Using Post Outcome Measurement Information in Censoring by Death Problems

Fan Yang

*University of Chicago, Chicago, USA.*

Dylan S. Small

*University of Pennsylvania, Philadelphia, USA.*

**Summary.** Many clinical studies on non-mortality outcomes such as quality of life suffer from the problem that the non-mortality outcome can be censored by death, i.e. the non-mortality outcome cannot be measured if the subject dies before the time of measurement. To address the problem that this censoring by death is informative, it is of interest to consider the average effect of the treatment on the non-mortality outcome among subjects whose measurement would not be censored under either treatment or control, called the survivor average causal effect (SACE). The SACE is not point identified under usual assumptions but bounds can be constructed. The previous literature on bounding the SACE uses only the survival information before the measurement of the non-mortality outcome. However, survival information after the measurement of the non-mortality outcome could also be informative. For randomized trials, we propose a set of ranked average score assumptions that make use of survival information before and after the measurement of the non-mortality outcome which are plausibly satisfied in many studies and develop a two-step linear programming approach to obtain the closed form for bounds on the SACE under our assumptions. We also extend our method to randomized trials with noncompliance or observational studies with a valid IV to obtain bounds on the complier survivor average causal effect which is presented in the Supplementary Material. We apply our method to a randomized trial of the effect of mechanical ventilation with lower tidal volume vs. traditional...
tial volume for acute lung injury patients. Our bounds on the SACE are much shorter than the bounds obtained using only the survival information before the measurement of the non-mortality outcome.

**Keywords**: Censoring by death; Causal inference; Instrumental variable; Quality of life.

1. **Introduction**

In many clinical studies, researchers are interested in the effect of a treatment on a non-mortality outcome such as complications or quality of life (QOL) in addition to mortality. However, the assessment of the causal effect on non-mortality outcomes of interest is often complicated by censoring by death. This censoring by death occurs because, by the time the non-mortality outcome is measured, some patients have died and thus the non-mortality outcome cannot be measured or is not well defined for these dead patients. For example, suppose we want to study the effect on intraventricular hemorrhage (IVH) of premature babies being delivered in a high-level neonatal intensive care unit (NICU) vs. a lower-level NICU. IVH is rarely present at birth but usually occurs in the first several days of life (Lee, 2013). If the baby died before being born (a fetal death) or shortly after birth, then whether the baby had IVH is not well-defined. Another example is that in cancer studies, QOL outcomes that might be measured six months or a year after treatment like incidence of fatigue, myelosuppression and treatment side-effects (e.g., Motzer et al., 2013) are important outcomes considered to assess the efficacy of a treatment. However, patients may die before the measurement of the QOL outcomes; for those patients, the QOL outcomes are not well-defined. Censoring by death is typically informative – patients who die usually would have had worse QOL than those who did not die even if the dead patients could have somehow been kept alive (Cox et al., 1992). Furthermore, those patients who are saved by a treatment are often sicker
patients on average than those patients who would live under both treatment and control. Consequently, a direct comparison of the non-mortality outcomes among the survivors in treatment vs. control would be biased. To address the fundamental problems that the non-mortality outcomes are not well defined for those who die before measurement and that the censoring of the measurement is informative, Rubin (2000), and Frangakis and Rubin (2002) proposed a well defined causal estimand – the survivor average causal effect (SACE) – which is the effect of treatment on the non-mortality outcome among patients who would survive under both treatment and control to the time point when the non-mortality outcome is measured. The group of people who would survive under both treatment and control to the time point when the non-mortality outcome is measured are called always survivors. Although in this paper, we focus on QOL outcomes, the phenomenon of censoring by death also arises in many other fields and the meaning of always survivors may vary from study to study. For example, in education, there is interest in comparing two educational programs’ effects on students’ final test scores in high school (or college admission test scores), the always survivors in this case refer to those students who would not drop out (would take the college admission test) regardless of program assignment (Zhang and Rubin, 2003; Angrist, Bettinger and Kremer, 2006); Another field where censoring by death arises is social policy interventions, for example, in studying the effect of an initiative to support marriage on quality of marriage, the quality of marriage is censored by divorce, and the always survivors here are the couples who would not divorce regardless of receiving the intervention or not (McConnell, Stuart and Devaney, 2008). In economics, censoring by death arises in the study of a job training program’s effect on wages; the always survivors are people who would be employed regardless of being assigned to training or not (Zhang, Rubin and Mealli, 2009). As pointed out in Zhang, Rubin and Mealli (2009), whether the always survivors group is of substantial interest depends on the context and on its relative size. In the job-training program Zhang, Rubin and Mealli
(2009) considered, the proportion of always survivors is estimated to be 51%; in the ARDSNet study we will apply our method to (section 6), the proportion of always survivors is estimated to be 65%. In both studies, always survivors constitute a significant part of the population and are of great interest. There are also studies where the always survivors only constitute a small proportion. In the HIV vaccine trial Gilbert, Bosch and Hudgens (2003) studied to assess the vaccine's impact on viral load after acquisition of HIV, the proportion of subjects who would always be infected regardless of being assigned to vaccine group or placebo group is less than 20%, but the evaluation of the SACE is an important goal because the viral load of an infected person predicts infectiousness (Quinn et al., 2000) and the rate of disease progression (Mellors et al., 1997).

Unfortunately, the SACE is not point identified without further assumptions. One type of assumption to obtain identifiability is through imposing parametric models, such as a mixture normal model on continuous outcomes proposed by Zhang, Rubin and Mealli (2009) and Frumento et al. (2012). However, when the outcome of interest is binary, a parametric binary mixture model is not sufficient to provide identifiability; in this case, Mattei and Mealli (2007) proposed a Bayesian approach to draw inference on the SACE. Another method to identify the SACE is by imposing assumptions on a pretreatment covariate as in Ding et al., (2011). These different sets of further assumptions that enable identifiability of the SACE are reasonable for the settings considered by the authors, but, for the setting of our study, we do not feel able to make such further assumptions. Although the SACE is not identified without strong untestable assumptions, with reasonable assumptions, an interval in which the SACE will lie is identified. Zhang and Rubin (2003) discussed various assumptions (ranked average score assumptions) that can be made to bound the SACE, and derived large sample bounds in a randomized trial. Imai (2008) provided an alternative proof that the bounds obtained in Zhang and Rubin (2003) are sharp and generalized the proof to obtain sharp bounds on the quan-
tile treatment effect. Chiba (2012) proposed a number of assumptions that are different from the ranked average score assumptions in Zhang and Rubin (2003) and derived the corresponding bounds. Another stream of work on drawing inference about the SACE is through sensitivity analysis procedures, for instance, Hayden et al. (2005), Egleston et al. (2007), and Chiba and VanderWeele (2011). A problem similar to censoring by death arises in evaluating the effect of vaccine vs. placebo on post-infection outcomes. Hudgens, Hoering and Self (2003) developed tests for the causal effect on viral load among the individuals who would be infected no matter whether they received the vaccine regimen or a placebo regimen. Gilbert, Bosch and Hudgens (2003) proposed a class of models indexed by an interpretable sensitivity parameter, where the SACE is identified given the sensitivity parameter.

In the previous literature on bounding the SACE, only the survival information before the measurement on the non-mortality outcome has been used. However, survival information after measurement may provide additional information as a proxy of health condition. This post measurement survival information is often available in studies that are interested in long term outcomes where patients were followed up for a considerable period of time, such as the ARDSNet trial introduced in the next paragraph, and the Awakening and Breathing controlled trial (Jackson et al. (2010)) where patients were followed up at three months and twelve months after discharge. In this paper, we develop a method to use both the survival information before and after the measurement of non-mortality outcome to sharpen inferences on the SACE in the setup of randomized experiments. We will also present in the online Supplementary Material an extension of our method to bound the complier survivor average causal effect (CSACE) in a randomized trial with noncompliance or an observational study where a binary instrumental variable (IV) is available. The CSACE refers to the average causal effect on an outcome of interest among a group of compliers (those who would take the treatment if encouraged to do so by an IV but not take the treatment if not encouraged) who would survive to the time point
of measurement under both the encouraging and not encouraging levels of the IV. Mattei and Mealli (2007) also considered the problem of treatment noncompliance in the presence of censoring by death by making inferences about treatment effects on always survivors among compliers.

We will apply our method to the ARDSNet study, a randomized clinical trial on the effect of mechanical ventilation with lower tidal volumes vs. traditional tidal volumes for patients suffering from acute lung injury (The Acute Respiratory Distress Syndrome Network, 2000). The trial found evidence that lower tidal volumes reduce mortality. The investigators were also interested in assessing the effect of lower tidal volumes on a QOL outcome, whether the patient was able to breathe without assistance by day 28. In the data, both survival at day 28, when the QOL is measured, and whether the patient was ultimately discharged home alive, which is post-QOL measurement survival information, are recorded. Utilizing the post QOL measurement survival information in addition to the pre-QOL measurement survival information, we are able to substantially narrow the bounds on the SACE for the effect of lower tidal volume on being able to breathe without assistance by day 28.

The rest of the paper is organized as follows. In section 2, we introduce notations and assumptions to set up the causal framework. In section 3, we present the derivations of the bounds on the SACE and provide some numerical examples to compare the bounds derived with the bounds without utilizing the post measurement survival information. In section 4, we discuss how to check the plausibility of our assumptions for the "large sample" data as well as the sample data. We discuss the estimation and inference for bounds for sample data in section 5, and we apply our approach to the ARDSNet study in section 6. Conclusions and discussions are presented in section 7. Supplementary material and the R code to implement our method to the ARDSNet study are available online: http://www-stat.wharton.upenn.edu/~dsmall
2. Notation and Assumptions

We consider two arm randomized experiments with perfect compliance. An extension of our method to randomized trial with noncompliance or an observational study where a binary IV is available is provided in the Supplementary Material.

2.1. Notation

We use the potential outcomes approach to define causal effects. Let \( D_i \) represent the binary treatment for the \( i^{th} \) subject; we call level 1 "the treatment" and level 0 "the control". Let \( D \) denote the vector of treatment assignment indicators for all subjects. Let \( S_1(d) \) be the potential survival indicator of subject \( i \) that would be observed if the treatment assignment were \( d \) at the first time point right after which the measurement of non-mortality outcome is taken, with 0 indicating death, 1 if alive. Let \( Y_i(d) \) represent the potential non-mortality binary outcome (for instance, complication of babies, QOL of participants) that would be observed under treatment assignment \( d \). The non-mortality outcome is measured after the first time point, thus if the subject would die before that time point (\( S_1(d) = 0 \)), \( Y_i(d) \) is not defined. For convenience, we assume that level 1 of the non-mortality outcome is worse than level 0 of the outcome, e.g., in the ARDSNet study, level 1 indicates that the patient was not able to breathe without assistance by day 28 and level 0 indicates the patient was able to breathe without assistance by day 28. We further define \( S_2(d) \) to be the potential indicator of survival at the second time point (a time point after the measurement of non-mortality outcome) for subject \( i \) that would be observed if under treatment assignment \( d \). If \( S_1(d) = 0 \), then \( S_2(d) = 0 \) by definition. We use \( D_i, S_1i, Y_i \) and \( S_2i \) to denote respectively the observed treatment received, observed survival indicator at the first time point, observed non-mortality outcome and observed survival indicator at the second time point for subject \( i \).
2.2. Assumptions

We assume that the following assumptions hold for randomized experiments.

Assumption 1. Stable unit treatment value assumption (SUTVA). Let \( d \) and \( d' \) be any two possible treatment assignments. If \( d_i = d'_i \), then \( S_{1i}(d) = S_{1i}(d') \), \( S_{2i}(d) = S_{2i}(d') \), and \( Y_i(d) = Y_i(d') \).

SUTVA means that there is one version of the treatment and that there is no interference between subjects so that a subject’s outcomes only depend on the subject’s own treatment. SUTVA is plausible for studies, such as the ARDSNet study, where one subject’s outcomes are not likely to be affected by other patients’ treatment assignment. Under SUTVA, each subject has two potential first time point survival outcomes \((S_{1i}(1), S_{1i}(0))\), based on values of which we can classify subjects into four groups:

- \( 11 = \{i| S_{1i}(1) = 1, S_{1i}(0) = 1\} \), always survivors: the subjects that would survive at least to the first time point under both treatment arms;
- \( 10 = \{i| S_{1i}(1) = 1, S_{1i}(0) = 0\} \), protected: the subjects that would survive at least to the first time point under treatment, but would die before then under control;
- \( 01 = \{i| S_{1i}(1) = 0, S_{1i}(0) = 1\} \), harmed: the subjects that would die before the first time point under treatment, but would survive at least to the first time point under control;
- \( 00 = \{i| S_{1i}(1) = 0, S_{1i}(0) = 0\} \), never survivors: the subjects that would die before the first time point under both treatment arms.

Assumption 2. The assignment \( D_i \) of each subject is independent of his/her potential outcomes.

When treatment is randomly assigned, Assumption 2 holds by the design of the experiment.
Two-stage Censoring by Death

Table 1. Fine Strata

<table>
<thead>
<tr>
<th>Probability</th>
<th>$S_{1i}(1)$</th>
<th>$S_{1i}(0)$</th>
<th>$S_{2i}(1)$</th>
<th>$S_{2i}(0)$</th>
<th>Principal Strata at Time Point 1</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\pi_{1111}$</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>Always survivors</td>
</tr>
<tr>
<td>$\pi_{1110}$</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>Always survivors</td>
</tr>
<tr>
<td>$\pi_{1100}$</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>Always survivors</td>
</tr>
<tr>
<td>$\pi_{1010}$</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>Protected</td>
</tr>
<tr>
<td>$\pi_{1000}$</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>Protected</td>
</tr>
<tr>
<td>$\pi_{0000}$</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>Never survivors</td>
</tr>
</tbody>
</table>

Assumption 3. Monotonicity: $S_{1i}(1) \geq S_{1i}(0), S_{2i}(1) \geq S_{2i}(0)$. There is no 01 (harmed) group.

The monotonicity assumption says that the treatment does not cause death. This assumption cannot be directly tested, but is often plausible in studies where the treatment benefits survival. Under this assumption, subjects could either be “always survivors”, “protected” or “never survivors”. The most meaningful inference of causal effect of treatment on $Y$ can be drawn only for the ”always survivors”, because it is the only group for which both $Y_{i}(1)$ and $Y_{i}(0)$ are well defined, see Rubin (2000) and Frangakis and Rubin (2002). Define the survivor average causal effect (SACE) as $E(Y_{i}(1) - Y_{i}(0) | 11)$, which is our quantity of interest.

We further create finer strata based on the possible combinations of potential survival at both the first (QOL measurement point) and second (post-QOL measurement point) time points, which are described in Table 1. These strata are principal strata which stratify the population based on the joint potential survival outcomes under treatment and under control. From now on, we call these strata fine strata to differentiate them with the stratification created based on survival only before the measurement, namely, always survivors, protected, harmed and never survivors.

The always survivors at time point 1 are divided into the following three subgroups: 1111, always survivors who would live at least to the second time point under both treatment arms; 1110, always survivors who would survive at least to the
second time point under treatment, but would die before then under control; 1100, always survivors who although can live at least to the first time point, would die before the second time point under both treatment arms. The protected at time point 1 are combinations of the following two subgroups: 1010, subjects who would live at least to the second time point under treatment, but would die before the first time point under control; 1000, subjects who if they receive treatment would live at least to the first time point but would die before the second time point, but if they receive control, would die even before the first time point. Never survivors comprise a single subgroup which we denote as 0000 because the second time point death indicator provides no additional information for them.

In terms of our fine strata, the SACE is expressed as:

\[
SACE = E(Y_i(1) - Y_i(0) | S_i(1) = S_i(0) = 1)
\]

\[
= \frac{(\pi_{1111}E(Y_i(1) | 1111) + \pi_{1110}E(Y_i(1) | 1110) + \pi_{1100}E(Y_i(1) | 1100))}{\pi_{1111} + \pi_{1110} + \pi_{1100}}
\]

\[
- \frac{(\pi_{1111}E(Y_i(0) | 1111) + \pi_{1110}E(Y_i(0) | 1110) + \pi_{1100}E(Y_i(0) | 1100))}{\pi_{1111} + \pi_{1110} + \pi_{1100}}.
\] (1)

Plausible assumptions can be made on data to tighten the bounds of SACE. Zhang and Rubin (2003) proposed the ranked average score assumption with one time point survival information, which is described in Assumption 4.

**Assumption 4.** Ranked average score assumption with one time point survival information (Zhang and Rubin, 2003). When assigned treatment, the probability of worse outcome for always survivors is not higher than that for the protected:

\[P(Y_i(1) = 1 | 11) \leq P(Y_i(1) = 1 | 10).\]

This is a plausible assumption for many studies since the protected would die under the control but the always survivors wouldn’t, thus it’s reasonable to assume that always survivors are on average healthier when measured than the protected so that they have lower risk of the bad QOL. This assumption uses only the information on death before the measurement of the non-mortality outcome. However,
Two-stage Censoring by Death

survival information after measurement of the non-mortality outcome may deliver finer information, making use of which can help us make more reasonable assumptions and sharpen inferences. We will refer to the following set of assumptions (Assumptions 5-7) as ranked average score assumptions with two time points survival information.

**Assumption 5.** Among always survivors at time point 1, the probability of worse outcome for group 1111 is the lowest, whereas the probability of worse outcome for group 1100 is the highest under both treatment arms:

\[ P(Y_i(1) = 1 | 1111) \leq P(Y_i(1) = 1 | 1110) \leq P(Y_i(1) = 1 | 1100), \] (2)

\[ P(Y_i(0) = 1 | 1111) \leq P(Y_i(0) = 1 | 1110) \leq P(Y_i(0) = 1 | 1100). \] (3)

**Assumption 6.** Among protected at time point 1, the probability of worse outcome for group 1010 is not higher than that for group 1000 under treatment:

\[ P(Y_i(1) = 1 | 1010) \leq P(Y_i(1) = 1 | 1000). \] (4)

**Assumption 7.** Under treatment, the probability of worse outcome for group 1100 is not lower than that for group 1010, but not higher than that for group 1000, and the probability of worse outcome for group 1110 is not higher than that for group 1010:

\[ P(Y_i(1) = 1 | 1110) \leq P(Y_i(1) = 1 | 1010) \leq P(Y_i(1) = 1 | 1100) \leq P(Y_i(1) = 1 | 1000). \] (5)

Assumptions 5, 6 and 7 are plausibly satisfied in many QOL studies. Consider the ARDSNet study of the effect of lower tidal volumes (treatment) vs. traditional tidal volumes (control) on being able to breathe without assistance by day 28 in the ICU described in the introduction, where the post-QOL measurement survival time point is being discharged home alive. Assumption 5 says, among patients who would survive to day 28 under both treatment and control, those patients who would be discharged home alive under both treatment and control are healthiest at
day 28 on average, and those who would be discharged home alive under treatment but not control are healthier at day 28, than those who would die in the hospital under both treatment and control. Assumption 6 says, among patients who would survive to day 28 only under treatment, those patients who would ultimately be discharged home alive under treatment are healthier on average than patients who would ultimately die in the hospital. Assumptions 5 and 6 are plausible because being discharged home alive is a proxy for health at day 28. Assumption 7 is a comparison of the 1010 patients who would die before day 28 under control but survive to day 28 and be discharged home alive under treatment, to the 1100 patients who would survive to day 28 under both treatment and control but die in the hospital after day 28 under both treatment and control. Assumption 7 says that under the treatment, the 1010 patients tend to be healthier than the 1100 patients at day 28. This is plausible for the ARDSNet study for the following reasons. The 1100 patients are likely to be fairly sick by day 28 under the treatment since these patients will die in the ICU. In contrast, the 1010 patients are likely to be less sick on day 28 under the treatment because they will be (or already have) discharged home alive. An example of a 1010 patient would be a patient who was healthy but suffered a gunshot wound that caused an acute lung injury. When the patient arrives at the ICU, the patient is in critical condition and only the treatment will save the patient, but if the patient receives the treatment, the patient’s health before the gunshot wound will enable the patient to recover well and be regaining his or her health by day 28. In summary, assumptions 5, 6 and 7 are plausible for the ARDSNet study.

The ranked average score assumption with one time point survival information is different from our ranked average score assumptions with two time points survival information. The major difference is that the one time point survival information assumption assumes that always survivors, on average, have better QOL outcome than the protected, whereas our two time points survival information assumptions assume that one particular always survivors group, 1100, has worse QOL
outcome than a particular protected group, 1010, on average under treatment, which is a more reasonable assumption for the ARDSNet study. The bounds obtained under the ranked average score assumption with one time point survival information and our two time points survival information assumptions are compared analytically in Section 3.3, and examples of the differences in the bounds are presented in numerical examples and the analysis of the ARDSNet study in section 3.4, 5 and 6 respectively.

3. Derivation of Bounds

Under Assumptions 1-3 and the ranked average score assumptions with two time points survival information (Assumptions 5-7), the SACE is not point identified based on the knowledge of the observable joint distribution of \((D_i, S_{1i}, S_{2i}, Y_i)\). However, we can use that joint distribution to obtain an interval in which the SACE must lie. We first derive the bounds for the proportions in each stratum, then for fixed proportions we derive the bounds for the SACE. In this section, we assume that the joint distribution of \((D_i, S_{1i}, S_{2i}, Y_i)\) is known; in section 5, we will discuss forming confidence intervals for the bounds to acknowledge sample uncertainty.

3.1. Bounds for proportions of each stratum

Notice that the observable strata of \((D_i, S_{1i}, S_{2i})\) are mixtures of fine strata (Table 1). Thus we can express the proportions of strata of \((D_i, S_{1i}, S_{2i})\) by proportions of fine strata. Combining this with the fact that all the proportions in the fine strata must lie between 0 and 1, we can obtain the bounds for each fine stratum’s proportion. We use \(p_{s_1s_2|d}\) to denote \(P(S_{1i} = s_1, S_{2i} = s_2 \mid D_i = d)\). The following identities hold:

\[
p_{11|1} = \pi_{1111} + \pi_{1110} + \pi_{1010},
\]

\[
p_{10|1} = \pi_{1100} + \pi_{1000},
\]
Yang and Small

\[ p_{00|0} = \pi_{0000}, \quad (8) \]

\[ p_{11|0} = \pi_{1111}, \quad (9) \]

\[ p_{10|0} = \pi_{1110} + \pi_{1100}, \quad (10) \]

\[ p_{00|0} = \pi_{1010} + \pi_{1000} + \pi_{0000}. \quad (11) \]

Further we have,

\[ 0 \leq \pi_{1111}, \pi_{1110}, \pi_{1100}, \pi_{1010}, \pi_{1100}, \pi_{1000}, \pi_{0000} \leq 1. \quad (12) \]

Given (6)-(11), we can express each \( \pi \) by functions of \( p_{ss|d} \) and \( \pi_{1100} \):

\[ \pi_{1111} = p_{11|0}, \]

\[ \pi_{1110} = p_{10|0} - \pi_{1100}, \]

\[ \pi_{1010} = p_{11|1} - p_{11|0} - p_{10|0} + \pi_{1100}, \]

\[ \pi_{1000} = p_{10|1} - \pi_{1100}, \]

\[ \pi_{0000} = p_{00|1}, \]

and subject to the constraint of (12), we have,

\[ \max\{0, p_{11|0} + p_{10|0} - p_{11|1}\} \leq \pi_{1100} \leq \min\{p_{10|0}, p_{10|1}\}. \quad (13) \]

The proportions of each principal stratum, namely, always survivor, never survivor, protected and harmed, are identified under the monotonicity assumption (Assumption 3) no matter whether the second time point survival information is utilized or not. For the fine strata (Table 1) created based on the possible combinations of potential survivals at both the first and second time points, by using the second time point survival information we can narrow their bounds as shown above. If we don’t use the second time point survival information, then for the fine strata, we only know that the proportions of strata 1111, 1110, 1100 sum up to be the proportion of always survivors, and that the proportions of strata 1010, 1000 sum up to be the proportion of protected.
3.2. Bounds for the SACE

In this step, we first derive the bounds for the SACE with known proportions of each fine stratum, then we will combine the result with the bounds obtained in section 3.1 to construct the final bounds for the SACE.

The observable strata of \((Y_i, S_{1i}, S_{2i} \mid D_i)\) are mixtures of potential outcomes from the fine strata. Letting \(q_{y s_1 s_2 \mid d}\) denote \(P(Y_i = y, S_{1i} = s_1, S_{2i} = s_2 \mid D_i = d)\), we have the following identities:

\[
q_{111|1} = \pi_{1111}E(Y_i(1) \mid 1111) + \pi_{1110}E(Y_i(1) \mid 1110) + \pi_{1010}E(Y_i(1) \mid 1010), \tag{14}
\]

\[
q_{110|1} = \pi_{1100}E(Y_i(1) \mid 1100) + \pi_{1000}E(Y_i(1) \mid 1000), \tag{15}
\]

\[
q_{111|0} = \pi_{1111}E(Y_i(0) \mid 1111), \tag{16}
\]

\[
q_{110|0} = \pi_{1110}E(Y_i(0) \mid 1110) + \pi_{1100}E(Y_i(0) \mid 1100). \tag{17}
\]

Recall that

\[
SACE = \frac{(\pi_{1111}E(Y_i(1) \mid 1111) + \pi_{1110}E(Y_i(1) \mid 1110) + \pi_{1100}E(Y_i(1) \mid 1100))}{\pi_{1111} + \pi_{1110} + \pi_{1100}}
- \frac{(\pi_{1111}E(Y_i(0) \mid 1111) + \pi_{1110}E(Y_i(0) \mid 1110) + \pi_{1100}E(Y_i(0) \mid 1100))}{\pi_{1111} + \pi_{1110} + \pi_{1100}}. \tag{18}
\]

Given \(\pi's\), \(\frac{(\pi_{1111}E(Y_i(0)\mid 1111) + \pi_{1110}E(Y_i(0)\mid 1110) + \pi_{1100}E(Y_i(0)\mid 1100))}{\pi_{1111} + \pi_{1110} + \pi_{1100}} = \frac{q_{111|0} + q_{110|0}}{\pi_{1111} + \pi_{1110} + \pi_{1100}}\) which is point identified. Thus to bound the SACE, we only need to bound \(\pi_{1111}E(Y_i(1) \mid 1111) + \pi_{1110}E(Y_i(1) \mid 1110) + \pi_{1100}E(Y_i(1) \mid 1100)\), which defines a linear programming problem:

\[
\text{min} / \text{max} \quad (\pi_{1111}E(Y_i(1) \mid 1111) + \pi_{1110}E(Y_i(1) \mid 1110) + \pi_{1100}E(Y_i(1) \mid 1100)) \mid \pi_{1100} \tag{19}
\]

Subject to:

\[
q_{111|1} = \pi_{1111}E(Y_i(1) \mid 1111) + \pi_{1110}E(Y_i(1) \mid 1110) + \pi_{1010}E(Y_i(1) \mid 1010), \tag{20}
\]

\[
q_{110|1} = \pi_{1100}E(Y_i(1) \mid 1100) + \pi_{1000}E(Y_i(1) \mid 1000). \tag{21}
\]
\[ E(Y_i(1) | 1111) \leq E(Y_i(1) | 1110) \leq E(Y_i(1) | 1100), \quad (22) \]
\[ E(Y_i(1) | 1101) \leq E(Y_i(1) | 1100), \quad (23) \]
\[ E(Y_i(1) | 1110) \leq E(Y_i(1) | 1101) \leq E(Y_i(1) | 1100) \leq E(Y_i(1) | 1000), \quad (24) \]
\[ 0 \leq E(Y_i(1) | 1111), E(Y_i(1) | 1110), E(Y_i(1) | 1100), E(Y_i(1) | 1010), E(Y_i(1) | 1000) \leq 1. \quad (25) \]

where constraints (22)-(24) are imposed by Assumptions 5-7.

The above linear programming problem has a solution if and only if \( \frac{q_{1111}}{p_{1111}} \geq \frac{q_{1110}}{p_{1110}} \), which is an inequality that must be satisfied based on Assumptions 5-7. For each possible value of \( \pi_{1100} \), we solve the above linear programming problem; then, combining this result with the bound for \( \pi_{1100} \) derived in section 3.1 that \( \pi_{1100} \in I \), where \( I = [\max\{0, p_{1110} + p_{1100} - p_{1111}\}, \min\{p_{1010}, p_{1011}\}] \), we have,

\[
\min \text{SACE} = \min_{\pi_{1100} \in I} \left[ \min\left( \frac{\pi_{1111} E(Y_i(1) | 1111) + \pi_{1110} E(Y_i(1) | 1110) + \pi_{1100} E(Y_i(1) | 1100) - (q_{1110} + q_{1100})}{\pi_{1111} + \pi_{1110} + \pi_{1100}} \right) \right]
\]
\[
= \begin{cases} 
\max\{0, \frac{q_{1111} + q_{1110} - p_{1101} - p_{1110} + p_{1100}}{p_{1110} + p_{1100}}, \frac{q_{1111} + q_{1110} + q_{1100}}{p_{1110} + p_{1100}}, \pi_{1111} \} - (q_{1110} + q_{1100}) & \text{if } p_{1110} + p_{1100} - p_{1111} \geq 0 \\
\frac{q_{1111} + q_{1110} + q_{1100}}{p_{1110} + p_{1100}} & \text{if } p_{1110} + p_{1100} - p_{1111} < 0 
\end{cases}
\quad (26)
\]

\[
\max \text{SACE} = \max_{\pi_{1100} \in I} \left[ \max\left( \frac{\pi_{1111} E(Y_i(1) | 1111) + \pi_{1110} E(Y_i(1) | 1110) + \pi_{1100} E(Y_i(1) | 1100) - (q_{1110} + q_{1100})}{\pi_{1111} + \pi_{1110} + \pi_{1100}} \right) \right]
\]
\[
= \frac{q_{1111}}{p_{1111}} - \frac{q_{1110} + q_{1100}}{p_{1110} + p_{1100}} + \frac{q_{1110} p_{1110} - q_{1111} p_{1101}}{p_{1101} + p_{1100}} \min\{p_{1010}, p_{1011}\} \quad (27)
\]

The details of the calculation for the bounds of SACE are provided in the Supplementary Material.

### 3.3. Comparisons with Bounds Under Different Sets of Assumptions

We will compare the bounds under (i) Assumptions 1 to 3 plus our ranked average score assumptions with two time points survival information (Assumptions 5-7) to (ii) Zhang and Rubin’s (2003) bounds which use Assumptions 1 to 3 as well as a ranked average score assumption with one time point survival information (Assumption 4); (iii) only Assumptions 1 to 3. For (ii), the bounds are (Zhang and Rubin...
For (iii), the lower bound on SACE is still (28), and the upper bound on SACE is given by

$$\max SACE = \min\{q_{1111} + q_{1100}, 1\} - \frac{q_{1110} + q_{1100}}{p_{1100}},$$

(30)

Through elementary calculation by comparing (26) and (28), one can easily show that the lower bound for the SACE under (i) will be at least equal to or larger than the lower bound for the SACE under the sets of assumptions (ii) or (iii). Thus when Assumptions 1-3 and ranked average score assumptions with two time points survival information are plausible, the lower bound on the SACE obtained under those assumptions may help identify cases where the treatment benefits survival but harms QOL compared to not utilizing post measurement survival information or not utilizing the survival information at all. Among the three upper bounds under the sets of assumptions (i), (ii) and (iii), the upper bound under (iii) is the largest since it has the least constraints. Depending on the values of \(p'\)'s and \(q'\)'s, the upper bound under (i) can be smaller or larger than the upper bound under (ii). It is smaller in Example 1 in the following section but larger in Example 2. The ranked average score assumptions with two time points survival information, do not imply the ranked average score assumption with one time point survival information. The ranked average score assumptions with two time points survival information allow for the possibility that the always survivors' (1111, 1110, 1100) probability of bad outcome exceed the protected's (1010, 1000) probability of bad outcomes which contradicts the ranked average score assumption with one time point survival information. Also, the ranked average score assumption with one time point survival information does not imply the ranked average score assumptions with two time points survival information. Since the ranked average score assumption with one time point survival
Table 2. Setup 1

<table>
<thead>
<tr>
<th>% of population</th>
<th>Fine Strata</th>
<th>% of $Y_1(1) = 1$</th>
<th>% of $Y_1(0) = 1$</th>
</tr>
</thead>
<tbody>
<tr>
<td>35</td>
<td>1111</td>
<td>20</td>
<td>5</td>
</tr>
<tr>
<td>10</td>
<td>1110</td>
<td>25</td>
<td>10</td>
</tr>
<tr>
<td>20</td>
<td>1100</td>
<td>40</td>
<td>25</td>
</tr>
<tr>
<td>5</td>
<td>1010</td>
<td>35</td>
<td>-</td>
</tr>
<tr>
<td>25</td>
<td>1000</td>
<td>60</td>
<td>-</td>
</tr>
<tr>
<td>5</td>
<td>0000</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

information and ranked average score assumptions with two time points survival information are not nested, the bounds could be narrower under either of the sets of assumptions. In cases where both sets of assumptions are plausible, then we take as the upper bound the smaller of the two upper bounds.

3.4. Numerical Examples

3.4.1. Example 1

Assume that the underlying truth about the population is described by Table 2. The $\text{SACE} = 0.15$, meaning that the treatment will increase the probability of the worse non-mortality outcome by 0.15 among always survivors who will survive at least to the first time point under both treatment and control.

Suppose that we have an infinite sample, then we would observe that

\[
p_{11|1} = 0.50 \quad p_{10|1} = 0.45 \quad p_{00|1} = 0.05 \quad p_{11|0} = 0.35 \quad p_{10|0} = 0.30 \quad p_{00|0} = 0.35 \quad (31)
\]

\[
q_{111|1} = 0.1125 \quad q_{110|1} = 0.23 \quad q_{111|0} = 0.0175 \quad q_{110|0} = 0.06. \quad (32)
\]

Given the constraints imposed by the observed data (31)-(32) and Assumptions 5-7, we obtain the bounds on the SACE: [0.106, 0.238], showing that the treatment increases the probability of the worse non-mortality outcome.

However, if we don't use the second time point survival information, the ob-
Two-stage Censoring by Death

Table 3. Setup 2

<table>
<thead>
<tr>
<th>% of population</th>
<th>Fine Strata</th>
<th>% of $Y_i(1) = 1$</th>
<th>% of $Y_i(0) = 1$</th>
</tr>
</thead>
<tbody>
<tr>
<td>40</td>
<td>1111</td>
<td>35</td>
<td>25</td>
</tr>
<tr>
<td>20</td>
<td>1110</td>
<td>35</td>
<td>25</td>
</tr>
<tr>
<td>20</td>
<td>1100</td>
<td>40</td>
<td>30</td>
</tr>
<tr>
<td>5</td>
<td>1010</td>
<td>35</td>
<td>-</td>
</tr>
<tr>
<td>10</td>
<td>1000</td>
<td>45</td>
<td>-</td>
</tr>
<tr>
<td>5</td>
<td>0000</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

served data would be:

$$P(S_{1i} = 1|D_i = 1) = 0.95 \quad P(S_{1i} = 1|D_i = 0) = 0.65$$  \quad (33)

$$P(Y_i = 1, S_{1i} = 1|D_i = 1) = 0.3425 \quad P(Y_i = 1, S_{1i} = 1|D_i = 0) = 0.0775.$$  \quad (34)

Then, given the constraints imposed by the observed data (33)-(34) and under Assumptions 1-3, the bounds on the SACE are [-0.054, 0.408] which is more than three times as wider as the bounds under Assumptions 1-3 and 5-7. If Assumptions 1-4 are used, then the bounds on the SACE are [-0.054, 0.241]. Based on these bounds that do not use the second time point survival information, we wouldn't know whether or not the treatment increases the probability of the worse non-mortality outcome even though the true SACE is positive. From this example, we see that making use of the survival information after measurement may provide us with more information and narrow the bounds on the SACE.

3.4.2. Example 2

Example 2 shows that the upper bound under the Assumptions 1-3 and 5-7 is not necessarily smaller than the upper bound under Assumptions 1-4.

Assume that the underlying truth about the population is described by Table 3. The true SACE is 0.1. If we have an infinite sample, then we would have the following
observed data:

\[ p_{11|1} = 0.65 \quad p_{10|1} = 0.3 \quad p_{00|1} = 0.05 \quad p_{11|0} = 0.4 \quad p_{10|0} = 0.4 \quad p_{00|0} = 0.2 \quad (35) \]

\[ q_{111|1} = 0.2275 \quad q_{110|1} = 0.125 \quad q_{111|0} = 0.1 \quad q_{110|0} = 0.11. \quad (36) \]

Given the constraints imposed by the observed data (35)-(36) and Assumptions 5-7, we obtain the bounds on the SACE: \([0.0875, 0.1125]\). If we don’t utilize the second time survival information, we would observe the following data:

\[ P(S_{1i} = 1|D_i = 1) = 0.95 \quad P(S_{1i} = 1|D_i = 0) = 0.8 \quad (37) \]

\[ P(Y_i = 1, S_{1i} = 1|D_i = 1) = 0.2525 \quad P(Y_i = 1, S_{1i} = 1|D_i = 0) = 0.21. \quad (38) \]

Then, given the constraints imposed by the observed data (37)-(38) and under Assumptions 1-3, the bounds on the SACE are \([-0.009, 0.178]\) which is more than seven times as wide as the bounds under Assumptions 1-3 and 5-7. Under Assumptions 1-4, the bounds we would obtain for the SACE are \([-0.009, 0.109]\). In this setup, the upper bound under the ranked average score assumptions with two time points survival information (Assumption 5-7) is larger than that under the ranked average score assumption with one time point survival information.

4. Checking the plausibility of ranked average score assumptions with two time points survival information

From the observable data, it cannot be determined whether our ranked average score assumptions with two time points survival information hold. However, there are some necessary conditions that the probability distribution of the observable data must satisfy when these assumptions are valid. If these conditions are violated, then we know our assumptions do not hold.

From the derivation of the bound for SACE in section 3, we know that the linear programming problem (19)-(25) under Assumptions 5-7 as well as the constraints
imposed by the observable "infinite sample" probability distribution has a solution if and only if
\[
\frac{q_{110|1}}{p_{10|1}} \geq \frac{q_{111|1}}{p_{11|1}}.
\]
This constraint says that the probability of the worse non-mortality outcome among the patients that are randomly assigned to treatment and that survive to the first time point but die before the second time point is equal to or larger than the probability of the worse non-mortality outcome among the patients that are randomly assigned to treatment and that survive at least to the second time point. This is a direct consequence from Assumptions 5-7 which say that
\[
E(Y_{i(1)|1111}) \leq E(Y_{i(1)|1110}) \leq E(Y_{i(1)|1010}) \leq E(Y_{i(1)|1100}) \leq E(Y_{i(1)|1000}).
\]
The first three expectations are for subjects who can survive at least to the second time point under treatment and the last two expectations are for subjects who die before the second time point.

The above constraint to check the plausibility of our assumptions are for "infinite sample" data. In practice, we can estimate the confidence with which the true observable population distribution satisfies the above constraints using a simple bootstrap procedure (Efron and Tibshirani, 1998). We bootstrap from the empirical distribution of the observed data and then count the percentage of the bootstrapped data sets for which the empirical distribution satisfies the constraints as an estimate of the confidence. Efron and Tibshirani (1998) provide some refinements on this simple bootstrap procedure that improve the accuracy of the estimated confidence.

5. Estimation and Inference

In section 3, the bounds we obtained are "infinite sample" bounds where we assume that the joint distribution of \((D_i, S_{1i}, S_{2i}, Y_i)\) is known. In empirical settings, this joint distribution is unknown and could only be estimated from the observed data, we need to account for the sampling uncertainty to draw inference. However, because the derived bounds involve minimum and maximum operators, the sam-
ple analog estimates of the bounds and the standard resampling techniques (e.g., the bootstrap methods) to obtain confidence intervals may fail to provide valid inference. Due to the concavity of the minimum operator and the convexity of the maximum operator, the sample analog estimates of the bounds tend to be narrower than the true bounds (Manski and Pepper, 2000, 2009), and Hirano and Porter (2012) showed when the estimands of interest are nondifferentiable functionals of the underlying data distribution, such as our case which involves minimum and maximum operators, achieving unbiased estimators is impossible in general. And also because of the minimum and maximum operators in the bounds, closed form characterization of the asymptotic distribution of these estimators is difficult to derive, and the canonical bootstrap techniques to construct confidence intervals are not necessarily uniformly valid (Andrews, 2000; Andrews and Guggenberger, 2009; Romano, 1989; Romano and Shaikh, 2008, 2010). These problems have generated a growing literature on inference methods for partially identified parameters (see Tamer, 2010 for a review).

In this paper, we will construct confidence intervals for the bounds using the method developed by Chernozhukov, Lee and Rosen (2013) (hereafter CLR) where they address the problem of inference for intersection bounds by proposing bias-corrected confidence intervals for the identification region. A brief summary of the CLR approach can be found in Flores and Flores-Lagunes (2013), and step by step implementation of CLR’s procedure is provided in Section 6 of CLR and the online Appendix of Flores and Flores-Lagunes (2013). CLR also proposed biased corrected estimators of the upper and lower bounds to address the problem of estimation of intersection bounds. Their proposed bias-corrected estimators are half-median unbiased in the sense that the upper bound estimator exceeds the true upper bound with probability at least one half asymptotically, and the reverse holds for the lower bound. The bounding functions involved in the intersection bounds in our case are simple functions of proportions, thus the sample analog estimators of those bound-
ing functions are consistent and asymptotically normally distributed. Theory is developed in CLR for large sample inference based on the strong approximation of a sequence of empirical processes by a sequence of Gaussian processes. This theory guarantees that the 95% CLR confidence intervals for our bounds have at least 95% coverage asymptotically.

We did some simulation studies to examine the performance of the CLR confidence intervals and their half-median unbiased estimators for finite samples. The results showed the proposed half-median unbiased estimator could be conservative (e.g. the upper bound estimator exceeds the true upper bound with probability around 0.7 in our simple simulation study presented in the Supplementary Materials). Detailed results of our simulation studies are presented below and in the Supplementary Materials.

In section 3.4, we discussed two numerical examples with the unrealistic “infinite sample” case. We now use these numerical examples to examine the finite sample performance of CLR 95% confidence intervals (CIs) and CLR half-median unbiased estimators. For each example, we considered sample size 500, 1000 and 2000. For each sample size, we simulated 2000 samples based on the known joint distribution of \((D_i, S_{1i}, S_{2i}, Y_i)\). Then for each simulated data set, we obtain the 95% CIs for the identification region and half-median unbiased estimates for upper and lower bounds. Since we know the underlying true bounds under the three different sets of assumptions (i), (ii) and (iii) described in section 3.3, we can count the proportions of the two thousand CIs that cover the true whole intervals under each set of assumptions and the proportions of the two thousand half-median unbiased estimates of the upper bounds exceeding the true upper bounds under each set of assumptions, similarly for the lower bounds. From Table 4, we see that the finite sample coverages of the 95% CLR CIs as well as the half-median unbiased estimators are good if slightly conservative.

From Table 4, we see that with finite samples, utilizing the second time point sur-
Survival information (Assumptions 5-7) also has a significant advantage over utilizing one time point survival information (Assumption 4) or not using survival information at all. The average lengths of the 95% CIs under Assumption 1-3 and the ranked average score assumptions with two time points survival information are the shortest for all the sample sizes and setups considered, and the reductions of the lengths are significant based on the standard deviations (SD. Length 95% CI). The last column of Table 4, % Inf. CI, reports the percentages of 95% CIs that provide informative bounds, i.e. correctly specifying whether treatment harms or benefits QOL by not covering 0 or any values opposite to the sign of the treatment effect. Using two time points survival information makes it much easier to find informative bounds than using only one time point survival information or no survival information, both of which could hardly ever detect informative bounds. As expected, it is harder to find informative bounds as the sample size decreases.
Table 4. Simulation results examining the performance of CLR 95% CIs and the CLR half-median unbiased estimates (HMUE) for the bounds of the SACE. % Inf. CI provide the percentages of 95% CIs giving informative bounds (not covering 0) for each combination of setup and sample size under different sets of assumptions.

<table>
<thead>
<tr>
<th>Setup</th>
<th>Sample Size</th>
<th>Assumptions</th>
<th>Cov. Prob</th>
<th>Ave. Length</th>
<th>SD. Length</th>
<th>Cov. Prob 95% CI</th>
<th>Cov. Prob 95% CI</th>
<th>Cov. Prob 95% CI</th>
<th>Cov. Prob HMUE Lower</th>
<th>Cov. Prob HMUE upper</th>
<th>% Inf. CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Example 1</td>
<td>2000</td>
<td>1-3&amp;5-7</td>
<td>96.25%</td>
<td>0.224</td>
<td>0.015</td>
<td>65.25%</td>
<td>58.85%</td>
<td>93.1%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>1-4</td>
<td>95.55%</td>
<td>0.404</td>
<td>0.020</td>
<td>67.75%</td>
<td>49.05%</td>
<td>0.05%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>1-3</td>
<td>94.85%</td>
<td>0.588</td>
<td>0.034</td>
<td>67.75%</td>
<td>48.40%</td>
<td>0.05%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1000</td>
<td>1-3&amp;5-7</td>
<td>97.7%</td>
<td>0.263</td>
<td>0.021</td>
<td>64.90%</td>
<td>58.75%</td>
<td>84.60%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>1-4</td>
<td>96.55%</td>
<td>0.439</td>
<td>0.024</td>
<td>67.60%</td>
<td>50.95%</td>
<td>0.1%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>1-3</td>
<td>95.75%</td>
<td>0.632</td>
<td>0.045</td>
<td>67.60%</td>
<td>51.05%</td>
<td>0.1%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>500</td>
<td>1-3&amp;5-7</td>
<td>96.55%</td>
<td>0.328</td>
<td>0.036</td>
<td>73.55%</td>
<td>59.80%</td>
<td>45.15%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>1-4</td>
<td>95.15%</td>
<td>0.482</td>
<td>0.032</td>
<td>65.80%</td>
<td>51.55%</td>
<td>0.3%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>1-3</td>
<td>94.95%</td>
<td>0.685</td>
<td>0.064</td>
<td>65.80%</td>
<td>50.30%</td>
<td>0.3%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Example 2</td>
<td>2000</td>
<td>1-3&amp;5-7</td>
<td>95.9%</td>
<td>0.122</td>
<td>0.013</td>
<td>62.25%</td>
<td>53.65%</td>
<td>93.25%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>1-4</td>
<td>94.8%</td>
<td>0.216</td>
<td>0.015</td>
<td>51.15%</td>
<td>51.05%</td>
<td>1.1%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>1-3</td>
<td>94.8%</td>
<td>0.293</td>
<td>0.023</td>
<td>51.15%</td>
<td>51.60%</td>
<td>1.1%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1000</td>
<td>1-3&amp;5-7</td>
<td>95.8%</td>
<td>0.163</td>
<td>0.019</td>
<td>60.75%</td>
<td>52.00%</td>
<td>64.60%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>1-4</td>
<td>94.9%</td>
<td>0.257</td>
<td>0.023</td>
<td>50.15%</td>
<td>49.65%</td>
<td>1.25%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>1-3</td>
<td>95.2%</td>
<td>0.336</td>
<td>0.034</td>
<td>50.15%</td>
<td>48.45%</td>
<td>1.25%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>500</td>
<td>1-3&amp;5-7</td>
<td>96.85%</td>
<td>0.230</td>
<td>0.039</td>
<td>62.70%</td>
<td>53.95%</td>
<td>32.25%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>1-4</td>
<td>94.55%</td>
<td>0.323</td>
<td>0.035</td>
<td>60.80%</td>
<td>51.55%</td>
<td>2.15%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>1-3</td>
<td>94.35%</td>
<td>0.407</td>
<td>0.051</td>
<td>60.80%</td>
<td>51.10%</td>
<td>2.15%</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
6. Application to ARDSNet Study

The ARDSNet study described in the introduction involved 861 patients with lung injury and acute respiratory distress syndrome who were randomized to receive mechanical ventilation with either lower tidal volumes or traditional tidal volumes. The non-mortality outcome variable we are interested in is whether patients were able to breathe without assistance by day 28 which is a measurement that reflects the quality of life for patients after treatment. We use $Y_i$ to represent this binary quality of life measurement, with $Y_i$ being 1 indicating that the $i^{th}$ patient was not able to breathe without assistance by day 28. Naturally, the first survival time point is day 28 after the treatment. If the patient died before day 28, then the non-mortality outcome could not be measured, thus will be undefined. The second time point survival indicator is whether the patient was eventually discharged home with unassisted breathing or not. We view the patients who received mechanical ventilation with lower tidal volume as the treatment group, and the patients who received mechanical ventilation with traditional tidal volume as the control group. Let $D_i$ equal 1 if the $i^{th}$ patient is randomized to treatment group, 0 if randomized to control group. Further details on the data are described in the Supplementary Material.

Table 5 presents the observed strata of $(D_i, S_{1i}, S_{2i}, Y_i)$. Among the survivors in the lower tidal volume group, the proportion of patients that cannot breathe without assistance by day 28 is 17.03% (which is 55/323); among the survivors in the traditional tidal volume group, the proportion of patients that cannot breathe without assistance by day 28 is 21.30% (which is 59/277). The difference of those two proportions $-4.27\%$ which is a direct comparison of the QOL among survivors in the lower tidal volume and survivors in the traditional tidal volume is likely an upward biased estimate for the SACE due to the informativeness of censoring by death.

The empirical distribution of $(D_i, S_{1i}, S_{2i}, Y_i)$ satisfies the constraint (39). Using the bootstrap procedure, all of the 2000 bootstrapped datasets satisfy the constraint (39), thus we are very confident that our set of two stage assumptions is plausible in
the sense that it does not violate the constraint (39).

Table 6 compares the CLR half-median unbiased estimates for the bounds of the SACE as well as the 95% confidence intervals obtained under three different sets of assumptions: (i) Assumptions 1-3 and ranked average score assumptions with two time points survival information (Assumptions 5-7); (ii) Assumptions 1-3 and ranked average score assumption with one time points survival information (Assumption 4); and Assumptions 1-3 only.

**Table 5.** Observed data for ARDSNet Study

<table>
<thead>
<tr>
<th>Number of Patients</th>
<th>$D_i$</th>
<th>$S_{1i}$</th>
<th>$S_{2i}$</th>
<th>$Y_i$</th>
</tr>
</thead>
<tbody>
<tr>
<td>258</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>29</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>10</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>26</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>109</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>–</td>
</tr>
<tr>
<td>211</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>34</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>7</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>25</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>152</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>–</td>
</tr>
</tbody>
</table>

**Table 6.** The CLR half-median unbiased estimates (HMUE) for the bounds of the SACE, and CLR 95% CIs of the SACE for the ARDSNet study under different sets of assumptions: Assumptions 1-3 and ranked average score assumptions with two time points survival information (Assumptions 5-7); Assumptions 1-3 and ranked average score assumption with one time points survival information (Assumption 4); and Assumptions 1-3 only.

<table>
<thead>
<tr>
<th>SACE</th>
<th>Assumptions 1-3 and 5-7</th>
<th>Assumptions 1-3 and 4</th>
<th>Assumptions 1-3</th>
</tr>
</thead>
<tbody>
<tr>
<td>CLR HMUE</td>
<td>$[-16.25%, -3.49%]$</td>
<td>$[-19.64%, -4.27%]$</td>
<td>$[-19.64%, -1.58%]$</td>
</tr>
<tr>
<td>CLR 95% CI</td>
<td>$[-23.89%, 2.42%]$</td>
<td>$[-27.09%, 2.55%]$</td>
<td>$[-27.09%, 5.58%]$</td>
</tr>
</tbody>
</table>
sumption 4); and (iii) Assumptions 1-3 only. According to the result of our two time points survival information analysis with the set of assumptions (i), among the patients with lung injury and the acute respiratory distress syndrome who would survive under both ventilation tidal volumes, the lower tidal volume would help reduce the probability of breathing with assistance by day 28 by an amount between 3.39% to 16.25%. This bound for the SACE is shorter, thus more informative, than the bound obtained through the one time point survival information analysis with the set of assumptions (ii) which estimates the reduction to be between [4.27%, 19.64%], and is also substantially shorter, thus more informative, than the bound obtained utilizing no survival information with the set of assumptions (iii) which estimates the reduction to be between [1.58%, 19.64%]. The same comparison also holds for the width of the 95% confidence intervals. The 95% confidence intervals under all sets of assumptions cover 0, meaning that there is not strong evidence that ventilation with lower tidal benefits patients in terms of the QOL outcome of breathing without assistance by day 28.

7. Conclusions and Discussions

The effect of treatment on a non-mortality outcome among always survivors is of interest in many clinical studies. The previous literature on bounding the SACE uses only the survival information before the measurement of the non-mortality outcome; however, in many cases, the survival information after the measurement of non-mortality outcome is informative. We proposed a set of ranked average score assumptions with two time points survival information which are plausibly satisfied in many quality of life studies and developed a two-step linear programming approach to obtain the closed form of the bounds of the SACE under our assumptions. Our method works not only for randomized trials with perfect compliance, but also can be extended to randomized trials with noncompliance or observational studies with a valid IV to obtain bounds on the complier survivor average causal effect.
We applied our method to the ARDSNet study. Making use of the post QOL measurement survival information (patients’ status when discharged home) in addition to the pre-QOL survival information (survival status at day 28) helps shorten the bound on the SACE – the effect of lower tidal volume on being able to breathe without assistance by day 28.

The SACE and CSACE are principal strata effects, causal effects on a subgroup of patients defined by the values that post-randomization variables would take under both treatment and control (Frangakis and Rubin, 2002). We have shown that bounds on these principal strata effects can be sharpened by using the further outcome information of survival after the non-mortality outcome is measured. In a different context, Mealli and Pacini (2013) showed that using further outcomes can narrow bounds on principal strata effects. Mealli and Pacini (2013) consider an outcome that is not affected by censoring by death in a randomized trial with noncompliance, and study bounds on the intention to treat effects for the compliers, always takers and never takers. Mealli and Pacini (2013) consider settings in which the exclusion restriction may not be satisfied and they show that a secondary outcome for which the exclusion restriction is satisfied can be used to narrow the bounds. The same idea of including information on a larger set of outcomes to tighten information on the causal estimand of interest is also exploited by Mattei, Li and Mealli (2013).

So far, we have assumed that the treatment assignment is random. Our method can also be naturally extended to the cases in which conditional on some discrete covariates there is ignorability such that the subjects are randomized. We can stratify the subjects into subsets defined by each level of covariates, and apply our method to obtain the bounds on the SACE within each subgroup. Then we can obtain the overall bounds on the SACE combining the proportions of each subgroup. See Freiman and Small (2014) for more details on this topic. How to deal with the case in which the covariates are continuous requires further research.
In this paper, we discussed ideas of utilizing post outcome measurement survival information in censoring by death studies in the context of binary quality of life. The ideas and the methodology can be naturally extended to ordinal quality of life outcomes. To save space, the generalization to ordinal quality of life outcomes is discussed in the Supplementary Material.

Although the paper is presented in the context of quality of life outcomes censored by the event “death”, the idea of using post outcome measurement information has a wide application to studies where the outcome of interest is only measured if another event does not occur before it. For example, in studying the effect of an initiative to support marriage on quality of marriage (introduced in the Introduction), suppose the quality of marriage is measured at three months after the initiative, then the censoring event is “divorce” before three months. If we also know whether each couple divorced or not at, say, 12 months after the initiative, this provides more detailed information on the quality of marriage scores since divorce at 12 months is related to the quality of the marriage at 3 months which makes Assumptions 5-7 reasonable.

The monotonicity assumption (Assumption 3) is often plausible for the studies we considered and have described in the paper, especially in prevention trials in which a treatment is compared with placebo. However, in a clinical trial in which a new treatment is launched and is compared with a standard treatment, monotonicity assumption may not be plausible. It is of future research interest to study to what extent the second time survival information could help tighten the bounds on the SACE when monotonicity does not hold.

In this study, we focus on studies where the non-mortality outcome is measured at a fixed time for all subjects. However, there are cases where the non-mortality outcome might be measured at different time for different subjects which complicates the analysis. For instance, IVH may happen at any time in the first several days of life of babies. How to handle the situation in which the non-mortality outcome
Two-stage Censoring by Death

could be measured at continuous times is a topic we are working on.

Acknowledgements:

We thank Scott Halpern, Mark Mikkelsen and Scott Lorch for helpful discussions of clinical issues relating to censoring by death. We thank Roland Ramsahai for insightful discussions on causal bounds.

References


Yang and Small


Freiman, M. and Small, D. (2014). Large sample bounds on the survivor average causal effect when outcomes are censored by death. (Under review)


