Sharpening Bounds on Principal Effects with Covariates

Dustin M. Long and
Department of Biostatistics, West Virginia University, Morgantown, WV 26506-9190, USA

Michael G. Hudgens
Department of Biostatistics, University of North Carolina, Chapel Hill, NC 27599-7420, USA

Dustin M. Long: dmlong@hsc.wvu.edu; Michael G. Hudgens: mhudgens@bios.unc.edu

Summary

Estimation of treatment effects in randomized studies is often hampered by possible selection bias induced by conditioning on or adjusting for a variable measured post-randomization. One approach to obviate such selection bias is to consider inference about treatment effects within principal strata, i.e., principal effects. A challenge with this approach is that without strong assumptions principal effects are not identifiable from the observable data. In settings where such assumptions are dubious, identifiable large sample bounds may be the preferred target of inference. In practice these bounds may be wide and not particularly informative. In this work we consider whether bounds on principal effects can be improved by adjusting for a categorical baseline covariate. Adjusted bounds are considered which are shown to never be wider than the unadjusted bounds. Necessary and sufficient conditions are given for which the adjusted bounds will be sharper (i.e., narrower) than the unadjusted bounds. The methods are illustrated using data from a recent, large study of interventions to prevent mother-to-child transmission of HIV through breastfeeding. Using a baseline covariate indicating low birth weight, the estimated adjusted bounds for the principal effect of interest are 63% narrower than the estimated unadjusted bounds.

Keywords

Bounds; Causal Effects; Partial Identifiability; Potential Outcomes; Principal Strata

1. Introduction

Often in randomized trials to evaluate the effect of a treatment, inference is hampered by possible selection bias induced by conditioning on or adjusting for a variable measured post-randomization. For instance, in randomized clinical trials where some study participants do not comply with their treatment assignment, investigators are often interested in the effect of treatment in participants who were compliant. One approach that avoids potential selection bias induced by conditioning on a post-randomization variable is to focus inference on the causal effect within a principal strata of interest, i.e., the principal effect (Frangakis and Rubin 2002). Principal strata are defined by the pair of potential outcomes under either

treatment assignment of the post-randomization variable. For instance, in the setting of non-compliance the principal stratum of interest may be study participants who would comply with their randomization assignment regardless of whether assigned to treatment or control (Angrist et al. 1996). In vaccine trials, a principal stratum of interest may be individuals who would be infected at a certain time regardless of vaccine status (Shepherd et al. 2011). In studies of interventions to prevent mother-to-child transmission (MTCT) of HIV through breastfeeding, a principal stratum of interest is infants who would be uninfected at a certain time regardless of treatment (Nolen and Hudgens 2011; Long and Hudgens 2012). There are many other settings where inference about treatment effects within certain principal strata may be of interest; see VanderWeele (2011) for a recent overview. A primary goal in these settings is often to better understand a treatment's effect by drawing inference about effects within subgroups defined by the principal strata. The principal stratification approach is less helpful for decision making, as principal stratum membership is generally not identifiable prior to treatment (Joffe 2011).

In many instances, even after treatment assignment individual principal strata membership is not identifiable from the observable data without strong assumptions because only one of the two post-randomization variable potential outcomes is ever observed for an individual. In turn, the principal effect of interest is not identifiable. One approach to cope with lack of identifiability is to conduct sensitivity analysis wherein some model is assumed, indexed by an unidentifiable parameter conditional on which the principal effect is identifiable from the observable data. Inference about the principal effect is then conducted conditional on some value of the unidentifiable parameter and sensitivity of the inference is examined by considering different values of the parameter. An alternative approach entails drawing inference about bounds on the principal effects, e.g., Zhang and Rubin (2003), Cheng and Small (2006). Informally, these extreme bounds provide the smallest and largest possible values of the principal effect consistent with the observed data distribution. This approach is appealing in that typically bounds can be obtained under minimal assumptions. However, in many cases the bounds may be quite wide and therefore not particularly informative about the principal effect.

Grilli and Mealli (2008) derived nonparametric bounds on principal effects under a number of different assumptions. They suggested these bounds can be improved (or narrowed) by creating bounds within strata defined by a baseline covariate and combining these stratum specific bounds by taking a weighted average to obtain new *adjusted* bounds on the principal effect. Grilli and Mealli employed this approach in the analysis of data from an employment study with mixed results: the adjusted bounds were an improvement on only one side of the unadjusted bounds, i.e., the adjusted upper bound was less than the unadjusted upper bound but the adjusted lower bound was also less than the unadjusted lower bound. The reason for only partial improvement was not addressed. More recent work by Lee (2009) and Mealli and Pacini (2012) indicate the adjusted bounds will never be wider than the unadjusted bounds and sometimes the adjusted bounds will be strictly narrower than the unadjusted bounds. In this paper we characterize the exact circumstances for which adjusting for a baseline covariate leads to improved bounds.

The outline of the remainder of this paper is as follows. In Section 2, notation and assumptions are introduced. Section 3 explains the lack of identifiability of the principal effect and in Section 4 the unadjusted bounds are reviewed. Section 5 defines the adjusted bounds based on a weighted average of bounds within levels of the baseline covariate. Section 6 contains the main result of this paper, giving necessary and sufficient conditions under which the covariate adjusted bounds improve upon (i.e., are narrower than) the unadjusted bounds. In Section 7 large sample inferential methods are discussed. In Section 8, the adjusted and unadjusted bounds are compared using data from a recent, large MTCT study. Conditions in which adjusting for the covariate leads to identification of the principal effect are discussed in Section 9. A brief discussion is given in Section 10. Proofs of the propositions in Section 6 are given in the Web Appendix A.

2. Notation and Assumptions

To motivate, throughout we consider the MTCT example where infants of HIV positive mothers are randomized at birth (i.e., baseline or time 0) to treatment or control. Suppose n infants are enrolled in a MTCT study and for $i = 1, \ldots, n$ let Z_i denote the randomization assignment for infant i. Without loss of generality let $Z_i = 0$ correspond to control and $Z_i = 1$ correspond to treatment. Let X_i be some binary variable measured at baseline (prior to randomization) taking on values 0 or 1. For simplicity X_i is assumed to be binary for now, although the results derived below will apply for any baseline categorical covariate with a finite number of levels. The primary endpoint in MTCT studies is typically HIV infection of the infant by some time point τ (e.g., six months) after baseline. Denote the presence or absence of the primary endpoint by Y_i, where $Y_i = 1$ indicates infant i became infected by τ and otherwise $Y_i = 0$. Because the goal of treatment is to prevent breast milk transmission of HIV, investigators are primarily interested in infant HIV infections that occur before τ but after some time $\tau_0 > 0$ (e.g., $\tau_0 = 2$ weeks), as infections prior to τ_0 are likely due to in utero or peripartum transmission and not breastfeeding. Let S_i denote whether infant i is infected by τ_0, where $S_i = 1$ if infant i is HIV infected by τ_0 and $S_i = 0$ otherwise. Let $S_i(z)$ denote the potential value of S_i when assigned treatment z for $z = 0, 1$ such that $S_i = (1 - Z_i)S_i(0) + Z_iS_i(1)$. Define $Y_i(z)$ similarly. Assume the treatment assignment of an infant does not affect the potential outcomes of other infants (i.e., there is no interference) and that there are not multiple forms of treatment.

An analysis of the effect of treatment that simply excludes infants infected by τ_0 (i.e., $S_i = 1$) is subject to selection bias. That is, because the set (or population) of infants that would be infected by τ_0 when assigned treatment $Z_i = 1$ is not necessarily the same as the set of infants that would be infected by τ_0 if assigned control $Z_i = 0$, direct comparisons between trial arms that exclude infants infected by τ_0 in general do not have a causal interpretation. To avoid such selection bias, the principal stratification framework may be adopted (Frangakis and Rubin 2002). Principal strata are defined by sets of infants with the same potential outcome pair $(S_i(0) = s_0, S_i(1) = s_1)$. Define the always infected (AI) principal stratum to be infants with $s_0 = s_1 = 1$, i.e., infants who would be infected at τ_0 regardless of randomization assignment. Similarly define the harmed stratum as those infants with $s_0 = 0, s_1 = 1$; the protected stratum as those infants with $s_0 = 1, s_1 = 0$; and the never infected (NI) stratum as those infants with $s_0 = s_1 = 0$. Based on considerations described above, investigators

conducting MTCT trials are interested in the NI stratum. The causal estimand of interest, the principal effect, is the effect of treatment on Y_i in infants who would be uninfected at τ_0 under either randomization assignment, namely

$$CE = \Pr[Y_i(1)=1|S_i(0)=S_i(1)=0] - \Pr[Y_i(0)=1|S_i(0)=S_i(1)=0].$$

Below we consider large sample bounds for *CE* that do and do not adjust for the baseline covariate X_i. Throughout we assume

> Assumption 1 (Independent Treatment Assignment): $Z_i \perp (X_i, S_i(0), S_i(1), Y_i(0), Y_i(1))$

> Assumption 2 (Monotonicity): $S_i(1) \quad S_i(0)$ for all i

where $\perp$ denotes independence. Assumption 1 will hold in randomized trials. Monotonicity assumes that treatment does no harm with respect to the intermediate variable S_i, i.e., there are no infants who would be infected by τ_0 only if treated. Under Assumption 2, there are only three possible principal strata: AI, NI, and protected.

3. Partial Identifiability

In this section we consider identifiability of *CE*. Let $\theta_{zst} = \Pr[Y_i(z) = 1|S_i(1) = s, S_i(0) = t]$, $\pi_z = \Pr[Y_i(z) = 1|S_i(z) = 0]$, and $\gamma = \Pr[S_i(0) = 0|S_i(1) = 0]$, such that $CE = \theta_{100} - \theta_{000}$. Assume $\gamma > 0$ as otherwise the NI stratum is empty with probability 1. Under Assumptions 1 and 2, $\theta_{000} = \Pr[Y_i = 1|S_i = 0, Z_i = 0]$, which is identifiable from the observed data. However, θ_{100} is not identifiable. Following Hudgens and Halloran (2006) note

$$\Pr[Y_i(1)=1|S_i(1)=0] = \Pr[Y_i=1|S_i=0, Z_i=1]$$
$$= \Pr[Y_i(1)=1|S_i(0)=0, S_i(1)=0]\Pr[S_i(0)=0|S_i(1)=0] + \Pr[Y_i(1)=1|S_i(0)=1, S_i(1)=0]\Pr[S_i(0)=1|S_i(1)=0],$$

i.e.,

$$\pi_1 = \gamma\theta_{100} + (1-\gamma)\theta_{101}. \quad (1)$$

Under Assumption 1, π_1 is identifiable. Under Assumptions 1 and 2, γ is identifiable because

$$\gamma = \frac{\Pr[S_i(0)=0]}{\Pr[S_i(1)=0]} = \frac{\Pr[S_i=0|Z_i=0]}{\Pr[S_i=0|Z_i=1]}.$$

On the other hand, θ_{100} and θ_{101} are not identifiable because infants who were assigned treatment and uninfected at τ_0 are a mixture of infants from the protected and NI strata. Solving (1) for θ_{100} yields

$$\theta_{100} = \frac{\pi_1}{\gamma} - \frac{1-\gamma}{\gamma}\theta_{101}. \quad (2)$$

Equation (2) describes a line with intercept π_1/γ and slope $-(1-\gamma)/\gamma$. Any pair of points $(\theta_{101}, \theta_{100})$ in the unit square that are on this line will give rise to the same observed data distribution.

Note that CE is identifiable if and only if $\gamma = 1$, $\pi_1 = 1$, or $\pi_1 = 0$. If $\gamma = 1$, then (2) is a horizontal line with intercept π_1 and thus $\theta_{100} = \pi_1$. Note $\gamma = 1$ is equivalent to $\Pr[S_i(0) = 0]$ $= \Pr[S_i(1) = 0]$, i.e., treatment has no effect on the intermediate variable S_i. If $\pi_1 = 1$, then (1) implies $\theta_{100} = 1$; geometrically this corresponds to the line (2) intersecting the unit square at the upper right corner (1,1). In words, $\pi_1 = 1$ means that all treated infants who are uninfected at τ_0 will become infected by τ. Likewise, if $\pi_1 = 0$, then (1) implies $\theta_{100} = 0$, corresponding to the line (2) intersecting the unit square at the origin (0,0). In words, $\pi_1 = 0$ means that all treated infants who are uninfected at τ_0 will not become infected by τ. Otherwise, if $\gamma < 1$ and $0 < \pi_1 < 1$, under Assumptions 1 and 2, CE is not identifiable from the observable random variables. On the other hand, θ_{000}, π_1, and γ are identifiable, implying the observed data distribution does reveal some information about possible values for CE, i.e., CE is partially identifiable. The focus of the sequel is on large sample bounds, i.e., the smallest and largest possible values of CE that are consistent with the observed data law.

4. Unadjusted Bounds

In this section, we present large sample bounds for CE (Zhang and Rubin 2003; Hudgens et al. 2003) that ignore the baseline covariate X. The large sample bounds for CE are found by first bounding θ_{100}. The upper bound for θ_{100} is obtained by assuming $\theta_{101} = 0$ or $\theta_{100} = 1$. Likewise, the lower bound for θ_{100} is obtained by assuming $\theta_{101} = 1$ or $\theta_{100} = 0$. These bounds can be envisaged as corresponding to the points where the line (2) intersects the unit square (Hudgens and Halloran 2006). In particular, the upper and lower bounds are

$$\theta_{100}^u = \min\left\{\frac{\pi_1}{\gamma}, 1\right\} \quad \text{and} \quad \theta_{100}^l = \max\left\{\frac{\pi_1 - (1-)}{\gamma}, 0\right\}. \quad (3)$$

Bounds for CE are found by replacing θ_{100} by θ_{100}^u and θ_{100}^l, i.e., $CE^u = \theta_{100}^u - \theta_{000}$ and $CE^l = \theta_{100}^l - \theta_{000}$. These bounds will be referred to as "unadjusted" bounds because no information from the covariate is used.

To illustrate, let the probabilities corresponding to a fictitious trial of MTCT of HIV be $\gamma = 0.95$, $\pi_1 = 0.02$, and $\pi_0 = 0.05$. Using (3), for this trial

$\theta_{100}^l = \max\{(0.02 - (1 - 0.95))/0.95, 0\} = 0$ and $\theta_{100}^u = \min\{0.02/0.95, 1\} = 0.021$ (note here and in the sequel that reported numerical results are rounded, in this case to the third decimal place). This gives the unadjusted bounds as $[CE^l, CE^u] = [0 - 0.05, 0.021 - 0.05] = [-0.05, -0.029]$ because $\theta_{000} = \pi_0$. In this example the bounds exclude zero, implying treatment reduces the risk of $Y = 1$ in the NI stratum. Let the probabilities for a second fictitious trial

be $\gamma = 0.80$, $\pi_1 = 0.85$, and $\pi_0 = 0.95$. Then $\theta_{100}^u = 1$ and $\theta_{100}^l = 0.813$, implying the unadjusted bounds are $[CE^l, CE^u] = [-0.137, 0.05]$. These two fictitious trials will be revisited in the next section.

5. Adjusted Bounds

Next we consider the method proposed by Grilli and Mealli (2008) for adjusting the large sample bounds using the binary baseline covariate X, i.e., bounds will be obtained within strata defined by X and then weighted averages of the stratum specific bounds will be computed. Let $\theta_{zstx} = \Pr[Y_i(z) = 1|S_i(1) = s, S_i(0) = t, X_i = x]$, $\gamma_x = \Pr[S_i(0) = 0|S_i(1) = 0, X_i = x]$, $\pi_{zx} = \Pr[Y_i(z) = 1|S_i(1) = 0, X_i = x]$, $\varphi_x = \Pr[X_i = x|S_i(1) = S_i(0) = 0]$, and $\lambda_x = \Pr[X_i = x| S_i(1) = 0]$. Throughout it is assumed that $\Pr[S_i(1) = 0, S_i(0) = t, X_i = x] > 0$ for $x, t = 0, 1$ such that the conditional probabilities θ_{z0tx}, γ_x, and π_{zx} are all well defined. This assumption also implies $\gamma_x > 0$, $\varphi_x > 0$ and $\lambda_x > 0$ for $x = 0, 1$. Note

$$\theta_{100} = \sum_x \theta_{100x} \phi_x, \quad (4)$$

where here and in the sequel $\sum_x = \sum_{x=0}^{1}$. As in the unadjusted case, θ_{100x} is not identifiable but using arguments analogous to (2) for $X_i = x$

$$\theta_{100x} = \frac{\pi_{1x}}{\gamma_x} - \frac{1-\gamma_x}{\gamma_x} \theta_{101x}, \quad (5)$$

and identifiable upper and lower bounds for θ_{100x} are

$$\theta_{100x}^u = \min\left\{ \frac{\pi_{1x}}{\gamma_x}, 1 \right\} \quad \text{and} \quad \theta_{100x}^l = \max\left\{ \frac{\pi_{1x} - (1-\gamma_x)}{\gamma_x}, 0 \right\}. \quad (6)$$

Under Assumptions 1 and 2, φ_x is identifiable because $\Pr[X_i = x|Z_i = 0, S_i = 0] = \Pr[X_i = x| S_i(0) = 0] = \Pr[X_i = x|S_i(1) = S_i(0) = 0]$. Therefore, identifiable bounds for θ_{100} can be obtained by combining (4) and (6), namely

$$\theta_{100X}^u = \sum_x \theta_{100x}^u \phi_x \quad \text{and} \quad \theta_{100X}^l = \sum_x \theta_{100x}^l \phi_x. \quad (7)$$

This leads to adjusted bounds $CE_X^u = \theta_{100X}^u - \theta_{000}$ and $CE_X^l = \theta_{100X}^l - \theta_{000}$.

Table 1 contains the values of two different binary baseline covariates, X_1 and X_2, for each of the fictional trials discussed in Section 4. For the first trial and X_1, by (6) we have $\theta_{1000}^u = \min\{0.035/0.995, 1\} = 0.035$, $\theta_{1000}^l = \max\{(0.035 - (1-0.995))/0.995, 0\} = 0.030$, $\theta_{1001}^u = 0.011$, and $\theta_{1001}^l = 0$. Thus, $\theta_{100X}^u = 0.035 \times 0.419 + 0.011 \times 0.581 = 0.021$ and $\theta_{100X}^l = 0.013$. Therefore, there is improvement to the lower bound of CE but not to the upper bound when adjusting for X_1, and thus the unadjusted bounds $[CE^l, CE^u] = [-0.05, -0.029]$ are wider

than the adjusted bounds $[CE^l_{X_1}, CE^u_{X_1}]=[-0.037, -0.029]$. On the other hand, adjusting by X_2 in the first trial does not yield an improvement since $\theta^u_{100X}=0.021=\theta^u_{100}$ and $\theta^l_{100X}=0=\theta^l_{100}$.

For the second trial and X_1, $\theta^u_{1000}=0.854$, $\theta^l_{1000}=0.730$, $\theta^u_{1001}=1$, and $\theta^l_{1001}=0.878$. Thus, $\theta^u_{100X}=0.935$ and $\theta^l_{100X}=0.813$. Here adjusting for X_1 yields a smaller upper bound resulting in narrower bounds, i.e., $[CE^l, CE^u] = [-0.137, 0.05]$ to $[CE^l_{X_1}, CE^u_{X_1}]=[-0.137, -0.015]$. Moreover, the adjusted upper bound is less than the null value of 0, indicating treatment has an effect in the NI principal stratum; such a conclusion was not possible prior to adjusting for X_1. On the other hand, adjusting for X_2 in the second trial yields no improvement in the bounds because $\theta^u_{100X}=1=\theta^u_{100}$ and $\theta^l_{100X}=0.813=\theta^l_{100}$.

A graphical depiction of the unadjusted and adjusted bounds is given in Figure 1. The unadjusted bounds correspond to where the solid lines intersect the unit square. Bounds within strata defined by X correspond to where the dashed and dotted lines intersect the unit square. The adjusted bounds, represented by ○ and +, are weighted averages of these stratum specific bounds. For example, in the upper left panel corresponding to trial 1 and X_1, we see that θ^l_{100X} is greater than θ^l_{100} because the vertical value of + is greater than zero, the point where the solid line intersects the horizontal axis.

6. Improvement of the Bounds

The examples in the preceding section illustrate that adjusting for a baseline covariate may or may not improve the bounds on *CE*. In this section, we give necessary and sufficient conditions for when the adjusted bounds (7) will be narrower than the unadjusted bounds of (3). Proofs of all propositions are given in Web Appendix A.

Proposition 1

$[\theta^l_{100X}, \theta^u_{100X}] \subseteq [\theta^l_{100}, \theta^u_{100}]$ for any baseline binary covariate *X*.

According to Proposition 1, the adjusted bounds will be at least as narrow as the unadjusted bounds no matter the choice of *X*. This proposition is analogous to Proposition 1b of Lee (2009) for *Y* continuous. To characterize the conditions under which the adjusted bounds are strictly narrower than the unadjusted bounds, consider the following two criteria:

$$\pi_{1x}<\gamma_x \text{ and } \pi_{1x'}>\gamma_{x'} \text{ for some } x \neq x' \quad (8)$$

and

$$\pi_{1x}>(1-\gamma_x) \text{ and } \pi_{1x'}<(1-\gamma_{x'}) \text{ for some } x \neq x' \quad (9)$$

where the value of *x* in (8) and (9) is not necessarily the same. In words, (8) and (9) indicate that X_i is informative about particular orderings between (i) the distribution of $S_i(0)$ given

$S_i(1) = 0$ and (ii) the distribution of $Y_i(1)$ given $S_i(1) = 0$. For the motivating example, these criteria indicate that the ordering of (i) the risk of infection (or not) by τ_0 when not treated and (ii) the risk of infection by τ when treated is different between strata defined by the levels of X_i among infants who would not be infected by τ_0 if treated. On the other hand, if X_i is uninformative about the relation between (i) and (ii) then neither (8) nor (9) will hold. For example, if X_i is independent of $S_i(0)$ given $S_i(1) = 0$ and if X_i is independent of $Y_i(1)$ given $S_i(1) = 0$, then neither (8) nor (9) will hold. Using (8) and (9), the following propositions characterize exactly the situations when the adjusted bounds will be narrower.

Proposition 2

$\theta^u_{100X} < \theta^u_{100}$ if and only if X satisfies (8).

Proposition 2 states (8) is a necessary and sufficient condition for the adjusted upper bound for θ_{100} to be less than the unadjusted bound. This proposition is exemplified in the second fictional trial from Section 5 using X_1, where $\pi_{10} < \gamma_0$ and $\pi_{11} > \gamma_1$. As shown in the lower left panel of Figure 1, (8) is satisfied because the dashed line has intercept greater than 1 and the dotted line has intercept less than 1. In contrast, for X_1 in the first fictional trial (8) is not satisfied because the dashed and dotted lines in the upper left panel of Figure 1 both have intercepts less than 1; thus the adjusted upper bound based on X_1 does not improve upon the unadjusted upper bound.

Proposition 3

$\theta^l_{100X} > \theta^l_{100}$ if and only if X satisfies (9).

Proposition 3 provides a necessary and sufficient condition for the adjusted lower bound to be greater than the unadjusted lower bound. This proposition is illustrated in the first fictional trial from Section 5 using X_1, where $\pi_{10} > (1 - \gamma_0)$ and $\pi_{11} < (1 - \gamma_1)$. As shown in the upper left panel of Figure 1, (9) is satisfied because the dashed line intersects the bottom of the unit square whereas the dotted line intersects the right side of the unit square. In contrast, for X_1 in the second ficitious trial (9) is not satisfied because the dashed and dotted lines in the lower left panel of Figure 1 both intersect the right side of the unit square; thus the adjusted lower bound equals the unadjusted lower bound.

Additional insight regarding criteria (8) and (9) can be obtained by constructing principal strata based on cross-classification of $\{S_i(0), S_i(1)\}$ as well as $\{Y_i(0), Y_i(1)\}$. Further details regarding this approach are given in Web Appendix B.

It follows immediately from Propositions 2 and 3 that if (8) and (9) both hold then the adjusted bounds are strictly contained within the unadjusted bounds and if neither hold, the adjusted and unadjusted bounds are equal. Note while it was assumed that X was a binary covariate, Propositions 1–3 can immediately be extended to any categorical baseline covariate with a finite number of levels k. Specifically, suppose adjusted bounds are calculated within each of the k strata and weighted averages of these bounds are computed analogous to (7). Then Proposition 1 will hold, and if there exists any two levels of X that satisfy either (8) or (9), then either Proposition 2 or Proposition 3 will hold respectively.

7. Inference

The unadjusted and adjusted bounds can be consistently estimated by plugging in consistent estimators of the component parameters of the bounds. Specifically, let

$$\hat{\gamma} = \min \left\{ \frac{\sum (1-S_i)(1-Z_i)/\sum(1-Z_i)}{\sum(1-S_i)Z_i/\sum Z_i}, 1 \right\}$$

and

$$\hat{\gamma}_x = \min \left\{ \frac{\sum (1-S_i)(1-Z_i)I(X_i=x)/\sum(1-Z_i)I(X_i=x)}{\sum(1-S_i)Z_iI(X_i=x)/\sum Z_iI(X_i=x)}, 1 \right\},$$

where $\sum = \sum_{i=1}^{n}$ and $I()$ is the usual indicator function. Likewise, let $\hat{\varphi_x} = \Sigma(1 - S_i)(1 - Z_i)I(X_i = x)/\Sigma(1 - S_i)(1 - Z_i)$, $\hat{\pi_z} = \Sigma Y_i(1 - S_i)I(Z_i = z)/\Sigma(1 - S_i)I(Z_i = z)$, and $\hat{\pi_{zx}} = \Sigma Y_i(1 - S_i)I(Z_i = z, X_i = x)/\Sigma(1 - S_i)I(Z_i = z, X_i = x)$. In the data example presented in Section 8 below, individuals drop out of the study prior to evaluation of Y_i so that the simple moment estimators $\hat{\pi_z}$ and $\hat{\pi_{zx}}$ given above will not be possible to compute. We ignore this complication for now and will return to this issue below.

Consistent estimators of the unadjusted bounds are $\hat{CE}^l = \hat{\theta}^l_{100} - \hat{\theta}_{000}$ and $\hat{CE}^u = \hat{\theta}^u_{100} - \hat{\theta}_{000}$ where $\hat{\theta}^l_{100}$ and $\hat{\theta}^u_{100}$ are obtained by plugging in $\hat{\pi_1}$ and $\hat{\gamma}$ into (3) and $\hat{\theta}_{000} = \hat{\pi_0}$. Similarly, letting $\hat{\theta}^l_{100x}$ and $\hat{\theta}^u_{100x}$ denote (6) evaluated at $\hat{\pi_{1x}}$ and $\hat{\gamma}_x$, consistent estimators of the unadjusted bounds are $\hat{CE}^l_X = \hat{\theta}^l_{100X} - \hat{\theta}_{000}$ and $\hat{CE}^u_X = \hat{\theta}^u_{100X} - \hat{\theta}_{000}$ where $\hat{\theta}^l_{100X}$ and $\hat{\theta}^u_{100X}$ are obtained by plugging in $\hat{\varphi_x}$, $\hat{\theta}^u_{100x}$, and $\hat{\theta}^l_{100x}$ into (7).

These estimated bounds reflect ignorance due to partial identifiability but do not account for uncertainty due to sampling variability; such uncertainty intervals can be constructed using the approach of Imbens and Manski (2004) and Vansteelandt et al. (2006). In particular, Lee (2009) proved the estimated unadjusted bounds $\hat{CE}^l$ and $\hat{CE}^u$ are asymptotically normal under standard regularity conditions and provided that $\Pr[S_i(1) < S_i(0)] > 0$, i.e., treatment has a protective effect on the intermediate variable S_i. Under these conditions, a $(1 - a) \times 100\%$ pointwise uncertainty interval for *CE* is given by

$$\left[\hat{CE}^l - c^*_{\alpha/2} \hat{var}\{\hat{CE}^l\}^{1/2}, \quad \hat{CE}^u + c^*_{\alpha/2} \hat{var}\{\hat{CE}^u\}^{1/2} \right] \quad (10)$$

where $c^*_{\alpha/2}$ can be computed using equation (4.3) of Vansteelandt et al. (2006), and $\hat{var}\{\hat{CE}^l\}$ and $\hat{var}\{\hat{CE}^u\}$ are consistent estimators of $var\{\hat{CE}^l\}$ and $var\{\hat{CE}^u\}$. As $n \to \infty$, the interval (10) will contain *CE* with probability $(1 - a)$. A pointwise uncertainty interval based on the estimated adjusted bounds can be constructed analogously.

8. Illustration

The Breastfeeding, Antiretroviral, and Nutrition (BAN) study was a randomized clinical trial to assess interventions for the prevention of breast milk transmission of HIV in 2369 HIV infected mothers and their infants in Lilongwe, Malawi (Chasela *et al.* 2010). There were three arms in the BAN study: daily antiretroviral therapy (ART) for the infant, daily ART for the mother, or control. While the primary analysis of the study considered comparisons of both ART arms to control, we will focus on comparing the infant ART and control arms only. The primary endpoint of BAN was infant HIV infection by $\tau = 28$ weeks, so we let $Y_i = 1$ if the infant was HIV positive by 28 weeks and $Y_i = 0$ otherwise. Per protocol, infants who died or were infected in the first two weeks post-treatment were to be excluded from the primary analysis. Let $S_i = 1$ if the infant became HIV positive or died by $\tau_0 = 2$ weeks, $S_i = 0$ otherwise. An analysis that compares randomization groups conditional on $S_i = 0$ is subject to selection bias. Instead we let the target of inference be *CE*, the change in risk of infection by 28 weeks due to infant ART in infants who would be HIV negative and alive at two weeks regardless of randomization assignment. Below estimated unadjusted bounds for *CE* are compared with estimated adjusted bounds based on a binary variable X_i indicating low birth weight (< 2.5 kg), i.e., $X_i = 1$ if the infant was of low birth weight, 0 otherwise.

Table 2 presents data on *S* and *X* from BAN by randomization arm. The proportion of infants who die or become HIV infected by two weeks is somewhat higher in the control group (5.7% versus 4.6%), although the difference is not statistically significant (Fisher's exact test two-sided $p = 0.35$). That the proportion was higher in the control group supports the monotonicity assumption. Because some mother-infant pairs dropped out of the study prior to 28 weeks, the simple moment estimators of π_z and π_{zx} given in Section 7 cannot be computed. In the absence of competing risks and informative censoring, π_x and π_{zx} can be consistently estimated using the Kalpan-Meier estimator. However, in the BAN study death and breastfeeding cessation prior to HIV infection were competing risks. Therefore the Aalen-Johansen estimator (Aalen and Johansen 1978) for the cumulative incidence function can be used to estimate π_z and π_{zx}; see Long and Hudgens (2012) for additional details. The Aalen-Johansen estimates (ignoring *X*) were $\hat{\pi}_0 = 0.0581$ and $\hat{\pi}_1 = 0.0141$. Using these estimates, the estimated unadjusted lower and upper bounds of *CE* are

$[\hat{CE}^l, \hat{CE}^u] = [-0.0556, -0.0438]$. Stratifying by *X*, the Aalen-Johansen estimators for each randomization arm were $\hat{\pi}_{00} = 0.0609$, $\hat{\pi}_{01} = 0.0233$, $\hat{\pi}_{10} = 0.0107$, and $\hat{\pi}_{11} = 0.0604$. Therefore the estimated adjusted lower and upper bounds are

$[\hat{CE}^l_X, \hat{CE}^u_X] = [-0.0482, -0.0431]$.

In this example, the estimated adjusted lower bound is greater than the estimated unadjusted lower bound, which is expected based on Proposition 3 and the fact X_i satisfies (9) empirically, i.e., $\hat{\pi}_{10} > (1 - \hat{\gamma}_0)$ and $\hat{\pi}_{11} < (1 - \hat{\gamma}_1)$. On the other hand, the estimated adjusted upper bound is greater than the estimated unadjusted upper bound, empirically contradicting Proposition 1. That the estimated adjusted upper bound is not less than the estimated unadjusted upper bound is actually not surprising based on Proposition 2 and the fact X_i does not satisfy (8) empirically, in particular $\hat{\pi}_{1x} < \hat{\gamma}_x$ for $x = 0, 1$. This suggests the adjusted and unadjusted upper bounds are equal, in which case no particular ordering between the

adjusted and unadjusted estimators might be expected. Grilli and Mealli (2008) reported a similar finding in an analysis of employment data of university students. This apparent contradiction between the estimated bounds and Proposition 1 suggests alternative adjusted estimators of the bounds, namely $\widetilde{CE}_X^l = \max\{\hat{CE}^l, \hat{CE}_X^l\}$ and $\widetilde{CE}_X^u = \min\{\hat{CE}^u, \hat{CE}_X^u\}$. These estimators are consistent for CE_X^l and CE_X^u and by construction will always empirically satisfy Proposition 1, i.e., will always be at least as narrow as the estimated unadjusted bounds. For the BAN data, $[\widetilde{CE}_X^l, \widetilde{CE}_X^u] = [-0.0482, -0.0438]$. That is, by adjusting for the baseline covariate X indicating lower birth weight, the estimated bounds on CE are 63% narrower than the estimated unadjusted bounds.

Uncertainty intervals corresponding to the estimated unadjusted and adjusted bounds were computed using (10) with bootstrap variance estimates (as in Long and Hudgens 2012). For the unadjusted bounds, the estimated 95% pointwise uncertainty interval equals [−0.076, −0.025]. Corresponding to the estimated adjusted bounds $[\widetilde{CE}_X^l, \widetilde{CE}_X^u]$ based on the low birth weight indicator, the estimated 95% pointwise uncertainty interval equals [−0.069, −0.024], i.e., adjusting for low birth weight also results in a narrower estimated uncertainty interval.

9. Identifiability

As noted at the end of Section 3, in the absence of covariate X, CE is identifiable if and only if one of the following three conditions occur: $\gamma = 1$, $\pi_1 = 1$, or $\pi_0 = 0$. When at least one of these conditions holds, CE is identifiable and $\theta_{100}^u = \theta_{100}^l$, i.e., the bounds collapse to a single point. In this section we consider conditions under which adjusting for the binary covariate X renders CE identifiable in the sense that the adjusted bounds collapse to a point, i.e., $CE_X^l = CE_X^u$. By the form of the adjusted bounds given in (7) and the assumption $\varphi_x > 0$ for $x = 0, 1$, it follows that the adjusted bounds yield a single point if and only if

$$\gamma_x = 1, \pi_{1x} = 1, \text{ or } \pi_{1x} = 0 \quad (11)$$

for $x = 0$ and $x = 1$.

Ding et al. (2011) also considered identifiability of a principal effect when outcomes are truncated by death, which is mathematically identical to the problem considered here. In addition to Assumptions 1 and 2 above, Ding et al. provided two additional assumptions which are sufficient for identifiability: (i) $X_i \perp Y_i | \{S_i(0), S_i(1), Z_i\}$ and (ii) $\Pr[X_i = x | S_i(0) = S_i(1) = 0]$ $\Pr[X_i = x | S_i(0) = 1, S_i(1) = 0]$. Unfortunately, assumption (i), which under Assumption 1 can be equivalently stated as $X_i \perp Y_i(z) | \{S_i(0), S_i(1)\}$ for $z = 0, 1$, is in general not subject to empirical test based on the observable data. Ding et al. also gave sufficient identifiability conditions that do not require (i) but instead require that X takes on at least three levels or is continuous and that the mean of Y satisfies a particular linear model.

In contrast, condition (11) can be assessed from the observable data because γ_x and π_{1x} are identifiable under Assumptions 1 and 2 only. Moreover, (11) suggests a strategy for

selecting X. In particular, if a covariate X can be found such that (11) holds for $x = 0, 1$, then CE will be identifiable. If no such covariate is available, then selecting X such that (11) approximately holds for $x = 0, 1$ should yield adjusted bounds with width close to zero. For instance, in the MTCT from the previous section, the low birth weight indicator covariate X yields $\hat{\gamma}_0 = 1.000$, i.e., (11) empirically holds for $x = 0$; while (11) does not hold empirically for $x = 1$, $\hat{\pi}_{11} = 0.0604$ is not too far from zero and indeed the birth weight adjusted bounds for CE are substantially narrower than the unadjusted bounds.

Finally, we note two special cases where X_i identifies CE. First, suppose $X_i = 1$ if and only if $S_i(0) = S_i(1) = 0$, i.e., X_i is a perfect predictor of membership in the NI principal stratum. Then trivially CE is identifiable under Assumption 1 alone by the stratum of individuals with $X_i = 1$. This first case is related to the "principal score," i.e., the probability an individual is within a principal stratum conditional on one or more covariates (Jo and Stuart, 2009). In practice, principal scores are not known but predicted based on fitted models using the observed data. For example, in the MTCT setting a model for $\Pr[S_i(0) = S_i(1) = 0|X_i = x]$ can be fit using data from infants assigned $Z_i = 0$ because under Assumptions 1 and 2 such infants are in the NI stratum if and only if $S_i = 0$. If a set of one or more baseline covariates (not necessarily binary or even discrete) can be found such that the principal scores for the NI stratum equal zero or one for each individual, then CE is identified by the stratum of individuals with principal scores equal to one. In practice this may not be possible; however, if covariates can be found such that the principal scores for the NI stratum are all close to zero or one, i.e., the principal scores are highly predictive of NI stratum membership, then the adjusted bounds constructed by stratifying on a dichotomization of the principal score should have width near zero. For the second special case, suppose $Y_i = 1$ if and only if $X_i = 1$, i.e., X_i is a perfect predictor of Y_i. Then $\pi_{10} = 0$ and $\pi_{11} = 1$ implying (11) holds for $x = 0$ and $x = 1$ and therefore $CE_X^l = CE_X^u$. In settings where Z_i has an effect on Y_i and Z_i is assigned randomly, no such perfect predictor X_i will exist (because X_i is measured pre-randomization), such that the second case seems to have little practical implication.

10. Discussion

In this paper we considered whether bounds on principal effects can be improved by adjusting for a categorical baseline covariate. Necessary and sufficient conditions were derived for which the covariate adjusted bounds will be sharper (i.e., narrower) than the unadjusted bounds. The methods were illustrated using data from a study of interventions to prevent mother-to-child transmission of HIV through breastfeeding. Using a baseline covariate indicating low birth weight, the estimated adjusted bounds for the principal effect of interest were 63% narrower than the estimated unadjusted bounds.

The veracity of the analysis of the BAN trial depends on several key assumptions. Assumption 1, independent treatment assignment, is reasonable because infants were randomly assigned treatment. Assumption 2, monotonicity, implies that the treatment is no worse than control for any individual in terms of the intermediate variable S. As mentioned in Section 8 this assumption is supported by the observed data. Additionally, the BAN study principal investigator, Dr. Charles van der Horst, has indicated that monotonicity is reasonable (personal communication). However, in other settings this assumption may be

unrealistic. For example, monotonicity might be considered questionable in an analysis comparing the two active arms of the BAN trial, i.e., infant ART versus maternal ART. In such settings, bounds that do not require the monotonicity assumption would be needed.

There are several possible avenues of related future research. One possible direction entails studying how adjusting for covariates affects the efficiency with which the bounds are estimated. For example, is it possible that certain covariates could be advantageous to adjust for with respect to sharpening bounds, but disadvantageous to adjust for in terms of efficiency? In the presence of multiple baseline covariates, future research could explore a formal approach for determining which covariates to adjust for (or include in a principal score model) and how to subsequently draw valid inference accounting for the covariate selection process. Related to Assumption 2, future research could examine relations between adjusted and unadjusted bounds when monotonicity is assumed for the primary endpoint Y.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

References

Aalen O, Johansen S. An empirical transition matrix for non-homogeneous Markov chains based on censored observations. Scandinavian Journal of Statistics. 1978; 5(3):141–150.

Angrist JD, Imbens GW, Rubin DB. Identification of causal effects using instrumental variables (disc: P456–472). Journal of the American Statistical Association. 1996; 91:444–455.

Chasela C, Hudgens M, Jamieson D, Kayira D, Hosseinipour M, Kourtis A, Martinson F, Tegha G, Knight R, Ahmed Y, Kamwendo D, Hoffman I, Ellington S, Kacheche Z, Soko A, Wiener J, Fiscus S, Kazembe P, Mofolo I, Chigwenembe M, Sichali D, van der Horst C. Maternal or infant antiretroviral drugs to reduce HIV-1 transmission. New England Journal of Medicine. 2010; 362(24):2271–2281. [PubMed: 20554982]

Cheng J, Small DS. Bounds on causal effects in three-arm trials with non-compliance. Journal of the Royal Statistical Society: Series B (Statistical Methodology). 2006; 68:815–836.

Ding P, Geng Z, Yan W, Zhou XH. Identifiability and estimation of causal effects by principal stratification with outcomes truncated by death. Journal of the American Statistical Association. 2011; 106:1578–1591.

Frangakis CE, Rubin DB. Principal stratification in causal inference. Biometrics. 2002; 58:21–29. [PubMed: 11890317]

Grilli L, Mealli F. Nonparametric bounds on the causal effect of university studies on job opportunities using principal stratification. Journal of Educational and Behavioral Statistics. 2008; 33(1):111–130.

Hudgens MG, Halloran ME. Causal vaccine effects on binary post-infection outcomes. Journal of the American Statistical Association. 2006; 101:51–64. [PubMed: 19096723]

Hudgens MG, Hoering A, Self SG. On the analysis of viral load endpoints in HIV vaccine trials. Statistics in Medicine. 2003; 22:2281–2298. [PubMed: 12854093]

Imbens GW, Manski CF. Confidence intervals for partially identified parameters. Econometrica. 2004; 72(6):1845–1857.

Jo B, Stuart E. On the use of propensity scores in principal causal effect estimation. Statistics in Medicine. 2009; 28(23):2857–2875. [PubMed: 19610131]

Joffe M. Principal stratification and attribution prohibition: Good ideas taken too far. International Journal of Biostatistics. 2011; 7(1):Article 35. [PubMed: 22049269]

Lee DS. Training, wages, and sample selection: Estimating sharp bounds on treatment effects. Review of Economic Studies. 2009; 76(3):1071–1102.

Long, DM. Doctoral Dissertation. Department of Biostatistics, University of North Carolina; Chapel Hill: 2012. Causal Inference and Principal Stratification: Competing Risks, Bounds, and Surrogates.

Long DM, Hudgens MG. Comparing competing risk outcomes within principal strata, with application to studies of mother-to-child transmission of HIV. Statistics in Medicine. 2012; 31(27):3406–3418. [PubMed: 22927321]

Mealli, F.; Pacini, B. Working Paper 4. Department of Statistics, University of Florence; 2012. Using secondary outcomes and covariates to sharpen inference in randomized experiments with noncompliance.

Nolen TL, Hudgens MG. Randomization-based inference within principal strata. Journal of the American Statistical Association. 2011; 106:581–593. [PubMed: 21987597]

Shepherd B, Gilbert P, Dupont C. Sensitivity analyses comparing time-to-event outcomes only existing in a subset selected postrandomization and relaxing monotonicity. Biometrics. 2011; 67(3):1100–1110. [PubMed: 21114663]

VanderWeele T. Principal stratification – uses and limitations. The International Journal of Biostatistics. 2011; 7(1):Article 28. [PubMed: 21841939]

Vansteelandt S, Goetghebeur E, Kenward MG, Molenberghs G. Ignorance and uncertainty regions as inferential tools in a sensitivity analysis. Statistica Sinica. 2006; 16:953–979.

Zhang JL, Rubin DB. Estimation of causal effects via principal stratification when some outcomes are truncated by "death". Journal of Educational and Behavioral Statistics. 2003; 28:353–368.
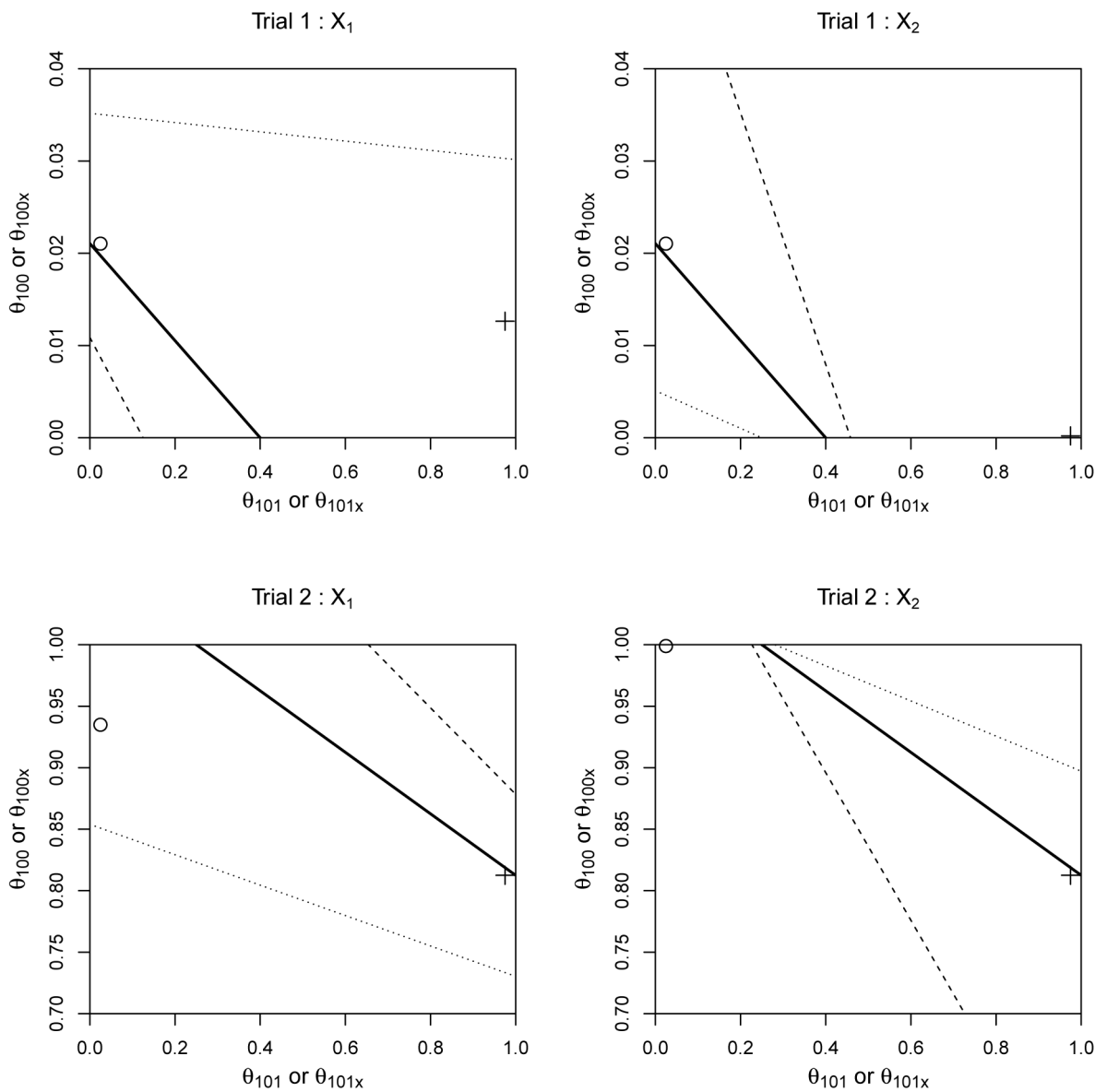
Figure 1.
Graphical depiction of bounds for the two fictional MTCT trials discussed in Section 5 with two different binary covariates X_1 and X_2. The solid lines depict equation (2) with $\pi_1 = 0.02$ and $\gamma = 0.95$ in the upper panels and $\pi_1 = 0.85$ and $\gamma = 0.8$ in the lower panels. The $\cdots$ ($-$ $-$ $-$) lines represent (5) for $X = 0$ ($X = 1$). The vertical value of $\bigcirc$ ($+$) corresponds to θ^u_{100X} (θ^l_{100X}).

Table 1

Probabilities from the fictional trials described in Section 4 stratified by X_1 and X_2.

Trial 1	X_1		X_2	
	0	**1**	**0**	**1**
γ_x	0.995	0.920	0.980	0.880
π_{1x}	0.035	0.010	0.005	0.055
λ_x	0.400	0.600	0.700	0.300
φ_x	0.419	0.581	0.722	0.278

Trial 2	X_1		X_2	
	0	**1**	**0**	**1**
γ_x	0.890	0.740	0.875	0.625
π_{1x}	0.760	0.910	0.910	0.710
λ_x	0.400	0.600	0.700	0.300
φ_x	0.445	0.555	0.766	0.234

Note by Bayes Theorem φ_x must equal $\gamma_x \lambda_x / (\gamma_0 \lambda_0 + \gamma_1 \lambda_1)$ for $x = 0, 1$; values of φ_x reported in the table have been rounded.

Table 2

Data from the BAN study stratified by X (low birth weight < 2.5 kg).

	Control ($Z = 0$)			Treatment ($Z = 1$)		
	$X = 0$	$X = 1$	Total	$X = 0$	$X = 1$	Total
$S = 1$	28	10	38	36	3	39
$S = 0$	584	46	630	751	62	813