr}(y_0, z|x_1) \tag{3}$$

for $z \in \{z_1, z_0\}$, which defines the range within which the ACDE must lie. It is remarkable that we get such a simple formula, consisting of only one additive expression in the lower bound and one additive expression in the upper bound.

To find when the lower bound coincides with the upper bound, we calculate their difference and obtain $\mathrm{pr}(z_{1-i}|x_0) + \mathrm{pr}(z_{1-i}|x_1)$ for $ACDE(z_i)$ ($z_i \in \{z_1, z_0\}$). Hence, in order to make the lower bound equal the upper bound, both $\mathrm{pr}(z_{1-i}|x_0)$ and $\mathrm{pr}(z_{1-i}|x_1)$ must be zero. This indicates that the upper bound cannot coincide with the lower bound in both $ACDE(z_0)$ and $ACDE(z_1)$ at the same time, because the probabilities in all the cells must be zero in order to achieve it. That is, the bounding interval never vanishes, regardless of the observations.

In addition, it should be noted that Eqn (3) provides a simple testable criterion for the existence of a direct effect, that is, if $\mathrm{pr}(y_0, z|x_0)+\mathrm{pr}(y_1, z|x_1) > 1$, then we are assured that $ACDE(z)$ is positive, and if $\mathrm{pr}(y_1, z|x_0) + \mathrm{pr}(y_0, z|x_1) > 1$, $ACDE(z)$ must be negative.

2.3 *Monotonic assumption case*

Kaufman et al. (2005) made two assumptions regarding the potential response types: (1) monotonic assumption, which means no unit-level causal effects of X on Z or of X on Y or of Z on Y can be negative, and (2) no-interaction assumption, which means that, for all units, the response of Y to change in X does not depend on the level at which we hold Z. There are 18 potential response types that satisfy monotonic assumption, that is, $\{q_{i11}, q_{i21}, q_{i22}, q_{i41}, q_{i42}, q_{i44} : i = 1, 2, 4\}$, and 12 potential response types that satisfy both monotonic and no-interaction assumptions, that is, $\{q_{i11}, q_{i22}, q_{i41}, q_{i44} : i = 1, 2, 4\}$ (Kaufman and Kaufman, 2006). By applying the Balke-Pearl method, we derive closed-form formulas for the tightest bounds on the ACDEs in the two cases. The following equations give the upper and lower bounds under monotonic assumption:

$$\max\left\{\begin{array}{c} 0 \\ \mathrm{pr}(y_0, z_1|x_0) - \mathrm{pr}(y_0, z_1|x_1) \end{array}\right\} \le ACDE(z_1) \le \mathrm{pr}(y_0|x_0) - \mathrm{pr}(y_0, z_1|x_1) \tag{4}$$

$$\max\left\{\begin{array}{c} 0 \\ \mathrm{pr}(y_1, z_0|x_1) - \mathrm{pr}(y_1, z_0|x_0) \end{array}\right\} \le ACDE(z_0) \le \mathrm{pr}(y_1|x_1) - \mathrm{pr}(y_1, z_0|x_0) \tag{5}$$

It is seen that the interval collapses when $\mathrm{pr}(y_0, z_1|x_0) = \mathrm{pr}(y_0|x_0)$ (or $\mathrm{pr}(y_0, z_0|x_0) = 0$) in Eqn (4) and $\mathrm{pr}(y_1, z_0|x_1) = \mathrm{pr}(y_1|x_1)$ (or $\mathrm{pr}(y_1, z_1|x_1) = 0$) in Eqn (5). In these cases, the ACDEs can be evaluated by the upper or lower bound.

On the basis of the lower bounds of Eqns (4) and (5), we can judge whether there exist positive direct effects under the monotonic assumption. That is, if $\mathrm{pr}(y_0, z_1|x_0) > \mathrm{pr}(y_0, z_1|x_1)$ and/or $\mathrm{pr}(y_1, z_0|x_1) > \mathrm{pr}(y_1, z_0|x_0)$, then we are assured that there exist positive direct effects.

Moreover, Eqns (4) and (5) provide a simple necessary test for the monotonic assumption. That is, if the monotonic assumption holds true, then the upper bounds should be no less than zero, because the lower bounds are nonnegative. Thus, if the observed quantities are $\mathrm{pr}(y_0|x_0) < \mathrm{pr}(y_0, z_1|x_1)$ or $\mathrm{pr}(y_1|x_1) < \mathrm{pr}(y_1, z_0|x_0)$, then the upper bounds would be negative, which indicates that the monotonic assumption does not hold in this situation.

On the other hand, the following equation gives the upper and lower bounds under both monotonic and no-interaction assumptions:

$$\max\left\{\begin{array}{c} 0 \\ \mathrm{pr}(y_0, z_1|x_0) - \mathrm{pr}(y_0, z_1|x_1) \\ \mathrm{pr}(y_1, z_0|x_1) - \mathrm{pr}(y_1, z_0|x_0) \\ \mathrm{pr}(y_0, z_1|x_0) - \mathrm{pr}(y_0, z_1|x_1) + \mathrm{pr}(y_1, z_0|x_1) - \mathrm{pr}(y_1, z_0|x_0) \end{array}\right\} \leq ACDE(z)$$

$$\leq \mathrm{pr}(y_1|x_1) - \mathrm{pr}(y_1|x_0) \tag{6}$$

for $z \in \{z_1, z_0\}$. It is seen that the upper bound is the total effect of X on Y. We can judge whether there exist positive direct effects from the lower bound of Eqn (6). That is, if either $\mathrm{pr}(y_0, z_1|x_0) > \mathrm{pr}(y_0, z_1|x_1)$ or $\mathrm{pr}(y_1, z_0|x_1) > \mathrm{pr}(y_1, z_0|x_0)$, then we are assured that there exist positive direct effects.

Further, Eqn (6) provides a simple necessary test for both monotonicity and no-interaction assumptions. Since $ACDE(z)$ must be nonnegative from the lower bounds of Eqn (6), then, if $\mathrm{pr}(y_1|x_1) < \mathrm{pr}(y_1|x_0)$, the upper bounds would be negative, which indicates that at least one of the two assumptions is violated.

One more thing to be mentioned is that, if the monotonic assumption and no-interaction assumption hold true, then the bounds under monotonic assumption should not be wider than those under no assumption, and similarly, the bounds under both monotonic and no-interaction assumptions should not be wider than those under monotonic assumption. Therefore, the bounds obtained from a given dataset are capable of indicating which assumption does not hold for that dataset.

2.4 *Estimation accuracy*

In section 2.2 and 2.3, we derive the estimators for the lower and upper bounds under three sets of assumptions. Another problem is the estimation accuracy of these estimators. For no assumption case, it is easy to obtain the exact variance for the lower and upper bounds. However, for the remaining two cases, it is very complicated to derive the variances for the lower bounds though it is easy to obtain the exact variances for the upper bounds. In Appendix B, we provide the variance estimators for the lower and upper bounds under the three cases. In addition, we evaluate the performance of the proposed variance estimators through simulation studies.

3. Stratified ACDE

The analysis of section 2 applies to situations where all confounders between Z and Y are unmeasured. However, if some of these confounders are observed, this information is helpful in narrowing the bounds on direct effects. In this section, we consider the directed acyclic graph with the set of confounders U in Fig. 1 being divided into two sets of variables: measured covariates S and unmeasured covariates W.

Then, we define the stratified ACDE as

$$ACDE(z|s) = \mathrm{pr}\{y_1|\mathrm{do}(x_1), \mathrm{do}(z), s\} - \mathrm{pr}\{y_1|\mathrm{do}(x_0), \mathrm{do}(z), s\} \qquad (7)$$

for $s \in \{s_1, \cdots, s_k\}$ and $z \in \{z_1, z_0\}$, where

$$\mathrm{pr}\{y|\mathrm{do}(x), \mathrm{do}(z), s\} = \frac{\mathrm{pr}\{y, s|\mathrm{do}(x), \mathrm{do}(z)\}}{\mathrm{pr}\{s|\mathrm{do}(x), \mathrm{do}(z)\}}.$$

Since S is a set of observed baseline covariates, it is not affected by X or Z, which implies $\mathrm{pr}\{s|\mathrm{do}(x), \mathrm{do}(z)\} = \mathrm{pr}(s)$ (Pearl, 2000). Let k be the number of categories of S, then 64^k potential response types are needed in order to obtain the tightest bound of ACDE. However, if we limit the potential responses to the case where $S = s$ is observed, there are in total 64 potential response types in each stratum. Therefore, we can apply the previous discussion to derive the bounds on the stratified ACDEs. When there exist 64 potential response types, the stratified ACDE is given by

$$\begin{aligned}
&\mathrm{pr}(y_0, z|x_0, s) + \mathrm{pr}(y_1, z|x_1, s) - 1 \\
&\qquad \leq \quad ACDE(z|s) \leq 1 - \mathrm{pr}(y_1, z|x_0, s) - \mathrm{pr}(y_0, z|x_1, s)
\end{aligned}$$

for $z \in \{z_1, z_0\}$. Similarly, if the monotonic assumption holds true in stratum s, we can obtain

$$\begin{aligned}
&\max\left\{\begin{array}{c} 0 \\ \mathrm{pr}(y_0, z_1|x_0, s) - \mathrm{pr}(y_0, z_1|x_1, s) \end{array}\right\} \\
&\quad \leq \quad ACDE(z_1|s) \leq \mathrm{pr}(y_0|x_0, s) - \mathrm{pr}(y_0, z_1|x_1, s),
\end{aligned}$$

$$\begin{aligned}
&\max\left\{\begin{array}{c} 0 \\ \mathrm{pr}(y_1, z_0|x_1, s) - \mathrm{pr}(y_1, z_0|x_0, s) \end{array}\right\} \\
&\quad \leq \quad ACDE(z_0|s) \leq \mathrm{pr}(y_1|x_1, s) - \mathrm{pr}(y_1, z_0|x_0, s).
\end{aligned}$$

In addition, if both the monotonic and no-interaction assumptions hold true in stratum s, we can obtain

$$\max \left\{ \begin{array}{c} 0 \\ \mathrm{pr}(y_0, z_1|x_0, s) - \mathrm{pr}(y_0, z_1|x_1, s) \\ \mathrm{pr}(y_1, z_0|x_1, s) - \mathrm{pr}(y_1, z_0|x_0, s) \\ \mathrm{pr}(y_0, z_1|x_0, s) - \mathrm{pr}(y_0, z_1|x_1, s) + \mathrm{pr}(y_1, z_0|x_1, s) - \mathrm{pr}(y_1, z_0|x_0, s) \end{array} \right\} \tag{8}$$

$$\leq ACDE(z|s) \leq \mathrm{pr}(y_1|x_1, s) - \mathrm{pr}(y_1|x_0, s)$$

for $z \in \{z_1, z_0\}$. Thus, since the $ACDE(z)$ can be obtained by

$$ACDE(z) = \sum_s ACDE(z|s)\mathrm{pr}(s),$$

letting $LB_s(z)$ and $UB_s(z)$ be the lower bound and the upper bound in stratum s respectively, the summarized bounds on the $ACDE(z)$ by using covariate information can be evaluated by

$$\sum_s LB_s(z)\mathrm{pr}(s) \leq ACDE(z) \leq \sum_s UB_s(z)\mathrm{pr}(s). \tag{9}$$

These summarized bounds on direct effects are not wider than the bounds derived in Section 2, a simple proof of which is provided in Appendix C.

We would like to point out some practical requirements for the observed covariates S. First of all, S must be baseline covariates in order for the method to be valid. Moreover, we can divide such baseline covariates into the following three cases: (a) S is a confounder between Z and Y; (b) S has an effect on Z but not on Y; (c) S has an effect on Y but not on Z. If the measured covariates S satisfies any of the three cases, then the summarized bounds of Eqn (9) should not be wider than those provided in Section 2.

4. Midpoint estimator

Kaufman et al. (2005) proposed the midpoint between the minimum and maximum values as an estimator of the ACDE, which is given as

$$\mathrm{mRD}(z) = \frac{UB(z) + LB(z)}{2}, \quad z \in \{z_0, z_1\}, \tag{10}$$

where $LB(z)$ and $UB(z)$ are the linear programming minimum and maximum values for $ACDE(z)$ derived from the observed probabilities using linear programming packages. With the derived formulas in Section 2, we can now present an exact formula of the midpoint estimator. For example, the midpoint estimator with no assumption is derived directly as

$$\mathrm{mRD}(z) = \frac{1}{2}\{\mathrm{pr}(y_0, z|x_0) + \mathrm{pr}(y_1, z|x_1) - \mathrm{pr}(y_1, z|x_0) - \mathrm{pr}(y_0, z|x_1)\} \tag{11}$$

based on Eqn (3). The midpoint estimators for the remaining two cases can be derived in the same way. Thus, we can calculate the midpoint estimator from the observed data without using linear programming packages.

When covariate information is available, we propose a new stratified midpoint estimator, which is given by

$$\mathrm{mRD}_s(z) = \sum_s \frac{LB_s(z) + UB_s(z)}{2} \mathrm{pr}(s), \tag{12}$$

where S is a set of observed baseline covariates discussed in Section 3.

The new stratified midpoint estimator is superior to the midpoint estimator when some covariates are observed. To see this, we consider a hypothetical example when both monotonic and no-interaction assumptions hold true, and there is a binary observed covariate S. Table 2 shows the true proportion of 12 potential response types in each stratum, and Table 3 shows the observed conditional probabilities $\mathrm{pr}(y, z|x, s)$ induced from Table 2. Here, $\mathrm{pr}(s_1)$ is set to be 0.45.

[Table 2 about here.]

[Table 3 about here.]

Then, according to Kaufman et al's method, the bounds on the direct effect are (0.050, 0.175), and the midpoint estimate is 0.112. On the other hand, the bounds are $(0.190, 0.230)$ in stratum s_1 and $(0.090, 0.130)$ in stratum s_0 according to our formula (8). Then, we calculate the summarized lower and upper bounds according to our summarized formula (9), which are (0.135, 0.175), and the stratified midpoint estimator according to formula (12), which is 0.155. Here, we can calculate the true stratified ACDE from Table 2, which is 0.220 in stratum s_1 and 0.120 in stratum s_0. In addition, the true ACDE is 0.165 from Table 2, which is included in both Kaufman et al's bounds and our bounds. However, it is seen that Kaufman et al's midpoint estimator is quite away from the true ACDE and outside our bounds, while the stratified midpoint estimator is close to the true ACDE.

5. Extension to multi-categorical case

In the discussion above, we consider the ACDE when observed variables are binary. In this section, we consider the case where X, Y and Z are multi-categorical variables. When the categorical treatment variable X is changed from x to x', we define the ACDE as

$$\mathrm{ACDE}(y, z, x, x') = \mathrm{pr}\{y|\mathrm{do}(x'), \mathrm{do}(z)\} - \mathrm{pr}\{y|\mathrm{do}(x), \mathrm{do}(z)\}.$$

where y and z are possible values of Y and Z, respectively. Then, we provide the lower and upper bounds on the ACDE under the multi-categorical case:

$$
\begin{aligned}
-1 + \mathrm{pr}(z|x) + \mathrm{pr}(y,z|x') - \mathrm{pr}(y,z|x) &\leq \mathrm{ACDE}(y,z,x,x') \\
\leq \; 1 - \mathrm{pr}(z|x') + \mathrm{pr}(y,z|x') - \mathrm{pr}(y,z|x).
\end{aligned}
$$

The proof is given in Appendix D. When X, Y and Z are binary variables, these bounds are consistent with Eqn (3). Kang and Tian (2006) provided a method to obtain the inequality constraint for causal effects from nonexperimental data in the presence of unobserved variables. The above bounds can also be obtained by using their method.

6. Empirical example

6.1 *Binary case*

We illustrate our results through the example given in Section 2. Kaufman et al. (2005) collapsed the serum cholesterol values into two categories from 5 original categories, based on the data in Freedman et al. (1992). We will discuss the five categories in next subsection.

Because treatment X is randomized, the total effect of X on Y can be unbiased estimated by the risk difference $\mathrm{pr}(y_1|x_1)$-$\mathrm{pr}(y_1|x_0)$=0.0876-0.0689=0.0187. On the other hand, the observed stratum-specific risk difference is $\mathrm{pr}(y_1|z_0,x_1) - \mathrm{pr}(y_1|z_0,x_0) = 0.0737 - 0.0637 = 0.0100$ in stratum z_0, and $\mathrm{pr}(y_1|z_1,x_1) - \mathrm{pr}(y_1|z_1,x_0) = 0.1092 - 0.0904 = 0.0188$ in stratum z_1. Thus, as noted in Kaufman et al. (2005), there appears to be a direct causative effect of not receiving cholestyramine on the risk of CHD in each stratum of intermediate.

The bounds on the ACDE under no assumption are [-0.1999, 0.3850] in stratum z_0, and [-0.7814, 0.6337] in stratum z_1, which are relatively wide. Here, according to Kaufman et al. (2005), it is reasonable to assume that neither cholestyramine nor absence of hyperlipidaemia may elevate risk of the outcomes, nor may cholestyramine elevate serum cholesterol, leading to 18 potential response types for consideration. In addition, the necessary test for the monotonic assumption in Section 2 shows that $\mathrm{pr}(y_0|x_0)$-$\mathrm{pr}(y_0, z_1|x_1)$=0.5823>0, and $\mathrm{pr}(y_1|x_1)$-$\mathrm{pr}(y_1, z_0|x_0)$=0.0362>0, which suggests that the monotonic assumption holds for the data. Then, according to our formulas, the bounds are [0,0.0362] in stratum z_0, and [0,0.5823] in stratum z_1. The upper bound in stratum z_1 can be as large as 0.5823, which is much larger than the total effect 0.0187. Even the midpoint estimator is 0.2912, larger than 0.0187. Therefore, it may not be helpful to calculate the relative contribution of the direct and indirect effects to the total effect, in order to validate the serum cholesterol level as a surrogate endpoint. One explanation is that there exists potential response type q_{442}, which contributes the $ACDE(z_1)$ value but does not contribute to the total effect.

When we restrict to 12 potential response types, again, the necessary test for no-interaction assumption holds, i.e., $\mathrm{pr}(y_1|x_1)-\mathrm{pr}(y_1|x_0)=0.0187 > 0$. The bounds on the $ACDE(z)$ are [0, 0.0187] in both strata. The upper bound equals the total effect, because the interactive potential response type q_{442} does not exist. The midpoint estimator gives an estimate 0.0094, which indicates that there may exist a direct effect of cholestyramine treatment on CHD without mediating serum cholesterol.

Moreover, it is noted that the bounds under both monotonic and no-interaction assumptions are narrower than those under monotonic assumption, and that the bounds under monotonic assumption are narrower than those under no assumption. Therefore, it is reasonable to make these two assumptions concerning this example.

6.2 *Multi-categorical case*

Freedman et al. (1992) provided the data from the LRC-CPPT study, where the serum cholesterol values (Z) have 5 categories, shown in Table 4. Based on our formulas in Section 5, we calculate the lower and upper bounds when the serum cholesterol is fixed at each of the 5 categories, which are shown in Table 4. When we compare the bounds in binary case with those in Table 4, it is seen that with the number of categories of Z increases, the observed probabilities become smaller and the bounds become wider, which indicates that the width of the bounds is dependent on the sparcity of the observations. However, the bounds are helpful if one is interested in the ACDE under more detailed categories, which the bounds of binary case cannot provide.

[Table 4 about here.]

7. Discussion

This paper applied the Balke-Pearl method to derive closed-form formulas for the upper and lower bounds on the ACDEs under three sets of assumptions. We also considered extensions to situations where the treatment, the intermediate and the outcome are multinomial, rather than dichotomous variables, as well as situations in which the confounding factors are partially observed, so that covariate-adjusted bounds and midpoint estimators can be obtained.

Since our approach is nonparametric and mainly based on observed information, the proposed bounds define a range within which the direct effect must lie. On the basis of these deterministic bounds, one can narrow the bound width substantially by introducing subject matter constraints. Therefore, these universal bounds are helpful for epidemiologists and clinical experimenters to assess the direct effect of treatment.

8. Supplementary materials

Web Appendices and Tables referenced in Sections 2, 3, 5 are available under the Paper Information link at the Biometrics website http://www.tibs.org/biometrics.

Acknowledgements

We would like to thank Prof. Sol Kaufman of University at Buffalo and Prof. Jay S. Kaufman of University of North Carolina for their helpful communication. We are also grateful to Prof. Tosiya Sato of Kyoto University for his valuable comments. This paper was partially supported by the Japan Society for the Promotion of Science, the Ministry of Education, Culture, Sports, Science and Technology of Japan, the Kayamori Foundation of Informational Science Advancement, the Kurara Foundation, and the NSF grants IIS-0535223 and IIS-0347846.

Résumé

The French version of the summary, if you want to translate it yourself.

References

Balke, A. (1995). Probabilistic counterfactuals: semantics, computation, and applications, UCLA Cognitive Systems Laboratory, Technical Report (R-242), Ph.D. Thesis.

Balke, A. and Pearl, J. (1997). Bounds on treatment effects from studies with imperfect compliance. *Journal of the American Statistical Association* **92**, 1171-1176.

Buyse, M. and Molenberghs, G. (1998). Criteria for the validation of surrogate endpoints in randomized experiments. *Biometrics* **54**, 1014-1029.

Freedman, L. S., Graubard, B. I., and Schatzkin, A. (1992). Statistical validation of intermediate endpoints for chronic diseases. *Statistics in Medicine* **11**, 167-178.

Greenland, S. and Robins, J. M. (1986). Identifiability, exchangeability, and epidemiological confounding. *International Journal of Epidemiology* **15**, 413-419.

Kang, C. and Tian, J. (2006). Inequality constraints in causal models with hidden variables. *Proceedings of the twenty second Conference on Uncertainty in Artificial Intelligence*, Boston, MA: UAI Press, 411-420.

Kaufman, S., Kaufman, J. S., MacLenose, R. F., Greenland, S., and Poole, C. (2005). Improved estimation of controlled direct effects in the presence of unmeasured confounding of intermediate variables. *Statistics in Medicine* **24**, 1683-1702. Correction, **25**, 3228.

Kaufman, S. and Kaufman, J. S. (2006). Personal Communication.

The Lipid Research Clinics Coronary Primary Prevention Trial Results (1984). I. Reduction in incidence of coronary heart disease. *Journal of the American Medical Association* **251**, 351-364.

Neyman, J. (1923). Justification of applications of the calculus of probabilities to the solutions of certain questions in agricultural experimentation. Excerpts English translation (1990) in *Statistical Science* **5**, 463-472.

Pearl, J. (2000). *Causality: Models, Reasoning, and Inference*. Cambridge University Press.

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

Petersen, M. L., Sinisi, S. E., and van der Laan, M. J. (2006). Estimation of direct causal effects. *Epidemiology* **17**, 276-284.

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

Rubin, D. B. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology* **66**, 688-701,

Rubin, D. B. (2004). Direct and indirect causal effects via potential outcomes. *Scandinavian Journal of Statistics* **31**, 161-170.

Taylor J. M. G., Wang Y., Thiebaut R. (2005). Counterfactual links to the proportion of treatment effect explained by a surrogate marker. *Biometrics* **61**, 1102-1111.

Wang, Y. and Taylor, J. M. G. (2002). A measure of the proportion of treatment effect explained by a surrogate marker. *Biometrics* **58**, 803-812.

Fig. 1. A directed acyclic graph with a measured treatment X, an intermediate Z and an outcome Y, and a set of unmeasured variables U (Kaufman et al., 2005)

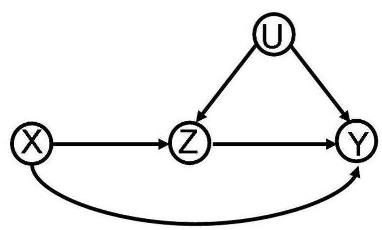

Fig. 1

Table 1. Definite CHD mortality or myocardial infarction events (Y) in the LRC-CPPT study according to randomized cholestyramine treatment group (X) and serum cholesterol (md/dl) at 1 year (Z) (Kaufman et al., 2005)

	placebo (x_1)			cholestyramine treatment (x_0)		
	cholesterol $\geq$280mg/dl (z_1)	cholesterol <280mg/dl (z_0)	total placebo	cholesterol $\geq$280mg/dl (z_1)	cholesterol <280mg/dl (z_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 2. A hypothetical example: proportion of potential response types in strata s_1 and s_0

s_1				s_0			
q_{111}	0.10	q_{141}	0.01	q_{111}	0.10	q_{141}	0.20
q_{211}	0.01	q_{241}	0.01	q_{211}	0.30	q_{241}	0.01
q_{411}	0.30	q_{441}	0.02	q_{411}	0.10	q_{441}	0.01
q_{122}	0.01	q_{144}	0.10	q_{122}	0.10	q_{144}	0.10
q_{222}	0.01	q_{244}	0.01	q_{222}	0.01	q_{244}	0.01
q_{422}	0.20	q_{444}	0.22	q_{422}	0.01	q_{444}	0.05

Table 3. Observed conditional probabilities $\mathrm{pr}(y, z|x, s)$ induced from Table 2

		s_1		s_0	
		y_1	y_0	y_1	y_0
x_1	z_1	0.15	0.11	0.72	0.11
	z_0	0.50	0.24	0.11	0.06
x_0	z_1	0.11	0.11	0.30	0.20
	z_0	0.31	0.47	0.40	0.10

Table 4. Data from definite CHD mortality or myocardial infarction events (Y) in the Lipid Research Clinics Coronary Primary Prevention Trial (Freedman et al., 1992) and the lower and upper bounds on the ACDE in 5 cholesterol categories

	placebo (x_1)		cholestyramine (x_0)		bounds	
cholesterol (Z)	y_1	y_0	y_1	y_0	lower	upper
<180	0	7	9	97	-0.949	0.992
180-230	8	83	34	641	-0.656	0.939
230-280	78	991	54	688	-0.595	0.455
280-330	64	572	23	281	-0.818	0.690
>330	18	97	10	51	-0.964	0.944