Received: Added at production | Revised: Added at production | Accepted: Added at production

DOI: xxx/xxxx

ARTICLE TYPE

Robust Method for Optimal Treatment Decision Making Based on Survival Data

Yuexin Fang[1] | Baqun Zhang[2] | Min Zhang*[3]

[1]Department of Mathematics, Shanghai Normal University, Shanghai, P.R. China
[2]School of Statistics and Management, Shanghai University of Finance and Economics, Shanghai, P.R. China
[3]Department of Biostatistics, University of Michigan, MI, USA

Correspondence
* Min Zhang
1415 Washington Heights, Ann Arbor, MI.
Email: mzhangst@umich.edu

Summary

Identifying the optimal treatment decision rule, where the best treatment for an individual varies according to his/her characteristics, is of great importance when treatment effect heterogeneity exists. We develop methods for estimating the optimal treatment decision rule based on data with survival time as the primary endpoint. Our methods are based on a flexible semiparametric accelerated failure time model, where only the treatment contrast (i.e., the difference in means between treatments) is parameterized and all other aspects are unspecified. An individual's treatment contrast is firstly estimated robustly by an augmented inverse probability weighted estimator (AIPWE). Then the optimal decision rule is estimated by minimizing the loss between the treatment contrast and the AIPWE contrast. Two loss functions with different strategies to account for censoring are proposed. The proposed loss functions distinguish from existing ones in that they are based on treatment contrasts, which completely determine the optimal treatment rule. Our methods can further incorporate a penalty term to select variables that are only important for treatment decision making, while taking advantage of all covariates predictive of outcomes to improve performance. Comprehensive simulation studies have been conducted to evaluate performances of the proposed methods relative to existing methods. The proposed methods are illustrated with an application to the ACTG 175 clinical trial on HIV-infected patients.

KEYWORDS:
Augmented inverse probability weighted estimator; Decision rule; Doubly robust; Optimal treatment regime; Subgroup identification; Variable selection.

1 | INTRODUCTION

The traditional framework for comparative effectiveness research has focused on average treatment effects. That is, the goal is to identify which treatment option, among a set of candidate options, will lead to the best outcomes on average. Within this framework implicitly one assumes that we should then treat every intended subject with the optimal treatment option. In practice, however, it has been long recognized that heterogeneity among patients may exist with regard to responses to different treatment options[1,2,3]. A treatment option that is better on average may not be useful and even be harmful to a subgroup of patients. When heterogeneity in treatment effects exists, alternative to identifying the optimal treatment on average, a more meaningful question is to learn the optimal treatment decision rule to guide future treatment decision making for an individual patient based

on his/her characteristics. Since more than a decade ago there has been an increasing interest in developing statistical methods for estimating the optimal treatment decision rule[4,5,6,7,8,9]. See the recent book on optimal dynamic treatment regimes for a comprehensive review[10].

The majority of the literature on estimating the optimal treatment regime has focused on continuous outcomes, with much less work on time-to-event outcomes. Since the primary outcome in many clinical trials is mortality or time to a major event, time-to-event outcomes are undoubtedly a very important type of outcome to study for optimal treatment decision making. Tian et al[11] proposed a clever modified covariate approach within a Cox model framework to study treatment effect heterogeneity. The modified covariate approach is further coupled with an augmentation procedure to improve efficiency. Extending the work of Lu et al[12] for uncensored outcomes to censored outcomes, Geng et al[13] studied a robust loss-based framework where the loss is weighted by the inverse probability of not being censored to account for censoring. Both methods of Tian et al[11] and Geng et al[13] are robust in the sense that they do not require working models for outcomes to be correctly specified. In addition, both methods are able to accommodate variable selection with regularization through an L_1 penalty. Other work on estimating the optimal treatment regimes for survival data includes, for example, Goldberg and Kosorok[14], Bai et al[15], Jiang et al[16], Hager, Tsiatis and Davidian[17], and Simoneau et al[18].

In this article, we propose methods to estimate the optimal treatment decision rule by directly targeting the treatment contrast function, defined as the difference in conditional expectations between treatments. As discussed in Section 2, the contrast function between treatments completely defines the optimal treatment decision rule. We propose to estimate the optimal decision rule by minimizing the empirical discrepancy between treatment contrasts and the estimated treatment contrasts for subjects. Specifically, we estimate the treatment contrast for an individual patient using an augmented inverse probability weighted approach, which possesses a robustness property in the setting of a randomized clinical trial and leads to good efficiency. As a result, the resulting method also enjoys robustness and good performances.

The remainder of the article is organized as follows. Section 2 introduces notations, formally defines the optimal treatment regime and discusses assumptions necessary to identify the optimal decision rule from the observed data. Section 3.1 introduce the model and Section 3.2 reviews some closely related methods that motivated the proposed methods. We present the proposed methods and two contrast function-based loss functions in Section 3.3, with details on modeling survival times for estimating treatment contrasts in Section 3.4. Section 3.5 studies asymptotic properties of the two loss functions and variable selection, and discusses the impacts of inverse probability censoring weighting versus not weighting. Comprehensive simulation studies and a real data application are reported in Sections 4 and 5, followed by a discussion in Section 6.

2 | NOTATIONS AND BACKGROUND

Consider a study with n subjects sampled from the patient population of interest. Let A_i, taking values 0 or 1, denote the treatment received by subject i and X_i denote a p-dimensional vector of covariates. When data are obtained from a randomized clinical trial, then we would have $A_i \perp X_i$ by randomization. More generally, the treatment assignment may depend on covariates in observational studies. As in Geng et al[13] we focus on the setting of a randomized clinical trial, but we also consider the more general setting where treatments are not randomized. Let T_i denote the survival time of interest and C_i the censoring time for subject i. Following Geng et al[13] and as is usually assumed for a randomized clinical trial, we make the independent censoring assumption, denoted by $C_i \perp (T_i, X_i, A_i)$. For survival data, one observes the minimum of survival time and censoring time, and we denote $U_i = \min(T_i, C_i) = T_i \wedge C_i$, where $a \wedge b$ stands for $\min(a, b)$. For convenience, we also define $Y_i = \log(T_i)$, and then equivalently one gets to observe $\widetilde{Y}_i = Y_i \wedge \log(C_i)$. The indicator for not being censored is defined as $\delta_i = I(T_i \leq C_i) = I(Y_i \leq \log(C_i))$. In summary, the observed data consist of $(\widetilde{Y}_i, U_i, X_i, A_i, \delta_i), i = 1, \ldots, n$, identically and independently across subject i.

To formally define the optimal treatment regime, we adopt the potential outcomes framework. Let $T_i^*(0), T_i^*(1)$ be the potential survival time had the patient, possibly contrary to fact, received treatment 0 and 1 respectively. Similarly, $Y_i^*(0), Y_i^*(1)$ are the potential log survival time for subject i. A treatment regime, $g(X)$, is a function of X that takes value 0 or 1 and maps values of covariates into a treatment option. It is a treatment decision rule that dictates how the treatment should be prescribed according to a given vector of covariates. For subject i following the treatment regime g, the corresponding potential survival time is defined as $T_i^*(g) = g(X_i)T_i^*(1) + \{1 - g(X_i)\}T_i^*(0)$. That is, it is equal to the potential outcome $T_i^*(a)$, where $a = g(X_i)$ is the treatment decision prescribed for subject i by the decision rule g. Similarly, one may define the potential log survival time as $Y_i^*(g) = g(X_i)Y_i^*(1)+\{1-g(X_i)\}Y_i^*(0)$. There are many possible decision rules and our goal is to estimate the optimal treatment regime based on the observed data. That is, we aim to use the observed data to learn the optimal treatment decision rule. In this

article the optimal treatment regime is defined as the one that maximizes the mean potential survival time and, equivalently the mean potential log survival time. Formally, the optimal treatment regime, g^{opt}, satisfies

$$g^{opt} = \arg\max_{g\in\mathcal{G}} E\{T^*(g)\} = \arg\max_{g\in\mathcal{G}} E\{Y^*(g)\},$$

where $\mathcal{G}$ denotes the set of all treatment regimes under consideration. The expectation $E\{Y^*(g)\}$ is commonly referred to as the value of the regime g.

The optimal treatment regime is defined in terms of potential outcomes, which are not completely observed even in the absence of censoring. In order for the optimal treatment regime to be identifiable from the observed data, we make the following standard assumptions[9].

(C1) Stable unit treatment value assumption: $Y = Y^*(1)A + Y^*(0)(1-A)$;

(C2) No unmeasured confounders assumption : $A \perp \{Y^*(0), Y^*(1)\}|X$.

When data are obtained from a randomized clinical trial, the no unmeasured confounders assumption is satisfied automatically by design. Under these two assumptions, the optimal treatment regime is identifiable from the observed data. Alternative definitions of the optimal treatment regime based on the observed data under the two assumptions and some further notations are given below; see, for example, Zhang, Tsiatis and Davidian[19] and Zhang et al[20] for more details. It is straightforward to see that

$$\begin{aligned} E\{Y^*(g)\} &= E\Big\{ E[g(X)Y^*(1) + \{1-g(X)\}Y^*(0)|X] \Big\} \\ &= E[g(X)E\{Y|A=1,X\} + \{1-g(X)\}E\{Y|A=0,X\}] \\ &= E[g(X)\{u_1(X) - u_0(X)\} + u_0(X)], \end{aligned}$$

where $u_a(X) = E(Y|A=a,X)$. Then the optimal treatment regime satisfies

$$\begin{aligned} g^{opt} &= \arg\max_{g\in\mathcal{G}} E[g(X)\{u_1(X) - u_0(X)\}] \\ &\equiv \arg\max_{g\in\mathcal{G}} E\{g(X)C(X)\}, \end{aligned}$$

where $C(X) = u_1(X) - u_0(X)$ and is referred to as the contrast function between treatments. Therefore, the optimal treatment regime can be identified through $u_a(X)$, the conditional mean of observed outcomes given the treatment and covariates. More specifically, the optimal treatment regime is determined by the treatment contrast $C(X)$. That is, by solving the above optimization problem, it is clear that

$$g^{opt}(X) = I(C(X) > 0)). \tag{1}$$

The discussion above assumes no censoring exists. For survival data, further assumption regarding censoring is needed for estimation of the optimal treatment regime from the observed data. In the following development, we assume the independent censoring assumption described earlier.

3 | METHODS

3.1 | Model

As discussed above, the optimal treatment regime is directly related to the treatment contrast function $C(X)$ and, equivalently, the conditional mean of outcomes given the treatment and covariates, $u_a(X) = E(Y|A=a,X)$. We assume that, given the treatment and covariates, the log survival time follows the following semiparametric model:

$$Y = \log(T) = M(X) + \gamma' H(X)A + \varepsilon, \tag{2}$$

where $M(X)$ is an unspecified function of X, $H(X)$ is a q-dimensional function of baseline covariates X and always includes an intercept, and ε is an error term such that $E(\varepsilon|X,A) = 0$. For simplicity, we denote $H(X)$ by H_X. Model (2) is a semiparametric accelerated failure time (AFT) model. Under this model, $u_0(X) = E(Y|X,A=0) = M(X)$, which is unspecified, and $u_1(X) = M(X) + \gamma' H_X$. The treatment contrast function $C(X)$ equals $u_1(X) - u_0(X) = \gamma' H_X$ and is modeled parametrically. This semiparametric model only makes assumptions on the parametric form of the treatment contrast function, leaving other aspects of the conditional mean function unspecified. Therefore, it is more robust and flexible than the usual regression model. This semiparametric model has been popular for estimating the optimal treatment regimes for outcomes without censoring[21], commonly referred to as A-learning. Under this model, it is easy to see that the optimal treatment regime is given by $g^{opt}(X) = I(\gamma' H_X > 0)$ according to (1). We comment that for censored survival data, the mean of survival time is usually not

estimable without making additional assumptions. However, under the assumed semiparametric accelerated failure time model, even though the mean log survival time is still not estimable since no assumptions on ϵ are made, the optimal treatment regime, i.e., $I(\gamma' H_X > 0)$, is estimable as it depends only on model coefficients but not on the underlying distribution of Y. A natural way to estimate the optimal treatment regime would be $I(\widehat{\gamma}' H_X > 0)$, where $\widehat{\gamma}$ is an estimator for γ. When censoring is absent, γ and hence the optimal treatment regime can be estimated by the G-estimation method[21]. The censoring poses additional challenges. Below we briefly introduce two methods that are closely related to and motivated the proposed methods.

3.2 | Related Methods

When no censoring exists, Lu et al[12] proposed a general and robust framework for variable selection and estimating the optimal treatment regime. Later Geng et al[13] generalized the framework to survival outcomes to account for censoring. Geng et al[13] considered the same model given in (2) and proposed to minimize the following inverse probability censoring weighted loss function:

$$\frac{1}{n}\sum_{i=1}^{n}\frac{\delta_i}{\widehat{G}(T_i)}\left\{Y_i - \phi(X_i;\widehat{\theta}) - \gamma' H_{X_i}(A_i - \widehat{\pi}_i)\right\}^2, \tag{3}$$

where $\widehat{G}(t)$ is the Kaplan-Meier estimator for the survival function of C at t, $\phi(X_i;\theta)$ is a posited parametric function for the baseline function $M(X)$, and $\widehat{\pi}_i \equiv \widehat{\pi}(X_i)$ is an estimate of the propensity score $P(A_i = 1|X_i)$. Geng et al[13] demonstrated that this method is robust for randomized clinical trials in the sense that it does not require the correct specification of the baseline function. For variable selection, they proposed to couple the above loss function with the adaptive LASSO penalty and minimize a penalized loss function.

When there is no censoring, Zhang et al[20] proposed a general direct optimization framework to estimate the optimal treatment regimes. Zhang et al[20] showed that popular existing methods for estimating the optimal treatment regime can fit into this direct optimization framework by a specific choice of the estimator for the contrast function $C(X)$. In particular, the robust method of Zhang et al[19], based on augmented inverse probability weighted estimators (AIPWE) for population means, is equivalent to estimating $C(X)$ by the following AIPWE estimator

$$\widehat{C}(X_i) = \frac{A_i}{\widehat{\pi}_i}Y_i - \frac{A_i - \widehat{\pi}_i}{\widehat{\pi}_i}\widehat{u}_1(X_i) - \left\{\frac{1 - A_i}{1 - \widehat{\pi}_i}Y_i - \frac{\widehat{\pi}_i - A_i}{1 - \widehat{\pi}_i}\widehat{u}_0(X_i)\right\}, \tag{4}$$

where $\widehat{u}_a(X), a = 0, 1$, is an estimate of $E(Y|A = a, X)$ that will be discussed in detail later. Zhang et al[20] and Zhang et al[19] have shown that the resulting estimator for the optimal treatment regime is doubly robust in the sense that it only requires the propensity score model or the outcome model, but not necessarily both, to be correct. For randomized clinical trials, the AIPWE for a population mean is always consistent as the model for the treatment probability (propensity score) can always be correct. Working models for $\widehat{u}_a(X_i)$ are used for improve efficiency because the corresponding AIPWE for a population mean is semiparametric efficient when $\widehat{u}_a(X), a = 0, 1$, are modeled correctly. As a result, robust AIPWE-based methods have superior performances as shown in Zhang et al[20] and Zhang et al[19].

3.3 | Proposed Loss Functions

Our proposed method has a close relationship with the two methods introduced in 3.2. Similar to Geng et al[13], we aim to estimate γ by minimizing a loss function that accounts for censoring. In our proposed loss function, we would like to take advantage of the general and robust AIPWE-based framework and the estimated contrast function (4). Our proposed method is based on the following considerations. First, as discussed at the end of Section 2, the optimal treatment regime is completely determined by the contrast function; that is, $g^{opt}(X) = I(C(X) > 0)$. Second, as briefly introduced in 3.2 and demonstrated by Zhang et al[20], popular existing methods essentially differ in ways to estimate the contrast function. In particular, the AIPWE approach (4) has been shown to be appealing in terms of robustness and efficiency[20,19]. Third, according to the semiparametric model (2), we assume the true contrast function is $C(X) = \gamma' H_X$. Due to these considerations, we propose to estimate γ and hence the optimal treatment regime by minimizing the following loss function

$$L_n(\gamma) = \frac{1}{n}\sum_{i=1}^{n}\frac{\delta_i}{\widehat{G}(T_i)}\{\widehat{C}(X_i) - \gamma' H_{X_i}\}^2, \tag{5}$$

where $\widehat{C}(X_i)$ is the robust AIPWE for the contrast function defined in (4). Compare the proposed loss function (5) with (3) studied by Geng et al[13]. We note that (3) minimizes a weighted discrepancy between outcomes and a function involving several components. The form of the function, although based on sound theoretical development, is not entirely intuitive and lacks an easy interpretation. Alternatively, the proposed loss function (5) directly minimizes a weighted discrepancy between the estimated treatment contrast and the contrast function implied by model (2) for all subjects. In construction of $\widehat{C}(X_i)$, one needs to estimate the treatment propensity $P(A = 1|X)$. For randomized clinical trials, the treatment propensity can be estimated by $\widehat{\pi}_i = \widehat{\pi} = n^{-1}\sum_i^n A_i$. Alternatively, as the treatment probability is known by design for a randomized trial and we may substitute $\widehat{\pi}_i$ in $\widehat{C}(X_i)$ directly by the known constant. For example, for a randomized trial with $\pi = 1/2$, it is easy to verify that

$$\widehat{C}(X) = 2(2A-1)\left\{Y - \frac{\widehat{u}_1(X) + \widehat{u}_0(X)}{2}\right\}.$$

In an observational study, when the treatment assignment is not randomized, one needs to model $P(A = 1|X)$. For example, a logistic regression model with lasso for variable selection may be used.

The loss function in (5) is weighted by the inverse probability of not being censored, which is intuitive and leads to good theoretic properties in large samples as shown in Section 3.5. However, based on our experience, inverse probability censoring weighting often leads to large variability in the loss function and, as a result, degraded performances empirically. This is likely due to the fact that a proportion of subjects may have small probability of not being censored by time $t = T_i$ and therefore receive large weights. Alternatively, we also propose to estimate γ by minimizing the unweighted loss function

$$L_n(\gamma) = \frac{1}{n}\sum_{i=1}^{n}\delta_i\{\widehat{C}_\delta(X_i) - \gamma' H_{X_i}\}^2, \tag{6}$$

where $\widehat{C}_\delta(X_i)$ mimics $\widehat{C}(X_i)$ and is defined as $\frac{A_i}{\widehat{\pi}_i^\delta}Y_i - \frac{A_i - \widehat{\pi}_i^\delta}{\widehat{\pi}_i^\delta}\widehat{u}_1(X_i) - \left\{\frac{1-A_i}{1-\widehat{\pi}_i^\delta}Y_i - \frac{\widehat{\pi}_i^\delta - A_i}{1-\widehat{\pi}_i^\delta}\widehat{u}_0(X_i)\right\}$, with $\widehat{\pi}_i^\delta \equiv \widehat{\pi}^\delta(X_i)$ estimating $P(A_i = 1|X_i, \delta_i = 1)$. Justification of this loss function is further discussed in Section 3.5 where we study the asymptotic property of the resulting estimator for γ. Both weighted and unweighted versions of the method is studied empirically by simulations below. Note that, although the loss function is not weighted by the inverse probability of not being censored, censoring is accounted for appropriately in construction of $\widehat{C}(X_i)$ as discussed in Section 3.4. For the treatment probability $P(A = 1|X, \delta = 1)$, a logistic model with lasso for variable selection could be fit using the uncensored data.

The estimator $\widehat{C}(X)$ also involves $\widehat{u}_a(X)$, which is estimate of $u_a(X)$. We discuss the details of estimation of $u_a(X)$ below.

3.4 | Estimation of $u_a(X)$

As the outcome Y is subject to censoring, naturally one can model it using models developed for survival outcomes to incorporate information from both censored and uncensored subjects. The Cox proportional hazards model and the accelerated failure time model are both extensively studied models for modeling time-to-event data subject to censoring. In our method we choose to model outcomes using a semiparametric AFT model because it directly relates covariates to the mean of Y, whereas the Cox model relates covariates to the hazard of Y instead. As our goal is to estimate the conditional mean $u_a(X)$, an AFT model is a more natural and direct choice. Several methods are available for fitting the AFT model and can be used for estimating $u_a(X)$. As discussed in Komárek et al[22], some existing methods may have less ideal computational properties. For example, they may fail to converge or become computationally intractable when the dimension of covariates is high. In our simulation studies we adopt the penalized Gaussian mixture method for fitting semiparametric AFT models[22], where the baseline distribution is left unspecified and modeled flexibly using Gaussian basis densities. Specifically, to estimate $u_a(X)$, for $a = 0$ and 1 separately we fit the following working AFT model based on subjects with $A_i = a$, i.e.,

$$Y = \log(T) = \beta'\widetilde{X} + \sigma\varepsilon,$$

where $\widetilde{X} = (1, X')'$, β is a $(p+1)$-dimensional vector of coefficients, σ is a scale parameter controlling the variance, and the density of the error term ε is unspecified and is denoted as $f(e)$. In the method of Komárek et al[22], $f(e)$ is modeled as a mixture of Gaussian densities, denoted as

$$f(e) = \sum_{j=1}^{g} c_j \varphi_{\mu_j, \sigma_0^2}(e),$$

where $\varphi_{\mu_j,\sigma_0^2}(e)$ is the Gaussian density with mean μ_j and variance σ_0^2 with values of $\mu_1, \ldots, \mu_g$ and σ_0^2 fixed by design and $(c_1, \ldots, c_g)$ are mixture coefficients to be estimated. The basis functions $\varphi_{\mu_j,\sigma_0^2}(e)$ are referred to as basis Gaussian densities. As explained in Komárek et al[22], the motivation for the method stems from the penalized B-spline smoothing method for modeling densities. The means of Gaussian densities $\mu_1, \ldots, \mu_g$ play similar roles as fixed knots in a spline smoothing method. Basis Gaussian densities can be viewed as the limiting case of B-spline smoothing and are advantageous in that it can model densities with infinite support. A penalized log-likelihood method was proposed to estimate unknown parameters. Then one can estimate $\widehat{u}_a(X) = \widehat{\beta}_a' \widetilde{X}$, where $\widehat{\beta}_a$ estimates the unknown coefficient in the AFT model for treatment group a. The implementation of the method can be easily carried out using the R package *smoothSurv*[23]. When the dimension of covariates are high, then variable selection is recommended. As the method of Komárek et al[22] does not incorporate a variable selection procedure, we suggest that we first conduct variable selection before fitting the AFT model with a smoothed error term. For example, one may conduct variable selection using the adaptive elastic net[24], implemented by the R-package *AdapEnetClass*. This strategy is used in our simulation studies reported in Section 4.

3.5 | Asymptotic Properties and Variable Selection

Let γ_0 be the true value of γ. We first consider the method where we minimize the inverse probability censoring weighted loss function . Let $\widehat{\gamma}$ be the solution that minimizes (5). Then the corresponding estimated optimal treatment regimes is $I(\widehat{\gamma}' H_X > 0)$. We have the following result regarding the asymptotic property of $\widehat{\gamma}$.

Result 1: Under certain regular conditions, for data obtained from a randomized clinical trial, regardless of whether the model for $E(Y|A, X)$ is correctly specified or not, as n goes to infinity, $\widehat{\gamma}$ converges in probability to γ_0.

The proof for this result is given in the Appendix. In the proof we show that $\widehat{\gamma}$ solves an estimating equation and the true γ_0 is the unique solution to the population analog of the estimating equation. Therefore, $\widehat{\gamma}$ is consistent for γ_0 by M-estimation/Z-estimation theory[25]. Therefore, the proposed method for estimating the optimal treatment regime enjoys a nice robustness property. This robustness is a result of randomization. For observational studies, the consistency of $\widehat{\gamma}$ requires an additional assumption that the propensity score $\pi(X) = P(A = 1|X)$ is modeled correctly. Results discussed above are based on model (2), which assumes the treatment contrast can be parameterized by γH_X. Regardless of whether this is a correct parameterization, the chosen parameterization defines a class of regimes indexed by parameter γ, and the proposed methods target the optimal decision rule within this class. Restricting to class of regimes indexed by parameters or of a specific form (e.g., decision-tree) to incorporate considerations of clinical knowledge and practice, cost and interpretability has been a common approach in literature[19,26].

When instead one minimizes the unweighted loss function (6), the minimizer, say $\widetilde{\gamma}$, still solves an estimating equation. However, γ_0 is the solution to the population analog of the estimating equation under the condition that the model for $P(A = 1|X, \delta = 1)$ is correct. As a result, the minimizer of (6) is consistent to γ_0 when $P(A = 1|X, \delta = 1)$ is correctly modeled. We comment that a weighted loss function introduces large variability in the loss function as well as the corresponding estimating function. This is especially true when $1/\widehat{G}(T_i)$ are very large for some subjects due to that $\widehat{G}(T_i)$ are close to zero. As a result, $\widehat{\gamma}$ may exhibit larger variability than $\widetilde{\gamma}$. And the unweighted loss function may lead to a better estimator of the optimal treatment regime in practice, although it requires modeling for $P(A = 1|X, \delta = 1)$ and the model is not necessarily correct even in a randomized trial. The above discussion provides a heuristic justification for the unweighted loss function and more details are in the Appendix. Our simulation studies reported in Section 4 show that indeed the unweighted loss function leads to better performances in general. When the treatment assignment is not randomized, then similar consistency results for $\widehat{\gamma}$ and $\widetilde{\gamma}$ hold under the condition that the corresponding propensity score model, $P(A = 1|X)$ or $P(A = 1|\delta = 1, X)$, is correctly specified, without requiring correct modeling for $E(Y|A, X)$.

When the dimension of covariates is high, variable selection is often necessary to reduce variability of estimation. Within the proposed framework, we can incorporate variable selection at two stages, each targeting selecting a different set of important covariates. As discussed previously in Section 3.4, we may carry out variable selection while modeling $E(Y|A, X)$. At this stage, the targeting covariates are factors that are predictive of outcomes, termed as predictive variables. Many existing methods are available for variable selection for predictive variables. Only predictive variables that interact qualitatively with treatment are variables important for treatment decision making and these variables are referred to as prescriptive variables. At the second stage, we target selecting prescriptive variables. To achieve this goal, as in Geng et al[13] one natural strategy is to include a

penalty term in the loss function used to estimate parameters in the contrast function. Specifically, one would then estimate γ by minimizing a penalized loss function, i.e.,

$$\min_{\gamma} L_n(\gamma) + \lambda_n ||\gamma||_1, \tag{7}$$

where $L_n(\gamma)$ is a weighted or unweighted loss function defined in (5) or (6), and $||\gamma||_1 = \sum_{j=1}^{p} |\gamma_j|$ and is a lasso-type penalty. Lasso-type penalty has been a popular choice for variable selection and has been extensively studied in statistics literature. For example, lasso penalty is also used in the modified covariates method of Tian et al[11] to select important variables within a Cox model framework. As the proposed loss functions are based on treatment contrasts only, this variable selection specifically targets only prescriptive variables important for treatment decision making. Variables that are not important for treatment decision making but predictive of outcomes are incorporated in augmentation terms in (1) and are useful for improving efficiency and performances.

4 | SIMULATION STUDIES

We conducted several simulation studies to evaluate the performance of the proposed methods and to compare them to the methods of Geng et al[13] and Tian et al[11]. Specifically, for the proposed methods, AIPWE_AFT and AIPWE_AFT* denote the resulting estimator by minimizing the inverse probability censoring weighted and unweighted loss functions (5) and (6) respectively. For each proposed method, we implement the penalized and unpenalized versions. In the implementation of each method, the corresponding model includes linear terms of covariates and their interactions with treatment are considered, but not interactions of covariates and higher order terms. Data were generated under various scenarios with different sample sizes, censoring percentages, and outcome models. As in previous studies, we focus on the setting of randomized clinical trials. Additional simulation studies mimicking observational studies are reported in the Supplementary Material. Specifically, for each scenario, data were generated with sample sizes $n = 400$ and 1000. The treatment A_i was generated as Bernoulli (0.5), mimicking a randomized clinical trial. The censoring C_i was generated as $\log(C_i) \sim \text{Uniform}(0, c)$, where c was chosen to induce 15% or 40% censoring rate. We denote $X_i = (X_{i1}, \cdots, X_{ip})'$, where p is the dimension of X_i and is set to 50 in our simulations. For simplicity, we define $\widetilde{X}_i = (1, X_i^T)^T$ and $\mathbf{m}_q$ as a q-dimensional vector consists of all m; for example, $\mathbf{0}_2 = (0, 0)^T$. Survival times were related to covariates and treatments according to three models detailed below.

• Model 1: This scenario is the fourth setting of Geng et al[13]. The survival time is related to covariates through an accelerated failure time model as follows.

$$\log(T_i) = 1 + 0.5(\beta_1' X_i)(\beta_2' X_i) + \gamma' \widetilde{X}_i A_i + \varepsilon_i,$$

where $\beta_1 = (1, 1, \mathbf{0}_{48})', \beta_2 = (1, \mathbf{0}_2, -1, \mathbf{0}_5, 1, \mathbf{0}_{40})', \gamma = (1, 1, \mathbf{0}_7, -0.9, 0.8, \mathbf{0}_{40})'$. Covariate X_i follows a multivariate normal distribution with mean 0, variance 1 and correlation $\text{Corr}(X_{ij}, X_{ik}) = 0.5^{|j-k|}$.

• Model 2: This scenario is adapted from Tian et al[11].

$$\log(T_i) = \left(\beta_0 + \sum_{j=1}^{p} \beta_j X_{ij}\right)^2 + \left(\gamma_0 + \sum_{j=1}^{p} \gamma_j X_{ij} + 0.8 X_{i1} X_{i2}\right) A_i + \sigma_0 \varepsilon_i,$$

where $\varepsilon_i \sim N(0, 1)$, $\beta_0 = (\sqrt{3})^{-1}$, $\beta_j = (2\sqrt{3})^{-1}$ for $j = 3, \cdots, 10$ and $\beta_j = 0$ otherwise, $\gamma = (0.4, 0.8, -0.8, 0.8, -0.8, \mathbf{0}_{46})$, $\sigma_0 = \sqrt{2}$. Covariate X_i follows an independent standard multivariate normal distribution.

• Model 3: Survival time T_i follows a Cox proportional hazards model.

$$\lambda(t|A_i, X_i) = 0.2 \exp(0.2t) \exp[-\{0.5(\beta_1' X_i)(\beta_2' X_i) + \gamma' \widetilde{X}_i A_i\}],$$

where $\beta_1 = (1, 1, \mathbf{0}_{48})', \beta_2 = (1, \mathbf{0}_2, -1, \mathbf{0}_5, 1, \mathbf{0}_{40})', \gamma = (1, 1, \mathbf{0}_7, -0.9, 0.8, \mathbf{0}_{40})'$. Covariate X_i follows a multivariate normal distribution with mean 0, variance 1.

For assessing the accuracy of the estimated treatment decision rules, PCD is the percentage of making correct treatment decisions of an estimated regime $\widehat{g}^{opt}$, defined as $n^{-1} \sum_{i=1}^{n} I\{\widehat{g}^{opt}(X_i) = g^{opt}(X_i)\}$. Additionally, we report the mean squared error of an estimator of γ, and the value of the estimated regime, i.e., the expectation of potential outcomes under the estimated regime. The Value of the true optimal regime is denoted by V_0. Following Geng et al[13], we report the following metrics to evaluate the variable selection performance: the number of non-zero coefficients incorrectly identified as zero (denoted by "Incor0"), the number of correct zero coefficients identified (denoted by "Corr0"), and the proportion of covering all the important variables

(denoted by "Cover"). The number of zero coefficients is 47 under Models 1 and 3, and is 46 under Model 2. Simulation results are based on 500 Monte Carlo data sets. Reported values of regimes are calculated based on 100,000 samples under each model.

Results for models 1-3 are summarized in Figure 1 and Table 1. Additional results on variable selection are reported in Table 2. We make the following comments regarding performances of various methods. First, under models 1 and 2 where survival times follow an AFT model, the two proposed methods and the "Geng et al" method perform considerably better than the "Tian et all" method. This result is not surprising as the true outcome models are AFT models but the method of "Tian et al" is based on the framework of Cox proportional hazards models. But we note that since the true outcome models involve complicated interaction and nonlinear terms of covariates, working models in the proposed methods as well as in the method of "Geng et al" are actually all incorrect. Second, under model 1, the proposed methods perform comparably as the method of "Geng et al" when the censoring rate is low and significantly better when the censoring rate is high. Third, under model 2 when the main effect of covariates is highly nonlinear, the proposed methods, especially AIPWE_AFT*, have much better performances than that of "Geng et al" regardless of the censoring rate and sample size. Fourth, under model 3 where survival times follow a Cox model with covariates, not surprisingly the method of "Tian et al" has the best performance because the posited model is also a Cox model but the posited working models in the proposed methods and the method of "Geng et al" are incorrect. The performance of AIPWE_AFT* under this scenario, although slightly worse than "Tian et al", is close to that of "Tian et al" and is significantly better than the method of "Geng et al" especially when $n = 1000$ or when the censoring rate is high. Finally, regarding penalization and variable selection, we note adding a penalty term when the dimensional of covariates is high improves performances for all methods in all scenarios with different outcome models, sample sizes and censoring rates. Overall we see that AIPWE_AFT*, the proposed method that minimizes the unweighted loss function, has the most robust performance across different scenarios. In terms of variable selection, all methods perform well when the sample size is large. When the sample size is small, the method of "Tian et al" tends to miss more important variables and have smaller proportion of covering all important variables.

Both AIPWE_AFT and the method of "Geng et al" are based on inverse probability censoring weighted loss functions and both utilize AFT models as working models. Based on our simulations, AIPWE_AFT consistently performs as well as and often significantly better than the method of "Geng et al". The difference lies in that the proposed method directly targets the contrast function, which determines the optimal treatment regime as discussed at the end of Section 2, and that AIPWE_AFT exploits a robust and efficient way to estimate the contrast function. In contrast, in the method of "Geng et al", the loss function also involves the main effect of covariates, i.e., $M(X_i)$. AIPWE_AFT*, which is based on minimizing an unweighted version of the loss function, further improves performances relative to AIPWE_AFT uniformly across all scenarios reported here and other unreported scenarios. In particular, it holds even when the censoring rate is relatively high. This result appears counter-intuitive initially. Because inverse probability of censoring weighting is a natural idea for handling censoring and, as shown in Result 1, it leads to consistent estimation of γ. As discussed in Section 3.5, inverse probability censoring weighting often introduces additional variability and leads to unstable estimators of γ and the optimal treatment regime. As shown in the Appendix, the unweighted estimator for γ is consistent under an additional assumption on the propensity score model. But it is much less variable and is better in terms of mean squared error in all scenarios we considered as shown in Tables 1 and 3. Due to the bias and variance trade-off, the unweighted version is often a better choice under reasonable sample sizes we typically see in practice. Based on our simulations, the inverse probability censoring weighted version only wins when the sample size is extremely large.

As Geng et al [13], we have also made the independence censoring assumption. Following Geng et al [13], we conducted sensitivity analysis to assess the impact of violation of this assumption. In each model (Models 1-3 considered previously), censoring was also generated according to $\log(C_i) = \tau_c + \eta' X_i + e_i$, where e_i followed the standard extreme value distribution, τ_c was chosen to control the censoring rate and $\eta = (1, \mathbf{0}_4, 1, \mathbf{0}_{44})$. Results for sensitivity analysis are reported in Figure 2 and Tables 3 and 4. Based on these results, we see that in general all methods are not much sensitive to violation of the independence censoring assumption and the comparative performances of various methods are very similar to what we have observed previously when the independence assumption holds.

Results on scenarios where the treatment probability depends on covariates are reported in the Supplementary Material. Under these scenarios, all methods perform reasonably well. Notable difference is that all methods have worse performance compared to when the treatment is randomized, which is expected as nonrandom treatment assignment increases variability and the difficulty to learn the optimal decision rule.

5 | DATA APPLICATION

We applied the various methods to data obtained from the AIDS Clinical Trials Group Protocol 175 (ACTG 175) on HIV-infected patients. Patients in ACTG 175 were randomly assigned with equal probability to receive one of four treatments: zidovudine (ZDV) monotherapy, ZDV+didanosine (ddI), ZDV+zalcitabine, and ddI monotherapy. The primary outcome was a composite endpoint corresponding to the first time a patient had a greater than 50% decline in CD4 cell count or death. Our analysis focused on the two combination treatments ZDV+ddI ($A = 1$) and ZDV+zalcitabline ($A = 0$), aiming to using the observed data to learn the optimal treatment decision rule to prescribe either ZDV+ddI or ZDV+alcitabline to the right patients. Our analysis was based on 522 and 524 patients who have received ZDV+ddI and ZDV+zalcitabline respectively. On these 1046 patients, about 79.7% of them were censored due to the end of the study or loss to follow-up.

As Zhang et al[27], our analysis considered 12 baseline covariates in deriving the optimal treatment regime: age (years), weight (kg), Karnofsky score (scale of 0-100), CD4 count (cells/mm^3), CD8 (cells/mm^3), all of which ae continuous variables; and binary variables indicating hemophilia (0=no, 1=yes), homosexual activity (0=no, 1=yes), history of intravenous drug use (0=no, 1=yes), race (0=white, 1=non-white), gender (0=female, 1=male), antiretroviral history (0=naive, 1=experienced) and symptomatic status (0=asymptomatic, 1=symptomatic).

Applying the two proposed methods, AIPWE_AFT* and AIPWE_AFT, and the two comparison methods Geng et al[13] and Tian et al[11] to the ACTG 175 data leads to the following estimated optimal treatment decision rules:

$$\begin{aligned}
\widehat{g}^{opt}_{AIPWE_AFT^*} &= I(-1.015 + 0.022\ \text{Age} + 0.367\ \text{Anti} >= 0),\\
\widehat{g}^{opt}_{AIPWE_AFT} &= I(-1.172 + 0.025\ \text{Age} + 0.010\ \text{Race} + 0.406\ \text{Anti} >= 0),\\
\widehat{g}^{opt}_{Geng\ et\ al} &= 1,\\
\widehat{g}^{opt}_{Tian\ et\ al} &= 1.
\end{aligned}$$

No covariates were selected in the methods of "Geng et al" and "Tian et al". They recommend all patients receive ZDV+ddI regardless of a patient's characteristics. Based on the estimated decision rules, the methods of AIPWE_AFT* and AIPWE_AFT recommends that 485 and 498 patients respectively, out of 1046 patients, in our observed data should receive ZDV+ddI.

Figure 3 plots the Kaplan-Meier survival curves for patients following the estimated optimal regimes from various methods. That is, if $A_i = \widehat{g}^{opt}(X_i)$ then the actual treatment received by subject i is consistent with the treatment option that would be prescribed to him/her had he/she followed the regime and therefore $T_i^*(\widehat{g}^{opt})$ is observed. In addition to Kaplan-Meier curves corresponding to the estimated regimes, Figure 3 also includes Kaplan-Meier curves corresponding to the two fixed regimes, each of which always prescribes ZDV+ddI or ZDV+alcitabline to all patients. Note, the estimated regimes by "Geng et al" and "Tian et al" are the same as the regime that always prescribes ZDV+ddI and there are a total of four Kaplan-Meier curves in Figure 3. The estimated regimes from the two proposed methods (AIPWE_AFT* and AIPWE_AFT) lead to very similar survival curves. In addition, it seems that the two regimes that always prescribe the same treatment to all patients, regardless of patients' characteristics, have lower survival probabilities.

As mean survival times are not estimable without imposing additional assumptions, following Geng et al[13], we also estimated nonparametrically the potential restricted mean log survival time corresponding to various estimated regimes and the two fixed regimes. The potential restricted mean log survival time of a regime g is defined as $E\{Y^*(g) \wedge L\}$ and $L = \tau_c$ is chosen to be the log of the maximum follow-up time. Other than a survival curve, the potential restricted mean log survival time of a regime g provides a scalar summary of the quality of a regime g. The estimated restricted mean log survival time corresponding to $\widehat{g}^{opt}_{AIPWE\_AFT^*}$ and $\widehat{g}^{opt}_{AIPWE\_AFT}$ are 6.432 (95% CI: 6.321, 6.542) and 6.432 (95% CI: 6.318, 6.545), respectively. The estimated restricted mean log survival time corresponding to the two fixed regimes, ZDV+ddI and ZDV+zalcitabline, are 6.372 (95% CI: 6.276, 6.467) and 6.373 (95% CI: 6.272, 6.474), respectively. These results are consistent with Kaplan-Meier curves, suggesting the estimated regimes from AIPWE_AFT* and AIPWE_AFT are very close and seem to be better than the two fixed regimes for this application. However, given the small sample size and large variability in estimation, no conclusive inferences can be made and the utilities of the estimated optimal treatment regimes should be studied and confirmed by future independent studies.

We conducted 200 times 10-fold cross-validation, where each time we randomly split the data into 10 subsamples with one being the test data set and the corresponding remaining 9 subsamples being the traning data set. The optimal treatment regime was estimated using the various methods on each training data set and the restricted mean lifetime under each of the estimated regimes was estimated nonparametrically on the corresponding test data set. The averages and standard deviations (in parenthesis) of the estimated restricted mean lifetime under various methods on test data sets are reported. Specifically, the average estimated restricted mean log survival time corresponding to the methods of AIPWE_AFT*, AIPWE_AFT, "Geng et

al" and "Tian et al" are 6.388(0.037), 6.412(0.024), 6.355(0.021) and 6.370(0.019) respectively. The estimated restricted mean log survival time corresponding to the two fixed regimes, ZDV+ddI and ZDV+zalcitabline, are 6.367(0.015) and 6.366(0.017), respectively. Results from cross-validation are consistent with results obtained on the entire data set.

6 | DISCUSSION

In this article, we have proposed new and robust methods for learning the optimal treatment regime for studies with a time-to-event outcome. The proposed methods minimize a (weighted or unweighted) empirical loss function involving the true treatment contrast function and augmented inverse probability weighted estimates of the contrast function. In constructing the AIPWE for the contrast function and, in particular, for the augmentation terms, one needs to build working models for the conditional expectation of survival times given the treatment and covariates. For data obtained from a randomized trial. the proposed methods enjoy good robustness property in the sense that the proposed methods do not require the working models for outcomes to be correctly specified. This robustness is achieved by taking advantage of the design and the robustness property of AIPWE. In randomized trials, the treatment assignment is independent of baseline covariates by design, which is the key that leads to the consistency of $\widehat{\gamma}$, the minimizer of the inverse censoring weighted loss function, as shown in Result 1. When the loss function is not inverse probability censoring weighted, the minimizer $\widetilde{\gamma}$ is consistent under the correct modeling of the propensity score for uncensored patients. Even though the methods are robust against misspecification of the outcome model, it is still important to model the relationship of outcomes with covariates and treatment well as it is essential for improving efficiency and performances in finite samples. We have proposed to adapt the well-known accelerated failure time model to our framework and in particular the accelerated failure time model with a smoothed error term. The reason is that an AFT model directly models the conditional expectation as opposed to the hazard function in the more popular Cox model. Therefore, it can be more easily incorporated into the augmentation terms. Our empirical studies have shown that overall the proposed methods are comparable or advantageous over existing methods. In general, the proposed method with an unweighted loss function exhibits the most robust performance across different scenarios.

The two proposed loss functions differ in whether inverse probability censoring weighting is used to account for censoring. Inverse probability censoring weighting is a natural strategy to account for censoring. In fact, it is necessary for consistent estimation of γ without imposing additional assumptions as shown in the Appendix. However, our ultimate goal is not to make inference on γ. The goal is to estimate the optimal treatment regime such that one can make correct treatment decisions for individual patients. Therefore, it is more important to consider the bias-and-variance trade-off in our setting than a typical inference problem. As we have discussed in Section 3.5 and Section 4, it is actually more advantageous to use the unweighted loss function based on their empirical performances. An intuitive explanation is that weighting greatly increases variance of estimation and therefore decreases the accuracy of the estimated optimal treatment decisions. We note that the discussion on accounting for censoring only pertains to the loss function and censoring is appropriately accounted for in both proposed methods in the stage of modeling for outcomes.

The proposed methods can handle high-dimensional covariates by incorporating variable selections at two stages. Variable selection can be carried out during the stage when one builds working AFT models for the time-to-events. In this stage, we would like to select all covariates that are predictive of outcomes to improve efficiency, i.e., covariates with main effects and/or treatment interaction effects. One also can incorporate variable selection during the minimization step by including a penalty term. In this step, the target of variable selection is on selecting variables that are important in the true optimal treatment regime, that is, variables interact with treatment. Based on our simulation studies, penalized methods do improve performances relative to the corresponding unpenalized methods when the number of covariates is large.

In this article, we have focused on the setting of a randomized clinical trial where the treatment probability is known and briefly discussed extension to observational studies. In principle, the proposed methods can be extended to observational studies by modeling the propensity score, i.e., the probability of treatment given covariates, and our simulation studies have illustrated the feasibility. However, when the estimated propensity scores vary considerably and may be close to zero or one, it may lead to numerical difficulty. This certainly warrants further more detailed study. Our study has focused on a learning method for learning the optimal decision rule from the observed data and we do not attempt to make statistical inferences. Making inferences is an important but challenging problem. Recent work mostly consider the setting without censoring[28,29,30]. It would be interesting to study inferences for optimal treatment regimes for survival data in the future.

DATA AVAILABILITY

The data that support the findings of this study are available from the corresponding author.

APPENDIX

We outline proofs of Result 1 in this Appendix. We consider a randomized clinical trial with randomization probability $P(A = 1) = \pi$. We consider the simpler case where π is known and used in constructing the contrast function. Proofs when $\widehat{\pi}$ is used is similar. We can rewrite equivalently

$$\begin{aligned}\widehat{C}(X_i) &= \frac{A_i - \pi}{\pi(1-\pi)}\left[\{A_i\pi - (1-A_i)(1-\pi)\}C(X_i) + \epsilon_i + e(X_i)\right] \\ &= C(X_i) + \frac{A_i - \pi}{\pi(1-\pi)}\epsilon_i + \frac{A_i - \pi}{\pi(1-\pi)}e(X_i) \\ &= \gamma_0' H_{X_i} + \frac{A_i - \pi}{\pi(1-\pi)}\epsilon_i + \frac{A_i - \pi}{\pi(1-\pi)}e(X_i).\end{aligned}$$

where $e(X) = (1-\pi)\{\mu_1(X) - \widehat{\mu}_1(X)\} + \pi\{\mu_0(X) - \widehat{\mu}_0(X)\}$. Recall $G(.)$ is the survival function of censoring time, and $\widehat{G}(.)$ is the Kaplan-Meier estimator of G. Let $L_i(\gamma) = w_i\{\widehat{C}(X_i) - \gamma H_{X_i}\}^2$, and $L_n(\gamma) = 1/n\sum_{i=1}^n L_i(\gamma)$.

The weighted loss function (5) corresponds to $w_i = \frac{\delta_i}{\widehat{G}(T_i)}$. We have

$$\begin{aligned}L_i^{\gamma}(\gamma, \widehat{u}_1, \widehat{u}_0) &\equiv \frac{\partial L_i(\gamma)}{\partial\gamma} \\ &= \frac{\delta_i}{\widehat{G}(T_i)}2\{\widehat{C}(X_i) - \gamma' H_{X_i}\}H_{X_i} \\ &= \frac{\delta_i}{\widehat{G}(T_i)}2\left\{\gamma_0' H_{X_i} + \frac{A_i - \pi}{\pi(1-\pi)}\epsilon_i + \frac{A_i - \pi}{\pi(1-\pi)}e(X_i) - \gamma' H_{X_i}\right\}H_{X_i},\end{aligned}$$

and

$$\begin{aligned}\frac{\partial L_n(\gamma)}{\partial\gamma} &= \frac{1}{n}\sum_{i=1}^n L_i^{\gamma}(\gamma, \widehat{u}_1, \widehat{u}_0) \\ &= \frac{1}{n}\sum_{i=1}^n \frac{\delta_i}{\widehat{G}(T_i)}2\left\{\gamma_0' H_{X_i} + \frac{A_i - \pi}{\pi(1-\pi)}\epsilon_i + \frac{A_i - \pi}{\pi(1-\pi)}e(X_i) - \gamma' H_{X_i}\right\}H_{X_i}.\end{aligned}$$

Suppose $\widehat{\mu}_1(X) \xrightarrow{P} \mu_1^*(X)$ and $\widehat{\mu}_0(X) \xrightarrow{P} \mu_0^*(X)$. Note, $\mu_1^*(X)$ and $\mu_0^*(X)$ may not necessarily be the truth. Let $e^*(X) = (1-\pi)\{\mu_1(X) - \mu_1^*(X)\} + \pi\{\mu_0(X) - \mu_0^*(X)\}$. Since $\widehat{G}(.)$ is a consistent estimator of $G(.)$, we can show that

$$\begin{aligned}&\frac{1}{n}\sum_{i=1}^n \frac{\delta_i}{\widehat{G}(T_i)}2\left\{\gamma_0' H_{X_i} + \frac{A-\pi}{\pi(1-\pi)}\epsilon + \frac{A-\pi}{\pi(1-\pi)}e(X_i) - \gamma' H_{X_i}\right\}H_{X_i} \\ &\xrightarrow{P} E\left[\frac{\delta}{G(T)}2\left\{\gamma_0' H_X + \frac{A-\pi}{\pi(1-\pi)}\epsilon + \frac{A-\pi}{\pi(1-\pi)}e^*(X) - \gamma' H_X\right\}H_X\right] \\ &= E\left[E\left\{\frac{\delta}{G(T)}|X, A, \varepsilon\right\}2\left\{\gamma_0' H_X + \frac{A-\pi}{\pi(1-\pi)}\epsilon + \frac{A-\pi}{\pi(1-\pi)}e^*(X) - \gamma' H_X\right\}H_X\right] \\ &= E\left[2\left\{\gamma_0' H_X + \frac{A-\pi}{\pi(1-\pi)}\epsilon + \frac{A-\pi}{\pi(1-\pi)}e^*(X) - \gamma' H_X\right\}H_X\right] \\ &= 2E\{(\gamma_0' H_X - \gamma' H_X)H_X\}.\end{aligned}$$

Therefore, we have

$$\frac{1}{n}\sum_{i=1}^n L_i^{\gamma}(\gamma, \widehat{u}_1, \widehat{u}_0) \xrightarrow{P} 2E\{(\gamma_0' H_X - \gamma' H_X)H_X\}.$$

As $\widehat{\gamma}$ is the unique solution to $\sum_{i=1}^{n} L_i^{\gamma}(\gamma, \widehat{u}_1, \widehat{u}_0) = 0$, and γ_0 is the unique solution to $E\{(\gamma_0' H_X - \gamma' H_X) H_X\} = 0$, this implies that $\widehat{\gamma} \xrightarrow{P} \gamma_0$ by M-estimation/Z-estimation theory.

The unweighted loss function (6) corresponds to $w_i = \delta_i$. Similarly, we have

$$L_i^{\gamma}(\gamma, \widehat{u}_1, \widehat{u}_0) = 2\delta_i \Big\{ \gamma_0' H_{X_i} + \frac{A_i - \widehat{\pi}_i^{\delta}}{\widehat{\pi}_i^{\delta}(1-\widehat{\pi}_i^{\delta})} \epsilon_i + \frac{A_i - \widehat{\pi}_i^{\delta}}{\widehat{\pi}_i^{\delta}(1-\widehat{\pi}_i^{\delta})} e(X_i) - \gamma' H_{X_i} \Big\} H_{X_i},$$

and

$$\begin{aligned} \frac{\partial L_n(\gamma)}{\partial \gamma} &= \frac{1}{n} \sum_{i=1}^{n} L_i^{\gamma}(\gamma, \widehat{u}_1, \widehat{u}_0) \\ &= \frac{1}{n} \sum_{i=1}^{n} 2\delta_i \Big\{ \gamma_0' H_{X_i} + \frac{A_i - \widehat{\pi}_i^{\delta}}{\widehat{\pi}_i^{\delta}(1-\widehat{\pi}_i^{\delta})} \epsilon_i + \frac{A_i - \widehat{\pi}_i^{\delta}}{\widehat{\pi}_i^{\delta}(1-\widehat{\pi}_i^{\delta})} e(X_i) - \gamma' H_{X_i} \Big\} H_{X_i}. \end{aligned}$$

Recall $\widehat{\pi}_i^{\delta} \equiv \widehat{\pi}^{\delta}(X_i)$. Suppose $\widehat{\pi}^{\delta}(x) \xrightarrow{P} \pi^{\delta *}(x)$, we can show that

$$\begin{aligned} &\frac{1}{n} \sum_{i=1}^{n} 2\delta_i \Big\{ \gamma_0' H_{X_i} + \frac{A_i - \widehat{\pi}_i^{\delta}}{\widehat{\pi}_i^{\delta}(1-\widehat{\pi}_i^{\delta})} \epsilon_i + \frac{A_i - \widehat{\pi}_i^{\delta}}{\widehat{\pi}_i^{\delta}(1-\widehat{\pi}_i^{\delta})} e(X_i) - \gamma' H_{X_i} \Big\} H_{X_i} \\ &\xrightarrow{P} E\Big[2\delta \Big\{ \gamma_0' H_X + \frac{A - \pi^{\delta *}(X)}{\pi^{\delta *}(X)\{1-\pi^{\delta *}(X)\}} \epsilon + \frac{A - \pi^{\delta *}(X)}{\pi^{\delta *}(X)\{1-\pi^{\delta *}(X)\}} e^*(X) - \gamma' H_X \Big\} H_X \Big] \\ &= 2P(\delta = 1) E\{(\gamma_0' H_X - \gamma' H_X) H_X | \delta = 1\} \\ &+ 2P(\delta = 1) E\Big\{ \frac{A - \pi^{\delta *}(X)}{\pi^{\delta *}(X)\{1-\pi^{\delta *}(X)\}} e^*(X) H_X | \delta = 1 \Big\}. \end{aligned}$$

Then we have

$$\begin{aligned} \frac{1}{n} \sum_{i=1}^{n} L_i^{\gamma}(\gamma, \widehat{u}_1, \widehat{u}_0) \xrightarrow{P} & \; 2P(\delta = 1) E\{(\gamma_0' H_X - \gamma' H_X) H_X | \delta = 1\} \\ & + \; 2P(\delta = 1) E\Big\{ \frac{A - \pi^{\delta *}(X)}{\pi^{\delta *}(X)\{1-\pi^{\delta *}(X)\}} e^*(X) H_X | \delta = 1 \Big\}. \end{aligned}$$

We see that γ_0 is still the unique solution to $2P(\delta = 1) E\{(\gamma_0' H_X - \gamma' H_X) H_X | \delta = 1\} = 0$. For the second term, we have

$$\begin{aligned} & E\Big\{ \frac{A - \pi^{\delta *}(X)}{\pi^{\delta *}(X)\{1-\pi^{\delta *}(X)\}} e^*(X) H_X | \delta = 1 \Big\} \\ &= E\Big[E\Big\{ \frac{A - \pi^{\delta *}(X)}{\pi^{\delta *}(X)\{1-\pi^{\delta *}(X)\}} e^*(X) H_X | X, \delta = 1 \Big\} | \delta = 1 \Big] \\ &= E\Big\{ \frac{\pi^{\delta}(X) - \pi^{\delta *}(X)}{\pi^{\delta *}(X)\{1-\pi^{\delta *}(X)\}} e^*(X) H_X | \delta = 1 \Big\}, \end{aligned}$$

where $\pi^{\delta}(X) = P(A = 1 | X, \delta = 1)$. If the propensity score model for the uncensored data is correct, then we have $\pi^{\delta *}(X) = \pi^{\delta}(X)$, and the second term is equal 0. By the same argument, we have $\widetilde{\gamma}$ converges in probability to γ_0.

When data are obtained from an observational study, the propensity score $\pi(X) = P(A = 1 | X)$ is unknown and can be estimated by $\widehat{\pi}(X)$. Using a similar argument as in proofs for the unweighted loss function , we can show that $\widehat{\gamma}$ is consistent for γ_0 if the model for $P(A = 1 | X)$ is correct.

References

1. Davison A. Treatment effect heterogeneity in paired data. *Biometrika* 1992; 79(3): 463–474.

2. Gail M, Simon R. Testing for qualitative interactions between treatment effects and patient subsets. *Biometrics* 1985: 361–372.

3. Lagakos SW, others . The challenge of subgroup analyses-reporting without distorting. *New England Journal of Medicine* 2006; 354(16): 1667.

4. Imai K, Ratkovic M, others . Estimating treatment effect heterogeneity in randomized program evaluation. *The Annals of Applied Statistics* 2013; 7(1): 443–470.

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

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

7. Moodie EE, Platt RW, Kramer MS. Estimating response-maximized decision rules with applications to breastfeeding. *Journal of the American Statistical Association* 2009; 104(485): 155–165.

8. Cai T, Tian L, Wong PH, Wei L. Analysis of randomized comparative clinical trial data for personalized treatment selections. *Biostatistics* 2011; 12(2): 270–282.

9. Schulte PJ, Tsiatis AA, Laber EB, Davidian M. Q-and A-learning methods for estimating optimal dynamic treatment regimes. *Statistical science: a review journal of the Institute of Mathematical Statistics* 2014; 29(4): 640.

10. Tsiatis AA. *Dynamic Treatment Regimes: Statistical Methods for Precision Medicine*. CRC Press . 2019.

11. Tian L, Alizadeh AA, Gentles AJ, Tibshirani R. A simple method for estimating interactions between a treatment and a large number of covariates. *Journal of the American Statistical Association* 2014; 109(508): 1517–1532.

12. Lu W, Zhang HH, Zeng D. Variable selection for optimal treatment decision. *Statistical methods in medical research* 2013; 22(5): 493–504.

13. Geng Y, Zhang HH, Lu W. On optimal treatment regimes selection for mean survival time. *Statistics in medicine* 2015; 34(7): 1169–1184.

14. Goldberg Y, Kosorok MR. Q-learning with censored data. *Annals of statistics* 2012; 40(1): 529.

15. Bai X, Tsiatis AA, Lu W, Song R. Optimal treatment regimes for survival endpoints using a locally-efficient doubly-robust estimator from a classification perspective. *Lifetime data analysis* 2017; 23(4): 585–604.

16. Jiang R, Lu W, Song R, Davidian M. On estimation of optimal treatment regimes for maximizing t-year survival probability. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 2017; 79(4): 1165–1185.

17. Hager R, Tsiatis AA, Davidian M. Optimal two-stage dynamic treatment regimes from a classification perspective with censored survival data. *Biometrics* 2018; 74(4): 1180–1192.

18. Simoneau G, Moodie EE, Nijjar JS, Platt RW, Investigators SERAIC, others . Estimating optimal dynamic treatment regimes with survival outcomes. *Journal of the American Statistical Association* 2020; 115(531): 1531–1539.

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

20. Zhang B, Tsiatis AA, Davidian M, Zhang M, Laber E. Estimating optimal treatment regimes from a classification perspective. *Stat* 2012; 1(1): 103–114.

21. Robins JM. Optimal structural nested models for optimal sequential decisions. In: Springer. ; 2004: 189–326.

22. Komárek A, Lesaffre E, Hilton JF. Accelerated failure time model for arbitrarily censored data with smoothed error distribution. *Journal of Computational and Graphical Statistics* 2005; 14(3): 726–745.

23. Komárek A, Komárek MA. Package smoothSurv. 2020.

24. Khan MHR, Shaw JEH. Variable selection for survival data with a class of adaptive elastic net techniques. *Statistics and Computing* 2016; 26(3): 725–741.

25. Stefanski LA, Boos DD. The calculus of M-estimation. *The American Statistician* 2002; 56(1): 29–38.

26. Zhang Y, Laber EB, Davidian M, Tsiatis AA. Interpretable dynamic treatment regimes. *Journal of the American Statistical Association* 2018; 113(524): 1541–1549.

27. Zhang M, Tsiatis AA, Davidian M. Improving efficiency of inferences in randomized clinical trials using auxiliary covariates. *Biometrics* 2008; 64(3): 707–715.

28. Luedtke AR, Van Der Laan MJ. Statistical inference for the mean outcome under a possibly non-unique optimal treatment strategy. *Annals of statistics* 2016; 44(2): 713.

29. Shi C, Lu W, Song R. Breaking the Curse of Nonregularity with Subagging—Inference of the Mean Outcome under Optimal Treatment Regimes. *Journal of Machine Learning Research* 2020; 21(176): 1–67.

30. Wager S, Athey S. Estimation and inference of heterogeneous treatment effects using random forests. *Journal of the American Statistical Association* 2018; 113(523): 1228–1242.

TABLE 1 Simulation results under Models 1-3 and independent censoring. PCD: the percentage of correct decisions; MSE: mean squared error for estimating γ for penalized methods; CR: censoring rate; Value: expectation of log survival time under the estimated optimal treatment regime; V_0 is the value of the true optimal treatment regime.

	Model 1 V_0=2.480				Model 2 V_0=1.877				Model 3 V_0=1.667			
			PCD				PCD				PCD	
Method	MSE	Value	Unpenalized	Penalized	MSE	Value	Unpenalized	Penalized	MSE	Value	Unpenalized	Penalized
					n=400, CR=15%							
AIPWE_AFT*	0.427	2.449	0.850	0.903	0.857	1.777	0.760	0.828	1.487	1.583	0.738	0.815
AIPWE_AFT	0.516	2.442	0.839	0.894	1.004	1.759	0.750	0.812	1.498	1.585	0.738	0.817
Geng et al	0.594	2.436	0.836	0.887	1.186	1.735	0.745	0.796	1.649	1.568	0.734	0.795
Tian et al	1.436	2.403	0.781	0.850	1.791	1.700	0.712	0.774	2.037	1.595	0.730	0.832
					n=400, CR=40%							
AIPWE_AFT*	0.643	2.438	0.829	0.883	1.098	1.737	0.738	0.803	1.675	1.551	0.711	0.783
AIPWE_AFT	0.866	2.420	0.814	0.864	1.290	1.709	0.729	0.782	1.782	1.542	0.707	0.775
Geng et al	1.425	2.368	0.782	0.816	2.224	1.595	0.684	0.709	2.604	1.462	0.667	0.691
Tian et al	1.435	2.398	0.766	0.842	1.679	1.687	0.710	0.770	2.040	1.570	0.715	0.804
					n=1000, CR=15%							
AIPWE_AFT*	0.162	2.475	0.908	0.943	0.282	1.853	0.841	0.911	1.132	1.636	0.822	0.894
AIPWE_AFT	0.182	2.473	0.898	0.938	0.359	1.842	0.825	0.893	1.166	1.636	0.823	0.895
Geng et al	0.191	2.473	0.898	0.939	0.396	1.831	0.824	0.887	1.250	1.635	0.821	0.892
Tian et al	1.110	2.443	0.841	0.889	1.425	1.816	0.787	0.868	1.772	1.630	0.797	0.886
					n=1000, CR=40%							
AIPWE_AFT*	0.259	2.469	0.894	0.939	0.472	1.834	0.813	0.891	1.135	1.626	0.795	0.877
AIPWE_AFT	0.465	2.455	0.872	0.913	0.567	1.821	0.804	0.871	1.327	1.616	0.790	0.862
Geng et al	0.610	2.447	0.862	0.902	0.936	1.759	0.770	0.820	1.578	1.595	0.771	0.831
Tian et al	1.007	2.437	0.832	0.886	1.247	1.810	0.787	0.864	1.687	1.623	0.786	0.871

TABLE 2 Selection results under independent censoring. Incorr0: the number of non-zero coefficients incorrectly identified as zero; Corr0: the number of correct zero coefficients identified; Cover: the proportion of covering all the important variables.

	Model 1			Model 2			Model 3		
Method	Incorr0	Corr0	Cover	Incorr0	Corr0	Cover	Incorr0	Corr0	Cover
				n=400, CR=15%					
AIPWE_AFT*	0.002	28.0	0.998	0.025	28.3	0.975	0.085	30.9	0.885
AIPWE_AFT	0.010	27.6	0.992	0.065	29.5	0.950	0.095	31.4	0.905
Geng et al	0.002	26.3	0.998	0.260	26.3	0.750	0.064	28.7	0.938
Tian et al	0.322	39.4	0.718	0.846	38.9	0.462	0.208	41.2	0.820
				n=400, CR=40%					
AIPWE_AFT*	0.010	26.7	0.990	0.080	29.8	0.935	0.215	31.6	0.785
AIPWE_AFT	0.050	28.4	0.955	0.185	30.0	0.870	0.305	31.8	0.796
Geng et al	0.024	16.3	0.976	0.606	18.3	0.524	0.296	18.0	0.772
Tian et al	0.316	38.9	0.724	0.906	38.6	0.378	0.324	40.3	0.726
				n=1000, CR=15%					
AIPWE_AFT*	0	33.7	1	0	35.6	1	0	36.6	1
AIPWE_AFT	0	34.6	1	0	34.5	0.998	0	36.5	1
Geng et al	0	32.5	1	0.042	31.8	0.958	0	34.5	1
Tian et al	0.012	37.3	0.988	0.080	37.1	0.922	0.004	39.7	0.996
				n=1000, CR=40%					
AIPWE_AFT*	0	33.6	1	0	34.2	1	0.020	37.4	0.980
AIPWE_AFT	0.005	32.6	0.995	0.032	34.0	0.982	0.058	36.4	0.954
Geng et al	0.004	26.6	0.998	0.590	28.8	0.638	0.068	30.9	0.936
Tian et al	0.020	36.7	0.980	0.214	36.4	0.790	0.018	38.7	0.982

TABLE 3 Sensitivity analysis: Simulation results under Models 1-3 and dependent censoring. PCD: the percentage of correct decisions; MSE: mean squared error for estimating γ for penalized methods; CR: censoring rate; Value: expectation of log survival time under the estimated optimal treatment regime; V_0 is the value of the true optimal treatment regime.

	Model 1 V_0=2.480				Model 2 V_0=1.877				Model 3 V_0=1.667			
			PCD				PCD				PCD	
Method	MSE	Value	Unpenalized	Penalized	MSE	Value	Unpenalized	Penalized	MSE	Value	Unpenalized	Penalized
					n=400, CR=15%							
AIPWE_AFT*	0.474	2.444	0.850	0.900	0.925	1.759	0.749	0.814	1.540	1.575	0.734	0.808
AIPWE_AFT	0.595	2.433	0.838	0.887	1.173	1.727	0.736	0.789	1.547	1.576	0.734	0.809
Geng et al	0.676	2.426	0.834	0.879	1.358	1.699	0.727	0.771	1.710	1.564	0.728	0.794
Tian et al	1.409	2.397	0.781	0.849	1.750	1.707	0.712	0.779	1.998	1.591	0.731	0.829
					n=400, CR=40%							
AIPWE_AFT*	0.658	2.423	0.827	0.879	1.201	1.730	0.726	0.794	1.757	1.538	0.699	0.771
AIPWE_AFT	0.951	2.400	0.810	0.855	1.590	1.670	0.710	0.752	1.819	1.536	0.698	0.770
Geng et al	1.393	2.365	0.789	0.822	2.559	1.580	0.673	0.697	2.267	1.486	0.676	0.718
Tian et al	1.521	2.373	0.761	0.829	1.620	1.703	0.711	0.778	2.119	1.562	0.703	0.799
					n=1000, CR=15%							
AIPWE_AFT*	0.226	2.468	0.909	0.940	0.373	1.849	0.834	0.904	1.206	1.628	0.816	0.881
AIPWE_AFT	0.279	2.461	0.893	0.926	0.527	1.824	0.810	0.871	1.241	1.627	0.817	0.879
Geng et al	0.294	2.463	0.895	0.930	0.597	1.808	0.808	0.860	1.354	1.630	0.814	0.885
Tian et al	1.057	2.434	0.841	0.884	1.379	1.819	0.790	0.870	1.773	1.622	0.795	0.873
					n=1000, CR=40%							
AIPWE_AFT*	0.256	2.466	0.893	0.934	0.501	1.831	0.804	0.883	1.209	1.613	0.786	0.860
AIPWE_AFT	0.489	2.445	0.865	0.901	0.812	1.786	0.773	0.835	1.384	1.601	0.786	0.844
Geng et al	0.526	2.448	0.865	0.905	1.074	1.739	0.753	0.800	1.570	1.604	0.777	0.847
Tian et al	1.110	2.417	0.823	0.865	1.278	1.800	0.778	0.854	1.791	1.605	0.772	0.851

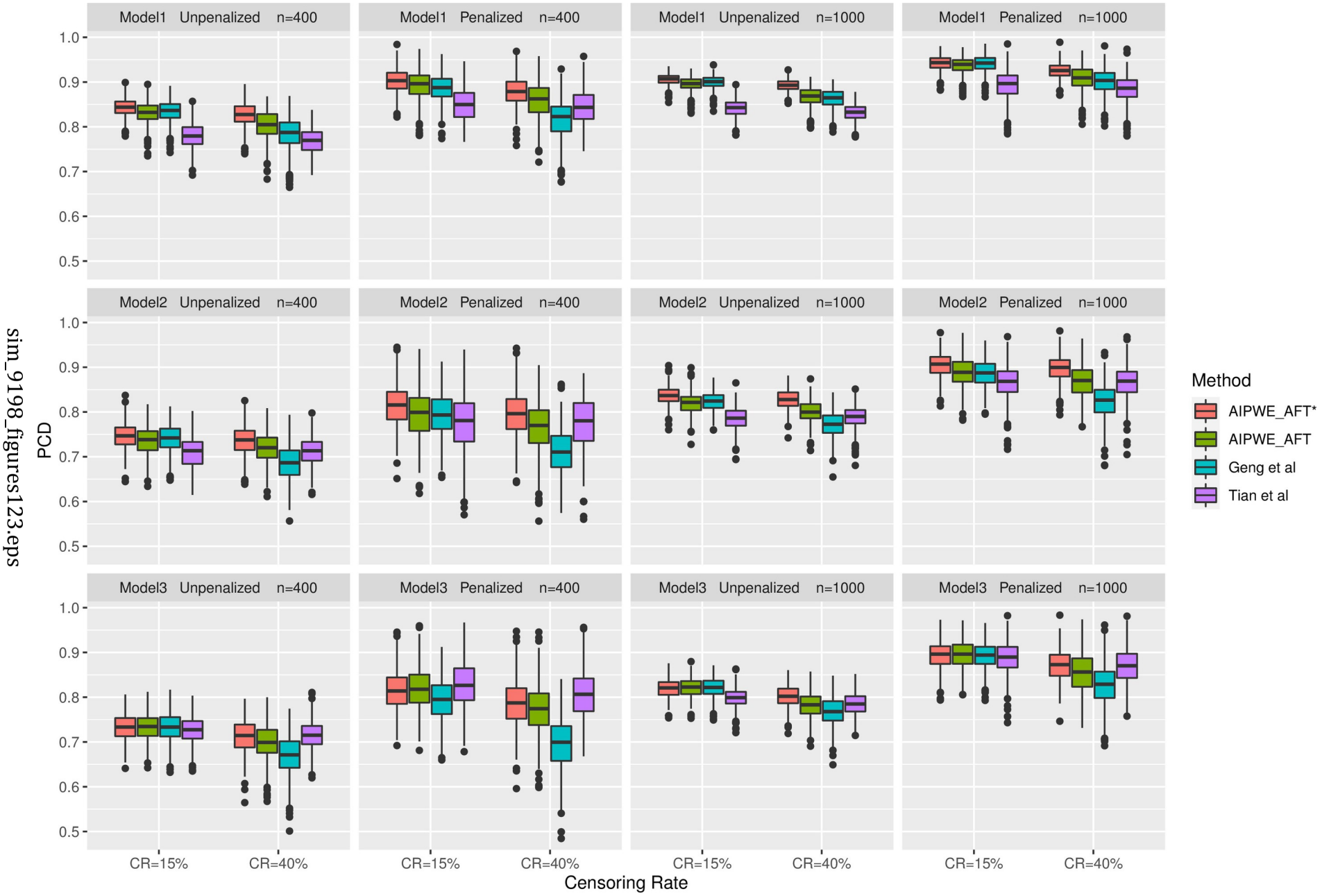
TABLE 4 Selection results under dependent censoring (Sensitivity analysis). Incorr0: the number of non-zero coefficients incorrectly identified as zero; Corr0: the number of correct zero coefficients identified; Cover: the proportion of covering all the important variables.

	Model 1			Model 2			Model 3		
Method	Incorr0	Corr0	Cover	Incorr0	Corr0	Cover	Incorr0	Corr0	Cover
				n=400, CR=15%					
AIPWE_AFT*	0.002	29.8	0.998	0.030	30.2	0.970	0.125	32.4	0.905
AIPWE_AFT	0.020	28.6	0.985	0.150	30.6	0.870	0.090	31.2	0.920
Geng et al	0.006	25.6	0.994	0.422	26.1	0.640	0.092	28.4	0.920
Tian et al	0.236	39.4	0.788	0.786	38.7	0.482	0.306	41.4	0.754
				n=400, CR=40%					
AIPWE_AFT*	0.010	26.8	0.992	0.075	30.0	0.930	0.300	31.0	0.740
AIPWE_AFT	0.090	28.0	0.915	0.460	29.9	0.720	0.360	31.2	0.715
Geng et al	0.022	16.0	0.978	0.634	16.8	0.546	0.256	21.2	0.770
Tian et al	0.464	40.0	0.622	1.074	38.8	0.244	0.564	41.6	0.566
				n=1000, CR=15%					
AIPWE_AFT*	0	33.1	1	0	34.5	1	0	35.9	1
AIPWE_AFT	0.016	32.0	1	0.010	34.4	0.990	0	35.6	1
Geng et al	0.002	31.4	0.998	0.178	31.8	0.824	0	34.0	1
Tian et al	0.010	37.1	0.990	0.094	37.1	0.910	0.008	39.9	0.992
				n=1000, CR=40%					
AIPWE_AFT*	0	32.7	1	0.010	36.0	0.995	0.010	36.2	0.970
AIPWE_AFT	0	31.5	1	0.085	34.7	0.925	0.035	35.4	0.965
Geng et al	0	27.5	1	0.596	26.9	0.602	0.012	32.9	0.988
Tian et al	0.034	36.7	0.966	0.510	37.4	0.492	0.060	39.7	0.942

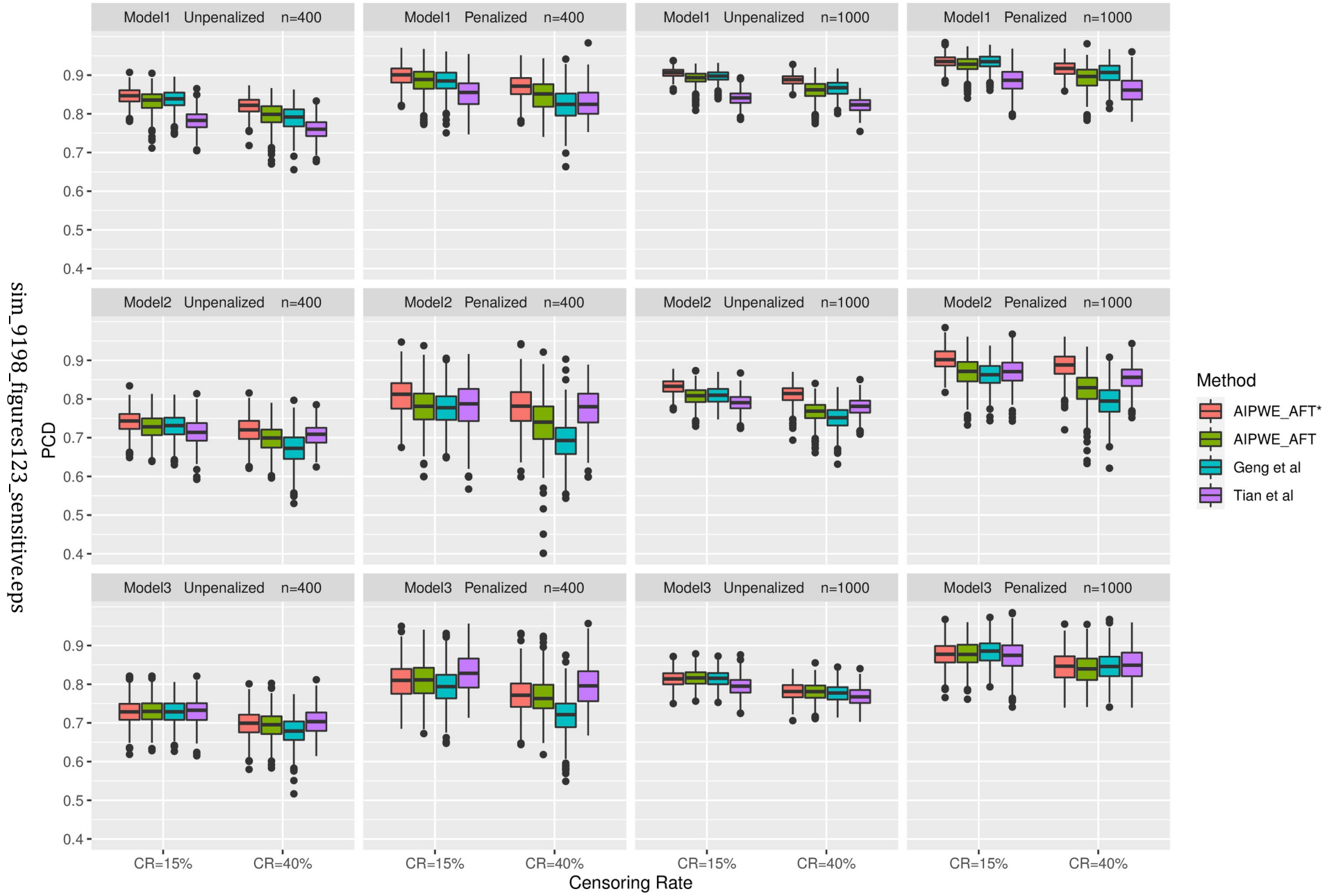
Figure 1: Boxplots of PCD for various methods under simulation models 1-3, independent censoring with sample size n = 400 and n = 1000 respectively.

Figure 2: Sensitivity analysis: boxplots of PCD for various methods under simulation models 1-3, dependent censoring with sample size n = 400 and n = 1000 respectively.

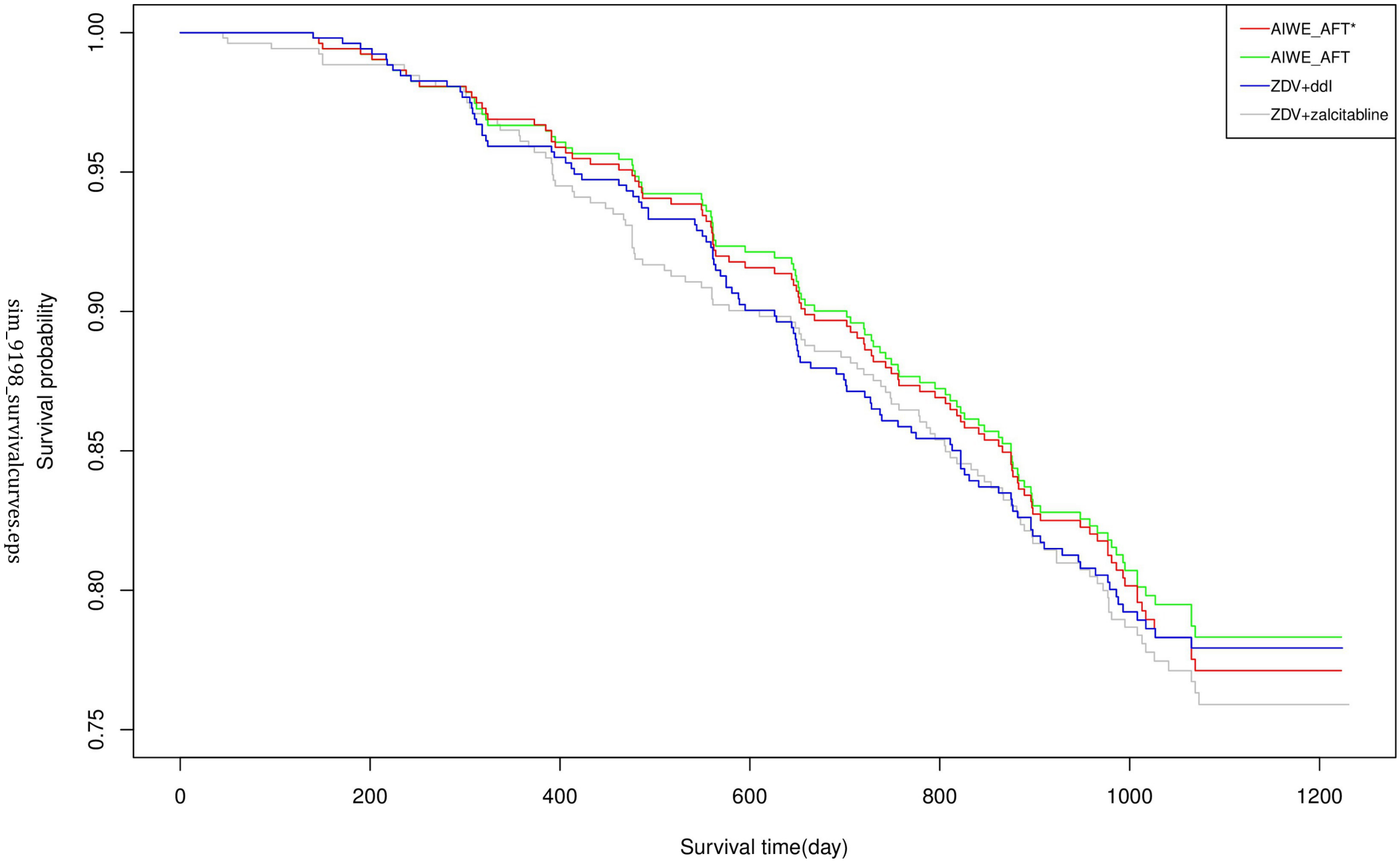
Figure 3: Kaplan-Meier survival curves for two treatment groups and patients following the estimated optimal treatment regime from various methods.



sim_9198_figures123.eps



sim_9198_figures123_sensitive.eps



sim_9198_survivalcurves.eps