University of Pennsylvania
ScholarlyCommons

Statistics Papers

Wharton Faculty Research

Large Sample Bounds on the Survivor Average Causal Effect in the Presence of a Binary Covariate with Conditionally Ignorable Treatment Assignment

Michael H. Freiman
U.S. Census Bureau

Dylan. S. Small
University of Pennsylvania

Follow this and additional works at: http://repository.upenn.edu/statistics_papers

Part of the Physical Sciences and Mathematics Commons

Recommended Citation

Large Sample Bounds on the Survivor Average Causal Effect in the Presence of a Binary Covariate with Conditionally Ignorable Treatment Assignment

Abstract

A common problem when conducting an experiment or observational study for the purpose of causal inference is "censoring by death," in which an event occurring during the experiment causes the desired outcome value – such as quality of life (QOL) – not to be defined for some subjects. One approach to this is to estimate the Survivor Average Causal Effect (SACE), which is the difference in the mean QOL between the treated and control arms, considering only those individuals who would have had well-defined QOL regardless of whether they received the treatment of interest, where the treatment is imposed by the researcher in an experiment or by the subject in the case of an observational study. Zhang and Rubin [5] (Estimation of causal effects via principal stratification when some outcomes are truncated by "death". J Educ Behav Stat 2003;28:353–68) have proposed a methodology to calculate large sample bounds – bounds on the SACE that assume that the exact QOL distribution for each arm is known or that the finite sample size can be ignored – in the case of a randomized experiment. We examine a modification of these bounds in the case where a binary covariate describing each of the subjects is available and assignment to the treatment or control group is ignorable conditional on the covariate. Using a dataset involving an employment training program, we find that the use of the covariate does not substantially change the bounds in this case, although it does weaken the assumptions about the sample and thus make the bounds more widely applicable. However, simulations show that the use of a binary covariate can in some cases dramatically narrow the bounds. Extensions and generalizations to more complicated variants of this situation are discussed, although the amount of computation increases very quickly as the number of covariates and the number of possible values of each covariate increase.

Disciplines
Physical Sciences and Mathematics

Research Article

Michael H. Freiman* and Dylan S. Small

Large Sample Bounds on the Survivor Average Causal Effect in the Presence of a Binary Covariate with Conditionally Ignorable Treatment Assignment

Abstract: A common problem when conducting an experiment or observational study for the purpose of causal inference is "censoring by death," in which an event occurring during the experiment causes the desired outcome value – such as quality of life (QOL) – not to be defined for some subjects. One approach to this is to estimate the Survivor Average Causal Effect (SACE), which is the difference in the mean QOL between the treated and control arms, considering only those individuals who would have had well-defined QOL regardless of whether they received the treatment of interest, where the treatment is imposed by the researcher in an experiment or by the subject in the case of an observational study. Zhang and Rubin [5] (Estimation of causal effects via principal stratification when some outcomes are truncated by "death". J Educ Behav Stat 2003;28:353–68) have proposed a methodology to calculate large sample bounds – bounds on the SACE that assume that the exact QOL distribution for each arm is known or that the finite sample size can be ignored – in the case of a randomized experiment. We examine a modification of these bounds in the case where a binary covariate describing each of the subjects is available and assignment to the treatment or control group is ignorable conditional on the covariate. Using a dataset involving an employment training program, we find that the use of the covariate does not substantially change the bounds in this case, although it does weaken the assumptions about the sample and thus make the bounds more widely applicable. However, simulations show that the use of a binary covariate can in some cases dramatically narrow the bounds. Extensions and generalizations to more complicated variants of this situation are discussed, although the amount of computation increases very quickly as the number of covariates and the number of possible values of each covariate increase.

Keywords: censoring by death; causal inference; principal stratification; large sample bounds

1 Introduction

A possible aim of a medical study is to assess a numerical value of quality of life (QOL) at a fixed time point. Interpretation of the results of such a study may be complicated by the phenomenon of *censoring by death*, in addition to other common phenomena such as non-ignorability of treatment assignment in an observational study. Censoring by death occurs when a patient dies before the measurement of QOL is taken. Such a patient does not have a well-defined QOL that is a real number, and this complicates the problem of finding a causal measurement that accurately represents the treatment's effect on QOL. This phenomenon appears in other biological contexts as well. A few examples follow:

- An experiment is run to test a treatment to mitigate the effects of HIV, with patients' viral loads measured at a fixed point in time. Patients who do not contract HIV do not have a viral load. Hence, Gilbert et al. [1] view this as a censoring by death problem, although it could be argued that a reasonable approach is to treat viral loads of HIV-negative patients as 0.

*Corresponding author: Michael H. Freiman,** Center for Disclosure Avoidance Research, U.S. Census Bureau, 4600 Silver Hill Road, Washington, DC 20233, USA, E-mail: michael.freiman@census.gov
Dylan S. Small, Department of Statistics, University of Pennsylvania, Philadelphia, PA 19104, USA, E-mail: dsmall@wharton.upenn.edu

- Egleston et al. [2] report the results of an observational study to examine the effect of vision loss on older patients' probability of experiencing "worsened emotional distress," with some participants dying before the study ends. This is somewhat different from our primary focus, as in this case, the variable of interest (worsened emotional distress) is binary, but otherwise it shares the same basic premise.
- An observational study or experiment examines treatments to lessen the frequency of drug use within the next year among a high-risk population. Subjects who are imprisoned, hospitalized, placed in a residential treatment facility or otherwise institutionalized for part or all of the post-treatment study period will have their opportunities to use drugs dramatically reduced or eliminated during the time of institutionalization. Here, institutionalization is somewhat akin to "death," although in this case, the key problem is not so much removal of those who "die" from the population as the property that institutionalization prevents subjects from exhibiting the drug usage patterns that they would have in the absence of institutionalization, and these are likely to be the patterns of primary interest. A complicating factor is that an individual may be institutionalized for only part of the study period; to make the analogy with censoring by death precise, it would be necessary to consider all individuals with any period of institutionalization during the relevant time period to have "died." This example is considered at length by McCaffrey et al. [3] and Griffin et al. [4], the latter of which defines institutionalization as spending at least one day during the study period "in a controlled environment."

Analogous examples also include studies to estimate the effect of an educational intervention on graduation test scores (where a dropout's test score is censored by death), or measuring the effect of an employment training program on hourly wages (where an unemployed person's wages are censored by death). Zhang and Rubin [5] point out that in some cases, it may be useful to consider QOL to assume values in the set $R \cup \{D\}$, where D is a non-real number denoting death.

In a case where everyone has a real QOL and treatment assignment is completely random, the causal effect of the treatment is easily estimated as the difference between the mean QOL of the treated group and the mean QOL of the control group. This measure is unbiased, if the study is an experiment with randomization performed properly, but does not apply in a censoring by death context or in an observational study. A tempting alternative is to eliminate all of those who die and calculate the difference in mean QOL between the treated survivors and the control survivors, a method that is sometimes called the "restricted analysis," as it analyzes the data in a way that is restricted to survivors. However, the restricted analysis is problematic, as the treatment's effects on mortality obscure somewhat the effect on QOL, a concern that will be examined briefly in Section 2.

This paper describes an established metric – the Survivor Average Causal Effect (SACE) – that allows the measurement of a causal effect on QOL in the presence of censoring by death and examines how well that metric can be estimated in the case in which a binary covariate is also available. Zhang and Rubin [5], Imai [6], and Chiba and VanderWeele [7] have considered bounds for the SACE in randomized experiments. The contribution of our paper is to consider bounds for the SACE in the context of an observational study in which there may not be perfect randomization, but in which the researcher is willing to make the assumption that treatment (or, more generally, the variable whose effect is to be evaluated) is randomly assigned conditional on a covariate (conditionally ignorable treatment assignment).

Section 2 describes the notation and assumptions that we will use in considering the SACE. Section 3 describes a dataset based on the National Supported Work Demonstration (NSW), considered in LaLonde [8], that has the structure and available covariates necessary for the analysis to be considered. Section 4 recaps the large sample bounds for the SACE from Zhang and Rubin [5]. In Section 5, we extend Zhang and Rubin's [5] methods to the case where a single binary covariate is available and consider the changes in formulas and assumptions that this creates. In Section 6, we apply the methods to data from an experiment on the effect of an employment training program on future earnings and find that in this case covariates may have little effect on the numerical values of the bounds that are derived, even though they may weaken

the assumptions underlying the bounds. In Section 7, we demonstrate the potential substantial utility of considering a binary covariate in a finding large sample bounds, using a simulated example. In Section 8, we conclude, discuss some other views on the topic, and suggest some extensions.

2 Notation and assumptions

A framework under which to consider the problem of estimating a causal effect of treatment on QOL arises from the idea of *potential outcomes* – that each subject has an outcome for each possible arm of the experiment in which he may be placed. In most experiments, each subject has two potential outcomes, a treated outcome and a control outcome, with each potential outcome consisting of a mortality outcome (life or death) and a QOL in the case of survival. For a given individual j, we let T_j be an indicator variable for treatment. We let S_0 be an indicator variable for survival given the control and S_1 be an indicator variable for survival given the treatment. Similarly, let Y_0 be the QOL of a given individual given the control and Y_1 be the QOL given treatment; for $i \in \{0, 1\}$, Y_i is defined only if $S_i = 1$. For each subject, one of the potential outcomes is observed, while the other is counterfactual and cannot be observed. In the case of an observational study performed with the hope of determining the effect of a binary predictor variable on a numerical response variable, the same notation may be used, with one of the values of the predictor designated as 0 (hence a survival outcome of S_0 and a QOL of Y_0) and one of the values of the predictor designated as 1 (hence a survival outcome of S_1 and a QOL of Y_1). For the purpose of Assumption 3, it is ideal to designate as 1 the value of the predictor that is hypothesized to lead to a higher average QOL; in many situations, this will be readily apparent.

Frangakis and Rubin [9] propose to solve the problem mentioned in Section 1 using the approach of *principal stratification*. It is assumed that a patient's survival status is a function solely of whether he receives the treatment. Hence each patient falls into one of four groups, called *principal strata*, depending on his potential survival outcomes under the two arms. The principal strata are *LL* (those who would survive regardless of arm), *LD* (those who would survive only if they receive the treatment), *DL* (those who would survive only if they receive the control), and *DD* (those who would not survive under any circumstances), with the first letter indicating life or death under treatment and the second indicating the same under the control. Pearl [10] refers to these four groups as "always healthy," "responders" (because they respond to the treatment), "hurt" (because they are hurt by the treatment), and "doomed;" other authors have proposed different names, depending on the specifics of the application at hand. The particular medical context Pearl examined was somewhat different from ours; in ours, the *LL* stratum might more reasonably be termed "always survivors."

Underlying any attempt to find a causal effect in a principal stratification framework is the following assumption.

Assumption 1 (Stable Unit Treatment Value Assumption, or SUTVA, Rubin [11]). For a given unit, the observed mortality and QOL (or other outcome of interest) are dependent only on the treatment that the unit received. In particular, the treatment is well-defined (not a treatment with different "versions"), and outcomes are independent both of other units' treatment assignments and of the treatment assignment mechanism.

To remedy the difficulty mentioned above of comparing the *LD* group with the *DL* group, Rubin [12] introduced the *SACE*, which measures the difference between QOL under treatment and QOL under control among only the *LL* stratum, which is the only stratum for which this difference can be defined. Following Chiba and VanderWeele [7], we may define the SACE as

$$\text{SACE} = E(Y_1 - Y_0 | S_1 = S_0 = 1).$$

The primary difficulty in measuring the SACE is that the principal stratum of a patient cannot be observed – if a patient is treated and survives, he may be in either the *LL* or *LD* stratum, while a control patient who survives

may be in either the *LL* or *DL* stratum. This suggests the difficulty with the restricted analysis in which the QOL of the survivors only is used to estimate the treatment effect on QOL, as the subjects from the treated group in the study are considered only if they are from two of the principal strata, while the subjects from the control group are considered only if they are from a different set of two principal strata. Since these do not represent the same population, no causal effect can be inferred, and bias may be introduced by differences between the *LD* and *DL* strata, either in size or in QOL.

Our goal in this paper is to consider *large sample bounds* on the SACE, and how they are affected by the need to condition on a binary covariate *X* for ignorability of treatment assignment to hold. Large sample bounds are bounds that assume that the only unknown information necessary to compute the SACE is the list of principal strata of the individual subjects. In particular, it is assumed that the sample size is large enough that effects of a finite sample can safely be ignored or that the distribution of QOL for each arm is known explicitly.

The bounds depend on what assumptions are made. Whereas SUTVA underlies any attempt to find causal effects within the principal stratification framework, Assumptions 2 and 3 may be made or not made, and these decisions will affect the results of the analysis.

Assumption 2 (Monotonicity Assumption). The *DL* principal stratum is empty.

Restated somewhat, Assumption 2 says that a treatment – or, in an observational study, one of the values of the predictor variable – should not kill someone who would otherwise have survived. This assumption seems reasonable in some cases, although of course there are many situations in which it is not true, such as a treatment with potentially fatal side effects.

Assumption 3 (Ranked Average Score Assumption, or RASA). The mean QOL for the members of the *LL* stratum in either arm is greater than or equal to the mean QOL for the members of the other surviving stratum (*LD* for treatment, *DL* for control) in the same arm.

The *LL* stratum seems to be the healthiest stratum, as evidenced by its survival regardless of treatment arm, so RASA also seems reasonable, although it too may not be universally true. If one or more covariates are available, then we may refine RASA into Assumption 4.

Assumption 4 (Stratified Ranked Average Score Assumption, or SRASA). Suppose certain covariates are available in addition to the survival and QOL outcomes. Let *S* be a subset of the potential population, where the definition of *S* depends on criteria that are entirely determined by the values of the covariates. Then, considering only the members of *S*, the mean QOL for the members of the *LL* stratum in the treatment arm is greater than or equal to the mean QOL for the members of the *LD* stratum in the treatment arm. Similarly, considering only the members of *S*, the mean QOL for the members of the *LL* stratum in the control arm is greater than or equal to the mean QOL for the members of the *DL* stratum in the control arm.

Restated somewhat, SRASA says that if a sub-universe is defined based solely on values of available covariates, then RASA applies to that sub-universe.

In this paper, we focus on the effect of the knowledge of a single binary covariate *X*, in which case Assumption 4 may be restated as "RASA holds when only those subjects for whom $X = 0$ are considered" and "RASA holds when only those subjects for whom $X = 1$ are considered." When such a covariate is available, we may make further assumptions, such as Assumptions 5–7. We assume that for the purpose of computing large sample bounds, the distribution of the covariate is known:

Assumption 5 (Knowledge of Covariate Distribution). The marginal probability of $X = 1$ across both arms of the experiment or observational study, $P_{X=1}$, is known.

Another assumption that is often made in analyzing experiments is completely random treatment assignment.

Assumption 6 (Completely Random Treatment Assignment). Treatment assignment is independent of potential outcomes.

For an observational study, the treatment assignment may not be completely random but might have a similar property after we condition on the covariate X, in which case we can replace Assumption 6 with the weaker Assumption 7.

Assumption 7 (Conditionally Ignorable Treatment Assignment, Rosenbaum and Rubin [13]). Conditional on X, treatment assignment is independent of potential outcomes.

Phrased more loosely, the assumption of conditionally ignorable treatment assignment states that any deviation of a given arm from being representative of the population is entirely explainable by X, as opposed to the random treatment assignment assumption, which states that the whole of each arm is representative of the population. In a well-run experiment, the completely random treatment assignment assumption is often reasonable, but the conditionally ignorable treatment assignment assumption may be useful if there is a possibility of an imbalance between the treated and control arms or if an observational study is being run. If the study is an experiment and the sample is large enough for large sample bounds to apply, then an imbalance of this sort is unlikely to happen by random chance, but may be a result of a problem with randomization, a deliberate attempt by the researcher to weight the two arms differently with respect to the covariate X, or rates of attrition and non-compliance that vary depending on X. Perhaps most importantly, conditional ignorability may be a more reasonable assumption than strong ignorability in the case of an observational study. We will often use Assumption 7 in combination with Assumption 5.

We now introduce some of the notation that will be used in the rest of the paper. Let π_{LL} denote the proportion of the population in the LL stratum, with analogous notation for the other strata. We let P_{CL} denote the proportion of individuals who would live if they received the control, with the obvious adaptations for treated individuals and/or those who die. When a binary covariate X is available, we let $\pi_{LL|X=1}$ denote the proportion in the LL stratum among those for whom $X = 1$, $P_{X=1}$ denote the proportion of the population for whom $X = 1$, $SACE_{X=1}$ denote the SACE among only those people for whom $X = 1$, and $P_{CL|X=1}$ denote the proportion of control individuals who live among the $X = 1$ population, with all notation again adapted in the obvious way for treated and/or non-surviving individuals. We adapt Zhang and Rubin's [5] notation to arrive at $\bar{Y}_{TL|X=1}(\max|p)$ to denote the mean QOL under treatment of a subset of the treated surviving group for whom $X = 1$, where the subset is chosen to be those individuals making up the proportion p of the $X = 1$ population who have the highest QOL, and the obvious modifications to the notation are made when a minimum is desired instead of a maximum, or when we are interested in control individuals or individuals with a different value of X. We acknowledge that, given the conditional probabilities involved, something similar to $P_{L|C,X=1}$ is more natural notation than the $P_{CL|X=1}$ that we have used, and similarly $\bar{Y}_{L|T,X=1}(\max|p)$ is more natural than $\bar{Y}_{TL|X=1}(\max|p)$. These notational decisions were made so that our notation would correspond more closely to Zhang and Rubin's [5].

A largely theoretical concern with Zhang and Rubin's [5] formulas is that sometimes, if the distribution of QOL is specified as a particular unbounded theoretical distribution, such as the normal, then the formulas may also give infinite bounds. In particular, one of the elements of Zhang and Rubin's [5] formulas in the cases without monotonicity is to take the mean of the QOL in one of the tails of the treated or control QOL distribution. The tails are defined so that they contain a certain amount of probability for that particular distribution. If this probability can get arbitrarily close to 0 and the distribution itself is unbounded, then the tail mean can become arbitrarily extreme, leading to the infinite bounds. Since the tail probabilities are decreasing functions of π_{DL}, a value of 0 can be obtained if and only if the tail probability is 0 when π_{DL} is at its maximum. The tail probabilities that arise in Zhang and Rubin's [5] formulas are $\frac{P_{CL}-\pi_{DL}}{P_{TL}}$ and $\frac{P_{CL}-\pi_{DL}}{P_{CL}}$, and these equal 0 when π_{DL} has the potential to equal P_{CL}. Setting the upper bound of π_{DL} greater than or equal to P_{CL}, we see that this happens when

$$\min(P_{CL}, 1 - P_{TL}) \geq P_{CL}$$
$$\Leftrightarrow P_{CL} \leq 1 - P_{TL}$$
$$\Leftrightarrow P_{CL} + P_{TL} \leq 1.$$

3 Motivating example: the National Supported Work Demonstration

For illustrative purposes, we will use data from the NSW, an experiment on the effect of an employment training program, which was conducted in the mid-1970s and is described in LaLonde [8]. The experiment randomly assigned 722 males either to receive the training or not to and recorded their subsequent earnings. Here, unemployment is akin to "death" and the log of earnings in 1978 is used in place of QOL. Some other covariates are available, including whether the individual had a high school diploma. There is some evidence that treatment assignment of those for whom data were recorded was not independent of the covariates, as will be discussed in Section 6. In addition to high school graduation, a covariate is available that indicates whether the person was unemployed in 1975, three years before the measurements of employment and income were taken, and before the treated subjects had the opportunity to participate in the NSW. The individual's marital status was also recorded.

4 Review of large sample bounds without covariates under completely random treatment assignment

In this section, we review Zhang and Rubin's [5] results on large sample bounds without covariates. The first assumption Zhang and Rubin [5] make is that the treatment assignment is completely random. This is an assumption that we aim to relax when a binary covariate is available, and this will be discussed in Section 5 and subsequently.

Zhang and Rubin [5] give large sample bounds for the SACE under a censoring by death setup, which Jemiai [14] proves are sharp. These bounds are estimands – albeit somewhat extreme ones – for the causal effect $E(Y_1 - Y_0 | S_1 = S_0 = 1)$. Zhang and Rubin [5] consider four possible situations, depending on what assumptions are made. In particular, RASA can be made or not made, and the monotonicity assumption can be made or not made, and for each combination of these assumptions, a different set of large sample bounds applies. In presenting Zhang and Rubin's [5] bounds, we provide intuitive explanations for them; more formal derivations are in the original paper.

The general approach followed by Zhang and Rubin [5] varies primarily depending on whether monotonicity is assumed. Under the monotonicity assumption, the sizes of the strata are known, and indeed the survivors in the control arm are known to be in the *LL* stratum, so finding the mean *LL* QOL for the control arm is straightforward, as this is just the overall mean QOL for the control survivors. Thus, if π_{LL} is the proportion of individuals in the *LL* stratum, then the largest possible mean QOL for the treated *LL* group is the mean of the π_{LL} proportion of treated observations with the highest QOL, while the smallest possible mean QOL is the mean of the π_{LL} proportion of treated observations with the lowest QOL. If RASA is also assumed, then all of these statements remain true, with the exception of the lower bound on the mean *LL* treated QOL, which now becomes the mean of all observed treated QOL values. The reasoning here is sufficient to give bounds on the SACE. A consequence of this is that when both monotonicity and RASA are assumed, the lower end of the large sample bounds of the SACE is given by the estimate from the restricted analysis. More formally, the bounds under monotonicity are

$$\left\{ \bar{Y}_{TL}\left(\min\left|\frac{P_{CL}}{P_{TL}}\right|\right) - \bar{Y}_{CL}, \bar{Y}_{TL}\left(\max\left|\frac{P_{CL}}{P_{TL}}\right|\right) - \bar{Y}_{CL} \right\},$$

and the bounds under both monotonicity and RASA are

$$\left\{ \bar{Y}_{TL} - \bar{Y}_{CL}, \bar{Y}_{TL}\left(\max\left| \frac{P_{CL}}{P_{TL}} \right| \right) - \bar{Y}_{CL} \right\}.$$

If monotonicity is not assumed, then the survivors from the control arm are split into two strata – the LL and DL strata – but the stratum memberships and sizes are unknown. However, if the size of one of the strata (in terms of its proportion in the full sample) is assumed known, then the size of all strata can be calculated exactly, and the methods in the previous paragraph can be adapted in a fairly straightforward manner to find bounds on the SACE conditional on this assumption. Zhang and Rubin [5] derive bounds on the size π_{DL} of the DL stratum; in particular,

$$\max(0, P_{CL} - P_{TL}) < \pi_{DL} < \min(P_{CL}, 1 - P_{TL}). \tag{1}$$

They then vary π_{DL} within these bounds. The overall minimum possible SACE is then the minimum, over all possible values of π_{DL}, of the minimum SACE given the value of π_{DL}. The same approach can be used to find the maximum SACE. From this, it follows that if neither assumption is made, the bounds are

$$\left\{ \min_{\pi_{DL}}\left(\bar{Y}_{TL}\left(\min\left| \frac{P_{CL} - \pi_{DL}}{P_{TL}} \right| \right) - \bar{Y}_{CL}\left(\max\left| 1 - \frac{\pi_{DL}}{P_{CL}} \right| \right) \right), \max_{\pi_{DL}}\left(\bar{Y}_{TL}\left(\min\left| \frac{P_{CL} - \pi_{DL}}{P_{TL}} \right| \right) - \bar{Y}_{CL}\left(\max\left| 1 - \frac{\pi_{DL}}{P_{CL}} \right| \right) \right) \right\},$$

and if only RASA is made, the bounds are

$$\left\{ \bar{Y}_{TL} - \max_{\pi_{DL}}\left(\bar{Y}_{CL}\left(\max\left| 1 - \frac{P_{DL}}{P_{TL}} \right| \right) \right), \max_{\pi_{DL}}\left(\bar{Y}_{TL}\left(\max\left| \frac{P_{CL} - \pi_{DL}}{P_{TL}} \right| \right) \right) - \bar{Y}_{CL} \right\}.$$

In both cases, π_{DL} is varied within the range in eq. (1) for the purposes of determining the minima and maxima in the formulas.

It is worth noting that any analysis involving the SACE works only in the event that the existence of an LL stratum is known. If the monotonicity assumption is not made and $P_{TL} + P_{CL} \leq 1$, then it is possible, if highly implausible, that all of the treated survivors are in the LD stratum and all of the control survivors are in the DL stratum. This will usually not be a concern in practice, as applications typically involve circumstances where most of the participants survive, and it is difficult to imagine a treatment that kills everyone who would otherwise have survived, while also saving some who would have died. However, if the numbers in a particular example are such that $P_{TL} + P_{CL} \leq 1$, then the large sample bounds may be infinitely wide.

5 Large sample bounds with a binary covariate under conditionally ignorable treatment assignment

In this section, we compute large sample bounds with a binary covariate under the assumption of conditionally ignorable treatment assignment.

Suppose that X is a binary covariate that is available for all individuals in a dataset of interest. The SACE is a weighted average of the SACE for the $X = 0$ subpopulation and for the $X = 1$ subpopulation, with the weights equal to their relative proportions in the LL group, and this is given by

$$\frac{\pi_{LL|X=1}P_{X=1}\text{SACE}_{X=1} + \pi_{LL|X=0}(1 - P_{X=1})\text{SACE}_{X=0}}{\pi_{LL|X=1}P_{X=1} + \pi_{LL|X=0}(1 - P_{X=1})}. \tag{2}$$

For the purpose of illustration, we use the example in which $X = 1$ denotes that the person in question has a high school diploma and $X = 0$ denotes no diploma.

The simpler case again arises when monotonicity is assumed. In this case, we know what proportion of each of the diploma subgroups is in the LL stratum. Thus it is straightforward to determine what proportion of the LL stratum has a diploma and what proportion does not. Bounds on the SACE for each of the groups

defined by X can be determined by Zhang and Rubin's [5] method, and then the appropriate weighted average can be taken according to the proportion of the *LL* stratum represented by each of the two diploma groups.

When monotonicity is not assumed, it is again the case that if the size of one stratum for each of the two diploma groups were known, then bounds on the SACE within each diploma group could be derived, conditional on the stratum sizes, using Zhang and Rubin's [5] approach. Furthermore, since the sizes of the *LL* strata within each diploma group would then be known, the proportion of the *LL* stratum accounted for by each of the diploma groups could be calculated using Bayes' Theorem, giving the appropriate weights to be used to combine the within-degree-group SACE bounds into overall SACE bounds. Hence, conditional on the known stratum proportions, we can again find bounds on the SACE. To find the unconditional SACE minimum, we simply vary the values of both *DL* stratum sizes within Zhang and Rubin's [5] bounds and take the minimum of the conditional minima. Similarly, to find the unconditional SACE maximum, we vary the values of both *DL* stratum sizes within the bounds and take the maximum of the conditional maxima. Zhang and Rubin's [5] formulas do not lend themselves to analytic optimization on empirical data, so we used an extensive grid search to determine the bounds in practice.

We now examine the formulas for the large sample bounds for the SACE in the presence of binary covariates. Although ordinarily one would start with fewer assumptions and progress to cases with more assumptions, we begin by assuming monotonicity, since the formulas are simpler in that case and the non-monotonicity case is an extension of the monotonicity case.

When only monotonicity is assumed, the proportion $\pi_{LL|X=0}$ of *LL* individuals among the $X = 0$ population is equal to the proportion of survivors $P_{CL|X=0}$ for the control arm of this population. Similarly, $\pi_{LL|X=1} = P_{CL|X=1}$. Hence $\pi_{LL|X=0}$ and $\pi_{LL|X=1}$ are known exactly. The proportion of the total population that is in the *LL* principal stratum and for which $X = 0$ is thus $P(\text{person in } X = 0 \text{ group})P(LL|\text{person in } X = 0 \text{ group}) = (1 - P_{X=1})P_{CL|X=0}$, while the proportion of the total population that is in the *LL* principal stratum and for which $X = 1$ is $P(\text{person in } X = 0 \text{ group})P(LL|\text{person in } X = 0 \text{ group}) = P_{X=1}P_{CL|X=1}$. The overall SACE is smallest when both the $X = 0$ population's SACE and the $X = 1$ population's SACE are minimized, and in this case, the overall SACE is just a weighted average of the $X = 0$ SACE and the $X = 1$ SACE, with weights proportional to the $X = 0$ group and $X = 1$ group's sizes within the *LL* principal stratum. Hence the weights should be proportional to $(1 - P_{X=1})\pi_{LL|X=0}$ and $P_{X=1}\pi_{LL|X=1}$, which in this case equals $(1 - P_{X=1})P_{CL|X=0}$ and $P_{X=1}P_{CL|X=1}$. The minimum overall SACE consistent with the observed data is a weighted average of the $X=0$ lower SACE bound and the $X=1$ lower SACE bound. Furthermore, the lower bound on the $X=0$ SACE can be found by applying Zhang and Rubin's [5] bounds to the subpopulation where $X=0$, and similarly for the $X=1$ SACE. Plugging all of this into expression (2) and using analogous reasoning for the upper bound gives us the following large sample bounds when only monotonicity is assumed:

$$\left\{ \frac{P_{CL|X=1}P_{X=1}\left(\bar{Y}_{TL|X=1}\left(\min\left|\frac{P_{CL|X=1}}{P_{TL|X=1}}\right|\right) - \bar{Y}_{CL|X=1}\right) + P_{CL|X=0}(1 - P_{X=1})\left(\bar{Y}_{TL|X=0}\left(\min\left|\frac{P_{CL|X=0}}{P_{TL|X=0}}\right|\right) - \bar{Y}_{CL|X=0}\right)}{P_{CL|X=1}P_{X=1} + P_{CL|X=0}(1 - P_{X=1})}, \right.$$

$$\left. \frac{P_{CL|X=1}P_{X=1}\left(\bar{Y}_{TL|X=1}\left(\max\left|\frac{P_{CL|X=1}}{P_{TL|X=1}}\right|\right) - \bar{Y}_{CL|X=1}\right) + P_{CL|X=0}(1 - P_{X=1})\left(\bar{Y}_{TL|X=0}\left(\max\left|\frac{P_{CL|X=0}}{P_{TL|X=0}}\right|\right) - \bar{Y}_{CL|X=0}\right)}{P_{CL|X=1}P_{X=1} + P_{CL|X=0}(1 - P_{X=1})} \right\}.$$

When monotonicity and SRASA are assumed, the same reasoning involving weighted averages gives the following bounds:

$$\left\{ \frac{P_{CL|X=1}P_{X=1}\left(\bar{Y}_{TL|X=1} - \bar{Y}_{CL|X=1}\right) + P_{CL|X=0}(1 - P_{X=1})\left(\bar{Y}_{TL|X=0} - \bar{Y}_{CL|X=0}\right)}{P_{CL|X=1}P_{X=1} + P_{CL|X=0}(1 - P_{X=1})}, \right.$$

$$\left. \frac{P_{CL|X=1}P_{X=1}\left(\bar{Y}_{TL|X=1}\left(\max\left|\frac{P_{CL|X=1}}{P_{TL|X=1}}\right|\right) - \bar{Y}_{CL|X=1}\right) + P_{CL|X=0}(1 - P_{X=1})\left(\bar{Y}_{TL|X=0}\left(\max\left|\frac{P_{CL|X=0}}{P_{TL|X=0}}\right|\right) - \bar{Y}_{CL|X=0}\right)}{P_{CL|X=1}P_{X=1} + P_{CL|X=0}(1 - P_{X=1})} \right\}.$$

When neither monotonicity nor SRASA is assumed, the range of possible SACEs becomes more complicated. There is no longer a direct formula as above, but instead, the range depends on the probability $\pi_{DL|X=1}$ of

an individual's being in the *DL* principal stratum given that $X = 1$ for that individual and the related probability $\pi_{DL|X=0}$ that an individual is in the *DL* principal stratum given that $X = 0$ for him or her. These two probabilities may be treated as varying parameters.

Zhang and Rubin [5] note that if π_{DL} is known and neither monotonicity nor RASA is made, then the bounds on the SACE are $\bar{Y}_{TL}\left(\min\left|\frac{P_{CL}}{P_{TL}} - \frac{\pi_{DL}}{P_{TL}}\right|\right) - \bar{Y}_{CL}\left(\max\left|1 - \frac{\pi_{DL}}{P_{CL}}\right|\right)$ as the minimum and $\bar{Y}_{TL}\left(\max\left|\frac{P_{CL}}{P_{TL}} - \frac{\pi_{DL}}{P_{TL}}\right|\right) - \bar{Y}_{CL}\left(\min\left|1 - \frac{\pi_{DL}}{P_{CL}}\right|\right)$ as the maximum. By the same reasoning as above, this means that the large sample bounds for the SACE with the covariate X are a weighted average of the $X = 0$ bounds and the $X = 1$ bounds, with weights proportional to $(1 - P_{X=1})\pi_{LL|X=0}$ and $P_{X=1}\pi_{LL|X=1}$, which reduces to $(1 - P_{X=1})(P_{CL|X=0} - \pi_{DL|X=0})$ and $P_{X=1}(P_{CL|X=1} - \pi_{DL|X=1})$, assuming $\pi_{DL|X=0}$ and $\pi_{DL|X=1}$ are known. Hence the weighted average that gives the SACE bounds conditional on $\pi_{DL|X=0}$ and $\pi_{DL|X=1}$ is

$$\left\{ \begin{array}{l} \dfrac{\begin{array}{l}(P_{CL|X=1} - \pi_{DL|X=1})P_{X=1}\left(\bar{Y}_{TL|X=1}\left(\min\left|\frac{P_{CL|X=1}-\pi_{DL|X=1}}{P_{TL|X=1}}\right|\right) - \bar{Y}_{CL|X=1}\left(\max\left|1 - \frac{\pi_{DL|X=1}}{P_{CL|X=1}}\right|\right)\right)\\ + (P_{CL|X=0} - \pi_{DL|X=0})(1 - P_{X=1})\left(\bar{Y}_{TL|X=0}\left(\min\left|\frac{P_{CL|X=0}-\pi_{DL|X=0}}{P_{TL|X=0}}\right|\right) - \bar{Y}_{CL|X=0}\left(\max\left|1 - \frac{\pi_{DL|X=0}}{P_{CL|X=0}}\right|\right)\right)\end{array}}{(P_{CL|X=1} - \pi_{DL|X=1})P_{X=1} + (P_{CL|X=0} - \pi_{DL|X=0})(1 - P_{X=1})},\\[3em] \dfrac{\begin{array}{l}(P_{CL|X=1} - \pi_{DL|X=1})P_{X=1}\left(\bar{Y}_{TL|X=1}\left(\max\left|\frac{P_{CL|X=1}-\pi_{DL|X=1}}{P_{TL|X=1}}\right|\right) - \bar{Y}_{CL|X=1}\left(\min\left|1 - \frac{\pi_{DL|X=1}}{P_{CL|X=1}}\right|\right)\right)\\ + (P_{CL|X=0} - \pi_{DL|X=0})(1 - P_{X=1})\left(\bar{Y}_{TL|X=0}\left(\max\left|\frac{P_{CL|X=0}-\pi_{DL|X=0}}{P_{TL|X=0}}\right|\right) - \bar{Y}_{CL|X=0}\left(\min\left|1 - \frac{\pi_{DL|X=0}}{P_{CL|X=0}}\right|\right)\right)\end{array}}{(P_{CL|X=1} - \pi_{DL|X=1})P_{X=1} + (P_{CL|X=0} - \pi_{DL|X=0})(1 - P_{X=1})} \end{array} \right\}.$$

Since $\pi_{DL|X=0}$ and $\pi_{DL|X=1}$ are unknown, we find the bounds by optimizing over all possible values of these two conditional probabilities, which are treated as parameters, just as Zhang and Rubin [5] optimized over π_{DL} when no covariate was used. Zhang and Rubin's [5] bounds on π_{DL} are given in eq. (1). By looking only at the $X = 0$ or $X = 1$ subpopulation, it is trivial to generalize this to

$$\max(0, P_{CL|X=0} - P_{TL|X=0}) < \pi_{DL|X=0} < \min(P_{CL|X=0}, 1 - P_{TL|X=0})$$

and

$$\max(0, P_{CL|X=1} - P_{TL|X=1}) < \pi_{DL|X=1} < \min(P_{CL|X=1}, 1 - P_{TL|X=1}),$$

and these are the lower and upper limits used for taking the minima and maxima below. We have not included these in the formulas below in the interest of not cluttering the presentation. With no assumptions, the weighted average – after the optimization above – becomes

$$\left\{ \begin{array}{l} \min_{\pi_{DL|X=1},\pi_{DL|X=0}} \dfrac{\begin{array}{l}(P_{CL|X=1} - \pi_{DL|X=1})P_{X=1}\left(\bar{Y}_{TL|X=1}\left(\min\left|\frac{P_{CL|X=1}-\pi_{DL|X=1}}{P_{TL|X=1}}\right|\right) - \bar{Y}_{CL|X=1}\left(\max\left|1 - \frac{\pi_{DL|X=1}}{P_{CL|X=1}}\right|\right)\right)\\ + (P_{CL|X=0} - \pi_{DL|X=0})(1 - P_{X=1})\left(\bar{Y}_{TL|X=0}\left(\min\left|\frac{P_{CL|X=0}-\pi_{DL|X=0}}{P_{TL|X=0}}\right|\right) - \bar{Y}_{CL|X=0}\left(\max\left|1 - \frac{\pi_{DL|X=0}}{P_{CL|X=0}}\right|\right)\right)\end{array}}{(P_{CL|X=1} - \pi_{DL|X=1})P_{X=1} + (P_{CL|X=0} - \pi_{DL|X=0})(1 - P_{X=1})},\\[3em] \max_{\pi_{DL|X=1},\pi_{DL|X=0}} \dfrac{\begin{array}{l}(P_{CL|X=1} - \pi_{DL|X=1})P_{X=1}\left(\bar{Y}_{TL|X=1}\left(\max\left|\frac{P_{CL|X=1}-\pi_{DL|X=1}}{P_{TL|X=1}}\right|\right) - \bar{Y}_{CL|X=1}\left(\min\left|1 - \frac{\pi_{DL|X=1}}{P_{CL|X=1}}\right|\right)\right)\\ + (P_{CL|X=0} - \pi_{DL|X=0})(1 - P_{X=1})\left(\bar{Y}_{TL|X=0}\left(\max\left|\frac{P_{CL|X=0}-\pi_{DL|X=0}}{P_{TL|X=0}}\right|\right) - \bar{Y}_{CL|X=0}\left(\min\left|1 - \frac{\pi_{DL|X=0}}{P_{CL|X=0}}\right|\right)\right)\end{array}}{(P_{CL|X=1} - \pi_{DL|X=1})P_{X=1} + (P_{CL|X=0} - \pi_{DL|X=0})(1 - P_{X=1})} \end{array} \right\}.$$

When only SRASA is assumed, we use the same approach of finding the bounds conditional on $\pi_{DL|X=0}$ and $\pi_{DL|X=1}$, and then optimizing over these parameters to find the overall bounds. By the same reasoning as the case where neither monotonicity nor SRASA was assumed, we end up with bounds of

$$
\left\{
\begin{aligned}
& \min_{\pi_{DL|X=1},\pi_{DL|X=0}} \frac{
\begin{aligned}
& (P_{CL|X=1}-\pi_{DL|X=1})P_{X=1}\left(\bar{Y}_{TL|X=1}-\bar{Y}_{CL|X=1}\left(\max\left|1-\frac{\pi_{DL|X=1}}{P_{CL|X=1}}\right|\right)\right)\\
& +(P_{CL|X=0}-\pi_{DL|X=0})(1-P_{X=1})\left(\bar{Y}_{TL|X=0}-\bar{Y}_{CL|X=0}\left(\max\left|1-\frac{\pi_{DL|X=0}}{P_{CL|X=0}}\right|\right)\right)
\end{aligned}
}{(P_{CL|X=1}-\pi_{DL|X=1})P_{X=1}+(P_{CL|X=0}-\pi_{DL|X=0})(1-P_{X=1})}, \\[2em]
& \max_{\pi_{DL|X=1},\pi_{DL|X=0}} \frac{
\begin{aligned}
& (P_{CL|X=1}-\pi_{DL|X=1})P_{X=1}\left(\bar{Y}_{TL|X=1}\left(\max\left|\frac{P_{CL|X=1}-\pi_{DL|X=1}}{P_{TL|X=1}}\right|\right)-\bar{Y}_{CL|X=1}\right)\\
& +(P_{CL|X=0}-\pi_{DL|X=0})(1-P_{X=1})\left(\bar{Y}_{TL|X=0}\left(\max\left|\frac{P_{CL|X=0}-\pi_{DL|X=0}}{P_{TL|X=0}}\right|\right)-\bar{Y}_{CL|X=0}\right)
\end{aligned}
}{(P_{CL|X=1}-\pi_{DL|X=1})P_{X=1}+(P_{CL|X=0}-\pi_{DL|X=0})(1-P_{X=1})}
\end{aligned}
\right\}.
$$

The general principle is to consider, for each value of X, all possible values of $\pi_{DL|X=0}$ and $\pi_{DL|X=1}$, and for each pair of these values, to find the SACE for the subpopulation defined by each value of X and then to weight these two SACEs according to the proportion of the LL stratum accounted for by each value of X given this pair.

It is important to note that, as indicated in eq. (1), the SACE when the binary covariate X is available is a weighted average of the SACE when $X = 0$ and the SACE when $X = 1$. Each of these individual SACEs is accurate if we make the assumption of conditionally ignorable treatment assignment (conditional on X). Since the weighting also considers the overall proportion of observations for which $X = 1$, the SACE also makes the assumption that the proportion of units with $X = 1$ in the population is the same as the overall proportion in the sample, a special case of the assumption of knowledge of covariate distribution. Because an observational study is not randomized, this weakened set of assumptions may lead to a more suitable analysis of an observational study.

As an example, we may be interested in studying the effect of consumption of a recommended quantity of vegetables per day on QOL. Here, vegetable consumption is treated as a binary, with below-recommended consumption being the control and consumption at or above the recommended level being the treatment. We may have confidence in subjects' reporting of their vegetable consumption, but not be able to impose such consumption on them. Zhang and Rubin's [5] bounds apply here under the assumption that whether someone eats vegetables or not has no relation to his or her potential outcomes under the vegetable regiment or under a non-vegetable regiment, i.e. that vegetable consumption could just as well have been randomly imposed. However, it may be suspected that gender plays a significant role in vegetable consumption. In this case, we may have reason to believe that there is a gender difference in consumption of the recommended quantity of vegetables, that whether a male eats vegetables has no relation to his potential outcomes, and that whether a female eats vegetables has no relation to her potential outcomes, but that if the two genders are combined, this independence does not necessarily remain. In other words, we assume that within each gender, vegetable consumption could just as well have been randomly imposed, but there may be differences across genders. Here, Zhang and Rubin's [5] bounds might be erroneous, but the bounds we propose are still applicable. However, such a study may have imbalances on a number of covariates, necessitating a more complicated adaptation of this method to account for these. We briefly discuss such adaptations in Section 8.

The use of a covariate may help us to understand the data in a more detailed way and thus has the potential to narrow the bounds from the original bounds given by Zhang and Rubin [5]. However, the weakening of assumptions has the potential to widen the bounds, and sometimes this will outweigh the effect of narrowing the bounds from the addition of a covariate. This is in keeping with the new bounds'

application to an observational study rather than an experiment, as we would expect that an experiment might give us a more precise causal measurement than an observational study.

When the monotonicity assumption is made, the bounds with the covariate are a weighted average of the bounds for the $X = 0$ group and the bounds for the $X = 1$ group. When the monotonicity assumption is not made, the bounds with the covariate are also a weighted average, but the weights are not known. In either case, it is possible that the weights in the covariate case will lead to somewhat wider bounds than the weights in the case without covariates.

In the case of an experiment where treatment is fully randomized, we would not expect the results of the experiment to depend on the relative ratio of treated to control, at least not in the large sample sense. Of course, in practice, most experiments have rather limited sample size, and so the proportion of treated versus control in the experiment may be manipulated to give more precision to estimates from a particular arm. Similarly, in the large sample bound case with a covariate, the proportion of treated within each of the two subpopulations defined by X should not change the large sample bounds when X is considered. This includes the case in which one of these subpopulations has a different proportion of treated units from the other.

Note that in the case of monotonicity (with or without RASA/SRASA), there are only two places where the control group enters into the bounds formula. One is in the subtraction of the control mean in determining the bounds, but since this is done at both ends of the boundary interval, it has no effect on the width of the bounds. The other is in determining what proportion of the observations is used to find the means of the upper and lower portions of the treated group. Hence, for a fixed control survival rate, the distribution of the control arm QOL has no effect on the width of the bounds.

6 Application of bounds to the National Supported Work Demonstration

We now examine the NSW data under Zhang and Rubin's [5] bounds and the bounds described above, again using the binary covariate indicating whether an individual has graduated from high school. We will consider the response variable to be the logarithm of income in 1978, and the censoring condition to be unemployment. Hence the SACE measures an additive causal effect on the log of income, which is equivalent to a multiplicative effect on income. We assume that the population distribution is equal to the empirical distribution of the NSW data. Since we will be dealing with the log of income among those who are employed, Figure 1 gives histograms of log-income among the treated and log-income among the control.

In Table 1, from left to right, the columns give the assumptions made regarding the principal strata, the range of possible SACEs using Zhang and Rubin's [5] methods (without covariates), the range of possible SACEs considering only individuals with a diploma, the range of SACEs considering only individuals without a diploma, and the overall SACE using the methods described above, assuming that the proportions of diploma-holders in the respective principal strata are unknown. Note that in the "Range" column, SRASA and RASA lead to equivalent results, since covariates are being ignored.

Table 1 shows how crude the large sample bounds are in this case, especially without assumptions. When no assumptions are made, the bounds are so wide that the possible effects range from the treatment multiplying the typical LL individual's income by 0.31 to the multiplicative effect being 3.16. Making SRASA narrows the bounds considerably, and monotonicity narrows them even more so, with the narrowest bounds found when both are assumed. The far right column (Overall range with covariate) is quite similar to the left column in all cases, although the degree covariate does allow us to narrow the bounds somewhat in the case with no assumptions, and the bounds in the right-hand column are always narrower than the bounds in the left-hand column, even if only slightly. However, the "Overall range with covariate" column is based on weaker assumptions than the "Range" column regarding the degree covariate, as described in Section 3. In this particular case, there does seem to be a difference between the two arms as regards the

Histogram of log income among the employed, treated arm

Histogram of log income among the employed, control arm

Figure 1 Marginal distribution of log of 1978 real earnings for each arm of the NSW experiment

Table 1 Large sample bounds for the SACE of income (measured on a natural log scale) for the NSW data, under different assumptions and considering the binary covariate that indicates whether participants had a high school diploma

Assumptions	Range	With degree	Without degree	Overall range with covariate
None	[−1.16, 1.15]	[−1.08, 1.06]	[−1.18, 1.14]	[−1.11, 1.07]
SRASA	[−0.56, 0.73]	[−0.54, 0.69]	[−0.58, 0.71]	[−0.53, 0.67]
Monotonicity	[−0.15, 0.26]	[−0.13, 0.18]	[−0.16, 0.26]	[−0.16, 0.24]
SRASA and monotonicity	[0.01, 0.26]	[−0.01, 0.18]	[−0.00, 0.26]	[−0.00, 0.24]

covariate, with high school graduates comprising 27% of those treated but only 19% of the control group, a highly significant difference which leads to a two-sided P-value of 0.00025 when applying a two-sample t-test. Whether this imbalance was deliberate or an unintended shortcoming of the experiment is unclear. One possible explanation is attrition. LaLonde [8] acknowledges the existence of attrition, but indicates a belief that it does not have a major effect on the integrity of the data.

Tables 2 and 3 give essentially the same results for other covariates. In Table 2, the binary covariate used is whether the individual was employed in 1975, three years before the measurements of employment and income were taken and before the treated subjects had the opportunity to participate in the NSW. In Table 3, the binary covariate shows whether the individual was married. (Although the NSW was offered, on

Table 2 Large sample bounds for the SACE of income (measured on a log scale) for the NSW data, under different assumptions and considering the binary covariate that indicates whether participants were employed in 1975

Assumptions	Range	Employed in 1975	Unemployed in 1975	Overall range with covariate
None	[−1.16, 1.15]	[−1.08, 0.91]	[−1.30, 1.54]	[−1.04, 1.00]
SRASA	[−0.56, 0.73]	[−0.59, 0.56]	[−0.50, 0.99]	[−0.53, 0.66]
Monotonicity	[−0.15, 0.26]	[−0.21, 0.16]	[−0.04, 0.40]	[−0.14, 0.25]
SRASA and monotonicity	[0.01, 0.26]	[−0.07, 0.16]	[0.14, 0.40]	[0.01, 0.25]

Table 3 Large sample bounds for the SACE of income (measured on a log scale) for the NSW data, under different assumptions and considering the binary covariate that indicates whether participants were married

Assumptions	Range	Not married	Married	Overall range with covariate
None	[−1.16, 1.15]	[−1.22, 1.18]	[−0.91, 1.00]	[−1.25, 1.23]
SRASA	[−0.56, 0.73]	[−0.61, 0.70]	[−0.35, 0.81]	[−0.53, 0.69]
Monotonicity	[−0.15, 0.26]	[−0.13, 0.16]	[−0.23, 0.56]	[−0.15, 0.23]
SRASA and monotonicity	[0.01, 0.26]	[−0.02, 0.16]	[0.11, 0.56]	[0.00, 0.23]

a random basis, both to men and to women, everyone in LaLonde's [8] dataset was male.) As before, the use of the covariate narrows the intervals slightly. The two exceptions occur when no assumptions are made. The covariate for employment in 1975 narrows the bounds fairly substantially in this case, while the covariate for marriage widens them slightly. Since the overall bounds with a covariate make weaker assumptions than the bounds without a covariate, it is not entirely unexpected that the bounds might widen to compensate, particularly in a case such as this one where the covariate (marital status) being considered arguably has only indirect relevance to the outcome (income).

Recall from Section 5 that if both SRASA and monotonicity are assumed, the lower bound of the large sample bounds without the covariate is equal to the restricted estimate. Tables 1–3 show that in this case, the restricted estimate is 0.01, so the restricted analysis, whose shortcomings we have discussed in Section 2, shows no substantial effect from the NSW.

7 Application of bounds to simulated data

In the examples in Section 6, the availability of covariates does not do much, if anything, to change the large sample bounds on the treatment effect, although the new bounds may still be useful because of the replacement of the assumption of strong ignorability of treatment assignment with the assumption of conditional ignorability of treatment assignment. Here, we examine a simulated case in which the availability of the covariate X may lead to narrower bounds, in addition to the weakening of assumptions. We analyze the experiment as if it were an observational study, but one with conditionally ignorable treatment assignment. In this example, the treated arm and the control arm have the same joint distribution of X, survival outcomes and QOL, which is given in Table 4.

Table 4 Frequencies and QOL distributions of the principal strata in the simulated example

Proportion of sample (%)	Value of X	Survival outcome	QOL distribution
40	1	L	$N(10,1)$
20	1	D	D
20	0	L	$N(0,1)$
20	0	D	D

Of course, we do not know what the real SACE is, but bounds are a reflection of what would happen if the SACE assumed the most extreme value consistent with the known data. For convenience of presentation, in the discussion that follows, we will refer to this as the "most extreme possibility" while acknowledging that in the strictest sense, there is only one possibility (the correct one), but this is unobservable. As usual, all uncertainty is due to incomplete knowledge on the part of the observer – in this case, regarding the unobserved potential outcomes and the resulting principal strata.

If no covariates are available, then the most extreme possibility is that the *LL* stratum of one arm consists of the individuals with the highest QOL and the *LL* stratum of the other arm consists of an equally large group with the lowest QOL. Hence, in this particular example, the most extreme possibility is that the *LL* stratum of one arm comes entirely from the region given by the leftmost portion of the normal curve centered around 0 (shown as the shaded area in the top graph in Figure 2), while the *LL* stratum of the other arm comes entirely from the rightmost portion of the normal curve centered around 10 (shown as the shaded area in the bottom graph in Figure 2). However, if *X* is available, then any subset of one arm's *LL* stratum that has a particular value of *X* must be matched by an equally sized subset of the other arm's *LL* stratum that has the same value of *X*. In this example, this means that any portion of the *LL* stratum that comes from one of the two normals must be matched on the other arm by an equally sized portion from the same normal. Hence the most extreme possibility is akin to what appears in Figure 3, where the QOL values corresponding to the shaded areas are much more similar than in Figure 2. (Figures 2 and 3 provide an illustration of the most extreme possibilities for particular values of π_{LL}. To find the actual bounds, we consider all values of π_{LL} – or, in Figure 3, all values of $\pi_{LL|X=0}$ and $\pi_{LL|X=1}$ – and then find the most extreme of the individual most extreme possibilities.)
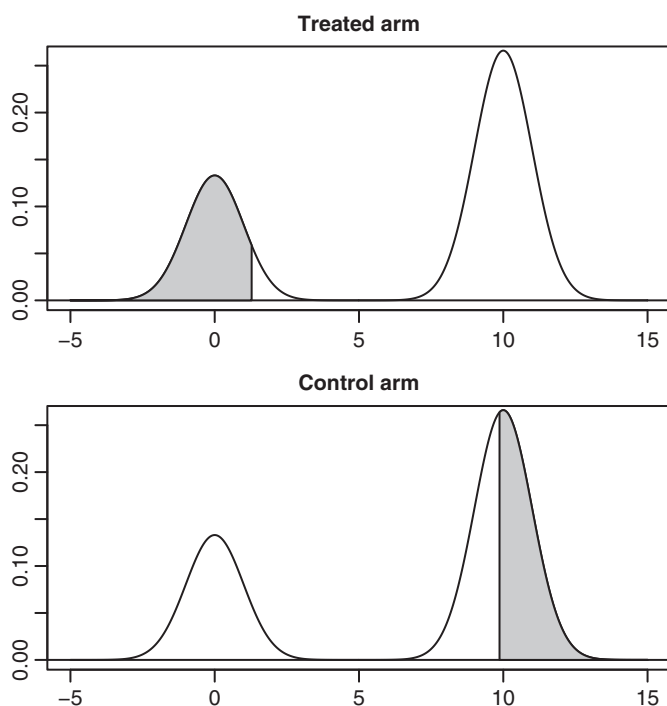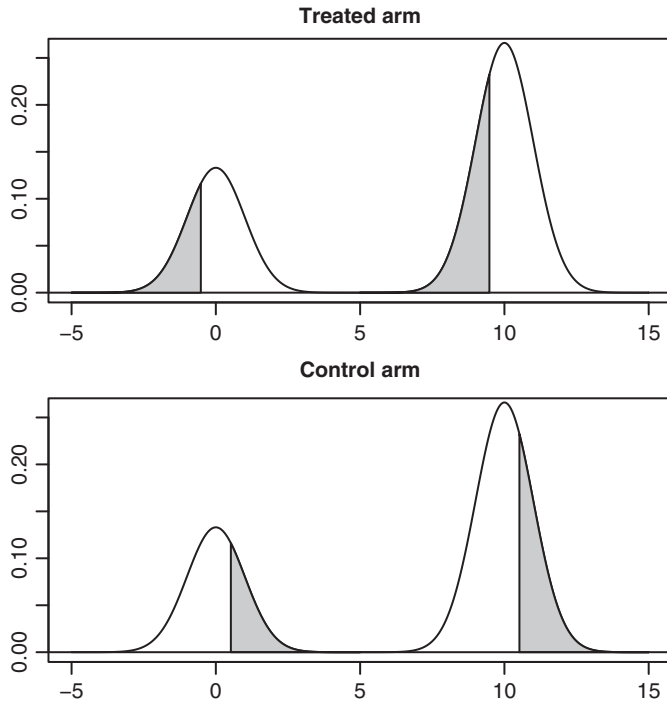

Figure 2 The most extreme possibility when bounding the SACE with no covariate. The shaded regions represent the *LL* strata of the respective arms. The QOL for each arm is a mixture of two normals with almost no overlap. The *LL* stratum of one arm – shown at the top – is composed of the observations with the lowest QOL among those in the lower normal in the mixture. The *LL* stratum of the other arm – shown at the bottom – is composed of the observations with the highest QOL among those in the higher normal of the mixture

Table 5 gives the large sample bounds in this particular case. Knowing the covariate helps immensely, decreasing the range of bounds by 83% when SRASA is not made and by 78% when SRASA is made. These improvements, in percentage terms, could be made arbitrarily close to 100% by moving the two components within each arm arbitrarily far apart.

We would expect that the covariate *X* would be more useful when it had a stronger relationship with the QOL outcome given the treatment arm. If *X* does not have a substantial conditional relationship of this sort with QOL, then the *X* = 0 and *X* = 1 large sample bounds should be very similar – both to each other and to the

Figure 3 The most extreme possibility when bounding the SACE with the covariate X available. The shaded regions represent the *LL* strata of the respective arms. The QOL for each arm is a mixture of two normals with almost no overlap, and with a binary covariate available which, in the surviving cases, exactly reflects the component of the mixture in which an observation lies. The *LL* stratum of one arm is composed of the observations with the lowest QOL within each of the components. The *LL* stratum of the other arm is composed of the observations with the highest QOL within each of the components

Table 5 A reconstruction of Tables 1–3, based on simulated data

Assumptions	Range	$X = 0$	$X = 1$	Overall range with covariate
None	[−10.80, 10.80]	[−6.85, 6.85]	[−1.62, 1.62]	[−1.82, 1.82]
SRASA	[−4.14, 4.14]	[−3.33, 3.33]	[−0.81, 0.81]	[−0.91, 0.91]
Monotonicity	[0.0]	[0.0]	[0.0]	[0.0]
SRASA and monotonicity	[0.0]	[0.0]	[0.0]	[0.0]

large sample bounds without a covariate – and since the SACE is a weighted average of these two values (albeit with the weights in the non-monotonicity cases determined by optimization), it follows that the large sample bounds with the covariate will be the same. A key point in the example above was that the variation in QOL in the $X = 0$ group and in the $X = 1$ group in the treated arm, taking the two groups individually, was much narrower than the variation in the entire treated arm, and the same phenomenon occurred in the control arm. This means that the $X = 0$ and $X = 1$ large sample bounds would be narrower, and as a result the weighted average would be narrower. Conversely, when the $X = 0$ and $X = 1$ groups are fairly similar to each other, we would expect that the weakening of assumptions associated with using the covariate would lead to a widening of the bounds as compared to the case where no covariate is used.

Perhaps surprisingly, the association of the covariate X with principal stratum is entirely irrelevant to the covariate's utility in narrowing the bounds. This is because the principal strata are not known, either before or after the analysis, and thus they never enter into the calculations of the large sample bounds. (This also makes it impossible to do any validation of the bounds based on a true value of the SACE.) Thus in generating the simulated data above, it was unnecessary to assign a principal stratum to any given simulated observation or group of simulated observations. One could imagine an extreme case in which all

of the $X = 1$ observations were in the *LL* stratum and all of the $X = 0$ observations were in the *LD* or *DL* stratum (depending on treatment arm), but it is hard to imagine such a stark situation in practice.

The Appendix provides some illustrative results of other simulations.

8 Conclusions

In this paper, we have described Zhang and Rubin's [5] approach to calculating large sample bounds on the SACE and have generalized their method to consider how, if at all, knowledge of covariates can affect these bounds. There are two reasons that covariates have utility in calculating large sample bounds on the SACE.

The first reason is that even when the bounds cannot be narrowed by using covariates, the assumptions that underlie the bounds are weakened somewhat, from completely random treatment assignment to conditionally ignorable treatment assignment. The extent to which this provides an advantage depends on the nature of the study and whether the assumption of completely random treatment assignment is reasonable. This is probably most useful in the case of an observational study, although in some cases it may still be a stronger assumption than is warranted.

Second, as seen in Section 7, there are circumstances in which a covariate can dramatically narrow the bounds. Application of the methodology to further datasets will be necessary to determine to what extent this narrowing tends to be realized in practice.

Pearl [10] analyzes the general underpinnings of the principal stratification approach. He points out that it is useful for an author to distinguish between the different possible reasons for interest in the *LL* (or other) stratum – "mathematical convenience, mathematical necessity (to achieve identification) or a genuine interest in the stratum under analysis." Our purpose is a mixture of mathematical necessity and genuine interest. A conceptual problem for further consideration is whether there is a meaningful way to include the non-*LL* survivors in a measure that maintains at least some of the SACE's convenient properties and meaningfulness.

Pearl [10] also points out the difficulties in using principal stratification to analyze direct effects, i.e. those that are not mediated by some intermediary variable. This is particularly problematic when there may be effects that have both a direct component and an indirect component.

The NSW data have a total sample size of 722. This is probably not enough for the large sample bounds to be valid. To make the bounds apply here, we might imagine that each observation in the real dataset represents a large number – perhaps 1,000 – of identical observations in a hypothetical dataset for which we wish to find bounds. Alternatively, we could adjust the bounds to compensate for the finite sample size, although here the problem we are solving changes somewhat. Cheng and Small [15] expand Zhang and Rubin's [5] ideas to a three-arm trial with a binary outcome and use a similar approach to derive large sample bounds for this situation. In this case, for a finite sample size, Cheng and Small find finite sample bounds such that, with asymptotic probability $1 - \alpha$, the true large sample bounds fall entirely within the finite sample bounds. (By "true" large sample bounds, we mean the bounds that could be derived if the distribution of QOL were known exactly for each arm, but the principal strata remained unknown.) They present two methods to accomplish this – one due to Horowitz and Manski [16] and one original approach. The original approach is based on the Bonferroni inequality, where the finite sample upper bound must be above the (unknown) upper bound based on the full distribution with probability at least $1 - \alpha/2$ and similarly with the lower bounds. Horowitz and Manski consider an approach using the full joint distribution of the estimated finite-sample bounds to ensure that these bounds have asymptotic coverage of $1 - \alpha$. In both cases, the bootstrap may be used to approximate the underlying distribution. Horowitz and Manski perform a simulation and find that their nominal 95% confidence interval for bounds has coverage that ranges from 93% to 96%. They focus on the case with missing outcomes and/or covariates; the censoring by death scenario may be fit into their framework if principal stratum is considered a missing covariate.

Another potential direction for further research involves generalizing the method to covariates with more than two categories, to multiple covariates at once, or to continuous covariates. For a covariate with n

categories, we must optimize with respect to n variables, each representing the proportion of patients in the *DL* stratum for one of the categories. Since the amount of computing time required to perform the grid search is exponential in the number of variables to be optimized, the number of calculations grows exponentially with n, so a complementary line of research involves whether there is a more efficient way to perform the optimization, for example, whether Newton–Raphson is effective in this case.

One further approach would be to consider the effects of a finite sample, which neither this paper nor Zhang and Rubin's [5] paper has done. This could involve finding either a confidence interval for the SACE or a confidence interval for the lower and/or upper large sample bounds on the SACE, although in the latter case, the creator of the estimates would have to be clear about what assumptions underlie particular large sample bounds in question. Imbens and Manski [17] have considered the more general case where a partially identified parameter (such as the SACE) is of interest, and the researcher desires to use a finite dataset to find a confidence interval. The confidence interval could be either for the range in which the parameter could fall – in our case, the bounds for the SACE, which in the more general setting would be called the identification region – or for the parameter itself – in our case, the unobservable SACE. Imbens and Manski are able to find a confidence interval with confidence level for the true parameter that is narrower than the interval that contains the entirety of the identification region with the same level of confidence.

Using multiple covariates to split the observations into strata can be handled in the same way as using a single covariate with arbitrarily many values, so that the stratification induced by two binary covariates can be considered a stratification with four levels in the same way as above. Under current methods, this quickly grows unwieldy, in the same way as with a single multiple-valued categorical covariate, especially since the number of parameters over which to optimize grows exponentially with respect to the number of variables. A particularly interesting variant of this problem involves whether and how one can exploit associations among the covariates and their effects on QOL. For example, it could be assumed that the SACE is an additive function with respect to the covariates.

A possible approach with a continuous covariate is to split the population into discrete groups according to the value of the covariate, thus essentially reducing the continuous case to the categorical case, although here again the proviso holds that the groups must be large enough so that each is considered a "large sample." Another possibility is to come up with large sample bounds for each observed value of the continuous covariate separately, applying Zhang and Rubin's [5] formulas to observations whose covariate value is close to the covariate value for the observation of interest, akin to a nearest-neighbor approach or a kernel smoother. This approach hinges on the assumption that for any given observation, there is a set of neighbors that are similar enough with respect to the covariate that there is only a negligible difference between the covariate's average causal effect on the individual observation and the covariate's average causal effect on the neighbors. At the same time, the set of neighbors must be large enough that the application of the large sample bounds is reasonable. As always when using a nearest-neighbor-type approach, there may be concerns about bias near the boundaries, as well as uncertainty about how to weight the different values of the nearest neighbors, so the appropriate weightings should be considered very carefully. Possibly some other simplifying assumption – such as that the SACE is linear with respect to the continuous covariate – could make calculating the bounds with a continuous variable a more reasonable prospect. For more generality, the researcher could make semiparametric assumptions, such as a relationship between the covariate and the SACE that has a specified smoothness and, in the more complicated case of multiple covariates, that the model giving the SACE has no interactions among the covariates. However, it is unclear whether calculating the SACE using a continuous covariate is feasible, even in the linear case with no interactions.

Appendix

We tried three other simulations of possible conditions to determine how much the bounds are narrowed (or widened) by the presence of a covariate given the different assumptions. In each simulation, we assumed that a binary covariate X is available. We fixed the proportion of the time that $X = 1$, as well as the survival probabilities conditional on the combination of X and the treatment arm. We assumed that for a given treatment status and a given value of the covariate, QOL among survivors is normally distributed with standard deviation 1. Hence when both values of the covariate X are considered together, the survivors for each treatment arm have QOL that follows a homoscedastic normal–normal mixture, with each normal representing one of the two values of X. We define the *separation* of this mixture to be the difference (in units of the common standard deviation) of the means of the two normals; if the $X = 1$ group has a greater mean than the $X = 0$ group, then we define the separation to be positive, while in the opposite case, we define the separation to be negative. We vary the separations for the two arms and consider the ratio of the width of the bounds without a covariate to the width of the bounds with a covariate. As noted in Section 5, the distribution of the surviving control arm QOL has no effect on the widths of the bounds in the case where monotonicity is assumed.

We performed three illustrative simulations, the results of which are below. For each simulation and each combination of assumptions and separations, we calculated the "width ratio" – the ratio of the width of the bounds with the covariate X to the width of the bounds without the covariate X. A lower width ratio shows more improvement from knowledge of the covariate.

Simulation 1: In this simulation, 60% of the treated individuals and 60% of the control individuals had $X = 1$, with the rest having $X = 0$. The survival probability was 60% if treated and 50% if receiving the control, regardless of X. Tables 6–8 show the results with no assumptions, with SRASA, and with monotonicity (both with and without SRASA), respectively.

Table 6 Width ratio as a function of treated and control separation, assuming neither monotonicity nor SRASA

Width ratio with no assumptions							Control separation	
		−1.5	−1	−0.5	0	0.5	1	1.5
Treatment separation	−1.5	0.805	0.848	0.878	0.890	0.879	0.851	0.815
	−1	0.850	0.897	0.932	0.944	0.932	0.898	0.853
	−0.5	0.883	0.934	0.971	0.985	0.971	0.934	0.883
	0	0.895	0.947	0.986	1.000	0.986	0.947	0.895
	0.5	0.883	0.934	0.971	0.985	0.971	0.934	0.883
	1	0.853	0.898	0.932	0.944	0.932	0.897	0.850
	1.5	0.815	0.851	0.879	0.890	0.878	0.848	0.805

Table 7 Width ratio as a function of treated and control separation, assuming SRASA but not monotonicity

Width ratio with SRASA							Control separation	
		−1.5	−1	−0.5	0	0.5	1	1.5
Treatment separation	−1.5	0.791	0.837	0.870	0.892	0.910	0.923	0.933
	−1	0.840	0.892	0.928	0.941	0.941	0.938	0.937
	−0.5	0.875	0.931	0.970	0.984	0.972	0.949	0.931
	0	0.897	0.945	0.985	1.000	0.986	0.951	0.915
	0.5	0.913	0.942	0.971	0.985	0.972	0.937	0.891
	1	0.924	0.937	0.945	0.948	0.935	0.903	0.860
	1.5	0.932	0.934	0.926	0.909	0.888	0.859	0.820

Table 8   Width ratio as a function of treated separation, with monotonicity assumed, and with and without SRASA

Treated separation	Width ratio with monotonicity	Width ratio with monotonicity and SRASA
−1.5	0.806	0.822
−1	0.898	0.904
−0.5	0.971	0.972
0	1.000	1.000
0.5	0.971	0.970
1	0.898	0.892
1.5	0.806	0.792

Simulation 2: In this simulation, 70% of the treated individuals and 60% of the control individuals have $X = 1$; the survival probability among those with $X = 0$ is 70% if treated and 50% if control; and the survival probability among those with $X = 1$ is 60% if treated and 50% if control. Tables 9–11 show the results with no assumptions, with SRASA, and with monotonicity (both with and without SRASA), respectively.

Table 9   Width ratio as a function of treated and control separation, assuming neither monotonicity nor SRASA

Width ratio with no assumptions							Control separation
		−1.5	−1	−0.5	0	0.5	1	1.5
Treatment separation	−1.5	0.648	0.682	0.707	0.718	0.715	0.700	0.677
	−1	0.683	0.721	0.749	0.759	0.751	0.730	0.700
	−0.5	0.708	0.750	0.779	0.790	0.779	0.752	0.716
	0	0.719	0.760	0.790	0.802	0.790	0.760	0.719
	0.5	0.716	0.752	0.779	0.790	0.779	0.750	0.708
	1	0.700	0.730	0.751	0.759	0.749	0.721	0.683
	1.5	0.677	0.700	0.715	0.718	0.707	0.682	0.648

Table 10   Width ratio as a function of treated and control separation, assuming SRASA but not monotonicity

Width ratio with SRASA							Control separation
		−1.5	−1	−0.5	0	0.5	1	1.5
Treatment separation	−1.5	0.635	0.672	0.706	0.739	0.772	0.800	0.823
	−1	0.675	0.717	0.745	0.765	0.781	0.797	0.812
	−0.5	0.710	0.747	0.779	0.790	0.789	0.786	0.789
	0	0.742	0.766	0.790	0.802	0.791	0.771	0.757
	0.5	0.770	0.781	0.788	0.791	0.780	0.752	0.722
	1	0.795	0.792	0.785	0.771	0.753	0.726	0.690
	1.5	0.818	0.807	0.786	0.758	0.725	0.694	0.661

Table 11   Width ratio as a function of treated separation, with monotonicity assumed, and with and without SRASA

Treated separation	Width ratio with monotonicity	Width ratio with monotonicity and SRASA
−1.5	0.819	0.840
−1	0.909	0.917
−0.5	0.979	0.980
0	1.006	1.006
0.5	0.979	0.977
1	0.909	0.901
1.5	0.819	0.799

Simulation 3: The following set of tables gives results for a somewhat more extreme situation. Here, only 10% of each arm consists of individuals for whom $X = 1$. For these few patients where $X = 1$, the treatment has a relatively small effect on the survival probability, raising it from 50% to 60%. However, for the majority for whom $X = 0$, the treatment has a dramatic effect, increasing the survival probability from 30% to 100%, thus turning likely death into certain survival. The same tables as before, adapted to this situation, are Tables 12–14.

Table 12 Width ratio as a function of treated and control separation, assuming neither monotonicity nor SRASA

Width ratio with No assumptions							Control separation	
		−1.5	−1	−0.5	0	0.5	1	1.5
Treatment separation	−1.5	0.746	0.755	0.766	0.788	0.807	0.821	0.831
	−1	0.764	0.774	0.780	0.788	0.806	0.821	0.831
	−0.5	0.783	0.787	0.793	0.795	0.799	0.814	0.824
	0	0.808	0.797	0.798	0.800	0.798	0.797	0.808
	0.5	0.824	0.814	0.799	0.795	0.793	0.787	0.783
	1	0.831	0.821	0.806	0.788	0.780	0.774	0.764
	1.5	0.831	0.821	0.807	0.788	0.766	0.755	0.746

Table 13 Width ratio as a function of treated and control separation, assuming SRASA but not monotonicity

Width ratio with SRASA							Control separation	
		−1.5	−1	−0.5	0	0.5	1	1.5
Treatment separation	−1.5	0.747	0.801	0.850	0.894	0.933	0.968	1.002
	−1	0.761	0.774	0.824	0.868	0.908	0.945	0.979
	−0.5	0.783	0.785	0.793	0.838	0.879	0.916	0.951
	0	0.828	0.801	0.798	0.800	0.841	0.878	0.914
	0.5	0.864	0.838	0.804	0.795	0.793	0.831	0.867
	1	0.888	0.863	0.830	0.791	0.777	0.773	0.809
	1.5	0.900	0.876	0.844	0.806	0.764	0.749	0.744

Table 14 Width ratio as a function of treated separation, with monotonicity assumed, and with and without SRASA

Treated separation	Width ratio with monotonicity	Width ratio with monotonicity and SRASA
−1.5	0.887	0.896
−1	0.914	0.917
−0.5	0.933	0.933
0	0.939	0.939
0.5	0.933	0.932
1	0.914	0.911
1.5	0.887	0.879

References

1. Gilbert PB, Bosch RJ, Hudgens MG. Sensitivity analysis for the assessment of causal vaccine effects on viral load in HIV vaccine trials. Biometrics 2003;59:531–41.
2. Egleston BL, Scharfstein DO, Freeman EE, West SK. Causal inference for non-mortality outcomes in the presence of death. Biostatistics 2007;8:526–45.

3. McCaffrey DF, Morral AR, Ridgeway G, Griffin BA. Interpreting treatment effects when cases are institutionalized after treatment. Drug Alcohol Dependence 2007;89:126–38.
4. Griffin BA, McCaffrey DF, Morral AR. An application of principal stratification to control for institutionalization at follow-up in studies of substance abuse treatment programs. Ann Appl Stat 2008;2:1034–55.
5. Zhang JL, Rubin DB. Estimation of causal effects via principal stratification when some outcomes are truncated by "death". J Educ Behav Stat 2003;28:353–68.
6. Imai K. Sharp bounds on the causal effects in randomized experiments with "truncation-by-death". Stat Probability Lett 2008;78:144–9.
7. Chiba Y, VanderWeele TJ. A simple method for principal strata effects when the outcome has been truncated due to death. Am J Epidemiol 2011;173:745–51.
8. LaLonde R. Evaluating the economic evaluation of training programs with experimental data. Am Econ Rev 1986;76:604–20.
9. Frangakis CE, Rubin DB. Principal stratification in causal inference. Biometrics 2002;58:21–9.
10. Pearl J. Principal stratification – a goal or a tool? Int J Biostat 2011;7:Article 20.
11. Rubin DB. Bayesian inference for causal effects: the role of randomization. Ann Stat 1978;6:34–58.
12. Rubin DB. More powerful randomization-based p-values in double-blind trials with non-compliance. Stat Med 1998;17:371–85.
13. Rosenbaum PR, Rubin DB. The central role of the propensity score in observational studies for causal effects. Biometrika 1983;70:41–55.
14. Jemiai Y. Semiparametric methods for the effect of treatment on an outcome existing only in a post-randomization selected subpopulation. Ph.D. thesis, Harvard University, Cambridge, MA, 2005.
15. Cheng J, Small DS. Bounds on causal effects in three-arm trials with non-compliance. J R Stat Soc Ser B Stat Methodol 2006;68:815–36.
16. Horowitz JL, Manski CF. Nonparametric analysis of randomized experiments with missing covariate and outcome data. J Am Stat Assoc 2000;95:77–84.
17. Imbens GW, Manski CF. Confidence intervals for partially identified parameters. Econometrica 2003;72:1845–57.