Published in final edited form as:

J R Stat Soc Series B Stat Methodol. 2017 September ; 79(4): 1165–1185. doi:10.1111/rssb.12201.

On Estimation of Optimal Treatment Regimes For Maximizing *t*-Year Survival Probability

Runchao Jiang, **Wenbin Lu**, **Rui Song**, and **Marie Davidian**

North Carolina State University, Raleigh, USA

Summary

A treatment regime is a deterministic function that dictates personalized treatment based on patients' individual prognostic information. There is increasing interest in finding optimal treatment regimes, which determine treatment at one or more treatment decision points so as to maximize expected long-term clinical outcome, where larger outcomes are preferred. For chronic diseases such as cancer or HIV infection, survival time is often the outcome of interest, and the goal is to select treatment to maximize survival probability. We propose two nonparametric estimators for the survival function of patients following a given treatment regime involving one or more decisions, i.e., the so-called value. Based on data from a clinical or observational study, we estimate an optimal regime by maximizing these estimators for the value over a prespecified class of regimes. Because the value function is very jagged, we introduce kernel smoothing within the estimator to improve performance. Asymptotic properties of the proposed estimators of value functions are established under suitable regularity conditions, and simulations studies evaluate the finite-sample performance of the proposed regime estimators. The methods are illustrated by application to data from an AIDS clinical trial.

Keywords

Inverse probability weighted estimation; Kaplan-Meier estimator; optimal treatment regime; personalized medicine; survival probability; value function

1. Introduction

For many complex diseases, such as cancer, HIV infection, and mental disorders, there is generally not a uniformly best treatment for all patients. Rather, different patients may benefit from different treatments due to individual heterogeneity. For example, in AIDS Clinical Trials Group (ACTG) Study 175 (Hammer et al., 1996), the primary composite outcome of interest was time to having a larger than 50% decline in CD4 count, a measure of immunological status; progression to AIDS; or death. For the comparison of two treatments, zidovudine plus didanosine (coded as 1) and zidovudine plus zalcitabine (coded as 0), the data suggest that zidovudine plus zalcitabine leads to more favorable outcomes for younger patients than zidovudine plus didanosine. Figure 1 shows treatment-specific Kaplan-Meier estimates of the survival function for the two age strata defined by the observed median age, 34 years, in ACTG 175. It is clear that, among younger patients with age ≤ 34, those receiving zidovudine plus zalcitabine have almost uniformly larger survival probabilities those receiving zidovudine plus didanosine, whereas the situation is reversed for older patients with age > 34.

This type of situation suggests that individual patient characteristics should be used when selecting treatments to maximize an expected long-term outcome of interest for which larger outcomes are preferred, such as t-year survival probability, and has heightened interest in derivation of optimal dynamic treatment regimes. Because in many chronic diseases treatment decisions may be made sequentially over time, a dynamic treatment regime is a set of one or more decision rules determine which treatment to give from among the available options based on accruing individual patient information, including baseline characteristics, intermediate outcomes between decisions, and previous treatments. An optimal regime is one that maximizes the expected outcome, or so-called value, if used by the entire patient population to select treatments.

There is a large literature on statistical methods to estimate an optimal treatment regime based on data from a clinical trial or observational study and non-survival outcomes. Q-learning (Watkins, 1989; Watkins and Dayan, 1992; Murphy, 2005; Zhao et al., 2009) and A-learning (Murphy, 2003; Robins, 2004) are two popular backward induction methods for estimating optimal dynamic treatment regimes based on regression-type modeling. The former involves positing parametric models for, roughly, the regression of outcome on accruing information and treatment, while the latter is based on semiparametric models in which only the part of the outcome regression representing contrasts among treatments is modeled parametrically, along with the propensity scores, the probabilities of observed treatment assignment given patient information at each decision point. Q-learning can be sensitive to misspecification of the required models, while A-learning enjoys the so-called double robustness property in that the corresponding estimating equations are asymptotically unbiased when either the propensity scores or main effects portion of the outcome models are correctly specified. An alternative class of approaches known as value or policy search methods is based on deriving and maximizing directly a consistent estimator for the value over a prespecified class of treatment regimes indexed by a finite-dimensional parameter. Zhang et al. (2012b) proposed inverse propensity score weighted (IPW) and augmented IPW (AIPW) estimators for the value in the case of a single decision point. Because the value estimator is nonsmooth, the optimization problem is challenging, and nonstandard optimization techniques are required. Zhao et al. (2012) and Zhang et al. (2012a) recast this approach as a weighted classification problem; the former refer to this method as outcome weighted learning. These approaches exploit approximations integrated into classification software to address the nonsmooth optimization problem, so that the class of regimes is dictated by a chosen classification method. Zhang et al. (2013) extended the value search methods of Zhang et al. (2012b) to more than one decision point, which share the computational challenges in the single decision case. Matsouaka et al. (2014) employed a kernel smoothing technique to nonparametrically estimate the conditional mean for the difference of the potential outcomes in a subgroup of patients and derived its associated treatment regime.

Although survival time is often the outcome of interest, to our knowledge there is relatively little development of methods for estimation of optimal treatment regimes where the goal is to maximize survival probability. Some work is focused on maximizing expected survival time. Goldberg and Kosorok (2012) developed a Q-learning method for censored survival data for estimating optimal dynamic treatment regimes and derived its associated finite sample risk bounds on the generalization error of the estimated regime, while Zhao et al. (2015) proposed a doubly robust estimator for expected survival time based on censored data and use outcome weighted learning to estimate an optimal regime. Bai et al. (2014) developed a locally-efficient doubly robust estimator for survival probability rather than mean survival time and estimate an optimal regime by extending the methods from a classification perspective of Zhang et al. (2012a). The latter two methods involve transforming maximization of the value to a weighted classification problem, which allows classification software to be used to address the optimization challenge and thus dictates the class of regimes. All of these methods are relevant to a single decision point only.

In this article, we propose a value search method for estimating an optimal treatment regime within a prespecified class for which the goal is to maximize survival probability that addresses the optimization challenges in a novel way and is relevant to more than one decision point. In particular, we develop a framework employing kernel smoothing techniques to smooth the estimator of the value prior to optimization, which we show greatly improves finite sample performance over the corresponding estimator with no smoothing. This approach is different from the smoothing technique used by Matsouaka et al. (2014), and, to the best of our knowledge, this is the first time smoothing has been integrated into estimation of the value function and its associated optimal treatment regimes in this way. Development of optimal treatment regimes for multiple decision points with censored survival data is challenging, as timing of observations, censoring, and events must be properly taken into account. In addition, we extend our smoothing approach to this setting.

In Sections 2 and 3, we introduce the statistical framework and estimators for a single decision point and multiple decisions, respectively. Asymptotic properties of the proposed estimators are given in Section 4. Finite sample performance is studied via simulation in Section 5, and Section 6 presents application of the methods to data from ACTG 175. Proofs are relegated to the Appendix.

2. Estimation of Optimal Treatment Regime for a Single Decision Time Point

2.1. Notation and Assumptions

Consider a study with two treatment options $\mathscr{A}$; = {0, 1} given at baseline. For the ith patient, $i = 1, \ldots, n$, let $\boldsymbol{X}_i$ denote the p-dimensional vector of baseline covariates taking values $\boldsymbol{x} \in \mathscr{X}$ and A_i denote the actual treatment received by the patient. Let T_i be the associated continuous survival time of interest, with conditional survival function $S_T(t|a, \boldsymbol{x}) \equiv P(T_i > t|A_i = a, \boldsymbol{X}_i = \boldsymbol{x})$ and corresponding conditional cumulative hazard function $\Lambda_T(t|a, \boldsymbol{x})$, where $a = 0, 1$. Let C_i denote right censoring time for patient i. The observed data are $\{(\boldsymbol{X}_i, A_i, \tilde{T}_i, \delta_i), i = 1, \ldots, n\}$, independent and identically distributed (iid) across i, where $\tilde{T}_i = \min\{T_i, C_i\}$ and $\delta_i = I\{T_i \le C_i\}$. We thus observe the counting process $N_i(t) = I(\tilde{T}_i \le t, \delta_i = 1)$ and the at risk process $Y_i(t) = I(\tilde{T}_i \ge t)$.

A treatment regime is a deterministic function that maps $\boldsymbol{x} \in \mathscr{X}$ to $\mathscr{A}$;. For simplicity, we assume the regimes of interest are from $\mathscr{G} = \{g_{\boldsymbol{\eta}} : g_{\boldsymbol{\eta}}(\boldsymbol{x}) = I\{\boldsymbol{\eta}^T \tilde{\boldsymbol{x}} \geq 0\}, \|\boldsymbol{\eta}\| = 1\}$, where $\tilde{\boldsymbol{x}} = (1, \boldsymbol{x}^T)^T$. However, the proposed method also applies to any other $\mathscr{G}$ indexed by finite-dimensional parameters. Denote the potential survival time of a patient if he/she were given treatment a, which may be contrary to fact, as $T^*(a)$. Accordingly, define the potential counting process $N^*(a; t)$ and at risk process $Y^*(a; t)$ under treatment a, where $N^*(a; t) = I\{\min(T^*(a), C) \leq t, T^*(a) \leq C\}$ and $Y^*(a; t) = I\{\min(T^*(a), C) \geq t\}$. If a patient follows a given regime $g_{\boldsymbol{\eta}}$, we can write the corresponding potential survival time as $T^*(g_{\boldsymbol{\eta}}) = T^*(1)g_{\boldsymbol{\eta}} + T^*(0)(1-g_{\boldsymbol{\eta}})$, whose survival function is given by $S^*(t; \boldsymbol{\eta}) = E(P[T^*\{g_{\boldsymbol{\eta}}(\boldsymbol{X})\} > t|\boldsymbol{X}])$, as well as the potential counting process $N^*(g_{\boldsymbol{\eta}}; t) = N^*(1; t)g_{\boldsymbol{\eta}} + N^*(0; t)(1 - g_{\boldsymbol{\eta}})$ and potential at risk process $Y^*(g_{\boldsymbol{\eta}}; t) = Y^*(1; t)g_{\boldsymbol{\eta}} + Y^*(0; t)(1 - g_{\boldsymbol{\eta}})$. We wish to find an optimal treatment regime in $\mathscr{G}$ that maximizes t-year survival probability; that is $g_{\boldsymbol{\eta}}^{\text{opt}}(\boldsymbol{x}) \equiv g(\boldsymbol{x};\boldsymbol{\eta}^{\text{opt}})$, where $\boldsymbol{\eta}^{\text{opt}} = \arg\max_{\|\boldsymbol{\eta}\|=1} S^*(t; \boldsymbol{\eta})$. Here, t is a pre-determined time point.

To find an optimal treatment regime, we first derive consistent estimators of $S^*(u; \boldsymbol{\eta})$ for any u. We make the uninformative censoring assumption: $\{T^*(1), T^*(0)\} \perp\!\!\!\perp C|A, \boldsymbol{X}$, where "$\perp\!\!\!\perp$" means "independent of". Let $S_C(t|a, \boldsymbol{x})$ denote the survival function of the censoring time given $A = a$ and $\boldsymbol{X} = \boldsymbol{x}$. If we were able to observe the $g_{\boldsymbol{\eta}}$-specific potential counting process $N_i^*(g_{\boldsymbol{\eta}};s)$ and at risk process $Y_i^*(g_{\boldsymbol{\eta}};s)$, an intuitive estimator for $S^*(u; \boldsymbol{\eta})$ is the inverse probability of censoring weighted Kaplan-Meier estimator

$$\hat{S}^*(u;\boldsymbol{\eta})=\prod_{s\leq u}\left(1-\frac{\sum_{i=1}^{n}[dN_i^*\{g_{\boldsymbol{\eta}}(\boldsymbol{X}_i);s\}/S_C\{s|g_{\boldsymbol{\eta}}(\boldsymbol{X}_i),\boldsymbol{X}_i\}]}{\sum_{i=1}^{n}[Y_i^*\{g_{\boldsymbol{\eta}}(\boldsymbol{X}_i);s\}/S_C\{s|g_{\boldsymbol{\eta}}(\boldsymbol{X}_i),\boldsymbol{X}_i\}]}\right). \qquad (1)$$

However, because $N_i^*(g_{\boldsymbol{\eta}};s)$ and $Y_i^*(g_{\boldsymbol{\eta}};s)$ are generally not observable, $\hat{S}^*(u; \boldsymbol{\eta})$ is not computable based on the observed data. To obtain proper estimators that are computable from the observed data, we make the following two assumptions, which are widely used in the causal inference literature (Rubin, 1974): (i) stable unit treatment value assumption (SUTVA); i.e. $T = T^*(1)A + T^*(0)(1-A)$; and (ii) no unmeasured confounders assumptions; i.e. $\{T^*(1), T^*(0)\} \perp\!\!\!\perp A|\boldsymbol{X}$.

2.2. Estimation Procedure

Following Zhang et al. (2012b), we cast estimation of $S^*(u; \boldsymbol{\eta})$ in a missing data framework. By SUTVA, for those patients whose actually received treatment matches the treatment dictated by $g_{\boldsymbol{\eta}}$, $N_i^*(g_{\boldsymbol{\eta}};s)=N_i(s)$ and $Y_i^*(g_{\boldsymbol{\eta}};s)=Y_i(s)$, which are observed. For other patients, they are missing. This motivates us to modify the estimator given in (1) by incorporating inverse propensity score weighting. Formally, the weight for the ith patient is given by

$$w_{\boldsymbol{\eta}i}=\frac{I[A_i=I\{\boldsymbol{\eta}^T\tilde{\boldsymbol{X}}\geq 0\}]}{\pi(\boldsymbol{X}_i)A_i+\{1-\pi(\boldsymbol{X}_i)\}(1-A_i)}=\frac{A_iI(\boldsymbol{\eta}^T\tilde{\boldsymbol{X}}\geq 0)+(1-A_i)\{1-I(\boldsymbol{\eta}^T\tilde{\boldsymbol{X}}\geq 0)\}}{\pi(\boldsymbol{X}_i)A_i+\{1-\pi(\boldsymbol{X}_i)\}(1-A_i)}, \qquad (2)$$

where $\pi(\boldsymbol{X}_i) = P(A_i = 1|\boldsymbol{X}_i)$ is the propensity score. In practice, $\pi(\boldsymbol{X}_i)$ is known by design, as in a randomized clinical trial, or must be modeled and estimated from the data as in observational studies. In the latter case, a parametric model, say a logistic regression is usually used for estimating $\pi(\boldsymbol{X}_i)$, specifically,

$$\text{logit}\{\pi(\boldsymbol{X}_i;\boldsymbol{\theta})\}=\boldsymbol{\theta}^T\tilde{\boldsymbol{X}}_i, \quad (3)$$

where logit(z) = log{z/(1−z)}. Let $\hat{\boldsymbol{\theta}}$ denote the maximum likelihood estimator of $\boldsymbol{\theta}$, and define $\hat{\pi}(\boldsymbol{X}_i) = \exp(\hat{\boldsymbol{\theta}}^T\tilde{\boldsymbol{X}}_i)/\{1 + \exp(\hat{\boldsymbol{\theta}}^T\tilde{\boldsymbol{X}}_i)\}$. If the logistic regression model is correctly specified, $\hat{\boldsymbol{\theta}}$ is a consistent estimator of $\boldsymbol{\theta}$.

To derive the estimator for $S^*(u; \boldsymbol{\eta})$, we also need to estimate the censoring time survival function $S_C(s|A_i, \boldsymbol{X}_i)$. In many clinical studies with satisfactory follow-up, it is reasonable to assume that censoring times are independent of treatment assignment and covariates, i.e. independent censoring. Here, the Kaplan-Meier estimator for censoring times consistently estimates $S_C(s|A_i, \boldsymbol{X}_i)$. For some applications, the independent censoring assumption may be restrictive, but can be relaxed to a certain extent. For example, if censoring times are assumed to depend only on treatment assignment, the stratified Kaplan-Meier estimator can be used to estimate the treatment-specific censoring time survival function. For more general dependence, we can build a semiparametric model, say a proportional hazards model for censoring times, and obtain the model based estimator of $S_C(s|A_i, \boldsymbol{X}_i)$. For simplicity, from now on we make the independent censoring assumption and let $\hat{S}_C(\cdot)$ denote the Kaplan-Meier estimator for censoring times.

Let $\hat{w}_{\boldsymbol{\eta}i}$ denote the estimator of $w_{\boldsymbol{\eta}i}$ obtained by replacing $\pi(\boldsymbol{X}_i)$ with $\hat{\pi}(\boldsymbol{X}_i)$ in $w_{\boldsymbol{\eta}i}$. We propose the inverse propensity score weighted Kaplan-Meier estimator (IPSWKME) for $S^*(u; \boldsymbol{\eta})$ given by

$$\hat{S}_I(u;\boldsymbol{\eta})=\prod_{s\le u}\left\{1-\frac{\sum_{i=1}^{n}\hat{w}_{\boldsymbol{\eta}i}dN_i(s)}{\sum_{i=1}^{n}\hat{w}_{\boldsymbol{\eta}i}Y_i(s)}\right\}. \quad (4)$$

Note that the IPSWKME dose not depend on the Kaplan-Meier estimator $\hat{S}_C(\cdot)$ for censoring times, as it cancels from numerator and denominator under the independent censoring assumption. In Section 4, we show that $\hat{S}_I(u; \boldsymbol{\eta})$ is a consistent estimator of $S^*(u; \boldsymbol{\eta})$ under certain conditions. Based on $\hat{S}_I(u; \boldsymbol{\eta})$, the estimated optimal treatment regime to maximize t-year survival probability is given by $g(\boldsymbol{x};\hat{\boldsymbol{\eta}}_I^{\text{opt}})$, where $\hat{\boldsymbol{\eta}}_I^{\text{opt}}=\arg\max_{\|\boldsymbol{\eta}\|=1}\hat{S}_I(t;\boldsymbol{\eta})$.

The IPSWKME (4) relies on correct specification of the propensity score model. If it is misspecified, the IPSWKME is inconsistent. To improve the robustness of the IPSWKME, we propose augmented IPSWKME (AIPSWKME) by incorporating assumed model information. For example, we may posit a proportional hazards (PH) model (Cox, 1972) for the conditional cumulative hazard function of T by

$$\Lambda_T(t|A,\boldsymbol{X})=\Lambda_0(t)\exp\{\boldsymbol{\beta}^T(\boldsymbol{X}^T,A,A\boldsymbol{X}^T)^T\}, \quad (5)$$

where $\Lambda_0(t)$ is the baseline cumulative hazard function, and $\boldsymbol{\beta}$ is a $(2p + 1)$-dimensional parameter. The term $w_{\boldsymbol{\eta}i}dN_i^*\{g_{\boldsymbol{\eta}}(\boldsymbol{X}_i);s\}$ is augmented by

$$\begin{aligned}&w_{\boldsymbol{\eta}i}dN_i^*\{g_{\boldsymbol{\eta}}(\boldsymbol{X}_i);s\}\\ &\quad+(1-w_{\boldsymbol{\eta}i})E[dN_i^*\{g_{\boldsymbol{\eta}}(\boldsymbol{X}_i);s\}|\boldsymbol{X}_i]\\ &\quad=w_{\boldsymbol{\eta}i}dN_i^*\{g_{\boldsymbol{\eta}}(\boldsymbol{X}_i);s\}\\ &\quad+(1-w_{\boldsymbol{\eta}i})S_T(s|g_{\boldsymbol{\eta}}(\boldsymbol{X}_i),\boldsymbol{X}_i)S_C(s)d\Lambda_T(s|g_{\boldsymbol{\eta}}(\boldsymbol{X}_i),\boldsymbol{X}_i),\end{aligned}$$

where $S_T(s|A_i, \boldsymbol{X}_i)$ and $S_C(s)$ are the conditional survival functions of T and C, respectively. Similarly, the term $w_{\boldsymbol{\eta}i}Y_i^*\{g_{\boldsymbol{\eta}}(\boldsymbol{X}_i);s\}$ is augmented by $w_{\boldsymbol{\eta}i}Y_i^*\{g_{\boldsymbol{\eta}}(\boldsymbol{X}_i);s\}+(1-w_{\boldsymbol{\eta}i})S_T(s|g_{\boldsymbol{\eta}}(\boldsymbol{X}_i),\boldsymbol{X}_i)S_C(s)$. It can be shown that the two augmented terms have the so-called double robustness property, i.e. they are unbiased for $E[dN_i^*\{g_{\boldsymbol{\eta}}(\boldsymbol{X}_i);s\}|\boldsymbol{X}_i]$ and $E[Y_i^*\{g_{\boldsymbol{\eta}}(\boldsymbol{X}_i);s\}|\boldsymbol{X}_i]$, respectively, when either the propensity score model or the posited PH model is correctly specified. Therefore, we propose the AIPSWKME for $S^*(u; \boldsymbol{\eta})$ as

$$\hat{S}_A(u;\boldsymbol{\eta})=\prod_{s\le u}\left(1-\frac{\sum_{i=1}^n[\hat{w}_{\boldsymbol{\eta}i}dN_i(s)+(1-\hat{w}_{\boldsymbol{\eta}i})\hat{S}_T\{s|g_{\boldsymbol{\eta}}(\boldsymbol{X}_i),\boldsymbol{X}_i\}\hat{S}_C(s)d\hat{\Lambda}_T\{s|g_{\boldsymbol{\eta}}(\boldsymbol{X}_i),\boldsymbol{X}_i\}]}{\sum_{i=1}^n[\hat{w}_{\boldsymbol{\eta}i}Y_i(s)+(1-\hat{w}_{\boldsymbol{\eta}i})\hat{S}_T\{s|g_{\boldsymbol{\eta}}(\boldsymbol{X}_i),\boldsymbol{X}_i\}\hat{S}_C(s)]}\right), \quad (6)$$

where $\hat{S}_T(s|A_i, \boldsymbol{X}_i)$ is the estimated survival function of T based on the fitted PH model and $\hat{S}_C(s)$ is the Kaplan-Meier estimator for censoring times. Based on $\hat{S}_A(u; \boldsymbol{\eta})$, the estimated optimal treatment regime to maximize t-year survival probability is given by $g(\boldsymbol{x};\hat{\boldsymbol{\eta}}_A^{\text{opt}})$, where $\hat{\boldsymbol{\eta}}_A^{\text{opt}}=\arg\max_{\|\boldsymbol{\eta}\|=1}\hat{S}_A(t;\boldsymbol{\eta})$. The asymptotic properties of $\hat{S}_A(u; \boldsymbol{\eta})$ and $\hat{S}_A(t;\hat{\boldsymbol{\eta}}_A^{\text{opt}})$ are studied in Section 4.

2.3. Computational Aspects

Note that $\hat{S}_I(t; \boldsymbol{\eta})$ and $\hat{S}_A(t; \boldsymbol{\eta})$ are not smooth functions of $\boldsymbol{\eta}$. As an illustration, we plot $\hat{S}_I(t; \boldsymbol{\eta})$ and $\hat{S}_A(t; \boldsymbol{\eta})$ as functions of η_1 in Figure 2 for a simple example with one covariate and the intercept term in $\boldsymbol{\eta}$ being set as 1. The estimates are very jagged, and direct maximization of them with respect to $\boldsymbol{\eta}$ is challenging and may lead to local maximizers. From our simulation studies in Section 5, estimated survival probabilities following the obtained optimal treatment regimes may show substantial biases. As studied in Matsouaka et al. (2014), cross-validation may be used to correct the finite sample biases of the unsmoothed estimators, but it may increase the computational burden.

To reduce the biases of the estimators, we propose to smooth the estimators $\hat{S}_I(t; \boldsymbol{\eta})$ and $\hat{S}_A(t; \boldsymbol{\eta})$ using kernel smoothers. Specifically, we replace $g_{\boldsymbol{\eta}}(\boldsymbol{x}_i) = I\{\boldsymbol{\eta}^T\tilde{\boldsymbol{x}}_i \geq 0\}$ in $\hat{S}_I(t; \boldsymbol{\eta})$ and $\hat{S}_A(t; \boldsymbol{\eta})$ with $\tilde{g}_{\boldsymbol{\eta}}(\boldsymbol{x}_i) = \Phi(\boldsymbol{\eta}^T\tilde{\boldsymbol{x}}_i/h)$ to obtain the smoothed IPSWKME (SIPSWKME) $\hat{S}_I(t; \boldsymbol{\eta})$ and smoothed AIPSWKME (S-AIPSWKME) $\hat{S}_A(t; \boldsymbol{\eta})$, where $\Phi(s)$ is the cumulative distribution function for the standard normal distribution, and h is a bandwidth parameter that goes to zero as n goes to infinity. For bandwidth selection, we set $h = c_0 n^{-1/3}\mathrm{sd}(\boldsymbol{\eta}^T\tilde{\boldsymbol{X}})$, where c_0 is a constant and $\mathrm{sd}(\boldsymbol{v})$ is the sample standard deviation of $\boldsymbol{v}$. In our numerical studies, we found that $c_0 = 4^{1/3}$ generally gives good results for all scenarios. We plot in Figure 2 the smoothed estimates with the chosen bandwidth parameter for the same example. The smoothed estimates approximate the original estimates well and have unique maximizers around the true value $\eta_1 = 0.5$. Because the treatment regime $I(\eta^T\tilde{X} \geq 0)$ remains the same when η is multiplied by k for any $k > 0$, choosing the bandwidth h to be a function of η, in particular, h being proportional to $sd(\eta^T\tilde{X})$, ensures the scale-free property of the regime, as the constant k cancels in $\Phi(\eta^T\tilde{X}/h)$. As shown in Figure 2, although the resulting smoothed value function is not convex in η, it generally has a unique mode, and the maximizer of the smoothed value function is much easier to obtain compared to the unsmoothed counterpart. In all our numerical studies, the non-convexity of the smoothed value function does not cause any difficulty in the maximization procedure. Such a bandwidth parameter has been widely used in the nonparametric smoothing literature and ensures that the original and smoothed estimators have the same asymptotic distribution (e.g. Heller, 2007). Let $\tilde{\boldsymbol{\eta}}_I^{\mathrm{opt}}$ and $\tilde{\boldsymbol{\eta}}_A^{\mathrm{opt}}$ denote the maximizers of $\hat{S}_I(t; \boldsymbol{\eta})$ and $\hat{S}_A(t; \boldsymbol{\eta})$, respectively. Then the associated estimated optimal treatment regimes are $g(\boldsymbol{x};\tilde{\boldsymbol{\eta}}_I^{\mathrm{opt}})$ and $g(\boldsymbol{x};\tilde{\boldsymbol{\eta}}_A^{\mathrm{opt}})$. In our implementation, we first conduct the optimization without the norm-one constraint. Instead, we search the maximizer in the domain $-1 \leq \eta_j \leq 1$ for all j's and then we rescale η to have norm one. This does not change the estimated value function, $\hat{S}_I$ and $\hat{S}_A$, and their smoothed counterparts.

3. Estimation of Optimal Treatment Regime for Multiple Decision Time Points

We now extend the foregoing methods to estimation of optimal dynamic treatment regimes incorporating multiple decision points. For simplicity, we illustrate for the case of two decision points. Specifically, treatments can be given at baseline and at a fixed interim time point s, $0 < s < t$. For the ith patient, let $\boldsymbol{X}_{0i}$ denote his or her p_0-dimensional vector of baseline covariates and $A_{0i} \in \mathscr{A}_0 = \{0, 1\}$ denote the initial treatment received at baseline. If the patient survives beyond s and is not censored before s, let $\boldsymbol{X}_{1i}$ denote his or her p_1-dimensional vector of intermediate covariates collected by s after assigning treatment A_{0i} and $A_{1i} \in \mathscr{A}_1 = \{0, 1\}$ denote the follow-up treatment given at s. Thus, the observed data are $\{\boldsymbol{X}_{0i}, A_{0i}, \boldsymbol{X}_{1i}I(\tilde{T}_i > s), A_{1i}I(\tilde{T}_i > s), \tilde{T}_i, \delta_i, i = 1, \ldots, n\}$ and iid across i.

As for a single decision point, we consider a class of linear dynamic treatment regimes for simplicity, i.e. $\mathscr{G} = \{\boldsymbol{g}_{\boldsymbol{\eta}} = (g_0, g_1)\}$, where

$$\begin{aligned} g_0(\boldsymbol{x}_0;\boldsymbol{\eta}_0) &= I\{\boldsymbol{\eta}_0^T(1,\boldsymbol{x}_0^T)\geq 0\}, \\ g_1(\boldsymbol{x}_0,\boldsymbol{x}_1;\boldsymbol{\eta}_1) &= I\{\boldsymbol{\eta}_1^T(1,\boldsymbol{x}_0^T,g_0(\boldsymbol{x}_0;\boldsymbol{\eta}_0),\boldsymbol{x}_1^T))\geq 0\}, \end{aligned}$$

$\boldsymbol{\eta}_0$ is a (p_0+1)-dimensional parameter with $\|\boldsymbol{\eta}_0\| = 1$, and $\boldsymbol{\eta}_1$ is a (p_0+p_1+2)-dimensional parameter with $\|\boldsymbol{\eta}_1\| = 1$. Here, a patient following a treatment regime $\boldsymbol{g}_{\boldsymbol{\eta}}$ is given treatment $g_0(\boldsymbol{X}_0;\ \boldsymbol{\eta}_0)$ at baseline, and, if he or she survives beyond s and is not censored before s, is given treatment $g_1(\boldsymbol{X}_0, \boldsymbol{X}_1;\ \boldsymbol{\eta}_1)$ at s. For patients whose initial treatments coincide with those assigned by $g_0(\boldsymbol{X}_0;\ \boldsymbol{\eta}_0)$ and who die before s, their treatment assignments are consistent with the regime $\boldsymbol{g}_{\boldsymbol{\eta}}$. However, for patients whose initial treatments coincide with those assigned by $g_0(\boldsymbol{X}_0;\ \boldsymbol{\eta}_0)$ but who are censored before s, it is not known whether their treatment assignments at the second decision follow the regime $\boldsymbol{g}_{\boldsymbol{\eta}}$. Let $T^*(\boldsymbol{g}_{\boldsymbol{\eta}})$ denote the potential survival time for a patient if he or she were given treatment according to $\boldsymbol{g}_{\boldsymbol{\eta}}(\boldsymbol{X}_0, \boldsymbol{X}_1)$. We are interested in finding the optimal dynamic treatment regime

$\boldsymbol{g}_{\boldsymbol{\eta}}^{\text{opt}}=\{g_0(\boldsymbol{X}_0;\boldsymbol{\eta}_0^{\text{opt}}),g_1(\boldsymbol{X}_0,\boldsymbol{X}_1;\boldsymbol{\eta}_1^{\text{opt}})\}$ in $\mathscr{G}$ that maximizes the t-year survival probability $S^{*(2)}(t;\ \boldsymbol{\eta}) = E(P[T^*\{g_{\boldsymbol{\eta}}(\boldsymbol{X}_0, \boldsymbol{X}_1)\} > t|\boldsymbol{X}_0, \boldsymbol{X}_1])$. As is standard in the causal inference literature for studying dynamic treatment regimes (e.g., Murphy, 2003), we assume: (i) SUTVA, i.e. a patient's observed outcome agrees with the corresponding potential outcome if his or her actually received treatments are consistent with the assigned treatments and (ii) sequential randomization assumption (SRA), i.e. the treatment assignment at current stage only depends on the past received treatments and observed covariates, but not the potential outcomes. Under these two assumptions, the above defined t-year survival probability can be estimated from the observed data.

We propose a similar inverse propensity score weighted Kaplan-Meier estimator for the survival function $S^{*(2)}(u;\ \boldsymbol{\eta})$ given any treatment regime $\boldsymbol{g}_{\boldsymbol{\eta}}$. However, the derivation of proper weights becomes more difficult, as some patients may be censored before s and whether their received treatments follow the regime $\boldsymbol{g}_{\boldsymbol{\eta}}$ is unknown. To take this into account, we define the following new weight for patient i, $i = 1, \ldots, n$:

$$\hat{w}_{\boldsymbol{\eta}i}^{(2)}=\frac{I(\tilde{T}_i\leq s)\times\delta_i}{\hat{S}_C(\tilde{T}_i)}\times\frac{I\{A_{0i}=g_0(\boldsymbol{X}_{0i};\boldsymbol{\eta}_0)\}}{\pi_{A_0}(\boldsymbol{X}_{0i})}+\frac{I(\tilde{T}_i>s)}{\hat{S}_C(s)}\times\frac{I\{A_{0i}=g_0(\boldsymbol{X}_{0i};\boldsymbol{\eta}_0),A_{1i}=g_1(\boldsymbol{X}_{0i},g_0(\boldsymbol{X}_{0i};\boldsymbol{\eta}_0),X_{1i};\boldsymbol{\eta}_i)\}}{\hat{\pi}_{A_0}(\boldsymbol{X}_{0i})\times\hat{\pi}_{A_1}(\boldsymbol{X}_{0i},A_{0i},\boldsymbol{X}_{1i})},$$

where $\hat{\boldsymbol{\pi}}_{A_0}(\boldsymbol{X}_{0i}) = \hat{\boldsymbol{\pi}}_0(\boldsymbol{X}_{0i})A_{0i}+\{1-\hat{\boldsymbol{\pi}}_0(\boldsymbol{X}_{0i})\}(1-A_{0i})$, $\hat{\boldsymbol{\pi}}_{A_1}\boldsymbol{X}_{0i}, A_{0i}, \boldsymbol{X}_{1i}) = \hat{\boldsymbol{\pi}}_1(\boldsymbol{X}_{0i}, A_{0i}, \boldsymbol{X}_{1i})A_{1i}+\{1-\hat{\boldsymbol{\pi}}_1(\boldsymbol{X}_{0i}, A_{0i}, \boldsymbol{X}_{1i})\}(1-A_{1i})$, and $\hat{\boldsymbol{\pi}}_0(\boldsymbol{X}_{0i})$ and $\hat{\boldsymbol{\pi}}_1(\boldsymbol{X}_{0i}, A_{0i}, \boldsymbol{X}_{1i})$ are the estimates of the propensity scores $P(A_{0i} = 1|\boldsymbol{X}_{0i})$ and $P(A_{1i} = 1|\boldsymbol{X}_{0i}, A_{0i}, \boldsymbol{X}_{1i}, \hat{T}_i > s)$, respectively. In randomized studies, $\hat{\boldsymbol{\pi}}_0$ and $\hat{\boldsymbol{\pi}}_1$ are known by design, while in observational studies, they must be estimated, e.g. using logistic regression. The IPSWKME for $S^*(u;\ \boldsymbol{\eta})$ is given by

$$\hat{S}_I^2(u;\boldsymbol{\eta})=\prod_{v\leq u}\left\{1-\frac{\sum_{i=1}^n\hat{w}_{\boldsymbol{\eta}i}^{(2)}dN_i(v)}{\sum_{i=1}^n\hat{w}_{\boldsymbol{\eta}i}^{(2)}Y_i(v)}\right\}. \quad (7)$$

Let $\hat{\boldsymbol{\eta}}_I^{\mathrm{opt},(2)}=(\hat{\boldsymbol{\eta}}_{I,0}^{\mathrm{opt},(2)},\boldsymbol{\eta}_{I,1}^{\mathrm{opt},(2)})=\arg\max_{\|\boldsymbol{\eta}_0\|=1,\|\boldsymbol{\eta}_1\|=1}\hat{S}_I^{(2)}(t;\boldsymbol{\eta})$. Then the estimated optimal dynamic treatment regime is given by $\hat{g}_{\boldsymbol{\eta}}^{\mathrm{opt},(2)}=\{g_0(\boldsymbol{X}_0;\hat{\boldsymbol{\eta}}_{I,0}^{\mathrm{opt},(2)}),g_1(\boldsymbol{X}_0,\boldsymbol{X}_1;\hat{\boldsymbol{\eta}}_{I,1}^{\mathrm{opt},(2)})\}$.

To improve the finite sample performance of the IPSWKME, we again introduce kernel smoothing and replace the indicator functions $g_0(\boldsymbol{X}_{0i}; \boldsymbol{\eta}_0)$ and $g_1(\boldsymbol{X}_{0i}, \boldsymbol{X}_{1i}; \boldsymbol{\eta}_1)$ in $\hat{S}_I^{(2)}(u;\boldsymbol{\eta})$ by $\Phi\left\{\boldsymbol{\eta}_0^T(1,\boldsymbol{X}_{0i}^T)/h_0\right\}$ and $\Phi\left[\boldsymbol{\eta}_1^T\{1,\boldsymbol{X}_0^T,g_0(\boldsymbol{X}_0;\boldsymbol{\eta}_0),\boldsymbol{X}_1^T\}/h_1\right]$, where the bandwidth parameters h_0 and h_1 are chosen as before. Let $\tilde{S}_I^{(2)}(u;\boldsymbol{\eta})$ denote the resulting smoothed IPSWKME and $\tilde{\boldsymbol{\eta}}_I^{\mathrm{opt},(2)}$ denote the maximizer of $\tilde{S}_I^{(2)}(t;\boldsymbol{\eta})$. To improve the robustness of IPSWKME, we can similarly derive the augmented IPSWKME based on a posited model for survival time, however, its formulation is very complicated and is not pursued here. Conceptually, the proposed IPSWKME can be generalized to accommodate more than two decision points. However, when there are more treatment decision points, the IPSWKME *Optimal Treatment Regimes for Survival Endpoint* may become less reliable because fewer patients will have assigned treatments consistent with a given dynamic treatment regime.

4. Asymptotic Properties

In this Section, we present the asymptotic properties of the proposed estimators in Theorems 1 – 3. Theorems 1 and 2 are for the cases with a single decision point while Theorem 3 is for the case with two decision points.

Theorem 1

Under conditions (A1)–(A6) in the Appendix*, if the propensity score model* (3) *is correctly specified, for any regime $g_{\boldsymbol{\eta}}$, we have, as $n \to \infty$,*

- **i.** $\hat{S}_I(u; \boldsymbol{\eta}) \to^p S^*(u; \boldsymbol{\eta})$ *for any* $0 < u \le t$;
- **ii.** $\sqrt{n}\{\hat{S}_I(u;\boldsymbol{\eta})-S^*(u;\boldsymbol{\eta})\}$ *converges weakly to a mean zero Gaussian process;*
- **iii.** $\sqrt{n}\{\hat{S}_I(t;\hat{\boldsymbol{\eta}}_I^{\mathrm{opt}})-S^*(t;\boldsymbol{\eta}^{\mathrm{opt}})\}\to^d N(0,\sum_I(t;\boldsymbol{\eta}^{\mathrm{opt}}))$, *where the expression of* $\Sigma_I(t; \boldsymbol{\eta}^{\mathrm{opt}})$ *is given in the* Appendix;
- **iv.** $\sqrt{n}\{\hat{S}_I(t;\hat{\boldsymbol{\eta}}_I^{\mathrm{opt}})-\tilde{S}_I(t;\tilde{\boldsymbol{\eta}}_I^{\mathrm{opt}})\}=o_p(1)$.

Theorem 2

Under condition (A1)–(A6) in the Appendix, *if either the propensity score model* (3) *or the proportional hazard model* (5) *is correctly specified, we have, as $n \to \infty$,*

- **i.** $\hat{S}_A(u; \boldsymbol{\eta}) \to^p S^*(u; \boldsymbol{\eta})$ *for any* $0 < u \le t$;
- **ii.** $\sqrt{n}\{\hat{S}_A(u;\boldsymbol{\eta})-S^*(u;\boldsymbol{\eta})\}$ *converges weakly to a mean zero Gaussian process;*
- **iii.** $\sqrt{n}\{\hat{S}_A(t;\hat{\boldsymbol{\eta}}_A^{\mathrm{opt}})-S^*(t;\boldsymbol{\eta}^{\mathrm{opt}})\}\to^d N(0,\sum_A(t;\boldsymbol{\eta}^{\mathrm{opt}}))$, *where the expression of* $\Sigma_A(t; \boldsymbol{\eta}^{\mathrm{opt}})$ *is given in the* Appendix;

iv. $\sqrt{n}\{\hat{S}_A(t;\hat{\boldsymbol{\eta}}_A^{\text{opt}})-\tilde{S}_A(t;\tilde{\boldsymbol{\eta}}_A^{\text{opt}})\}=o_p(1)$.

Theorem 3

Under certain regularity conditions, if the two propensity score models $\pi_0(\cdot)$ *and* $\pi_1(\cdot)$ *are correctly specified, for any regime* $g_{\boldsymbol{\eta}}$, *we have, as* $n \to \infty$,

i. $\hat{S}_I^{(2)}(u;\boldsymbol{\eta}) \to^p S^{*(2)}(u;\boldsymbol{\eta})$ *for any* $0 < u \le t$;

ii. $\sqrt{n}\{\hat{S}_I^{(2)}(u;\boldsymbol{\eta})-S^{*(2)}(u;\boldsymbol{\eta})\}$ *converges weakly to a mean zero Gaussian process;*

iii. $\sqrt{n}\{\hat{S}_I^{(2)}(t;\hat{\boldsymbol{\eta}}_I^{\text{opt},(2)})-S^*(t;\boldsymbol{\eta}^{\text{opt},(2)})\}\to^d N(0,\sum_I^{(2)}(t;\boldsymbol{\eta}^{\text{opt},(2)}))$, *where* $\boldsymbol{\eta}^{\text{opt},(2)}=(\boldsymbol{\eta}_0^{\text{opt}},\boldsymbol{\eta}_1^{\text{opt}})$;

iv. $\sqrt{n}\{\hat{S}_I^{(2)}(t;\hat{\boldsymbol{\eta}}_I^{\text{opt},(2)})-\tilde{S}_I^{(2)}(t;\tilde{\boldsymbol{\eta}}_I^{\text{opt},(2)})\}=o_p(1)$.

Here the asymptotic variances $\Sigma_I(t;\boldsymbol{\eta}^{\text{opt}})$, $\Sigma_A(t;\boldsymbol{\eta}^{\text{opt}})$ and $\sum_I^{(2)}(t;\boldsymbol{\eta}^{\text{opt},(2)})$ can be consistently estimated from the observed data using the usual plug-in method. The proofs of Theorems 1–3 are given in the Appendix.

5. Simulation Studies

We examine the finite sample performance of the proposed estimators by simulation. We first consider scenarios with a single treatment decision time point at baseline. For each patient, baseline covariates X_1 and X_2 are independently and uniformly distributed on $(-2, 2)$. Given X_1 and X_2, the binary treatment indicator A is generated from the logistic model $\text{logit}\{\pi(X_1, X_2)\} = X_1 - 0.5X_2$. The survival time T is generated from a linear transformation model (Cheng et al., 1995), $h(T) = -0.5X_1 + A(X_1 - X_2) + \varepsilon$, where $h(s) = \log(e^s - 1) - 2$ is an increasing function, and the error term ε follows some known distribution, either the extreme value distribution or the logistic distribution, which corresponds to a proportional hazards and proportional odds model, respectively. The covariate-independent censoring time C is uniformly distributed on $(0, C_0)$, where C_0 is chosen to achieve the censoring rate of 15% and 40%. The optimal treatment regime maximizing t-year survival probability is $g_{\boldsymbol{\eta}}^{\text{opt}}(X_1,X_2)=I\{X_1-X_2 \ge 0\}$ for any t. We search the optimal treatment regime in the class of regimes given by $\mathscr{G} = \{g_{\boldsymbol{\eta}} : g_{\boldsymbol{\eta}}(X_1, X_2) = I\{\eta_0 + \eta_1X_1 + \eta_2X_2 \ge 0\}, \boldsymbol{\eta} = (\eta_0, \eta_1, \eta_2)^T\}$, which contains the true optimal treatment regime with $\boldsymbol{\eta}^{\text{opt}} = (0, 0.707, -0.707)$.

To implement the proposed estimators, it is necessary to posit a model for the propensity scores. We consider both a correctly specified model, $\text{logit}\{\pi_A(X_1, X_2)\} = \theta_0+\theta_1X_1+\theta_2X_2$, and a misspecified model, $\text{logit}\{\pi_A(X_1, X_2)\} = \theta_0$. For the augmented estimators, we must posit a model for the survival time T. Here, we always use the proportional hazard model $\lambda(t/X_1, X_2) = \lambda_0(t)\exp\{\beta_{11}X_1+\beta_{12}X_2+A(\beta_{20}+\beta_{21}X_1+\beta_{22}X_2)\}$. Note that when ε follows the extreme value distribution, the posited survival model is correctly specified. On the other hand, when ε follows the logistic distribution, this model is misspecified. We compare the performance of the IPSWKME ($\hat{S}_I$) and AIPSWKME ($\hat{S}_A$), as well as their smoothed versions, S-IPSWKME ($\tilde{S}_I$) and S-AIPSWKME ($\tilde{S}_A$), under different combinations of the assumed propensity score (PS) model, error term distribution, censoring rate, sample size (n = 250 or 500) and time point of interest (t = 1 or 2). For each scenario, we ran 1000 replications and used a genetic algorithm to do the optimization, which is implemented by the R function genoud within the package rgenoud (Mebane, Jr. and Sekhon, 2011).

We report results for the scenarios with $n = 250$ and $t = 2$, which are given in Tables 1 and 2 for the extreme value error and logistic error distributions, respectively. Results for other scenarios are similar. In the tables, we report the mean of estimated $\boldsymbol{\eta}^{\text{opt}}$, the mean of estimated t-year survival probability following the estimated optimal treatment regime, namely the estimated optimal t-year survival probability (denoted by $\hat{S}(\hat{\boldsymbol{\eta}}^{\text{opt}})$), the mean of estimated standard error of $\hat{S}(\hat{\boldsymbol{\eta}}^{\text{opt}})$ using the plug-in method based on the asymptotic variances established in Theorems 1–2 (denoted by SE), the empirical coverage probability of 95% confidence interval for the t-year survival probability following the true optimal treatment regime $S(\boldsymbol{\eta}^{\text{opt}})$ (denoted by CP), the mean of simulated true t-year survival probability following the estimated optimal treatment regime (denoted by $S(\hat{\boldsymbol{\eta}}^{\text{opt}})$), and the mean of misclassification rate by comparing the true and estimated optimal treatment regimes (denoted by MR). The numbers given in parenthesis are the standard deviations of the corresponding estimates. Here, $S(\boldsymbol{\eta}^{\text{opt}})$ and $S(\hat{\boldsymbol{\eta}}^{\text{opt}})$ are computed using simulated survival times following the given treatment regime based on a large random sample of 5×10^6 patients. We have $S(\boldsymbol{\eta}^{\text{opt}}) = 0.605$ for the extreme value error distribution and $S(\boldsymbol{\eta}^{\text{opt}}) = 0.672$ for the logistic distribution. The misclassification rate for one simulation is calculated as the proportion of patients for which the true and estimated optimal treatment regimes do not select the same treatment.

From the results, when the PS model is correctly specified, all estimators of $\boldsymbol{\eta}^{\text{opt}}$ have relatively small biases, in particular, the mean of $\hat{\eta}_0^{\text{opt}}$ is close to zero while the mean ratio of $\hat{\eta}_1^{\text{opt}}$ to $\hat{\eta}_2^{\text{opt}}$ is very close to negative one. The means of simulated true t-year survival probability following the estimated optimal treatment regimes, i.e. $S(\hat{\boldsymbol{\eta}}^{\text{opt}})$, are all close to the true values. In addition, the estimates of $\boldsymbol{\eta}^{\text{opt}}$ based on the AIPSWKME and S-AIPSWKME of t-year survival probability generally have smaller standard deviation than those based on IPSWKME and S-IPSWKME. The unsmoothed IPSWKME and AIPSWKME of the optimal t-year survival probability have relatively large biases mainly due to the very jagged estimates of t-year survival probability, as illustrated in Figure 2, and as a consequence, the associated coverage probability of 95% confidence interval is much lower than the nominal level. The smoothed S-IPSWKME and S-AIPSWKME of the optimal t-year survival probability greatly reduce the biases and thus give the proper coverage probability. In addition, the unsmoothed and smoothed estimators of the optimal t-year survival probability have nearly the same standard deviation. When the PS model is misspecified, the IPSWKME and S-IPSWKME generally have relatively large biases as expected, while the AIPSWKME and S-AIPSWKME greatly reduce the biases and give much smaller MR. In particular, when the posited survival model is correctly specified under the extreme value error distribution, the S-AIPSWKME yields proper coverage probability. On the other hand, when the posited survival model is misspecified under the logistic error distribution, although the S-AIPSWKME is not consistent in general, it still gives small biases with reasonable coverage probability. Performance of the estimators improves as the censoring rate decreases and sample size increases.

We also compare the proposed method with the methods of Zhao et al. (2013) and Zhao et al. (2015). For the comparison with the method of Zhao et al. (2013), we consider randomized studies with known propensity scores, i.e. $\pi_A \equiv 0.5$, sample size $n = 250$, decision time point of interest $t_0 = 2$, and censoring rate of 15%. When implementing the method of Zhao et al. (2013), we set the threshold $\xi = 0, 0.1, \ldots, 0.6$ and find the associated treatment regime for each ξ value.

Table 3 summarizes the simulation results for the extreme value and logistic error distributions based on 1000 replications. The performance of the method of Zhao et al. (2013) depends on the choice of the threshold value ξ. For the extreme value error distribution, the best choice is $\xi = 0.4$, while for the logistic error distribution, the best choice is $\xi = 0.3$. In practice, the best threshold value to use is unknown and must be estimated from data, which may not be straightforward. Moreover, even with the best choice of ξ value, the performance of the method by Zhao et al. (2013) is still worse than that of our proposed smoothed estimators, S-IPSWKME and S-AIPSWKME, under all the considered settings.

For the comparison with the method of Zhao et al. (2015), we consider the same simulation settings as in Tables 1 and 2 with sample size $n = 250$, decision time point of interest $t_0 = 2$, and censoring rate of 15%. For both methods, we consider the augmented estimation. Table 4 summarizes the simulation results based on 1000 replications. The proposed methods and the method of Zhao et al. (2015) lead to comparable survival probabilities under the estimated treatment rules, while the proposed methods yield smaller misclassification rates under all the considered settings. In summary, the proposed methods demonstrate very competitive performance compared with existing approaches.

Next, we consider scenarios with two treatment decision time points, one at the baseline and the other at $s = 1$. The initial treatment assignment A_0 and the follow-up treatment assignment A_1, if applicable, are generated independently from a Bernoulli distribution with success probability of 0.5. A single baseline covariate X_0 is generated from a uniform distribution on (0, 4). To generate the survival time T, we first generate a time T_1 given A_0 and X_0 from an exponential distribution with the rate function $\lambda_1(A_0, X_0)$. The censoring time C is generated from a uniform distribution on $(0, C_0)$. If a patient is neither dead nor censored at time $s = 1$ (i.e. $\min(T_1, C) > 1$), we generate a single intermediate covariate X_1 for this patient as $X_1 = 0.5\, X_0 - 0.4(A_0 - 0.5) + e$, where e is uniformly distributed on (0, 2). Then we generate another time T_2 given A_0, A_1, X_0 and X_1 from an exponential distribution with the rate function $\lambda_2(A_0, A_1, X_0, X_1)$. The survival time T of interest is defined as $T = T_1$ if $T_1 \le 1$ and $T = 1 + T_2$ otherwise. The observed survival time is $\tilde{T} = \min(T, C)$ with the censoring indicator $\delta = I(T \le C)$. Here, C_0 is chosen to achieve censoring rates of 15% and 40%. We consider three scenarios for the rate functions λ_1 and λ_2: (i) $\lambda_1(A_0, X_0) = 0.5 \exp\{1.75(A_0 - 0.5)(X_0 - 2)\}$ and $\lambda_2(A_0, A_1, X_0, X_1) = 0.3 \exp\{2.5(A_1 - 0.4)(X_1 - 2) - A_0(X_1 - 2)\}$; (ii) $\lambda_1(A_0, X_0) = 0.1 \exp\{2(A_0 - 0.5)(X_0 - 2)\}$ and $\lambda_2(A_0, A_1, X_0, X_1) = 0.2$

$\exp\{3(A_1 - 0.4)(X_1 - 2) - 3(A_0 - 0.5)(X_0 - 2)\}$; (iii) $\lambda_1(A_0, X_0) = 0.2 \exp\{1.5(A_0 - 0.3)(X_0 - 3)\}$ and $\lambda_2(A_0, A_1, X_0, X_1) = 0.3 \exp\{2(A_1 - 0.5)(X_1 - 2) + 0.5(A_0 - 0.7)(X_0 - 1)\}$.

For the above three scenarios, the true optimal rule for maximizing t-year survival probability ($t > 1$) at time $s = 1$ is given by $g_1^{\text{opt}}(x_1)=I(2-x_1>0)$. However, the true optimal rule $g_0^{\text{opt}}(x_0)$ at time $s = 0$ is a complicated nonlinear function of x_0, which can be derived using backward induction as in Q-learning. In our implementation, for computation simplicity, we search for the optimal dynamic treatment regime in a class involving linear decision rules, specifically, $\mathscr{G}_{\boldsymbol{\eta}} = \{g_0(x_0) = I\{\eta_1 + \eta_2 x_0 > 0\}, g_1(x_1) = I\{\eta_3 + \eta_4 x_1 > 0\}, \|(\eta_1, \eta_2)\| = 1, \|(\eta_3, \eta_4)\| = 1\}$. Then, the true optimal rule $g_1^{\text{opt}}(x_1)$ at time $s = 1$ corresponds to $(\eta_3^{\text{opt}}, \eta_4^{\text{opt}})=(0.894, -0.447)$ for all three scenarios.

For scenarios (i) and (iii), we take $t = 3$, while for (ii) we take $t = 6$. We use simulation to find the true optimal rule at $s = 0$ in $\mathscr{G}_{\boldsymbol{\eta}}$ to maximize t-year survival probability. Specifically, we first generate X_0, and for a given (η_1, η_2), we set A_0 by the regime $g_0(X_0)$. Then, we generate X_1 given A_0 and X_0 the same way as in our design, and set A_1 by the optimal rule g_1^{opt}. Finally, we generate T_1 and T_2, and define T the same way as before. Using generated T's for a large random sample of 5×10^6 patients, we compute the associated empirical t-year survival probability. We find ($\eta_1^{\text{opt}}, \eta_2^{\text{opt}}$) to maximize the empirical t-year survival probability, which gives the true optimal rule g_0^{opt} in $\mathscr{G}_{\boldsymbol{\eta}}$. Here, we use grid search method to find ($\eta_1^{\text{opt}}, \eta_2^{\text{opt}}$). Since $\|(\eta_1^{\text{opt}}, \eta_2^{\text{opt}})\| = 1$, we only need to do grid search for η_1. We have $(\eta_1^{\text{opt}}, \eta_2^{\text{opt}})=(0.890, -0.456)$ and $S(3; \boldsymbol{\eta}^{\text{opt}}) = 0.567$ for scenario 1, $(\eta_1^{\text{opt}}, \eta_2^{\text{opt}})=(-0.891, 0.454)$ and $S(6; \boldsymbol{\eta}^{\text{opt}}) = 0.624$ for scenario 2, and $(\eta_1^{\text{opt}}, \eta_2^{\text{opt}})=(0.908, -0.419)$ and $S(3; \boldsymbol{\eta}^{\text{opt}}) = 0.702$ for scenario 3. Here $\boldsymbol{\eta}^{\text{opt}}=(\eta_1^{\text{opt}}, \eta_2^{\text{opt}}, \eta_3^{\text{opt}}, \eta_4^{\text{opt}})^T$ and $S(t; \boldsymbol{\eta}^{\text{opt}})$ is the t-year survival probability following the optimal dynamic treatment regime defined by $\boldsymbol{\eta}^{\text{opt}}$.

We compare the unsmoothed and smoothed estimators. For both estimators, the propensity score models $\boldsymbol{\pi}_0$ and $\boldsymbol{\pi}_1$ are assumed known as for randomized clinical trials. Simulation results for 1000 replications are summarized in Table 3. From the results, we observe: (i) both unsmoothed and smoothed estimation methods give nearly unbiased estimators of $\boldsymbol{\eta}^{\text{opt}}$, and the t-year survival probability following the estimated optimal treatment regime (denoted by $S(\hat{\boldsymbol{\eta}}^{\text{opt}})$ in the table) is very close to the t-year survival probability following the true optimal treatment regime $\boldsymbol{\eta}^{\text{opt}}$; (ii) the mean of estimated standard error (SE) of $\hat{S}(\hat{\boldsymbol{\eta}}^{\text{opt}})$ based on the established theory is close to the standard deviation of the estimates given in the parenthesis; (iii) The unsmoothed estimator for the t-year survival probability following the estimated optimal treatment regime (denoted by $\hat{S}(\hat{\boldsymbol{\eta}}^{\text{opt}})$) has relatively large bias and the associated coverage probability (CP) is below the nominal level; and (iv) the smoothed estimator for the t-year survival probability following the estimated optimal treatment regime has largely reduced bias and thus lead to proper coverage probability.

6. Application to ACTG 175

We illustrate the proposed methods for a single decision with the data from the ACTG Study 175 (Hammer et al., 1996). Subjects were randomized to four treatment groups with equal probability: zidovudine (ZDV) monotherapy, ZDV plus didanosine (ddI), ZDV plus zalcitabine (zal), and ddI monotherapy. A primary composite endpoint of interest is the time to having a larger than 50% decline in the CD4 count, or progressing to AIDS, or death, whichever comes first. From treatment-specific Kaplan-Meier curves, it can be clearly seen that treatments ZDV+ddI, ZDV+zal and ddI only are uniformly better than treatment ZDV only in terms of survival. In addition, treatments ZDV+ddI and ZDV+zal are overall the two best treatments giving the highest survival probabilities especially after day 400. For simplicity, we only consider two treatment options in our analysis, $A = 1$ for ZDV +ddI and $A = 0$ for ZDV+zal, which involves 1046 subjects. For each subject, there are 12 baseline clinical covariates; preliminary analysis results showed that Karnofsky score (Karnof), baseline CD4 count (CD40), and age (Age) are three important risk predictors and may have interaction effects with treatments. We include these three covariates in constructing treatment regimes. Our goal is to find the optimal treatment regime from the class of linear regimes defined by $\mathscr{G} = \{g_{\boldsymbol{\eta}} = I(\eta_0 + \eta_1 x_1 + \eta_2 x_2 + \eta_3 x_3 \geq 0) : \boldsymbol{\eta} = (\eta_0, \eta_1, \eta_2, \eta_3)^T, \|\boldsymbol{\eta}\| = 1\}$ to maximize t-year survival probability, x_1 is Karnof, x_2 is CD40, and x_3 is Age. Because the data come from a randomized study, we use a constant model for the propensity score and estimate this constant from data. For augmented estimation, we posit the proportional hazard model as given in (5). We estimate optimal treatment regimes at day $t = 400$, 600, 800 and 1000.

The estimated optimal treatment regimes and the associated t-year survival probabilities are presented in Table 6. We only present the results for S-IPSWKME and S-AIPSWKME, as they have better numerical performance than their nonsmoothed counterparts based on our simulation studies. The numbers in the columns of Intercept, Karnof, CD40 and Age are the parameter estimates $\tilde{\boldsymbol{\eta}}^{\text{opt}}$ defining the optimal treatment regimes, and $\tilde{S}(t; \tilde{\boldsymbol{\eta}}^{\text{opt}})$ is the estimated t-year survival probability following the estimated optimal treatment regime. From the Table, the estimated optimal treatment regime for an earlier time may be different from that for a later time. For example, comparing the obtained optimal treatment regimes for $t = 600$ and $t = 800$, the S-IPSWKME assigns a set of 353 patients to treatment 0 and another set of 583 patients to treatment 1 for both time points. However, it assigns a set of 52 patients to treatment 0 for $t = 600$ but to treatment 1 for $t = 800$. On the other hand, it assigns another set of 58 patients to treatment 1 for $t = 600$ but to treatment 0 for $t = 800$. For the S-AIPSWKME, the findings are similar. S-IPSWKME and S-AIPSWKME yield very different parameter estimates $\tilde{\boldsymbol{\eta}}^{\text{opt}}$. However, the corresponding optimal treatment regimes are similar. Using the results for day 600 as an example, among the 1046 subjects, there are only 57 subjects whose assigned treatments are different by the estimated optimal treatment regimes based on S-IPSWKME and S-AIPSWKME. In addition, the estimated t-year survival probabilities following the estimated optimal treatment regimes are nearly the same based on S-IPSWKME and S-AIPSWKME.

Next, we compare the estimated optimal regimes with the simple regimes that assign all subjects to the same treatment. Specifically, we construct 95% Wald-type confidence intervals for the difference between the estimated *t*-year survival probabilities under the estimated optimal treatment regimes and the simple regimes based on the derived asymptotic normal distribution. The results are also given in Table 6. The confidence intervals either stay above zero or zero is very close to the left end point of the intervals when it is contained. This implies that the increase in value realized by following the estimated optimal treatment regimes comparing with the simple regimes is significant or at least marginally significant. The Kaplan-Meier curves for patients following the estimated optimal treatment regimes (not shown here) are all uniformly better than those for each single treatment.

We have also estimated the optimal treatment regimes using the proposed methods based all twelve covariates when smoothing is and is not employed. We do not report on this here for brevity; however, we note that the results for smoothed estimators when using three versus twelve covariates are comparable, demonstrating the adaptivity of the smoothed estimators to incorporating relatively many covariates. The unsmoothed estimators can lead to slightly different optimal treatment rules but with similar estimated survival probabilities. In addition, the estimated survival probabilities show relatively larger differences between the cases with three and twelve covariates, which is likely due to the instability in maximizing the unsmoothed value functions.

7. Discussion

We have proposed Kaplan-Meier type estimators for the survival function of patients following a given (dynamic) treatment regime and introduce kernel smoothing to improve their performance. An optimal (dynamic) treatment regime within a class of prespecified treatment regimes may then be estimated by maximizing the estimator of the associated *t*-year survival probability. We consider the case when there are two treatment options at each decision time point. However, the proposed methods can be generalized to incorporate multiple treatment options at each decision by defining a treatment regime using multiple indexes instead of a single indicator function. In addition, current methods find the optimal (dynamic) treatment regime to maximize *t*-year survival probability, which can also be generalized to maximize other clinical outcomes of interest. Specifically, using the IPSWKME, $\hat{S}_I(\cdot;\, \boldsymbol{\eta})$, as an illustration, we can find the optimal treatment regime to maximize $f\{\hat{S}_I(\cdot;\, \boldsymbol{\eta})\}$, where f is a specified function of interest; e.g., $f\{\hat{S}_I(\cdot;\boldsymbol{\eta})\}=\int_0^L \hat{S}_I(u;\boldsymbol{\eta})du$ corresponds to restricted mean survival time under a given treatment regime. Likewise $f\{\hat{S}_I(\cdot;\, \boldsymbol{\eta})\} = \sup\{u : \hat{S}_I(u;\, \boldsymbol{\eta}) \geq 0.5\}$ corresponds to the median survival time under a given treatment regime.

In this paper, we study the asymptotic distributions of the estimated value function under the derived optimal treatment regimes. The asymptotic properties of $\hat{\eta}$ in the treatment regime function are very challenging to obtain. The convergence rate of $\hat{\eta}$ is slower than the classical $n^{1/2}$-rate due to the indicator function $I(\boldsymbol{\eta}^T \tilde{X} \geq 0)$, and the resulting limiting distribution is not standard. Matsouaka et al. (2014) studied a special case where the estimated value function depends on a single threshold value and showed that the estimator of the threshold that maximizes the estimated value function has the $n^{1/3}$-rate. We conjecture that our estimator $\hat{\eta}$ should also have $n^{1/3}$-rate. This is an interesting problem that warrants future research.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

The authors are grateful to two referees and an Associate Editor for their thoughtful and suggestive comments, which have helped to greatly improve on an earlier manuscript. The work was partially supported by National Institutes of Health grants R01 CA140632 and P01 CA142538.

References

Bai X, Tsiatis AA, Lu W, Song R. Optimal treatment regimes for survival endpoints using a locally-efficient doubly-robust estimator from a classification perspective. Technical Report. 2014

Cheng SC, Wei LJ, Ying Z. Analysis of transformation models with censored data. Biometrika. 1995; 82(4):835–845.

Cox DR. Regression models and life-tables. Journal of the Royal Statistical Society Series B (Methodological). 1972; 34(2):187–220.

Goldberg Y, Kosorok MR. Q-learning with censored data. Annals of Statistics. 2012; 40:529–560. [PubMed: 22754029]

Hammer SM, Katzenstein DA, Hughes MD, Gundacker H, Schooley RT, Haubrich RH, Henry WK, Lederman MM, Phair JP, Niu M, Hirsch MS, Merigan TC. A trial comparing nucleoside monotherapy with combination therapy in hiv-infected adults with cd4 cell counts from 200 to 500 per cubic millimeter. New England Journal of Medicine. 1996; 335(15):1081–1090. [PubMed: 8813038]

Heller G. Smoothed rank regression with censored data. Journal of the American Statistical Association. 2007; 102(478):552–559.

Matsouaka RA, Li J, Cai T. Evaluating marker-guided treatment selection strategies. Biometrics. 2014; 70:489–499. [PubMed: 24779731]

Mebane WR Jr, Sekhon JS. Genetic optimization using derivatives: The rgenoud package for R. Journal of Statistical Software. 2011; 42(11):1–26.

Murphy SA. Optimal dynamic treatment regimes. Journal of the Royal Statistical Society: Series B (Statistical Methodology). 2003; 65(2):331–355.

Murphy SA. An experimental design for the development of adaptive treatment strategies. Statistics in medicine. 2005; 24(10):1455–1481. [PubMed: 15586395]

Robins, JM. Optimal structural nested models for optimal sequential decisions. Proceedings of the second seattle Symposium in Biostatistics; Springer; 2004. p. 189-326.

Rubin DB. Estimating causal effects of treatments in randomized and nonrandomized studies. Journal of educational Psychology. 1974; 66(5):688–701.

Shorack, GR., Wellner, JA. Empirical processes with applications to statistics. Vol. 59. SIAM; 2009.

Watkins C, Dayan P. Q-learning. Machine Learning. 1992; 8(3–4):279–292.

Watkins, CJ. PhD thesis. University of Cambridge; England: 1989. Learning from delayed rewards.

Zhang B, Tsiatis AA, Davidian M, Zhang M, Laber EB. Estimating optimal treatment regimes from a classification perspective. Stat. 2012a; 1(1):103–114. [PubMed: 23645940]

Zhang B, Tsiatis AA, Laber EB, Davidian M. A robust method for estimating optimal treatment regimes. Biometrics. 2012b; 68(4):1010–1018. [PubMed: 22550953]

Zhang B, Tsiatis AA, Laber EB, Davidian M. Robust estimation of optimal dynamic treatment regimes for sequential treatment decisions. Biometrika. 2013; 100:681–694.

Zhao L, Tian L, Cai T, Claggett B, Wei LJ. Effectively selecting a target population for a future comparative study. Journal of the American Statistical Association. 2013; 108:527539.

Zhao Y, Kosorok MR, Zeng D. Reinforcement learning design for cancer clinical trials. Statistics in Medicine. 2009; 28(26):3294–3315. [PubMed: 19750510]

Zhao Y, Zeng D, Laber E, Song R, Yuan M, Kosorok M. Doubly robust learning for estimating individualized treatment with censored data. Biometrika. 2015; 102:151–168. [PubMed: 25937641]

Zhao Y, Zeng D, Rush AJ, Kosorok MR. Estimating individualized treatment rules using outcome weighted learning. Journal of the American Statistical Association. 2012; 107(499):1106–1118. [PubMed: 23630406]

A. Proof of Theorems

To establish the asymptotic results given in Theorems 1–2, we need to assume some regularity conditions. Recall that a working logistic model (3) is assumed for the propensity scores with parameters $\boldsymbol{\theta}$ for the IPSWKME and a working proportional hazards model (5) is further assumed for the survival time T for the AIPSWKME with parameters $\boldsymbol{\beta}$ and Λ_0. Let $\boldsymbol{\nu}_{Ai}=(\boldsymbol{X}_i^T, A_i, A_i\boldsymbol{X}_i^T)^T$ and $\boldsymbol{\nu}_{\boldsymbol{\eta}i}=(\boldsymbol{X}_i^T, g_{\boldsymbol{\eta}}(\boldsymbol{X}_i), g_{\boldsymbol{\eta}}(\boldsymbol{X}_i)\boldsymbol{X}_i^T)^T$. Define

$$K_1^I(\boldsymbol{X}, A, \tilde{T}, \delta; \boldsymbol{\eta})=\int_0^t \frac{(2A-1)dN(u)}{\pi^* E\{w_{\boldsymbol{\eta}}^* Y(u)\}},$$
$$K_2^I(\boldsymbol{X}, A, \tilde{T}, \delta; \boldsymbol{\eta})=\int_0^t \frac{(2A-1)Y(u)E[\{(2A-1)g_{\boldsymbol{\eta}}(\boldsymbol{X})+(1-A)\}dN(u)]}{[\pi^* E\{w_{\boldsymbol{\eta}}^* Y(u)\}]^2},$$

where $w_{\boldsymbol{\eta}}^*=[Ag_{\boldsymbol{\eta}}(\boldsymbol{X})+(1-A)\{1-g_{\boldsymbol{\eta}}(\boldsymbol{X})\}]/\pi^*$ and $\boldsymbol{\pi}^* = \boldsymbol{\pi}(\boldsymbol{X};\ \boldsymbol{\theta}^*)A+\{1-\boldsymbol{\pi}(\boldsymbol{X};\ \boldsymbol{\theta}^*)\}(1-A)$. In addition, define

$$K_1^A(\boldsymbol{X}, A, \tilde{T}, \delta; \boldsymbol{\eta})=\int_0^t \frac{J_1^A(u)-J_0^A(u)}{E\left[\{L_1^A(u)-L_0^A(u)\}g_{\boldsymbol{\eta}}(\boldsymbol{X})+L_0^A(u)\right]},$$
$$K_2^A(\boldsymbol{X}, A, \tilde{T}, \delta; \boldsymbol{\eta})=\int_0^t \frac{\{L_1^A(u)-L_0^A(u)\}E\left[\{J_1^A(u)-J_0^A(u)\}g_{\boldsymbol{\eta}}(\boldsymbol{X})+J_0^A(u)\right]}{\left(E\left[\{L_1^A(u)-L_0^A(u)\}\ g_{\boldsymbol{\eta}}(\boldsymbol{X})+L_0^A(u)\right]\right)^2},$$

$$\begin{aligned} J_k^A(u)=&\frac{1-k-(-1)^k A}{\pi^*}dN(u)\\ &+e_k\ \left(1-\frac{1-k-(-1)^k A}{\pi^*}\right)\ \exp\{\\ &-\Lambda_0^*(u)e_k\}\ S_C(u)d\Lambda_0^*(u), L_k^A(u)=\frac{1-k-(-1)^k A}{\pi^*}Y(u)+\left(1-\frac{1-k-(-1)^k A}{\pi^*}\right)\ \exp\{\end{aligned}$$

where $-\Lambda_0^*(u)e_k\}\ S_C(u)$ $e_k =$ exp $\{\boldsymbol{\beta}^{*T}(\boldsymbol{X}^T, k, k\boldsymbol{X}^T)^T\}$, $k = 0, 1$. We assume the following conditions.

A1 The covariates $\boldsymbol{X}$ are bounded.

A2 The propensity score $\boldsymbol{\pi}(\boldsymbol{X})$ is bounded away from 0 and 1 for all possible values of $\boldsymbol{X}$.

A3 The equation $E\left[\left\{A-\frac{\exp(\boldsymbol{\theta}^T\tilde{\boldsymbol{X}})}{1+\exp(\boldsymbol{\theta}^T\tilde{\boldsymbol{X}})}\right\}\tilde{\boldsymbol{X}}\right]=0$ has a unique solution $\boldsymbol{\theta}^*$.

A4 The equation

$$E\left(\int_0^\tau\left[\boldsymbol{\nu}_{Ai}-\frac{E\left\{Y_i(s)\exp(\boldsymbol{\beta}^T\boldsymbol{\nu}_{Ai})\boldsymbol{\nu}_{Ai}\right\}}{E\left\{Y_i(s)\exp(\boldsymbol{\beta}^T\boldsymbol{\nu}_{Ai})\right\}}\right]\times dN_i(s)\right)=0.$$

has a unique solution $\boldsymbol{\beta}^*$, where $\boldsymbol{\tau} > t$ is a prespecified time point satisfying $P(\tilde{T}_i \geq \boldsymbol{\tau}) > 0$. Let $\Lambda_0^*(u)=E[\int_0^u dN_i(s)/E\{Y_i(s)\exp(\boldsymbol{\beta}^{*T}\boldsymbol{\nu}_{Ai})\}]$ and it satisfies $\Lambda_0^*(\tau)<\infty$.

A5 $\sup_{\|\boldsymbol{\eta}\|=1}E[\{K_j^I(\boldsymbol{X},A,\tilde{T},\delta;\boldsymbol{\eta})\}^2]<\infty$ and $\sup_{\|\boldsymbol{\eta}\|=1}E[\{K_j^A(\boldsymbol{X},A,\tilde{T},\delta;\boldsymbol{\eta})\}^2]<\infty$, $j = 1, 2$.

A6 $nh \to \infty$ and $nh^4 \to 0$ as $n \to \infty$.

Under assumed regularity conditions A1 – A4, we have the following asymptotic representations:

$$\sqrt{n}(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}^*)=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\phi_{1i}+o_p(1), \qquad \sqrt{n}(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}^*)=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\phi_{2i}+o_p(1),$$
$$\sqrt{n}\{\hat{\Lambda}_0(u)-\Lambda_0^*(u)\}=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\phi_{3i}(u)+o_p(1), \qquad \sqrt{n}\{\hat{S}_C(u)-S_C(u)\}=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\phi_{4i}(u)+o_p(1),$$

where $\boldsymbol{\phi}_{1i}$'s and $\boldsymbol{\phi}_{2i}$'s are independently and identically distributed mean-zero vectors, and $\boldsymbol{\phi}_{3i}(u)$ and $\boldsymbol{\phi}_{4i}(u)$ are independent mean-zero processes. Moreover, consistent estimators $\hat{\boldsymbol{\phi}}_{1i}$, $\hat{\boldsymbol{\phi}}_{2i}$, $\hat{\boldsymbol{\phi}}_{3i}(u)$ and $\hat{\boldsymbol{\phi}}_{4i}(u)$ of $\boldsymbol{\phi}_{1i}$, $\boldsymbol{\phi}_{2i}$, $\boldsymbol{\phi}_{3i}(u)$ and $\boldsymbol{\phi}_{4i}(u)$ can be easily obtained.

In the following, we give a sketch for the proof of Theorem 1. The detailed proofs for Theorems 1–2 are provided in the Supplementary Appendix.

A.1. Proof of Theorem 1

For any given regime $g_{\boldsymbol{\eta}}$, we first derive the asymptotic properties for the corresponding inverse propensity score weighted (IPSW) Nelson-Aalen estimator. Specifically,

$$\hat{\Lambda}_I(u;\boldsymbol{\eta}) \equiv \hat{\Lambda}_I(u;\boldsymbol{\eta},\hat{\boldsymbol{\theta}})=\int_0^u\frac{\sum_{i=1}^n\hat{w}_{\boldsymbol{\eta}i}dN_i(s)}{\sum_{i=1}^n\hat{w}_{\boldsymbol{\eta}i}Y_i(s)}. \qquad \text{(A.1)}$$

It is easy to show that $\hat{S}_I(u;\ \boldsymbol{\eta})$ and $\exp\{-\hat{\Lambda}_I(u;\ \boldsymbol{\eta})\}$ are asymptotically equivalent for any given $\boldsymbol{\eta}$. Therefore, the asymptotic properties of $\hat{S}_I(u;\ \boldsymbol{\eta})$ easily follows those of $\hat{\Lambda}_I(u;\ \boldsymbol{\eta})$.

When the propensity score model is correctly specified, we have $\boldsymbol{\theta}^* = \boldsymbol{\theta}$ and $w_{\boldsymbol{\eta}i}^*=w_{\boldsymbol{\eta}i}$. Then $n^{-1}\sum_{i=1}^n\hat{w}_{\boldsymbol{\eta}i}Y_i(s)\to_p E\{w_{\boldsymbol{\eta}i}Y_i(s)\}=E[Y^*\{g_{\boldsymbol{\eta}}(\boldsymbol{X});s\}]$ uniformly for $s\in[0,\ \boldsymbol{\tau}]$ as $n\to\infty$. Similarly, we have $n^{-1}\sum_{i=1}^n\hat{w}_{\boldsymbol{\eta}i}dN_i(s)\to_p E\{w_{\boldsymbol{\eta}i}dN_i(s)\}=E[dN^*\{g_{\boldsymbol{\eta}}(\boldsymbol{X});s\}]$ uniformly for $s\in[0,\ \boldsymbol{\tau}]$ as $n\to\infty$. Therefore,

$$\begin{aligned}\hat{\Lambda}_I(u;\boldsymbol{\eta})\to_p\int_0^u\frac{E[dN^*\{g_{\boldsymbol{\eta}}(\boldsymbol{X});s\}]}{E[Y^*\{g_{\boldsymbol{\eta}}(\boldsymbol{X});s\}]}=\int_0^u\frac{S_C(s)dP[T^*\{g_{\boldsymbol{\eta}}(\boldsymbol{X})\}\leq s]}{S_C(s)P[T^*\{g_{\boldsymbol{\eta}}(\boldsymbol{X})\}\geq s]}\\ =-\log\{S^*(u;\boldsymbol{\eta})\} \equiv \Lambda^*(u;\boldsymbol{\eta}),\end{aligned}$$

which establish the consistency given in (i) of Theorem 1.

Next, we derive the asymptotic distribution of $\Lambda_I(u; \boldsymbol{\eta})$. By applying the first-order Taylor expansion of $\hat{\Lambda}_I(u; \boldsymbol{\eta})$ with respect to parameter $\boldsymbol{\theta}$ and some empirical process approximation techniques, we have

$$\begin{aligned}\sqrt{n}\{\hat{\Lambda}_I(u;\boldsymbol{\eta})-\Lambda^*(u;\boldsymbol{\eta})\}&=n^{-1/2}\sum_{i=1}^{n}\left(\int_0^u \frac{w_{\boldsymbol{\eta}i}dM_i^*\{g_{\boldsymbol{\eta}}(\boldsymbol{X});s\}}{E[Y^*\{g_{\boldsymbol{\eta}}(\boldsymbol{X});s\}]}+D_1(u)^T\phi_{1i}\right)+o_p(1)\\ &\equiv n^{-1/2}\sum_{i=1}^{n}\zeta_i(u;\boldsymbol{\eta})+o_p(1),\end{aligned}$$

where $M_i^*\{g_{\boldsymbol{\eta}}(\boldsymbol{X});s\}=N_i^*\{g_{\boldsymbol{\eta}}(\boldsymbol{X});s\}-\int_0^s Y_i^*\{g_{\boldsymbol{\eta}}(\boldsymbol{X});v\}d\Lambda^*(v;\boldsymbol{\eta})$ is a mean-zero martingale process and $D_1(u) = \lim_{n\to\infty} \partial\hat{\Lambda}_I(u; \boldsymbol{\eta}, \boldsymbol{\theta})/\partial\boldsymbol{\theta}$. By delta method, we have $\sqrt{n}\{\hat{S}_I(u;\boldsymbol{\eta})-S^*(u;\boldsymbol{\eta})\}=-S^*(u;\boldsymbol{\eta})n^{-1/2}\sum_{i=1}^{n}\zeta_i(u;\boldsymbol{\eta})+o_p(1)$, which converges weakly to a mean-zero Gaussian process by applying the empirical process theory. This proves (ii) of Theorem 1.

Since $\hat{\boldsymbol{\eta}}_I^{\text{opt}}$ is the maximizer of $\hat{S}_I(t; \boldsymbol{\eta})$ and $\boldsymbol{\eta}^{\text{opt}}$ is the maximizer of $S^*(t; \boldsymbol{\eta})$, following the similar arguments in Zhang et al. (2012b), we have

$$\sqrt{n}\{\hat{S}_I(t;\hat{\boldsymbol{\eta}}_I^{\text{opt}})-S^*(t;\boldsymbol{\eta}^{\text{opt}})\}-\sqrt{n}\{\hat{S}_I(t;\boldsymbol{\eta}^{\text{opt}})-S^*(t;\boldsymbol{\eta}^{\text{opt}})\}=o_p(1).$$

It follows that $\sqrt{n}\{\hat{S}_I(t;\hat{\boldsymbol{\eta}}_I^{\text{opt}})-S^*(t;\boldsymbol{\eta}^{\text{opt}})\}\to^d N(0,\sum_I(t;\boldsymbol{\eta}^{\text{opt}}))$, where $\sum_I(t;\boldsymbol{\eta}^{\text{opt}})=\{S^*(t;\boldsymbol{\eta}^{\text{opt}})\}^2E\{\zeta_i^2(t;\boldsymbol{\eta}^{\text{opt}})\}$. This proves (iii) of Theorem 1.

Finally, we show that $\hat{S}_I(t;\hat{\boldsymbol{\eta}}_I^{\text{opt}})$ and $\tilde{S}_I(t;\tilde{\boldsymbol{\eta}}_I^{\text{opt}})$ are asymptotically equivalent. For any given $\boldsymbol{\eta}$, we have

$$\sqrt{n}\{\tilde{\Lambda}_I(t;\boldsymbol{\eta})-\hat{\Lambda}_I(t;\boldsymbol{\eta})\}=\sqrt{n}\times\frac{1}{n}\sum_{i=1}^{n}\left\{\Phi\left(\frac{\boldsymbol{\eta}^T\boldsymbol{X}_i}{h}\right)-I\left(\boldsymbol{\eta}^T\boldsymbol{X}_i\ge 0\right)\right\}\times K_1^I(\boldsymbol{X}_i,A_i,\tilde{T}_i,\delta;\boldsymbol{\eta}) \tag{A.2}$$

$$\begin{aligned}&+\sqrt{n}\times\frac{1}{n}\sum_{i=1}^{n}\left\{\Phi\left(\frac{\boldsymbol{\eta}^T\boldsymbol{X}_i}{h}\right)-I\left(\boldsymbol{\eta}^T\boldsymbol{X}_i\ge 0\right)\right\}\times K_2^I(\boldsymbol{X}_i,A_i,\tilde{T}_i,\delta;\boldsymbol{\eta})\\ &+o_p(1).\end{aligned} \tag{A.3}$$

Following the similar arguments in Heller (2007), we can show that $\sup_{\|\boldsymbol{\eta}\|=1} |(A.2)| = o_p(1)$ and $\sup_{\|\boldsymbol{\eta}\|=1} |(A.3)| = o_p(1)$. Therefore, we have $\sqrt{n}\{\tilde{\Lambda}_I(t;\boldsymbol{\eta})-\hat{\Lambda}_I(t;\boldsymbol{\eta})\}=o_p(1)$ uniformly in $\boldsymbol{\eta}$, which implies $\sqrt{n}\{\tilde{S}_I(t;\boldsymbol{\eta})-\hat{S}_I(t;\boldsymbol{\eta})\}=o_p(1)$ uniformly in $\boldsymbol{\eta}$. It follows that $\sqrt{n}\{\tilde{S}_I(t;\tilde{\boldsymbol{\eta}}_I^{\text{opt}})-\hat{S}_I(t;\hat{\boldsymbol{\eta}}_I^{\text{opt}})\}=o_p(1)$, which proves (iv) of Theorem 1.

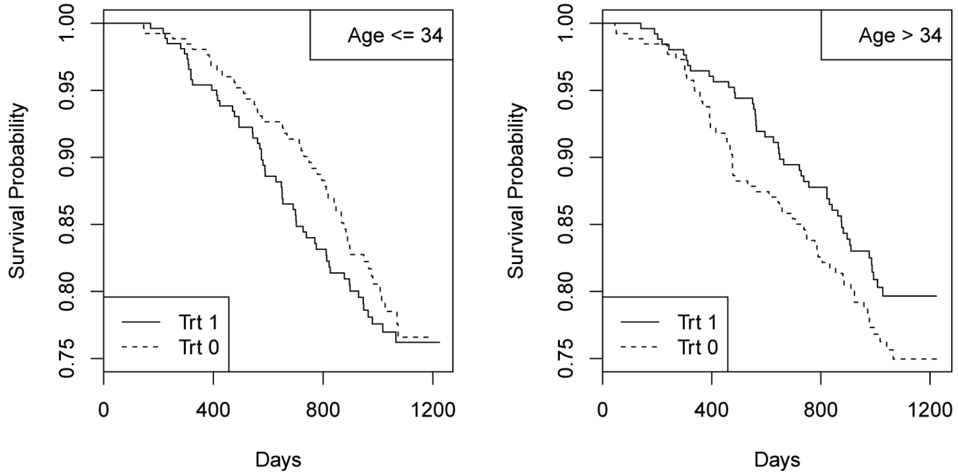

Fig. 1.

Treatment specific Kaplan-Meier curves by age.

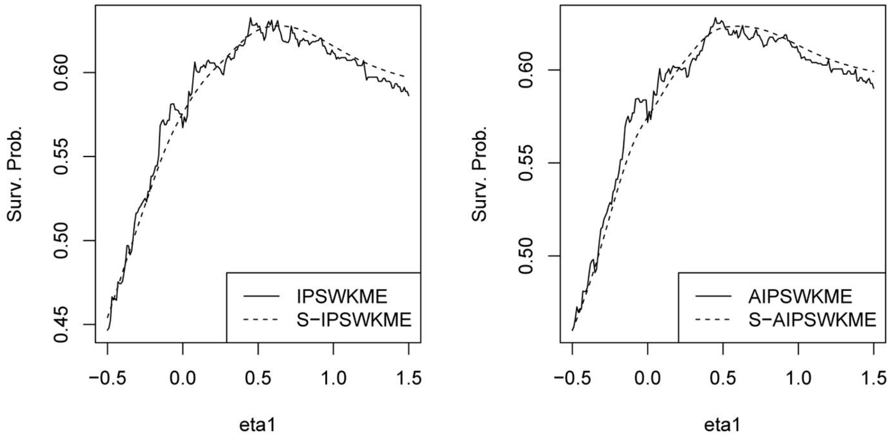

Fig. 2.

Plots for the original and smoothed estimates, where the original estimates are in black and the smoothed estimates are in red. In addition, the IPW and AIPW estimates are given in the left and right panels, respectively.

Table 1

Simulation results for the extreme value error distribution with $n = 250$ and $t = 2$.

	PS	$\hat{\eta}_0$	$\hat{\eta}_1$	$\hat{\eta}_2$	$\hat{S}(\hat{\eta}_{\text{opt}})$	SE	CP	$S(\hat{\eta}_{\text{opt}})$	MR
				Censor Rate = 15%					
$\hat{S}_I$	T	0.010 (0.298)	0.633 (0.192)	−0.665 (0.178)	0.645 (0.037)	0.040	0.839	0.590 (0.016)	0.118 (0.063)
$\tilde{S}_I$	T	−0.005 (0.263)	0.652 (0.179)	−0.667 (0.171)	0.612 (0.036)	0.040	0.968	0.593 (0.014)	0.107 (0.057)
$\hat{S}_A$	T	−0.002 (0.287)	0.639 (0.171)	−0.676 (0.155)	0.639 (0.037)	0.040	0.866	0.592 (0.014)	0.109 (0.058)
$\tilde{S}_A$	T	0.002 (0.256)	0.654 (0.169)	−0.675 (0.152)	0.609 (0.036)	0.040	0.969	0.594 (0.013)	0.102 (0.055)
$\hat{S}_I$	F	−0.031 (0.423)	0.408 (0.327)	−0.697 (0.249)	0.666 (0.036)	0.039	0.659	0.565 (0.040)	0.193 (0.100)
$\tilde{S}_I$	F	−0.051 (0.403)	0.426 (0.285)	−0.714 (0.252)	0.643 (0.035)	0.039	0.844	0.569 (0.034)	0.184 (0.090)
$\hat{S}_A$	F	−0.014 (0.278)	0.660 (0.151)	−0.662 (0.161)	0.635 (0.038)	0.041	0.886	0.593 (0.012)	0.107 (0.055)
$\tilde{S}_A$	F	−0.002 (0.246)	0.675 (0.141)	−0.665 (0.148)	0.607 (0.038)	0.041	0.968	0.596 (0.010)	0.096 (0.050)
				Censor Rate = 40%					
$\hat{S}_I$	T	0.008 (0.311)	0.616 (0.214)	−0.661 (0.202)	0.650 (0.041)	0.044	0.850	0.588 (0.019)	0.127 (0.068)
$\tilde{S}_I$	T	−0.002 (0.285)	0.637 (0.202)	−0.660 (0.191)	0.613 (0.040)	0.045	0.958	0.590 (0.017)	0.118 (0.064)
$\hat{S}_A$	T	0.006 (0.310)	0.623 (0.203)	−0.663 (0.189)	0.645 (0.041)	0.045	0.879	0.589 (0.019)	0.123 (0.068)
$\tilde{S}_A$	T	0.001 (0.282)	0.643 (0.192)	−0.661 (0.183)	0.612 (0.040)	0.044	0.965	0.591 (0.017)	0.115 (0.062)
$\hat{S}_I$	F	0.002 (0.448)	0.388 (0.349)	−0.676 (0.267)	0.671 (0.039)	0.043	0.677	0.560 (0.045)	0.206 (0.109)
$\tilde{S}_I$	F	−0.024 (0.432)	0.403 (0.311)	−0.694 (0.271)	0.645 (0.039)	0.043	0.867	0.564 (0.038)	0.200 (0.095)
$\hat{S}_A$	F	−0.005 (0.299)	0.655 (0.169)	−0.650 (0.176)	0.641 (0.043)	0.046	0.896	0.591 (0.014)	0.115 (0.060)
$\tilde{S}_A$	F	−0.005 (0.270)	0.664 (0.162)	−0.656 (0.173)	0.609 (0.041)	0.046	0.964	0.593 (0.012)	0.109 (0.054)

[†] PS, the propensity score model. Here T means the correctly specified PS model while F means the misspecified PS model. Recall that $S(\eta_{\text{opt}}) = 0.605$.

Table 2

Simulation results for the logistic error distribution with $n = 250$ and $t = 2$.

	PS	$\hat{\eta}_0$	$\hat{\eta}_1$	$\hat{\eta}_2$	$\hat{S}(\hat{\eta}_{opt})$	SE	CP	$S(\hat{\eta}_{opt})$	MR
				Censor Rate = 15%					
$\hat{S}_I$	T	0.010 (0.370)	0.566 (0.272)	−0.641 (0.241)	0.716 (0.034)	0.038	0.791	0.652 (0.022)	0.155 (0.089)
$\tilde{S}_I$	T	−0.004 (0.341)	0.593 (0.262)	−0.640 (0.235)	0.685 (0.034)	0.039	0.955	0.655 (0.020)	0.145 (0.082)
$\hat{S}_A$	T	0.007 (0.363)	0.578 (0.260)	−0.639 (0.240)	0.713 (0.034)	0.039	0.818	0.653 (0.020)	0.151 (0.084)
$\tilde{S}_A$	T	−0.006 (0.341)	0.595 (0.251)	−0.642 (0.233)	0.684 (0.034)	0.039	0.962	0.655 (0.020)	0.143 (0.081)
$\hat{S}_I$	F	0.041 (0.461)	0.340 (0.389)	−0.662 (0.284)	0.729 (0.033)	0.037	0.649	0.632 (0.040)	0.224 (0.120)
$\tilde{S}_I$	F	−0.001 (0.461)	0.375 (0.350)	−0.667 (0.283)	0.707 (0.033)	0.037	0.846	0.636 (0.035)	0.216 (0.107)
$\hat{S}_A$	F	−0.025 (0.337)	0.630 (0.198)	−0.637 (0.210)	0.723 (0.036)	0.040	0.753	0.658 (0.013)	0.133 (0.068)
$\tilde{S}_A$	F	−0.029 (0.320)	0.633 (0.204)	−0.642 (0.204)	0.695 (0.036)	0.040	0.926	0.659 (0.012)	0.130 (0.064)
				Censor Rate = 40%					
$\hat{S}_I$	T	0.013 (0.395)	0.545 (0.293)	−0.625 (0.266)	0.721 (0.036)	0.041	0.785	0.649 (0.027)	0.168 (0.097)
$\tilde{S}_I$	T	−0.008 (0.362)	0.581 (0.274)	−0.626 (0.255)	0.687 (0.036)	0.041	0.948	0.652 (0.022)	0.155 (0.087)
$\hat{S}_A$	T	0.004 (0.381)	0.558 (0.277)	−0.635 (0.255)	0.718 (0.036)	0.042	0.807	0.651 (0.023)	0.160 (0.089)
$\tilde{S}_A$	T	−0.016 (0.361)	0.578 (0.270)	−0.634 (0.246)	0.686 (0.036)	0.042	0.955	0.653 (0.022)	0.153 (0.086)
$\hat{S}_I$	F	0.061 (0.471)	0.325 (0.413)	−0.640 (0.299)	0.733 (0.035)	0.039	0.661	0.628 (0.042)	0.235 (0.124)
$\tilde{S}_I$	F	0.021 (0.482)	0.355 (0.370)	−0.639 (0.312)	0.709 (0.035)	0.040	0.842	0.631 (0.038)	0.229 (0.114)
$\hat{S}_A$	F	−0.012 (0.350)	0.625 (0.206)	−0.631 (0.217)	0.722 (0.038)	0.042	0.785	0.657 (0.014)	0.138 (0.070)
$\tilde{S}_A$	F	−0.022 (0.331)	0.628 (0.214)	−0.634 (0.221)	0.692 (0.038)	0.043	0.939	0.658 (0.013)	0.136 (0.067)

[†] PS, the propensity score model. Here T means the correctly specified PS model while F means the misspecified PS model. Recall that $S(\eta_{opt}) = 0.672$.

Table 3

Results for comparison with the method of Zhao et al. (2013).

error	method	Surv. Prob.	MR
extreme value	Zhao et al. (2013) w. $\xi = 0$	0.445 (0.030)	0.467 (0.048)
	Zhao et al. (2013) w. $\xi = 0.1$	0.499 (0.046)	0.373 (0.089)
	Zhao et al. (2013) w. $\xi = 0.2$	0.555 (0.035)	0.245 (0.099)
	Zhao et al. (2013) w. $\xi = 0.3$	0.585 (0.027)	0.143 (0.091)
	Zhao et al. (2013) w. $\xi = 0.4$	0.590 (0.028)	0.112 (0.093)
	Zhao et al. (2013) w. $\xi = 0.5$	0.543 (0.066)	0.241 (0.162)
	Zhao et al. (2013) w. $\xi = 0.6$	0.542 (0.045)	0.275 (0.109)
	S-IPSWKME	0.594 (0.011)	0.107 (0.052)
	S-AIPSWKME	0.595 (0.009)	0.099 (0.047)
logistic	Zhao et al. (2013) w. $\xi = 0$	0.552 (0.028)	0.456 (0.061)
	Zhao et al. (2013) w. $\xi = 0.1$	0.606 (0.040)	0.323 (0.113)
	Zhao et al. (2013) w. $\xi = 0.2$	0.643 (0.029)	0.200 (0.111)
	Zhao et al. (2013) w. $\xi = 0.3$	0.650 (0.030)	0.164 (0.117)
	Zhao et al. (2013) w. $\xi = 0.4$	0.630 (0.037)	0.246 (0.132)
	Zhao et al. (2013) w. $\xi = 0.5$	0.590 (0.039)	0.373 (0.110)
	Zhao et al. (2013) w. $\xi = 0.6$	0.590 (0.030)	0.382 (0.079)
	S-IPSWKME	0.659 (0.012)	0.130 (0.063)
	S-AIPSWKME	0.660 (0.011)	0.126 (0.061)

†Surv. Prob., the simulated survival probability at $t_0 = 2$; MR, the misclassification rate.

The true optimal survival probabilities are 0.605 and 0.672 for the extreme value and logistic error, respectively.

Values in the parenthesis are the standard deviations over 1000 simulations.

Table 4

Results for comparison with the method of Zhao et al. (2015).

error	method	PS	Surv. Prob.	MR
extreme value	Zhao et al. (2015)	T	0.587 (0.022)	0.136 (0.065)
	AIPSWKM	T	0.592 (0.014)	0.109 (0.058)
	S-AIPSWKM	T	0.594 (0.013)	0.102 (0.055)
	Zhao et al. (2015)	F	0.590 (0.008)	0.134 (0.044)
	AIPSWKM	F	0.593 (0.012)	0.107 (0.055)
	S-AIPSWKM	F	0.596 (0.010)	0.096 (0.050)
logistic	Zhao et al. (2015)	T	0.652 (0.027)	0.159 (0.090)
	AIPSWKM	T	0.653 (0.020)	0.151 (0.084)
	S-AIPSWKM	T	0.655 (0.020)	0.143 (0.081)
	Zhao et al. (2015)	F	0.659 (0.007)	0.141 (0.047)
	AIPSWKM	F	0.658 (0.013)	0.133 (0.068)
	S-AIPSWKM	F	0.659 (0.012)	0.130 (0.064)

[†] PS, the propensity score model. Here T means the correctly specified PS model while F means the misspecified PS model.

[†] Surv. Prob., the simulated survival probability at $t_0 = 2$; MR, the misclassification rate.

The true optimal survival probabilities are 0.605 and 0.672 for the extreme value and logistic error, respectively.

Values in the parenthesis are the standard deviations over 1000 simulations.

Table 5

Simulation results for estimating optimal dynamic treatment regimes.

C%	S	$\hat{\eta}_1^{\text{opt}}$	$\hat{\eta}_2^{\text{opt}}$	$\hat{\eta}_3^{\text{opt}}$	$\hat{\eta}_4^{\text{opt}}$	$\hat{S}(\hat{\boldsymbol{\eta}}_{\text{opt}})$	SE	CP	$S(\hat{\boldsymbol{\eta}}_{\text{opt}})$	MR
Senario 1: $\boldsymbol{\eta}_{\text{opt}}$ = (0.890,−0.456, 0.894,−0.447); $S(3;\ \boldsymbol{\eta}_{\text{opt}})$ = 0.567										
15	F	0.881 (0.035)	−0.468 (0.062)	0.893 (0.017)	−0.449 (0.033)	0.591 (0.028)	0.030	0.887	0.559 (0.008)	0.107 (0.054)
	T	0.884 (0.029)	−0.463 (0.052)	0.894 (0.013)	−0.448 (0.026)	0.570 (0.028)	0.030	0.955	0.561 (0.006)	0.088 (0.048)
40	F	0.878 (0.042)	−0.471 (0.072)	0.890 (0.022)	−0.453 (0.042)	0.600 (0.036)	0.037	0.841	0.555 (0.011)	0.125 (0.061)
	T	0.884 (0.034)	−0.463 (0.061)	0.892 (0.018)	−0.450 (0.035)	0.574 (0.035)	0.038	0.955	0.558 (0.009)	0.108 (0.056)
Senario 2: $\boldsymbol{\eta}_{\text{opt}}$ = (−0.891, 0.454, 0.894,−0.447); $S(6;\ \boldsymbol{\eta}_{\text{opt}})$ = 0.624										
15	F	−0.888 (0.025)	0.456 (0.045)	0.891 (0.018)	−0.452 (0.035)	0.645 (0.025)	0.027	0.890	0.615 (0.008)	0.099 (0.052)
	T	−0.889 (0.018)	0.456 (0.034)	0.893 (0.014)	−0.450 (0.028)	0.624 (0.024)	0.027	0.967	0.618 (0.005)	0.079 (0.042)
40	F	−0.886 (0.030)	0.459 (0.053)	0.890 (0.020)	−0.453 (0.038)	0.650 (0.027)	0.029	0.855	0.613 (0.010)	0.110 (0.055)
	T	−0.888 (0.022)	0.457 (0.040)	0.892 (0.016)	−0.451 (0.032)	0.626 (0.027)	0.030	0.972	0.617 (0.007)	0.091 (0.048)
Senario 3: $\boldsymbol{\eta}_{\text{opt}}$ = (0.908,−0.419, 0.894,−0.447); $S(3;\ \boldsymbol{\eta}_{\text{opt}})$ = 0.702										
15	F	0.897 (0.037)	−0.434 (0.069)	0.892 (0.020)	−0.449 (0.039)	0.728 (0.026)	0.027	0.829	0.692 (0.009)	0.134 (0.067)
	T	0.900 (0.031)	−0.430 (0.060)	0.893 (0.016)	−0.448 (0.031)	0.707 (0.026)	0.027	0.952	0.695 (0.007)	0.116 (0.061)
40	F	0.895 (0.041)	−0.437 (0.075)	0.891 (0.023)	−0.451 (0.043)	0.732 (0.028)	0.029	0.809	0.691 (0.010)	0.142 (0.073)
	T	0.899 (0.035)	−0.431 (0.066)	0.893 (0.019)	−0.449 (0.036)	0.709 (0.028)	0.030	0.951	0.693 (0.008)	0.126 (0.066)

†*C*% denotes the censoring rate; *S* indicates whether the smoothing technique is applied (T) or not (F).

Table 6

Estimation results for the ACTG175 data.

t	Method	Intercept	Karnof	CD40	Age	$\tilde{S}(t;\ \tilde{\eta}_{\text{opt}})$	CI$_1$	CI$_0$
400	I	−0.303	−0.340	0.024	0.890	0.965 (0.008)	(−0.002, 0.023)	(−0.003, 0.044)
	A	−0.729	−0.240	0.018	0.640	0.965 (0.008)	(−0.002, 0.022)	(−0.003, 0.043)
600	I	0.975	−0.082	0.001	0.206	0.923 (0.012)	(0.000, 0.045)	(−0.006, 0.052)
	A	0.909	−0.137	0.000	0.392	0.922 (0.012)	(0.000, 0.043)	(−0.009, 0.052)
800	I	0.871	−0.133	−0.010	0.473	0.887 (0.014)	(0.008, 0.058)	(−0.002, 0.069)
	A	0.874	−0.131	−0.009	0.469	0.886 (0.014)	(0.006, 0.057)	(−0.003, 0.068)
1000	I	−0.210	−0.185	−0.035	0.959	0.824 (0.017)	(0.004, 0.060)	(−0.006, 0.081)
	A	0.001	−0.187	−0.037	0.982	0.823 (0.017)	(0.002, 0.059)	(−0.007, 0.080)

[†] I denotes the S-IPSWKME and A denotes the S-AIPSWKME; the numbers in the parenthesis are the estimated standard errors; CI$_1$ and CI$_0$ denote the 95% confidence intervals for the difference of the value functions obtained under the estimated optimal treatment regime and the simple treatment regime assigning all to treatment 1 and 0, respectively.