Published in final edited form as:

Lifetime Data Anal. 2017 October ; 23(4): 585–604. doi:10.1007/s10985-016-9376-x.

Optimal treatment regimes for survival endpoints using a locally-efficient doubly-robust estimator from a classification perspective

Xiaofei Bai, Anastasios A. Tsiatis, Wenbin Lu, and Rui Song

Abstract

A treatment regime at a single decision point is a rule that assigns a treatment, among the available options, to a patient based on the patient's baseline characteristics. The value of a treatment regime is the average outcome of a population of patients if they were all treated in accordance to the treatment regime, where large values are desirable. The optimal treatment regime is a regime which results in the greatest value. Typically, the optimal treatment regime is estimated by positing a regression relationship for the outcome of interest as a function of treatment and baseline characteristics. However, this can lead to suboptimal treatment regimes when the regression model is misspecified. We instead consider value search estimators for the optimal treatment regime where we directly estimate the value for any treatment regime and then maximize this estimator over a class of regimes. For many studies the primary outcome of interest is survival time which is often censored. We derive a locally efficient, doubly robust, augmented inverse probability weighted complete case estimator for the value function with censored survival data and study the large sample properties of this estimator. The optimization is realized from a weighted classification perspective that allows us to use available off the shelf software. In some studies one treatment may have greater toxicity or side effects, thus we also consider estimating a quality adjusted optimal treatment regime that allows a patient to trade some additional risk of death in order to avoid the more invasive treatment.

Keywords

Classification; Doubly-robust; Observational survival study; Optimal treatment regime; Value search

1 Introduction

Surgical revascularization (coronary artery bypass grafting) or catheter-based revascularization (percutaneous coronary intervention) are two major treatment options for patients with coronary artery disease. An important clinical question is to determine which of these two treatments should be given to a patient with coronary artery disease based on the patient's characteristics at the time they present for treatment.

A treatment regime for this setting is a decision rule that takes an individual's baseline information and dictates the treatment s/he should receive among the available options. A treatment will be assessed by a health outcome of interest which is coded so that larger values reflect greater benefit. In this paper the health outcome of interest will be a monotone

increasing function of the survival time measured from time to treatment until death. The value of a treatment regime is the average outcome of a population of patients if they were all treated in accordance with that treatment regime. An optimal treatment regime is the decision rule that leads to the most beneficial outcome on average or the greatest value. Therefore, it is of great interest to estimate the optimal treatment regime from available data. Specifically, we use data from the ASCERT study to guide us on which treatment, bypass surgery or percutaneous coronary intervention, should be recommended for patients with coronary artery disease based on their baseline characteristics.

The ASCERT study was a retrospective study of patients who had either two or three vessel coronary artery disease and were treated with bypass surgery or percutaneous coronary intervention. This was an observational study where the treatment assignment was not random and the primary outcome of interest was survival time after treatment which was subject to censoring. Twenty-eight baseline covariates were collected and used to determine which treatment, percutaneous coronary intervention or bypass surgery should be recommended. Moreover, we demonstrate and derive the quality adjusted optimal treatment regimes to allow the patients to trade off some additional risk in order to avoid a more invasive treatment (in the case of ASCERT study, surgical revascularization). More details on the specifics of the data analysis are given in section 6.

In the past decades, there has been substantial research on statistical methodology for estimating optimal treatment regimes based on data from clinical trials and observational studies. Most of these methods involve positing models for the outcome regressed on the treatment assignment and baseline covariates (e.g., Murphy, 2003; Robins, 2004; Moodie et al., 2007) or using inverse propensity score weighting methods (e.g., Robins et al., 2008; Brinkley et al., 2009; Zhao et al., 2009; Orellana et al., 2010). Consequently, the consistency of these estimators depends on correct specification of the proposed outcome regression or propensity score models. If the models are misspecified, the derived optimal treatment regimes are no longer guaranteed to be consistent.

Moreover, in clinical trials and observational studies, the primary outcomes of interest are often times-to-event, for example, overall survival or disease-free survival, and are usually censored for some study participants. There is little work on estimating optimal regimes from such data nor on regimes that focus on maximizing probability of survival at or beyond a threshold time point and related quantities of interest. In this paper, we exploit Bai et al. (2013) to develop methods for estimating optimal treatment regimes from censored survival data. Under the no unmeasured confounders assumption, we show the derived estimators are doubly robust. They will be consistent if either of the postulated model for the survival distribution as a function of covariates or those for the propensity score and censoring distribution are correctly specified, even if the other is not. Thus, the proposed estimators will also offer protection against model misspecification.

In the next section, we introduce the notation and assumptions used in this paper. In Section 3, we derive a doubly robust estimator for the value of a treatment regime and the corresponding asymptotic variance and use this to derive the value search estimator of the optimal regime within a restricted class of regimes. We describe how to cast this

optimization problem from a classification perspective in Section 4. The performance of the proposed estimators is demonstrated by simulation studies in Section 5. In Section 6, we apply our method to the motivating example from the ASCERT study. In this section we also describe how to modify these methods to define and estimate a quality adjusted optimal treatment regime that allows a patient to trade some additional risk of death in order to avoid the more invasive treatment of bypass surgery. The paper ends with discussion and conclusion in Section 7.

2 Notation and Assumptions

We begin by introducing the notation and assumptions used throughout the paper. Let A denote treatment assignment, in this case we let $A = 1$ or 0 (bypass surgery or percutaneous coronary intervention). Let X denote the vector of baseline covariates. Let N denote the number of observations included in our sample. For each $i = 1, \dots, N$, $T_i^*(a)$ denotes the potential survival time and $C_i^*(a)$ the potential censoring time if individual i were given treatment (possibly contrary to fact) a , $a = 0, 1$. Assume the consistency assumption that $T_i = A_i T_i^*(1) + (1 - A_i) T_i^*(0)$, where T_i is the potential time to death of patient i given his/her assigned treatment A_i . Similarly, $C_i = A_i C_i^*(1) + (1 - A_i) C_i^*(0)$ is the potential censoring time of patient i given his/her assigned treatment A_i . Let $U_i = \min(T_i, C_i)$ denote the observed time to death or censoring and $\delta_i = I(T_i \leq C_i)$ the failure indicator for individual i , where $I(\cdot)$ denotes the indicator function. The observed data from a study can be summarized as $(U_i, \delta_i, X_i, A_i)$, $i = 1, \dots, N$. They are assumed independently and identically distributed.

We make the usual assumption of non-informative censoring that the potential censoring time is conditionally independent of the potential survival time given the baseline covariates, that is, $C^*(a) \perp\!\!\!\perp T^*(a) | X$, $a = 0, 1$, where “ $\perp\!\!\!\perp$ ” denotes the symbol of statistically independent or conditionally independent. We also make the strong ignorability assumption (Rubin, 1978), which is also referred to as the no unmeasured confounders assumption, that treatment assignment A is conditionally independent of the potential outcomes $T^*(a)$, $C^*(a)$ given covariates X ; that is, $A \perp\!\!\!\perp \{T^*(a), C^*(a)\} | X$. Such an assumption is necessarily true in a randomized clinical trial where treatment A is assigned at random, but must be evaluated critically in an observational study where this assumption would hold true if the important covariates X that a patient and physician used to decide on treatment are collected in the database. The no unmeasured confounders assumption together with the non-informative censoring assumption imply that $A \perp\!\!\!\perp T^*(a) \perp\!\!\!\perp C^*(a) | X$, for $a = 0, 1$.

The baseline covariates X take on values $x \in \mathcal{X}$, the sample space for X . A treatment regime $d: \mathcal{X} \rightarrow (0, 1)$ is a function that maps from $\mathcal{X}$ to the values 0 or 1 so that a patient with baseline information $X = x$ would receive treatment 1 if $d(x) = 1$ and treatment 0 if $d(x) = 0$. We denote by $\mathcal{D}$ the class of all possible treatment regimes. Accordingly, the potential survival time of a patient with baseline covariate X if s/he receives treatment following treatment regime d is $T^*(d) = d(X) T^*(1) + \{1 - d(X)\} T^*(0)$. We define the primary outcome of interest as $f\{T^*(d)\}$, where $f(\cdot)$ is a function of the survival time that is of interest. For example, $f\{T^*(d)\}$ may be equal to $I\{T^*(d) \leq u\}$ or $\min\{T^*(d), L\}$ to denote the indicator of

whether a patient survives through time u , or the truncated survival time up through time L , respectively, if the patient received treatment following treatment regime d . Because in most studies patients are followed for a finite amount of time we are limited to considering the truncated lifetime.

For these two examples of $f(\cdot)$, $E[f(T^*(d))]$ denotes the survival probability at time u and the mean truncated survival time, respectively, for the population if all patients were to receive treatment according to d . $E[f(T^*(d))]$ is also referred to as the value of d or $V(d)$. An optimal treatment regime $d^{opt} \in \mathcal{D}$ satisfies $V(d) \leq V(d^{opt})$ for all $d \in \mathcal{D}$.

3 Value Search Estimators for Optimal Restricted Regimes

It is straightforward to show that under the assumption of no unmeasured confounders and the consistency assumption that the optimal treatment regime is given as

$$d^{opt}(x) = I[E\{f(T)|A=1, X=x\} > E\{f(T)|A=0, X=x\}]. \quad (1)$$

An obvious strategy is to develop a model for the conditional distribution of T given A and X through parameters θ , using either parametric or semiparametric models such as Cox's proportional hazards model. This model is used to derive $E\{f(T)|A, X\} = Q(X, A; \theta)$, in which case, the estimator of the optimal regime is $\hat{d}^{opt}(x) = I\{Q(x, 1; \hat{\theta}) > Q(x, 0; \hat{\theta})\}$, where $\hat{\theta}$ is an estimator for θ . We refer to this as an outcome regression model and the outcome regression estimator for the value function $V(d)$ would be

$\hat{V}^{OR}(d) = N^{-1} \sum_{i=1}^N [d(X_i)Q(X_i, 1; \hat{\theta}) + \{1 - d(X_i)\}Q(X_i, 0; \hat{\theta})]$ and the outcome regression estimator for the optimal regime would be $\hat{V}^{OR}(\hat{d}^{opt})$.

A drawback of this approach is that a misspecified model can lead to an estimator that is far from the optimal regime, see Zhang et al. (2012a). Another strategy is to limit the class of regimes to a reasonable restricted class of regimes indexed by a parameter η . Even when we posit models $Q(X, A; \theta)$, as indicated above, we are implicitly considering a restricted class of regimes $I\{Q(X, 1; \theta) > Q(X, 0; \theta)\}$ indexed by parameter η defined through θ . For example, if we consider a proportional hazards model where the hazard function for the survival time at r given X and A is given by

$$\lambda_T(r|X, A) = \lambda_{0T}(r) \exp\{\gamma_0 + \gamma_1^T X - A(\eta_0 + \eta_1^T X)\}, \quad (2)$$

then for such a model, $I\{Q(x, 1; \theta) > Q(x, 0; \theta)\}$, where $\theta = \{\lambda_{0T}(\cdot), \gamma, \eta\}$, is equivalent to $d(x, \eta) = I(\eta_0 + \eta_1^T x > 0)$ and this would be the case for any choice of the function $f(\cdot)$ which is monotone increasing. This demonstrates that positing a model can be viewed as defining a class $\mathcal{D}_\eta$, say, whose elements are indexed by a parameter η . Obviously, $\mathcal{D}_\eta \in \mathcal{D}$. If in fact the model is correctly specified then $d^{opt} \in \mathcal{D}$ is also such that $d^{opt} \in \mathcal{D}_\eta$; if the model is not correctly specified, then d^{opt} may or may not be in $\mathcal{D}_\eta$. If a misspecified model is fitted to

the data, the resulting estimated optimal regime may not necessarily estimate an optimal regime within $\mathcal{D}_\eta$ (defined formally below).

The above reasoning suggests that we consider a restricted class of regimes $\mathcal{D}_\eta$ with elements $d(x; \eta) = d_\eta(x)$ chosen based on interpretability, feasibility in practice or cost. For example, rules involving cut-off or threshold such as $d(x; \eta) = \mathbb{I}(x_1 < \eta_1, x_2 < \eta_2)$ are straightforward to interpret and implement. The restricted class $\mathcal{D}_\eta$ may or may not contain $d^{opt} \in \mathcal{D}$, but an optimal regime in $\mathcal{D}_\eta$ is still of interest when the focus is on regimes with certain features. An optimal regime in $\mathcal{D}_\eta$, d_η^{opt} , should maximize the value $E[f\{T^*(d_\eta)\}] = V(d_\eta)$ among regimes $d_\eta \in \mathcal{D}_\eta$. Formally, $d_\eta^{opt}(x) = d(x; \eta^{opt})$, $\eta^{opt} = \arg \max_\eta V(d_\eta)$. If for any regime d_η we can estimate $V(d_\eta)$, say $\hat{V}(d_\eta)$, using the observed data, then the estimated optimal restricted regime is $\hat{\eta}^{opt} = \arg \max_\eta \hat{V}(d_\eta)$. This is referred to as a value search estimator.

To implement this strategy we desire a robust efficient estimator of the value function for any regime d . Here we consider a semiparametric model. That is, we make no assumptions regarding the conditional distribution of T given A and X ; i.e., this is the nonparametric component of the model. We denote the propensity score as $\pi(X) = P(A = 1|X)$ and the conditional distribution of censoring given A and X by $K_c(a, r, X) = P(C \leq r | X, A = a)$, $a = 0, 1$. If we take the point of view that $\pi(X)$ and $K_c(a, r, X)$ are known or can be estimated consistently with appropriate models then using the semiparametric theory as outlined in chapters 7–9 of (Tsiatis, 2006), it can be shown that, under these assumptions, all regular asymptotically linear estimators for $E[f\{T^*(a)\}]$ for $a = 0, 1$ are given by

$N^{-1} \sum_{i=1}^N IF_i(a, h)$, where

$$IF_i(a, h) = \frac{A_i^a (1 - A_i)^{1-a} \Delta_i f(U_i)}{\{\pi(X_i)\}^a \{1 - \pi(X_i)\}^{1-a} K_c(a, U_i, X_i)} - \left\{ \frac{(2a - 1) \{A_i - \pi(X_i)\}}{\{\pi(X_i)\}^a \{1 - \pi(X_i)\}^{1-a}} \right\} h_1(a, X_i) \\ + \frac{A_i^a (1 - A_i)^{1-a}}{\{\pi(X_i)\}^a \{1 - \pi(X_i)\}^{1-a}} \int_0^\infty \frac{dM_c(a, r, X_i)}{K_c(a, r, X_i)} h_2(a, r, X_i), \quad (3)$$

$dM_c(a, r, X)$ is the martingale increment for the potential censoring time for treatment a ; namely, $dM_c(a, r, X) = dN_c(r) - \lambda_c(a, r, X) Y(r)$, $N_c(r) = \mathbb{I}(U \leq r, C = 0)$, $Y(r) = \mathbb{I}(U \leq r)$ and $\lambda_c(a, r, X) = -d \log K_c(a, r, X)/dr$ is the censoring time hazard function given X and $A = a$, and $h_1(a, X)$ and $h_2(a, r, X)$ are arbitrary functions.

This class of estimators can be motivated heuristically by the following considerations. For any fixed treatment a and any patient i , we note that we will observe the endpoint of interest $f\{T_i^*(a)\}$ if $A_i = a$; that is, if patient i receives treatment a , and if $C_i > U_i$, or equivalently ($C_i > U_i$). The probability of seeing such a complete case for an individual with covariate X_i is given by $\{\pi(X_i)\}^a \{1 - \pi(X_i)\}^{1-a} K_c(a, U_i, X_i)$ and is the inverse of the weight used in the first term of (3). If we choose $h_1(a, X) = 0$ and $h_2(a, r, X) = 0$, then the resulting estimator is

referred to as the inverse probability weighted complete-case (IPWCC) estimator. We shall define $IF_i^{IPW}(a)$ to be the leading term on the right hand side of (3); that is, when $h_1(a, X) = h_2(a, r, X) = 0$, used to derive the IPWCC estimator.

The class of estimators using $IF_i(a, h)$ above with arbitrary $h_1(a, X)$ and $h_2(a, r, X)$ are referred to as augmented inverse probability weighted complete-case (AIPWCC) estimators. The second and third term on the right-hand side of (3) are referred to as augmentation terms which have mean zero under our assumptions and lead to consistent asymptotically normal estimators for $E\{f(T^*(a))\}$. The choice of $h_1(a, X)$ and $h_2(a, r, X)$ will however affect the asymptotic variance and thus the efficiency of the corresponding estimator.

It is shown in Theorem 10.4 of (Tsiatis, 2006) that the optimal choice for $h_1(a, X)$ and $h_2(a, r, X)$ are $E\{f(T)|X, A = a\}$ and $E\{f(T)|T \geq r, X, A = a\}$, respectively. As a result it was shown by Hubbard et al. (1999) and Bai et al. (2013), that the locally efficient estimator for $E\{f(T^*(a))\}$ for $a = 0, 1$ is given by choosing

$$IF_i(a) = \frac{A_i^a(1 - A_i)^{1-a} \Delta_i f(U_i)}{\{\pi(X_i)\}^a \{1 - \pi(X_i)\}^{1-a} K_c(a, U_i, X_i)} - \left\{ \frac{(2a - 1)\{A_i - \pi(X_i)\}}{\{\pi(X_i)\}^a \{1 - \pi(X_i)\}^{1-a}} \right\} E\{f(T)|X_i, A=a\} \\ + \frac{A_i^a(1 - A_i)^{1-a}}{\{\pi(X_i)\}^a \{1 - \pi(X_i)\}^{1-a}} \int_0^\infty \frac{dM_c(a, r, X_i)}{K_c(a, r, X_i)} E\{f(T)|T \geq r, X_i, A=a\}, \quad (4)$$

Again, a heuristic motivation for the optimal $IF_i(a)$ is that for an individual i in our sample that does not receive treatment a , then we observe their baseline covariates X_i but the endpoint $f\{T_i^*(a)\}$ is missing. The second term in (4) and the first augmentation term is used to capture back some information on the endpoint of interest for such individuals. For patients that receive treatment a but are censored, $A_i = a$ and $i = 0$, then we observe the baseline covariates X_i and the partial information that their survival time $T_i^*(a)$ was greater than C_i . The third term in (4) and the second augmentation term captures back some information on the endpoint of interest for these patients.

In order to obtain an estimator for $E\{f(T^*(a))\}$, we need to estimate $\pi(X)$, $K_c(a, r, X)$ and $E\{f(T)|T \geq r, X, a\}$ for $a = 0, 1$. The above development was predicated on the assumption that the propensity score and conditional censoring distribution can be consistently estimated. In a randomized clinical trial, where treatments are assigned at random with known probabilities, the propensity score is known by design. Also, in many randomized clinical trials, censoring will not depend on either the covariates X or the treatment assignment A . In that case the censoring distribution can be consistently estimated using the Kaplan-Meier estimator for censoring derived by reversing the roles of death and censoring. For observational studies, such as the one motivating this research, the propensity score and conditional censoring distribution are not known and need to be estimated from the data using posited models. Because A is binary we often posit a logistic regression model for $\pi(X)$ and

estimate the parameters using $\{(A_i, X_i), i = 1, \dots, N\}$. We denote the resulting estimator by $\hat{\pi}(X)$. To estimate $K_c(a, r, X)$ and $\lambda_c(a, r, X)$, one generally posits a proportional hazards model, stratified by treatment, and obtains estimators using the data $\{(U_i(1 - \delta_i), X_i), i: A_i = a\}$. The resulting estimators are denoted by $\hat{K}_c(a, r, X)$ and $\hat{\lambda}_c(a, r, X)$ which then can be used to estimate $d\hat{M}_c(a, r, X)$. Finally, we need an estimator for $E\{f(T)|T \geq r, X, a\}$, which necessitates that we posit a model for the conditional distribution of T given X and A . With censored data one generally uses the proportional hazards model such as that in (2). We suggest using proportional hazards models, stratified by treatment, with which we can obtain estimators $\hat{H}(a, r, X)$ for $H(a, r, X) = P\{T \geq r|X, A = a\}$ using data $\{(U_i(1 - \delta_i), X_i), i: A_i = a\}$. In that case

$$\hat{E}\{f(T)|T \geq r, X, a\} = \int_r^\infty f(z) \left\{ \frac{-d\hat{H}(a, z, X)}{\hat{H}(a, r, X)} \right\}.$$

For example, if $f\{T^*(a)\} = f\{T^*(a) - u\}$ then $\hat{E}\{f(T)|T \geq r, X, a\} = \hat{H}(a, u, X)/\hat{H}(a, r, X)$ for $r \geq u$.

For randomized studies the AIPWCC estimators described above, including the IPWCC estimator will lead to consistent estimators for $E[f\{T^*(a)\}]$. For observational studies these estimators will be consistent if the posited models for the propensity score and conditional censoring distribution are correctly specified otherwise, the resulting estimator may be biased. However, one of the properties of the locally efficient estimator using (4) is that of double robustness. As shown in Online Resource 1, $IF_i(a)$ is a doubly robust predictor of $E[f\{T_i^*(a)\}|X_i]$ in the sense that

$$E\{IF_i(a)|X_i\} = E[f\{T_i^*(a)\}|X_i], \quad (5)$$

if either the propensity score model $\pi(X)$ and the model for censoring distribution $K_c(a, r, X)$ are correctly specified or the regression model $H(a, r, X) = P\{T \geq r|X, A = a\}$ is correctly specified, for $a = 0, 1$. Consequently, $E\{IF_i(a)\} = E[f\{T_i^*(a)\}]$ and therefore the corresponding estimator will be doubly robust; that is, will be a consistent asymptotically normal estimator for either scenario above.

Because $f\{T^*(a)\} = d(X)f\{T^*(1)\} + \{1 - d(X)\}f\{T^*(0)\}$, then it follows by (5) that the locally efficient doubly robust estimator for $V(d) = E[f\{T^*(d)\}]$ is given by

$N^{-1} \sum_{i=1}^N IF_i(d)$, where

$$IF_i(d) = d(X_i)IF_i(1) + \{1 - d(X_i)\}IF_i(0) = d(X_i)\{IF_i(1) - IF_i(0)\} + IF_i(0). \quad (6)$$

The estimator for $V(d)$ is given by $\hat{V}(d) = N^{-1} \sum_{i=1}^N \widehat{IF}_i(d)$, where $\widehat{IF}_i(d)$, an estimator for $IF_i(d)$ given by (6), is derived by substituting $\hat{\pi}(X)$, $\hat{K}_c(a, r, X)$ and $\hat{E}\{f(T)|T \geq r, X, A = a\}$, a

= 0, 1 into (4). For any regime $d_\eta \in \mathcal{D}_\eta$ we can derive the estimator $\hat{V}(d_\eta)$ as described above, in which case the optimal restricted regime estimator is given by

$$\hat{\eta}^{\text{opt}} = \arg \max_{\eta} \hat{V}(d_\eta). \quad (7)$$

We could have also estimated the value function $V(d)$ using the inverse probability weighted estimator $\hat{V}^{IPW}(d) = N^{-1} \sum_{i=1}^N \widehat{IF}_i^{IPW}(d)$, where $\widehat{IF}_i^{IPW}(d)$, an estimator for $IF_i^{IPW}(d) = d(X_i)IF_i^{IPW}(1) + \{1 - d(X_i)\}IF_i^{IPW}(0)$, derived by substituting estimates for the propensity score and conditional censoring distribution as above. For any fixed regime d the estimator $\hat{V}^{IPW}(d)$ will be consistent only if the propensity score and conditional censoring distribution are estimated consistently (i.e., it does not have the double-robustness property) and, moreover, will be less efficient than $\hat{V}(d)$. Again, this estimator will be consistent in randomized clinical trials but will not enjoy the double robustness property for observational studies.

As described by Zhang et al. (2012a), under certain regularity conditions,

$$n^{1/2} \{ \hat{V}(d_{\hat{\eta}^{\text{opt}}}) - V(d_{\eta^{\text{opt}}}) \} = n^{1/2} \{ \hat{V}(d_{\eta^{\text{opt}}}) - V(d_{\eta^{\text{opt}}}) \} + o_p(1).$$

Based on this result, we can derive the asymptotic normality of the value function estimator. The asymptotic variance of $\hat{V}(d_{\eta^{\text{opt}}})$ can be approximated by the asymptotic variance of $\hat{V}(d_{\eta^{\text{opt}}})$, which can be estimated by the standard sandwich variance formula

$$N^{-1} \sum_{i=1}^N \left\{ \widehat{IF}_i(d_{\hat{\eta}^{\text{opt}}}) - \hat{V}(d_{\hat{\eta}^{\text{opt}}}) \right\}^2. \quad (8)$$

We must however be a little careful here. In order that the above hold we need $V(d_\eta)$ to be maximized at $\eta = \eta^{\text{opt}}$ and the gradient of $V(d_\eta)$, as a function in η , be zero at $\eta = \eta^{\text{opt}}$. In general the above conditions will hold under what is called a non exceptional law. That is when the set $\{x: E\{\mathcal{K}(T)|A=1, X=x\} - E\{\mathcal{K}(T)|A=0, X=x\} = 0\}$ has probability zero. Such a result was demonstrated by van der Laan & Luedtke (2014).

Consequently, if the null hypothesis of no treatment effect holds, where $V(d_\eta)$ is constant for all η , then the regularity conditions may not hold. We will examine the impact of the proposed asymptotic methods under this scenario in our simulation studies.

Although the estimator $\hat{V}(d_{\eta^{\text{opt}}})$ has a double robustness property, the variance estimator (8) is no longer guaranteed to be valid under some model misspecification. By the semiparametric theory (Tsiatis, 2006), if both $\pi(X)$ and $K_c(a, r, X)$ are consistently estimated but $H(a, r, X)$ is not, then the sandwich variance is expected to be a bit conservative; if $H(a, r, X)$ is consistently estimated but either $\pi(X)$ or $K_c(a, r, X)$ is not, then theoretically, we do not know in which direction the variance estimator (8) will be biased.

We study this issue later in the simulation studies. In the case of possible model misspecification, a non-parametric bootstrap procedure might be used to obtain the variance estimator.

Moreover, we can compare the value function following the optimal treatment regime and fixed treatments. The difference between $V(d_{\eta^{opt}})$ and $V(a)$ for $a = 0, 1$ can be estimated by

$$N^{-1} \sum_{i=1}^N \left\{ \widehat{IF}_i(d_{\eta^{opt}}) - \widehat{IF}_i(a) \right\}, \quad (9)$$

and its variance can be estimated by

$$N^{-1} \sum_{i=1}^N \left[\widehat{IF}_i(d_{\eta^{opt}}) - \widehat{IF}_i(a) - \{ \hat{V}(d_{\eta^{opt}}) - \hat{V}(a) \} \right]^2. \quad (10)$$

Of course, the previous discussion regarding regularity conditions and the arguments on double robustness and the effect of model misspecification still hold.

In the development above we were interested in finding optimal treatment regimes. In some cases one treatment may have worse adverse effects than the other; for example, bypass surgery is a more invasive procedure than percutaneous coronary intervention and often associated with severe side effects. In such situations, some patients may be willing to consider a quality adjusted outcome of interest; namely, $f\{T^*(d)\} - \mu d$, $\mu > 0$, where μ is the amount a patient is willing to trade on their survival outcome of interest to avoid the potential side effects of treatment $d = 1$ (bypass surgery). Consequently, the μ -quality adjusted optimal restricted regime is defined as $\eta^{opt}(\mu) = \arg \max_{\eta} E[f\{T^*(d_{\eta})\} - \mu d_{\eta}(X)]$ and is estimated by maximizing in η

$$N^{-1} \sum_{i=1}^N \{ \widehat{IF}_i(d_{\eta}) - \mu d_{\eta}(X_i) \}. \quad (11)$$

Deriving an estimator for the μ -quality adjusted optimal restricted regime is similar to finding the estimator for the optimal restricted regime; however, there are some issues that have to be addressed if we want to consider simultaneously a range of μ . We defer this discussion until section 6 where we consider the analysis of the ASCERT data and now just concentrate on the optimal restricted regime.

Because $\hat{V}(d_{\eta})$ is a non-smooth function of η , standard optimization techniques cannot be used to carry out the maximization in η to derive $\hat{\eta}^{opt}$. Zhang et al. (2012a) report on successful use of a genetic algorithm for this purpose. However, these methods are feasible only when the dimension of the parameter η is small and becomes computationally prohibitive otherwise. This leads us to the classification perspective.

4 Classification Perspective

Zhang et al. (2012b) and Zhao et al. (2012) observed that it is possible to cast estimation of an optimal regime as a weighted classification problem. It was also shown in Zhang et al. (2012b) that $\hat{V}(d_\eta)$ may be written as

$$\hat{V}(d_\eta) = N^{-1} \sum_{i=1}^N d(X_i; \eta) \widehat{CF}(X_i) + \text{terms not involving } d, \quad (12)$$

where $\widehat{CF}(X_i) = \widehat{IF}_i(1) - \widehat{IF}_i(0)$. We showed in (5) that $E\{IF_i(1) - IF_i(0) | X_i\} = CF(X_i)$, where $CF(X_i)$, referred to as the contrast function, is equal to $E\{I\{T^*(1)\} | X_i\} - E\{I\{T^*(0)\} | X_i\}$ and that this is true if either the models for $\pi(X)$ and $K_\zeta(a, r, X)$ are both correctly specified or if the model for $H(a, r, X)$ is correctly specified. Consequently, we can view $\widehat{CF}(X_i) = \widehat{IF}_i(1) - \widehat{IF}_i(0)$ as a doubly-robust predictor of the i -th contrast function (details on contrast function and the residual term not involving d is presented in Online Resource 2).

Note that $d^{opt} \in \mathcal{D}$ in (1) can be expressed as $d^{opt}(x) = I\{CF(x) > 0\}$. By further algebra as shown by Zhang et al. (2012b),

$$d(X_i; \eta) \widehat{CF}(X_i) = -|\widehat{CF}(X_i)| [I\{\widehat{CF}(X_i) > 0\} - d(X_i; \eta)]^2 + |\widehat{CF}(X_i)| I\{\widehat{CF}(X_i) > 0\},$$

so that

$$\hat{\eta}^{opt} = \arg \min_{\eta} \sum_{i=1}^N |\widehat{CF}(X_i)| [I\{\widehat{CF}(X_i) > 0\} - d(X_i; \eta)]^2 = \arg \min_{\eta} \sum_{i=1}^N W_i \{Z_i - d(X_i; \eta)\}^2. \quad (13)$$

From (13), value search estimation may be viewed as a weighted classification problem with “class” or “label” $I\{\widehat{CF}(X_i) > 0\}$ identified with the binary response Z_i , “weight” $|\widehat{CF}(X_i)|$ identified with W_i and “classifier” $d(X_i; \eta)$.

Representation (13) suggests that an optimal restricted treatment regime may be estimated by minimizing the weighted classification error. This formulation underlies machine learning algorithms, specifically, supervised learning methods such as support vector machines (Cortes & Vapnik, 1995) which lead to classifiers in the form of hyperplanes; or recursive partitioning techniques such as classification and regression trees yielding rectangular regions. In the rest of this paper we will focus on linear classifiers using weighted support vector machines.

Specifically, our goal is to find a linear classifier $d_\eta(X) = I(\eta_0 + \eta_1^T X > 0)$ with coefficients

$(\eta_0, \eta_1^T)^T$ that minimize $\sum_{i=1}^N \hat{W}_i \psi\{Y_i(\eta_0 + \eta_1^T X_i)\} + c\|\eta_1\|$, where $\hat{W}_i = |\widehat{CF}(X_i)|$, $Y_i = 2I\{\widehat{CF}(X_i) > 0\} - 1$, $\|\eta_1\|$ denotes the norm of η_1 hence, $c\|\eta_1\|$ is a penalty term that penalizes coefficients that are far from zero with c being a tuning parameter (often determined via cross validation). Ideally, estimating the optimal treatment regime would take $\psi(t) = 1$ if $t < 0$; $\psi(t) = 0$ if $t \geq 0$ because such a choice of ψ would minimize the weighted classification error given by (13). However, such $\psi(t)$ is a non-convex function and the minimization problem would be difficult to solve. Therefore, weighted support vector machine uses the convex Hinge loss function, $\psi(t) = \max(1 - t, 0)$ and the goal is to find η_0 and η_1 that minimize

$$\sum_{i=1}^N \hat{W}_i \times \max\{1 - Y_i(\eta_0 + \eta_1^T X_i), 0\} + c\|\eta_1\|.$$

We consider L_1 norm $\|\cdot\|_1$ and solve the convex optimization problem using linear programming. A detailed description on implementing the weighted support vector machine method, including how the tuning parameter c is chosen using cross-validation, can be found in Online Resource 3.

We also note that the μ -quality adjusted value estimator given by (11) is

$$N^{-1} \sum_{i=1}^N d(X_i; \eta) \{\widehat{CF}(X_i) - \mu\} + \text{terms not involving } d,$$

consequently, finding the μ -quality adjusted optimal restricted regime involves replacing the label above by $I\{\widehat{CF}(X_i) - \mu > 0\}$ and the weight $|\widehat{CF}(X_i) - \mu|$. In the section on the analysis of the ASCERT study we will discuss how classification methods could be used when considering estimating μ -quality regimes over a range of values for μ .

In Zhao et al. (2012) they also considered using the weighted support vector machine to approximate the weighted classification error. However, their formulation which they called outcome weighted learning (OWL) is equivalent to the method above but using the less efficient IPW estimators $\widehat{IF}_i^{IPW}(a)$, $a = 0, 1$ to derive the estimator of the contrast function. Moreover, in their work they did not consider issues regarding censored observations and assumed the treatments were assigned by randomization where the propensity score is known by design. Therefore, OWL as derived in Zhao et al. (2012), could not be used directly on the problem we are considering. Nonetheless, because of the interest in OWL, we considered the IPW estimator for the value function to estimate the optimal restricted regime as the natural OWL competitor for comparison. Because our proposed doubly-robust estimator is a more efficient estimator than one which uses only inverse probability weighting we expect that this will result in a better estimator for the optimal treatment regime. This will be evaluated in the next section on simulation studies.

5 Simulation

Several simulation studies are carried out to evaluate the performance of the proposed estimators, each including 1000 Monte-Carlo replications and 500 observations. In the first set of simulations, we consider observational study scenarios where the propensity score is not known by design. We generate the covariate $X = (X_1, X_2)^T$, where X_1 follows a uniform distribution on $[-2, 2]$ and X_2 follows a uniform distribution on $[0, 2]$ independent of X_1 . The treatment assignment propensity is $P(A = 1|X) = \exp(-0.5 - 0.5X_1 + 0.5X_2) / \{1 + \exp(-0.5 - 0.5X_1 + 0.5X_2)\}$. The hazard function is generated by $\lambda(t|X, A) = e^{t \times \exp\{-0.5 + X_1 - X_2 - A \times (1.5 + 3X_1 - 2X_2)\}}$; hence, the optimal treatment regime $d^{opt}(X) = \mathbb{I}(1.5 + 3X_1 - 2X_2 > 0)$. The censoring variable follows a uniform distribution on $[0, 10]$ and the censoring rate is approximately 12%. The goal of this simulation study is to find the optimal treatment regime that maximizes the survival probability $S(u, d) = P\{T^*(d) > u\}$ at two time points of interest, $u = 2$ and $u = 3$.

We consider three estimators for $S(u, d^{opt})$; AIPWCC will refer to the doubly robust estimator using $IF_f(a)$ as defined by (4) and uses weighted support vector machine classification, IPWE will refer to the inverse probability weighted estimator which does not include the augmentation terms and uses weighted support vector machine classification, and the outcome regression estimator. For the outcome regression estimator we estimate $S(u, d)$ by $\hat{S}^{OR}(u, d) = N^{-1} \sum_{i=1}^N [d(X_i) \hat{H}(1, u, X_i) + \{1 - d(X_i)\} \hat{H}(0, u, X_i)]$. We consider different scenarios for model misspecification. For the “all models correct” case, the covariate $X = (X_1, X_2)^T$ is used to fit the models for $\pi(X)$, $K_{\mathcal{A}}(a, r, X)$ and $H(a, r, X)$; for the “wrong propensity model” case, only the covariate X_2 is used to fit the models for $\pi(X)$, $K_{\mathcal{A}}(a, r, X)$ and the covariate X is used to fit $H(a, r, X)$; for the “wrong regression model” case, the covariate X is used to fit the models for $\pi(X)$, $K_{\mathcal{A}}(a, r, X)$ while the covariate X_2 is used to fit $H(a, r, X)$, finally, for the “all models wrong” case, the covariate X_2 is used to fit all models for $\pi(X)$, $K_{\mathcal{A}}(a, r, X)$ and $H(a, r, X)$.

Table 1 shows the results from the simulation studies. We consider estimating both $S(u, d^{opt})$ and the difference $\{S(u, d^{opt}) - S(u, a)\}$, $a = 0, 1$. The estimated standard error (SE) is computed using the square root of the sandwich variance estimator given by (8) and (10), respectively, and the coverage rate is computed using the 95% confidence interval derived as the point estimator plus/minus 1.96 times the estimated SE. The underlying true probabilities, i.e., $S(u, d^{opt})$, $S(u, 1)$ and $S(u, 0)$, are computed via Monte-Carlo simulation with 100,000 replicates.

These results demonstrate that the point estimators using AIPWCC are almost unbiased if either the propensity model or the regression model is correctly specified; thus confirming the doubly robustness property. In the “all correct model” scenario, the sandwich variance estimator is consistent with the Monte-Carlo variance, yielding coverage close to 0.95.

With misspecified propensity, the sandwich variance estimator is not guaranteed to be valid. In the simulation scenarios we considered, the estimated standard error underestimated the Monte-Carlo standard error resulting in under coverage of the nominal level of 0.95. As described in Online Resource 4 we also estimated the standard error using a nonparametric

bootstrap for this scenario and the results are presented in Table 1 of the Online Resource. The resulting estimated standard error was very close to the Monte-Carlo standard error with coverage probability close to the nominal level. For the case of misspecified regression, according to the semiparametric theory, the coverage rate should be conservative and in most cases this indeed is what we observe. Again, for this scenario we also estimated the standard error using a nonparametric bootstrap and again the estimated standard error was very close to the Monte-Carlo standard error with coverage probability close to the nominal level.

The IPW estimator is consistent when the propensity score model is correctly specified but the AIPWCC estimator is slightly more efficient whether or not the outcome regression model is misspecified. When the outcome regression model is correctly specified the outcome regression estimator is consistent and is slightly more efficient than the AIPWCC estimator. As expected, when the outcome regression model is misspecified the outcome regression estimator is biased, when the propensity score model is misspecified the IPW estimator is biased, and when both models are misspecified the AIPWCC estimator is biased. We do notice however, that the bias for the AIPWCC estimator is smaller than the others possibly due to the protection offered by double-robustness.

In addition, we calculate the proportion of subjects being assigned to the true underlying optimal treatment regime if they follow the estimated optimal regime,

$N^{-1} \sum_{i=1}^N I\{d(X_i; \eta^{\text{opt}}) = d(X_i; \hat{\eta}^{\text{opt}})\}$. The proportion of individuals getting the correct assignment in our all-correct simulation scenario ranges from around 93% to 97%.

We also considered the estimates of the parameters of the linear boundary of the classifier or the support vector $\eta_0 + \eta_1^T x$. The summary statistics of the estimated coefficients of the linear boundary are given in Online Resource 5. These estimators were highly variable; however, one must keep in mind that the main goal of this paper is to give the best treatment possible to patients leading to the largest value. Classification methods attempt to classify correctly as evidenced by the high proportion getting the correct treatment.

We also considered how these various methods compared in the case of randomized studies where the issue of confounding or misspecification of the propensity score is not of concern. The survival data and censoring data were generated exactly as before, however, treatment assignment was made at random; that is, $\pi(x) = P(A = 1|X = x) = 0.5$ for all x . Here we

estimated the propensity score $\hat{\pi}(X) = N^{-1} \sum_{i=1}^N A_i$ using the sample proportion receiving treatment 1 and estimated the censoring distribution by the treatment-specific Kaplan-Meier estimator making the assumption that the censoring distribution does not depend on covariates. We compared the AIPWCC, IPW and outcome regression estimators both when the survival regression model was and wasn't correctly specified. The results are presented in Table 3 of Online Resource 6 where we demonstrate that no matter if the regression model is correct or not, the resulting survival estimates for AIPWCC and IPW are consistent. The outcome regression estimator showed substantial bias when the outcome regression model was misspecified

In the previous simulation studies, the parameters were chosen so that the regularity conditions for the asymptotic theory would hold. In Online Resource 7, we also explore the impact on the asymptotic theory for the AIP-WCC estimator when the regularity conditions do not hold; that is, when the non exceptional law is violated. In these simulations, we generate the data by letting $\eta_1 = 0$; specifically, the hazard function is generated as $\lambda(t|X, A) = e^t \times \exp\{-0.5 + X_1 - X_2 - \eta_0 \times A\}$. The censoring variable follows a uniform distribution on $[0, 10]$ as before and the censoring rate is close to 20%. Also the propensity score model was generated as previously described. We consider both the cases when $\eta_0 = 0$ (the null hypothesis of no treatment difference) and $\eta_0 > 0$ where $d^{pt}(X) = 1$. The AIPWCC estimator is derived exactly as before. When $\eta_0 > 0$ the optimal treatment regime is $d^{pt}(X) = 1$; that is, to assign every individual to treatment 1. The results when $\eta_0 > 0$ are summarized in Table 4 of Online Resource 7. In this scenario, the estimators for $S(u, d^{pt})$ and $\{S(u, d^{pt}) - S(u, 0)\}$ are consistent, with consistent estimators for the standard errors and correct coverage probability. We note that we do not consider the case of estimating $\{S(u, d^{pt}) - S(u, 1)\}$ as our estimated optimal regime is to give everyone treatment 1. In the case of the null hypothesis where $\eta_0 = 0$ and the underlying truth is $S(u, d^{pt}) = S(u, 1) = S(u, 0)$, the regularity conditions for the asymptotic theory do not hold and the classification procedure estimates a slightly “over-optimistic” value for the optimal treatment regime. These results are found in Table 5 of Online Resource 7.

All the simulations above were conducted with two covariates. We also considered the case where we included additional covariates that were not important. The reason of this investigation was two-fold; (i) to see if the methods would apply when considering more covariates and (ii) to assess the impact of the penalty term in the SVM which was derived using cross-validation. The data were generated exactly as before but we also generated an additional 25 independent standard normal random variables which were independent of all other data (i.e., 25 additional unimportant covariates). We considered both the case with no penalty term in the SVM ($c = 0$) and when the penalty term c was computed via cross-validation. The results are summarized in Tables 6–9 in Online Resource 8. We see the estimated coefficient of the unimportant covariates are consistent with the truth (i.e., 0). We also see that the Monte-Carlo standard error of the estimates for the unimportant variables were substantially smaller (i.e, more concentrated about zero) when we used cross-validation for the penalty term. Although there is some impact on the estimates for the important covariates, the estimated survival probability is close to the truth, however, the coverage of the confidence interval was slightly lower than the nominal level.

In all the simulations above the censoring distribution was generated independent of X and A . We also conducted additional simulations with dependent censoring where we allowed the censoring distribution to depend on both X and A . The description and results are given in Online Resource 9. The results are similar to the case of independent censoring.

6 ASCERT Analysis

In this section, we apply the proposed method to data from the ASCERT study. The ASCERT study was a retrospective analysis of patients who had either two-vessel or three-vessel coronary artery disease and were treated either by surgical revascularization (coronary

artery bypass grafting) or catheter-based revascularization (percutaneous coronary intervention). As in most observational survival studies, the treatment assignment is not random and the outcome of this study is censored survival time. For this analysis, we considered the subset of 7,391 eligible patients from 54 hospitals as described by Bai et al. (2013). Twenty-eight covariates were used on the sample which included demographics (e.g., age, sex), risk factors (e.g., body mass index, smoking), symptoms and history of cardiovascular disease (e.g., chest pain, congestive heart failure), and comorbidities (e.g., diabetes).

Of the 7,391 patients, 1,016 were observed to die during the course of the study of which 966 died before 4 years. Therefore, for the primary outcome we considered survival at 4 years; that is, $I\{T^*(d) \leq 4\}$ and the goal of the analysis was to estimate the optimal treatment regime that maximizes the survival probability at 4 years. We estimated the value function $\hat{V}(\hat{d}_{opt})$ using the ASCERT dataset. The logistic regression model was used to estimate $\pi(X)$, and stratified proportional hazards regression models were used to estimate $K(a, r, X)$ and the regression model $H(a, r, X)$ as described in the simulation section. We considered the restricted class of regimes in the form of hyperplanes and used the weighted support vector machine method as the classification technique to obtain estimators for the parameters in the hyperplane, as described in the previous simulation section. In the analysis all the covariates were standardized to have mean zero and variance one. Five of the 28 variables had estimated coefficients that were virtually zero and were eliminated. The list of the 23 covariates used and the estimated coefficients of the optimal hyperplane (on the original scale) are given in Online Resource 10.

The estimated survival probability at four years for the estimated optimal regime is equal to $\hat{S}(4, \hat{d}^{opt}) = 0.862$, whereas, the estimated survival probability at four years on bypass surgery and percutaneous coronary intervention are $\hat{S}(4, 1) = 0.841$ and $\hat{S}(4, 0) = 0.816$ respectively. We estimated the difference between the value of the optimal restricted regime and the values of each of the treatment-specific regimes and the corresponding standard errors using (9) and (10). The results are shown in Table 2. We conclude that the survival probability at four years following the optimal restricted regime is significantly larger than each of the treatment specific survival probabilities.

As a result of this analysis, 5024 or 68% patients are recommended to receive bypass surgery treatment; i.e., $\hat{\eta}_0 + \hat{\eta}_1^T X_i > 0$ and 2367 or 32% patients to receive percutaneous coronary intervention treatment; i.e., $\hat{\eta}_0 + \hat{\eta}_1^T X_i \leq 0$. For the 5024 patients recommended to bypass surgery, we used the methods in this paper to estimate the potential survival probability at 4 years had these patients all received bypass surgery or percutaneous coronary intervention respectively; namely, $\hat{E}[f\{T^*(a)\} | \hat{\eta}_0 + \hat{\eta}_1^T X_i > 0]$, $a = 0, 1$, where $I\{T^*(a) \leq 4\}$, and the results are shown in Table 3. For this subset of patients we estimate a 6.7% increase in the survival probability at four years on bypass surgery compared to percutaneous coronary intervention. Similarly, for those 2367 patients recommended to receive percutaneous coronary intervention we estimate a 7.8% increase in the survival probability at four years on percutaneous coronary intervention compared to bypass surgery.

Because bypass surgery is a more invasive procedure than percutaneous coronary intervention and often associated with severe side effects, some patients may be willing to consider a quality adjusted outcome of interest; namely, $I\{T^*(d) - \mu d, \mu > 0\}$, where μ is the amount a patient is willing to trade on their survival outcome of interest to avoid the potential side effects of treatment $d = 1$ (bypass surgery). In that case, the μ -quality adjusted optimal regime is given by $d^{opt,\mu}(X) = I\{CF(X) - \mu > 0\} = I\{CF(X) > \mu\}$. That is, a patient with covariate X would choose treatment 1 if their expected increase in their outcome on treatment 1 would exceed their expected outcome on treatment 0 by at least μ units. Consequently, the μ -quality adjusted optimal restricted regime is estimated by maximizing,

in η , $N^{-1} \sum_{i=1}^N d(X_i, \eta) \{\widehat{CF}(X_i) - \mu\}$ and the resulting estimator is denoted by $\hat{\eta}^{opt,\mu}$. As before, $\hat{\eta}^{opt,\mu}$ can be obtained by using weighted support vector machine.

The amount μ may vary by patient and in order to give some guidance to patients we consider a range of μ 's, $0 = \mu_1 < \mu_2 < \dots < \mu_K$. For each μ_k , $k = 1, \dots, K$, the weighted support vector machine finds η_{0k} , η_{1k} that minimizes

$$\sum_{i=1}^N W_{ik} \psi\{Y_{ik}(\eta_{0k} + \eta_{1k}^T X_i)\} + c \|\eta_{1k}\|_1, \text{ where}$$

$W_{ik} = |\widehat{CF}(X_i) - \mu_k|$, $Y_{ik} = 2I\{\widehat{CF}(X_i) - \mu_k > 0\} - 1$, $\psi(t) = \max(1 - t, 0)$ is the Hinge loss function, c is the tuning parameter and $\|\eta_{1k}\|_1$ is the L_1 norm of vector η_{1k} . The difficulty however is that for a series of μ 's, $0 = \mu_1 < \mu_2 < \dots < \mu_K$, it is not guaranteed that the resulting estimated treatment regimes will be nested in the sense that $d(x, \hat{\eta}^{opt,k}) \leq d(x, \hat{\eta}^{opt,k'})$ if $k < k'$ for all x ; that is, if a patient with covariate x is recommended treatment 0 for some value of μ_k , then they should also be recommended treatment zero for any $\mu_{k'} > \mu_k$. Therefore, to guarantee the proper nesting, we propose for a predetermined set of μ 's, $0 = \mu_1 < \mu_2 < \dots < \mu_K$, to find η_{0k} , $k = 1, \dots, K$ and η_1 that minimizes

$$\sum_{k=1}^K \sum_{i=1}^N W_{ik} \psi\{Y_{ik}(\eta_{0k} + \eta_1^T X_i)\} + c \|\eta_1\|_1, \text{ subject to the restriction } \eta_{01} \leq \eta_{02} \leq \dots \leq \eta_{0K}. \text{ With common } \eta_1 \text{ and monotone } \eta_{0k}, \text{ the hierarchical order of the resulting optimal treatment regime is now guaranteed. Consequently we can derive a score } s(x) = \eta_{01} + \eta_1^T x \text{ and then the estimated } \mu_k\text{-quality adjusted optimal restricted regime would be given by } I\{s(x) > \alpha_k\}, \text{ where } \alpha_k = \eta_{01} - \eta_{0k}, k = 1, \dots, K.$$

Zhao et al. (2013) also considered different scoring methods for estimating $CF(X)$ and identifying subgroups of patients for which $\widehat{CF}(X) \geq \mu$ for various values of μ . However, in their case their interest focused on estimating the overall treatment difference within this subgroup; namely, $AD(\mu) = E[f\{T^*(1)\} - f\{T^*(0)\} | \widehat{CF}(X) \geq \mu]$; whereas, our focus was to derive the optimal restricted regime d_η that maximizes $E[I\{T^*(d_\eta)\} - \mu d_\eta]$ in an efficient manner as possible.

A secondary objective of our analysis was to estimate the μ_k -quality adjusted optimal treatment regimes that maximize the survival probability at 4 years with different level of μ_k . We chose $\mu = (0, 0.005, 0.010, 0.015, 0.020, 0.025)$. Results are shown for each of these in Table 4 together with the estimates for the treatment-specific survival probability at four years for the different subsets of patients. As expected, more patients are recommended to switch from bypass surgery treatment to percutaneous coronary intervention treatment as the

value of μ increases. For different levels μ_k , the coefficients that make up the score $s(x)$ are given in Online Resource 11 and the corresponding α_k is given in Table 5. Equipped with their score, an individual can use this table of α 's to decide which treatment they should receive based on how much of the survival probability they are willing to trade in order to avoid the possible complications of bypass surgery.

7 Conclusion and Discussion

From a classification perspective, we derived the optimal treatment regime maximizing a survival related parameter of interest, including but not limited to survival probability, mean truncated lifetime, etc. Following the framework of Zhang et al. (2012b), the optimal treatment regime problem is transformed into a weighted classification problem. The weights are calculated using the powerful semiparametric theory and hence they are locally-efficient and doubly-robust. We also derived an estimator for the asymptotic variance of the estimated parameters. The simulation studies showed that the large sample properties of these estimators and corresponding confidence intervals worked well in most cases where we expected consistent estimators.

Moreover, like in the case of our motivating example of the ASCERT data, we can derive quality adjusted optimal treatment regimes as well. This allows the patients/physicians to choose the treatment depending on how much risk they are willing to trade in order to avoid a more invasive or toxic treatment.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This research was supported by NIH grants R01 HL118336 and P01 CA142538. The authors gratefully acknowledge Dr. Sean O'Brien, Dr. William Weintraub, Dr. Fred Edwards, and the ASCERT investigator team for assembling the study database.

References

- Bai X, Tsiatis AA, O'Brien SM. Doubly-robust estimators of treatment-specific survival distributions in observational studies with stratified sampling. *Biometrics*. 2013; 69:830–839. [PubMed: 24117096]
- Brinkley J, Tsiatis AA, Anstrom KJ. A generalized estimator of the attributable benefit of an optimal treatment regime. *Biometrics*. 2009; 21:512–522.
- Cortes C, Vapnik V. Support-vector networks. *Machine Learning*. 1995; 20(3):273–297.
- Hubbard, AE., van der Laan, MJ., Robins, JM. Nonparametric locally efficient estimation of the treatment specific survival distributions with right censored data and covariates in observational studies. In: Halloran, E., Berry, D., editors. *Statistical Models in Epidemiology: The Environment and Clinical Trials*. New York: Springer; 1999. p. 134178
- Moodie E, Richardson TS, Stephens D. Demystifying optimal dynamic treatment regimes. *Biometrics*. 2007; 63:447–455. [PubMed: 17688497]
- Murphy SA. Optimal dynamic treatment regimes (with discussion). *Journal of the Royal Statistical Society, Series B*. 2003; 65:331–366.

- Orellana L, Rotnitzky A, Robins JM. Dynamic regime marginal structural mean models for estimation of optimal treatment regimes, part I: Main content. *International Journal of Biostatistics*. 2010; 6(2) Article 8.
- Robins, JM. Optimal structured nested models for optimal sequential decisions. In: Lin, DY., Heagerty, PJ., editors. *Proceedings of the Second Seattle Symposium on Biostatistics*. New York: Springer; 2004. p. 189-326.
- Robins J, Orellana L, Rotnitzky A. Estimation and extrapolation of optimal treatment and testing strategies. *Statistics in Medicine*. 2008; 27:4678–4721. [PubMed: 18646286]
- Rubin DB. Bayesian inference for causal effects: The role of randomization. *Annals of Statistics*. 1978; 6:34–58.
- Tsiatis, AA. *Semiparametric Theory and Missing Data*. New York: Springer; 2006.
- van der Laan MJ, Luedtke AR. Targeted Learning of the Mean Outcome Under an Optimal Dynamic Treatment Rule. U.C. Berkeley Division of Biostatistics Working Paper Series. Working Paper. 2014:325.
- Zhang B, Tsiatis AA, Laber EB, Davidian M. A robust method for estimating optimal treatment regimes. *Biometrics*. 2012a; 68:1010–1018. [PubMed: 22550953]
- Zhang B, Tsiatis AA, Davidian M, Zhang M, Laber EB. Estimating optimal treatment regimes from a classification perspective. *Stat*. 2012b; 1:103–114. [PubMed: 23645940]
- 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; 108:527–539. [PubMed: 24058223]
- Zhao Y, Zeng D, Rush AJ, Kosorok MR. Estimating individualized treatment rules using outcome weighted learning. *Journal of the American Statistical Association*. 2012; 107:1106–1118. [PubMed: 23630406]
- Zhao Y, Kosorok MR, Zeng D. Reinforcement learning design for cancer clinical trials. *Statistics in Medicine*. 2009; 28:3294–3315. [PubMed: 19750510]

Table 1

Linear Optimal Treatment Regime, Weighted Support Vector Machine Classification.

	$\hat{S}(u, \hat{d}^{pr})$			MC SE of $\hat{S}(u, \hat{d}^{pr})$			Est. SE of $\hat{S}(u, \hat{d}^{pr})$			Coverage		
	$u = 2$	$u = 3$		$u = 2$	$u = 3$		$u = 2$	$u = 3$		$u = 2$	$u = 3$	
AIPWCC, all models correct	0.62	0.33		0.035	0.037		0.035	0.036		0.95	0.95	
AIPWCC, wrong propensity model	0.62	0.33		0.033	0.034		0.028	0.027		0.91	0.87	
AIPWCC, wrong regression model	0.62	0.33		0.035	0.039		0.046	0.045		0.99	0.97	
AIPWCC, all models wrong	0.62	0.33		0.037	0.034		0.033	0.030		0.08	0.23	
IPWE, correct model	0.62	0.33		0.035	0.039		-	-		-	-	
IPWE, wrong model	0.62	0.33		0.039	0.032		-	-		-	-	
Outcome regression, correct model	0.62	0.33		0.029	0.032		-	-		-	-	
Outcome regression, wrong model	0.62	0.33		0.028	0.024		-	-		-	-	
<hr/>												
	$\hat{S}(u, \hat{d}^{pr}) - \hat{S}(u, 1)$			MC SE of $\hat{S}(u, \hat{d}^{pr}) - \hat{S}(u, 1)$			Est. SE of $\hat{S}(u, \hat{d}^{pr}) - \hat{S}(u, 1)$			Coverage		
	$u = 2$	$u = 3$		$u = 2$	$u = 3$		$u = 2$	$u = 3$		$u = 2$	$u = 3$	
AIPWCC, all models correct	0.29	0.13		0.031	0.027		0.030	0.026		0.93	0.93	
AIPWCC, wrong propensity model	0.29	0.13		0.029	0.025		0.025	0.019		0.91	0.84	
AIPWCC, wrong regression model	0.29	0.13		0.033	0.029		0.034	0.030		0.95	0.94	
AIPWCC, all models wrong	0.29	0.13		0.029	0.023		0.026	0.020		0.80	0.79	
IPWE, correct model	0.29	0.13		0.033	0.029		-	-		-	-	
IPWE, wrong model	0.29	0.13		0.030	0.021		-	-		-	-	
Outcome regression, correct model	0.29	0.13		0.026	0.024		-	-		-	-	
Outcome regression, wrong model	0.29	0.13		0.025	0.016		-	-		-	-	
<hr/>												
	$\hat{S}(u, \hat{d}^{pr}) - \hat{S}(u, 0)$			MC SE of $\hat{S}(u, \hat{d}^{pr}) - \hat{S}(u, 0)$			Est. SE of $\hat{S}(u, \hat{d}^{pr}) - \hat{S}(u, 0)$			Coverage		
	$u = 2$	$u = 3$		$u = 2$	$u = 3$		$u = 2$	$u = 3$		$u = 2$	$u = 3$	
AIPWCC, all models correct	0.30	0.20		0.028	0.027		0.029	0.027		0.96	0.95	

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

AIPWCC, wrong propensity model	0.30	0.20	0.30	0.20	0.027	0.025	0.027	0.021	0.94	0.89
AIPWCC, wrong regression model	0.30	0.20	0.30	0.20	0.029	0.030	0.036	0.035	0.99	0.97
AIPWCC, all models wrong	0.30	0.20	0.26	0.16	0.029	0.026	0.027	0.024	0.63	0.55
IPWE, correct model	0.30	0.20	0.30	0.20	0.030	0.031	-	-	-	-
IPWE, wrong model	0.30	0.20	0.20	0.13	0.030	0.025	-	-	-	-
Outcome regression, correct model	0.30	0.20	0.30	0.20	0.024	0.024	-	-	-	-
Outcome regression, wrong model	0.30	0.20	0.07	0.07	0.024	0.022	-	-	-	-

Table 2

ASCERT analysis with original contrast function, inference on the difference between optimal regime and treatment-specific survivals.

$\hat{S}(u, \hat{d}^{opt})$	$\hat{S}(u, \hat{d}^{opt}) - \hat{S}(u, 1)(SE)$	CI of $S(u, \hat{d}^{opt}) - \hat{S}(u, 1)$	$\hat{S}(u, \hat{d}^{opt}) - \hat{S}(u, 0)(SE)$	CI of $S(u, \hat{d}^{opt}) - S(u, 0)$
0.862	0.021 (0.008)	(0.005, 0.036)	0.046 (0.009)	(0.028, 0.064)

Table 3

ASCERT analysis with original contrast function. CABG stands for coronary artery bypass grafting; PCI stands for percutaneous coronary intervention.

Assigned Treatment	Number of patient	Survival Probability (%)	
		CABG	PCI
CABG	5024	86.9	80.2
PCI	2367	76.5	84.3

Table 4

ASCERT analysis with quality adjusted contrast function, new $\widehat{CF}(X) = \widehat{CF}(X) - \mu$, monotone regime.
 CABG stands for coronary artery bypass grafting; PCI stands for percutaneous coronary intervention.

μ	Assigned Treatment	Number of patient	Survival Probability (%)	
			CABG	PCI
0	CABG	4511	85.8	78.5
	PCI	2880	80.1	86.1
0.005	CABG	4234	85.4	77.4
	PCI	3157	81.6	86.4
0.010	CABG	4046	86.0	78.7
	PCI	3345	80.5	85.8
0.015	CABG	4033	85.4	77.4
	PCI	3358	81.6	86.4
0.020	CABG	3859	85.2	77.4
	PCI	3532	82.3	86.0
0.025	CABG	3646	85.1	77.0
	PCI	3745	82.6	85.8

Table 5

ASCERT analysis, relationship between score and μ values.

μ	0	0.005	0.010	0.015	0.020	0.025
σ	0	0.050	0.087	0.090	0.123	0.161