HHS Public Access

Author manuscript

J Biopharm Stat. Author manuscript; available in PMC 2019 January 01.

Published in final edited form as:

J Biopharm Stat. 2018 ; 28(2): 362–381. doi:10.1080/10543406.2017.1380036.

Estimating the Optimal Personalized Treatment Strategy Based on Selected Variables to Prolong Survival via Random Survival Forest with Weighted Bootstrap

Jincheng Shen[*], **Lu Wang**[†], **Stephanie Daignault**[†], **Daniel E. Spratt**[‡], **Todd M. Morgan**[§], and **Jeremy M.G. Taylor**[†]

[*]Department of Population Health Sciences, University of Utah School of Medicine, Salt Lake City, UT 84108

[†]Department of Biostatistics, University of Michigan, Ann Arbor, MI 48109

[‡]Department of Radiation Oncology, University of Michigan, Ann Arbor, MI 48109

[§]Department of Urology, University of Michigan, Ann Arbor, MI 48109

Abstract

A personalized treatment policy requires defining the optimal treatment for each patient based on their clinical and other characteristics. Here we consider a commonly encountered situation in practice, when analyzing data from observational cohorts, that there are auxiliary variables which affect both the treatment and the outcome, yet these variables are not of primary interest to be included in a generalizable treatment strategy. Furthermore, there is not enough prior knowledge of the effect of the treatments or of the importance of the covariates for us to explicitly specify the dependency between the outcome and different covariates, thus we choose a model that is flexible enough to accommodate the possibly complex association of the outcome on the covariates. We consider observational studies with a survival outcome and propose to use Random Survival Forest with Weighted Bootstrap (RSFWB) to model the counterfactual outcomes while marginalizing over the auxiliary covariates. By maximizing the restricted mean survival time, we estimate the optimal regime for a target population based on a selected set of covariates. Simulation studies illustrate that the proposed method performs reliably across a range of different scenarios. We further apply RSFWB to a prostate cancer study.

Keywords

Inverse Probability Weighting; Optimal Treatment Regime; Random Survival Forest; Weighted Bootstrap

1 Introduction

It has been shown that patients can exhibit significant heterogeneity in response to treatments in many different diseases (Ishigooka et al. 2000; Rothwell 2005). The emerging field of personalized medicine, which focuses on making treatment decisions for an individual patient based on his/her clinical, genomic, and other information, is of considerable interest, as it has the potential of maximizing the treatment benefits (Piquette-

Miller and Grant 2007). When customized therapy for each patient is assigned, individuals could benefit the most from the treatment they receive, therefore the optimal treatment effect can be achieved for the whole population.

Studies of biological mechanisms have identified several biomarkers that affect patients' responses to certain drugs and have been successfully applied clinically in treatment assignment (Ellsworth et al. 2010). To better understand the treatment mechanism, developing statistical approaches to study the heterogeneity of the treatment from data is of great importance. Various methods have been developed for testing and identifying subgroups of patients who can effectively respond to the treatment under investigation from randomized controlled trial data (Foster et al. 2011; Doove et al. 2014). However, observational studies are the most commonly available data source in practice. Qian and Murphy (2011), among others, have proposed causal inference methods to learn about the optimal regime from observational data. Commonly, methods in this category always involve postulating some parametric structure for the counterfactual models (e.g. Zhang et al. 2012), however, the optimal treatment regime identified from such models could suffer from bias if those parametric models are misspecified. Recently, machine learning based approaches are becoming more and more popular in this area, as they generally put less structural assumptions on the models and thus also acquire more flexibility on the form of the optimal regime (Zhao et al. 2012; Laber and Zhao 2015; Zhao et al. 2015; Lu et al. 2017).

One important feature of observational studies is that they typically have the tendency to collect as many variables as possible, particularly in situations where the underlying mechanisms of the treatments and the disease are not well understood. In certain situations, there could be some confounders of the effect of treatment on the outcome that may not be considered as primary or desirable factors to include for a generalizable treatment rule. For example, variables such as social economic status and health insurance plan may be collected and used when making treatment assignment in a clinic, however, such factors are not helpful in situations where there may be changes in the health policies and economic environment. The primary interest of a future treatment assignment rule should be based on more intrinsic factors, such as the patient's biological and clinical characteristics. Variables which are associated with treatment assignment in the observational data but not desirable to be included in a generalizable treatment rule for the targeted population, are called auxiliary variables. Another situation when auxiliary variables exist is when some markers are hard to obtain for the targeted population due to economic, logistical or other reasons. Directly excluding such auxiliary variables or their interactions with treatment from the outcome models may lead to severe model misspecification, especially when the auxiliary variables are correlated with the confounders of interest.

When the primary outcome of interest is a censored survival time, the task of properly accounting for the auxiliary variables is even more challenging. In this paper, we propose the Random Survival Forest with Weighted Bootstrapping (RSFWB) method to correctly model the survival outcome conditional on the selected subset of covariates. The method consists of a modified version of Random Survival Forest to provide a flexible model, and an inverse probability weighted bootstrap procedure to account for the potential selection bias on the auxiliary variables. The optimal regime is identified as the one that maximizes the estimated

counterfactual restricted mean survival time, which allows comparison of survival time nonparametrically. We introduce the notation and more details of the proposed method in the next section. In Section 3, we present the numerical results under different simulation schemes, and in Section 4, we apply the proposed method to a prostate cancer study and estimate the generalizable optimal regime to lengthen the time to clinical failure after treatment. Some concluding points and a discussion are included in Section 5.

2 Notation and Methods

Consider a study of n patients. Let $\mathbf{W}_i = \left(\mathbf{X}_i^T, \mathbf{Z}_i^T\right)^T$ denote a d-dimensional vector of baseline covariates for patient i, $i = 1,\ldots, n$, where $\mathbf{X}_i$ denotes the subset of d_1 covariates that would be included to form the generalizable treatment regime ($d_1 < d$), and $\mathbf{Z}_i$ is a d_2 -dimensional vector ($d_2 = d - d_1$) denoting the auxiliary variables. We consider two possible treatments, which could both be active treatments, but for convenience in this paper we will arbitrarily denote one of them as treatment and the other as no treatment (control). Let A_i be the observed treatment status, with $A_i = 1$ if the patient received the treatment, and $A_i = 0$ if the patient was on the control arm. We consider the case where the treatment decision is only made at time zero, and therefore A_i is time-independent. For the outcome, let T_i^0 denote the survival time if subject i did not receive the treatment (control), and T_i^1 denote the survival time if subject i received the treatment, one of which is counter to the fact, because the patient can only be assigned to one arm. Let T_i denote the actual survival time for subject i and C_i is the censoring time. The observed outcome is then (Y_i, Δ_i) where $Y_i = \min(T_i, C_i)$ and $\Delta_i = I(T_i \leq C_i)$. We can use the framework of causal inference to connect the observational data to the counterfactual outcomes T_i^0 and T_i^1, and then the outcomes of patients for arbitrary regimes. Specifically, to facilitate the estimation for these counterfactual quantities, we need to impose the following causal inference assumptions:

1. Consistency assumption: $T_i = A_i T_i^1 + (1 - A_i) T_i^0$;
2. Positivity assumption: $P(A_i|\mathbf{W}_i) > \varepsilon$ for some small positive ε for $A_i = 0,1$;
3. No unmeasured confounders assumption (NUCA): $T_i^a \coprod A_i | \mathbf{W}_i$, for $a = 0,1$.

In addition, we assume that the counterfactual survival time is independent of the censoring mechanism given the other covariates, i.e., $T_i^a \coprod C_i | \mathbf{X}_i$, for $a = 0,1$, which also guarantees that $T_i \coprod C_i | \mathbf{X}_i$. We focus on identifying the optimal regime based only on $\mathbf{X}_i$, which we denote by $g(\mathbf{X}_i) \in \{0,1\}$. For simplicity, we will suppress the patient index i in the future when no confusion exists.

2.1 Flexible Modeling of the Counterfactual Outcomes

The optimal treatment regime would assign each individual the treatment that leads to the most favorable counterfactual outcome. However, in the observational study, one and only one of the counterfactual outcomes can be observed for each individual. Therefore, one needs to employ causal models to make statistical inferences on the counterfactual

outcomes. Typically, some parametric models are used in the modeling strategies, and therefore the method's performance depends on whether those parametric models are specified correctly or not. When the outcome is subject to censoring, the Cox proportional hazard regression model (Cox 1972) and its extensions have been widely used, where the semiparametric structure of the model can accommodate various families of distribution functions. However, a Cox model typically uses a linear combination of covariates, and this may not be appropriate. If it is desired to model the counterfactual outcomes in a less constrained way, a survival tree (LeBlanc and Crowley 1995) is one method that allows for more flexibility. In tree methods, a binary tree is grown by dividing patients at each node into two groups, where the split is chosen to maximize a criterion function which measures heterogeneity. It has the feature of automatically identifying interactions between the variables without prespecifying their structure. While tree methods can be unstable, ensemble approaches can effectively compensate for the instability of tree based models by combining many trees (Breiman 2001; Bou-Hamad et al. 2011). For right-censored survival data, the Random Survival Forest method (Ishwaran et al. 2008) draws B bootstrap samples from the original data, survival trees are then grown for each bootstrap sample. At each node of each tree, a random set of p candidate variables are chosen and the node is split using the candidate variable that maximizes the survival difference between daughter nodes. After growing the tree to full size, the Cumulative Hazard Function (CHF) is obtained from each tree, and the ensemble estimation is then calculated by averaging over these CHFs.

2.2 Inverse Probability Weighting

In observational studies, the covariates may have different distributions across different subgroups for both counterfactual and observational data, this would apply to both $\mathbf{X}$ and $\mathbf{Z}$ in our setting. In general, for regime g (X) based on $\mathbf{X}$, we have $E[I\{A=g(\mathbf{X})\}T|\mathbf{X},\mathbf{Z}] \neq E[I\{A=g(\mathbf{X})\}T|\mathbf{X}]$unless T is independent of the auxiliary variables $\mathbf{Z}$ given $\mathbf{X}$. Since our primary interest is in $E[I\{A=g(\mathbf{X})\}T|\mathbf{X}]$ for regime $g(\mathbf{X})$, we need to correctly model the counterfactuals conditional on $\mathbf{X}$. To this end, it is important to account for the discrepancy of the distribution of $\mathbf{Z}$ between the observational data and the counterfactual data, especially for nonparametric and/or semi-parametric models where the model structure is not fully specified. Inverse probability weighting can be employed to account for the discrepancy between observational data distribution law and counterfactual data distribution law, and correctly estimate the conditional mean counterfactuals for each subject. Here, given the NUCA assumption, $T^a \coprod A \,|\, \mathbf{W}$, we can estimate the probability of treatment assignment conditional on $\mathbf{W}$, $\boldsymbol{\pi}(\mathbf{W})$, using the following logistic model,

$$\text{logit}P(A=1|\mathbf{W}) = \text{logit}\{\pi(\mathbf{W})\} = \theta_0 + \mathbf{W}^T\mathbf{q}_1 \quad (1)$$

Note that here $\mathbf{Z}$ is included as part of $\mathbf{W}$. Then $\hat{\pi}(\mathbf{W}) = \hat{P}(A=1|\mathbf{W})$, which we denote by $\hat{p}$, can be used to calculate the estimated weights for regime $g=1$ and $g=0$ given by

$$wt^1 = \frac{I(A=1)}{\hat{p}} \text{ and } wt^0 = \frac{I(A=0)}{1-\hat{p}}$$

To illustrate the role of these weights, let $\mathcal{P}$ denote the measure generated by the observational data, and $\mathcal{P}^g$ denote the corresponding measure generated by the counterfactual data under any arbitrary regime g. Because $d\mathcal{P} = p(T|\mathbf{X},\mathbf{Z},A)p(A|\mathbf{X},\mathbf{Z})p(\mathbf{X}|\mathbf{Z})p(\mathbf{Z})$ and $d\mathcal{P}^g = p(T|\mathbf{X},\mathbf{Z},A)I\{A = g(\mathbf{X})\}p(\mathbf{X}|\mathbf{Z})p(\mathbf{Z})$, we have

$$\frac{d\mathcal{P}^g}{d\mathcal{P}} = \frac{p(T|\mathbf{X},\mathbf{Z},A)I\{A = g(\mathbf{X})\}p(\mathbf{X}|\mathbf{Z})p(\mathbf{Z})}{p(T|\mathbf{X},\mathbf{Z},A)p(A|\mathbf{X},\mathbf{Z})p(\mathbf{X}|\mathbf{Z})p(\mathbf{Z})} = \frac{I\{A = g(\mathbf{X})\}}{p(A|\mathbf{X},\mathbf{Z})} \equiv wt^g$$

For T, the outcome in the observational data and the corresponding regime g specific counterfactual outcome T^g, we have

$$E(T^g) = E^g(T) = \int T d\mathcal{P}^g = \int T\frac{d\mathcal{P}^g}{d\mathcal{P}}d\mathcal{P} = E\left[\frac{I\{A = g(\mathbf{X})\}}{p(A|\mathbf{X},\mathbf{Z})}T\right]$$

Similarly,

$$E(T^g|\mathbf{X}) = E\left[\frac{I\{A = g(\mathbf{X})\}}{p(A|\mathbf{X},\mathbf{Z})}T\middle|\mathbf{X}\right]$$

In general, this is not equal to $I\{A \sim g(\mathbf{X})\}\cdot E(T|\mathbf{X})$ or $I\{A \sim g(\mathbf{X})\}\cdot E(T|\mathbf{X}) / p(A|\mathbf{X})$, which correspond to the case of using either no weights or weights based only on $\mathbf{X}$. Thus the weights defined above (wt^g 's) provide the correct adjustment when estimating $E(T^g|\mathbf{X})$ from the observational data.

2.3 Estimating the Optimal Treatment Regime with RSFWB

In order to make valid causal inference while providing the flexibility to account for potentially complicated heterogeneity in the outcome model, we incorporate wt^g into Random Survival Forest (RSF). A possible implementation is to incorporate the weights into the splitting criterion. This may not be appropriate for our purpose here, because it would lead to the model where the hypothetical group of patients in the weighted population which are represented by one single patient in the original cohort would always end up in the same leaf for any tree. Here, we propose the RSFWB method, where we propose to directly implement the weights through a weighted bootstrap procedure. The intuition behind the weighted bootstrap is the following: in the standard bootstrap the weights are $1/n$, it thus gives a sample resembling one from the original population, while for weighted bootstrap, observations are resampled with probability proportional to weights thus giving a sample resembling one from a different population. The idea of weighted bootstrap has been used before (Norazan et al. 2009; Makarenkov et al. 2010; Barbe and Bertail 2012), with a major application being to account for potential sampling bias (Nahorniak et al. 2015; Xu et al. 2016). As mentioned earlier, when modeling $E(T^g|\mathbf{X})$, we need to marginalize Z in the counterfactual data, while the conditional distribution of $\{Z|\mathbf{X} = \mathbf{x}\}$ in regime g specific counterfactual data would be different than the one in the observed regime g compliant sample $\{Z|\mathbf{X} = \mathbf{x}, A = g(\mathbf{x})\}$. Here we use this weighted bootstrap resampling procedure to mimics the scenario that the samples were drawn from regime g specific counterfactual data

and adjust for the selection bias when sampling from observational data. RSF models are then built on each weighted bootstrap sample. The final estimate will then be obtained by counting the contribution of each RSF estimate equally. We implement the proposed method in R, where we use *cforest* () function to serve as the base leaner on each weighted bootstrap sample. This is one of the earliest implementation of a random forest type method for survival data, where the ensemble forest models are built based on conditional inference trees (Hothorn et al. 2006; Strobl et al. 2009). To proceed, we propose to use the RSFWB method separately for the two counterfactual outcomes. For treatment counterfactual T^1, we draw bootstrap samples with p_i^1 as the weight for subject i, where $p_i^1 = wt_i^1/(\sum_{j=1}^n wt_j^1)$. The R function *treeresponse*() is used to obtain the survival probabilities $S^{1(m)}(t)$ (Kaplan-Meier estimator) from the model built on the m th weighted bootstrap sample, $m = 1,\ldots, B$. The final survival estimate is obtained by averaging the cumulative hazard function (CHF) at each time point with equal weights over the B bootstrap sample specific estimates,

$$\hat{S}^1(t) = \exp\{-\sum_{m=1}^{B} -\log\hat{S}^{1(m)}(t)\}$$

The survival estimate for the counterfactual of being assigned to control arm $\hat{S}^0(t)$ can be constructed following the same procedure with weights $p_i^0 = wt_i^0/(\sum_{j=1}^n wt_j^0)$.

The performance of machine learning methods always depends on the choice of tuning parameters. For our proposal, the tuning parameters we have are the number of covariates to be considered at each split for the conditional inference tree (*mtry*) and the number of trees (the number of RSF models and the number of trees in each model) to grow. For the later parameter, due to practical concern of computation load, we fix the number of weighted bootstrap samples (the number of RSF models to build) as $B = 200$, and perform selection on *mtry* and *ntree* based on single forest model. Following Mogensen et al. (2012), we use the Integrated Brier Score (IBS) to assess the model performance over a grid of tuning parameter values:

$$\text{IBS}(\text{BS}, \tau) = \frac{1}{\tau}\int_0^\tau BS(t, \hat{S})dt$$

where $\text{BS}(t, \hat{S}) = n^{-1}\sum_{i=1}^n \{Y_i(t) - \hat{S}(t|\mathbf{X}_i)\}^2$ is the Brier score at time t, with $Y_i(\text{t}) = I(T_i \leq \text{t})$ denoting whether subject i is at risk at time t. A smaller IBS would suggest better prediction accuracy in $\hat{S}$. 5-fold cross-validation is used to avoid the potential issue of overfitting. Although growing more trees is likely to give better prediction accuracy, it also substantially increases the computation burden. Thus we limit our grid search for *ntree* only from {50,100}, and *mtry* is also chosen from a given set of values ({2,4,8,16} for simulation, {1,2,3,4,5,6} for data application). The one set of parameters that yields the smallest IBS is then applied to RSFs across all weighted bootstrap samples. More details on tuning parameter selection procedure and results can be found in Appendix A4.

By comparing $\hat{S}^0(t|\mathbf{x}_i)$ and $\hat{S}^1(t|\mathbf{x}_i)$, one determines the optimal treatment for subject i. Since the primary outcome is right censored in our case, we choose to compare the restricted mean survival time for the regime g -specific counterfactual $\mu^g = E(\min\{T^g, \tau\})$ where we choose $\tau > 0$ to be the longest follow-up time. This may lead to slightly different conclusion than the optimal regime determined base on $E(T^g)$ (more details can be found in Appendix A2). The optimal regime g^{opt} is then the one that provides the largest μ^g over the regime space $\mathcal{G} = \{g : g(\mathbf{X}) \in 0,1\}$, that is, $g^{\text{opt}} = \arg\max_{g \in \mathcal{G}} \mu^g$. Note that

$$\begin{aligned}
\mu^g &= E\left[E\left\{\min\left(T^g, \tau\right)|\mathbf{X}\right\}\right] = E\left[E\left\{\min\left(g(\mathbf{X})T^1 + \left\{1 - g(\mathbf{X})\right\}T^0, \tau\right)|\mathbf{X}\right\}\right] \\
&= E\left[g(\mathbf{X})E\left\{\min\left(T^1, \tau\right)|\mathbf{X}\right\} + \left\{1 - g(\mathbf{X})\right\}E\left\{\min\left(T^0, \tau\right)|\mathbf{X}\right\}\right] \\
&= E\left[g(\mathbf{X})\mu^1(\mathbf{X}) + \left\{1 - g(\mathbf{X})\right\}\mu^0(\mathbf{X})\right] \\
&= E\left[\mu^0(\mathbf{X}) + g(\mathbf{X})\left\{\mu^1(\mathbf{X}) - \mu^0(\mathbf{X})\right\}\right]
\end{aligned}$$

where $\mu^0(\mathbf{X}) = E\{\min(T^0, \tau)|\ \mathbf{X}\}$ and $\mu^1(\mathbf{X}) = E\{\min(T^1, \tau)|\ \mathbf{X}\}$ are the counterfactual restricted conditional mean survival time given $\mathbf{X}$. Thus, the optimal regime takes the form $g^{\text{opt}}(\mathbf{X}) = I\{\mu^1(\mathbf{X}) > \mu^0(\mathbf{X})\}$. We can estimate these restricted conditional mean survival times from the corresponding counterfactual survival function estimates, such that,

$$\hat{\mu}^1(\mathbf{X}) = \int_0^\tau \hat{S}^1(t)dt \quad \text{and} \quad \hat{\mu}^0(\mathbf{X}) = \int_0^\tau \hat{S}^0(t)dt\,.$$

Then the estimated optimal regime is $\hat{g}^{\text{opt}}(\mathbf{X}) = I\{\hat{\mu}^1(\mathbf{X}) > \hat{\mu}^0(\mathbf{X})\}$.

3 Simulations

To assess the performance of the proposed method, we conduct simulation studies in various scenarios and compare the results with other commonly used methods.

Three competing methods are considered here for the counterfactual models. First, as a widely used approach for survival analysis, we consider a standard Cox model, where we fit $(Y, \boldsymbol{\Delta}) \sim \mathbf{X} + A + \mathbf{X} \times A$. Then we estimate the counterfactuals as $\hat{S}^1(t, \mathbf{X}) = \hat{P}(T > t|A = 1, \mathbf{X})$ and $\hat{S}^0(t, \mathbf{X}) = \hat{P}(T > t|A = 0, \mathbf{X})$. As the second approach, we consider the weighted Cox models for the treatment and control counterfactuals, where we fit a Cox model $(Y, \boldsymbol{\Delta}) \sim \mathbf{X}$ with weight $\hat{w}t^1$ to calculate the treatment specific $\hat{S}^1(t, \mathbf{X})$ and with weight $\hat{w}t^0$ to calculate the control specific $\hat{S}^0(t, \mathbf{X})$. The third method we consider is a regular RSF model, where we fit $(Y, \Delta) \sim \mathbf{X} + A$ using the unweighted version of RSF model (in this case, *cforest* (), the same model as the base learner used in RSFWB), and then calculate $\hat{S}^1(t, \mathbf{X})$ and $\hat{S}^0(t, \mathbf{X})$ by arbitrarily setting A to either 1 or 0 similarly as in the first method. For all methods, the counterfactual restricted conditional mean survival time and the optimal regime are then calculated following the same procedure as described for RSFWB.

3.1 Simulation Schemes

For simulation studies, we generate $\mathbf{X} = \{X_1, X_2, \ldots, X_{20}\}$ with dimension $d_1 = 20$ from independent $N(0,1)$, and a scalar auxiliary variable Z ($d_2 = 1$) which is correlated with $\mathbf{X}$, specifically with X_2 as $Z \mid X_2 = x_2 \sim N(x_2, 1)$. The observed treatment indicator A is then generated from a logistic model

$$P(A = 1 \mid \mathbf{X}, Z) = \theta_0 + \theta_1 X_1 + \ldots + \theta_{20} X_{20} + \theta_z Z, \quad (2)$$

and the two counterfactual survival outcomes are generated from log-normal models of the form

$$\log T^0 = \beta_0 + \beta_1 X_1 + \ldots + \beta_{20} X_{20} + \beta_z Z + \varepsilon^0, \quad (3)$$

$$\log T^1 = \beta_0 + \beta_1 X_1 + \ldots + \beta_{20} X_{20} + \beta_z Z + h(\mathbf{X}, Z) + \varepsilon^1, \quad (4)$$

where ε^0 and ε^1 are generated from $N(0, \sigma^2)$. The observed survival time can then be calculated as $T = AT^1 + (1\text{- A})T^0$. The censoring time is then generated from uniform distribution $C \sim \text{Uniform}(0, \tau)$. We observe time $Y = \min(T, C)$ and the event indicator $\Delta = I(T \leq C)$. The true optimal regime $g^{\text{opt}}(\mathbf{X})$ would mainly depend on the form of $h(\mathbf{X}, Z)$. As mentioned in the previous section, it may also be affected by the choice of the time boundary τ. More on the calculation details of $g^{\text{opt}}(\mathbf{X})$ based on models (3) and (4) can be found in Appendix A1 and A2.

3.2 Different Scenarios

We consider multiple simulation scenarios with different $h(\mathbf{X}, Z)$ functions to mimic different type of optimal treatment rules. In Scenario 1, we set $\theta_0 = 0.1$, $\theta_1 = 0.5$, $\theta_2 = -1$, $\theta_z = -2$, and all other θ_j set to 0 (for $j = 3,4,\ldots,20$). We generate the counterfactual survival outcome from models (3) and (4), with $\sigma^2 = 1$, $\beta_0 = 0.5$, $\beta_1 = 1$, $\beta_2 = -1$, $\beta_z = 2$, all other $\beta_j = 0$ (for $j = 3,4,\ldots,20$). Simple linear interaction is considered here as $h(\mathbf{X}, Z) = 2(X_1 - X_2)$. If we ignore the influence of τ and directly exam the counterfactual survival time, the optimal regime can be approximated as to assign treatment $A = 1$ in the region of patient characteristic space defined by the linear combination $X_1 - X_2 > 0$. This regime fell in the regime space considered by the Cox models used in the competing methods, thus this scenario favors the model specification in the Cox model based methods.

For Scenario 2, we set coefficients in treatment model (2) as $\theta_0 = 0.1$, $\theta_1 = 0.5$, $\theta_2 = -0.2$, $\theta_z = -1.2$, and all other θ_j set to 0 (for $j = 3,4,\ldots,20$). For the counterfactual outcome model, we set $\sigma^2 = 1$, $\beta_0 = 0.8$, $\beta_1 = 0.5$, $\beta_2 = -0.5$, $\beta_z = 1.2$, all other $\beta_j = 0$ ($j = 3,4,..,20$), and $h(\mathbf{X}, Z) = 1.5X_1^2 - 0.6X_1 - 0.5X_2 + 0.3Z - 0.74$ in models (3) and (4). When the influence of τ

is small, the optimal regime can be approximated as to assign the treatment to patients with $h(\mathbf{X}, Z) = 1.5(X_1 - 0.2)^2 - 0.2(X_2 + 4) > 0$.

In Scenario 3, we consider treatment assignment model with θ_0=0.5, $\theta_1 = -0.5$, $\theta_2 = -0.5$, $\theta_z = 2$, and all other θ_j set to 0 (for $j = 3, 4, \ldots, 20$). For models (3) and (4), we set σ^2 =4, β_0 =0.5, $\beta_1 = -0.2$, β_2 =0.3, β_z =0.5, all other $\beta_j = 0$ ($j = 3,4,\ldots, 20$), and $h(\mathbf{X}, Z) = 2.5 \cdot I(X_1 > -0.5) \cdot I(X_2 < 0.5) - 1$. Here the optimal regime can be approximated as to treat patients with $X_1 > -0.5$ and $X_2 < 0.5$, which is a tree-type decision rule.

For each scenario, we consider two cases with different censoring time distributions, where the values of τ are chosen to create one case with about 20% censoring and another case with about 45% censoring.

3.3 Simulation Results

For each simulation setting, we generate data with $n = 500$ patients for 200 replicates, and apply the proposed method and the other three competing methods to each replicate. We compare different methods through both the fitting of counterfactual models and the performance of the estimated regime. To evaluate the model fit, we calculate the Root Mean Squared Difference (RMSD) as follows:

$$\text{RMSD}(0) = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left\{\hat{\mu}^0(\mathbf{X}_i) - \mu^0(\mathbf{X}_i)\right\}^2} \quad \text{and} \quad \text{RMSD}(1) = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left\{\hat{\mu}^1(\mathbf{X}_i) - \mu^1(\mathbf{X}_i)\right\}^2},$$

It indicates how close the model estimates are to the true counterfactual restricted conditional mean survival times. Table 1 shows the average RMSDs from all four methods, where we can see that, in Scenario 1, Cox model based methods tend to give slightly smaller RMSDs than RSF based methods, while in Scenarios 2 and 3, RSF based methods achieve smaller RMSD. One reason for this phenomenon is that in Scenario 1, the data generating procedure creates an optimal treatment regime with a linear boundary $X_1 = X_2$, and the hazards are approximately proportional for each covariate, thus the Cox model structure would perform nicely in revealing $\mathbf{X}$ - A interactions which are linear in $\mathbf{X}$. The differences in RMSDs for all methods in Scenario 1 are not large, which suggests that even in the scenario that favors the Cox models, the proposed method can still provide comparable model fit. In Scenarios 2 and 3, the Cox models are severely misspecified, thus yield much larger RMSDs than regular RSF and RSFWB. In addition, RSFWB tends to yield better fits of the counterfactual outcomes than the regular RSF. Compared to regular RSF, the inverse probability of treatment weighting can effectively reduce selection bias and the additional bootstrap procedure actually increases the randomness and thus the overall model fit for RSFWB is expected to be improved. One may notice that, compared to the cases with 20% censoring, we have much smaller RMSDs when the censoring rates are 45% in Table 1. This is related to the fact that when similar data generating procedures are used in one scenario, much smaller τ would be needed in order to yield 45% censoring case comparing to the one used for 20% censoring. This would in turn lead to smaller restricted conditional mean survival time μ.

The optimal treatment regime will benefit the whole population. Therefore in order to evaluate the performance of the estimated optimal regime, we also calculate the regime specific restricted mean survival time under the estimated optimal treatment regime $\hat{g}^{\text{opt}}$, which can be calculated based on the true counterfactuals $\mu^0(X_i)$'s and $\mu^1(X_i)$'s:

$$\mu^{\hat{g}} = \frac{1}{n}\sum_{i=1}^{n}\left[\hat{g}(X_i)\mu^0(X_i) + \{1-\hat{g}(X_i)\}\mu^1(X_i)\right]$$

This can be interpreted as the population average outcome when everybody in the cohort follows regime $\hat{g}$. A better regime would yield larger $\mu^{\hat{g}}$. The results from regimes estimated from different methods are summarized in Table 2, where we can see that in all scenarios, the true optimal regime g^{opt} always yields the largest μ^g. By comparing to the last two columns where all patients in the cohort are either assigned to the treatment or the control arm, the results suggest that applying the personalized treatment regime does benefit the study cohort compared to using a unified treatment assignment strategy. Since this is the underlying true optimal regime, it can serve as the upper bound that one can achieve by applying any treatment regime. In Scenario 1, RSFWB yields very close yet slightly smaller $\mu^{\hat{g}}$ than the Cox model based methods. In Scenarios 2 and 3, RSFWB yields larger $\mu^{\hat{g}}$ than the Cox model based methods and the regular RSF, suggesting that the proposed method can lead to better regime estimation for such nonlinear cases.

Furthermore, since in all scenarios, we make the true regime based on only X_1 and X_2, we can plot the estimated optimal treatment regime in the X_1 - X_2 plane. When the true optimal regime has a simple shape, for example an approximately linear boundary as in Scenario 1, all four methods can yield a partition very close to the true optimal regime (results not shown). When the true optimal regime has a more complicated and nonlinear shape, Figure 1 shows the plots of the estimated treatment regimes for Scenario 2 with 20% censoring, and Figure 2 is the result for Scenario 3 with 45% censoring. We can see that in both cases, the Cox model based methods tend to provide a partition of treatment decisions with a close to linear boundary on the X_1 - X_2 plane, while for both regular RSF and RSFWB, the partition does not seem to have any restrictions on its form, and thus can yield estimation much closer to the shape of the true optimal. Similar results are seen in the other cases in Scenario 2 and 3 (results not shown). These plots suggest that the estimated treatment regime from the proposed method is close to the true optimal for most of the patients. Similar results are also seen in other simulation settings along with higher censoring rate cases (see Appendix A3 for additional simulation studies). In summary, across all the scenarios considered here, the proposed method performs robustly in modeling the counterfactual outcomes and estimating the optimal treatment regime.

4 Analysis of the Prostate Cancer Data

In this section, we apply the proposed RSFWB to a prostate cancer dataset, which included a total of 4544 patients with clinically localized prostate cancer. Each patient received either surgery or radiation therapy at time zero at the University of Michigan. Both treatments are widely used to treat prostate cancer and neither has been established to be superior to the

other. The two treatments do have different potential adverse side effects and there may also be other personal or logistical reasons that lead a patient to prefer one treatment over the other. There is a tendency for older and less healthy patients who have more advanced disease to choose radiation therapy and for others to receive surgery, but whether this approach is the best for reducing the chance of the prostate cancer recurring after treatment is unknown. In the available data, pretreatment prognostic factors are measured at baseline, including age, prostate cancer stage (with values T1-T3), PSA (ng/ml), the Charlson comorbidity score, Gleason grade, presence of perineural invasion at biopsy (PNI) and date of the treatment (1996–2013). After treatment, all the patients were followed for clinical failure, which is defined as the occurrence of distant metastasis. We focus on the subset of 3540 patients with complete records of all the variables listed above, of which 702 patients (19.8%) received radiation therapy. The median follow-up time for the whole cohort was 5.6 years, with 5.13 years for surgery group and 7.56 years for radiation group. 93 patients (2.6%) were observed to experience a clinical failure during the study, which includes 47 patients from surgery group (1.7%) and 46 patients from radiation group (6.6%). Table 3 lists some summary statistics of the dataset. Note that the PSA values were already log transformed, i.e. log(PSA +1). As can be seen, patients who were older and with worse tumor characteristics (higher PSA, Gleason, Stage, PNI) tended to be assigned to receive radiation therapy.

It is of great interest to learn from the observational data about how different treatments are expected to perform in the future and whether some common guidelines could be suggested regarding whether to recommend one therapy versus the other for the purpose of prolonging the survival time until clinical failure. In the current observational study, one of the variables available is the treatment initiation time (the date of the treatment). The distributions of the treatment initiation time of the two treatment groups are quite different (data not shown), and the treatment time is also correlated with the time to clinical failure, with p-value < 0.001 from the likelihood ratio test in a marginal univariate Cox model. Thus, treatment initiation time is a confounder for both treatment assignment and outcome. However, for obvious reasons, it is not useful for assigning a treatment for future patients. Thus, we apply RSFWB while treating treatment initiation as an auxiliary variable. We choose $\tau = 12$ years as only 6.2% patients had follow-up time longer than 12 years. For more stable estimation we also truncate the weights at 15, which changes the extreme weights for 2.03% patients.

–Figure 3 shows the estimated optimal regime using RSFWB. Figure 3a is the estimated optimal treatment assignment on the PSA-Age plane, which is not partitioned linearly. Thus it is not likely to be captured by the Cox model with assumptions of simple linear interactions. Besides, it is clear from the figure that other variables in addition to age and PSA are important in defining the optimal regime. Figure 3b shows the boxplot of age in different treatment groups according to the estimated optimal regime. We can see that the marginal distributions of age are not identical in the two groups, and older patients are more likely to be assigned to radiation therapy. Figure 3 c shows the distribution of recommended treatment by age and PNI, which suggests that, for PNI negative group, surgery is recommended for most patients under 45, and radiation is recommended for most patients over 75. While for PNI positive group, radiation therapy is preferred for a large portion of the patient population. Figures 3d and 3e illustrate the magnitude of the difference in e

stimated restricted conditional mean survival time between radiation and surgery therapy ($\hat{\mu}^1(\boldsymbol{X}_i) - \hat{\mu}^0(\boldsymbol{X}_i)$). In Figures 3d, we colored the points differently when the time differences are larger than 0.5 years, and a histogram of these time differences are shown in Figures 3e. We can see that, for many patients, the (estimated) benefit of choosing one therapy over the other was rather small. Considering the uncertainty in the estimation procedure, there may not be a clear gain for these patients to switch treatment in practice. Thus we would only recommend a treatment switch if the estimated gain for the patient's restricted conditional mean survival time is more than 0.1 years. The recommended therapy thus estimated would suggest 182 patients (25.9%) who were observed to receive radiation therapy to switch to surgery, while 259 patients (9.1%) in the surgery group should switch to the radiation therapy. Such recommendation yields a maximal restricted mean failure free survival time of $\hat{\mu}^{\hat{g}} = 11.779$ years, while for the as-treated regime it was $\hat{\mu} = 11.759$ years. Comparing to the case where all patients were assigned to surgery ($\hat{\mu}^0 = 11.752$ years), and where all patients were assigned to radiation therapy ($\hat{\mu}^1 = 11.731$ years), there is a gain by following the recommended regime. We further use the bootstrap with 500 iterations to calculate the 95% confidence interval for $\hat{\mu}^{\hat{g}} - \hat{\mu}^0$ which is (0.032,0.207), and 95% confidence interval for $\hat{\mu}^{\hat{g}} - \hat{\mu}^1$ which is (0.017,0.114). Both confidence intervals do not include 0, thus, there is significant gain in customizing the treatment assignment according to the recommended regime comparing to treating everyone with the same therapy.

5 Discussion

The idea of personalized treatment regimes is very attractive as it maximizes the treatment effect for the entire cohort. In this paper, we focus on observational data when the outcome is subject to right censoring, and show that the proposed RSFWB method can effectively marginalize over the auxiliary variables and correctly identify the optimal regime based on a subset of clinically important covariates. The machine learning based modeling technique employed in RSFWB provides flexibility in capturing the complex variable dependency and puts less constraints on the regimes under consideration. This would be a desirable feature for clinical studies where the disease and treatment mechanisms are not well understood. Furthermore, RSFWB is essentially a regression tree based algorithm, which would be suitable for detecting interactions among variables. Thus it would be a good choice in estimating heterogeneity in treatment effect and thus identifying the optimal regime (Loh 2002). An attractive feature of Random Forest models is their ability to deal with high dimensional data (Genuer et al. 2010). As an ensemble of RSFs, RSFWB can also provide variable importance measures using similar procedures as for individual RSF, thus variable selection could also be performed in a similar fashion for RSFWB. This could facilitate variable selection for interactions and lead to a more parsimonious model and simple treatment rules. To this end, Wager and Athey (2017) provides more discussion on applying Random Forest type methods in treatment regime construction for high dimensional settings.

In addition, literature has suggested that heterogeneity and randomness induced by bootstrap resampling in Random Forest type ensemble methods, can substantially increase the diversity of the trees and thus improve the performance of the predictions (Dietterich 2000;

Biau et al. 2008; Zhu and Kosorok 2012). RSFWB's resampling based implementation of the weighting procedure allows a varying number of co pies of each subject to be included in one leaf of a tree, which increases the randomness of the tree building process. Although the theoretical properties of Random Forest based methods have not been fully explored, the consistency of tree based regression models has been studied since Gordon and Olshen (1984). In addition, theories from bagging methods have suggested the connection between the asymptotics of the ensemble estimator and its base learners (Biau et al. 2008; Biau and Devroye 2010; Wager et al. 2014; Scornet et al. 2015). Intuitively, the weighted bootstrapping scenario used here is equivalent to sampling from the corresponding counterfactual data, which would be expected to yield unbiased counterfactual models.

In this paper we have focused on using a non-parametric approach for the outcome model and used a simple logistic model to estimate the weights. In the cases when the mechanism of treatment assignment is not fully understood, we may also consider semi-parametric or nonparametric models for weight estimation, an example of which is provided in Shen et al. (2016).

In RSFWB, we propose to build the counterfactual models for treatment and control separately. Thus it is straightforward to extend RSFWB to problems when more than two lines of treatments are available for each patient. The optimal regime can then be identified by comparing all possible counterfactual outcomes.

We propose to incorporate the inverse probability weights through a weighted bootstrap resampling scheme. Although we chose to use the conditional inference forest in R as the base learner, the same idea can be easily implemented with other RSF algorithms in commonly used statistical software. This weighted bootstrap scheme provides a natural solution, because most existing RSF codes either do not have an option to include weights or do not weight samples in a way that was appropriate for our purpose here. In addition, this proposal actually provides a rather general framework for people to implement weights with methods other than RSF as base learner. However, for a particular algorithm or software, it may be possible to incorporate the weights in a more computationally efficient way.

6 Acknowledgements

This work was partially supported by National Institutes of Health grants CA129102 and CA199338.

Appendix

A1. True Conditional Quantities for Simulation Models

In the simulation studies, we generate survival time based on lognormal model. To identify the true optimal regime, ideally, we would like to find the regime which gives the largest conditional mean survival time $E\{T|\mathbf{X}\}$ for each subject. However, since the observed survival time is subject to censoring, we make comparison of different regimes through the restricted conditional mean survival time $\mu = E\{\min(T,\tau)|\mathbf{X}\}$ for some $\tau>0$. From now on, we will denote it as $\mu(\tau,\mathbf{X})$ to emphasis its dependency on both $\mathbf{X}$ and τ. As a first step to

understand the true optimal regime here, we investigate the relationship between various conditional expectations and the covariates under the lognormal models.

Without loss of generality, we start with simple model where both X and Z are one dimensional. Consider the lognormal model

$$\log T = \beta_0 + \beta_1 X + \beta_2 Z + \varepsilon$$

where $Z = \eta X + \varepsilon_z$, $X \sim N\left(0, \sigma_1^2\right)$, $\varepsilon_z \sim N\left(0, \sigma_2^2\right)$, and $\varepsilon \sim N\left(0, \sigma_0^2\right)$ are independently distributed. Now the quantity of interest is $\mu(\boldsymbol{\tau}, \mathbf{X})$. As the first step, the conditional mean for $\log T$ is

$$E\left\{\log T \mid X = x\right\} = \beta_0 + \left(\beta_1 + \eta\beta_2\right)x$$

and then the conditional mean survival time is

$$E\left\{T \mid X = x\right\} = \iint e^{\beta_0 + \beta_1 x + \beta_2 z + \varepsilon} f_{Z \mid X}(z \mid x) f_{\varepsilon}(\varepsilon) dz d\varepsilon$$
$$= \iint e^{\beta_0 + \left(\beta_1 + \eta\beta_2\right)x + \beta_2 \varepsilon_z + \varepsilon} f_{\varepsilon_z}\left(\varepsilon_z\right) f_{\varepsilon}(\varepsilon) d\varepsilon_z d\varepsilon = e^{\beta_0 + \left(\beta_1 + \eta\beta_2\right)x} \int e^{\beta_2 \varepsilon_z} f_{\varepsilon_z}\left(\varepsilon_z\right) d\varepsilon_z \int e^{\varepsilon} f_{\varepsilon}(\varepsilon) d\varepsilon$$

Where $f_{\mathbf{Z}|\mathbf{X}}(\cdot \mid \cdot)$, $f_{\varepsilon}(\cdot)$ and $f_{\varepsilon_z}(\cdot)$ are the density function of the corresponding random variables, and

$$\int e^{\varepsilon} f_{\varepsilon}(\varepsilon) d\varepsilon = \int e^{\varepsilon} \frac{1}{\sigma_0 \sqrt{2\pi}} e^{-\frac{\varepsilon^2}{\sigma_0^2}} d\varepsilon = e^{\frac{\sigma_0^2}{2}}$$

$$\int e^{\beta_2 \varepsilon_z} f_{\varepsilon_z}\left(\varepsilon_z\right) d\varepsilon_z = \int e^{\beta_2 \varepsilon_z} \frac{1}{\sigma_2 \sqrt{2\pi}} e^{-\frac{\varepsilon_z^2}{\sigma_2^2}} d\varepsilon_z = e^{\frac{\beta_2^2 \sigma_2^2}{2}}$$

thus

$$E\{T \mid X = x\} = e^{\beta_0 + \left(\beta_1 + \eta\beta_2\right)x + \frac{1}{2}\left(\beta_2^2 \sigma_2^2 + \sigma_0^2\right)}$$

The restricted mean survival (conditional on $\mathbf{W} = (X, Z)^{\mathrm{T}}$) is:

$$E\left\{\min(T,\tau)|X,Z\right\} = \int_0^\infty \min(t,\tau) f_{T|X,Z}(t|x,z)dt$$

$$= \int_0^\tau t f_{T|X,Z}(t|x,z)dt + \int_\tau^\infty \tau f_{T|X,z}(t|x,z)dt = \int_0^\tau t f_{T|X,Z}(t|x,z)dt + \tau P(T>\tau|X,Z)$$

$$= \left(e^{\beta_0+\beta_1 x+\beta_2 z+\frac{1}{2}\sigma_0^2}\right)\Phi\left(\frac{\log\tau-\left(\beta_0+\beta_1 x+\beta_2 z\right)-\sigma_0^2}{\sigma_0}\right)+\tau\Phi\left(\frac{\left(\beta_0+\beta_1 x+\beta_2 z\right)-\log\tau}{\sigma_0}\right)$$

with $\Phi(\cdot)$ be the cumulative distribution function (cdf) for the standard normal distribution. Next, we can calculate the restricted conditional mean survival (conditional on X):

$$E\left\{\min(T,\tau)|X=x\right\} = \int_0^\tau t f_{T|X}(t|x)dt + \int_\tau^\infty \tau f_{T|X}(t|x)dt$$

$$= \int_{-\infty}^{+\infty}\int_{-\infty}^{u} e^{\beta_0+\left(\beta_1+\eta\beta_2\right)x+\beta_2\varepsilon_z+\varepsilon} f_\varepsilon(\varepsilon)d\varepsilon f_{\varepsilon_z}\left(\varepsilon_z\right)d\varepsilon_z + \tau\int_{-\infty}^{+\infty}\int_{u}^{\infty} f_\varepsilon(\varepsilon)d\varepsilon f_{\varepsilon_z}\left(\varepsilon_z\right)d\varepsilon_z$$

$$= e^{\beta_0+\left(\beta_1+\eta\beta_2\right)x}\int_{-\infty}^{+\infty} e^{\beta_2\varepsilon_z}\int_{-\infty}^{u} e^{\varepsilon} f_\varepsilon(\varepsilon)d\varepsilon f_{\varepsilon_z}\left(\varepsilon_z\right)d\varepsilon_z + \tau\int_{-\infty}^{+\infty}\int_{u}^{\infty} f_\varepsilon(\varepsilon)d\varepsilon f_{\varepsilon_z}\left(\varepsilon_z\right)d\varepsilon_z$$

$$= e^{\beta_0+\left(\beta_1+\eta\beta_2\right)x+\frac{\sigma_0^2}{2}}\int_{-\infty}^{+\infty} e^{\beta_2\varepsilon_z}\Phi\left(\frac{\log\tau-\beta_0-\left(\beta_1+\eta\beta_2\right)x-\beta_2\varepsilon_z-\sigma_0^2}{\sigma_0}\right) f_{\varepsilon_z}\left(\varepsilon_z\right)d\varepsilon_z + \tau\int_{-\infty}^{+\infty}\Phi$$

$$\left(-\frac{\log\tau-\beta_0-\left(\beta_1+\eta\beta_2\right)x-\beta_2\varepsilon_z}{\sigma_0}\right) f_{\varepsilon_z}\left(\varepsilon_z\right)d\varepsilon_z$$

$$= e^{\beta_0+\left(\beta_1+\eta\beta_2\right)x+\frac{\sigma_0^2}{2}}\int_{-\infty}^{+\infty} e^{\beta_2\sigma_2 v}\Phi\left(\frac{\log\tau-\beta_0-\left(\beta_1+\eta\beta_2\right)x-\beta_2\sigma_2 v-\sigma_0^2}{\sigma_0}\right)\phi(v)dv$$

$$+\tau\int_{-\infty}^{+\infty}\Phi\left(-\frac{\log\tau-\beta_0-\left(\beta_1+\eta\beta_2\right)x-\beta_2\sigma_2 v}{\sigma_0}\right)\phi(v)dv$$

where $f_{T|X}(\cdot\,|\,\cdot)$ is the density function for the conditional distribution of $T|\,X$, $\phi(\cdot)$ is the density function (pdf) for the standard normal distribution. The integral here does not generally have a closed form. In practice, we propose to use Gaussian quadrature to calculate the above quantity numerically using the function *gauss.quad.prob*() in the {*statmod*} package in R. Following this procedure, we can calculate the true restricted

conditional mean survival time $\mu^0(\tau, \mathbf{X})$ and $\mu^1(\tau,\mathbf{X})$ from the counterfactual models, and thus numerically identify the true optimal regime as defined.

A2. The Relationship Between the Optimal Regime and τ

As shown in the previous section, $\mu^0(\tau, \mathbf{X})$ and $\mu^1(\tau,\mathbf{X})$ depend on both the values of τ and $\mathbf{X}$, so the optimal treatment regime defined by the restricted conditional mean survival time would also depend on the choice of τ. Here we study the influence of τ on the optimal regime under lognormal models. Consider the counterfactual models:

$$\log T^0 = \beta_0 + \beta_1 X + \beta_2 Z + \varepsilon^0 \text{ and } \log T^1 = \beta_0 + \beta_1 X + h(X, Z) + \beta_2 Z + \varepsilon^1$$

where, again, for simplicity, we consider scalar X and Z. Similar results can be obtained when both $\mathbf{X}$ and $\mathbf{Z}$ are multi-dimensional. Then the regime which yields the largest mean survival time would be to give the treatment when $E[h(X, Z)|X] > 0$. However, when we define the optimal regime with the restricted mean survival time. The difference between the restricted conditional mean survivals is

$$\mu^1(\tau; x) - \mu^0(\tau; x)$$

$$= e^{\beta_0 + (\beta_1 + \eta\beta_2)x + \frac{\sigma_0^2}{2}} \int_{-\infty}^{+\infty} e^{h(x, \eta x + \varepsilon_z) + \beta_2 \varepsilon_z} \Phi\left(\frac{\log\tau - \beta_0 - (\beta_1 + \eta\beta_2)x - h(x, \eta x + \varepsilon_z) - \beta_2\varepsilon_z - \sigma_0^2}{\sigma_0}\right) f_{\varepsilon_z}(\varepsilon_z) d\varepsilon_z$$

$$+ \tau \int_{-\infty}^{+\infty} \Phi\left(-\frac{\log\tau - \beta_0 - (\beta_1 + \eta\beta_2)x - h(x, \eta x + \varepsilon_z) - \beta_2\varepsilon_z}{\sigma_0}\right) f_{\varepsilon_z}(\varepsilon_z) d\varepsilon_z$$

$$- e^{\beta_0 + (\beta_1 + \eta\beta_2)x + \frac{\sigma_0^2}{2}} \int_{-\infty}^{+\infty} e^{\beta_2\varepsilon_z} \Phi\left(\frac{\log\tau - \beta_0 - (\beta_1 + \eta\beta_2)x - \beta_2\varepsilon_z - \sigma_0^2}{\sigma_0}\right) f_{\varepsilon_z}(\varepsilon_z) d\varepsilon_z$$

$$- \tau \int_{-\infty}^{+\infty} \Phi\left(-\frac{\log\tau - \beta_0 - (\beta_1 + \eta\beta_2)x - \beta_2\varepsilon_z}{\sigma_0}\right) f_{\varepsilon_z}(\varepsilon_z) d\varepsilon_z$$

$$= A + B$$

where we let

$$A = e^{\beta_0 + (\beta_1 + \eta\beta_2)x + \frac{\sigma_0^2}{2}} \int_{-\infty}^{+\infty} \left\{ e^{h(x,\eta x+\varepsilon_z)} \Phi\left(\frac{\log\tau - \beta_0 - (\beta_1 + \eta\beta_2)x - h(x,\eta x+\varepsilon_z) - \beta_2\varepsilon_z - \sigma_0^2}{\sigma_0} \right) \right.$$
$$\left. - \Phi\left(\frac{\log\tau - \beta_0 - (\beta_1 + \eta\beta_2)x - \beta_2\varepsilon_z - \sigma_0^2}{\sigma_0} \right) \right\} \cdot e^{\beta_2\varepsilon_z} f_{\varepsilon_z}(\varepsilon_z) d\varepsilon_z$$

and

$$B = \tau \int_{-\infty}^{+\infty} \left\{ \Phi\left(-\frac{\log\tau - \beta_0 - (\beta_1 + \eta\beta_2)x - h(x,\eta x+\varepsilon_z) - \beta_2\varepsilon_z}{\sigma_0} \right) \right.$$
$$\left. - \Phi\left(-\frac{\log\tau - \beta_0 - (\beta_1 + \eta\beta_2)x - \beta_2\varepsilon_2}{\sigma_0} \right) \right\} \phi(\varepsilon_z) d\varepsilon_z$$

In general, the true optimal regime $g(x) = I(\mu^1(\tau, x) - \mu^0(\tau; x) > 0)$ would have a complicated shape, which is not equal to $h(x, \eta x) > 0$ or $E[h(X, Z) \mid X = x] > 0$. If $\tau \to \infty$ and $h(X)$ only depends on X then $\mu^1(\tau, X) = \mu^0(\tau; X)$ if and only if $h(X) = 0$. In practice, people usually choose time to study end or the largest follow-up time, i.e. τ is likely to be large, in such cases, if the interaction term is mainly about X, i.e. $h(X)$, then the optimal regime defined by restricted conditional mean survival can be well approximated by the shape of $h(X)$.

A3. Additional Simulation with High Censoring Rate

In this section, we conduct further simulation studies to investigate the performance of the proposed method in cases with higher censoring rate. We generate data similarly as Scenario 3, but with slightly different parameter setting to make the lower tail of T not too small and to guarantee that there would be enough events in the data to better illustrate the different performance for various methods especially under higher censoring cases. In detail, we generate A, T^0 and T^1 as

$$P(A = 1 | \mathbf{X}, Z) = 0.1 - 0.5X_1 - 0.5X_2 + 2Z,$$

$$\log T^0 = 2.5 - 0.2X_1 + 0.3X_2 + 0.5Z + \varepsilon^0,$$

$$\log T^1 = 2.5 - 0.2X_1 + 0.3X_2 + 0.5Z + \left\{ 2 \cdot I(X_1 > -0.5) \cdot I(X_2 < 0.5) - 1 \right\} + \varepsilon^1.$$

with ε^0 and ε^1 are generated from $N(0,4)$. We generate the censoring time C from Uniform$(0, \tau)$, with τ chosen to yield cases with 20%, 45%, 70% and 85% censoring, respectively. Tables A1 and A2 show the results from different estimation methods. Similar

to the results shown in Scenario 3 of main text, the proposed method yields the smallest RMSD(0) and RMSD(1), and the largest regime specific restricted mean survival time among all four methods. This is true even for 70% and 85% censoring cases. As shown in Figure 3, for the case of 85% censoring, the shape of the estimated optimal regime yielded by RSFWB is the one closest to the true optimal. This suggests that the proposed method works robustly when the censoring rate is high.

A4. Effect of Tuning Parameters in RSFWB

The choice of tuning parameters always has a big impact on the performance of machine learning based methods. For RSFWB, we fix the number of base learners, i.e. the number of weighted bootstrap samples, as 200. Then the tuning parameters to select for better performance are the number of candidate covariates at each split (*mtry*), and the number of trees in one forest (*ntree*). The same set of parameters will be applied to each base learner (conditional inference survival forest). We propose to use grid search to select the set (*mtry,ntree*) that gives the smallest cumulative prediction error, i.e IBS. In practice, to limit the computation load for tuning parameter selection, we use the IBS calculated from a single forest model for the full dataset (both treatment and control arm data) with covariates (X_i, A_i) as a proximal criteria to measure the prediction. For a given set of (*mtry,ntree*), the {*pec*} package in R is employed to calculate such IBS from 5-fold cross-validation mogensen2012evaluating. In this process, we choose the marginal Kaplan-Meier estimator for the censoring weight and default values for other parameters. In simulation studies, we perform the grid search for *mtry* ∈ {2, 4,8,16} and *ntreee* {50,100}. Table A3 illustrates an example of the tuning parameter selection, where the approximated prediction error (IBS) calculated from the procedure described here, as well as the RMSD(0), RMSD(1) calculated from RSFWB with the same set of tuning parameters are presented. The data generation models for this example are the same as the ones used for additional simulations in Appendix A3. The mean and standard deviation of these three measures from 200 simulations are shown. We can see the set of tuning parameters that minimizes the approximated prediction error, also yields the smallest RMSD(0) and RMSD(1). Thus this procedure seems helpful in optimizing RMSD(0) and RMSD(1), and thus improve the accuracy of regime estimation.

Table A1:

Root Mean Squared Differences (RMSD) for Additional Simulation Studies. The mean and standard deviation of RMSDs for different censoring rates are recorded.

		Cox		**weighted Cox**		**RSF**		**RSFWB**	
censoring		mean	(SD)	mean	(SD)	mean	(SD)	Mean	(SD)
20%	RMSD(1)	26.37	(1.67)	27.55	(2.29)	16.94	(1.29)	16.12	(1.36)
	RMSD(0)	12.15	(2.09)	14.24	(2.80)	17.39	(1.99)	16.84	(2.57)
45%	RMSD(1)	9.57	(0.43)	10.10	(0.58)	6.60	(0.61)	6.52	(0.60)
	RMSD(0)	5.32	(0.93)	5.64	(0.99)	6.80	(0.87)	6.52	(1.03)
70%	RMSD(1)	2.53	(0.16)	2.71	(0.20)	2.08	(0.16)	1.95	(0.23)
	RMSD(0)	1.67	(0.27)	1.61	(0.28)	1.72	(0.22)	1.67	(0.25)

		Cox		weighted Cox		RSF		RSFWB	
85%	RMSD(1)	1.00	(0.08)	1.09	(0.11)	0.85	(0.06)	0.81	(0.10)
	RMSD(0)	0.73	(0.13)	0.69	(0.13)	0.63	(0.09)	0.61	(0.10)

Table A2:

Regime Specific Restricted Mean Survival Time for Additional Simulation.

	Cox	Weighted Cox	RSF	RSFWB	opt	all ctrl	all trt
20% censoring	37.77	36.89	37.99	38.02	43.96	26.50	33.05
45% censoring	21.16	21.07	22.11	22.39	24.18	17.14	19.23
70% censoring	9.43	9.46	9.52	9.83	10.32	8.63	8.95
85% censoring	5.30	5.31	5.38	5.41	5.63	5.09	5.14

Table A3:

Effect of the Tuning Parameters. RMSDs and approximated cumulative prediction errors (IBS) calculated from 5-fold cross-validation are obtained for different tuning parameter settings (*mtry* and *ntree*), and the mean and standard deviation of these measures from 200 simulations with the setting for additional simulation scenario described in Appendix A3 are listed.

		IBS		RMSD(1)		RMSD(0)	
mtry	ntree	mean	(SD)	mean	(SD)	mean	(SD)
2	50	0.1126	(0.0068)	19.53	(0.78)	19.75	(1.82)
	100	0.1123	(0.0067)	19.41	(0.72)	19.77	(1.78)
4	50	0.1079	(0.0064)	18.35	(0.99)	18.43	(2.15)
	100	0.1077	(0.0063)	18.25	(0.93)	18.42	(2.14)
8	50	0.0976	(0.0067)	16.16	(2.28)	15.89	(2.82)
	100	0.0973	(0.0067)	16.05	(2.24)	15.84	(2.82)
16	50	0.1024	(0.0064)	17.08	(1.42)	16.84	(2.60)
	100	0.1020	(0.0065)	16.95	(1.35)	16.83	(2.56)

References

Barbe P and Bertail P (2012). The Weighted Bootstrap, volume 98 Springer Science & Business Media.

Biau G and Devroye L (2010). On the layered nearest neighbour estimate, the bagged nearest neighbour estimate and the random forest method in regression and classification. Journal of Multivariate Analysis, 101(10):2499–2518.

Biau G, Devroye L, and Lugosi G (2008). Consistency of random forests and other averaging classifiers. The Journal of Machine Learning Research, 9:2015–2033.

Bou-Hamad I, Larocque D, Ben-Ameur H, et al. (2011). A review of survival trees. Statistics Surveys, 5:44–71.

Breiman L (2001). Random forests. Machine Learning, 45(1):5–32.

Cox DR (1972). Regression models and life tables (with discussion). Journal of the Royal Statistical Society, 34:187–220.

Dietterich TG (2000). Ensemble methods in machine learning. In International Workshop on Multiple Classifier Systems, pages 1–15. Springer.

Doove L, Dusseldorp E, Van Deun K, and Van Mechelen I (2014). A comparison of five recursive partitioning methods to find person subgroups involved in meaningful treatment-subgroup interactions. Advances in Data Analysis and Classification, 8(4):403–425.

Ellsworth RE, Decewicz DJ, Shriver CD, and Ellsworth DL (2010). Breast cancer in the personal genomics era. Current Genomics, 11(3):146–161. [PubMed: 21037853]

Foster JC, Taylor JMG, and Ruberg SJ (2011). Subgroup identification from randomized clinical trial data. Statistics in Medicine, 30(24):2867–2880. [PubMed: 21815180]

Genuer R, Poggi J-M, and Tuleau-Malot C (2010). Variable selection using random forests. Pattern Recognition Letters, 31(14):2225–2236.

Gordon L and Olshen RA (1984). Almost surely consistent nonparametric regression from recursive partitioning schemes. Journal of Multivariate Analysis, 15(2):147–163.

Hothorn T, Hornik K, and Zeileis A (2006). Unbiased recursive partitioning: A conditional inference framework. Journal of Computational and Graphical statistics, 15(3):651–674.

Ishigooka J, Murasaki M, Miura S, and the Olanzapine Late-Phase II Study Group (2000). Olanzapine optimal dose: Results of an open-label multicenter study in schizophrenic patients. Psychiatry and Clinical Neurosciences, 54(4):467–478. [PubMed: 10997865]

Ishwaran H, Kogalur UB, Blackstone EH, and Lauer MS (2008). Random survival forests. The Annals of Applied Statistics, 2(3): 841–860.

Laber EB and Zhao Y (2015). Tree-based methods for individualized treatment regimes. Biometrika, 102(3):501–514. [PubMed: 26893526]

LeBlanc M and Crowley J (1995). A review of tree-based prognostic models In Recent Advances in Clinical Trial Design and Analysis, pages 113–124. Springer.

Loh W-Y (2002). Regression tress with unbiased variable selection and interaction detection. Statistica Sinica, 12:361–386.

Lu M, Sadiq S, Feaster DJ, and Ishwaran H (2017). Estimating individual treatment effect in observational data using random forest methods. arXiv preprint arXiv:1701.05306.

Makarenkov V, Boc A, Xie J, Peres-Neto P, Lapointe F-J, and Legendre P (2010). Weighted bootstrapping: a correction method for assessing the robustness of phylogenetic trees. BMC Evolutionary Biology, 10(1):1. [PubMed: 20044934]

Mogensen UB, Ishwaran H, and Gerds TA (2012). Evaluating random forests for survival analysis using prediction error curves. Journal of Statistical Software, 50(11):1. [PubMed: 25317082]

Nahorniak M, Larsen DP, Volk C, and Jordan CE (2015). Using inverse probability bootstrap sampling to eliminate sample induced bias in model based analysis of unequal probability samples. PloS One, 10(6):e0131765. [PubMed: 26126211]

Norazan M, Habshah M, Imon A, and Chen S (2009). Weighted bootstrap with probability in regression. In WSEAS International Conference. Proceedings Mathematics and Computers in Science and Engineering. World Scientific and Engineering Academy and Society.

Piquette-Miller M and Grant D (2007). The art and science of personalized medicine. Clinical Pharmacology and Therapeutics, 81(3):311–315. [PubMed: 17339856]

Qian M and Murphy SA (2011). Performance guarantees for individualized treatment rules. Annals of Statistics, 39(2):1180. [PubMed: 21666835]

Rothwell PM (2005). Subgroup analysis in randomised controlled trials: importance, indications, and interpretation. The Lancet, 365(9454):176–186.

Scornet E, Biau G, and Vert J-P (2015). Consistency of random forests. Annals of Statistics, 43(4): 1716–1741.

Shen J, Wang L, and Taylor JMG (2016). Estimation of the optimal regime in treatment of prostate cancer recurrence from observational data using flexible weighting models. Biometrics, page in press.

Strobl C, Malley J, and Tutz G (2009). An introduction to recursive partitioning: rationale, application, and characteristics of classification and regression trees, bagging, and random forests. Psychological Methods, 14(4):323. [PubMed: 19968396]

Wager S and Athey S (2017). Estimation and inference of heterogeneous treatment effects using random forests. Journal of the American Statistical Association, (just-accepted).

Wager S, Hastie T, and Efron B (2014). Confidence intervals for random forests: the jackknife and the infinitesimal jackknife. Journal of Machine Learning Research, 15(1):1625–1651. [PubMed: 25580094]

Xu R, Nettleton D, and Nordman DJ (2016). Case-specific random forests. Journal of Computational and Graphical Statistics, 25(1):49–65.

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

Zhao Y, Zeng D, Laber EB, Song R, Yuan M, and Kosorok MR (2015). Doubly robust learning for estimating individualized treatment with censored data. Biometrika, 102(1):151. [PubMed: 25937641]

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

Zhu R and Kosorok MR (2012). Recursively imputed survival trees. Journal of the American Statistical Association, 107(497):331–340. [PubMed: 23125470]

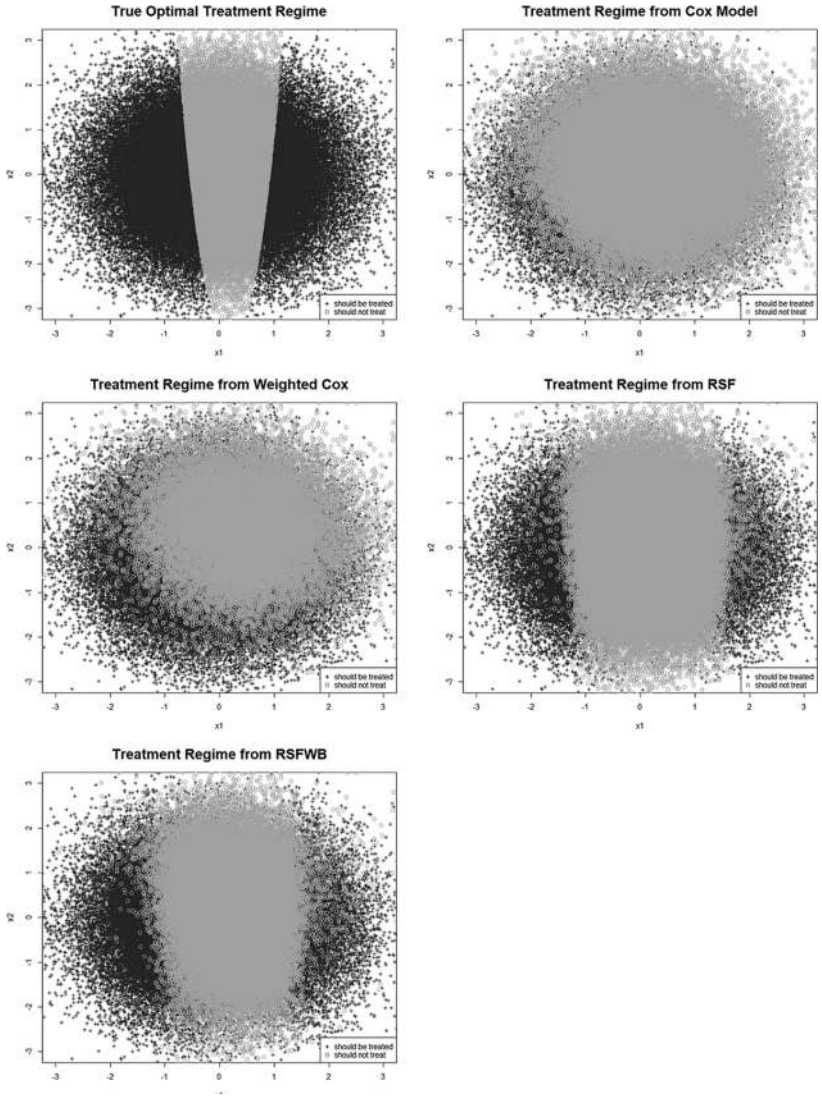


Figure 1:

Cumulative Plot for Estimated Optimal Treatment Regime for Scenario 2 with 20% Censoring. In each plot, the treatment regime is presented in term of the treatment assignment for all patients over the 200 replicates, black plus symbols are the ones assigned to treatment arm, and light grey circles are the ones assigned to control arm. The five regimes are the true optimal regime and the ones estimated by the four methods (Cox, weighted Cox, RSF and RSFWB).

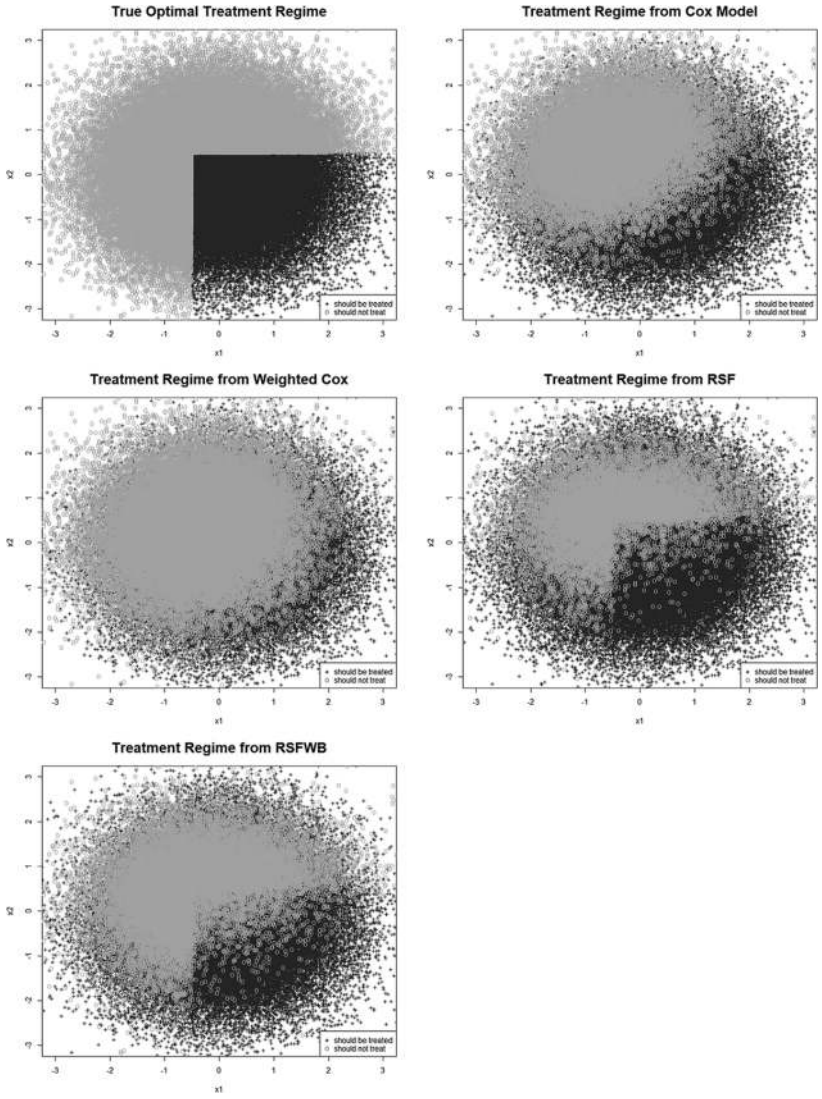
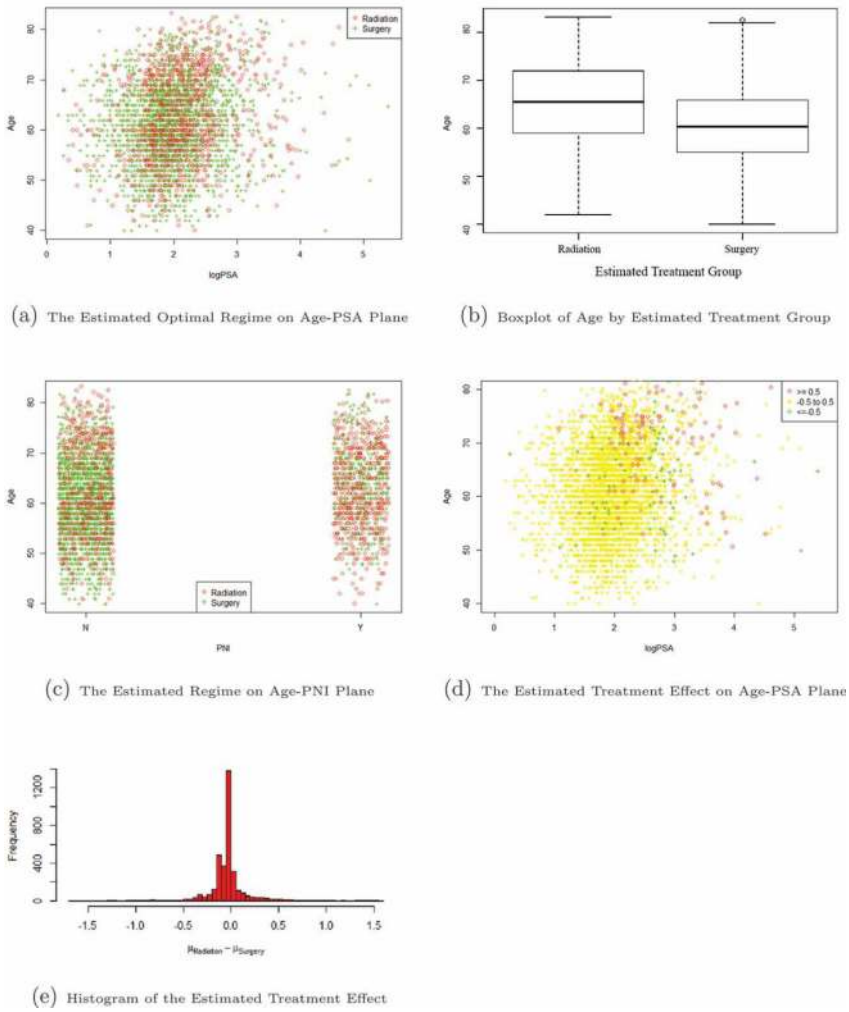


Figure 2:

Cumulative Plot for Estimated Optimal Treatment Regime for Scenario 3 with 45% Censoring. Similar as in Figure 1, the true optimal regime and the ones estimated by the four methods (Cox, weighted Cox, RSF and RSFWB) are plotted separately.



(a) The Estimated Optimal Regime on Age-PSA Plane (b) Boxplot of Age by Estimated Treatment Group

(c) The Estimated Regime on Age-PNI Plane (d) The Estimated Treatment Effect on Age-PSA Plane

(e) Histogram of the Estimated Treatment Effect

Figure 3:

Illustration of the Estimated Optimal Treatment Regime from the Prostate Cancer Study. Panel (a) shows the optimal regime assignment estimated by RSFWB on the Age- PSA plane. The circle symbols stand for assignments of radiation therapy, and the plus symbols stand for assignments of surgery. Panel (b) shows the boxplot of age for patients for each treatment arm in the estimated optimal treatment regime. Panel (c) is the estimated optimal treatment assignment on the Age-PNI plane. Panel (d) shows the estimated treatment effect on the Age-PSA plane, where the treatment effect is defined by the difference in restricted conditional mean time to clinical failure between radiation therapy and surgery. The color and shape of each point indicates its value of the estimated treatment effect as shown in the legend. Panel (e) is the histogram of the treatment effect.

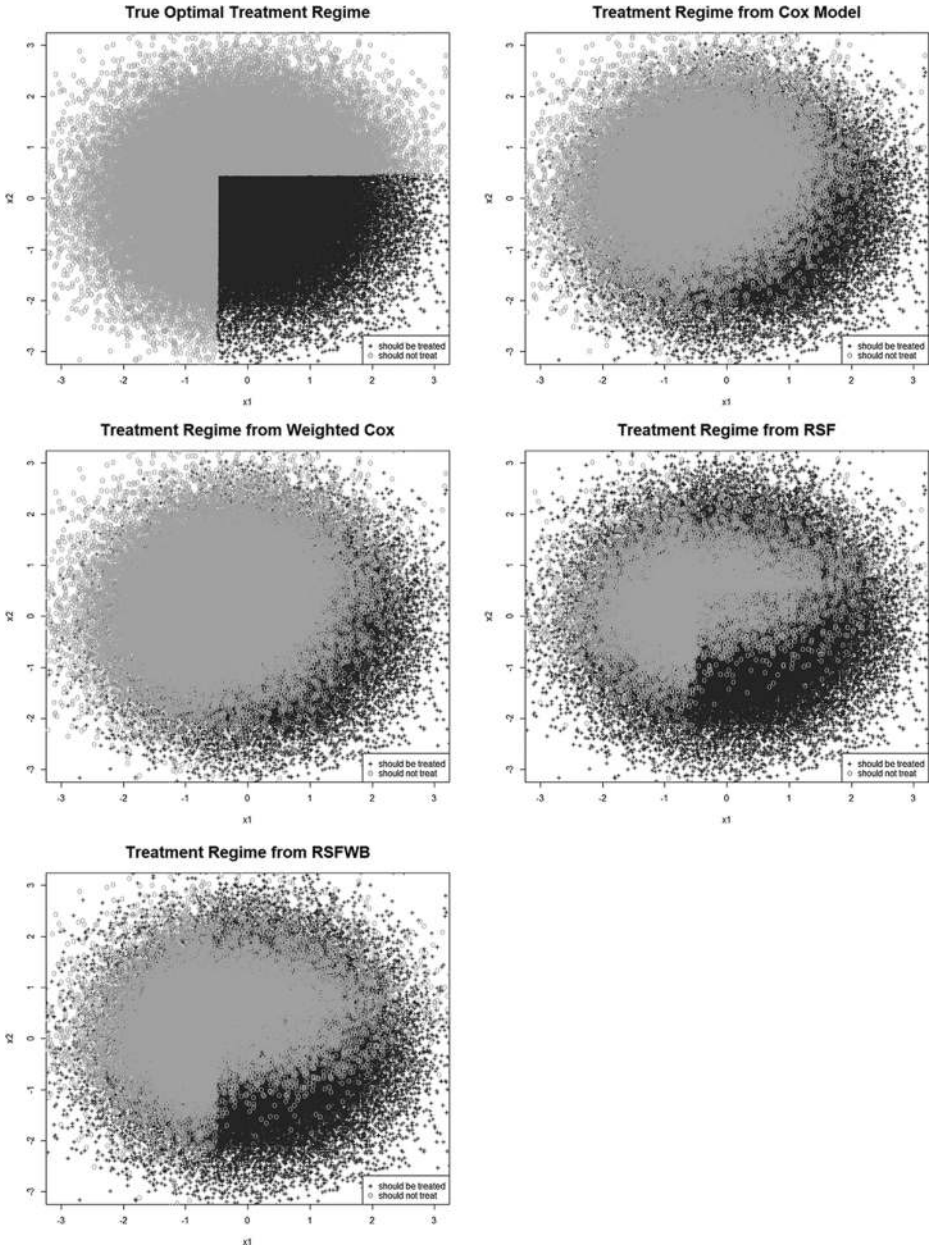


Figure A1:

Cumulative Plot for Estimated Optimal Treatment Regime for the Additional Simulation Study with 85% Censoring. Similar as in Figure 1, the true optimal regime and the ones estimated by the four methods (Cox, weighted Cox, RSF and RSFWB) are plotted separately.

Table 1:

Root Mean Squared Differences (RMSD) for Different Counterfactual Models: For each setting, 4 methods are compared, Cox: standard Cox model; weighted Cox: weighted Cox counterfactual models; RSF: regular RSF model (without weights); RSFWB: proposed method. For each method, the mean and standard deviation of RMSDs are recorded.

		Cox		**weighted Cox**		**RSF**		**RSFWB**	
		mean	**(SD)**	**mean**	**(SD)**	**mean**	**(SD)**	**mean**	**(SD)**
Scenario 1	20% censoring								
	RMSD(1)	17.03	(1.82)	16.00	(2.67)	17.10	(1.76)	16.80	(2.17)
	RMSD(0)	19.40	(4.21)	15.99	(4.52)	22.66	(2.89)	16.86	(2.51)
	45% censoring								
	RMSD(1)	2.42	(0.29)	2.28	(0.32)	1.92	(0.15)	1.90	(0.22)
	RMSD(0)	3.56	(0.53)	2.82	(0.63)	3.54	(0.29)	3.16	(0.39)
Scenario 2	20% censoring								
	RMSD(1)	14.54	(0.71)	14.52	(0.74)	10.97	(0.93)	10.65	(1.12)
	RMSD(0)	6.13	(1.28)	4.70	(1.04)	5.56	(0.87)	4.15	(0.82)
	45% censoring								
	RMSD(1)	3.63	(0.19)	3.51	(0.17)	2.76	(0.22)	2.60	(0.28)
	RMSD(0)	2.10	(0.35)	1.77	(0.32)	1.89	(0.24)	1.77	(0.29)
Scenario 3	20% censoring								
	RMSD(1)	6.16	(0.57)	6.76	(0.90)	4.57	(0.36)	4.38	(0.39)
	RMSD(0)	3.98	(0.58)	5.27	(1.07)	3.84	(0.58)	3.87	(0.71)
	45% censoring								
	RMSD(1)	1.36	(0.11)	1.51	(0.14)	1.11	(0.09)	1.05	(0.10)
	RMSD(0)	0.98	(0.16)	1.21	(0.19)	0.97	(0.14)	0.93	(0.17)

Table 2:

Regime Specific Restricted Mean Survival Time from Different Methods: The regimes under consideration include: the optimal regime estimated for all four methods (Cox, weighted Cox, RSF, and RSFWB), the true optimal treatment regime (opt), and the regime where everybody does not receive treatment (all ctrl) and the regime where everybody receives the treatment (all trt).

		Cox	**weighted Cox**	**RSF**	**RSFWB**	**opt**	**all ctrl**	**all trt**
Scenario 1	20% censoring	29.07	31.39	28.89	29.85	33.53	16.06	28.28
	45% censoring	6.35	6.75	5.92	6.23	7.07	4.94	5.67
Scenario 2	20% censoring	10.32	11.63	12.64	12.97	14.35	7.38	12.83
	45% censoring	4.93	5.20	5.32	5.46	6.06	4.26	5.39
Scenario 3	20% censoring	9.51	9.18	9.58	9.97	11.12	7.58	8.78
	45% censoring	3.34	3.28	3.37	3.42	3.77	2.93	3.17

Table 3:

Summary Statistics of the Prostate Cancer Study Data

		All Patients	Received Surgery	Received Radiation
Variable		**n=3540**	**n=2838 (80.2%)**	**n=702 (19.8%)**
		Mean/Count (SD/Freq)	**Mean/Count (SD/Freq)**	**Mean/Count (SD/Freq)**
Age		61.4 (8.0)	59.9 (7.2)	67.6 (8.1)
Stage	T1	2508 (70.8%)	2104 (74.1%)	404 (57.6%)
	T2	986 (27.9%)	722 (25.4%)	264 (37.6%)
	T3	46 (1.3%)	12 (0.5%)	34 (4.8%)
PSA		2.01 (0.58)	1.94 (0.54)	2.30 (0.66)
#Comorbidity	1	274 (7.7%)	266 (9.4%)	8 (1.2%)
	2	1088 (30.7%)	993 (35.0%)	95 (13.5%)
	3	1103 (31.2%)	942 (33.2%)	161 (22.9%)
	4+	1075 (30.4%)	637 (22.4%)	438 (62.4%)
Gleason	5–7	2711 (76.6%)	2281 (80.4%)	430 (61.3%)
Grade	7.5	437 (12.3%)	316 (11.1%)	121 (17.2%)
	8–10	392 (11.1%)	241 (8.5%)	151 (21.5%)
PNI	N	2579 (72.9%)	2139 (75.4%)	440 (62.7%)
	Y	961 (27.1%)	699 (24.6%)	262 (37.3%)
Treatment Initiation		10.6 (3.9)	11.2 (3.6)	8.1 (3.9)

Note: PSA is logarithm transformed by log(PSA+1); Treatment Initiation is the time from 1996/01/01 to the date of the treatment (in years); All 7 variables are statistically different between the radiation and surgery groups ($p < 0.01$).