

1-21-2011

# Evaluating Markers for Treatment Selection Based on Survival Time

Xiao Song
*University of Geogia*, xsong@uga.edu

Xiao-Hua Zhou
*University of Washington*, azhou@u.washington.edu

# Evaluating Markers for Treatment Selection Based on Survival Time

Xiao Song[1,*,†] and Xiao-Hua Zhou[2]

[1] *Department of Epidemiology and Biostatistics, College of Public Health, University of Georgia, U.S.A.*

[2] *Puget Sound Health Care System and University of Washington, Seattle, U.S.A.*

SUMMARY

For many medical conditions several treatment options may be available for treating patients. We consider evaluating markers based on a simple treatment selection policy that incorporates information on the patient's marker value exceeding a threshold. For example, colon cancer patients may be treated by surgery alone or surgery plus chemotherapy. The c-myc gene expression level may be used as a biomarker for treatment selection. Although traditional regression methods may assess the effect of the marker and treatment on outcomes, it is appealing to quantify more directly the potential impact on the population of using the marker to select treatment. A useful tool is the selection impact (SI) curve proposed by Song and Pepe for binary outcomes [1]. However, the current SI method does not deal with continuous outcomes, nor does it allow to adjust for other covariates that are important for treatment selection. In this

* Correspondence to: Xiao Song, Department of Epidemiology and Biostatistics, College of Public Health, University of Georgia, Paul Coverdell Center, Room 129c, Athens, GA 30602, U.S.A.

† Email: xsong@uga.edu

Contract/grant sponsor: University of Georgia (Song) and the VA Clinical R&D (Zhou and Song).

paper, we extend the SI curve for general outcomes, with a specific focus on survival time. We further propose the covariate specific SI curve to incorporate covariate information in treatment selection. Nonparametric and semiparametric estimators are developed accordingly. We show that the proposed estimators are consistent and asymptotically normal. The performance is illustrated by simulation studies and through an application to data from a cancer clinical trial.

KEY WORDS: biomarker; restricted survival time; selection impact curve; statistical interaction.

## 1. INTRODUCTION

Selecting an appropriate treatment for patients is important when several treatment options are available. Treatment selection may be facilitated by evaluating clinical characteristics or biomarker measurements of patients at diagnosis. The rapidly expanding biotechnologies, including gene expression arrays and imaging modalities, show promises in providing useful biomarkers that may be used for selection of the optimal treatment of disease [2]. For example, patients with colon cancer can be treated by surgery alone or surgery plus chemotherapy. Surgery alone is less expensive and has less side effects than surgery plus chemotherapy, but it may be less effective as well, at least for some patients. It is desirable to identify the patients who may benefit more from surgery based on biomarkers. A possible useful biomarker is the c-myc gene, which is overexpressed in approximately 70% of human colonic tumors [3]. Based on a study conducted by the Eastern Cooperative Oncoloty Group (ECOG), Augenlicht et al. suggested that the c-myc gene may be of clinical prognostic importance in patients with colon cancer [4]. Using a subset of the cases from this

clinical trial, Li and Ryan indicated that there is an interaction between the c-myc gene expression level and the two treatments on overall survival and disease progression free survival [5]. Using the same dataset, we estimated the overall survival and the disease progression free survival by treatment and c-myc gene expression level. Figure 1 shows the Kaplan-Meier estimates of the disease progression survival for the four combinations of the two treatments and whether the c-myc gene expression level exceeding 1.05, the 25% sample percentile. Although surgery plus chemotherapy seems prolonging disease progression free survival better than surgery alone for patients with c-myc level $> 1.05$, it is not that clear which treatment is better for patients with c-myc level $\leq 1.05$. The Kaplan-Meier estimates for the overall survival show a similar pattern (not shown). This motivated us to assess using the c-myc gene expression level for treatment selection for colon cancer patients.

In clinical protocols, treatment selection is often based on whether a marker value exceeds a threshold. Some common examples include serum creatine $> 1.3$ mg/dL, cholesterol $> 200$ mg/dL, and serum PSA $> 4.0$ ng/mL. As an analogy, we consider selection of surgery alone versus chemotherapy plus surgery based on c-myc level exceeding a threshold. To evaluate such policies, Song and Pepe proposed a graphical tool, the selection impact (SI) curve, for selection between two treatments based on binary response rate using a biomarker [1]. Let $A = 0$ denote one treatment and $A = 1$ denote the other. The outcome $T$ is dichotomous denoting success ($T = 1$) or failure ($T = 0$) in curing a disease. Let $Y$ be a continuous biomarker and larger values of $Y$ are potentially associated with better performance of treatment 1 versus 0 ($Y$ can be recoded if necessary to achieve this ). Consider the following treatment policy that determines which treatment the patient receives based on a patient's marker

measure $Y$ exceeding a threshold:

$$\begin{aligned} &\text{if } Y > c \quad , \quad \text{select treatment } A = 1; \\ &\text{if } Y \leq c \quad , \quad \text{select treatment } A = 0, \end{aligned} \tag{1}$$

This implies that $Y$ is known before treatment selection and generally a baseline measurement. The population response (success) rate of the outcome corresponding to this policy is

$$\theta(v) \quad = \quad \Pr\{T = 1 | \text{ treatment policy (1) } \}$$

where $v = \Pr[Y \leq c]$ is the proportion of subjects with the marker value below $c$ and hence assigned to treatment 0. That is, $\theta$ is the proportion of subjects in the population who respond if the treatment policy in effect is to assign a subject to treatment 1 if his marker value exceeds $c$ but to assign him to treatment 0 otherwise. Observe that when $c = -\infty$ or equivalently $v = 0$, the policy is that all patients receive treatment 1 and none receive 0, while for $c = \infty$ or equivalently $v = 1$ all patients receive treatment 0. As $c$ increases from $-\infty$ to $\infty$, the proportion of subjects assigned to treatment 0 increases from 0 to 1. There are two reasons for defining $\theta$ as a function of quantile $v$. First, in evaluating a treatment policy of this sort, it will be important to know the fractions of patients potentially assigned to treatment $A = 0$ versus $A = 1$ by the policy, $1 - v$ and $v$, respectively. Second, the display on this scale allows one to compare policies based on different markers. In particular, even if the markers are measured in different units, we can still compare the treatment policies at a common percentile $v$, the larger $\theta(v)$ the better. The same idea has been adopted for the receiver operative curve (ROC), which is widely used in evaluation of

diagnostic tests [6, 7]. When the two treatments are comparable in all other aspects including cost and side effects, then the optimal threshold maximizes the SI curve $\theta(v)$ such that the overall success rate achieves the maximum. As illustrated in Figure 2, based on the biomarker $Y_1$, the best criteria would be to assign 40% patients to treatment 0 whose biomarker value is less than the 40% percentile, while $Y_3$ indicates it would be best to assign all patients to treatment 1. Based on the biomarker $Y_2$, for any $v$ between 0 and 0.4, the success rate would be the same if we assign patients to treatment 0 whose biomarker value falls below the $v$th percentile. However, if treatment 1 is more invasive, more expensive or has more side effects, we may want to assign 40% patients to treatment 0 such that the overall success rate achieves the maximum. Comparing the three biomarkers, it is obviously that policy based on $Y_1$ achieves the best success rate.

The SI curve is essentially a curve of the population response rate versus the percentile threshold. Compared to traditional regression models, this approach has the advantage of quantifying more directly the potential impact on the population of using the biomarker to select treatment. Specifically, we may choose an optimal threshold based on the the SI curve. However, there is a need for further improvement for wider applicability. First, the current SI curve methodology considers only binary outcomes. It cannot be applied to continuous outcomes such as survival time, which is frequently encountered in practice. For example, in the ECOG study, the outcome of interest is the overall survival time and the disease progression free survival of the colon cancer patients [5]. It may sometimes be desirable to select treatment based on a utility measure that incorporates notions of cost and benefit when those factors differ for the two treatments. Second, the current SI curve does not allow to adjust for

other covariates, which may contain additional important information for treatment selection. For colon cancer patients, the c-myc gene may be indicative for which patients are likely to benefit more from surgery alone. The effect of c-myc gene may be further impacted by covariates like gender and stage of cancer. For example, the optimal threshold for c-myc gene expression level may be different for patients with different covariate values. For example, Figures 3(a) and (b) show the Kaplan-Meier estimates of disease progression free survival by treatment and c-myc gene expression level above or below 1.05 for females and males separately. For either females or males with c-myc level $> 1.05$, surgery plus chemotherapy seems better than surgery alone in terms of prolonging disease progression free survival. However, for patients with c-myc level $\leq 1.05$, females seem benefited more from surgery plus chemothrapy while it is not clear which treatment is better for males. In addition, the SI curve methodology was not applied to any real dataset in [1]. It is of great interest to demonstrate this method in real applications such as the ECOG study.

To overcome the limitations of the current SI curve methodology, we generalize the current SI curve definition in two steps. First, we propose the SI curve for general outcomes for evaluating markers on treatment selection; Second, we extend the SI curve to adjust for covariates. In this paper, we focus specifically on survival time, which is the outcome of interest in the ECOG data and is more challenging compared to discrete and continuous outcomes without censoring. The same technique can be easily adapted to the latter with only minor modifications. We propose the SI curve based on the mean restricted survival time up to a given time $L$. The reason of using mean restricted survival time is to avoid the infeasibility of estimating the mean unrestricted survival time when censoring exists, that is, the mean unrestricted

survival time may not be estimated if the largest observed survival time is censored without some tail correction on the estimated survival function [8]. The technique of restricting survival time has been used previously in estimating the mean lifetime and quality-adjusted lifetime (see [9, 10] and the references therein). The restricted survival time has also been widely used in practice, for example, in cancer statistics, five year survival has been commonly used. Due to the existence of censoring, the inference is more challenging than the binary case. A nonparametric estimator is proposed to estimate the SI curve with no model assumptions on survival time. To adjust for covariate effects, we further propose the covariate specific SI curve and develop semiparametric estimators based on the proportional hazards model. Asymptotic properties of the estimators are derived using empirical process and U-process theories. The approach can be adapted to uncensored continuous outcomes with some modifications.

This paper is organized as follows. In Section 2, we define the SI curve for survival time, develop the nonparametric estimator and derive its asymptotic properties. In Section 3, we define the covariate specific SI curve. We further derive the semiparametric estimators and their asymptotic properties. The finite sample performance of these estimators is assessed via simulation studies in Section 4 and illustrated by application to the ECOG data in Section 5. The paper concludes with discussions in Section 6.

## 2. SELECTION IMPACT CURVE

### 2.1. *Definition*

We extend the SI curve to a general outcome $W$, that is, $W$ can be either discrete or continuous. We define the SI curve as

$$\theta(v) \quad = \quad E\left\{W \,|\text{treatment policy } (1)\right\}.$$

Here $v = \Pr[Y \leq c]$ is the proportion of patients assigned to treatment 0 under policy (1). The SI curve considered in [1] is a special case of (2) when $W$ is dichotomous. In this paper, we focus on SI curve for survival time hereafter.

Using the notion of potential responses [11–13], for $a = 0, 1$, we define $T^a$ as the survival time if a patient receives treatment $A = a$. It is impossible to observe $T^0$ and $T^1$ on the same patient; instead, we can only observe $T = AT^1 + (1 - A)T^0$. Let $T^{*a} = \min(T^a, L)$, be the corresponding restricted survival time by time $L$. Similarly, $T^{*0}$ and $T^{*1}$ can not be observed on the same patient and only $T^* = AT^{*1} + (1-A)T^{*0}$ is observable. Both $T$ and $T^*$ are subject to censoring. We define the SI curve as the mean restricted survival time corresponding to policy (1), that is,

$$\theta(v) = E\left\{T^* \,|\text{treatment policy } (1)\right\}.$$

Noting that $c$ is the $v$-th percentile $y_v$ of the biomarker $Y$ and

$$\begin{aligned} E\left\{T^* \,|\text{treatment policy } (1)\right\} \quad &= \quad P(Y > y_v)E\left\{T^* \,|\text{treatment policy } (1) \text{ and} Y > y_v\right\} \\ &\quad + P(Y \leq y_v)E\left\{T^* \,|\text{treatment policy } (1) \text{ and } Y \leq y_v\right\}, \end{aligned}$$

under the common assumption that the potential responses $(T^{*1}, T^{*0})$ are independent

from the treatment assignment $A$, it follows that

$$\theta(v) = (1-v)E\left\{T^{*1}\,|Y > y_v\right\} + vE\left\{T^{*0}\,|Y \le y_v\right\} \tag{2}$$

$$\tag{3}$$

When the two treatments are comparable, the optimal percentile $v_{opt}$ maximizes $\theta(v)$ and the optimal threshold $c_{opt} = F^{-1}(v_{opt})$. Otherwise, a utility measure incorporating the survival time, cost and side effects can be used. We illustrate an application of a simple utility measure in Section 5.

### 2.2. *Estimation*

We now consider estimation of $\theta(v)$ using data obtained from a randomized trial where the failure time is subject to censoring. Let $C$ denote the censoring time. The observed survival data consist of $X = \min(T, C)$ and $\Delta = I(T \le C)$. Suppose the observations $(X_i, \Delta_i, Y_i, A_i)$, $i = 1, \ldots, n$, are an i.i.d. sample from the distribution of $(X, \Delta, Y, A)$. We make the following assumptions: (i) patients enrolled in the trial are a simple random sample from the population of interest; (ii) the treatment assignment for patients in the trial does not depend on his marker value, that is, $A_i$ and $Y_i$ are independent.

Noting that $E\left\{T^{*a}\,|Y > y_v\right\} = E\left\{T\,|Y > y_v, T = a\right\}$ $(a = 0, 1)$, together with the fact that for a non-negative random variable $W$, $E(W) = \int_0^\infty P(W \ge w)dw$ [14, Ch 20, Sec 21], (2) can be rewritten as

$$\theta(v) = (1-v)\int_0^L S^{(1)}(t, y_v)dt + v\int_0^L S^{(0)}(t, y_v)dt,$$

where $S^{(1)}(t, y) = P(T \ge t | A = 1, Y > y)$ and $S^{(0)}(t, y) = P(T \ge t | A = 0, Y \le y)$. Therefore, an estimator of $\theta(v)$ can be obtained by substituting estimators for

$S^{(a)}(t,y)$ $(a=0,1)$ and $y_v$. Here we use the Kaplan-Meier estimator of $S^{(a)}(t,y)$,

$$\hat{S}^{(a)}(t,y) = \prod\!\!\!\!\prod_0^t \left\{ 1 - \frac{N_i^{(a)}(ds,y)}{R_i^{(a)}(s,y)} \right\},$$

where $\prod\!\!\!\!\prod$ is the product integral notation, $N_i^{(1)}(t,y) = I(X_i \le t, \Delta_i = 1, A_i = 1, Y_i > y)$, $N_i^{(0)}(t,y) = I(X_i \le t, \Delta_i = 1,\ A_i = 0, Y_i \le y), R_i^{(1)}(t,y) = I(X_i \ge t, A_i = 1, Y_i > y)$, and $R_i^{(0)}(t,y) = I(X_i \ge t, A_i = 0, Y_i \le y)$, $N_i^{(a)}(ds,y) = N_i^{(a)}(s,y) - N_i^{(a)}(s-,y)$. To estimate $y_v$, we use the nonparametric estimator $\hat{y}_v = \hat{F}_Y^{-1}(v)$, where $\hat{F}_Y(y) = n^{-1}\sum_{i=1}^n I(Y_i \le y)$ is the empirical distribution of $Y$. Thus an estimator of $\theta(v)$ can be written as

$$\hat{\theta}(v) = (1-v)\int_0^L \hat{S}^{(1)}(t,\hat{y}_v)dt + v\int_0^L \hat{S}^{(0)}(t,\hat{y}_v)dt.$$

This is a nonparametric estimator. It does not require any parametric assumptions on the survival time and the biomarker. Moreover, it is invariant to any monotone increasing transformation of the biomarker.

Now we derive the asymptotic properties of $\hat{\theta}(v)$ using empirical process theory. The idea is to first show that $n^{1/2}\left\{\hat{S}^{(a)}(t,y) - S^{(a)}(t,y)\right\}$ converges to a Gaussian process under some regularity conditions. Since $\hat{\theta}(v)$ is a composite function of $\hat{S}^{(0)}$, $\hat{S}^{(1)}$ and $\hat{F}_Y$, the asymptotic distribution of $\hat{\theta}(v)$ is then derived by the functional delta method [15, Ch 3.9]. Specifically, we can show that $n^{1/2}\left\{\hat{\theta}(v) - \theta(v)\right\}$ converges to a Gaussian process. The details of the asymptotic distribution and the proof are given in [16].

The covariance formula for $\hat{\theta}(v)$ contains the density function of $Y$ and the derivative of the cumulative hazard functions of $T$ conditional on $(A = 1, Y > y)$ or $(A = 0, Y \le y)$. Usually smoothing techniques are needed to estimate these quantities. In applications, to avoid the complexity of the smoothing approaches, for

simplicity, we use the bootstrap method to calculate the standard error and confidence band for $\theta(v)$. This is justified by the empirical process theory for the bootstrap given in [15, Ch 3.9]. Let $\Theta = \{(X_i, \Delta_i, Y_i, A_i) : i = 1, \ldots, n\}$ be the observed data set, and $\Theta_B$ be the $B^{th}$ resampling bootstrap dataset, where $B = 1, \ldots, M$. Let $\hat{\theta}^B(v)$ be the estimator based on $\Theta_B$. Then $\sup_{v\in[p,q]} |\hat{\theta}^B(v) - \hat{\theta}(v)|$ given $\Theta$ is asymptotically equivalent to $\sup_{v\in[p,q]} |\hat{\theta}(v) - \theta(v)|$, $0 < p < q < 1$. Let $c_\alpha$ be the $1-\alpha$ quantile of $\sup_{v\in[p,q]} |\hat{\theta}^B(v) - \hat{\theta}(v)|$, then a level $1-\alpha$ confidence band for $\theta(v)$ is $(\hat{\theta}^B(v) - c_\alpha, \hat{\theta}^B(v) + c_\alpha)$. The standard error of $\hat{\theta}(v)$ can be estimated by the standard deviation of $\{\hat{\theta}^B(v) : B = 1, \ldots, M\}$.

The optimal percentile $v_{opt}$ can be estimated by $\hat{v}_{opt}$, which maximizes $\hat{\theta}(v)$. When $v_{opt}$ is an interior point in the interval $[0, 1]$, under some regularity conditions, it can be shown that $\hat{v}_{opt}$ is asymptically normal. The standard errors of $\hat{v}_{opt}$ can be obtained via bootstrap. The Optimal threshold $c_{opt}$can be estimated by $\hat{c}_{opt} = \hat{F}_Y^{-1}(\hat{v}_{opt})$.

## 3. COVARIATE SPECIFIC SELECTION IMPACT CURVE

So far, we have considered treatment selection based only on the biomarker. There are situations that covariates other than the biomarker may impact treatment selection. Thus it is important to adjust for such covariates. For example, if we use c-myc gene expression level to select patients for surgery alone or surgery plus chemotherapy, we may want to adjust for sex. If the covariates are discrete, it is possible to consider a separate SI curve for each covariate combination. However, this will not work when there exist continuous covariates. In addition, even if all covariates are discrete, the sample size for some covariate combination may be too small to obtain a reliable SI curve estimate. It is noticeable that the SI curve bears some similarities to the ROC

curve used to evaluate diagnostic tests based on a biomarker [17]. To account for covariate effects, the ROC curve has been extended to the covariate specific ROC curve [6, 7]. This motivates us to propose the covariate specific SI curve by analogy to the covariate specific ROC curve.

### 3.1. *Definition*

Let $\boldsymbol{Z}$ denote the vector of $K$ covariates that may impact the treatment selection other than the biomarker. To incorporate the covariates, we consider the following treatment policy given $\boldsymbol{Z} = \boldsymbol{z}$:

$$\begin{aligned} &\text{if } Y > y_{v,\boldsymbol{z}} \quad , \quad \text{select treatment } A = 1; \\ &\text{if } Y \leq y_{v,\boldsymbol{z}} \quad , \quad \text{select treatment } A = 0. \end{aligned} \tag{4}$$

Here $y_{v,\boldsymbol{z}}$ is the $v$th quantile of the conditional distribution of $Y$ given $\boldsymbol{Z} = \boldsymbol{z}$. The corresponding covariate specific SI curve is defined as

$$\theta(v|\boldsymbol{z}) \;\; = \;\; E\left\{T^* \,|\text{treatment policy (4)}\right\}$$

Using similar arguments as those for the unadjusted SI curve $\theta(v)$, we can show that

$$\theta(v|\boldsymbol{z}) = (1-v)E\left\{T^{*1}\,|Y > y_{v,\boldsymbol{z}}, \boldsymbol{Z} = \boldsymbol{z}\right\} + vE\left\{T^{*0}\,|Y \leq y_{v,\boldsymbol{z}}, \boldsymbol{Z} = \boldsymbol{z}\right\}. \tag{5}$$

This definition is general to any outcomes, include binary and continuous outcomes. It reduces to the unadjusted SI curve given in (2) when there is no covariate.

The covariate specific SI curve considers treatment selection conditional on the covariate $\boldsymbol{Z}$, which may be different for various values of $\boldsymbol{Z}$. For a given value of $\boldsymbol{z}$, the optimal threshold $\hat{v}_{opt}(z)$ can be obtained by maximizing $\theta(v|\boldsymbol{z})$. It provides a useful tool for policy makers to choose different biomarker thresholds based on the values of $\boldsymbol{Z}$.

3.2. *Estimation*

Now we consider estimation of $\theta(v|\boldsymbol{z})$ using data obtained from randomized trials, which satisfies condition (i) and (ii) described in Section 2.2 and (iii)$'$ $A$ is independent of $Y$ given $\boldsymbol{Z}$. Suppose we have independent and identically distributed observations $(X_i, \Delta_i, Y_i, A_i, \boldsymbol{Z}_i)$, $i = 1, \ldots, n$.

With some simple algebra, it can be shown that (5) can be rewritten as

$$\theta(v|\boldsymbol{z}) = \int_v^1 \int_0^L S(t|u, \boldsymbol{z}, 1)dtdu + \int_0^v \int_0^L S(t|u, \boldsymbol{z}, 0)dtdu, \tag{6}$$

where $S(t|u, \boldsymbol{z}, a) = \mathrm{P}\{T \geq t|Y = y_{u,\boldsymbol{z}}, \boldsymbol{Z} = \boldsymbol{z}, A = a\}$. An estimator of $\theta(v|\boldsymbol{z})$ can be obtained based on (6).

Assume that the hazard of failure follows the proportional hazards model

$$\lambda_i(t|V_i, \boldsymbol{Z}_i, A_i) = \lambda_0(t) \exp\left[\beta_0^T g\left\{V_i, \boldsymbol{Z}_i, A_i\right\}\right], \tag{7}$$

where $V_i = F_{Y|\boldsymbol{Z}}(Y_i|\boldsymbol{Z}_i)$, $F_{Y|\boldsymbol{Z}}(\cdot|\boldsymbol{z})$ is the conditional distribution function of $Y$ given $\boldsymbol{Z} = \boldsymbol{z}$, and $g(v, \boldsymbol{z}, a)$ is a known $r$-dimensional function. It is useful to write the model in terms of $V_i$, since $\theta(v)$ is considered a function of $v$ for the reasons mentioned earlier. For example we might use the model

$$\begin{aligned}\lambda_i(t|V_i, \boldsymbol{Z}_i, A_i) &= \lambda_0(t) \exp\left[\beta_1 D(V_i, \boldsymbol{Z}_i) + \beta_2^T \boldsymbol{Z}_i + \beta_3 A_i + \beta_4 D(V_i, \boldsymbol{Z}_i) A_i\right. \\ &\quad \left. + \beta_5^T D(V_i, \boldsymbol{Z}_i) Z_i + \beta_6 Z_i A_i + \beta_7^T D(V_i, \boldsymbol{Z}_i) Z_i A_i\right],\end{aligned}$$

where $D(v, z)$ is a function of $v$ and $z$. When $D(v, z) = F_{Y|\boldsymbol{Z}}^{-1}(v|z)$, $D(v_i, \boldsymbol{Z}_i)$ is just the biomarker $Y_i$. Further assuming that $F_{Y|\boldsymbol{Z}}(\cdot|\boldsymbol{z})$ is known, based on (6), we can estimate $\theta(v|\boldsymbol{z})$ by

$$\tilde{\theta}(v|\boldsymbol{z}) = \int_v^1 \int_0^L \hat{S}(t|u, \boldsymbol{z}, 1)dtdu + \int_0^v \int_0^L \hat{S}(t|u, \boldsymbol{z}, 0)dtdu,$$

where

$$\hat{S}(t|v,\boldsymbol{z},a)=\exp\left\{-\hat{\Lambda}_0(t)\exp\left(\hat{\beta}^T g(v,\boldsymbol{z},a)\right)\right\}$$

is an estimator for the survival function $S(t|v,\boldsymbol{z},a)$, $\hat{\beta}$ is the partial likelihood estimator of $\beta_0$, and $\hat{\Lambda}_0(t)$ is the Breslow estimator of the baseline cumulative hazard function $\Lambda_0(t)$.

In practice $F_{Y|\boldsymbol{Z}}$ is usually unknown. We assume the conditional distribution of $Y$ given $\boldsymbol{Z}$ following the semiparametric model [18, 19],

$$F_{Y|\boldsymbol{Z}}(y|\boldsymbol{z})=h(y-\gamma_0^T\boldsymbol{z}), \tag{8}$$

where $h(\cdot)$ is an unknown distribution function. The estimator $\hat{\gamma}$ of $\gamma_0$ can be obtained by solving

$$\sum_{i=1}^{n}(Y_i-\gamma^T\boldsymbol{Z}_i)\boldsymbol{Z}_i=0.$$

The function $h(y)$ can be estimated by

$$\hat{h}(y;\hat{\gamma})=n^{-1}\sum_{i=1}^{n}I(Y_i-\hat{\gamma}^T\boldsymbol{Z}_i\le y).$$

Let $\tilde{\beta}$ be the solution to the partial likelihood estimating equation with $F_{Y|\boldsymbol{Z}}(y|\boldsymbol{z})$ replaced by $\hat{h}(y-\hat{\gamma}^T\boldsymbol{z};\hat{\gamma})$ and $\tilde{\Lambda}_0(t)$ be the corresponding Breslow estimator of $\Lambda_0(t)$. An estimator of $\theta(v|\boldsymbol{z})$ is

$$\tilde{\theta}^*(v|\boldsymbol{Z})=\int_0^L\int_v^1\tilde{S}(t|v,\boldsymbol{z},1)dudt+\int_0^L\int_0^v\tilde{S}(t|v,\boldsymbol{z},0)dudt,$$

where

$$\tilde{S}(t|v,\boldsymbol{z},a)=\exp\left\{-\tilde{\Lambda}_0(t)\exp\left(\tilde{\beta}^T g(v,\boldsymbol{z},a)\right)\right\}.$$

Note that both $\tilde{\theta}$ and $\tilde{\theta}^*$ can also be used in the case of no covariates. Since these estimators are semiparametric estimators, they are more efficient than the nonparametric

estimator $\hat{\theta}$ under the correct model assumptions. However, if the model assumptions are violated, inference based on these estimators can be misleading. We may need to transform the biomarker appropriately to assure that the proportional hazards assumption holds. In contrast, the nonparametric approach is invariant to monotone increasing transformations.

Under some regularity conditions, we can show that both $n^{1/2}\left\{\tilde{\theta}(v|\boldsymbol{z}) - \theta(v|\boldsymbol{z})\right\}$ and $n^{1/2}\left\{\tilde{\theta}^*(v|\boldsymbol{z}) - \theta(v|\boldsymbol{z})\right\}$ converge to zero-mean Gaussian processes. The derivation is involved with application of empirical process and U-process theories. The details of the asymptotic distribution and the proof are given in [16]. A consistent estimator for the variance of $\tilde{\theta}(v|\boldsymbol{z})$ is given in equation (B.4), which involves complex integrals. The variance formula for $\tilde{\theta}^*(v)$ is even more complicated and involves the derivative of the unknown function $h(\cdot)$. In practice, by analogy to the nonparametric case, we can use the bootstrap method to compute estimates of the standard errors and confidence bands for $\tilde{\theta}(v|\boldsymbol{z})$ and $\tilde{\theta}^*(v|\boldsymbol{z})$.

The optimal percentile $v_{opt}(z)$ can be estimated by $\tilde{v}_{opt}(z)$ and $\tilde{v}^*_{opt}(z)$, which maximizes $\tilde{\theta}(v|z)$ and $\tilde{\theta}^*(v|z)$ respectively. When $v_{opt}(z)$ is an interior point in the interval $[0, 1]$, under some regularity conditions, it can be shown that $\tilde{v}_{opt}(z)$ and $\tilde{v}^*_{opt}(z)$ are asymptotically normal. The variance of $\tilde{v}_{opt}(z)$ and and $\tilde{v}^*_{opt}(z)$ can be obtained via bootstrap. The optimal threshold $v_{opt}(z)$ can be estimated by $\tilde{c}_{opt}(z) = F_{Y|Z}(\tilde{v}_{opt}(z)|z)$ and $\tilde{c}^*_{opt}(z) = \hat{F}_{Y|Z}(\tilde{v}^*_{opt}(z)|z)$ correspondingly.

Note: We have derived the asymptotic properties of estimators of the overall SI curve and the covariate adjusted SI curve in an interval excludes $v = 0$ and $v = 1$. However, the asymptotic normality still holds at $v = 0$ and $v = 1$ by noticing that the nonparametric estimator essentially does not include the term $y_v$ at $v = 0$ and 1 and

the proof can be easily modified for the asymptotic normality of the semiparametric estimators.

## 4. SIMULATION STUDIES

Simulation studies were conducted to assess the finite sample properties of the estimators. We assess the performance of the estimators at boundary points $v = 0$ and $v = 1$ as well as internal $v$ points.

*Case I*: We first considered the simple case of no covariates when both the nonparametric and semiparametric estimators can be used. We generated data for 100 and 300 patients in a randomized trial with $\mathrm{P}[A = 1] = 0.5$. The marker $Y$ was generated from the standard normal distribution. The survival time was generated according to the proportional hazards model (7) with $g(v, \boldsymbol{z}, a) = (v, a, va)^T$, $\beta = (1, 0.5, -2.6)^T$, and $\lambda_0(t) = 0.1$. Censoring time was generated from an exponential distribution with mean 30 and truncated at 20, leading to a censoring rate of about 30%. Consider $L = 15$, and the SI curve is shown in Fig. 4.

We estimated the SI curve using $\hat{\theta}$, $\tilde{\theta}$ and $\tilde{\theta}^*$ for 500 simulated data sets. The estimated standard errors were computed by the bootstrap method using 100 resampled data sets for all the estimators. For each estimator, the 95% Wald confidence intervals were calculated. Table 1 presents the results for $v = 0.0, 0.1, \ldots, 1.0$. All the estimators exhibit negligible bias, and the standard errors track the true standard deviations of the estimators well. The coverage probabilities are close to their nominal levels. As expected, $\tilde{\theta}^*$ is less efficient than $\tilde{\theta}$ because the true $F_Y$ is used for the former while $F_Y$ is estimated for the latter, but the efficiency loss is small. Moreover, $\tilde{\theta}^*$ is more efficient than the nonparametric estimator $\hat{\theta}$ as expected.

*Case II*: To investigate the robustness of the estimators, we also conducted simulations under similar scenarios as above except that the survival time was generated from a gamma distribution with the shape parameter equal to 4 and the scale parameter equal to $\{5\lambda_i(t, Y_i, \boldsymbol{Z}_i)\}^{-1}$, where $\lambda_i(t, Y_i, \boldsymbol{Z}_i)$ has the same form of the hazard in Case I with $\beta = (-4, -2, 1)^T$ and $\lambda_0(t) = 2$. In this case, the survival model is misspecified for the semiparametric estimators. The results are shown in Table 2. The nonparametric estimator $\hat{\theta}$ still works well while the semiparametric estimators are obviously biased with poor coverage probabilities for the confidence intervals.

Case III: We then considered the case when a covariate $\boldsymbol{Z}$ was included. The covariate $\boldsymbol{Z}$ was generated from a standard normal distribution. The biomarker $Y$ equals $\boldsymbol{Z}$ plus a standard normal error. The survival time followed the proportional hazards model (7) with $g(v, \boldsymbol{z}, a) = (v, a, z, va, vz, az, vaz)^T$, $\beta = (1, -0.7, -0.1, -0.9, -0.3, 0.8, -0.9)$, and $\lambda_0(t) = 0.2$. Censoring time was generated from an exponential distribution with mean 20 and truncated at 8, leading to a censoring rate of about 30%. Consider $L = 5$ and the covariate SI curves for $z = -1, 0, 1$ are shown in Fig. 5. We estimated the covariate specific SI curves for 500 simulated datasets with $n = 100$ and 200 using $\tilde{\theta}$ and $\tilde{\theta}^*$. The results in Table 3 show similar patterns as above and both estimators work well.

We all assessed the estimation of the optimal percentile under case I, where the optimal percentile $v_{opt} = 0.196$. The results are shown in table IV. The confidence interval are computed via bootstrap, which works better than the asymptotic confidence interval at these sample sizes in our numerical study. Overall the estimators work reasonably well. The nonparametric estimator is less efficient than the semiparametric estimators.

In summary, the nonparametric estimator $\hat{\theta}$ is robust while the semiparametric estimators $\tilde{\theta}^*$ and $\hat{\theta}$ depend on the correct specification of the models. Obvious deviances between these estimators may indicate violation of the model assumptions for the semiparametric estimators. We recommend using the nonparametric estimator when there is no covariates. Extension of the semiparametric estimators to more flexible models is discussed in Section 6.

## 5. APPLICATION

As an illustration, we applied the proposed approaches to a subset of the ECOG clinical trial, which was analyzed in [5]. In this subset, disease progression free survival and c-myc expression level was measured on 92 patients randomized to receive surgery alone or surgery plus chemotherapy, among which 47 were males and 45 were females. The analysis in [5] focused on assessing whether there was a c-myc effect and/or a treatment/c-myc interaction. This is different from our goal here, which is to assess whether the c-myc expression level can be used to select patients for surgery alone versus surgery plus chemotherapy. We estimated the SI curve using both the nonparametric and semiparametric methods. For the semiparametric approach, we included log c-myc expression level $Y$, treatment indicator $A = I(\text{surgery plus chemotherapy})$ and their interaction in the the proportional hazards model. This model was also used in [5]. This corresponds to $g(v, \boldsymbol{z}, a) = (v, a, va)^T$ in the proportional hazards model (7). We consider the mean disease progression free survival time restricted to five years. Since the distribution $F_Y(y)$ is unknown, we estimated the SI curve using the estimators $\hat{\theta}(v)$ and $\tilde{\theta}^*(v)$. The estimates and the 95% pointwise confidence intervals and simultaneous confidence bands are shown in the left panel of Figure 6.

The semiparametric estimate is essentially a smoothed version of the nonparametric estimate, which indicates that the corresponding model assumptions are appropriate for this dataset. The estimated SI curves seem to decrease with $v$ with the estimated optimal percentile 0.00 (se= 0.153). That is, assigning all patients to surgery plus chemotherapy may achieve the maximum mean survival time within 5 years. However, this may not be the optimal treatment policy considering that surgery plus chemotherapy has more side effects than surgery alone and the estimated SI curve is almost horizonal for $v$ in $(0, 0.3)$.

To take into account of the side effect of chemotherapy, we may consider a utility measure $U$, for example, $U = T^* - dA$, where $d$ is a nonnegative weight denoting the deteriorating effect of chemotherapy on the survival time; that is, a person may prefer surgery alone if the additional chemotherapy does not lengthen the survival time by at least $d$. The SI curve based on $U$ can be obtained by shifting the SI curve based on $T^*$ by $w(1-v)$ and hence can be easily estimated. The optimal percentile can be achieved by maximizing the SI curve based on $U$. The right panel of Figure 6 shows the corresponding nonparametric and semiparametric estimates with $d = 0.5$. The estimated optimal percentile is 0.322 (se= 0.245) and corresponding estimated threshold is 0.118 based on the semiparametric SI curve estimate. This indicates that, considering the c-myc expression level and the side effect of chemotherapy, the optimal treatment policy may be the one that assigns patients whose log c-myc expression level falls below 0.118 (32.2% percentile) to surgery alone. The maximum point of the nonparametric estimate may be less stable since the non-smoothness of the curve. We do not see significant difference on either estimate across $v$ which may due to the small sample size of this dataset. Whether the difference is clinically important may

be worth future investigation with larger clinical trials.

As we have discussed in Section 1, it is also of interest to assess whether the gender of the patient may affect the treatment selection. Although we may estimate a separate SI curve for males and females, the estimate may be unreliable because of the small sample size in each group. We thus estimated the covariate specific SI curve based on $T$ adjusting for $Z = I(\text{gender=male})$ using the semiparametric estimator $\tilde{\theta}^*(v|\boldsymbol{z})$ by further including $Z$ and all the two-way and three-way interactions in the proportional hazards model considered above. The left panel of Figure 7 shows the results. For male patients, the SI curve seems to be horizontal when $v$ is small and decline thereafter, while for female patients the SI curve tends to decrease for $v$ from 0 to 1. This indicates the optimal thresholds treatment selection should be different for males and females. The estimated optimal percentiles are 0.00 (se= 0.265) are females and 0.130 (se= 0.207) with the corresponding estimated threshold $-0.033$ for males. To take into account of the side effect of chemotherapy, the covariate specific SI curves based on $U$ were also estimated as shown in the right panel of Figure 7. The estimated optimal percentiles are 0.00 (se= 0.345) for females and 0.239 (se= 0.252) with the corresponding estimated threshold 0.077 for males. Thus we may want to assign almost all females to surgery plus chemotherapy while assign male patients whose log c-myc expression level fall below 0.077 (23.3% percentile) to surgery alone.

To ensure the validity of these estimators, we checked the proportional hazards assumptions for the models using the method in [20, Ch 6.2]. There were no evidences against the proportional hazards assumptions.

## 6. DISCUSSION

We have proposed the SI curve for the survival time to evaluate the impact of a treatment selection policy based on $Y > y_v$. Both nonparametric and semiparametric estimators are derived. We recommend using the nonparametric estimator when it is not necessary to adjust for covariates because of its robustness. The semiparametric estimators are less robust, but can be more efficient under the correctly specified model. In addition, they can easily incorporate covariates in estimation. Whether to use the unadjusted or adjusted SI curve depends on the specific treatment policy, that is, whether we would like to select the treatment based on the biomarker alone.

For the semiparametric approach, we have used the standard proportional hazards model to characterize the relationship between the survival time and the marker and the covariates. This can be easily adapted to more flexible models. For example, we can use other survival models, such as the stratified proportional hazards model, the accelerated failure time model and the transformation model, as long as we can obtain consistent estimators for the survival distribution. The nonparametric transformation model [21] may be an attractive extension as it includes most popular survival models as special cases, such as the proportional hazards model and the accelerated failure time model. The semiparametric location model (8) is used only for estimating the conditional distribution function $F_{Y|\boldsymbol{Z}}$. We can remove this assumption and estimate $F_{Y|\boldsymbol{Z}}$ by kernel smoothing method when the number of covariates is small and the sample size is relatively large, since the kernel smoothing method may not work well otherwise.

Although the SI curve is proposed based on the restricted mean survival time, it can be extended to other statistical measures. For example, a utility function that

incorporates notions of cost and quality of life might be employed. Although we have focused on survival time as outcomes, the approach can be adapted to discrete and continuous outcomes with minor modifications.

In this paper, we consider SI curves based on a single biomarker. In practice, there may exist multiple biomarkers. For example, multiple genes may be related to a specific disease or affect the survival time. An important issue is how to combine these biomarkers for treatment selection. On the other hand, it may be needed to select among more than two treatments. These issues will be investigated in our future research.

## ACKNOWLEDGEMENTS

This work is supported in part by the UGARF grant from the University of Georgia (Song) and the VA Clinical R&D funding (Zhou and Song). Zhou's work is supported in part by AHRQ grant R01HS013105 and VA HSR&D grant EPID-006-07F. This paper presents the findings and conclusions of the authors. It does not necessarily represent those of VA HSR&D Service. We would like to thank Dr. Yi Li and ECOG for providing the cancer data.

## REFERENCES

1. Song X, Pepe MS. Evaluating markers for selecting a patient's treatment. *Biometrics* 2004; **60**:874–883.

2. Elmer-Dewitt P, Lemonick M, Park A, Nash M. Medicine: the future of drugs. *Time* 2001; **57**:56–102.

3. Erisman MD, Scott JK, Watt RA, Astrin SM. The c-myc protein is constitutively expressed at elevated levels in colorectal carcinoma cell lines. *Oncogene* 1988; **2**:367–378.

4. Augenlicht L, Wadler S, Corner G, Richards C, Ryan L, Multani A, Pathak S, Benson A, Hailer D, Heerdt B. Low-level c-myc amplification in human colonic carcinoma cell lines and tumors: a frequent, p53-independent mutation associated with improved outcome in a randomized multi-institutional trial. *Cancer Research* 1997; **57**:1769–1775.

5. Li Y, Ryan L. Inference on survival data with covariate measurement error an imputation-based approach. *Scandinavian Journal of Statistics* 2006; **33**:169–190.

6. Zhou XH, Obuchowski NA, McClish DK. *Statistical Methods in Diagnostic Medicine*. New York: Wiley 2002.

7. Pepe MS. *The Statistical Evaluation of Medical Tests for Classification and Prediction*. New York: Oxford University 2003.

8. Klein JP, Moeschberger ML. *Survival Analysis: Techniques for Censored and Truncated Data*. New York: Springer 2003.

9. Zhao H, Tsitais AA. A consistent estimator for the distribution of quality adjusted lifetime. *Biometrika* 1997; **84**:339–348.

10. Chen PY, Tsitais AA. Causal inference on the difference of the restricted mean lifetime between two groups. *Biometrics* 2001; **57**:1030–1038.

11. Rubin DB. Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology* 1974; **66**:688–701.

12. Rubin DB. Bayesian inference for causal effects: The role of randomization. *Annals of Statistics* 1978; **6**:34–58.

13. Holland P. Statistics and causal inference. *Journal of the American Statistical Association* 1986; **81**:945–960.

14. Billingsley P. *Probability and Measure*. New York: Wiley 1995.

15. Van der Vaart AW, Wellner JA. *Weak Convergence and Empirical Processes*. New York: Springer-Verlag 2000.

16. Song X, Zhou XH. Evaluating markers for treatment selection based on survival time. *UW Biostatistics Working Paper Series* 2009; :Working Paper 349, http://www.bepress.com/uwbiostat/paper349.

17. Pepe MS. Receiver operating characteristic methodology. *Journal of the American Statistical Association* 2000; **95**:308–311.

18. Pepe MS. Three approaches to regression analysis of receiver operating characteristic curves in medical diagnostic testing. *Biometrics* 1998; **54**:124–135.

19. Heagerty P, Pepe M. Semiparametric estimation of regression quantiles with application to standarizing weight for height and age in U.S. children. *Applied Statistics* 1999; **48**:533–551.

20. Therneau TM, Grambsch PM. *Modeling Survival Data: Extending the Cox Model*. New York: Springer-Verlag 2000.

21. Song X, Ma S, Huang J, Zhou XH. A semiparametric approach for the nonparametric transformation survival model with multiple covariates. *Biostatistics* 2007; **8**:197–211.

Table I. Results of simulation studies for overall SI under the proportional hazards model (case I). SD, empirical standard deviation across simulated data sets; SE, average of bootstrap standard errors; CP, coverage probability of the 95% Wald confidence interval.

| | | | $\hat{\theta}_{np}$ | | | | $\hat{\theta}_{p}$ | | | | $\tilde{\theta}_{p}$ | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $n$ | $v$ | $\theta$ | est | SD | SE | CP | est | SD | SE | CP | est | SD | SE | CP |
| 100 | 0.0 | 8.922 | 8.896 | 0.875 | 0.823 | 0.928 | 8.954 | 0.806 | 0.766 | 0.934 | 8.945 | 0.832 | 0.783 | 0.928 |
| | 0.1 | 9.087 | 9.069 | 0.871 | 0.838 | 0.932 | 9.127 | 0.756 | 0.727 | 0.936 | 9.118 | 0.777 | 0.739 | 0.930 |
| | 0.2 | 9.136 | 9.104 | 0.845 | 0.838 | 0.944 | 9.186 | 0.737 | 0.716 | 0.932 | 9.177 | 0.752 | 0.723 | 0.926 |
| | 0.3 | 9.069 | 9.075 | 0.837 | 0.827 | 0.938 | 9.128 | 0.732 | 0.718 | 0.934 | 9.119 | 0.742 | 0.722 | 0.936 |
| | 0.4 | 8.887 | 8.887 | 0.805 | 0.815 | 0.948 | 8.951 | 0.727 | 0.720 | 0.942 | 8.944 | 0.735 | 0.722 | 0.936 |
| | 0.5 | 8.592 | 8.610 | 0.782 | 0.798 | 0.948 | 8.659 | 0.716 | 0.714 | 0.940 | 8.653 | 0.722 | 0.715 | 0.940 |
| | 0.6 | 8.188 | 8.206 | 0.774 | 0.782 | 0.948 | 8.255 | 0.697 | 0.698 | 0.944 | 8.252 | 0.702 | 0.699 | 0.940 |
| | 0.7 | 7.681 | 7.679 | 0.754 | 0.767 | 0.954 | 7.747 | 0.674 | 0.678 | 0.948 | 7.746 | 0.679 | 0.680 | 0.944 |
| | 0.8 | 7.076 | 7.106 | 0.749 | 0.748 | 0.942 | 7.143 | 0.655 | 0.661 | 0.942 | 7.145 | 0.662 | 0.664 | 0.942 |
| | 0.9 | 6.380 | 6.414 | 0.749 | 0.732 | 0.922 | 6.451 | 0.650 | 0.657 | 0.938 | 6.456 | 0.659 | 0.662 | 0.942 |
| | 1.0 | 5.601 | 5.645 | 0.720 | 0.710 | 0.928 | 5.681 | 0.668 | 0.678 | 0.946 | 5.689 | 0.681 | 0.685 | 0.934 |
| 300 | 0.0 | 8.922 | 8.901 | 0.478 | 0.473 | 0.928 | 8.913 | 0.420 | 0.432 | 0.946 | 8.907 | 0.436 | 0.445 | 0.952 |
| | 0.1 | 9.087 | 9.061 | 0.459 | 0.476 | 0.940 | 9.086 | 0.390 | 0.409 | 0.932 | 9.081 | 0.401 | 0.419 | 0.942 |
| | 0.2 | 9.136 | 9.122 | 0.460 | 0.474 | 0.938 | 9.144 | 0.380 | 0.403 | 0.946 | 9.140 | 0.387 | 0.410 | 0.960 |
| | 0.3 | 9.069 | 9.077 | 0.456 | 0.470 | 0.948 | 9.084 | 0.381 | 0.405 | 0.946 | 9.080 | 0.385 | 0.410 | 0.958 |
| | 0.4 | 8.887 | 8.891 | 0.449 | 0.462 | 0.942 | 8.905 | 0.384 | 0.407 | 0.950 | 8.903 | 0.386 | 0.409 | 0.954 |
| | 0.5 | 8.592 | 8.606 | 0.440 | 0.456 | 0.944 | 8.612 | 0.384 | 0.404 | 0.960 | 8.610 | 0.385 | 0.406 | 0.958 |
| | 0.6 | 8.188 | 8.201 | 0.419 | 0.448 | 0.958 | 8.208 | 0.379 | 0.396 | 0.964 | 8.207 | 0.380 | 0.398 | 0.964 |
| | 0.7 | 7.681 | 7.681 | 0.430 | 0.438 | 0.956 | 7.699 | 0.371 | 0.385 | 0.966 | 7.699 | 0.372 | 0.387 | 0.972 |
| | 0.8 | 7.076 | 7.079 | 0.409 | 0.430 | 0.944 | 7.091 | 0.365 | 0.377 | 0.956 | 7.093 | 0.365 | 0.380 | 0.962 |
| | 0.9 | 6.380 | 6.375 | 0.403 | 0.421 | 0.946 | 6.393 | 0.365 | 0.376 | 0.946 | 6.396 | 0.366 | 0.380 | 0.950 |
| | 1.0 | 5.601 | 5.593 | 0.386 | 0.408 | 0.944 | 5.611 | 0.376 | 0.387 | 0.938 | 5.615 | 0.378 | 0.394 | 0.952 |

Table II. Results of simulation studies for overall SI under the Gamma survival model (case II). SD, empirical standard deviation across simulated data sets; SE, average of bootstrap standard errors; CP, coverage probability of the 95% Wald confidence interval.

| | | | $\hat{\theta}_{np}$ | | | | $\hat{\theta}_{p}$ | | | | $\tilde{\theta}_{p}$ | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $n$ | $v$ | $\theta$ | est | SD | SE | CP | est | SD | SE | CP | est | SD | SE | CP |
| 100 | 0.0 | 10.249 | 10.232 | 0.749 | 0.755 | 0.954 | 6.850 | 1.723 | 1.527 | 0.442 | 7.117 | 1.927 | 1.851 | 0.582 |
| | 0.1 | 9.953 | 9.941 | 0.695 | 0.715 | 0.950 | 6.559 | 1.719 | 1.515 | 0.440 | 6.825 | 1.912 | 1.828 | 0.576 |
| | 0.2 | 9.561 | 9.547 | 0.628 | 0.648 | 0.944 | 6.190 | 1.692 | 1.478 | 0.432 | 6.451 | 1.873 | 1.779 | 0.564 |
| | 0.3 | 9.046 | 9.104 | 0.540 | 0.565 | 0.942 | 5.748 | 1.623 | 1.402 | 0.406 | 5.998 | 1.793 | 1.689 | 0.544 |
| | 0.4 | 8.388 | 8.401 | 0.449 | 0.473 | 0.932 | 5.260 | 1.500 | 1.282 | 0.394 | 5.491 | 1.659 | 1.551 | 0.524 |
| | 0.5 | 7.599 | 7.613 | 0.376 | 0.392 | 0.946 | 4.773 | 1.326 | 1.125 | 0.364 | 4.981 | 1.475 | 1.375 | 0.510 |
| | 0.6 | 6.743 | 6.826 | 0.352 | 0.364 | 0.968 | 4.350 | 1.119 | 0.952 | 0.368 | 4.532 | 1.264 | 1.185 | 0.524 |
| | 0.7 | 5.932 | 5.979 | 0.387 | 0.413 | 0.972 | 4.037 | 0.918 | 0.798 | 0.398 | 4.195 | 1.064 | 1.017 | 0.566 |
| | 0.8 | 5.290 | 5.341 | 0.493 | 0.519 | 0.940 | 3.840 | 0.767 | 0.686 | 0.464 | 3.979 | 0.917 | 0.896 | 0.646 |
| | 0.9 | 4.886 | 4.931 | 0.616 | 0.633 | 0.942 | 3.733 | 0.679 | 0.622 | 0.536 | 3.857 | 0.833 | 0.827 | 0.714 |
| | 1.0 | 4.687 | 4.707 | 0.718 | 0.701 | 0.930 | 3.679 | 0.635 | 0.592 | 0.586 | 3.795 | 0.792 | 0.796 | 0.748 |
| 300 | 0.0 | 10.249 | 10.242 | 0.430 | 0.434 | 0.944 | 7.321 | 1.037 | 0.982 | 0.214 | 7.424 | 1.206 | 1.166 | 0.352 |
| | 0.1 | 9.953 | 9.946 | 0.402 | 0.407 | 0.952 | 7.028 | 1.037 | 0.982 | 0.214 | 7.131 | 1.198 | 1.160 | 0.356 |
| | 0.2 | 9.561 | 9.552 | 0.361 | 0.366 | 0.950 | 6.646 | 1.027 | 0.971 | 0.218 | 6.749 | 1.179 | 1.141 | 0.356 |
| | 0.3 | 9.046 | 9.049 | 0.318 | 0.320 | 0.942 | 6.174 | 0.993 | 0.936 | 0.212 | 6.275 | 1.136 | 1.097 | 0.346 |
| | 0.4 | 8.388 | 8.372 | 0.269 | 0.266 | 0.934 | 5.639 | 0.925 | 0.868 | 0.198 | 5.734 | 1.058 | 1.018 | 0.318 |
| | 0.5 | 7.599 | 7.582 | 0.220 | 0.219 | 0.938 | 5.095 | 0.821 | 0.768 | 0.186 | 5.181 | 0.944 | 0.907 | 0.304 |
| | 0.6 | 6.743 | 6.753 | 0.194 | 0.197 | 0.948 | 4.612 | 0.691 | 0.646 | 0.176 | 4.684 | 0.806 | 0.775 | 0.314 |
| | 0.7 | 5.932 | 5.942 | 0.219 | 0.225 | 0.958 | 4.253 | 0.557 | 0.526 | 0.184 | 4.309 | 0.667 | 0.647 | 0.348 |
| | 0.8 | 5.290 | 5.280 | 0.287 | 0.289 | 0.936 | 4.034 | 0.452 | 0.435 | 0.224 | 4.075 | 0.560 | 0.550 | 0.420 |
| | 0.9 | 4.886 | 4.870 | 0.363 | 0.358 | 0.932 | 3.920 | 0.389 | 0.381 | 0.326 | 3.951 | 0.498 | 0.493 | 0.514 |
| | 1.0 | 4.687 | 4.670 | 0.421 | 0.405 | 0.924 | 3.892 | 0.467 | 0.466 | 0.556 | 3.892 | 0.467 | 0.466 | 0.556 |

Table III. Results of simulation studies for the covariate specific SI curve (case III). SD, empirical standard deviation across simulated data sets; SE, average of bootstrap standard errors; CP, coverage probability of the 95% Wald confidence interval.

| | | | $n = 100$ | | | | | | | | $n = 200$ | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | $\tilde{\theta}$ | | | | $\tilde{\theta}^*$ | | | | $\tilde{\theta}$ | | | | $\tilde{\theta}^*$ | | | |
| $z$ | $v$ | $\theta$ | est | SD | SE | CP | est | SD | SE | CP | est | SD | SE | CP | est | SD | SE | CP |
| −1 | 0.0 | 3.943 | 3.945 | 0.295 | 0.309 | 0.944 | 3.952 | 0.297 | 0.308 | 0.942 | 3.946 | 0.201 | 0.206 | 0.974 | 3.951 | 0.202 | 0.207 | 0.976 |
| | 0.2 | 3.640 | 3.642 | 0.296 | 0.310 | 0.950 | 3.648 | 0.297 | 0.309 | 0.946 | 3.647 | 0.201 | 0.207 | 0.968 | 3.652 | 0.203 | 0.209 | 0.964 |
| | 0.4 | 3.295 | 3.291 | 0.321 | 0.339 | 0.970 | 3.298 | 0.322 | 0.341 | 0.970 | 3.303 | 0.219 | 0.225 | 0.964 | 3.309 | 0.222 | 0.227 | 0.964 |
| | 0.6 | 2.914 | 2.900 | 0.330 | 0.352 | 0.976 | 2.907 | 0.332 | 0.356 | 0.976 | 2.921 | 0.227 | 0.232 | 0.970 | 2.927 | 0.231 | 0.236 | 0.972 |
| | 0.8 | 2.508 | 2.489 | 0.320 | 0.337 | 0.982 | 2.498 | 0.324 | 0.344 | 0.976 | 2.516 | 0.222 | 0.225 | 0.974 | 2.522 | 0.227 | 0.231 | 0.974 |
| | 1.0 | 2.091 | 2.088 | 0.328 | 0.338 | 0.974 | 2.096 | 0.331 | 0.345 | 0.980 | 2.108 | 0.230 | 0.228 | 0.980 | 2.114 | 0.235 | 0.234 | 0.982 |
| 0 | 0.0 | 3.895 | 3.928 | 0.233 | 0.235 | 0.930 | 3.927 | 0.232 | 0.235 | 0.930 | 3.918 | 0.157 | 0.158 | 0.962 | 3.917 | 0.159 | 0.159 | 0.954 |
| | 0.2 | 3.714 | 3.747 | 0.223 | 0.225 | 0.938 | 3.747 | 0.221 | 0.224 | 0.938 | 3.738 | 0.146 | 0.151 | 0.950 | 3.738 | 0.147 | 0.151 | 0.952 |
| | 0.4 | 3.480 | 3.508 | 0.233 | 0.241 | 0.950 | 3.509 | 0.230 | 0.241 | 0.956 | 3.504 | 0.153 | 0.159 | 0.958 | 3.504 | 0.153 | 0.160 | 0.956 |
| | 0.6 | 3.190 | 3.211 | 0.236 | 0.250 | 0.968 | 3.213 | 0.233 | 0.251 | 0.976 | 3.213 | 0.159 | 0.164 | 0.966 | 3.214 | 0.159 | 0.166 | 0.962 |
| | 0.8 | 2.845 | 2.858 | 0.232 | 0.249 | 0.978 | 2.862 | 0.229 | 0.252 | 0.982 | 2.865 | 0.162 | 0.164 | 0.974 | 2.868 | 0.163 | 0.166 | 0.972 |
| | 1.0 | 2.446 | 2.456 | 0.246 | 0.262 | 0.984 | 2.462 | 0.244 | 0.265 | 0.978 | 2.465 | 0.179 | 0.174 | 0.974 | 2.470 | 0.180 | 0.177 | 0.974 |
| 1 | 0.0 | 3.768 | 3.759 | 0.323 | 0.344 | 0.958 | 3.752 | 0.325 | 0.349 | 0.954 | 3.777 | 0.222 | 0.220 | 0.958 | 3.773 | 0.225 | 0.223 | 0.956 |
| | 0.2 | 3.748 | 3.753 | 0.290 | 0.308 | 0.962 | 3.748 | 0.287 | 0.310 | 0.968 | 3.763 | 0.190 | 0.197 | 0.956 | 3.758 | 0.190 | 0.198 | 0.958 |
| | 0.4 | 3.637 | 3.650 | 0.309 | 0.320 | 0.956 | 3.647 | 0.305 | 0.321 | 0.962 | 3.655 | 0.197 | 0.206 | 0.966 | 3.651 | 0.196 | 0.206 | 0.966 |
| | 0.6 | 3.441 | 3.455 | 0.323 | 0.332 | 0.966 | 3.453 | 0.320 | 0.334 | 0.964 | 3.458 | 0.208 | 0.214 | 0.962 | 3.456 | 0.206 | 0.215 | 0.962 |
| | 0.8 | 3.163 | 3.178 | 0.331 | 0.337 | 0.970 | 3.178 | 0.330 | 0.340 | 0.976 | 3.180 | 0.219 | 0.221 | 0.968 | 3.181 | 0.219 | 0.222 | 0.970 |
| | 1.0 | 2.809 | 2.830 | 0.369 | 0.369 | 0.970 | 2.832 | 0.368 | 0.371 | 0.966 | 2.827 | 0.252 | 0.248 | 0.976 | 2.832 | 0.252 | 0.249 | 0.974 |

Table IV. Estimation of the optimal percentile under case I.

| | | $n = 100$ | | | | $n = 300$ | | | |
|---|---|---|---|---|---|---|---|---|---|
| | $v_{opt}$ | est | SD | SE | CP | est | SD | SE | CP |
| $\hat{v}_{opt}$ | 0.192 | 0.271 | 0.133 | 0.133 | 0.938 | 0.239 | 0.095 | 0.097 | 0.954 |
| $\tilde{v}_{opt}$ | 0.192 | 0.187 | 0.118 | 0.105 | 0.916 | 0.190 | 0.077 | 0.071 | 0.926 |
| $\tilde{v}^*_{opt}$ | 0.192 | 0.189 | 0.120 | 0.108 | 0.928 | 0.190 | 0.079 | 0.073 | 0.928 |

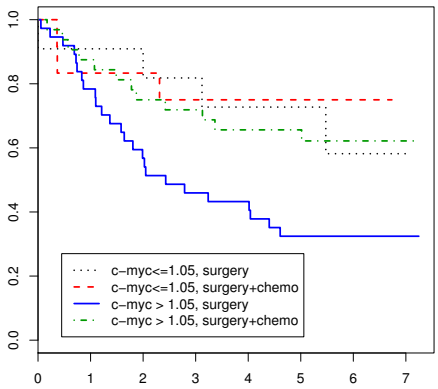



Figure 1. Kaplan-Meier estimates of disease progression free survival by treatment and c-myc level

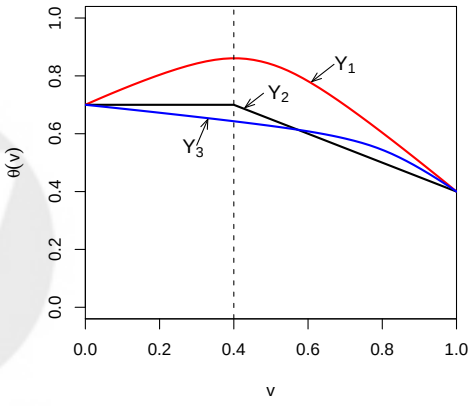
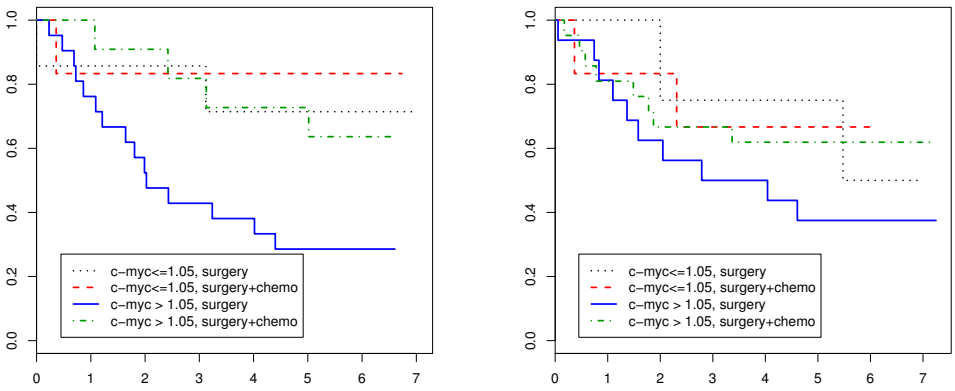



Figure 2. Example of SI curves




(a) Female (b) Male

Figure 3. Kaplan-Meier estimates of disease progression free survival by treatment and c-myc level for different genders.

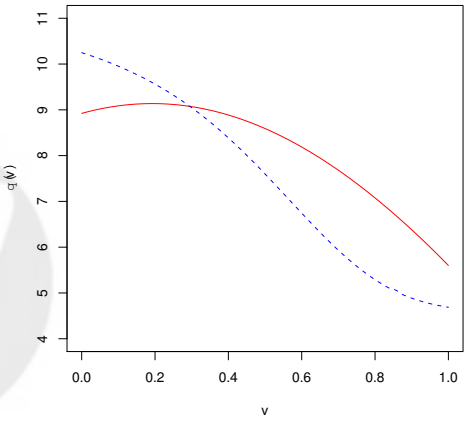



Figure 4. True SI curve in simulation. Solid line, proportional hazards model; dashed line, Gamma survival distribution.

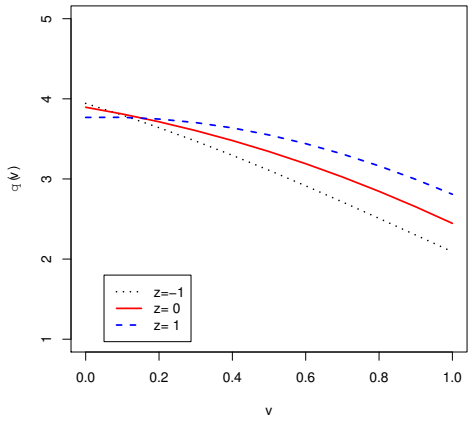



Figure 5. True covariate specific SI curve in simulation.

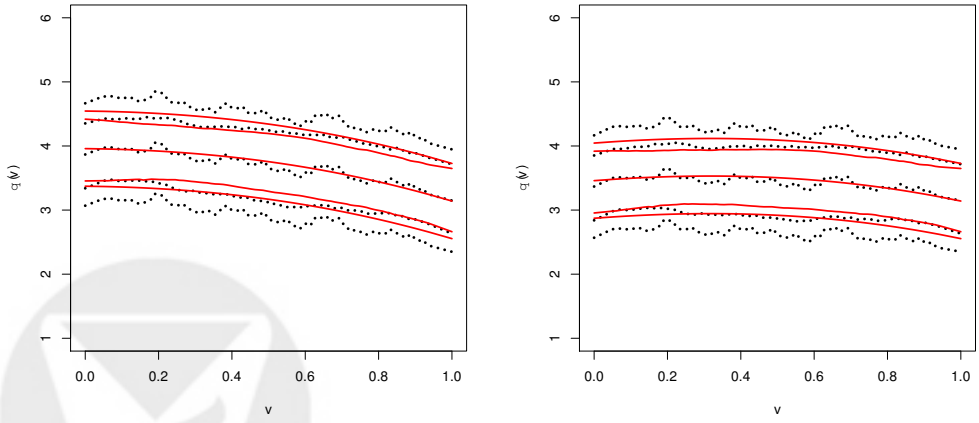
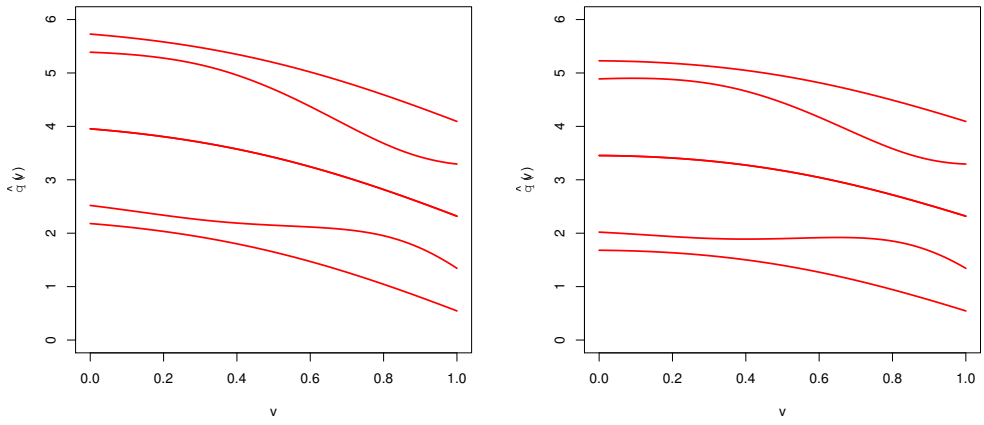
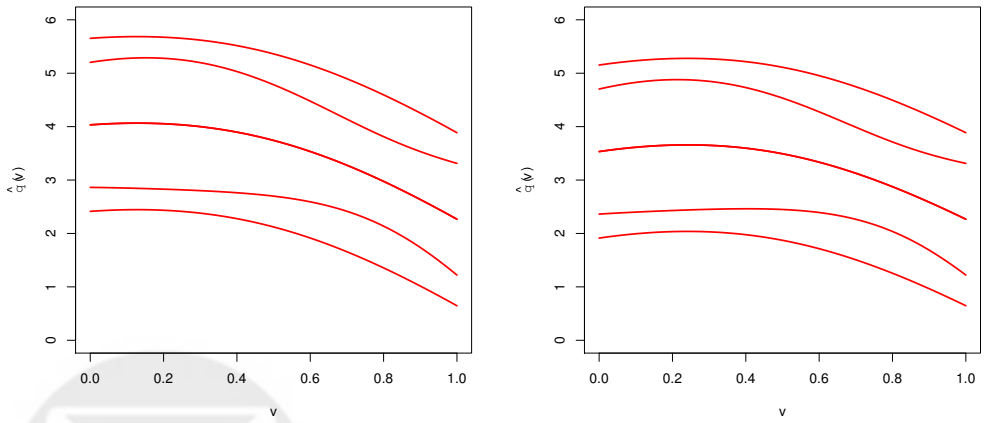



Figure 6. Estimated SI curve for ECOG data. Left panel is based on 5 year restricted survival time $T^*$, right panel is based on the utility function $U = T^* - 0.5A$. Nonparametric estimate $\hat{\theta}$, dotted curve; semiparametric estimate $\tilde{\theta}$, solid curve. 95% confidence bands are shown with the outer curves, 95% pointwise confidence intervals are shown with the intermediate curves, the estimates themselves are shown with the center curves.




(a) Female




(b) Male

Figure 7. Estimated SI curves adjusted for gender for ECOG data. Left panel is based on 5 year restricted survival time $T^*$, right panel is based on the utility function $U = T^* - 0.5A$. 95% confidence bands are shown with the outer curves, 95% pointwise confidence intervals are shown with the intermediate curves, the estimates themselves are shown with the center curves.