CAUSAL EFFECTS OF TREATMENTS FOR INFORMATIVE MISSING DATA DUE TO PROGRESSION/DEATH

Keunbaik Lee,[1,*] **Michael J. Daniels**[2,†] **and Daniel J. Sargent**[3,‡]

[1]Assistant Professor, Biostatistics Program, Louisiana State University-Health Sciences Center, New Orleans, LA 70112, USA

[2]Professor, Department of Statistics, University of Florida, Gainesville FL 32611, U.S.A.

[3]Professor of Biostatistics and Oncology, Division of Biomedical Statistics and Informatics, Mayo Clinic, Rochester, MN 55905, U.S.A.

August 21, 2009

Summary.

In longitudinal clinical trials, when outcome variables at later time points are only defined for patients who survive to those times, the evaluation of the causal effect of treatment is complicated. In this paper, we describe an approach that can be used to obtain the causal effect of three treatment arms with ordinal outcomes in the presence of death using a principal stratification approach. We introduce a set of flexible assumptions to identify the causal effect and implement a sensitivity analysis for non-identifiable assumptions which we parameterize parsimoniously. Methods are illustrated on quality of life data from a recent colorectal cancer clinical trial.

Key words: Principal stratification; QOL; Ordinal data; Sensitivity analysis.

[*] *email:* klee4@lsuhsc.edu
[†] *email:* mdaniels@stat.ufl.edu
[‡] *email:* sargent.daniel@mayo.edu

1. Introduction

In longitudinal clinical trials, outcome variables at later time points are only defined for patients who survive to those times. This complicates analyses, since those who survive on each treatment arm to any given time point are not a random sample of those randomized to each treatment arm. Standard approaches to account for missing data in the literature implicitly 'impute' values of outcome after dropout. For quality of life (QOL) data, if a subject drops out due to deteriorating disease status or death, the QOL will not be defined after the dropout time. If we restrict the analyses to subgroup of patients who survived, the resulting estimate of the treatment effect will no longer have a causal interpretation.

One way to conduct analyses in this circumstance is to construct the joint distribution of the longitudinal responses and death times (Hogan and Laird, 1997; Pauler et al., 2003; Kurland and Heagerty, 2005; Lee and Daniels, 2007). Hogan and Laird (1997) proposed a likelihood-based mixture model approach to model the joint distribution for incomplete repeated measurements and right-censored event times. Pauler et al. (2003) proposed a pattern mixture model for longitudinal quality of life data with nonignorable missingness due to dropout and censorship by death. Recently, Kurland and Heagerty (2005) explored regression models conditioning on being alive as a valid target of inference using regression models that condition on survival status rather than a specific survival time. Lee and Daniels (2007) used a pattern mixture approach with patterns defined on the observed tumor progression or death times which is the similar to ideas in Hogan and Laird (1997) and Pauler et al. (2003). Each of these approaches to drawing inference in the presence of death in effect, stratify the analysis using events that occur after randomization, and do not maintain the causal interpretation of the treatment effect.

To obtain the causal effect of treatment, we start by defining potential outcomes. Potential outcomes are all the outcomes that would be observed if each of the treatments

had been applied to each of the subjects (Rubin, 1974, 1978; Holland, 1986). Frangakis and Rubin (2002) used the concept of potential outcomes in an approach called 'principal stratification'. Principal stratification partitions subjects into sets with respect to posttreatment variables. For example, consider two treatments, Tx $= 0, 1$ and a binary response $Y(\text{Tx})$. The causal effect of the treatment on response is $Y(1) - Y(0)$. However, we only observe either $Y(1)$ or $Y(0)$ for each subject. Let $D(\text{Tx})$ be a binary posttreatment variable. In this case, there are four principal strata defined by the pairs of potential values of $D(\text{Tx})$, (i) $(\{i|D(0) = 0, D(1) = 0\})$, (ii) $(\{i|D(0) = 0, D(1) = 1\})$, (iii) $(\{i|D(0) = 1, D(1) = 0\})$, and (iv) $(\{i|D(0) = 1, D(1) = 1\})$. The principal strata are not affected by treatment assignment. Causal effects are defined within these principal strata.

Frangakis and Rubin (2002) discussed causal effects using studies where the outcome does not exist due to death. Rubin (2000) and Hayden et al. (2005) referred to the estimand in Frangakis and Rubin (2002) as 'the survivors average causal effect (SACE)'. Egleston et al. (2007) proposed assumptions to identify the SACE and implemented a sensitivity analysis for some of those assumptions among survivors at a fixed time point in the cohort study of older adults. Rubin (2006) introduced the causal effect of a treatment on a outcome that is censored by death in a Quality of Life study. Recently, Lee and Daniels (2008) used a 'principal stratification' approach to estimate the SACE for ordinal data in a two treatment arm setting.

The identification of the $SACE$ relies on untestable assumptions. For example, the widely used monotonicity assumption (in the context of death) states that "no subject would be dead if randomized to an active treatment, but would be alive if randomized to a control treatment". This assumption (and variations) has been used in a variety of settings, including death and noncompliance (Zhang and Rubin, 2003; Gilbert et al., 2003;

Cheng and Small, 2006; Egleston et al., 2007; Lee and Daniels, 2008; Cheng, 2009; Imai 2008; Jin and Rubin, 2008). Stochastic monotonicity assumptions have been developed recently as a weaker and more realistic alternative in the context of compliance (Roy et al., 2008) and instrumental variables (Small and Tan, 2008).

In this paper, we describe an approach that can be used to obtain the causal effect of treatment in the presence of death in a setting of three treatment arms with longitudinal ordinal outcomes. We replace the common deterministic monotonicity assumption with a more flexible set of assumptions that allows incorporation of subject specific covariates to increase its applicability to our setting. Overall, we provide a realistic framework that has only four sensitivity parameters for longitudinal ordinal data involving comparisons among three treatments. Although our main interest will be treatment differences at the end of the study, we will make efficient use of all available longitudinal data which is essential due to the high rate of dropouts unrelated to death.

The paper is arranged as follows. In section 2, we describe the clinical trial that motivated our analyses. In section 3, we formally define the causal effects of interest, provide some assumptions that are necessary to identify them, and then derive how the causal effects can be estimated from these assumptions along with modeling assumptions for the observed data in this complex setting. We also propose a sensitivity analysis procedure for some of these untestable assumptions. Methods are illustrated on the quality of life data in section 4. Section 5 contains conclusions and a discussion.

2. Quality of Life Data in Clinical Trials

As a secondary outcome, many clinical trials measure aspects of a patient's sense of well-being and ability to perform various tasks to assess the effects that cancer and its treatment have on the patient. Such quality of life (QOL) outcomes in clinical trials are generally obtained repeatedly over time using questionnaires that can be completed by

the patient without direct supervision. These standardized questionnaires rate factors such as pain, treatment side-effects, mood, energy level, family and social interactions, sexual function, ability to work, and ability to maintain routine daily activities. Here, we describe the QOL data set that we will work.

We analyze QOL data from a recent colorectal cancer clinical trial (Trial N9741, Goldberg et al., 2004). The trial randomized patients with advanced colorectal cancer to compare three different two drug combinations. The main objective of the trial was to find a better treatment for colorectal cancer. The trial led to a new standard of care.

A total of 795 patients with colorectal cancer were randomly assigned to one of three treatments between May 1999 and April 2001. Patients received either irinotecan and bolus fluorouracil plus leucovorin (IFL arm), oxaliplatin and infused fluorouracil plus leucovorin (FOLFOX arm), or irinotecan and oxaliplatin (IROX arm). At the time the trial was initiated, the IFL regimen was the FDA approved standard of care (Saltz et al., 2000). Based on the primary endpoints of time to progression and overall survival, the FOLFOX arm was identified as the superior treatment and FOLFOX was also comparatively safe relative to the other treatments.

However, given that the toxicity profiles were quite different on the three treatment arms, it was of interest to see if there was a negative impact of the 'better' treatment (FOLFOX) on patient's QOL. In this paper we compare the quality of life data on the three treatments.

2.1 *QOL outcomes*

The QOL outcomes measured here included fatigue, nausea, vomiting, diarrhea, and dehydration. Previous analyses split the QOL scores into 12 week windows. The patient's visit is the window number during which the survey was filled out (0=baseline, 1=1-84 days after going on study, 2=85-168 days after going on study,..., 5=337-420 days after

going on study). We will analyze the data from the first five windows (up to about one year).

For illustration, we focus on one QOL measure, fatigue. Fatigue is measured on a 5 point ordinal scale (1: I am usually not tired at all; 2: I am occasionally rather tired; 3: There are frequently periods when I am quite tired; 4: I am usually very tired; 5: I feel exhausted most of the time). Because very few patients reported category 5, we collapsed category 4 and 5 into one category.

2.2 *Dropouts*

Similar to other longitudinal studies, the QOL data collection in this clinical trial had substantial dropout. The overall dropout rates were similar in the three treatment groups (96% for IFL; 98% for IROX; 98% for FOLFOX). About 95% of dropouts occurred between baseline and the fourth visit. Dropout was related to whether cancer had progressed, death, toxicity, and other factors. A large number of subjects dropped out due to tumor progression or death (40.5%). The dropout rates due to progression/death were marginally higher on the IFL and IROX arms than on the FOLFOX arm. A interesting feature of this dataset is that we have the actual progression/death times for all subjects including those who dropped out for other reasons. More detailed description of the pattern of dropouts in this study is given in Lee and Daniels (2007).

We will use this data to illustrate the models and methods introduced in the subsequent sections.

3. Principal Stratification Approach

Consider a three-arm trial. In our application the three treatments were IFL (standard of care) and FOLFOX and IROX (both experimental). Let Tx_i be the treatment indicator

for subject i,

$$\mathrm{Tx}_i = \begin{cases} 0, & \text{if subject } i \text{ is assigned to IFL;} \\ 1, & \text{if subject } i \text{ is assigned to IROX;} \\ 2, & \text{if subject } i \text{ is assigned to FOLFOX,} \end{cases}$$

and let $D_{it}(\mathrm{Tx}_i)$ be tumor-progression/death indicator for subject i at visit t,

$$D_{it}(\mathrm{Tx}_i) = \begin{cases} 0, & \text{alive;} \\ 1, & \text{tumor progressed or dead.} \end{cases}$$

Obviously, if $D_{it}(\mathrm{Tx}_i) = 1$, then $D_{is}(\mathrm{Tx}_i) = 1$ for $s > t$. To maintain clarity, we will refer to the tumor progression/death indicator $D_{it}(\cdot)$ as the death indicator for rest of this paper.

We define the following 8 (principal) strata:

1. $A_t(0) = \{i|(D_t(0), D_t(1), D_t(2)) = (0, 0, 0)\}$: the subjects who would be alive under all three arms at visit t.

2. $A_t(1) = \{i|(D_t(0), D_t(1), D_t(2)) = (1, 0, 0)\}$: the subjects who would be alive under the two experimental treatments but not alive under the standard of care (IFL) at visit t.

3. $A_t(2) = \{i|(D_t(0), D_t(1), D_t(2)) = (1, 1, 0)\}$: the subjects who would be alive under FOLFOX but not alive under the standard of care (IFL) and IROX at visit t.

4. $A_t(3) = \{i|(D_t(0), D_t(1), D_t(2)) = (1, 0, 1)\}$: the subjects who would be alive under IROX but not alive under the standard of care (IFL) and FOLFOX at visit t.

5. $A_t(4) = \{i|(D_t(0), D_t(1), D_t(2)) = (0, 0, 1)\}$: the subjects who would be alive under the standard of care (IFL) and IROX but not alive under FOLFOX at visit t.

6. $A_t(5) = \{i|(D_t(0), D_t(1), D_t(2)) = (0, 1, 0)\}$: the subjects who would be alive under the standard of care (IFL) and FOLFOX but not alive under IROX at visit t.

7. $A_t(6) = \{i|(D_t(0), D_t(1), D_t(2)) = (0, 1, 1)\}$: the subjects who would be alive under standard of care (IFL) but not alive under the two experimental treatments at visit t.

8. $A_t(7) = \{i|(D_t(0), D_t(1), D_t(2)) = (1, 1, 1)\}$: the subjects who would not be alive under any treatment at visit t.

Now, let $Y_{it}(\text{Tx}_i)$ be the potential outcome for subject i at time t. The full set of potential outcomes at visit t is

$$\mathcal{P}_t = \{D_t(0), (Y_t(0); D_t(0) = 0), D_t(1), (Y_t(1); D_t(1) = 0), D_t(2), (Y_t(2); D_t(2) = 0)\},$$

where, for treatment tx, $Y_t(\text{tx})$ is the potential outcome if a subject is alive at visit t ($D_t(\text{tx}) = 0$). If the patient is alive ($D_{it}(\text{Tx}_i) = 0$), define R_{it} to be the indicator that $Y_{it} = Y_{it}(\text{Tx}_i)$ is observed. In the following, we assume monotone dropout ($R_{it} = 1 \Rightarrow R_{it-1} = 1$). W_i and S_i respectively denote vector of baseline covariates and progression/death time ($S_i = j \Leftrightarrow \{D_{ij} = 1, D_{ij-1} = 0\}$). So, the observed data for individual i is

$$\mathcal{O}_{it} = \{W_i, \text{Tx}_i, S_i, D_{it}, (R_{it}; D_{it} = 0), (Y_{it}; D_{it} = 0, R_{it} = 1)\}.$$

Unfortunately, given $\mathcal{O}_{it}$ alone, we cannot determine to which strata individual i belongs.

In the next section, we formally define the causal effects of interest and then state some additional assumptions that are necessary to estimate them.

3.1 *Defining the Causal Effects*

Suppress i here for clarity. We assume each individual at visit t belongs to one of 8 basic principal strata (Frangakis and Rubin, 2002) defined by the unique combinations of $(D_t(0), D_t(1), D_t(2))$. Some of these strata will be empty because of the monotonicity assumptions (Assumption I in section 3.2).

For each pairwise treatment comparison, we define the three causal effects of interest at time T,

1.

$$\begin{aligned}
SACE_{k,1}(1,0) &= \frac{odds\left\{Y_T(1) > k | (D_T(0), D_T(1), D_T(2)) = (0,0,0)\right\}}{odds\left\{Y_T(0) > k | (D_T(0), D_T(1), D_T(2)) = (0,0,0)\right\}} \\
&= \frac{odds\left\{Y_T(1) > k | A_T(0)\right\}}{odds\left\{Y_T(0) > k | A_T(0)\right\}}, \\
SACE_{k,2}(1,0) &= \frac{odds\left\{Y_T(1) > k | (D_T(0), D_T(1), D_T(2)) = (0,0,1)\right\}}{odds\left\{Y_T(0) > k | (D_T(0), D_T(1), D_T(2)) = (0,0,1)\right\}} \\
&= \frac{odds\left\{Y_T(1) > k | A_T(4)\right\}}{odds\left\{Y_T(0) > k | A_T(4)\right\}}, \\
SACE_{k,3}(1,0) &= \frac{odds\left\{Y_T(1) > k | A_T(0) \cup A_T(4)\right\}}{odds\left\{Y_T(0) > k | A_T(0) \cup A_T(4)\right\}},
\end{aligned}$$

2.

$$\begin{aligned}
SACE_{k,1}(2,0) &= \frac{odds\left\{Y_T(2) > k | (D_T(0), D_T(1), D_T(2)) = (0,0,0)\right\}}{odds\left\{Y_T(0) > k | (D_T(0), D_T(1), D_T(2)) = (0,0,0)\right\}} \\
&= \frac{odds\left\{Y_T(2) > k | A_T(0)\right\}}{odds\left\{Y_T(0) > k | A_T(0)\right\}}, \\
SACE_{k,2}(2,0) &= \frac{odds\left\{Y_T(2) > k | (D_T(0), D_T(1), D_T(2)) = (0,1,0)\right\}}{odds\left\{Y_T(0) > k | (D_T(0), D_T(1), D_T(2)) = (0,1,0)\right\}} \\
&= \frac{odds\left\{Y_T(2) > k | A_T(5)\right\}}{odds\left\{Y_T(0) > k | A_T(5)\right\}}, \\
SACE_{k,3}(2,0) &= \frac{odds\left\{Y_T(2) > k | A_T(0) \cup A_T(5)\right\}}{odds\left\{Y_T(0) > k | A_T(0) \cup A_T(5)\right\}},
\end{aligned}$$

3.

$$
\begin{aligned}
SACE_{k,1}(2,1) &= \frac{odds\left\{Y_T(2) > k | (D_T(0), D_T(1), D_T(2)) \in (0,0,0)\right\}}{odds\left\{Y_T(1) > k | (D_T(0), D_T(1), D_T(2)) \in (0,0,0)\right\}} \\
&= \frac{odds\left\{Y_T(2) > k | A_T(0)\right\}}{odds\left\{Y_T(1) > k | A_T(0)\right\}}, \\
SACE_{k,2}(2,1) &= \frac{odds\left\{Y_T(2) > k | (D_T(0), D_T(1), D_T(2)) \in (1,0,0)\right\}}{odds\left\{Y_T(1) > k | (D_T(0), D_T(1), D_T(2)) \in (1,0,0)\right\}} \\
&= \frac{odds\left\{Y_T(2) > k | A_T(1)\right\}}{odds\left\{Y_T(1) > k | A_T(1)\right\}}, \\
SACE_{k,3}(2,1) &= \frac{odds\left\{Y_T(2) > k | A_T(0) \cup A_T(1)\right\}}{odds\left\{Y_T(1) > k | A_T(0) \cup A_T(1)\right\}},
\end{aligned}
$$

There are three SACEs for comparing IROX and IFL, three for comparing FOLFOX and IFL, and three for comparing IROX and FOLFOX. $SACE_{k,1}(1,0)$ is the odds ratio that the outcome at time T is larger than k for comparing IROX and IFL for subjects who would be alive under all three treatment arms. $SACE_{k,2}(1,0)$ is the odds ratio that the outcome at time T is larger than k for comparing IROX and IFL for subjects who would be alive under IROX and IFL but not alive under FOLFOX. $SACE_{k,3}(1,0)$ is the odds ratio that the outcome at time T is larger than k for comparing IROX and IFL for subjects who would be alive either under all three treatment arms or only under IROX and IFL. $SACE_{k,\cdot}(2,0)$ and $SACE_{k,\cdot}(2,1)$ are defined similarly for comparisons of FOLFOX and IFL, and FOLFOX and IROX, respectively.

3.2 *Assumptions for Identifiability of a Causal Effect*

Define $\overline{\mathcal{P}}_t = (\mathcal{P}_1, \cdots, \mathcal{P}_t)$ to be the history of potential outcomes up to and including the outcomes at visit t. Before introducing the assumptions, it is convenient to define some additional quantities. Let $g_T(\text{tx}, w) = P(D_T(\text{tx}) = 0 | w)$ for $\text{tx} = 0, 1, 2$ where w is

vector of baseline covariates. We assume that

$$\text{logit} g_T(\text{tx}, w) = w^T \psi_{\text{tx}}. \tag{1}$$

That is, the probability of being alive on the treatments tx $= 0, 1, 2$ is a function of baseline covariates with coefficients depending on treatment tx. Also, let

$$p_T^*(1, w) = P(D_T(1) = 0 | D_T(0) = 0, w), \tag{2}$$

$$p_T^*(2, w) = P(D_T(2) = 0 | D_T(0) = 0, w), \tag{3}$$

$$q_T^*(w) = P(D_T(2) = 1 | D_T(1) = 1, D_T(0) = 1); \tag{4}$$

thus, $p_T^*(1, w)$ is the probability of a patient with baseline covariates w being alive on IROX at time T, given that they would be alive on IFL at time T. Similarly, $p_T^*(2, w)$ represents the probability of being alive on FOLFOX at time T, given that they would be alive on IFL at time T for a patient with baseline covariates, w. $q_T^*(w)$ is the probability of a patient with baseline covariates w being dead on FOLFOX at time T given that they would be dead on the other two arms at time T.

We first make the common Stable Unit Treatment Value Assumption (SUTVA, Rubin, 1978) which states that a patient's potential outcomes are unrelated to the treatment status of other patients. Next, we list the assumptions necessary to identify the causal effect of interest:

Assumption 1 (Monotonicity) :

a) Deterministic:

$$P(D_T(1) = 1 | D_T(0) = 0, D_T(2) = 1) = 0, \tag{5}$$

This assumption implies that there will be no subjects who die on both experimental treatments and survive on the standard of care.

b) Stochastic I: We assume $p_T^*(j,w)$ for $j = 1, 2$ are given by

$$p_T^*(j,w) = g_T(j,w) + \rho\left\{U_j(w) - g_T(j,w)\right\}, \quad 0 \leq \rho \leq 1, \tag{6}$$

where $U_j(w) = \min\left\{1, \frac{g_T(j,w)}{g_T(0,w)}\right\}$. To understand $U_j(w)$, note that given $g_T(j,w)$, we have the following bounds on the conditional distribution,

$$g_T(j,w) \leq p_T^*(w) \leq \min\left\{1, \frac{g_T(j,w)}{g_T(0,w)}\right\}.$$

This modeling assumption was used in Roy et al. (2008) in the context of compliance in a smoking cessation trial. Note that if $\rho = 0$, then $p_T^*(j,w) = g_T(j,w)$, which implies that $D_T(1)$ is independent of $D_T(0)$ and $D_T(2)$ is also independent of $D_T(0)$. If $\rho = 1$, then $p_T^*(j,w) = U_j(w)$, which is the largest possible probability that is compatible with the marginal distributions.

c) Stochastic II:

$$q_T^*(w) = (1 - g_T(2,w)) + \nu\left\{U_3(w) - (1 - g_T(2,w))\right\}, \quad 0 \leq \nu \leq 1, \tag{7}$$

where $U_3(w) = \min\left\{1, \frac{1-g_T(2,w)}{P(D_T(1)=1, D_T(0)=1)}\right\}$. Similarly, we have

$$(1 - g_T(2,w) \leq q_T^*(w) \leq \min\left\{1, \frac{(1 - g_T(2,w))}{P(D_T(1) = 1, D_T(0) = 1)}\right\}.$$

This assumption is similar to b) above, but addresses all three treatments. It says that dying on FOLFOX is positively correlated with dying on both IROX and IFL; that is, patients more likely to die on IROX and IFL (standard of care) would be more likely to die on FOLFOX.

Both stochastic monotonicity assumptions address the concern of a deterministic monotonicity assumption not being reasonable between two treatment arms and implicitly

allows the assumption to vary with subject specific covariates, w. In addition, they both imply a positive covariance (correlation) between the event on the left and right of the conditioning. The sensitivity parameters in Assumption I are ρ and ν.

Assumption 2 (Ignorability): $\text{Tx} \perp \bar{\mathcal{P}}_T$.

Recall $\bar{\mathcal{P}}_t$ is all the potential outcome up to and including time t. Assumption 2 states that the treatment arm is unrelated to the set of potential outcomes; this holds by randomization of the treatments.

Assumption 3 (Sensitivity Parameter Reduction via Proportional odds):

$$\frac{odds(Y_T(0) > k | A_T(j))}{odds(Y_T(0) > k | A_T(0))} \stackrel{\text{let}}{=} \tau, \quad \text{if } j = 4, 5 \tag{8}$$

$$\frac{odds(Y_T(1) > k | A_T(j))}{odds(Y_T(1) > k | A_T(0))} = \tau, \quad \text{if } j = 1, 4 \tag{9}$$

$$\frac{odds(Y_T(2) > k | A_T(j))}{odds(Y_T(2) > k | A_T(0))} = \tau, \quad \text{if } j = 1, 5 \tag{10}$$

$$\frac{odds(Y_T(1) > k | A_T(3))}{odds(Y_T(1) > k | A_T(0))} = \lambda, \tag{11}$$

and

$$\frac{odds(Y_T(2) > k | A_T(2))}{odds(Y_T(2) > k | A_T(0))} = \lambda, \tag{12}$$

Note that $A_T(1)$, $A_T(4)$, or $A_T(5)$ in equations (8)-(10) are the principal strata in which subjects would be dead under one of three arms. In contrast, $A_T(2)$ or $A_T(3)$ in equations (11)-(12) are the principal strata in which subjects would be dead under two of three arms. The sensitivity parameters, τ and λ correspond to the odds ratios formed from conditioning on the subjects dying on either one or two treatment arms respectively (vs. surviving on all three arms). More specifically, the first sensitivity parameter, τ corresponds to the ratio of odds of fatigue, for a given treatment, for a patient in a principal stratum in which they would survive on only two arms to the odds for a patient in a principal stratum in which they would survive on all three arms. The

second sensitivity parameter, λ corresponds to the ratio of odds of fatigue, for a given treatment, for a patient in a principal stratum in which they would survive on only one arm to the odds for a patient in a principal stratum in which they would survive on all three arms.

Note also that the two sensitivity parameters, τ and λ, do not depend on k. As such, the assumption is referred to as proportional odds. In total, Assumption 3 has *two* sensitivity parameters.

Assumptions like Assumption 3, which involve a few sensitivity parameters, are fundamental to the ability to reasonably elicit sensitivity parameters to conduct inference (Egleston et al., 2009). The semiparametric models of Robins and colleagues (Robins et al., 1995; Rotnitzky et al., 1998; Scharfstein et al., 1999) for MNAR missingness also make simplifying assumptions to reduce the sensitivity parameters (in that context, in the missing data mechanism). Egleston et al. (2007) and Lee and Daniels (2008) use similar assumptions to Assumption 3 in much simpler contexts. Other variations on Assumption 3 are possible and would be specific to the application. The assumptions in this case were carefully formulated with the expertise and guidance of Dr. Sargent, who was the statistician on the N9741 clinical trial.

3.3 *Identification and Estimation of the SACE*

Our goal in the following will be to use the observed data, $\mathcal{O}_t$, along with the three assumptions in Section 3.2 to draw inference about the *SACE*. For clarity of presentation, we demonstrate the identification of the SACE in two parts. First, we identify the joint distribution of $\mathbf{D_T} = (D_T(0), D_T(1), D_T(2))$; and second, we identify the conditional distribution of Y_T given $\mathbf{D_T}$.

3.3.1 Identification and Calculation of Joint Distribution of D_T We identify the marginal and joint distributions of $\mathbf{D_T}$ using Theorem I.

Theorem. I. (Identification of the joint distribution of $\mathbf{D_T}$)
Under Assumptions 1 and 2, $P(\mathbf{D_T})$ is identified.

A detailed proof of this result can be found in the Appendix.

Using Theorem I, the marginal and joint probabilities of $\mathbf{D_T}$ can be written as functions of (1)-(4) as follows:

- The marginal distributions, $P(D_T(\text{tx}) = 0)$ for $\text{tx} = 0, 1, 2$ can be expressed as

$$P(D_T(\text{tx}) = 0) = P(D_T(\text{tx}) = 0) = E_w \left\{ g_T(\text{tx}, w) \right\}, \tag{13}$$

- The joint distribution, $P(\mathbf{D_T})$, which corresponds to $P(A_T(j))$ for $j = 0, \cdots, 6$ can be expressed as

$$\begin{aligned}
P(A_T(0)) &= E_w \left\{ p_T^*(2, w) g_T(0, w) \right\} - P(A_T(5)), \\
P(A_T(1)) &= E_w \left\{ g_T(2, w) - g_T(0, w) p_T^*(2, w) \right\} - P(A_T(2)), \\
P(A_T(2)) &= E_w \left[(1 - q_T^*(w)) \left\{ 1 - g_T(0, w) \left(1 - p_T^*(1, w) \right) - g_T(1, w) \right\} \right], \\
P(A_T(3)) &= E_w \left\{ g_T(1, w) - p_T^*(1, w) g_T(0, w) \right\} - P(A_T(1)), \\
P(A_T(4)) &= E_w \left\{ p_T^*(1, w) g_T(0, w) \right\} - P(A_T(0)), \\
P(A_T(5)) &= E_w \left[g_T(0, w) \left\{ 1 - p_T^*(1, w) \right\} \right], \\
P(A_T(6)) &= 0.
\end{aligned}$$

These expressions are used for the identification and estimation of the causal effects in the subsequent sections.

3.3.2 Identification of SACE We now consider identification of the conditional distributions of Y_T given $\mathbf{D_T}$. Note that Assumption 2 implies

$$P(Y_T(\text{tx}) > k | D_T(\text{tx}) = 0) \quad = \quad P(Y_T > k | D_T = 0, \text{Tx} = \text{tx}) \stackrel{\text{let}}{=} h_{T,\text{tx}}(k),$$

for $\text{tx} = 0, 1, 2$. We demonstrate the identification of the $SACE$s using this result, Theorem I, and Assumption 3.

Theorem. II.

Under assumptions 1, 2, and 3, $SACE_{k,\cdot}(1,0)$ is identified.

A simple proof of this result starts by showing that the numerators and denominators of $SACE_{k,\cdot}(1,0)$ can all be written as functions of $P(Y_T(0) > k|A_T(0))$ and $P(Y_T(1) > k|A_T(0))$ using Assumption 3. We show this next.

From Assumption 3, we have

$$P(Y_T(0) > k|A_T(4)) = \frac{\tau P(Y_T(0) > k|A_T(0))}{1 + (\tau - 1)\, P(Y_T(0) > k|A_T(0))}, \tag{14}$$

$$P(Y_T(1) > k|A_T(4)) = \frac{\tau P(Y_T(1) > k|A_T(0))}{1 + (\tau - 1)\, P(Y_T(1) > k|A_T(0))}. \tag{15}$$

By identifying $P(Y_T(0) > k|A_T(0))$, we identify (14) and by identifying $P(Y_T(1) > k|A_T(0))$, we identify (15).

For $SACE_{k,3}(1,0)$, we need

$$P(Y_T(0) > k|A_T(0) \cup A_T(4)) = \frac{P(Y_T(0) > k|A_T(0))P(A_T(0)) + P(Y_T(0) > k|A_T(4))P(A_T(4))}{P(A_T(0) \cup A_T(4))}, \tag{16}$$

$$P(Y_T(1) > k|A_T(0) \cup A_T(4)) = \frac{P(Y_T(1) > k|A_T(0))P(A_T(0)) + P(Y_T(1) > k|A_T(4))P(A_T(4))}{P(A_T(0) \cup A_T(4))}, \tag{17}$$

which are functions of $P(A_T(0) \cup A_T(4))$, $P(A_T(0))$, and $P(A_T(4))$ (all identified by Theorem I). We showed above that $P(Y_T(0) > k|A_T(4))$ and $P(Y_T(1) > k|A_T(4))$ can be written as functions of $P(Y_T(0) > k|A_T(0))$ and $P(Y_T(1) > k|A_T(0))$. The details on

the identification of $P(Y_T(0) > k|A_T(0))$ and $P(Y_T(1) > k|A_T(0))$ can be found in the appendix.

Similar arguments can be used to identify $SACE_{k,\cdot}(2,0)$ and $SACE_{k,\cdot}(2,1)$. We state this formally in the following two corollaries.

Corollary. I

Under assumptions 1, 2, and 3, $SACE_{k,\cdot}(2,0)$ is identified.

Given the identification results in Theorems I and II and following a similar argument to Theorem I, all the relevant numerators and denominators can be written as a function of the quantities necessary to compute $SACE_{k,\cdot}(1,0)$ except for $P(Y_T(2) > k|A_T(0))$. The appendix contains a proof of the identification of this probability.

Corollary. II

Under assumptions 1, 2, and 3, $SACE_{k,\cdot}(2,1)$ is identified.

The entire proof is in the appendix.

3.4 *Estimation and modeling of $g_T(tx,w)$, $p^*_T(j,w)$, $q^*_T(w)$, and $h_{T,tx}(k)$*

As shown in the previous two subsections, the $SACE$ is a function of $g_T(\text{tx},w)$, $p^*_T(j,w)$, $q^*_T(w)$, and $h_{T,\text{tx}}(k)$. The parameters for $g_T(\text{tx},w)$, ψ_{tx}, can be estimated using maximum likelihood based on $N_T(\text{tx})$ subjects, where $N_T(\text{tx})$ is the number of subjects who were alive after the study was complete under Tx = tx (see Table 1). We estimate $p^*_T(j,w)$ in Assumption 1 using $\hat{g}_T(1,w)$, $\hat{g}_T(2,w)$, and $\hat{g}_T(0,w)$ with fixed ρ. We estimate $q^*_T(w)$ using $\hat{g}_T(\cdot,w)$ with fixed ν.

Note that we can use the empirical expectation to estimate $E_w\{f(w;\text{tx})\}$ for an arbitrary function $f(\cdot)$ which is given by

$$\hat{E}_w\{f(w;\text{tx})\} = \frac{1}{N}\sum_{i=1}^{N}\hat{f}(w_i;\text{tx}).$$

For example, $E_w\{g_T(\text{tx}, w)\}$ and $E_w\{p_T^*(j, w)g_T(\text{tx}, w)\}$ for $j = 1, 2$; tx $= 0, 1, 2$ can be estimated by

$$\hat{E}_w\{g_T(\text{tx}, w)\} = \frac{1}{N}\sum_{i=1}^{N}\hat{g}_T(\text{tx}, w_i),$$

$$\hat{E}_w\{p_T^*(j, w)g_T(\text{tx}, w)\} = \frac{1}{N}\sum_{i=1}^{N}\hat{p}_T^*(j, w_i)\hat{g}_T(\text{tx}, w_i).$$

To estimate $h_{T,\text{tx}}(k)$, we need to specify models for the observed data and missing data due to loss to follow-up (not progression/death). Before stating the models, we specify one additional assumption that allows us to estimate the above quantities from the observed data. Recall that S is the progression/death time. So, $S = j \Leftrightarrow \{D_j = 1, D_{j-1} = 0\}$.

Assumption 4 (Conditional MAR): $R_t \perp \{Y_t, \ldots, Y_{j-1}\}|S = j, \bar{Y}_{t-1}, \text{Tx}$ for $t < j$.

This assumption states that missingness of outcome is independent of the value of the outcome given the previous history of outcomes and Tx for all subjects with progress/death time j. It is MAR conditional on progression/death time. Missingness assumptions like this have recently been termed *partially ignorable* (Harel and Schafer, 2009). This assumption will allow us to use all the available data on survivors for the analysis.

We specify IOMTM(2) models (Lee and Daniels, 2007) conditional on progression/death time as follows, S,

$$\log\frac{P(Y_t > k|S = j, \text{tx})}{1 - P(Y_t > k|S = j, \text{tx})} = \beta_{0k}(j) + \beta(j, \text{tx}), \tag{18}$$

$$\log\frac{P(Y_t = k|Y_{i,t-1}, Y_{it-2}, S = j, \text{tx})}{P(Y_t = K|Y_{i,t-1}, Y_{i,t-2}, S = j, \text{tx})} = \triangle_{itk}(j, \text{tx}) + \gamma_{11}^{(k)}(j, \text{tx})Y_{i,t-1} + \gamma_{12}^{(k)}(j, \text{tx})Y_{i,t-1}^2 + \gamma_{21}^{(k)}(j, \text{tx})Y_{i,t-2} + \gamma_{22}^{(k)}(j, \text{tx})Y_{i,t-2}^2, \tag{19}$$

where $\beta(j, \text{tx})$ are the coefficients of treatment arm tx, $\beta_{01}(j) > \cdots > \beta_{0K-1}(j)$, $\gamma_{1l}^{(k)}(j, \text{tx})$ and $\gamma_{2l}^{(k)}(j, \text{tx})$ for $l = 1, 2$ are the linear and quadratic coefficients of of the lag one and two responses, respectively. Model (18) is a cumulative logit model, a common model for ordinal data. However, the link function in the conditional probabilities (19) is a

multinomial logit because we expect the $K-1$ conditional probabilities to have different coefficients $\gamma_{1l}^{(k)}(\cdot,\cdot)$ depending on k.

Clearly, for models (18) and (19) to form a coherent probability model, the $\triangle_{itk}(j,\text{tx})$ will be constrained. $\triangle_{itk}(j,\text{tx})$ is found using the following identity

$$\begin{aligned} P(Y_{it}=k|S=j,\text{tx}) &= \sum_{g_1=1}^{K}\sum_{g_2=1}^{K} P(Y_{it}=k|Y_{it-1}=g_1,Y_{it-2}=g_2,S=j,\text{tx})\times \\ &\qquad P(Y_{it-1}=g_1,Y_{it-2}=g_2|S=j,\text{tx}). \end{aligned}$$

We refer the reader to Lee and Daniels (2007) for more details on these models. Within this modeling framework, $h_{T,\text{tx}}(k)$ is given by

$$\begin{aligned} h_{T,\text{tx}}(k) &= P(Y_T>k|S>T-1,\text{tx}) \\ &= \frac{\sum_{j>T-1}P(Y_T>k|S=j,\text{tx})P(S=j|\text{tx})}{\sum_{j>T-1}P(S=j|\text{tx})}, \end{aligned} \tag{20}$$

where $P(S=j|\text{tx})$ is a multinomial mass function for S, and $P(Y_T>k \mid S=j,\text{tx})$ is calculated from (18) given tx and is only defined for $t<j$. Via assumption 4, we assume that for a given progression/death time (pattern), that missing data before the progression/death time is MAR (conditional on pattern); this situation occurs when dropout is not due to progression/death. These models were chosen since they were shown to fit the observed data well for non-causal inference in Lee and Daniels (2007).

3.5 *Standard Errors for SACE*

To calculate standard errors for *SACE*, we use the delta method. Let

$$\theta^T=(\psi_0^T,\psi_1^T,\psi_2^T,\beta_0^T(\cdot),\beta^T(\cdot)),$$

where $\beta_0^T(\cdot)=(\beta_0^T(0),\beta_0^T(1),\beta_0^T(2))^T$ and $\beta^T(\cdot)=(\beta^T(0),\beta^T(1),\beta^T(2))$. The variance of the $SACE_{K-1}(\text{Tx},\text{Tx'})$ for Tx $\neq$ Tx' is given by

$$var(SACE_{K-1}(\text{Tx},\text{Tx'}))=\Delta SACE_{K-1}^T(\text{Tx},\text{Tx'})var(\hat{\theta})\Delta SACE_{K-1}(\text{Tx},\text{Tx'}),$$

where $\Delta SACE_{K-1}(\text{Tx, Tx'})$ is the gradient of $SACE_{K-1}(\text{Tx, Tx'})$ with respect to θ. Denoting $\theta_1^T = (\psi_0^T, \psi_1^T, \psi_2^T)$ and $\theta_2^T = (\beta_0^T(\cdot), \beta^T(\cdot))$, we have

$$var(\hat{\theta}) = \begin{pmatrix} var(\hat{\theta}_1(\cdot)) & 0 \\ 0 & var(\hat{\theta}_2(\cdot)) \end{pmatrix}.$$

Details on computing the gradient and variance of $\hat{\theta}$ are given in the appendix.

3.6 *Sensitivity Analysis*

The identification of the *SACE* relies on untestable assumptions (in particular, Assumptions 1 and 3). We will use a sensitivity analysis procedure to draw inference about *SACE* by varying the sensitivity parameters (τ, λ) in Assumption 3, and $(\rho,\ \nu)$ in Assumption 1. We discuss reasonable ranges for the former two sensitivity parameters in Section 4.1.

4. Analysis of the QOL Data

We use the approach proposed in Section 3 to analyze the QOL data from Section 2. We start out with a discussion of the elicitation of reasonable ranges for the sensitivity parameters.

4.1 *Elicitation of Sensitivity Parameters*

Dr. Sargent, the primary statistician for this clinical trial, provided the expertise on the sensitivity parameters for the analysis (including the expertise needed for the specification of Assumption 3). Dr. Sargent is both a PhD level statistician and a professor of oncology. Having both the quantitative and clinical knowledge is an important asset in eliciting values for sensitivity parameters. The feasibility of the eliciting of sensitivity parameters in general has been discussed in Shepherd et al. (2007). The following decisions regarding reasonable values for the sensitivity parameters were elicited through extensive discussion:

- τ and λ are unlikely to be outside the range of $(1/2, 1.5)$. That is, $\tau,\ \lambda \in [1/2, 1.5]$

The key to these ranges is that it was thought very unlikely for more than a 50% increase or decrease in these odds ratios.

We conduct a sensitivity analysis and draw inferences in the next section.

4.2 *Causal inference results*

We are interested in subjects who would be alive on at least two treatment arms until the end of the study. Table 1 gives the number of patients alive at the end of the study by treatment.

4.2.1 Estimation of survival and fatigue probabilities Baseline covariates related to survival, w, included in the analyses were age (AGE), patient performance status (PS), and number of sites of metastatic disease (NOSITE). These three factors are well established to impact patient survival in this setting. To simplify computations of the empirical expectations in Section 3.4, we categorized each of the covariates. The cutpoints were chosen by Dr. Sargent using his extensive experience with the trial and the disease. Age was discretized as < 70 ($AGE = 0$) and ≥ 70 ($AGE = 1$). Performance status (PS) is a measurement of the patient's ability to carry out activities of daily living, measured on a 5 point ordinal scale, where 0 represents full health and 4 represents death. Thus a higher PS indicates *worse* health. We discretized performance status into high $\{0$ or $1\}$ ($PS = 0$) vs. low $\{2, 3, 4\}$ ($PS = 1$). The number of sites of metastatic disease was broken into three categories, $\{1, 2, > 2\}$ and coded using two indicator variables, NOSITE1 and NOSITE2, respectively as $\{(0, 0), (1, 0), (0, 1)\}$.

We fit the logistic regression models in (1) using these covariates. The regression coefficient estimates, $\hat{\psi}$, are given in Table 2. For FOLFOX, the estimated coefficients of performance status and number of metastatic sites were significant; for IFL, performance status was significant and for IROX, number of metastatic sites was significant. For all

three arms, the coefficients indicated that the estimated survival probability decreased for patients with lower performance status and decreased for patients with more sites of metastases.

[Table 1 about here.]

[Table 2 about here.]

The MLE's of $h_{T,\cdot}(K-1)$ based on the IOMTM(2) (specified in (18) and (19)), are given in Table 3. The probability of fatigue for survivors under IROX was highest among three arms. The probability under IFL was the lowest.

[Table 3 about here.]

4.2.2 Estimation of SACE In the following, we present inferential results and sensitivity analyses for $\log SACE_{K-1,\cdot}(1,0)$, $\log SACE_{K-1,\cdot}(2,0)$, and $\log SACE_{K-1,\cdot}(2,1)$. Note that under Assumption 3, we have following equalities

$$\log SACE_{K-1,1}(1,0) \equiv \log SACE_{K-1,2}(1,0),$$
$$\log SACE_{K-1,1}(2,0) \equiv \log SACE_{K-1,2}(2,0),$$
$$\log SACE_{K-1,1}(2,1) \equiv \log SACE_{K-1,2}(2,1).$$

As such, we only focus $SACE_{K-1,1}(\cdot,\cdot)$ and $SACE_{K-1,3}(\cdot,\cdot)$ in the following. For each given set of sensitivity parameters, $\log SACE_{k,1}(1,0)$ and $\log SACE_{k,3}(1,0)$ are the log odds ratios that the fatigue score at time T is larger than k for comparing IROX and IFL (standard of care) for subjects who would be alive under all three treatment arms and who would be alive either under all three treatment arms or only under IFL (standard of care) and IROX. $\log SACE_{k,1}(2,0)$ and $\log SACE_{k,3}(2,0)$ are the log odds ratios similarly

defined for comparing IFL and FOLFOX; and $\log SACE_{k,1}(2,1)$ and $\log SACE_{k,3}(2,1)$, for comparing the two experimental treatments, IROX and FOLFOX.

The estimated values of $\log SACE_{K-1,1}(1,0)$ and associated p-values testing whether they are different from zero over the range of τ and λ elicited in Section 4.1 are given in Figure 1 for various values of ρ and ν. When $(\rho,\nu)=(0,0)$, which corresponds to the probability of dying on IROX being independent of the probability of dying on IFL, the probability of dying on FOLFOX being independent of the probability of dying on IFL, and the probability of dying on FOLFOX being independent of the probability of dying on both IROX and IFL, respectively, the estimated values of $\log SACE_{K-1,1}(1,0)$ increased as τ and λ respectively increased and decreased. The estimated values of $\log SACE_{K-1,1}(1,0)$ were greater than 0 for all τ and λ (Figure 1(a)), implying that the probability of fatigue on the IROX treatment was greater than the probability of a patient's fatigue on the IFL treatment. The estimated values of $\log SACE_{K-1,1}(1,0)$ increased as λ decreased for $(\rho,\nu)=(0.5,0.5)$ which corresponds to a moderate correlation between death probabilities on the two treatments and three treatments, respectively (Figure 1(c)). In contrast, when $(\rho,\nu)=(1,1)$ (highest correlation), the $\log SACE_{K-1,1}(1,0)$ decreased as τ increased (Figure 1(e)). However, for all values of (ρ,ν) there were no significant differences between IROX and IFL. (Figure 1(b), 1(d), 1(f)).

[Figure 1 about here.]

Figure 2 presents estimated values of $\log SACE_{K-1,3}(1,0)$ and corresponding p-values. The pattern was similar for $\log SACE_{K-1,1}(1,0)$ (Figure 2(a), 2(c), 2(e)) and there were no significant differences between the two treatment arms for all values of (ρ,ν) (Figure 2(b), 2(d), 2(f)). We note that the plots for estimated values of $\log SACE_{K-1,1}(1,0)$ and $\log SACE_{K-1,3}(1,0)$ looked similar. However, there were small difference between these estimated values.

[Figure 2 about here.]

Figures 3 and 4 show similar results for $\log SACE_{K-1,1}(2,0)$ and $\log SACE_{K-1,3}(2,0)$. The estimated values of $\log SACE_{K-1,1}(2,0)$ and $\log SACE_{K-1,3}(2,0)$ were greater than 0 for all values of (ρ, ν). However, there were no significant differences between FOLFOX and IFL.

[Figure 3 about here.]

[Figure 4 about here.]

Figure 5 and 6 give the estimated values of $\log SACE_{K-1,1}(2,1)$ and $\log SACE_{K-1,3}(2,1)$ and the corresponding p-values. The estimated values of $SACE_{K-1,1}(2,1)$ were less than 0 for all values of (ρ, ν), implying that the probability of fatigue on IROX was larger than the probability of fatigue on FOLFOX. However, again the differences were not significant.

[Figure 5 about here.]

[Figure 6 about here.]

5. Summary

In this paper, we have proposed a principal stratification approach to estimate the causal effect of treatment in the presence of death for a three arm study with two active treatments and a longitudinal ordinal outcome. Our formulation of sensitivity parameters in Assumption 3 is very much context specific. Based on this approach, we concluded that patients' fatigue levels were not influenced by treatment, an important finding that allows the use of the treatment with the most favorable anti-cancer activities. However, we note that the power was low for this comparison with few patients still on study at one year.

In our approach, instead of just using the data Y_T at the end of the study, we used all available longitudinal data using the models in Section 3.4 along with Assumption 4. This was essential to the viability of our approach as only 18 subjects in total had the QOL measure at time T, but more than 45% of the subjects were alive at the end of the study. Assumption 4 (Conditional MAR) could be weakened by including covariates in the model specified in (18)-(19).

A key assumption in our proposed approach is the deterministic monotonicity assumption Ia, which assumes that there are no patients who would die on both experimental arms but survive on the control arm. While this is weaker than a more typical assumption that no patient would die on either experimental arm but be alive on the control arm, it still requires justification. In this trial, both experimental arms were superior for overall survival compared to the control, and were also less toxic. In addition, despite extensive analyses of both baseline clinical and blood-based predictors (pharmacogenomics) reported elsewhere (Sanoff et al., 2008; McLeod et al., 2003), no predictors of differential treatment efficacy between the three treatments has been identified. Thus, while personalized medicine has the hope to identify the optimal therapy for each patient (which for some patients may be the control), at the present time for this trial, assumption Ia is not unreasonable.

A difficult research problem is the best way to 'honestly' elicit sensitivity parameters in causal inference problems, often from physicians. Here we had the benefit of a statistician, Dr. Sargent, who is also a professor of oncology and was intimately involved in this trial. However eliciting from experts *without* this dual expertise remains a challenge.

In the current approach, we compute the large sample variance of SACE using the delta method. We plan in future work to examine its operating characteristic in small samples. We also plan to explore a fully Bayesian approach and as such, use informative priors

for the sensitivity parameters, $(\tau, \lambda, \rho, \nu)$ to obtain a single inference that characterizes our uncertainty about the sensitivity parameters. This will also allow us to easily make non-large sample inferences. This is ongoing work.

Acknowledgments

We would like to thank Ms. Erin Green of the Mayo Clinic for her help in data collection and clarifying issues with the data. The authors gratefully acknowledge the associate editor and two referees for insightful comments that have greatly improved the manuscript. This project was supported by NIH grants CA85295, HL079457, and CA25224.

Appendix

Proof of Theorem I

We know that the marginal distributions, $P(D_T(j) = 0)$ for $j = 0, 1, 2$ are identified via Assumption 2 and (1),

$$P(D_T(\text{tx}) = 0) = P(D_T = 0|\text{Tx} = \text{tx}) = E_w \left\{P(D_T = 0|\text{Tx} = \text{tx}, w)\right\} = E_w \left\{g_T(\text{tx}, w)\right\}. \tag{21}$$

Now we move to identification of the bivariate distributions $P(D_T(j) = l, D_T(j^{'}) = l^{'})$, $l, l^{'} \in \{0, 1\}$, $j, j^{'} \in \{0, 1, 2\}$. From Assumption 1 and (2), we have the following results

$$\begin{aligned}
P(D_T(1) = 0, D_T(0) = 0) &= P(D_T(1) = 0|D_T(0) = 0)P(D_T(0) = 0) \\
&= E_w \left\{p_T^*(1, w)g_T(0, w)\right\}, \\
P(D_T(1) = 1, D_T(0) = 0) &= P(D_T(1) = 1|D_T(0) = 0)P(D_T(0) = 0) \\
&= E_w \left\{(1 - p_T^*(1, w))\, g_T(0, w)\right\}.
\end{aligned}$$

Therefore, both $P(D_T(1)=0, D_T(0)=0)$ and $P(D_T(1)=1, D_T(0)=0)$ are identified. Since

$$\begin{aligned} & P(D_T(1)=0) \\ = \; & P(D_T(1)=0|D_T(0)=0)P(D_T(0)=0) + P(D_T(1)=0|D_T(0)=1)P(D_T(0)=1) \\ = \; & p_T^*(1,w)g_T(0,w) + P(D_T(1)=0|D_T(0)=1)\left(1-g_T(0,w)\right), \end{aligned}$$

$P(D_T(1)=0|D_T(0)=1)$ is identified. From the following result

$$\sum_{j,j'} P(D_T(1)=j, D_T(0)=j') = 1,$$

$P(D_T(1)=1, D_T(0)=1)$ is identified.

We now show that

$$P(D_T(2)=j, D_T(0)=j') \text{ for } j,j'=0,1,$$

are identified. From Assumption 1 and (3), we have the following results

$$\begin{aligned} P(D_T(2)=0, D_T(0)=0) &= P(D_T(2)=0|D_T(0)=0)P(D_T(0)=0) \\ &= E_w\left\{p_T^*(2,w)g_T(0,w)\right\}, \\ P(D_T(2)=1, D_T(0)=0) &= \left\{1-P(D_T(2)=0|D_T(0)=0)\right\}P(D_T(0)=0) \\ &= E_w\left\{\left(1-p_T^*(2,w)\right)g_T(0,w)\right\}. \end{aligned}$$

Therefore, both $P(D_T(2)=0, D_T(0)=0)$ and $P(D_T(2)=1, D_T(0)=0)$ are identified. Since

$$\begin{aligned} & P(D_T(2)=0) \\ = \; & P(D_T(2)=0|D_T(0)=0)P(D_T(0)=0) + P(D_T(2)=0|D_T(0)=1)P(D_T(0)=1) \\ = \; & p_T^*(2,w)g_T(0,w) + P(D_T(2)=0|D_T(0)=1)\left(1-g_T(0,w)\right), \end{aligned}$$

$P(D_T(2)=0|D_T(0)=1)$ is identified. Finally, $P(D_T(2)=1|D_T(0)=1)$ is identified since

$$\sum_{j,j'} P(D_T(2)=j, D_T(0)=j') = 1.$$

Now we will prove $P(A_T(j))$ for $j = 0, 1, \cdots, 6$ are identified. More specifically, we just need to show that

$$P(D_T(0) = j, D_T(1) = l, D_T(2) = k),$$

for $j, l, k = 0, 1$ are identified (except $(j, l, k) = (1, 1, 1)$). We first identify $P(A_T(5)) = P(D_T(0) = 0, D_T(1) = 1, D_T(2) = 0)$,

$$\begin{aligned}
& P(D_T(0) = 0, D_T(1) = 1, D_T(2) = 0) \\
= \; & P(D_T(0) = 0) - P(D_T(0) = 0, D_T(1) = 0) \\
= \; & P(D_T(0) = 0) - P(D_T(1) = 0 | D_T(0) = 0) P(D_T(0) = 0) \\
= \; & E_w \left[g_T(0, w) \left\{ 1 - p_T^*(1, w) \right\} \right]. \text{ Assumption Ib.}
\end{aligned}$$

We now identify $P(A_T(0)) = P(D_T(0) = 0, D_T(1) = 0, D_T(2) = 0)$.

$$\begin{aligned}
& P(D_T(0) = 0, D_T(1) = 0, D_T(2) = 0) \\
= \; & P(D_T(0) = 0, D_T(2) = 0) - P(D_T(0) = 0, D_T(1) = 1, D_T(2) = 0) \\
= \; & P(D_T(2) = 0 | D_T(0) = 0) P(D_T(0) = 0) - P(A_T(5)) \\
= \; & E_w \left\{ p_T^*(2, w) g_T(0, w) \right\} - P(A_T(5)).
\end{aligned}$$

To identify $P(A_T(2)) = P(D_T(0) = 1, D_T(1) = 1, D_T(2) = 0)$, we have

$$\begin{aligned}
& P(D_T(0) = 1, D_T(1) = 1, D_T(2) = 0) \\
= \; & P(D_T(2) = 0 | D_T(0) = 1, D_T(1) = 1) P(D_T(0) = 1, D_T(1) = 1) \\
= \; & P(D_T(2) = 0 | D_T(0) = 1, D_T(1) = 1) \left\{ 1 - P(D_T(0) = 0, D_T(1) = 0) \right. \\
& \quad \left. - P(D_T(0) = 0, D_T(1) = 1) - P(D_T(0) = 1, D_T(1) = 0) \right\} \\
= \; & P(D_T(2) = 0 | D_T(0) = 1, D_T(1) = 1) \left[1 - P(D_T(1) = 0 | D_T(0) = 0) P(D_T(0) = 0) \right. \\
& \quad - P(D_T(1) = 1 | D_T(0) = 0) P(D_T(0) = 0) \\
& \quad \left. - \left\{ P(D_T(1) = 0) - P(D_T(1) = 0 | D_T(0) = 0) P(D_T(0) = 0) \right\} \right] \\
= \; & E_w \left[(1 - q_T^*(w)) \left\{ 1 - g_T(0, w) - g_T(1, w) + p_T^*(1, w) g_T(0, w) \right\} \right]. \text{ Assumption 1c).}
\end{aligned}$$

To identify $P(A_T(1)) = P(D_T(0) = 1, D_T(1) = 0, D_T(2) = 0)$, we need the following equation

$$\begin{aligned}
& P(D_T(0) = 1, D_T(1) = 0, D_T(2) = 0) \\
= \; & P(D_T(2) = 0, D_T(0) = 1) - P(D_T(0) = 1, D_T(1) = 1, D_T(2) = 0) \\
= \; & \{1 - P(D_T(0) = 0 | D_T(2) = 0)\} P(D_T(2) = 0) - P(A_T(2)) \\
= \; & P(D_T(2) = 0) - P(D_T(0) = 0) P(D_T(2) = 0 | D_T(0) = 0) - P(A_T(2)) \\
= \; & E_w \{g_T(2, w) - g_T(0, w) p_T^*(2, w)\} - P(A_T(2)).
\end{aligned}$$

To identify $P(A_T(4)) = P(D_T(0) = 0, D_T(1) = 0, D_T(2) = 1)$, we use the following relations,

$$\begin{aligned}
& P(D_T(0) = 0, D_T(1) = 0, D_T(2) = 1) \\
= \; & P(D_T(0) = 0, D_T(1) = 0) - P(D_T(0) = 0, D_T(1) = 0, D_T(2) = 0) \\
= \; & P(D_T(1) = 0 | D_T(0) = 0) P(D_T(0) = 0) - P(A_T(0)) \\
= \; & E_w \{p_T^*(1, w) g_T(0, w)\} - P(A_T(0)).
\end{aligned}$$

To identify $P(A_T(3)) = P(D_T(0) = 1, D_T(1) = 0, D_T(2) = 1)$, we have following equalities

$$\begin{aligned}
& P(D_T(0) = 1, D_T(1) = 0, D_T(2) = 1) \\
= \; & P(D_T(0) = 1, D_T(1) = 0) - P(D_T(0) = 1, D_T(1) = 0, D_T(2) = 0) \\
= \; & \{1 - P(D_T(0) = 0 | D_T(1) = 0)\} P(D_T(1) = 0) - P(A_T(1)) \\
= \; & P(D_T(1) = 0) - P(D_T(1) = 0 | D_T(0) = 0) P(D_T(0) = 0) - P(A_T(1)) \\
= \; & E_w \{g_T(1, w) - p_T^*(1, w) g_T(0, w)\} - P(A_T(1)).
\end{aligned}$$

Finally, to identify $P(A_T(6)) = P(D_T(0) = 0, D_T(1) = 1, D_T(2) = 0)$, we use the following equality

$$\begin{aligned}
& P(D_T(0) = 0, D_T(1) = 1, D_T(2) = 1) \\
= \; & P(D_T(1) = 1 | D_T(0) = 0, D_T(2) = 1) P(D_T(0) = 0, D_T(2) = 1) = 0. \text{ Assumption 1a).}
\end{aligned}$$

Proof of Theorem II

In addition to (14) and (15), we have from Assumption 3

$$P(Y_T(0) > k|A_T(5)) = \frac{\tau P(Y_T(0) > k|A_T(0))}{1 + (\tau - 1)\, P(Y_T(0) > k|A_T(0))}, \tag{22}$$

$$P(Y_T(1) > k|A_T(1)) = \frac{\tau P(Y_T(1) > k|A_T(0))}{1 + (\tau - 1)\, P(Y_T(1) > k|A_T(0))}, \tag{23}$$

$$P(Y_T(1) > k|A_T(3)) = \frac{\lambda P(Y_T(1) > k|A_T(0))}{1 + (\lambda - 1)\, P(Y_T(1) > k|A_T(0))}. \tag{24}$$

To identify $P(Y_T(0) > k|A_T(0))$, we use the following equality,

$$\begin{aligned} P(Y_T(0) > k|D_T(1) = 0) &= P(Y_T(0) > k|A_T(0))P(A_T(0)|D_T(0) = 0) \\ &+ P(Y_T(0) > k|A_T(4))P(A_T(4)|D_T(0) = 0) + P(Y_T(0) > k|A_T(5))P(A_T(5)|D_T(0) = 0). \end{aligned}$$

Now we replace $P(Y_T(0) > k|A_T(4))$ and $P(Y_T(0) > k|A_T(5))$ with (14) and (22), respectively. Then we have the following equality,

$$A_0 P^3(Y_T(0) > k|A_T(0)) + B_0 P^2(Y_T(0) > k|A_T(0)) + C_0 P(Y_T(0) > k|A_T(0)) - h_{T,0}(k) = 0, \tag{25}$$

where

$$\begin{aligned} A_0 &= P(A_T(0)|D_T(0) = 0)\,(\tau - 1)^2\,, \\ B_0 &= P(A_T(0)|D_T(0) = 0)\,(2\tau - 2) - h_{T,0}(k)\,(\tau - 1)^2\,, \\ &\quad + \tau\,(\tau - 1)\,P(A_T(4)|D_T(0) = 0) + \tau\,(\tau - 1)\,P(A_T(5)|D_T(0) = 0) \\ C_0 &= P(A_T(0)|D_T(0) = 0) + \tau P(A_T(4)|D_T(0) = 0) + \tau P(A_T(5)|D_T(0) = 0) \\ &\quad - (2\tau - 2)\,h_{T,0}(k). \end{aligned}$$

The probabilities $P(A_T(j)|D_T(0) = 0)$ for $j = 0, 4, 5$ are identified by Theorem I so the coefficients (A_0, B_0, C_0) are identified. Here, we need a solution of the cubic equation (25) in $[0, 1]$. In our example, only one solution was in this range.

To identify $P(Y_T(1) > k|A_T(0))$, we use following equation

$$\begin{aligned} & P(Y_T(1) > k|D_T(1) = 0) \\ = \; & P(Y_T(1) > k|A_T(0))P(A_T(0)|D_T(1) = 0) + P(Y_T(1) > k|A_T(1))P(A_T(1)|D_T(1) = 0) \\ & + P(Y_T(1) > k|A_T(3))P(A_T(3)|D_T(1) = 0) + P(Y_T(1) > k|A_T(4))P(A_T(4)|D_T(1) = 0). \end{aligned}$$

We replace $P(Y_T(1) > k|A_T(1))$, $P(Y_T(1) > k|A_T(3))$, and $P(Y_T(1) > k|A_T(4))$ with (23), (24) and (15), respectively. Then we have the following equation,

$$A_1 P^4(Y_T(1) > k|A_T(0)) + B_1 P^3(Y_T(1) > k|A_T(0)) + C_1 P^2(Y_T(1) > k|A_T(0))$$
$$+D_1 P(Y_T(1) > k|A_T(0)) - h_{T,1}(k) = 0, \tag{26}$$

where

$$\begin{aligned}
&A_1 = P(A_T(0)|D_T(1) = 0)\,(\tau - 1)^2\,(\lambda - 1)\,,\\
&B_1 = P(A_T(0)|D_T(1) = 0)\,\{(\tau - 1)\,(\lambda - 1) + (\tau + \lambda - 2)\,(\tau - 1)\}\\
&+P(A_T(1)|D_T(1) = 0)\tau\,(\lambda - 1)\,(\tau - 1) + P(A_T(3)|D_T(1) = 0)\lambda\,(\tau - 1)^2\\
&+P(A_T(4)|D_T(1) = 0)\tau\,(\tau - 1)\,(\lambda - 1) - h_{T,1}(k)\,(\tau - 1)^2\,(\lambda - 1)\,,\\
&C_1 = P(A_T(0)|D_T(1) = 0)\,(2\tau + \lambda - 3)\\
&+P(A_T(1)|D_T(1) = 0)\tau\,(\tau + \lambda - 2) + P(A_T(3)|D_T(1) = 0)\lambda\,(2\tau - 2)\\
&+P(A_T(4)|D_T(1) = 0)\tau\,(\tau + \lambda - 2)\\
&-h_{T,1}(k)\,\{(\tau - 1)\,(\lambda - 1) + (\tau + \lambda - 2)\,(\tau - 1)\}\,,\\
&D_1 = P(A_T(0)|D_T(1) = 0) + P(A_T(1)|D_T(1) = 0)\tau + P(A_T(3)|D_T(1) = 0)\lambda\\
&+P(A_T(4)|D_T(1) = 0)\tau - h_{T,1}(k)\,(2\tau + \lambda - 3)\,.
\end{aligned}$$

The probabilities $P(A_T(0)|D_T(0) = 0)$, $P(A_T(4)|D_T(0) = 0)$, $P(A_T(1)|D_T(0) = 0)$, $P(A_T(5)|D_T(0) = 0)$, $P(A_T(0)|D_T(1) = 0)$, $P(A_T(1)|D_T(1) = 0)$, $P(A_T(3)|D_T(1) = 0)$, and $P(A_T(4)|D_T(1) = 0)$ are identified by Theorem I so the coefficients (A_1, B_1, C_1, D_1) are identified. Similarly to the solution to (25), we have only one solution in $[0, 1]$.

Identification of $SACE_{k,\cdot}(2, 0)$

Via the following equations,

$$P(Y_T(0) > k|A_T(0) \cup A_T(5)) = \frac{P(Y_T(0) > k|A_T(0))P(A_T(0)) + P(Y_T(0) > k|A_T(5))P(A_T(5))}{P(A_T(0) \cup A_T(5))}, \tag{27}$$
$$P(Y_T(2) > k|A_T(0) \cup A_T(5)) = \frac{P(Y_T(2) > k|A_T(0))P(A_T(0)) + P(Y_T(2) > k|A_T(5))P(A_T(5))}{P(A_T(0) \cup A_T(5))}, \tag{28}$$

and the fact that $P(A_T(0) \cup A_T(5))$, $P(A_T(0))$, and $P(A_T(5))$ are identified from Theorem 1, the identification of $P(Y_T(0) > k|A_T(0) \cup A_T(5))$ and $P(Y_T(2) > k|A_T(0) \cup A_T(5))$

only requires the additional identification of $P(Y_T(2) > k|A_T(0))$ for $SACE_{k,1}(2,0)$, and $P(Y_T(2) > k|A_T(5))$ for $SACE_{k,2}(2,0)$, respectively.

From Assumption 3, we have following relations

$$P(Y_T(2) > k|A_T(1)) = \frac{\tau P(Y_T(2) > k|A_T(0))}{1 + (\tau - 1)\, P(Y_T(2) > k|A_T(0))}, \tag{29}$$

$$P(Y_T(2) > k|A_T(2)) = \frac{\lambda P(Y_T(2) > k|A_T(0))}{1 + (\lambda - 1)\, P(Y_T(2) > k|A_T(0))}, \tag{30}$$

$$P(Y_T(2) > k|A_T(5)) = \frac{\tau P(Y_T(2) > k|A_T(0))}{1 + (\tau - 1)\, P(Y_T(2) > k|A_T(0))}. \tag{31}$$

Therefore, we can identify $P(Y_T(2) > k|A_T(1))$, $P(Y_T(2) > k|A_T(2))$, and $P(Y_T(2) > k|A_T(5))$ by identifying $P(Y_T(2) > k|A_T(0))$. The identification of this final quantity is given in Corollary I, whose proof follows.

Proof of Corollary I

To identify $P(Y_T(2) > k|A_T(0))$, we have following equation

$$\begin{aligned} & P(Y_T(2) > k|D_T(2) = 0) \\ = \; & P(Y_T(2) > k|A_T(0))P(A_T(0)|D_T(2) = 0) + P(Y_T(2) > k|A_T(1))P(A_T(1)|D_T(2) = 0) \\ & + P(Y_T(2) > k|A_T(2))P(A_T(2)|D_T(2) = 0) + P(Y_T(2) > k|A_T(5))P(A_T(5)|D_T(2) = 0). \end{aligned}$$

Now we replace $P(Y_T(2) > k|A_T(1))$ and $P(Y_T(2) > k|A_T(5))$ with (10) and $P(Y_T(2) > k|A_T(2))$ with (12), respectively. Then we have the following equations,

$$\begin{aligned} & A_2 P^4(Y_T(2) > k|A_T(0)) + B_2 P^3(Y_T(2) > k|A_T(0)) + C_2 P^2(Y_T(2) > k|A_T(0)) \\ & + D_2 P(Y_T(2) > k|A_T(0)) - h_{T,2}(k) = 0, \end{aligned} \tag{32}$$

where

$$
\begin{aligned}
&A_2 = P(A_T(0)|D_T(2)=0)\,(\tau-1)^2\,(\lambda-1)\,,\\
&B_2 = P(A_T(0)|D_T(2)=0)\,\{(\tau-1)\,(\lambda-1)+(\tau+\lambda-2)\,(\tau-1)\}\\
&+P(A_T(1)|D_T(2)=0)\tau\,(\lambda-1)\,(\tau-1)+P(A_T(2)|D_T(2)=0)\lambda\,(\tau-1)^2\\
&+P(A_T(5)|D_T(2)=0)\tau\,(\tau-1)\,(\lambda-1)-h_{T,2}(k)\,(\tau-1)^2\,(\lambda-1)\,,\\
&C_2 = P(A_T(0)|D_T(2)=0)\,(2\tau+\lambda-3)\\
&+P(A_T(1)|D_T(2)=0)\tau\,(\tau+\lambda-2)+P(A_T(2)|D_T(2)=0)\lambda\,(2\tau-2)\\
&+P(A_T(5)|D_T(2)=0)\tau\,(\tau+\lambda-2)\\
&-h_{T,2}(k)\,\{(\tau-1)\,(\lambda-1)+(\tau+\lambda-2)\,(\tau-1)\}\,,\\
&D_2 = P(A_T(0)|D_T(2)=0)+P(A_T(1)|D_T(2)=0)\tau+P(A_T(2)|D_T(2)=0)\lambda\\
&+P(A_T(5)|D_T(2)=0)\tau-h_{T,2}(k)\,(2\tau+\lambda-3)\,.
\end{aligned}
$$

The probabilities $P(A_T(0)|D_T(2)=0)$, $P(A_T(1)|D_T(2)=0)$, $P(A_T(2)|D_T(2)=0)$, and $P(A_T(5)|D_T(2)=0)$ are identified from Theorem 1. In the application, we only found one solution to the quartic equation in $[0,1]$.

Identification of $SACE_{k,\cdot}(2,1)$ (Corollary II)

Similar to (16), (17), (27), and (28), we have following expressions for $SACE_{k,3}(2,1)$,

$$P(Y_T(1)>k|A_T(0)\cup A_T(1)) = \frac{P(Y_T(1)>k|A_T(0))P(A_T(0))+P(Y_T(1)>k|A_T(1))P(A_T(1))}{A_T(0)\cup A_T(1)}, \tag{33}$$

$$P(Y_T(2)>k|A_T(0)\cup A_T(1)) = \frac{P(Y_T(2)>k|A_T(0))P(A_T(0))+P(Y_T(2)>k|A_T(1))P(A_T(1))}{A_T(0)\cup A_T(1)}. \tag{34}$$

In the process of identifying $SACE_{k,3}(1,0)$ and $SACE_{k,3}(2,0)$, we identified $P(Y_T(1)>k|A_T(0))$, $P(Y_T(1)>k|A_T(1))$, $P(Y_T(2)>k|A_T(0))$, and $P(Y_T(2)>k|A_T(5))$. In addition, the mixing probabilities were identified in Theorem I.

Detailed Calculations for the Covariance Matrix of $SACE_{k,\cdot}(1,0)$

We first describe the calculation for the covariance matrix for the log of $SACE_{k,1}(1,0)$. The log of $SACE_{k,1}(1,0)$ is

$$
\begin{aligned}
\log \mathrm{SACE}_{k,1}(1,0) \;=\;& \log P(Y_T(1)>k|A_T(0)) - \log\{1-P(Y_T(1)>k|A_T(0))\}\\
&-\log P(Y_T(0)>k|A_T(0)) + \log\{1-P(Y_T(0)>k|A_T(0))\}\,.
\end{aligned}
$$

The derivatives of $\log \mathrm{SACE}_k(1,0)$ with respect to $\beta_0(0)$, $\beta_0(1)$, $\beta_0(2)$, $\beta(0)$, $\beta(1)$, and $\beta(2)$ are, respectively, given by

$$\frac{\partial \log SACE_{k,1}(1,0)}{\partial \beta_0(\mathrm{tx})} = \frac{\partial \log SACE_{k,1}(1,0)}{\partial h_{T,\mathrm{tx}}(k)} \frac{\partial h_{T,\mathrm{tx}}(k)}{\partial \beta_0(\mathrm{tx})},$$
$$\frac{\partial \log SACE_{k,1}(1,0)}{\partial \beta(\mathrm{tx})} = \frac{\partial \log SACE_{k,1}(1,0)}{\partial h_{T,\mathrm{tx}}(k)} \frac{\partial h_{T,\mathrm{tx}}(k)}{\partial \beta(\mathrm{tx})}.$$

$\frac{\partial h_{T,\mathrm{tx}}(k)}{\partial \beta_0(\mathrm{tx})}$ and $\frac{\partial h_{T,\mathrm{tx}}(k)}{\partial \beta(\mathrm{tx})}$ are given in Lee and Daniels (2007). The derivatives of $\log \mathrm{SACE}_{k,1}(1,0)$ with respect to ψ_0, ψ_1, ψ_2, $h_{T,0}(k)$, $h_{T,1}(k)$, and $h_{T,2}(k)$ are given by

$$\begin{aligned}
\frac{\partial \log SACE_{k,1}(1,0)}{\partial \psi_0} &= \frac{\frac{\partial P(Y_T(1)>k|A_T(0))}{\partial \psi_0}}{P(Y_T(1)>k|A_T(0))\{1-P(Y_T(1)>k|A_T(0))\}} \\
&\quad - \frac{\frac{\partial P(Y_T(0)>k|A_T(0))}{\partial \psi_0}}{P(Y_T(0)>k|A_T(0))\{1-P(Y_T(0)>k|A_T(0))\}}, \qquad (35) \\
\frac{\partial \log SACE_{k,1}(1,0)}{\partial \psi_1} &= \frac{\frac{\partial P(Y_T(1)>k|A_T(0))}{\partial \psi_1}}{P(Y_T(1)>k|A_T(0))\{1-P(Y_T(1)>k|A_T(0))\}} \\
&\quad - \frac{\frac{\partial P(Y_T(0)>k|A_T(0))}{\partial \psi_1}}{P(Y_T(0)>k|A_T(0))\{1-P(Y_T(0)>k|A_T(0))\}}, \qquad (36) \\
\frac{\partial \log SACE_{k,1}(1,0)}{\partial \psi_2} &= \frac{\frac{\partial P(Y_T(1)>k|A_T(0))}{\partial \psi_2}}{P(Y_T(1)>k|A_T(0))\{1-P(Y_T(1)>k|A_T(0))\}} \\
&\quad - \frac{\frac{\partial P(Y_T(0)>k|A_T(0))}{\partial \psi_2}}{P(Y_T(0)>k|A_T(0))\{1-P(Y_T(0)>k|A_T(0))\}}, \qquad (37) \\
\frac{\partial \log SACE_{k,1}(1,0)}{\partial h_{T,0}(k)} &= - \frac{\frac{\partial P(Y_T(0)>k|A_T(0))}{\partial h_{T,0}(k)}}{P(Y_T(0)>k|A_T(0))\{1-P(Y_T(0)>k|A_T(0))\}}, \qquad (38) \\
\frac{\partial \log SACE_{k,1}(1,0)}{\partial h_{T,1}(k)} &= \frac{\frac{\partial P(Y_T(1)>k|A_T(0))}{\partial h_{T,1}(k)}}{P(Y_T(1)>k|A_T(0))\{1-P(Y_T(1)>k|A_T(0))\}}, \qquad (39) \\
\frac{\partial \log SACE_{k,1}(1,0)}{\partial h_{T,2}(k)} &= 0,
\end{aligned}$$

where the numerators in (35)-(39) are calculated using numerical derivatives.

We now describe the calculation for the covariance matrix for the log of $SACE_{k,2}(1,0)$.

The log of $SACE_{k,2}(1,0)$ is

$$\begin{aligned}\log \mathrm{SACE}_{k,2}(1,0) &= \log P(Y_T(1) > k|A_T(4)) - \log\left\{1 - P(Y_T(1) > k|A_T(4))\right\} \\ &\quad - \log P(Y_T(0) > k|A_T(4)) + \log\left\{1 - P(Y_T(0) > k|A_T(4))\right\}.\end{aligned}$$

The derivatives of $\log \mathrm{SACE}_{k,2}(1,0)$ with respect to $\beta_0(0)$, $\beta_0(1)$, $\beta_0(2)$, $\beta(0)$, $\beta(1)$, and $\beta(2)$ are, respectively, given by

$$\begin{aligned}\frac{\partial \log SACE_{k,2}(1,0)}{\partial \beta_0(\mathrm{tx})} &= \frac{\partial \log SACE_{k,2}(1,0)}{\partial h_{T,\mathrm{tx}}(k)} \frac{\partial h_{T,\mathrm{tx}}(k)}{\partial \beta_0(\mathrm{tx})}, \\ \frac{\partial \log SACE_{k,2}(1,0)}{\partial \beta(\mathrm{tx})} &= \frac{\partial \log SACE_{k,2}(1,0)}{\partial h_{T,\mathrm{tx}}(k)} \frac{\partial h_{T,\mathrm{tx}}(k)}{\partial \beta(\mathrm{tx})}.\end{aligned}$$

The derivatives of $\log \mathrm{SACE}_{k,2}(1,0)$ with respect to ψ_0, ψ_1, ψ_2, $h_{T,0}(k)$, $h_{T,1}(k)$, and $h_{T,2}(k)$ are given by

$$\begin{aligned}\frac{\partial \log SACE_{k,2}(1,0)}{\partial \psi_0} &= \frac{\frac{\partial P(Y_T(1)>k|A_T(4))}{\partial \psi_0}}{P(Y_T(1) > k|A_T(4))\left\{1 - P(Y_T(1) > k|A_T(4))\right\}} \\ &\quad - \frac{\frac{\partial P(Y_T(0)>k|A_T(4))}{\partial \psi_0}}{P(Y_T(0) > k|A_T(4))\left\{1 - P(Y_T(0) > k|A_T(4))\right\}}, \\ \frac{\partial \log SACE_{k,2}(1,0)}{\partial \psi_1} &= \frac{\frac{\partial P(Y_T(1)>k|A_T(4))}{\partial \psi_1}}{P(Y_T(1) > k|A_T(4))\left\{1 - P(Y_T(1) > k|A_T(4))\right\}} \\ &\quad - \frac{\frac{\partial P(Y_T(0)>k|A_T(4))}{\partial \psi_1}}{P(Y_T(0) > k|A_T(4))\left\{1 - P(Y_T(0) > k|A_T(4))\right\}}, \\ \frac{\partial \log SACE_{k,2}(1,0)}{\partial \psi_2} &= \frac{\frac{\partial P(Y_T(1)>k|A_T(4))}{\partial \psi_2}}{P(Y_T(1) > k|A_T(4))\left\{1 - P(Y_T(1) > k|A_T(4))\right\}} \\ &\quad - \frac{\frac{\partial P(Y_T(0)>k|A_T(4))}{\partial \psi_2}}{P(Y_T(0) > k|A_T(4))\left\{1 - P(Y_T(0) > k|A_T(4))\right\}}, \\ \frac{\partial \log SACE_{k,2}(1,0)}{\partial h_{T,0}(k)} &= -\frac{\frac{\partial P(Y_T(0)>k|A_T(4))}{\partial h_{T,0}(k)}}{P(Y_T(0) > k|A_T(4))\left\{1 - P(Y_T(0) > k|A_T(4))\right\}}, \\ \frac{\partial \log SACE_{k,2}(1,0)}{\partial h_{T,1}(k)} &= \frac{\frac{\partial P(Y_T(1)>k|A_T(4))}{\partial h_{T,1}(k)}}{P(Y_T(1) > k|A_T(4))\left\{1 - P(Y_T(1) > k|A_T(4))\right\}}, \\ \frac{\partial \log SACE_{k,2}(1,0)}{\partial h_{T,2}(k)} &= 0,\end{aligned}$$

where

$$\frac{\partial P(Y_T(j) > k|A_T(4))}{\partial \psi_{\text{tx}}} = \frac{\tau \frac{\partial P(Y_T(j)>k|A_T(0))}{\partial \psi_{\text{tx}}}}{\{1 + (\tau - 1)P(Y_T(j) > k|A_T(0))\}^2},$$
$$\frac{\partial P(Y_T(0) > k|A_T(4))}{\partial h_{T,0}(k)} = \frac{\tau \frac{\partial P(Y_T(0)>k|A_T(0))}{\partial h_{T,0}(k)}}{\{1 + (\tau - 1)P(Y_T(0) > k|A_T(0))\}^2},$$
$$\frac{\partial P(Y_T(1) > k|A_T(4))}{\partial h_{T,1}(k)} = \frac{\tau \frac{\partial P(Y_T(1)>k|A_T(0))}{\partial h_{T,1}(k)}}{\{1 + (\tau - 1)P(Y_T(1) > k|A_T(0))\}^2},$$

for $j = 0, 1$ and $\text{tx} = 0, 1, 2$.

We now describe the calculation for the covariance matrix for the log of $SACE_{k,3}(1, 0)$. The log of $SACE_{k,3}(1, 0)$ is

$$\begin{aligned}\log \text{SACE}_{k,3}(1, 0) &= \log P(Y_T(1) > k|A_T(0) \cup A_T(4)) - \log \{1 - P(Y_T(1) > k|A_T(0) \cup A_T(4))\} \\ &\quad - \log P(Y_T(0) > k|A_T(0) \cup A_T(4)) + \log \{1 - P(Y_T(0) > k|A_T(0) \cup A_T(4))\}.\end{aligned}$$

The derivatives of $\log \text{SACE}_{k,3}(1, 0)$ with respect to $\beta_0(0)$, $\beta_0(1)$, $\beta_0(2)$, $\beta(0)$, $\beta(1)$, and $\beta(2)$ are, respectively, given by

$$\frac{\partial \log SACE_{k,3}(1, 0)}{\partial \beta_0(\text{tx})} = \frac{\partial \log SACE_{k,3}(1, 0)}{\partial h_{T,\text{tx}}(k)} \frac{\partial h_{T,\text{tx}}(k)}{\partial \beta_0(\text{tx})},$$
$$\frac{\partial \log SACE_{k,3}(1, 0)}{\partial \beta(\text{tx})} = \frac{\partial \log SACE_{k,3}(1, 0)}{\partial h_{T,\text{tx}}(k)} \frac{\partial h_{T,\text{tx}}(k)}{\partial \beta(\text{tx})}.$$

The derivatives of $\log \text{SACE}_{k,3}(1, 0)$ with respect to ψ_0, ψ_1, ψ_2, $h_{T,0}(k)$, $h_{T,1}(k)$, and $h_{T,2}(k)$

are given by

$$\begin{aligned}
\frac{\partial \log SACE_{k,3}(1,0)}{\partial \psi_0} &= \frac{\frac{\partial P(Y_T(1)>k|A_T(0)\cup A_T(4))}{\partial \psi_0}}{P(Y_T(1)>k|A_T(0)\cup A_T(4))\{1-P(Y_T(1)>k|A_T(0)\cup A_T(4))\}} \\
&\quad - \frac{\frac{\partial P(Y_T(0)>k|A_T(0)\cup A_T(4))}{\partial \psi_0}}{P(Y_T(0)>k|A_T(0)\cup A_T(4))\{1-P(Y_T(0)>k|A_T(0)\cup A_T(4))\}}, \\
\frac{\partial \log SACE_{k,3}(1,0)}{\partial \psi_1} &= \frac{\frac{\partial P(Y_T(1)>k|A_T(0)\cup A_T(4))}{\partial \psi_1}}{P(Y_T(1)>k|A_T(0)\cup A_T(4))\{1-P(Y_T(1)>k|A_T(0)\cup A_T(4))\}} \\
&\quad - \frac{\frac{\partial P(Y_T(0)>k|A_T(0)\cup A_T(4))}{\partial \psi_1}}{P(Y_T(0)>k|A_T(0)\cup A_T(4))\{1-P(Y_T(0)>k|A_T(0)\cup A_T(4))\}}, \\
\frac{\partial \log SACE_{k,3}(1,0)}{\partial \psi_2} &= \frac{\frac{\partial P(Y_T(1)>k|A_T(0)\cup A_T(4))}{\partial \psi_2}}{P(Y_T(1)>k|A_T(0)\cup A_T(4))\{1-P(Y_T(1)>k|A_T(0)\cup A_T(4))\}} \\
&\quad - \frac{\frac{\partial P(Y_T(0)>k|A_T(0)\cup A_T(4))}{\partial \psi_2}}{P(Y_T(0)>k|A_T(0)\cup A_T(4))\{1-P(Y_T(0)>k|A_T(0)\cup A_T(4))\}}, \\
\frac{\partial \log SACE_{k,3}(1,0)}{\partial h_{T,0}(k)} &= -\frac{\frac{\partial P(Y_T(0)>k|A_T(0)\cup A_T(4))}{\partial h_{T,0}(k)}}{P(Y_T(0)>k|A_T(0)\cup A_T(4))\{1-P(Y_T(0)>k|A_T(0)\cup A_T(4))\}}, \\
\frac{\partial \log SACE_{k,3}(1,0)}{\partial h_{T,1}(k)} &= \frac{\frac{\partial P(Y_T(1)>k|A_T(0)\cup A_T(4))}{\partial h_{T,1}(k)}}{P(Y_T(1)>k|A_T(0)\cup A_T(4))\{1-P(Y_T(1)>k|A_T(0)\cup A_T(4))\}}, \\
\frac{\partial \log SACE_{k,3}(1,0)}{\partial h_{T,2}(k)} &= 0,
\end{aligned}$$

where

$$\frac{\partial P(Y_T(0)>k|A_T(0)\cup A_T(4))}{\partial\psi_0}=\frac{1}{P(A_T(0)\cup A_T(4))}\left\{\frac{\partial P(Y_T(0)>k|A_T(0))}{\partial\psi_0}P(A_T(0))+P(Y_T(0)>k|A_T(0))\frac{\partial P(A_T(0))}{\partial\psi_0}\right.$$
$$\left.+\frac{\partial P(Y_T(0)>k|A_T(4))}{\partial\psi_0}P(A_T(4))+P(Y_T(0)>k|A_T(4))\frac{\partial P(A_T(4))}{\partial\psi_0}\right\}$$
$$-\frac{P(Y_T(0)>k|A_T(0))P(A_T(0))+P(Y_T(0)>k|A_T(4))P(A_T(4))}{\{P(A_T(0)\cup A_T(4))\}^2}\frac{\partial P(A_T(0)\cup A_T(4))}{\partial\psi_0},$$
$$\frac{\partial P(Y_T(1)>k|A_T(0)\cup A_T(4))}{\partial\psi_0}=\frac{1}{P(A_T(0)\cup A_T(4))}\left\{\frac{\partial P(Y_T(1)>k|A_T(0))}{\partial\psi_0}P(A_T(0))+P(Y_T(1)>k|A_T(0))\frac{\partial P(A_T(0))}{\partial\psi_0}\right.$$
$$\left.+\frac{\partial P(Y_T(1)>k|A_T(4))}{\partial\psi_0}P(A_T(4))+P(Y_T(1)>k|A_T(4))\frac{\partial P(A_T(4))}{\partial\psi_0}\right\}$$
$$-\frac{P(Y_T(1)>k|A_T(0))P(A_T(0))+P(Y_T(1)>k|A_T(4))P(A_T(4))}{\{P(A_T(0)\cup A_T(4))\}^2}\frac{\partial P(A_T(0)\cup A_T(4))}{\partial\psi_0},$$
$$\frac{\partial P(Y_T(0)>k|A_T(0)\cup A_T(4))}{\partial\psi_1}=\frac{1}{P(A_T(0)\cup A_T(4))}\left\{\frac{\partial P(Y_T(0)>k|A_T(0))}{\partial\psi_1}P(A_T(0))+P(Y_T(0)>k|A_T(0))\frac{\partial P(A_T(0))}{\partial\psi_1}\right.$$
$$\left.+\frac{\partial P(Y_T(0)>k|A_T(4))}{\partial\psi_1}P(A_T(4))+P(Y_T(0)>k|A_T(4))\frac{\partial P(A_T(4))}{\partial\psi_1}\right\}$$
$$-\frac{P(Y_T(0)>k|A_T(0))P(A_T(0))+P(Y_T(0)>k|A_T(4))P(A_T(4))}{\{P(A_T(0)\cup A_T(4))\}^2}\frac{\partial P(A_T(0)\cup A_T(4))}{\partial\psi_1},$$
$$\frac{\partial P(Y_T(1)>k|A_T(0)\cup A_T(4))}{\partial\psi_1}=\frac{1}{P(A_T(0)\cup A_T(4))}\left\{\frac{\partial P(Y_T(1)>k|A_T(0))}{\partial\psi_1}P(A_T(0))+P(Y_T(1)>k|A_T(0))\frac{\partial P(A_T(0))}{\partial\psi_1}\right.$$
$$\left.+\frac{\partial P(Y_T(1)>k|A_T(4))}{\partial\psi_1}P(A_T(4))+P(Y_T(1)>k|A_T(4))\frac{\partial P(A_T(4))}{\partial\psi_1}\right\}$$
$$-\frac{P(Y_T(1)>k|A_T(0))P(A_T(0))+P(Y_T(1)>k|A_T(4))P(A_T(4))}{\{P(A_T(0)\cup A_T(4))\}^2}\frac{\partial P(A_T(0)\cup A_T(4))}{\partial\psi_1},$$
$$\frac{\partial P(Y_T(0)>k|A_T(0)\cup A_T(4))}{\partial\psi_2}=\frac{1}{P(A_T(0)\cup A_T(4))}\left\{\frac{\partial P(Y_T(0)>k|A_T(0))}{\partial\psi_2}P(A_T(0))+P(Y_T(0)>k|A_T(0))\frac{\partial P(A_T(0))}{\partial\psi_2}\right.$$
$$\left.+\frac{\partial P(Y_T(0)>k|A_T(4))}{\partial\psi_2}P(A_T(4))+P(Y_T(0)>k|A_T(4))\frac{\partial P(A_T(4))}{\partial\psi_2}\right\},$$
$$\frac{\partial P(Y_T(1)>k|A_T(0)\cup A_T(4))}{\partial\psi_2}=\frac{1}{P(A_T(0)\cup A_T(4))}\left\{\frac{\partial P(Y_T(1)>k|A_T(0))}{\partial\psi_2}P(A_T(0))+P(Y_T(1)>k|A_T(0))\frac{\partial P(A_T(0))}{\partial\psi_2}\right.$$
$$\left.+\frac{\partial P(Y_T(1)>k|A_T(4))}{\partial\psi_2}P(A_T(4))+P(Y_T(1)>k|A_T(4))\frac{\partial P(A_T(4))}{\partial\psi_2}\right\},$$
$$\frac{\partial P(Y_T(0)>k|A_T(0)\cup A_T(4))}{\partial h_{T,0}(k)}=\frac{1}{P(A_T(0)\cup A_T(4))}\left\{\frac{\partial P(Y_T(0)>k|A_T(0))}{\partial h_{T,0}(k)}P(A_T(0))+\frac{\partial P(Y_T(1)>k|A_T(4))}{\partial h_{T,1}(k)}P(A_T(4))\right\},$$
$$\frac{\partial P(Y_T(1)>k|A_T(0)\cup A_T(4))}{\partial h_{T,1}(k)}=\frac{1}{P(A_T(0)\cup A_T(4))}\left\{\frac{\partial P(Y_T(1)>k|A_T(0))}{\partial h_{T,1}(k)}P(A_T(0))+\frac{\partial P(Y_T(1)>k|A_T(4))}{\partial h_{T,1}(k)}P(A_T(4))\right\},$$

with

$$\frac{\partial P(A_T(0))}{\partial \psi_0} = E_w \left[\frac{\partial g_T(0,w)}{\partial \psi_0} \left\{ p_T^*(2,w) - p_T^*(1,w) + 1 \right\} + g_T(0,w) \left\{ \frac{\partial p_T^*(2,w)}{\partial \psi_0} - \frac{\partial p_T^*(1,w)}{\partial \psi_0} \right\} \right],$$
$$\frac{\partial P(A_T(4))}{\partial \psi_0} = E_w \left\{ \frac{\partial p_T^*(1,w)}{\partial \psi_0} + p_T^*(1,w) \frac{\partial g_T(0,w)}{\partial \psi_0} \right\} - \frac{\partial P(A_T(0))}{\partial \psi_0},$$
$$\frac{\partial P(A_T(0) \cup A_T(4))}{\partial \psi_0} = \frac{\partial P(A_T(0))}{\partial \psi_0} + \frac{\partial P(A_T(4))}{\partial \psi_0},$$
$$\frac{\partial P(A_T(0))}{\partial \psi_1} = -E_w \left\{ g_T(0,w) \frac{\partial p_T^*(1,w)}{\partial \psi_1} \right\},$$
$$\frac{\partial P(A_T(4))}{\partial \psi_1} = 2E_w \left\{ g_T(0,w) \frac{\partial p_T^*(1,w)}{\partial \psi_1} \right\},$$
$$\frac{\partial P(A_T(0) \cup A_T(4))}{\partial \psi_1} = \frac{\partial P(A_T(0))}{\partial \psi_1} + \frac{\partial P(A_T(4))}{\partial \psi_1},$$
$$\frac{\partial P(A_T(0))}{\partial \psi_2} = E_w \left\{ g_T(0,w) \frac{\partial p_T^*(2,w)}{\partial \psi_2} \right\},$$
$$\frac{\partial P(A_T(4))}{\partial \psi_2} = -\frac{\partial P(A_T(0))}{\partial \psi_2}.$$

Detailed Calculations for the Covariance Matrix of $SACE_{k,\cdot}(2,0)$

The log of $\mathrm{SACE}_{k,1}(2,0)$ is

$$\begin{aligned} \log \mathrm{SACE}_{k,1}(2,0) &= \log P(Y_T(2) > k | A_T(0)) - \log \left\{ 1 - P(Y_T(2) > k | A_T(0)) \right\} \\ &\quad - \log P(Y_T(0) > k | A_T(0)) + \log \left\{ 1 - P(Y_T(0) > k | A_T(0)) \right\}. \end{aligned}$$

Similar to the calculation of derivatives in $\mathrm{SACE}_{k,1}(1,0)$, the derivatives of $\log \mathrm{SACE}_{k,1}(2,0)$ with respect to $\beta_0(0)$, $\beta_0(1)$, $\beta_0(2)$, $\beta(0)$, $\beta(1)$, and $\beta(2)$ are, respectively, given by

$$\frac{\partial \log SACE_{k,1}(2,0)}{\partial \beta_0(\mathrm{tx})} = \frac{\partial \log SACE_{k,1}(2,0)}{\partial h_{T,\mathrm{tx}}(k)} \frac{\partial h_{T,\mathrm{tx}}(k)}{\partial \beta_0(\mathrm{tx})},$$
$$\frac{\partial \log SACE_{k,1}(2,0)}{\partial \beta(\mathrm{tx})} = \frac{\partial \log SACE_{k,1}(2,0)}{\partial h_{T,\mathrm{tx}}(k)} \frac{\partial h_{T,\mathrm{tx}}(k)}{\partial \beta(\mathrm{tx})}.$$

The derivatives of $\log \mathrm{SACE}_{k,1}(2,0)$ with respect to ψ_0, ψ_1, ψ_2, $h_{T,0}(k)$, $h_{T,1}(k)$, and $h_{T,2}(k)$

are, respectively, given by

$$\begin{aligned}
&\frac{\partial \log SACE_{k,1}(2,0)}{\partial \psi_0} \\
&= \frac{\frac{\partial P(Y_T(2)>k|A_T(0))}{\partial \psi_0}}{P(Y_T(2)>k|A_T(0))\{1-P(Y_T(2)>k|A_T(0))\}} - \frac{\frac{\partial P(Y_T(0)>k|A_T(0))}{\partial \psi_0}}{P(Y_T(0)>k|A_T(0))\{1-P(Y_T(0)>k|A_T(0))\}}, \qquad (40) \\
&\frac{\partial \log SACE_{k,1}(2,0)}{\partial \psi_1} \\
&= \frac{\frac{\partial P(Y_T(2)>k|A_T(0))}{\partial \psi_1}}{P(Y_T(2)>k|A_T(0))\{1-P(Y_T(2)>k|A_T(0))\}} - \frac{\frac{\partial P(Y_T(0)>k|A_T(0))}{\partial \psi_1}}{P(Y_T(0)>k|A_T(0))\{1-P(Y_T(0)>k|A_T(0))\}}, \qquad (41) \\
&\frac{\partial \log SACE_{k,1}(2,0)}{\partial \psi_2} \\
&= \frac{\frac{\partial P(Y_T(2)>k|A_T(0))}{\partial \psi_2}}{P(Y_T(2)>k|A_T(0))\{1-P(Y_T(2)>k|A_T(0))\}} - \frac{\frac{\partial P(Y_T(0)>k|A_T(0))}{\partial \psi_2}}{P(Y_T(0)>k|A_T(0))\{1-P(Y_T(0)>k|A_T(0))\}}, \qquad (42) \\
&\frac{\partial \log SACE_{k,1}(2,0)}{\partial h_{T,0}(k)} = -\frac{\frac{\partial P(Y_T(0)>k|A_T(0)}{\partial h_{T,0}(k)}}{P(Y_T(0)>k|A_T(0))\{1-P(Y_T(0)>k|A_T(0))\}}, \qquad (43) \\
&\frac{\partial \log SACE_{k,1}(2,0)}{\partial h_{T,1}(k)} = 0, \\
&\frac{\partial \log SACE_{k,1}(2,0)}{\partial h_{T,2}(k)} = \frac{\frac{\partial P(Y_T(2)>k|A_T(0))}{\partial h_{T,2}(k)}}{P(Y_T(2)>k|A_T(0))\{1-P(Y_T(2)>k|A_T(0))\}}. \qquad (44)
\end{aligned}$$

where the numerator parts in (40)-(44) are calculated using numerical derivatives.

We now describe the calculation for the covariance matrix for the log of $SACE_{k,2}(2,0)$. The log of $SACE_{k,2}(2,0)$ is

$$\begin{aligned}
\log \mathrm{SACE}_{k,2}(2,0) &= \log P(Y_T(2)>k|A_T(5)) - \log\{1-P(Y_T(2)>k|A_T(5))\} \\
&\quad - \log P(Y_T(0)>k|A_T(5)) + \log\{1-P(Y_T(0)>k|A_T(5))\}.
\end{aligned}$$

The derivatives of $\log \mathrm{SACE}_{k,2}(2,0)$ with respect to $\beta_0(0)$, $\beta_0(1)$, $\beta_0(2)$, $\beta(0)$, $\beta(1)$, and $\beta(2)$ are, respectively, given by

$$\begin{aligned}
\frac{\partial \log SACE_{k,2}(2,0)}{\partial \beta_0(\mathrm{tx})} &= \frac{\partial \log SACE_{k,2}(2,0)}{\partial h_{T,\mathrm{tx}}(k)} \frac{\partial h_{T,\mathrm{tx}}(k)}{\partial \beta_0(\mathrm{tx})}, \\
\frac{\partial \log SACE_{k,2}(2,0)}{\partial \beta(\mathrm{tx})} &= \frac{\partial \log SACE_{k,2}(2,0)}{\partial h_{T,\mathrm{tx}}(k)} \frac{\partial h_{T,\mathrm{tx}}(k)}{\partial \beta(\mathrm{tx})}.
\end{aligned}$$

The derivatives of $\log \mathrm{SACE}_{k,2}(2,0)$ with respect to ψ_0, ψ_1, ψ_2, $h_{T,0}(k)$, $h_{T,1}(k)$, and $h_{T,2}(k)$

are given by

$$\begin{aligned}
\frac{\partial \log SACE_{k,2}(2,0)}{\partial \psi_0} &= \frac{\frac{\partial P(Y_T(2)>k|A_T(5))}{\partial \psi_0}}{P(Y_T(2)>k|A_T(5))\left\{1-P(Y_T(2)>k|A_T(5))\right\}} \\
&\quad - \frac{\frac{\partial P(Y_T(0)>k|A_T(5))}{\partial \psi_0}}{P(Y_T(0)>k|A_T(5))\left\{1-P(Y_T(0)>k|A_T(5))\right\}}, \\
\frac{\partial \log SACE_{k,2}(2,0)}{\partial \psi_1} &= \frac{\frac{\partial P(Y_T(2)>k|A_T(5))}{\partial \psi_1}}{P(Y_T(2)>k|A_T(5))\left\{1-P(Y_T(2)>k|A_T(5))\right\}} \\
&\quad - \frac{\frac{\partial P(Y_T(0)>k|A_T(5))}{\partial \psi_1}}{P(Y_T(0)>k|A_T(5))\left\{1-P(Y_T(0)>k|A_T(5))\right\}}, \\
\frac{\partial \log SACE_{k,2}(2,0)}{\partial \psi_2} &= \frac{\frac{\partial P(Y_T(2)>k|A_T(5))}{\partial \psi_2}}{P(Y_T(2)>k|A_T(5))\left\{1-P(Y_T(2)>k|A_T(5))\right\}} \\
&\quad - \frac{\frac{\partial P(Y_T(0)>k|A_T(5))}{\partial \psi_2}}{P(Y_T(0)>k|A_T(5))\left\{1-P(Y_T(0)>k|A_T(5))\right\}}, \\
\frac{\partial \log SACE_{k,2}(2,0)}{\partial h_{T,0}(k)} &= -\frac{\frac{\partial P(Y_T(0)>k|A_T(5))}{\partial h_{T,0}(k)}}{P(Y_T(0)>k|A_T(5))\left\{1-P(Y_T(0)>k|A_T(5))\right\}}, \\
\frac{\partial \log SACE_{k,2}(2,0)}{\partial h_{T,1}(k)} &= 0, \\
\frac{\partial \log SACE_{k,2}(2,0)}{\partial h_{T,2}(k)} &= \frac{\frac{\partial P(Y_T(2)>k|A_T(5))}{\partial h_{T,2}(k)}}{P(Y_T(2)>k|A_T(5))\left\{1-P(Y_T(2)>k|A_T(5))\right\}},
\end{aligned}$$

where

$$\begin{aligned}
\frac{\partial P(Y_T(j)>k|A_T(5))}{\partial \psi_{\mathrm{tx}}} &= \frac{\tau \frac{\partial P(Y_T(j)>k|A_T(0))}{\partial \psi_{\mathrm{tx}}}}{\left\{1+(\tau-1)P(Y_T(j)>k|A_T(0))\right\}^2}, \\
\frac{\partial P(Y_T(0)>k|A_T(5))}{\partial h_{T,0}(k)} &= \frac{\tau \frac{\partial P(Y_T(0)>k|A_T(0))}{\partial h_{T,0}(k)}}{\left\{1+(\tau-1)P(Y_T(0)>k|A_T(0))\right\}^2}, \\
\frac{\partial P(Y_T(2)>k|A_T(5))}{\partial h_{T,2}(k)} &= \frac{\tau \frac{\partial P(Y_T(2)>k|A_T(0))}{\partial h_{T,2}(k)}}{\left\{1+(\tau-1)P(Y_T(2)>k|A_T(0))\right\}^2},
\end{aligned}$$

for $j = 0, 2$ and $\mathrm{tx} = 0, 1, 2$.

We now describe the calculation for the covariance matrix for the log of $SACE_{k,3}(2,0)$.

The log of $SACE_{k,3}(2,0)$ is

$$\begin{aligned}\log \mathrm{SACE}_{k,3}(2,0) &= \log P(Y_T(2) > k | A_T(0) \cup A_T(5)) - \log \{1 - P(Y_T(2) > k | A_T(0) \cup A_T(5))\} \\ &\quad - \log P(Y_T(0) > k | A_T(0) \cup A_T(5)) + \log \{1 - P(Y_T(0) > k | A_T(0) \cup A_T(5))\}.\end{aligned}$$

The derivatives of $\log \mathrm{SACE}_{k,3}(2,0)$ with respect to $\beta_0(0)$, $\beta_0(1)$, $\beta_0(2)$, $\beta(0)$, $\beta(1)$, and $\beta(2)$ are, respectively, given by

$$\begin{aligned}\frac{\partial \log SACE_{k,3}(2,0)}{\partial \beta_0(\mathrm{tx})} &= \frac{\partial \log SACE_{k,3}(2,0)}{\partial h_{T,\mathrm{tx}}(k)} \frac{\partial h_{T,\mathrm{tx}}(k)}{\partial \beta_0(\mathrm{tx})}, \\ \frac{\partial \log SACE_{k,3}(2,0)}{\partial \beta(\mathrm{tx})} &= \frac{\partial \log SACE_{k,3}(2,0)}{\partial h_{T,\mathrm{tx}}(k)} \frac{\partial h_{T,\mathrm{tx}}(k)}{\partial \beta(\mathrm{tx})}.\end{aligned}$$

The derivatives of $\log \mathrm{SACE}_{k,3}(2,0)$ with respect to ψ_0, ψ_1, ψ_2, $h_{T,0}(k)$, $h_{T,1}(k)$, and $h_{T,2}(k)$ are given by

$$\begin{aligned}\frac{\partial \log SACE_{k,3}(2,0)}{\partial \psi_0} &= \frac{\frac{\partial P(Y_T(2)>k|A_T(0)\cup A_T(5))}{\partial \psi_0}}{P(Y_T(2) > k|A_T(0) \cup A_T(5)) \{1 - P(Y_T(2) > k|A_T(0) \cup A_T(5))\}} \\ &\quad - \frac{\frac{\partial P(Y_T(0)>k|A_T(0)\cup A_T(5))}{\partial \psi_0}}{P(Y_T(0) > k|A_T(0) \cup A_T(5)) \{1 - P(Y_T(0) > k|A_T(0) \cup A_T(5))\}}, \\ \frac{\partial \log SACE_{k,3}(2,0)}{\partial \psi_1} &= \frac{\frac{\partial P(Y_T(2)>k|A_T(0)\cup A_T(5))}{\partial \psi_1}}{P(Y_T(2) > k|A_T(0) \cup A_T(5)) \{1 - P(Y_T(1) > k|A_T(0) \cup A_T(5))\}} \\ &\quad - \frac{\frac{\partial P(Y_T(0)>k|A_T(0)\cup A_T(5))}{\partial \psi_1}}{P(Y_T(0) > k|A_T(0) \cup A_T(5)) \{1 - P(Y_T(0) > k|A_T(0) \cup A_T(5))\}}, \\ \frac{\partial \log SACE_{k,3}(2,0)}{\partial \psi_2} &= \frac{\frac{\partial P(Y_T(2)>k|A_T(0)\cup A_T(5))}{\partial \psi_2}}{P(Y_T(2) > k|A_T(0) \cup A_T(5)) \{1 - P(Y_T(2) > k|A_T(0) \cup A_T(5))\}} \\ &\quad - \frac{\frac{\partial P(Y_T(0)>k|A_T(0)\cup A_T(5))}{\partial \psi_2}}{P(Y_T(0) > k|A_T(0) \cup A_T(5)) \{1 - P(Y_T(0) > k|A_T(0) \cup A_T(5))\}}, \\ \frac{\partial \log SACE_{k,3}(2,0)}{\partial h_{T,0}(k)} &= - \frac{\frac{\partial P(Y_T(0)>k|A_T(0)\cup A_T(5))}{\partial h_{T,0}(k)}}{P(Y_T(0) > k|A_T(0) \cup A_T(5)) \{1 - P(Y_T(0) > k|A_T(0) \cup A_T(5))\}}, \\ \frac{\partial \log SACE_{k,3}(2,0)}{\partial h_{T,1}(k)} &= 0, \\ \frac{\partial \log SACE_{k,3}(2,0)}{\partial h_{T,2}(k)} &= \frac{\frac{\partial P(Y_T(2)>k|A_T(0)\cup A_T(5))}{\partial h_{T,2}(k)}}{P(Y_T(2) > k|A_T(0) \cup A_T(5)) \{1 - P(Y_T(2) > k|A_T(0) \cup A_T(5))\}},\end{aligned}$$

where

$$\frac{\partial P(Y_T(0)>k|A_T(0)\cup A_T(5))}{\partial \psi_0} = \frac{1}{P(A_T(0)\cup A_T(5))}\left\{\frac{\partial P(Y_T(0)>k|A_T(0))}{\partial \psi_0}P(A_T(0)) + P(Y_T(0)>k|A_T(0))\frac{\partial P(A_T(0))}{\partial \psi_0}\right.$$
$$\left. + \frac{\partial P(Y_T(0)>k|A_T(5))}{\partial \psi_0}P(A_T(5)) + P(Y_T(0)>k|A_T(5))\frac{\partial P(A_T(5))}{\partial \psi_0}\right\}$$
$$- \frac{P(Y_T(0)>k|A_T(0))P(A_T(0)) + P(Y_T(0)>k|A_T(5))P(A_T(5))}{\{P(A_T(0)\cup A_T(5))\}^2}\frac{\partial P(A_T(0)\cup A_T(5))}{\partial \psi_0},$$
$$\frac{\partial P(Y_T(2)>k|A_T(0)\cup A_T(5))}{\partial \psi_0} = \frac{1}{P(A_T(0)\cup A_T(5))}\left\{\frac{\partial P(Y_T(2)>k|A_T(0))}{\partial \psi_0}P(A_T(0)) + P(Y_T(2)>k|A_T(0))\frac{\partial P(A_T(0))}{\partial \psi_0}\right.$$
$$\left. + \frac{\partial P(Y_T(2)>k|A_T(5))}{\partial \psi_0}P(A_T(5)) + P(Y_T(2)>k|A_T(5))\frac{\partial P(A_T(5))}{\partial \psi_0}\right\}$$
$$- \frac{P(Y_T(2)>k|A_T(0))P(A_T(0)) + P(Y_T(2)>k|A_T(5))P(A_T(5))}{\{P(A_T(0)\cup A_T(5))\}^2}\frac{\partial P(A_T(0)\cup A_T(5))}{\partial \psi_0},$$
$$\frac{\partial P(Y_T(0)>k|A_T(0)\cup A_T(5))}{\partial \psi_1} = \frac{1}{P(A_T(0)\cup A_T(5))}\left\{\frac{\partial P(Y_T(0)>k|A_T(0))}{\partial \psi_1}P(A_T(0)) + P(Y_T(0)>k|A_T(0))\frac{\partial P(A_T(0))}{\partial \psi_1}\right.$$
$$\left. + \frac{\partial P(Y_T(0)>k|A_T(5))}{\partial \psi_1}P(A_T(5)) + P(Y_T(0)>k|A_T(5))\frac{\partial P(A_T(5))}{\partial \psi_1}\right\},$$
$$\frac{\partial P(Y_T(2)>k|A_T(0)\cup A_T(5))}{\partial \psi_1} = \frac{1}{P(A_T(0)\cup A_T(5))}\left\{\frac{\partial P(Y_T(2)>k|A_T(0))}{\partial \psi_1}P(A_T(0)) + P(Y_T(2)>k|A_T(0))\frac{\partial P(A_T(0))}{\partial \psi_1}\right.$$
$$\left. + \frac{\partial P(Y_T(2)>k|A_T(5))}{\partial \psi_1}P(A_T(5)) + P(Y_T(2)>k|A_T(5))\frac{\partial P(A_T(5))}{\partial \psi_1}\right\},$$
$$\frac{\partial P(Y_T(0)>k|A_T(0)\cup A_T(5))}{\partial \psi_2} = \frac{1}{P(A_T(0)\cup A_T(5))}\left\{\frac{\partial P(Y_T(0)>k|A_T(0))}{\partial \psi_2}P(A_T(0)) + P(Y_T(0)>k|A_T(0))\frac{\partial P(A_T(0))}{\partial \psi_2}\right.$$
$$\left. + \frac{\partial P(Y_T(0)>k|A_T(5))}{\partial \psi_2}P(A_T(5))\right\}$$
$$- \frac{P(Y_T(0)>k|A_T(0))P(A_T(0)) + P(Y_T(0)>k|A_T(5))P(A_T(5))}{\{P(A_T(0)\cup A_T(5))\}^2}\frac{\partial P(A_T(0)\cup A_T(5))}{\partial \psi_2},$$
$$\frac{\partial P(Y_T(2)>k|A_T(0)\cup A_T(5))}{\partial \psi_2} = \frac{1}{P(A_T(0)\cup A_T(5))}\left\{\frac{\partial P(Y_T(2)>k|A_T(0))}{\partial \psi_2}P(A_T(0)) + P(Y_T(2)>k|A_T(0))\frac{\partial P(A_T(0))}{\partial \psi_2}\right.$$
$$\left. + \frac{\partial P(Y_T(2)>k|A_T(5))}{\partial \psi_2}P(A_T(5))\right\}$$
$$- \frac{P(Y_T(2)>k|A_T(0))P(A_T(0)) + P(Y_T(2)>k|A_T(5))P(A_T(5))}{\{P(A_T(0)\cup A_T(5))\}^2}\frac{\partial P(A_T(0)\cup A_T(5))}{\partial \psi_2},$$
$$\frac{\partial P(Y_T(0)>k|A_T(0)\cup A_T(5))}{\partial h_{T,0}} = \frac{1}{P(A_T(0)\cup A_T(5))}\left\{\frac{\partial P(Y_T(0)>k|A_T(0))}{\partial h_{T,0}(k)}P(A_T(0)) + \frac{\partial P(Y_T(0)>k|A_T(5))}{\partial h_{T,0}(k)}P(A_T(5))\right\},$$
$$\frac{\partial P(Y_T(2)>k|A_T(0)\cup A_T(5))}{\partial h_{T,2}} = \frac{1}{P(A_T(0)\cup A_T(5))}\left\{\frac{\partial P(Y_T(2)>k|A_T(0))}{\partial h_{T,2}(k)}P(A_T(0)) + \frac{\partial P(Y_T(2)>k|A_T(5))}{\partial h_{T,2}(k)}P(A_T(5))\right\},$$

with

$$\frac{\partial P(A_T(5))}{\partial \psi_0} = E_w \left\{ \frac{\partial g_T(0,w)}{\partial \psi_0} \left(1 - p_T^*(1,w)\right) - g_T^*(0,w) \frac{\partial p_T^*(1,w)}{\partial \psi_0} \right\},$$
$$\frac{\partial P(A_T(0) \cup A_T(5))}{\partial \psi_0} = \frac{\partial P(A_T(0))}{\partial \psi_0} + \frac{\partial P(A_T(5))}{\partial \psi_0},$$
$$\frac{\partial P(A_T(5))}{\partial \psi_1} = -E_w \left\{ g_T(0,w) \frac{\partial p_T^*(1,w)}{\partial \psi_1} \right\},$$
$$\frac{\partial P(A_T(0) \cup A_T(5))}{\partial \psi_2} = \frac{\partial P(A_T(0))}{\partial \psi_2}.$$

Detailed Calculations for the Covariance Matrix of $SACE_{k,\cdot}(2,1)$

The log of $\mathrm{SACE}_{k,1}(2,1)$ is

$$\begin{aligned} \log \mathrm{SACE}_{k,1}(2,1) &= \log P(Y_T(2) > k | A_T(0)) - \log\left\{1 - P(Y_T(2) > k | A_T(0))\right\} \\ &\quad - \log P(Y_T(1) > k | A_T(0)) + \log\left\{1 - P(Y_T(1) > k | A_T(0))\right\}. \end{aligned}$$

The derivatives of $\log \mathrm{SACE}_{k,1}(2,1)$ with respect to $\beta_0(0)$, $\beta_0(1)$, $\beta_0(2)$, $\beta(0)$, $\beta(1)$, and $\beta(2)$ are, respectively, given by

$$\frac{\partial \log SACE_{k,1}(2,1)}{\partial \beta_0(\mathrm{tx})} = \frac{\partial \log SACE_{k,1}(2,1)}{\partial h_{T,\mathrm{tx}}(k)} \frac{\partial h_{T,\mathrm{tx}}(k)}{\partial \beta_0(\mathrm{tx})},$$
$$\frac{\partial \log SACE_{k,1}(2,1)}{\partial \beta(\mathrm{tx})} = \frac{\partial \log SACE_{k,1}(2,1)}{\partial h_{T,\mathrm{tx}}(k)} \frac{\partial h_{T,\mathrm{tx}}(k)}{\partial \beta(\mathrm{tx})}.$$

The derivatives of $\log \mathrm{SACE}_{k,1}(2,1)$ with respect to ψ_0, ψ_1, ψ_2, $h_{T,0}(k)$, $h_{T,1}(k)$, and $h_{T,2}(k)$

are given by

$$\begin{aligned}
&\frac{\partial \log SACE_{k,1}(2,1)}{\partial \psi_0} \\
&= \frac{\frac{\partial P(Y_T(2)>k|A_T(0))}{\partial \psi_0}}{P(Y_T(2)>k|A_T(0))\{1-P(Y_T(2)>k|A_T(0))\}} - \frac{\frac{\partial P(Y_T(1)>k|A_T(0))}{\partial \psi_0}}{P(Y_T(1)>k|A_T(0))\{1-P(Y_T(1)>k|A_T(0))\}}, \\
&\frac{\partial \log SACE_{k,1}(2,1)}{\partial \psi_1} \\
&= \frac{\frac{\partial P(Y_T(2)>k|A_T(0))}{\partial \psi_1}}{P(Y_T(2)>k|A_T(0))\{1-P(Y_T(2)>k|A_T(0))\}} - \frac{\frac{\partial P(Y_T(1)>k|A_T(0))}{\partial \psi_1}}{P(Y_T(1)>k|A_T(0))\{1-P(Y_T(1)>k|A_T(0))\}}, \\
&\frac{\partial \log SACE_{k,1}(2,1)}{\partial \psi_2} \\
&= \frac{\frac{\partial P(Y_T(2)>k|A_T(0))}{\partial \psi_2}}{P(Y_T(2)>k|A_T(0))\{1-P(Y_T(2)>k|A_T(0))\}} - \frac{\frac{\partial P(Y_T(1)>k|A_T(0))}{\partial \psi_2}}{P(Y_T(1)>k|A_T(0))\{1-P(Y_T(1)>k|A_T(0))\}}, \\
&\frac{\partial \log SACE_{k,1}(2,1)}{\partial h_{T,0}(k)} = 0, \\
&\frac{\partial \log SACE_{k,1}(2,1)}{\partial h_{T,1}(k)} = -\frac{\frac{\partial P(Y_T(1)>k|A_T(0)}{\partial h_{T,1}(k)}}{P(Y_T(1)>k|A_T(0))\{1-P(Y_T(1)>k|A_T(0))\}}, \\
&\frac{\partial \log SACE_{k,1}(2,1)}{\partial h_{T,2}(k)} = \frac{\frac{\partial P(Y_T(2)>k|A_T(0))}{\partial h_{T,2}(k)}}{P(Y_T(2)>k|A_T(0))\{1-P(Y_T(2)>k|A_T(0))\}}.
\end{aligned}$$

We now describe the calculation for the covariance matrix for the log of $SACE_{k,2}(2,1)$. The log of $SACE_{k,2}(2,1)$ is

$$\begin{aligned}
\log \mathrm{SACE}_{k,2}(2,1) &= \log P(Y_T(2)>k|A_T(1)) - \log\{1-P(Y_T(2)>k|A_T(1))\} \\
&\quad - \log P(Y_T(1)>k|A_T(1)) + \log\{1-P(Y_T(1)>k|A_T(1))\}.
\end{aligned}$$

The derivatives of $\log \mathrm{SACE}_{k,2}(2,1)$ with respect to $\beta_0(0)$, $\beta_0(1)$, $\beta_0(2)$, $\beta(0)$, $\beta(1)$, and $\beta(2)$ are, respectively, given by

$$\begin{aligned}
\frac{\partial \log SACE_{k,2}(2,1)}{\partial \beta_0(\mathrm{tx})} &= \frac{\partial \log SACE_{k,2}(2,1)}{\partial h_{T,\mathrm{tx}}(k)} \frac{\partial h_{T,\mathrm{tx}}(k)}{\partial \beta_0(\mathrm{tx})}, \\
\frac{\partial \log SACE_{k,2}(2,1)}{\partial \beta(\mathrm{tx})} &= \frac{\partial \log SACE_{k,2}(2,1)}{\partial h_{T,\mathrm{tx}}(k)} \frac{\partial h_{T,\mathrm{tx}}(k)}{\partial \beta(\mathrm{tx})}.
\end{aligned}$$

The derivatives of $\log \mathrm{SACE}_{k,2}(2,1)$ with respect to ψ_0, ψ_1, ψ_2, $h_{T,0}(k)$, $h_{T,1}(k)$, and $h_{T,2}(k)$

are given by

$$\begin{aligned}
\frac{\partial \log SACE_{k,2}(2,1)}{\partial \psi_0} &= \frac{\frac{\partial P(Y_T(2)>k|A_T(1))}{\partial \psi_0}}{P(Y_T(2)>k|A_T(1))\left\{1-P(Y_T(2)>k|A_T(1))\right\}} \\
&\quad - \frac{\frac{\partial P(Y_T(1)>k|A_T(1))}{\partial \psi_0}}{P(Y_T(1)>k|A_T(1))\left\{1-P(Y_T(1)>k|A_T(1))\right\}}, \\
\frac{\partial \log SACE_{k,2}(2,1)}{\partial \psi_1} &= \frac{\frac{\partial P(Y_T(2)>k|A_T(1))}{\partial \psi_1}}{P(Y_T(2)>k|A_T(1))\left\{1-P(Y_T(2)>k|A_T(1))\right\}} \\
&\quad - \frac{\frac{\partial P(Y_T(1)>k|A_T(1))}{\partial \psi_1}}{P(Y_T(1)>k|A_T(1))\left\{1-P(Y_T(1)>k|A_T(1))\right\}}, \\
\frac{\partial \log SACE_{k,2}(2,1)}{\partial \psi_2} &= \frac{\frac{\partial P(Y_T(2)>k|A_T(1))}{\partial \psi_2}}{P(Y_T(2)>k|A_T(1))\left\{1-P(Y_T(2)>k|A_T(1))\right\}} \\
&\quad - \frac{\frac{\partial P(Y_T(1)>k|A_T(1))}{\partial \psi_2}}{P(Y_T(1)>k|A_T(1))\left\{1-P(Y_T(1)>k|A_T(1))\right\}}, \\
\frac{\partial \log SACE_{k,2}(2,1)}{\partial h_{T,0}(k)} &= 0, \\
\frac{\partial \log SACE_{k,2}(2,1)}{\partial h_{T,1}(k)} &= -\frac{\frac{\partial P(Y_T(1)>k|A_T(1))}{\partial h_{T,1}(k)}}{P(Y_T(1)>k|A_T(1))\left\{1-P(Y_T(1)>k|A_T(5))\right\}}, \\
\frac{\partial \log SACE_{k,2}(2,1)}{\partial h_{T,2}(k)} &= \frac{\frac{\partial P(Y_T(2)>k|A_T(1))}{\partial h_{T,2}(k)}}{P(Y_T(2)>k|A_T(1))\left\{1-P(Y_T(2)>k|A_T(1))\right\}},
\end{aligned}$$

where

$$\begin{aligned}
\frac{\partial P(Y_T(j)>k|A_T(1))}{\partial \psi_{\mathrm{tx}}} &= \frac{\tau \frac{\partial P(Y_T(j)>k|A_T(0))}{\partial \psi_{\mathrm{tx}}}}{\left\{1+(\tau-1)P(Y_T(j)>k|A_T(0))\right\}^2}, \\
\frac{\partial P(Y_T(1)>k|A_T(1))}{\partial h_{T,1}(k)} &= \frac{\tau \frac{\partial P(Y_T(1)>k|A_T(0))}{\partial h_{T,1}(k)}}{\left\{1+(\tau-1)P(Y_T(1)>k|A_T(0))\right\}^2}, \\
\frac{\partial P(Y_T(2)>k|A_T(1))}{\partial h_{T,2}(k)} &= \frac{\tau \frac{\partial P(Y_T(2)>k|A_T(0))}{\partial h_{T,2}(k)}}{\left\{1+(\tau-1)P(Y_T(2)>k|A_T(0))\right\}^2},
\end{aligned}$$

for $j = 1, 2$ and $\mathrm{tx} = 0, 1, 2$.

We finally describe the calculation for the covariance matrix for the log of $SACE_{k,3}(2,1)$.

The log of $SACE_{k,3}(2,1)$ is

$$\begin{aligned}\log \mathrm{SACE}_{k,3}(2,1) &= \log P(Y_T(2) > k | A_T(0) \cup A_T(1)) - \log \{1 - P(Y_T(2) > k | A_T(0) \cup A_T(1))\} \\ &\quad - \log P(Y_T(1) > k | A_T(0) \cup A_T(1)) + \log \{1 - P(Y_T(1) > k | A_T(0) \cup A_T(1))\}.\end{aligned}$$

The derivatives of $\log \mathrm{SACE}_{k,3}(2,1)$ with respect to $\beta_0(0)$, $\beta_0(1)$, $\beta_0(2)$, $\beta(0)$, $\beta(1)$, and $\beta(2)$ are, respectively, given by

$$\begin{aligned}\frac{\partial \log SACE_{k,3}(2,1)}{\partial \beta_0(\mathrm{tx})} &= \frac{\partial \log SACE_{k,3}(2,1)}{\partial h_{T,\mathrm{tx}}(k)} \frac{\partial h_{T,\mathrm{tx}}(k)}{\partial \beta_0(\mathrm{tx})}, \\ \frac{\partial \log SACE_{k,3}(2,1)}{\partial \beta(\mathrm{tx})} &= \frac{\partial \log SACE_{k,3}(2,1)}{\partial h_{T,\mathrm{tx}}(k)} \frac{\partial h_{T,\mathrm{tx}}(k)}{\partial \beta(\mathrm{tx})}.\end{aligned}$$

The derivatives of $\log \mathrm{SACE}_{k,3}(2,1)$ with respect to ψ_0, ψ_1, ψ_2, $h_{T,0}(k)$, $h_{T,1}(k)$, and $h_{T,2}(k)$ are given by

$$\begin{aligned}\frac{\partial \log SACE_{k,3}(2,1)}{\partial \psi_0} &= \frac{\frac{\partial P(Y_T(2)>k|A_T(0)\cup A_T(1))}{\partial \psi_0}}{P(Y_T(2) > k|A_T(0) \cup A_T(1)) \{1 - P(Y_T(2) > k|A_T(0) \cup A_T(1))\}} \\ &\quad - \frac{\frac{\partial P(Y_T(1)>k|A_T(0)\cup A_T(1))}{\partial \psi_0}}{P(Y_T(1) > k|A_T(0) \cup A_T(1)) \{1 - P(Y_T(1) > k|A_T(0) \cup A_T(1))\}}, \\ \frac{\partial \log SACE_{k,3}(2,1)}{\partial \psi_1} &= \frac{\frac{\partial P(Y_T(2)>k|A_T(0)\cup A_T(1))}{\partial \psi_1}}{P(Y_T(2) > k|A_T(0) \cup A_T(1)) \{1 - P(Y_T(2) > k|A_T(0) \cup A_T(1))\}} \\ &\quad - \frac{\frac{\partial P(Y_T(1)>k|A_T(0)\cup A_T(1))}{\partial \psi_1}}{P(Y_T(1) > k|A_T(0) \cup A_T(1)) \{1 - P(Y_T(1) > k|A_T(0) \cup A_T(1))\}}, \\ \frac{\partial \log SACE_{k,3}(2,1)}{\partial \psi_2} &= \frac{\frac{\partial P(Y_T(2)>k|A_T(0)\cup A_T(1))}{\partial \psi_2}}{P(Y_T(2) > k|A_T(0) \cup A_T(1)) \{1 - P(Y_T(2) > k|A_T(0) \cup A_T(1))\}} \\ &\quad - \frac{\frac{\partial P(Y_T(1)>k|A_T(0)\cup A_T(1))}{\partial \psi_2}}{P(Y_T(1) > k|A_T(0) \cup A_T(1)) \{1 - P(Y_T(1) > k|A_T(0) \cup A_T(1))\}}, \\ \frac{\partial \log SACE_{k,3}(2,1)}{\partial h_{T,0}(k)} &= 0, \\ \frac{\partial \log SACE_{k,3}(2,1)}{\partial h_{T,1}(k)} &= - \frac{\frac{\partial P(Y_T(1)>k|A_T(0)\cup A_T(1))}{\partial h_{T,1}(k)}}{P(Y_T(1) > k|A_T(0) \cup A_T(1)) \{1 - P(Y_T(1) > k|A_T(0) \cup A_T(1))\}}, \\ \frac{\partial \log SACE_{k,3}(2,1)}{\partial h_{T,2}(k)} &= \frac{\frac{\partial P(Y_T(2)>k|A_T(0)\cup A_T(1))}{\partial h_{T,2}(k)}}{P(Y_T(2) > k|A_T(0) \cup A_T(1)) \{1 - P(Y_T(2) > k|A_T(0) \cup A_T(1))\}},\end{aligned}$$

where

$$\frac{\partial P(Y_T(1)>k|A_T(0)\cup A_T(1))}{\partial \psi_0} = \frac{1}{P(A_T(0)\cup A_T(1))}\left\{\frac{\partial P(Y_T(1)>k|A_T(0))}{\partial \psi_0}P(A_T(0)) + P(Y_T(1)>k|A_T(0))\frac{\partial P(A_T(0))}{\partial \psi_0}\right.$$
$$\left. + \frac{\partial P(Y_T(1)>k|A_T(1))}{\partial \psi_0}P(A_T(1)) + P(Y_T(1)>k|A_T(1))\frac{\partial P(A_T(1))}{\partial \psi_0}\right\}$$
$$- \frac{P(Y_T(1)>k|A_T(0))P(A_T(0)) + P(Y_T(1)>k|A_T(1))P(A_T(1))}{\{P(A_T(0)\cup A_T(1))\}^2}\frac{\partial P(A_T(0)\cup A_T(1))}{\partial \psi_0},$$

$$\frac{\partial P(Y_T(2)>k|A_T(0)\cup A_T(1))}{\partial \psi_0} = \frac{1}{P(A_T(0)\cup A_T(1))}\left\{\frac{\partial P(Y_T(2)>k|A_T(0))}{\partial \psi_0}P(A_T(0)) + P(Y_T(2)>k|A_T(0))\frac{\partial P(A_T(0))}{\partial \psi_0}\right.$$
$$\left. + \frac{\partial P(Y_T(2)>k|A_T(1))}{\partial \psi_0}P(A_T(1)) + P(Y_T(2)>k|A_T(1))\frac{\partial P(A_T(1))}{\partial \psi_0}\right\}$$
$$- \frac{P(Y_T(2)>k|A_T(0))P(A_T(0)) + P(Y_T(2)>k|A_T(1))P(A_T(1))}{\{P(A_T(0)\cup A_T(1))\}^2}\frac{\partial P(A_T(0)\cup A_T(1))}{\partial \psi_0},$$

$$\frac{\partial P(Y_T(1)>k|A_T(0)\cup A_T(1))}{\partial \psi_1} = \frac{1}{P(A_T(0)\cup A_T(1))}\left\{\frac{\partial P(Y_T(1)>k|A_T(0))}{\partial \psi_1}P(A_T(0)) + P(Y_T(1)>k|A_T(0))\frac{\partial P(A_T(0))}{\partial \psi_1}\right.$$
$$\left. + \frac{\partial P(Y_T(1)>k|A_T(1))}{\partial \psi_1}P(A_T(1)) + P(Y_T(1)>k|A_T(1))\frac{\partial P(A_T(1))}{\partial \psi_1}\right\}$$
$$- \frac{P(Y_T(1)>k|A_T(0))P(A_T(0)) + P(Y_T(1)>k|A_T(1))P(A_T(1))}{\{P(A_T(0)\cup A_T(1))\}^2}\frac{\partial P(A_T(0)\cup A_T(1))}{\partial \psi_1},$$

$$\frac{\partial P(Y_T(2)>k|A_T(0)\cup A_T(1))}{\partial \psi_1} = \frac{1}{P(A_T(0)\cup A_T(1))}\left\{\frac{\partial P(Y_T(2)>k|A_T(0))}{\partial \psi_1}P(A_T(0)) + P(Y_T(2)>k|A_T(0))\frac{\partial P(A_T(0))}{\partial \psi_1}\right.$$
$$\left. + \frac{\partial P(Y_T(2)>k|A_T(1))}{\partial \psi_1}P(A_T(1)) + P(Y_T(2)>k|A_T(1))\frac{\partial P(A_T(1))}{\partial \psi_1}\right\}$$
$$- \frac{P(Y_T(2)>k|A_T(0))P(A_T(0)) + P(Y_T(2)>k|A_T(1))P(A_T(1))}{\{P(A_T(0)\cup A_T(1))\}^2}\frac{\partial P(A_T(0)\cup A_T(1))}{\partial \psi_1},$$

$$\frac{\partial P(Y_T(1)>k|A_T(0)\cup A_T(1))}{\partial \psi_2} = \frac{1}{P(A_T(0)\cup A_T(1))}\left\{\frac{\partial P(Y_T(1)>k|A_T(0))}{\partial \psi_2}P(A_T(0)) + P(Y_T(1)>k|A_T(0))\frac{\partial P(A_T(0))}{\partial \psi_2}\right.$$
$$\left. + \frac{\partial P(Y_T(1)>k|A_T(1))}{\partial \psi_2}P(A_T(1)) + P(Y_T(1)>k|A_T(1))\frac{\partial P(A_T(1))}{\partial \psi_2}\right\}$$
$$- \frac{P(Y_T(1)>k|A_T(0))P(A_T(0)) + P(Y_T(1)>k|A_T(1))P(A_T(1))}{\{P(A_T(0)\cup A_T(1))\}^2}\frac{\partial P(A_T(0)\cup A_T(1))}{\partial \psi_2},$$

$$\frac{\partial P(Y_T(2)>k|A_T(0)\cup A_T(1))}{\partial \psi_2} = \frac{1}{P(A_T(0)\cup A_T(1))}\left\{\frac{\partial P(Y_T(2)>k|A_T(0))}{\partial \psi_2}P(A_T(0)) + P(Y_T(2)>k|A_T(0))\frac{\partial P(A_T(0))}{\partial \psi_2}\right.$$
$$\left. + \frac{\partial P(Y_T(2)>k|A_T(1))}{\partial \psi_2}P(A_T(1)) + P(Y_T(2)>k|A_T(1))\frac{\partial P(A_T(1))}{\partial \psi_2}\right\}$$
$$- \frac{P(Y_T(2)>k|A_T(0))P(A_T(0)) + P(Y_T(2)>k|A_T(1))P(A_T(1))}{\{P(A_T(0)\cup A_T(1))\}^2}\frac{\partial P(A_T(0)\cup A_T(1))}{\partial \psi_2},$$

$$\frac{\partial P(Y_T(1)>k|A_T(0)\cup A_T(1))}{\partial h_{T,1}(k)} = \frac{1}{P(A_T(0)\cup A_T(1))}\left\{\frac{\partial P(Y_T(1)>k|A_T(0))}{\partial h_{T,1}(k)}P(A_T(0)) + \frac{\partial P(Y_T(1)>k|A_T(1))}{\partial h_{T,1}(k)}P(A_T(1))\right\},$$

$$\frac{\partial P(Y_T(2)>k|A_T(0)\cup A_T(1))}{\partial h_{T,2}(k)} = \frac{1}{P(A_T(0)\cup A_T(1))}\left\{\frac{\partial P(Y_T(2)>k|A_T(0))}{\partial h_{T,2}(k)}P(A_T(0)) + \frac{\partial P(Y_T(2)>k|A_T(1))}{\partial h_{T,2}(k)}P(A_T(1))\right\},$$

with

$$\frac{\partial P(A_T(1))}{\partial \psi_0} = -E_w \left\{ \frac{\partial g_T(0,w)}{\partial \psi_0} p_T^*(2,w) + g_T(0,w) \frac{\partial p_T^*(2,w)}{\partial \psi_0} \right\} - \frac{\partial P(A_T(2))}{\partial \psi_0},$$

$$\frac{\partial P(A_T(2))}{\partial \psi_0} = E_w \left[-\frac{\partial q_T^*(w)}{\partial \psi_0} \{1 - g_T(0,w) - g_T(1,w) + p_T^*(1,w) g_T(0,w)\} \right.$$
$$\left. + (1 - q_T^*(w)) \left\{ -\frac{\partial g_T(0,w)}{\partial \psi_0} + \frac{\partial p_T^*(1,w)}{\partial \psi_0} g_T(0,w) + p_T^*(1,w) \frac{\partial g_T(0,w)}{\partial \psi_0} \right\} \right],$$

$$\frac{\partial P(A_T(0) \cup A_T(1))}{\partial \psi_0} = \frac{\partial P(A_T(0))}{\partial \psi_0} + \frac{\partial P(A_T(1))}{\partial \psi_0},$$

$$\frac{\partial P(A_T(1))}{\partial \psi_1} = -E_w \left[\frac{\partial q_T^*(w)}{\partial \psi_1} \{1 - g_T(0,w) - g_T(1,w) + p_T^*(1,w) g_T(0,w)\} \right.$$
$$\left. + (1 - q_T^*(w)) \left\{ \frac{\partial g_T(1,w)}{\partial \psi_1} - g_T(0,w) \frac{\partial p_T^*(1,w)}{\partial \psi_1} \right\} \right],$$

$$\frac{\partial P(A_T(0) \cup A_T(1))}{\partial \psi_1} = \frac{\partial P(A_T(0))}{\partial \psi_1} + \frac{\partial P(A_T(1))}{\partial \psi_1},$$

$$\frac{\partial P(A_T(1))}{\partial \psi_2} = E_w \left\{ \frac{\partial g_T(2,w)}{\partial \psi_2} - g_T(0,w) \frac{\partial p_T^*(2,w)}{\partial \psi_2} \right\} - \frac{\partial P(A_T(2))}{\partial \psi_2},$$

$$\frac{\partial P(A_T(2))}{\partial \psi_2} = -E_w \left[\frac{\partial q_T^*(w)}{\partial \psi_2} \{1 - g_T(0,w) - g_T(1,w) + p_T^*(1,w) g_T(0,w)\} \right],$$

$$\frac{\partial P(A_T(0) \cup A_T(1))}{\partial \psi_2} = \frac{\partial P(A_T(0))}{\partial \psi_2} + \frac{\partial P(A_T(1))}{\partial \psi_2},$$

Numerical derivatives

Numerical derivatives are computed as local slopes based on small perturbations of parameter estimates (Press et al., 1992). Let $f(\omega)$ be the left hand side of (25), (26), or (32) where $\omega = (\psi_0, \psi_1, \psi_2, h_{T,1}(k), h_{T,2}(k))$, and let $\omega^{(-i)}$ be a vector with the ith element of ω removed. The ith element of ν, $\omega^{(i)}$, is perturbed as follows:

$$\begin{aligned} \delta_i &= c_1 \left(\frac{\omega^{(i)}}{\omega^{(i)} + c_2} \right), \\ \omega_1^{(i)} &= \begin{cases} \omega^{(i)} - \delta_i, & \delta_i \geq c_2; \\ \omega^{(i)} - c_2, & \delta_i < c_2, \end{cases} \quad \omega_2^{(i)} = \begin{cases} \omega^{(i)} + \delta_i, & \delta_i \geq c_2; \\ \omega^{(i)} + c_2, & \delta_i < c_2, \end{cases} \end{aligned}$$

where c_1 and c_2 are small (here, we set $c_1 = 0.01$ and $c_2 = 0.001$). Using the current

estimates of the parameters ω, the derivative of $\omega^{(i)}$ is

$$\frac{\partial P(Y_T(\text{tx}) > k | A_T(0))}{\partial \omega^{(i)}} = \frac{f(\omega^{(-i)}, \omega_2^{(i)}) - f(\omega^{(-i)}, \omega_1^{(i)})}{2\delta_i},$$

for $\text{tx} = 0, 1, 2$.

References

Cheng, J. (2009) Estimation and inference for the causal effect of receiving treatment on a multinomial outcome, *Biometrics*, 65, 96-103.

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

Egleston, B. L., Scharfstein, D. O., Freeman, E. R. and West, S. K. (2007). Causal inference for non-mortality outcomes in the presence of death. *Biostatistics*, **8**, 526-545.

Egleston, B. L., Scharfstein, D. O., and MacKenzie, E. (2009). On estimation of the survivor average causal effect in observational studies when important confounders are missing due to death. *Biometrics* **65**, 497–504.

Frangakis, C. E. and Rubin, D. B. (2002). Principal stratification in causal inference. *Biometrics* **58**, 21–29.

Gilbert, P. B., Ronald, J. B., and Hudgens, M. G. (2003). Sensitivity analysis for the assessment of causal vaccine effects on viral load in HIV vaccine trials. *Biometrics* **59**, 531–541.

Goldberg, R. M., Sargent, D. J., Morton, R. F., Fuchs, C. S., Ramanathan, R. K., Williamson, S. K., Findlay, B. P., Pitot, H. C, and Albets, S. R (2004). A randomized controlled trial of fluorouracil plus leucovorin, irinotecan, and oxaliplatin combinations in patients with previously untreated metastatic colorectal cancer. *Journal of Clinical Oncology* **22**, 23–30.

Harel, O. and Schafer, J.L. (2009) Partial and latent ignorability in missing-data problems. *Biometrika*, **96**, 37–50.

Hayden, D., Pauler, D. K., and Schoenfeld, D. (2005). An estimator for treatment comparisons among survivors in randomized trials. *Biometrics* **61**, 305–310.

Hogan, J. W. and Laird, N. M. (1997). Mixture models for the joint distribution of repeated measures and event times. *Statistics in Medicine* **16**, 239–257.

Holland, P. (1986). Statistics and causal inference. *Journal of the American Statistical Association* **81**, 945–961.

Imai, K. (2008) Sharp bounds on the causal effects in randomized experiments with truncation-by-death. *Statistics and Probability Letters*. **78**, 144-149.

Jin, H. and Rubin, D. (2008) Principal stratification for causal inference with extended partial compliance. *Journal of the American Statistical Association*. **103**, 101-110.

Kurland, B. F. and Heagerty, P. J.(2005). Directly parameterized regression conditioning on being alive: analysis of longitudinal data truncated by deaths. *Biostatistics* **6**, 241–258.

Lee, K. and Daniels, M. (2007). A class of Markov models for longitudinal ordinal data. *Biometrics* **63**, 1060–1067.

Lee, K. and Daniels, M. (2008). Marginalized Models for Longitudinal Ordinal Data with Application to Quality of Life Studies. *Statistics in Medicine*, **27**, 4359-4380.

McLeod, H.L., Sargent, D.J., Marsh, S., Fuchs, C., Ramanathan, R.K., Williamson, S., Findlay, B., Thibodeau, S., Petersen, G., and Goldberg, R. (2003) Pharmaco-

genetic analysis of systemic toxicity and response after 5-fluorouracil (5FU)/CPT-11, 5FU/oxaliplatin (oxal), or CPT-11/oxal therapy for advanced colorectal cancer (CRC): Results from an intergroup trial. *Proc Am Soc Clin Oncol*, 22: abstract 1013.

Pauler, D. K., McCoy, S., and Moinpour, C. (2003). Pattern mixture models for longitudinal quality of life studies in advanced stage disease. *Statistics in Medicine* **22**, 795–809.

Press, W. H., Teukolsky, S. A., Vetterling, W. T., and Flannery, B. P. (1992). *Numerical Recipes in C: The Art of Scienific Computing*. Cambridge University Press, Cambridge, UK, second edition.

Robins, J. M., Rotnitzky, A., and Zhao, L. P. (1995). Analysis of semiparametric regression models for repeated outcomes in the presence of missing data. *Journal of the American Statistical Association*. **90**, 106–121.

Rotnitzky, A, Robins, J. M., and Scharfstein D. O. (1998) Semiparametric regression for repeated outcomes with nonignorable nonresponse. *Journal of the American Statistical Association*. **93**, 1321–1339.

Roy, J., Hogan, J.W., and Marcus, B.H. (2008). Principal stratification with predictors of compliance for randomized trials with 2 active treatments. *Biostatistics* **9**, 277–289.

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

Rubin, D. B. (1978). Bayesian inference for causal effects. *Annals of Statistics* **6**, 34–58.

Rubin, D. B. (2000). Comment on "Causal inference without counterfactuals", by A. P. Dawid. *Journal of the American Statistical Association* **6**, 34–58.

Rubin, D. B. (2006). Causal inference through potential outcomes and principal stratification: Application to studies with 'censoring' due to death. *Statistical Science* **21**, 299–309.

Saltz, L.B., Cox, J.V., Blanke, C., Rosen, L.S., Fehrenbacher, L., Moore, M.J., Maroun, J.A., Ackland, S.P., Locker, P.K., Pirotta, N., Elfring, G.L., Miller, L.L. for the Irinotecan Study Group. (2000) Irinotecan plus fluorouracil and leucovorin for metastatic colorectal cancer, *New England Journal of Medicine*, **343**, 905-914.

Sanoff, H.K., Sargent, D.J., Campbell, M.E., Morton, R.F., Fuchs, C.S., Ramanathan, R.K., Williamson, S.K., Findlay, B.P., Pitot, H.C., and Goldberg, R.M. (2008). Five-Year Data and Prognostic Factor Analysis of Oxaliplatin and Irinotecan Combinations for Advanced Colorectal Cancer: N9741, *Journal of Clinical Oncology*, **26**, 5721-5727.

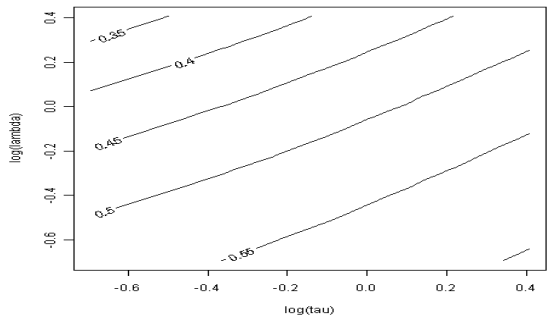
Scharfstein, D. O., Rotnitzky A., Robins J. M. (1999). Adjusting for Non-ignorable Drop-out Using Semiparametric Non-response Models (with Discussion). *Journal of the American Statistical Association.* **94**, 1096–1146.

Shepherd, B.E, Gilbert, P. B., and Mehrotra, D.V. (2007). Eliciting a counterfactual sensitivity parameter. *American Statistician*, **61**, 56–63.

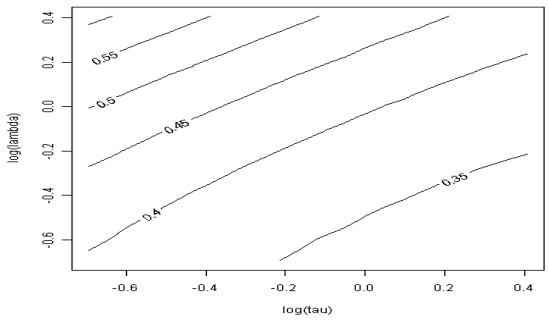
Small, D. and Tan, Z. (2008). A stochastic monotonicity assumption for the instrumental variables method. *Working paper.*

Zhang, J. L. and Rubin, D. B. (2003). Estimation of causal effects via principal stratification when some outcomes are truncated by "death". *Journal of Educational and*

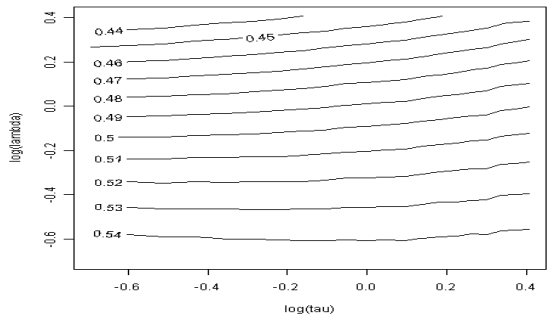
Behavioral Statistics **28**, 353–368.



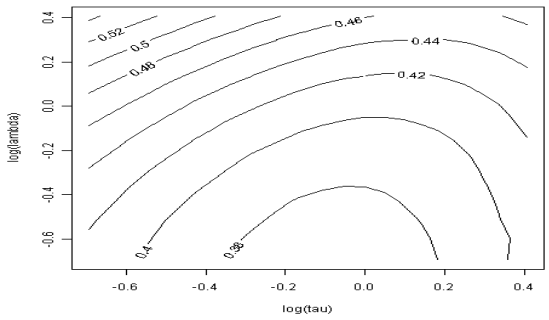
(a) $\log SACE_{K-1,1}(1,0)$ with $(\rho,\nu)=(0,0)$



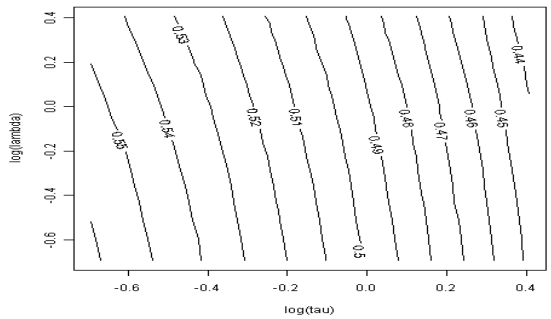
(b) p-values with $(\rho,\nu)=(0,0)$



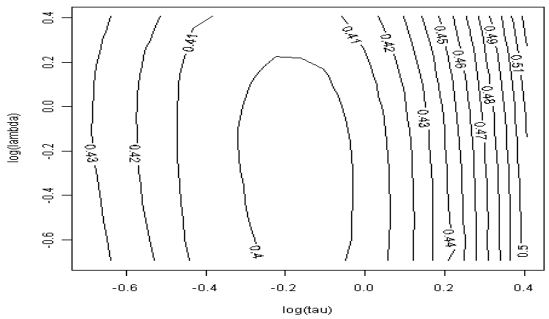
(c) $\log SACE_{K-1,1}(1,0)$ with $(\rho,\nu)=(0.5,0.5)$



(d) p-values with $(\rho,\nu)=(0.5,0.5)$

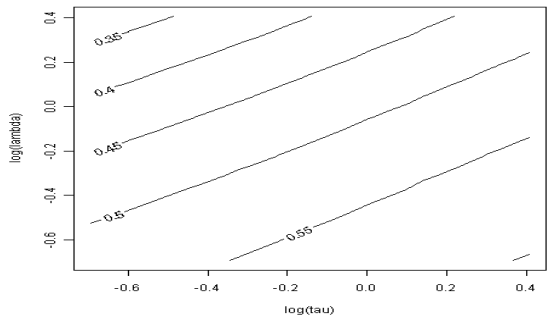


(e) $\log SACE_{K-1,1}(1,0)$ with $(\rho,\nu)=(1,1)$

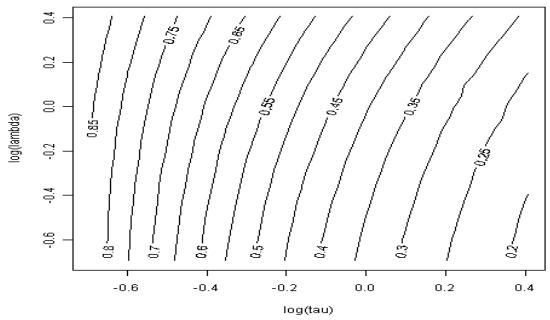


(f) p-values with $(\rho,\nu)=(1,1)$

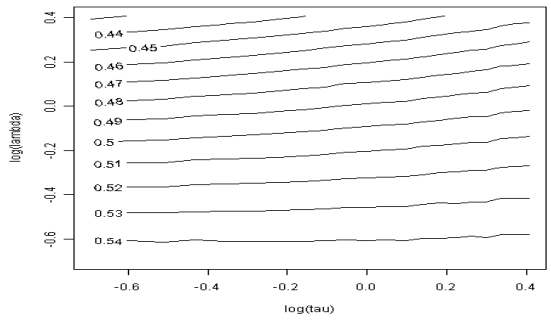
Figure 1. Maximum likelihood estimates of $\log SACE_{K-1,1}(1,0)$ with $\rho,\nu \in \{0,0.5,1.0\}$, respectively, and corresponding p-values for $\log SACE_{K-1,1}(1,0)$.



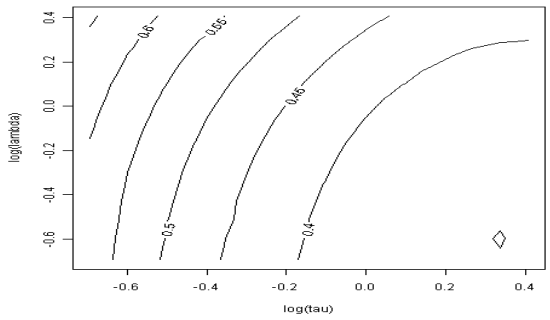
(a) $\log SACE_{K-1,3}(1,0)$ with $(\rho,\nu)=(0,0)$



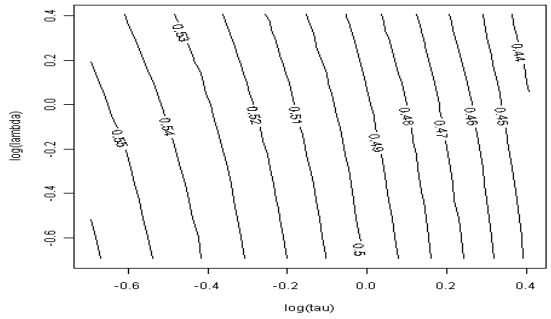
(b) p-values with $(\rho,\nu)=(0,0)$



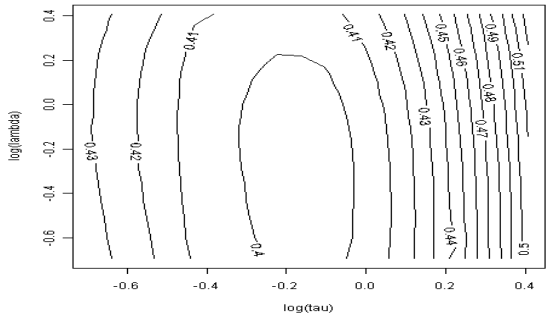
(c) $\log SACE_{K-1,3}(1,0)$ with $(\rho,\nu)=(0.5,0.5)$



(d) p-values with $(\rho,\nu)=(0.5,0.5)$

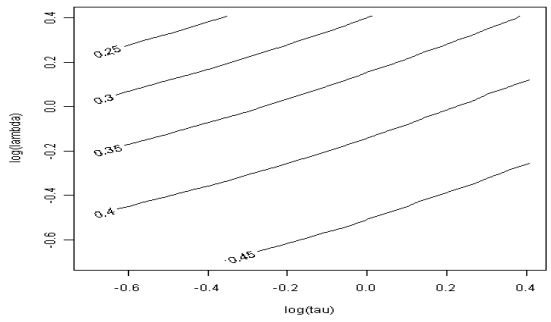


(e) $\log SACE_{K-1,3}(1,0)$ with $(\rho,\nu)=(1,1)$

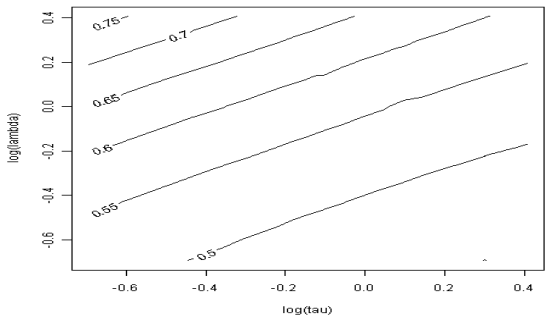


(f) p-values with $(\rho,\nu)=(1,1)$

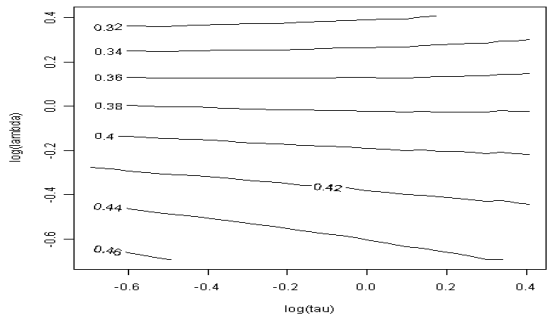
Figure 2. Maximum likelihood estimates of $\log SACE_{K-1,1}(1,0)$ with $\rho,\nu \in \{0,0.5,1.0\}$, respectively, and corresponding p-values for $\log SACE_{K-1,3}(1,0)$.



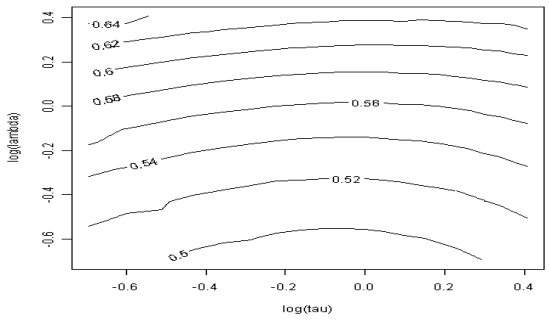
(a) $\log SACE_{K-1,1}(2,0)$ with $(\rho,\nu)=(0,0)$



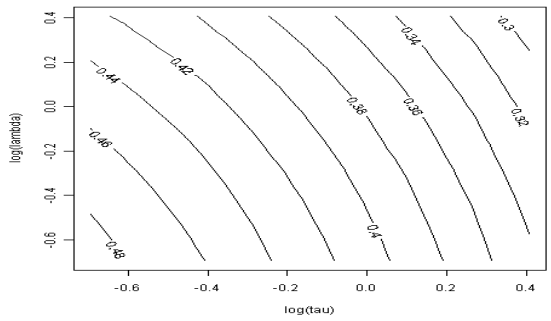
(b) p-values with $(\rho,\nu)=(0,0)$



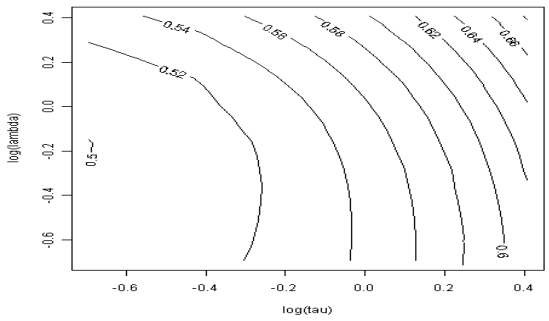
(c) $\log SACE_{K-1,1}(2,0)$ with $(\rho,\nu)=(0.5,0.5)$



(d) p-values with $(\rho,\nu)=(0.5,0.5)$

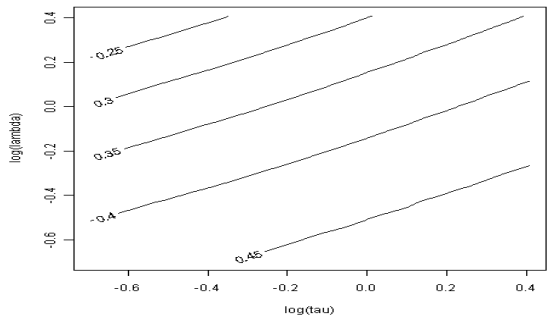


(e) $\log SACE_{K-1,1}(2,0)$ with $(\rho,\nu)=(1,1)$

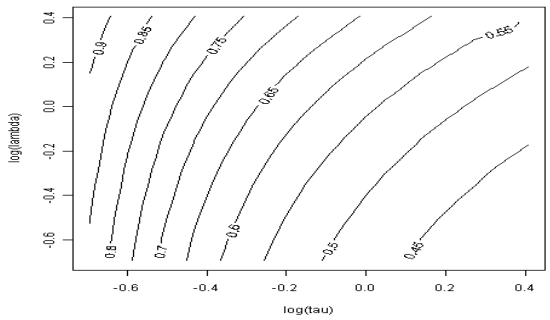


(f) p-values with $(\rho,\nu)=(1,1)$

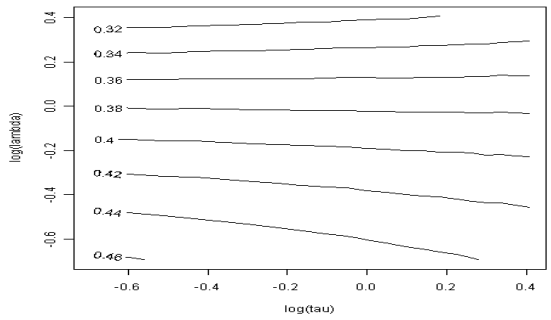
Figure 3. Maximum likelihood estimates of $\log SACE_{K-1,1}(2,0)$ with $\rho,\nu \in \{0,0.5,1.0\}$, respectively, and corresponding p-values for $\log SACE_{K-1,1}(2,0)$.



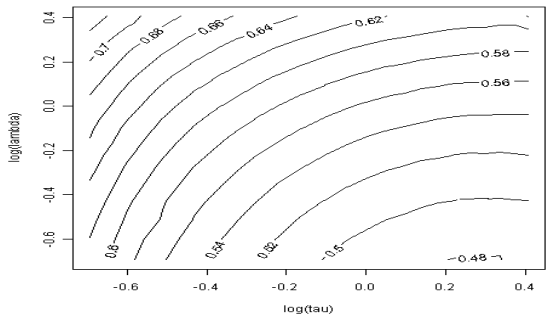
(a) $\log SACE_{K-1,3}(2,0)$ with $(\rho,\nu)=(0,0)$



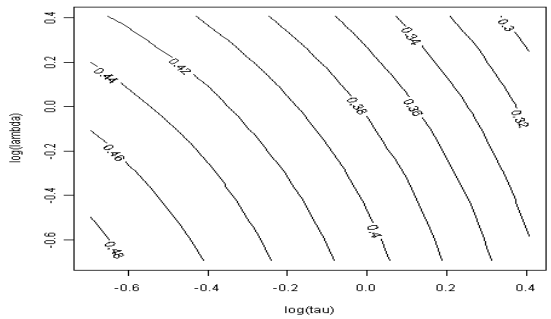
(b) p-values with $(\rho,\nu)=(0,0)$



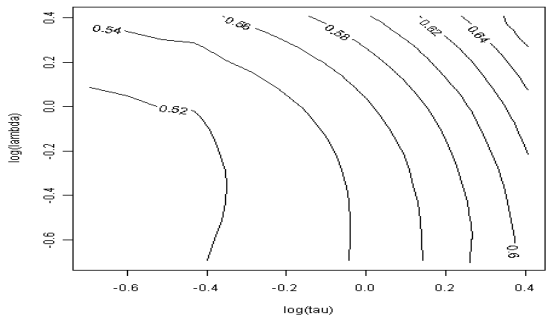
(c) $\log SACE_{K-1,3}(2,0)$ with $(\rho,\nu)=(0.5,0.5)$



(d) p-values with $(\rho,\nu)=(0.5,0.5)$

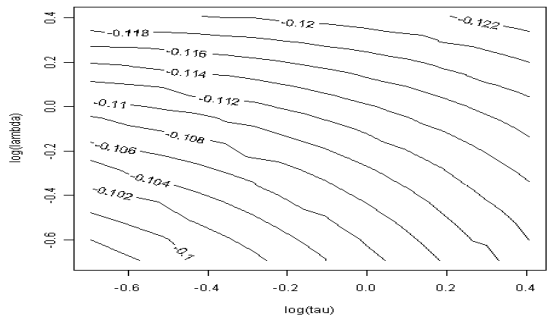


(e) $\log SACE_{K-1,3}(2,0)$ with $(\rho,\nu)=(1,1)$

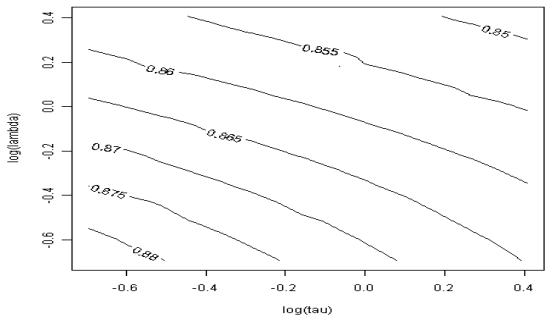


(f) p-values with $(\rho,\nu)=(1,1)$

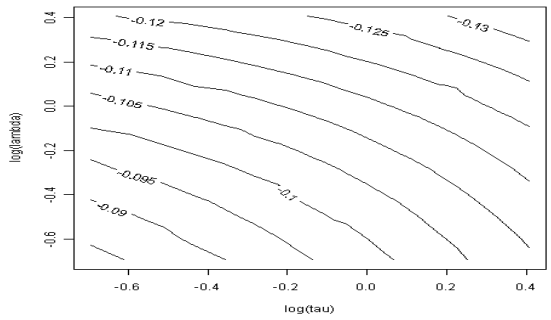
Figure 4. Maximum likelihood estimates of $\log SACE_{K-1,3}(2,0)$ with $\rho,\nu \in \{0,0.5,1.0\}$, respectively, and corresponding p-values for $\log SACE_{K-1,3}(2,0)$.



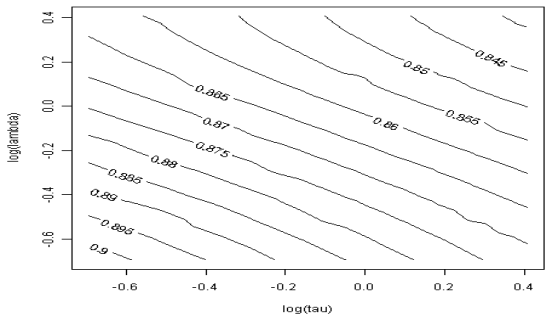
(a) $\log SACE_{K-1,1}(2,1)$ with $(\rho,\nu)=(0,0)$



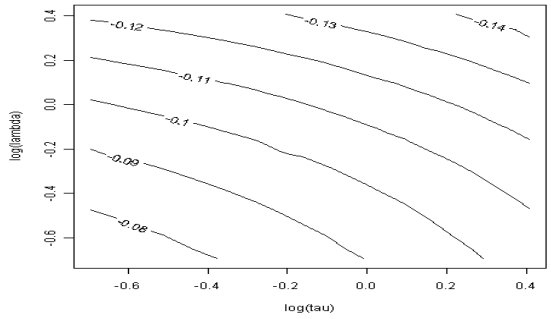
(b) p-values with $(\rho,\nu)=(0,0)$



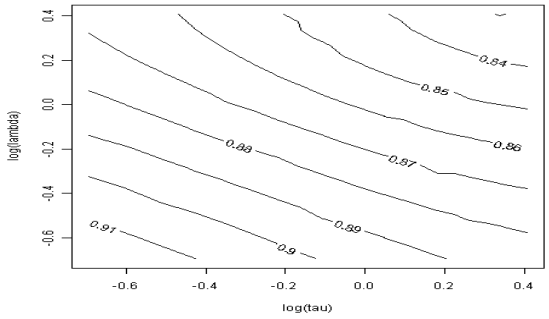
(c) $\log SACE_{K-1,1}(2,1)$ with $(\rho,\nu)=(0.5,0.5)$



(d) p-values with $(\rho,\nu)=(0.5,0.5)$

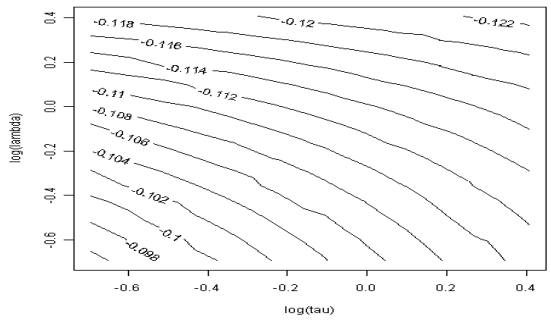


(e) $\log SACE_{K-1,1}(2,1)$ with $(\rho,\nu)=(1,1)$

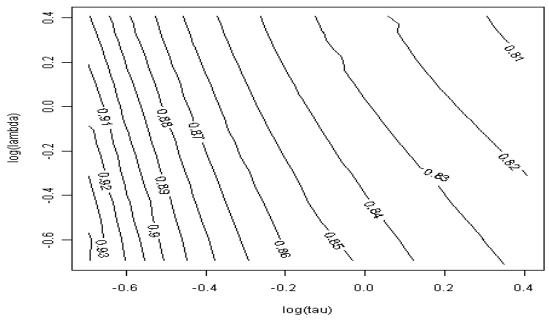


(f) p-values with $(\rho,\nu)=(1,1)$

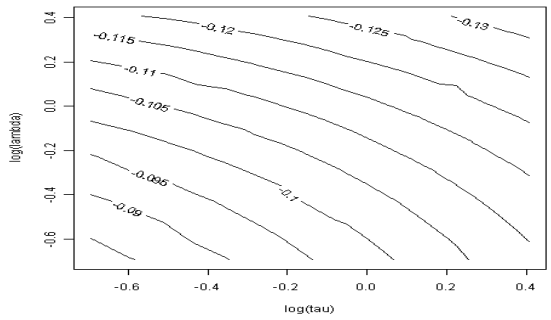
Figure 5. Maximum likelihood estimates of $\log SACE_{K-1,1}(2,1)$ with $\rho,\nu \in \{0,0.5,1.0\}$, respectively, and corresponding p-values for $\log SACE_{K-1,1}(2,1)$.



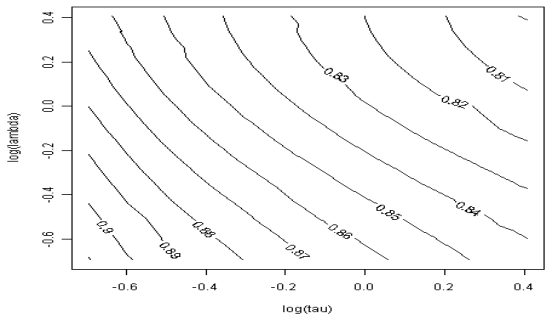
(a) $\log SACE_{K-1,3}(2,1)$ with $(\rho,\nu)=(0,0)$



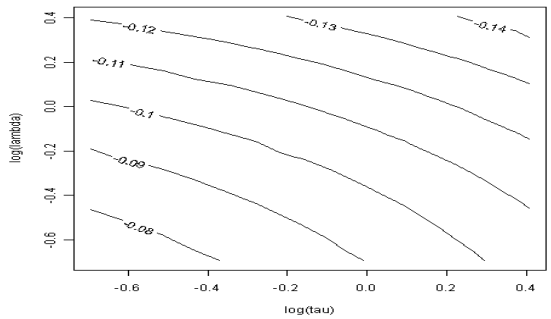
(b) p-values with $(\rho,\nu)=(0,0)$



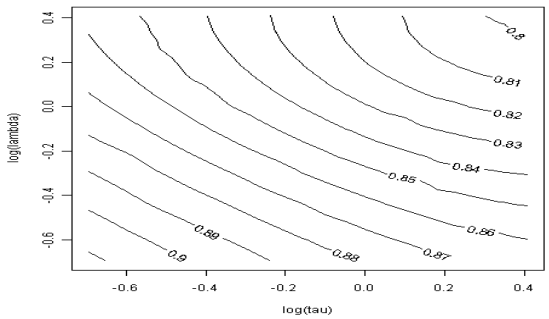
(c) $\log SACE_{K-1,3}(2,1)$ with $(\rho,\nu)=(0.5,0.5)$



(d) p-values with $(\rho,\nu)=(0.5,0.5)$



(e) $\log SACE_{K-1,3}(2,1)$ with $(\rho,\nu)=(1,1)$



(f) p-values with $(\rho,\nu)=(1,1)$

Figure 6. Maximum likelihood estimates of $\log SACE_{K-1,3}(2,1)$ with $\rho,\nu \in \{0,0.5,1.0\}$, respectively, and corresponding p-values for $\log SACE_{K-1,3}(2,1)$.

Table 1

Breakdown of subjects who were alive by treatment group at the time of study completion. Proportions are in parentheses

	Alive	Prog/Death	Total
IFL	86 (0.366)	149 (0.634)	235
IROX	105 (0.439)	134 (0.561)	239
FOLFOX	129 (0.554)	104 (0.446)	233
Total	320	387	707

Table 2

Maximum likelihood estimates of ψ_{tx}. Standard errors are in parentheses. AGE=$I\{age < 70\}$; PS=$I\{$performance status $\in (2,3,4)\}$; NOSITE1 = $I\{\#$ of sites $= 2\}$; NOSITE2 = $I\{\#$ of sites $> 2\}$.

	IFL	IROX	FOLFOX
Intercept	0.506 (0.230)	1.029 (0.247)	1.318 (0.275)
AGE	−0.391 (0.333)	−0.011 (0.330)	−0.288 (0.353)
PS	−2.413* (1.064)	−0.909 (0.602)	−1.526* (0.661)
NOSITE1	−0.037 (0.317)	−0.992* (0.328)	−0.547 (0.349)
NOSITE2	−0.669 (0.342)	−0.599 (0.342)	−0.907* (0.364)

* indicates significance with 95% confidence level.

Table 3

Maximum Likelihood Estimates of $h_{T,tx}(K-1)$) (standard errors) where $Tx \in \{IFL,\ IROX,\ FOLFOX\}$.

Tx	IFL	IROX	FOLFOX
$h_{T,\mathrm{tx}}(K-1)$	0.073	0.114	0.103
	(0.028)	(0.046)	(0.041)