

# Robustifying trial-derived optimal treatment rules for a target population

Ying-Qi Zhao\*

*Public Health Sciences Division, Fred Hutchinson Cancer Research Center,  
Seattle, WA, 98109*

*e-mail: [yqzhao@fredhutch.org](mailto:yqzhao@fredhutch.org)*

Donglin Zeng

*Department of Biostatistics, University of North Carolina at Chapel Hill,  
Chapel Hill, NC, 27599*

*e-mail: [dzeng@email.unc.edu](mailto:dzeng@email.unc.edu)*

Catherine M. Tangen and Michael L. Leblanc

*Public Health Sciences Division, Fred Hutchinson Cancer Research Center,  
Seattle, WA, 98109*

*e-mail: [ctangen@fredhutch.org](mailto:ctangen@fredhutch.org); [mleblanc@fredhutch.org](mailto:mleblanc@fredhutch.org)*

**Abstract:** Treatment rules based on individual patient characteristics that are easy to interpret and disseminate are important in clinical practice. Properly planned and conducted randomized clinical trials are used to construct individualized treatment rules. However, it is often a concern that trial participants lack representativeness, so it limits the applicability of the derived rules to a target population. In this work, we use data from a single trial study to propose a two-stage procedure to derive a robust and parsimonious rule to maximize the benefit in the target population. The procedure allows a wide range of possible covariate distributions in the target population, with minimal assumptions on the first two moments of the covariate distribution. The practical utility and favorable performance of the methodology are demonstrated using extensive simulations and a real data application.

**Keywords and phrases:** Primary Minimax linear decision, biased sample, classification, individualized treatment rules, personalized medicine.

Received January 2018.

## Contents

1	Introduction . . . . .	1718
2	Methodology . . . . .	1719
2.1	Background and the optimal treatment rule . . . . .	1719
2.2	A General Quality Value of Treatment Rules for the Target Population . . . . .	1721
2.3	Learning Robust Linear Rules for the Target Population . . . . .	1721

---

\*Corresponding author

2.4	Estimating robust rules using empirical data . . . . .	1724
2.5	Choices of $(\nu, \rho_j)$ . . . . .	1726
3	Theoretical results . . . . .	1726
4	Simulation Studies . . . . .	1727
5	Data Analysis . . . . .	1730
6	Discussion . . . . .	1734
A	Appendix . . . . .	1735
	Acknowledgments . . . . .	1740
	Supplementary Material . . . . .	1741
	References . . . . .	1741

## 1. Introduction

In the new era of personalized medicine, it has been advocated that treatments should be recommended according to individual patient characteristics to account for considerable heterogeneity among patients' responses to different treatments (Hayes et al., 2007; Hamburg and Collins, 2010). Randomized clinical trials (RCTs) are ideal for constructing such rules, since they provide internal validity by ensuring consistency, positivity and no unmeasured confounders (Greenland, 1990; Hernán and Robins, 2006) that may be violated in observational studies.

Sophisticated statistical methods have been developed to estimate optimal individualized treatment rules using data from randomized trials. Regression-based methods estimate outcome as a function of patient covariates and treatment, and then select the treatment that maximizes the predicted outcome for each individual (Brinkley et al., 2010; Qian and Murphy, 2011; Kang et al., 2014). Some recent developments directly search for the individualized treatment rules that maximize the benefit for future patients (Zhang et al., 2012; Zhao et al., 2012). When the primary outcome of interest is survival time subject to right censoring, some methods have been proposed in this regard using either regression-based methods (Goldberg and Kosorok, 2012; Huang et al., 2014) or direct search methods (Zhao et al., 2015).

The derived optimal rules sometimes are complex and nonlinear, so they are highly variable and may not be practically useful. To better inform clinical practice, it is more desirable that a recommended treatment rule be easy to interpret and disseminate. As suggested in Orellana et al. (2010), the class of practically enforceable candidate regimes is significantly smaller than the class of arbitrary functions of covariates. Usually these regimes are comprised of functions that only depend on a small set of covariates, and are indexed by a set of finite dimensional parameters. Statistically, Qian and Murphy (2011) used a rich linear basis for better modeling the outcome, with a sparsity penalty imposed to preserve the parsimoniousness of the resulting rule, while most works considered and recommended rules from a linear functional class.

Unfortunately, it has been well known that the RCT may have limited generalizability when the distribution of treatment effect modifiers in trial partic-

ipants differs from the one in a target population (Buchanan et al., 2016). A parsimonious treatment rule constructed from trial data using current methods cannot be directly carried over to a target population, when the population characteristics are different between the two. Some attempts have been made to adjust population difference using inverse probability-of-selection weight method (Cole and Stuart, 2010). This, however, requires complete knowledge of population selection mechanism, and it is yet to be extended to the context of developing optimal treatment rules.

In this paper, we aim to robustify the treatment rule from a trial so that the robustified rule is (1) parsimonious (linear in patient features in particular); (2) still maximizes a general benefit criterion when applied to the target population. We provide a new framework, called minimax linear decision (MiLD), to robustify the treatment rule. MiLD enables the construction of a linear rule that optimizes the general benefit function in the target population, allowing differences between the target and the trial populations in the means and covariances in treatment effect modifiers. It can be further extended to construct nonlinear rules using the ‘kernel trick’, which avoids the explicit feature mapping but only relies on a kernel function to learn a nonlinear decision boundary. Regardless, MiLD requires no further assumptions beyond the first two moments of the covariate distributions. Moreover, our developed framework is applicable to all types of outcomes.

The remainder of the paper is organized as follows. In Section 2, we formulate the problem of finding a robust and parsimonious treatment rule using data from a clinical trial. The proposed method, MiLD, is then developed to optimize a general benefit function in the target population, allowing the covariate distribution to be different between the two populations. Consistency and convergence rate results are established for the proposed method. We present simulation studies to evaluate performances of the proposed method in Section 4. We further illustrate the method using a data example from a US NIH funded SWOG trial on castration-resistant prostate cancer patients in Section 5. The proofs of theoretical results are given in the Appendix.

## 2. Methodology

### 2.1. Background and the optimal treatment rule

Let  $T$  denote the outcome of interest,  $A \in \{-1, 1\}$  denote a binary treatment, and  $X = (X_1, \dots, X_m)^\top \in \mathbb{R}^m$  denote patient covariates. We assume the dimension is small to moderate, and fixed. The outcome could be a continuous, discrete, or time-to-event outcome subject to censoring. A treatment regime  $d \in \mathcal{D}$  is a function mapping from the space of  $X$  to the space of treatments. A future patient with  $X = x$  is then assigned with treatment  $d(x)$ . Let  $T(a)$  denote the potential outcome under treatment  $a \in \{-1, 1\}$  (Rubin, 1978), and  $T(d) = T\{d(X)\}$  be the potential outcome under the rule  $d$  for the given covariate  $X$ . The quality of  $d$  can be measured by the marginal mean outcome  $E\{T(d)\}$ , which represents the trial population mean outcome were all patients

in the trial population to receive treatment according to  $d$ . We denote this quantity as  $\mathcal{V}(d)$ , which is also called the value function of  $d$ . The optimal treatment regime  $d^*$  maximizes  $\mathcal{V}(d)$  over all possible  $d$ . We assume that in the trial,

- (A1)  $P(A = a|X) > 0$  with probability one for  $a \in \{-1, 1\}$ ;
- (A2)  $\{T(-1), T(1)\}$  are independent of  $A$  conditional  $X$ ;
- (A3) consistency so that  $T(a) = \sum_a I(A = a)T$ .

Since  $\mathcal{V}(d) = E\{T(d)\} = E_X(E\{T\{d(X)\}|X\})$ , where the outer expectation  $E_X(\cdot)$  is taken with respect to the marginal distribution of  $X$  in the trial population, the optimal treatment for any patient with characteristics  $x$  is  $d^*(x) = \operatorname{argmax}_{a \in \{-1, 1\}} E\{T(a)|X = x\}$  when there is no restriction of the functional form of the rule. These assumptions are typically standard and satisfied in randomized clinical trials (Robins et al., 2000), because application of the intervention to any individual is under the control of the investigator. This is however, not guaranteed in observational studies. For example, consistency assumption can be violated given that there could be variations of treatment, and each of them may have a different causal effect on the outcome. In addition, noncompliance can be severe in observational studies, where some subjects may never take what they are asked to take. If (A1)-(A3) are satisfied, the optimal treatment rule is

$$d^*(x) = \operatorname{sign}\{f^*(x)\}, \text{ where } f^*(x) = E(T|X = x, A = 1) - E(T|X = x, A = -1).$$

To assess the overall benefit of the obtained rule when applied to the target population, we let  $p_X(x)$  denote the distribution of  $X$  in the trial population, and  $q_X(x)$  denote the distribution of  $X$  in the target population, which is not necessarily the same as  $p_X(x)$ . Furthermore, we assume

- (A4) the support of  $q_X(x)$  is contained in the support of  $p_X(x)$ , i.e., there is no under-coverage in the trial study;
- (A5) the potential outcome mean given  $X$  is the same between the trial population and the target population. Thus, the treatment works the same in both populations.

Thus, the only difference between the two populations is due to the covariate distributions. Let  $P^q$  denote the distribution of  $X$  in the target population and  $P^p$  denote the distribution of  $X$  in the trial population. Let  $\tilde{E}$  denote the expectation is taken with respect to the distribution of  $X$  in the target population. Under (A4) and (A5), it is easy to observe that the expected outcome of rule  $d(X)$  for the target population is

$$\begin{aligned} \tilde{\mathcal{V}}(d) &= \tilde{E}[T\{d(X)\}] = \int E[T\{d(X)\}|X] dP^q \\ &= \int E[T\{d(X)\}|X] \frac{dP^q}{dP^p} dP^p = E_X \left( E[T\{d(X)\}|X] \frac{q_X(X)}{p_X(X)} \right), \end{aligned}$$

where  $E_X[\cdot]$  denotes the expectation under  $X \sim p_X(x)$ . Consequently, the optimal rule obtained from the trial population,  $d^*$ , is also the optimal for the target population if there is not any restriction to the formulation in  $d^*$ .

## 2.2. A General Quality Value of Treatment Rules for the Target Population

We propose a general criterion assessing the quality of a decision rule for the target population, which includes both the value function and the correct allocation rate to the optimal rule as special cases. For a given rule  $d(X)$ , the proposed quality assessment in the target population is defined as

$$\tilde{\mathcal{B}}(d) = \tilde{E}[W(X)I\{d(X) = d^*(X)\}],$$

where  $W(X)$  is a non-negative function, essentially a reward if the treatment rule  $d(x)$  is the optimal.

The quality value in the definition has a different interpretation depending on the choice of  $W(x)$ . For example, let  $W(x) = W_1(x) = E\{T|A = d^*(x), X = x\} - E\{T|A \neq d^*(x), X = x\} = |f^*(x)|$ . That is, for subject with covariate  $X = x$ ,  $W(x)$  is the gain if he/she follows treatment rule  $d^*$ .  $\tilde{\mathcal{B}}(d)$  achieves optimal if  $d = d^*$ . This is because  $\tilde{\mathcal{B}}(d) = \tilde{\mathcal{V}}(d) + \tilde{E}\{T(-d^*)\}$ . Hence, maximizing  $\tilde{\mathcal{B}}(d)$  is equivalent to maximizing the value function in the target population. If we set  $W(X) = W_2(X) \equiv 1$ , then  $\tilde{\mathcal{B}}(d) = \tilde{P}\{d(X) = d^*(X)\}$ , and the criterion corresponds to the correct allocation rate of the optimal treatment. Additionally, in practice,  $W(X)$  could allow more general trade off between the benefit of optimal treatment and the relative cost of giving optimal treatment to patients instead of standard care. In this paper, we will focus on the choice of  $W = W_1$  and  $W_2$ .

## 2.3. Learning Robust Linear Rules for the Target Population

In many practices, since the set of enforceable treatment rules are usually restrictive and cannot be arbitrary, a parsimonious and interpretable decision rule is preferred. In particular, we focus on a linear decision rule, i.e.,  $d(x) = \text{sign}\{f(x)\}$  where  $f(x) = x^\top \beta_1 + \beta_0$ . A patient  $x$  is assigned to treatment 1 if  $x^\top \beta_1 + \beta_0 \geq 0$  and treatment  $-1$  otherwise. For example, let the potential outcome model be  $E[\log\{\tilde{T}(a)\}|X] = (2I[\{3X/4 + \sin(X)/4 - 2\}^2 - 1] - 1)a$ . The optimal rule  $d^*(x) = 1$  if  $3x/4 + \sin(x)/4 - 1 \leq 0$  or  $3x/4 + \sin(x)/4 - 3 > 0$ , and  $d^*(x) = -1$  otherwise. Consider the space of linear decision rules with  $\mathcal{D}_L = \{d(X) = \text{sign}(\beta_0 + \beta_1 X), \beta_0, \beta_1 \in \mathbb{R}\}$ . Therefore  $d^*(x)$  is highly nonlinear, and  $d^*(x) \notin \mathcal{D}_L$ . Assume that  $X \sim N(2, 1)$  and  $X \sim N(4, 1)$  in the trial and target population, respectively. Then the optimal linear rule is  $d_L^*(X) = \text{sign}(X - 0.9)$  using the trial data. However,  $\text{sign}(-X + 0.28)$  should be the optimal linear rule in the target population, which lead to a benefit of 0.997, versus a benefit of -0.387 using the optimal rule derived from the trial data. Hence, the optimal linear rules could be substantially different between the two populations.

We should note that unless the true optimal rule is linear itself, once restricted to such a class of linear rules, the optimal rule within this class may no longer be the optimal for the target population, since both  $p_X(x)$  and  $q_X(x)$  can be very

different. It is thus desirable to guarantee that  $\tilde{\mathcal{B}}(d)$  is not small, regardless of  $X$  and which treatment is optimal; and ideally, the larger the better. To this end, we propose the following method, namely, Minimax Linear Decisions (MiLD). Note that

$$\begin{aligned}\tilde{\mathcal{B}}(d) &= \tilde{E}[W(X)I\{d(X) = 1\}|d^*(X) = 1]\tilde{P}\{d^*(X) = 1\} \\ &\quad + \tilde{E}[W(X)I\{d(X) = -1\}|d^*(X) = -1]\tilde{P}\{d^*(X) = -1\}.\end{aligned}$$

We introduce a lower bound  $\alpha$  on  $\tilde{E}\{W(X)I\{\text{sign}(X^\top\beta_1 + \beta_0) = j\}|d^*(X) = j\}$ ,  $j = \pm 1$ , which represents the expected benefit that would have been obtained if the patients were to receive treatment  $j$ , whose optimal treatments would indeed be  $j$  in the target population. Hence,  $\alpha$  controls the worst case overall benefits for each group, and we set the same lower bound for both quantities for simplicity. We then consider the following optimization problem:

$$\begin{aligned}\max_{\alpha, \beta_1, \beta_0} \quad &\alpha \quad \text{subject to} \quad \inf_{X \sim f_1} \tilde{E}\{W(X)I(X^\top\beta_1 + \beta_0 \geq 0)|d^*(X) = 1\} \geq \alpha, \quad (1) \\ &\inf_{X \sim f_{-1}} \tilde{E}\{W(X)I(X^\top\beta_1 + \beta_0 < 0)|d^*(X) = -1\} \geq \alpha,\end{aligned}$$

where  $f_1$  and  $f_{-1}$  are the density of  $X$  in patients whose optimal treatment are 1 or  $-1$  respectively. That is, we want to guarantee the quality of being given optimal treatment as large as possible among those who should indeed be treated with the same treatment.

Let  $\tilde{q}_j(x)$  denote the density of  $X$  for patients with  $d^*(X) = j$ ,  $j = \pm 1$  in the target population. Then

$$\begin{aligned}\tilde{E}\{W(X)I(X^\top\beta_1 + \beta_0 \geq 0)|d^*(X) = j\} &= \int W(x)I(x^\top\beta_1 + \beta_0 \geq 0)q_j(x)dx \\ &\propto \check{P}(X^{\dagger\top}\beta_1 + \beta_0 \geq 0),\end{aligned}$$

where the density of  $X^\dagger$  is proportional to  $\tilde{q}_j(x^\dagger)W(x^\dagger)$ . Hence, (1) is equivalent to

$$\begin{aligned}\max_{\alpha, \beta_1, \beta_0} \quad &\alpha \quad \text{subject to} \quad \inf_{X^\dagger \sim f_1^\dagger} \check{P}(X^{\dagger\top}\beta_1 + \beta_0 \geq 0) \geq \alpha, \quad (2) \\ &\inf_{X^\dagger \sim f_{-1}^\dagger} \check{P}(X^{\dagger\top}\beta_1 + \beta_0 < 0) \geq \alpha,\end{aligned}$$

where  $f_1^\dagger$  and  $f_{-1}^\dagger$  are the density of  $X^\dagger$  in patients with optimal treatment being 1 or  $-1$  respectively.

However,  $f_1^\dagger$  and  $f_{-1}^\dagger$  could be very different, and difficult to characterize based on the trial data. We propose to quantify the difference between the target population and the trial population in terms of the first two moment conditions, without making any specific distributional assumptions for the two populations. In other words, the covariate moments of  $X^\dagger$  in patients with  $d^*(x) = j$ ,  $j = \pm 1$  in the target population could be different from that of the trial population to some degree.

Suppose that the means of covariate  $X^\dagger$  are  $\tilde{\mu}_j^\dagger$ , and intraclass covariance matrices are  $\tilde{\Sigma}_j$  for patients with  $d^*(X) = j$  in the target population, and  $\mu_j^\dagger$  and  $\Sigma_j^\dagger, j = \pm 1$ , respectively, in the trial population under the new density. We assume that the quantities in the target population belong to

$$\mathcal{U}_j^\dagger = \{(\tilde{\mu}_j^\dagger, \tilde{\Sigma}_j^\dagger) : (\tilde{\mu}_j^\dagger - \mu_j^\dagger)^\top \tilde{\Sigma}_j^\dagger (\tilde{\mu}_j^\dagger - \mu_j^\dagger) \leq \nu^2, \|\tilde{\Sigma}_j^\dagger - \Sigma_j^\dagger\|_F \leq \rho_j\}, j = \pm 1. \quad (3)$$

Here,  $\nu \geq 0$  and  $\rho_j \geq 0$  are known constants, and  $\|\cdot\|_F$  is the Frobenius norm defined as  $\|M\|_F^2 = \text{Tr}(M^\top M)$ . Such conditions define the closeness of the target population to the trial population.  $\mathcal{U}_j^\dagger$  suggests that the mean  $\tilde{\mu}_j^\dagger$  in the target population belongs to an elliptical region around  $\mu_j^\dagger$  with shape determined by the covariance  $\tilde{\Sigma}_j^\dagger, j = \pm 1$ . The covariance matrix  $\tilde{\Sigma}_j^\dagger$ , centered around  $\Sigma_j^\dagger$ , can also vary to certain degrees. Clearly, the larger  $\nu$  or  $\rho_j$ , the more different the two populations are. (2) is consequently written as

$$\begin{aligned} \max_{\alpha, \beta_1, \beta_0} \alpha \text{ subject to} \quad & \inf_{X^\dagger \sim (\tilde{\mu}_1^\dagger, \tilde{\Sigma}_1^\dagger) \in \mathcal{U}_1^\dagger} \check{P}(X^{\dagger\top} \beta_1 + \beta_0 \geq 0) \geq \alpha, \quad (4) \\ & \inf_{X^\dagger \sim (\tilde{\mu}_{-1}^\dagger, \tilde{\Sigma}_{-1}^\dagger) \in \mathcal{U}_{-1}^\dagger} \check{P}(X^{\dagger\top} \beta_1 + \beta_0 < 0) \geq \alpha. \end{aligned}$$

This yields a linear decision rule that safeguards against the possible difference of the distribution of  $X$  between the trial and the target populations. Such robustness of the treatment regime estimation is mainly due to the minimum requirements on the means and covariances. The above objective function has a similar form to the minimax probability machine techniques developed in Lanckriet et al. (2003), and their techniques for deriving the optimal linear rule can be employed if  $(\tilde{\mu}_j^\dagger, \tilde{\Sigma}_j^\dagger), j = \pm 1$ , were known. The key step is to recognize that the condition  $\inf_{X^\dagger \sim (\tilde{\mu}_j^\dagger, \tilde{\Sigma}_j^\dagger)} P(X^{\dagger\top} \beta_1 + \beta_0 \geq 0) \geq \alpha$ , holds if and only if

$$\beta_0 + \tilde{\mu}_j^{\dagger\top} \beta_1 \geq \kappa(\alpha) \sqrt{\beta_1^\top \tilde{\Sigma}_j^\dagger \beta_1}, j = \pm 1, \quad (5)$$

by applying the generalized Chebychev inequality (Marshall and Olkin, 1960), where  $\kappa(\alpha) = \sqrt{\alpha/(1-\alpha)}$  (Lemma 1, Lanckriet and others (2003)). The constraint (5) is imposed on the distance with respect to the mean of the covariates within the class, taking into account the effect of the covariance matrices, which could be representative of the class. In other words, we search a line such that the normalized margin between the means of classes is as large as possible. Consequently, (4) is equivalent to

$$\begin{aligned} \max_{\alpha, \beta_1, \beta_0} \alpha \text{ subject to} \quad & \beta_0 + \tilde{\mu}_1^{\dagger\top} \beta_1 \geq \kappa(\alpha) \sqrt{\beta_1^\top \tilde{\Sigma}_1^\dagger \beta_1} \\ & \beta_0 + \tilde{\mu}_{-1}^{\dagger\top} \beta_1 \geq \kappa(\alpha) \sqrt{\beta_1^\top \tilde{\Sigma}_{-1}^\dagger \beta_1}. \end{aligned}$$

As shown in the Appendix, a further simplified objective function can be obtained as

$$\min_{\beta_1} \sqrt{\beta_1^\top (\Sigma_1^\dagger + \rho_1 I_p) \beta_1} + \sqrt{\beta_1^\top (\Sigma_{-1}^\dagger + \rho_{-1} I_p) \beta_1}, \quad (6)$$

subject to  $\beta_1^\top(\mu_1^\dagger - \mu_{-1}^\dagger) = 1$ . We can eliminate the equality constraint in (6) by letting  $\beta_1 = \beta_{10} + Fu$ , where  $u \in \mathbb{R}^{p-1}$ ,  $\beta_{10} = (\mu_1^\dagger - \mu_{-1}^\dagger) / \|\mu_1^\dagger - \mu_{-1}^\dagger\|_2$ .  $F$  is a  $p \times (p-1)$  matrix. Let  $h$  denote the position of the maximum element in  $\mu_1^\dagger - \mu_{-1}^\dagger$ .  $F$  is constructed from a  $(p-1) \times (p-1)$  identity matrix, with  $-(\mu_1^\dagger - \mu_{-1}^\dagger) / \max(\mu_1^\dagger - \mu_{-1}^\dagger)$  inserted into the  $h^{\text{th}}$  row. We can see that  $F$  is an orthogonal matrix whose columns span the subspace of vectors orthogonal to  $\mu_1^\dagger - \mu_{-1}^\dagger$ . Hence, the optimization problem can be written as

$$\min_{\beta_1} \sqrt{(\beta_{10} + Fu)^\top (\Sigma_1^\dagger + \rho_1 I_p) (\beta_{10} + Fu)} \\ + \sqrt{(\beta_{10} + Fu)^\top (\Sigma_{-1}^\dagger + \rho_{-1} I_p) (\beta_{10} + Fu)}.$$

The lower bound on the worst case allocation rate  $\alpha^* = (\kappa^* - \nu)^2 / \{1 + (\kappa^* - \nu)^2\}$ , and  $\nu$  should not exceed  $\kappa^*$ . Here,

$$\kappa^* = \left\{ \sqrt{\beta_1^{*\top} (\tilde{\Sigma}_{-1}^\dagger + \rho_{-1} I_p) \beta_1^*} + \sqrt{\beta_1^{*\top} (\tilde{\Sigma}_1^\dagger + \rho_1 I_p) \beta_1^*} \right\}^{-1}$$

and  $\nu$  is defined in (3). More details can be found in the Appendix.

**Remark 1** *Although the MiLD is proposed to identify a robust linear rule, it can be easily generalized using nonlinear kernel functions. We seek a decision rule in the form of  $f(x) = \varphi(x)^\top \beta_1 + \beta_0$ , where  $\varphi(x)$  is a set of basis functions. The kernelization of the proposed approach is possible because the objective function (6) can be expressed in terms of inner products between different  $X$ 's. Hence, the objective and constraint can be expressed in terms of inner products of  $\varphi(X)$ . Subsequently, we can compute the robust nonlinear rule by slightly modifying the algorithm.*

#### 2.4. Estimating robust rules using empirical data

To solve for the minimax linear decisions given the observed data from a clinical trial, we first need to estimate  $(\mu_j^\dagger, \Sigma_j^\dagger)$ ,  $j = \pm 1$ , which depends on both the optimal treatment rule  $d^*(x)$  and the weight  $W(X)$ . A three-step procedure is outlined below.

*Step 1.* Estimate  $d^*(x)$  using a nonparametric method with the trial data, denoted by  $\hat{d}(x)$ .

*Step 2.* Estimate  $(\tilde{\mu}_j^\dagger, \tilde{\Sigma}_j^\dagger)$  using the initial estimate  $\hat{d}(x)$ , denoted by  $(\hat{\mu}_j^\dagger, \hat{\Sigma}_j^\dagger)$ .

*Step 3.* Implement MiLD based on the estimated  $(\hat{\mu}_j^\dagger, \hat{\Sigma}_j^\dagger)$ .

In Step 1, there are several ways to estimate  $d^*$ . Indirect estimation methods model the response as a function of  $X$  and  $A$ , and select treatment, which maximizes the predicted mean outcome (Qian and Murphy, 2011). Since a consistent estimator  $\hat{d}(x)$  is required to guarantee the performance, a flexible nonparametric method in Step 1 is preferred. For example, support vector machine and support vector regression can be used to estimate  $E(T|X, A)$  for binary and continuous outcomes respectively. Alternatively, outcome weighted learning proposed

in Zhao et al. (2012) circumvents the two-step procedure by directly optimizing the value function. We can employ a kernel function to induce nonlinearity in obtaining the initial estimator  $\hat{d}$ .

In a cancer clinical trial, it is common that the primary endpoint is survival time that subjects to censoring. Given that  $d^*(x)$  is invariant over the covariate distribution, we suggest utilizing flexible nonparametric machine learning methods. In particular, we will use the random forest survival tree method (Ishwaran et al., 2008) to estimate  $E(T|X, A)$ , an ensemble tree method that extended random forest method (Breiman, 2001) for analysis of right-censored survival data. It uses independent bootstrap samples to grow trees by randomly selecting a subset of variables at each node and splitting the node using a survival criterion adjusting for censoring status. The ensemble estimated cumulative hazard function is the average of the Nelson-Aalen estimator for each case's terminal node. Easy-to-use software is available on R CRAN (<https://cran.r-project.org>). We can estimate  $E(T|X, A)$  correspondingly via  $E(T|X, A) = \int_0^T P(T > t|X, A)dt$ . Then  $\hat{d}(x) = \text{sign}\{\hat{f}(x)\}$ , where  $\hat{f}(x) = \hat{E}(T|X = x, A = 1) - \hat{E}(T|X = x, A = -1)$ .

To estimate  $(\mu_j^\dagger, \Sigma_j^\dagger)$  with a general weight  $W(x)$ , we treat it as a multiplicative adjustment to  $X$  where we now sample from a density proportional to  $p_X(x)W(x)$  instead of  $p_X(x)$ . This motivates us to employ techniques from importance sampling to estimate  $(\mu_j^\dagger, \Sigma_j^\dagger)$ ,  $j = \pm 1$ . In particular,  $\mu_j^\dagger$  can be estimated by

$$\hat{\mu}_j^\dagger = \frac{\sum_{i=1}^n X_i W(X_i) I\{\hat{d}(X_i) = j\}}{\sum_{i=1}^n W(X_i) I\{\hat{d}(X_i) = j\}},$$

and  $\Sigma_j^\dagger$  can be estimated by

$$\hat{\Sigma}_j^\dagger = \sum_{i=1}^n \left[ \frac{W(X_i) I\{\hat{d}(X_i) = j\}}{\sum_{i=1}^n W(X_i) I\{\hat{d}(X_i) = j\}} \right]^2 (X_i - \hat{\mu}_j^\dagger)^\top (X_i - \hat{\mu}_j^\dagger).$$

In our case, we choose

$$\begin{aligned} W(x) = \widehat{W}_1(x) &= \widehat{E}\{T|A = d^*(x), X = x\} - \widehat{E}\{T|A \neq d^*(x), X = x\} \\ &= \left| \widehat{E}\{T|A = 1, X = x\} - \widehat{E}\{T|A = -1, X = x\} \right| \\ &= |\hat{f}(X)| \end{aligned}$$

or  $W(X) = W_2(X) \equiv 1$ .

Provided with the estimated  $(\hat{\mu}_j^\dagger, \hat{\Sigma}_j^\dagger)$ ,  $j = \pm 1$ , we solve for the robust linear treatment regimes in Step 3 using the procedure outlined in Section 2.3. In particular, estimates will be plugged in for  $\mu_j^\dagger$  and  $\Sigma_j^\dagger$ ,  $j = \pm 1$  and related quantities. We obtain  $\hat{\beta}_1$  and  $\hat{\beta}_0$  accordingly.

### 2.5. Choices of $(\nu, \rho_j)$

Sometimes pilot data from the target population or the general patient population data are available. We can use this information to choose  $\rho_j$  and  $\nu$ . First we obtain  $\hat{d}(x)$  as an initial estimate of  $d^*(x)$  using the trial data. Based on this preliminary decision boundary  $\hat{d}(x)$ , we can estimate  $\tilde{\mu}_j^\dagger$  and  $\tilde{\Sigma}_j^\dagger, j = \pm 1$ , denoted as  $\hat{\mu}_j$  and  $\hat{\Sigma}_j$ , for a target population using the pilot data. Then we can set  $\rho_j = b\|\hat{\Sigma}_j\|_F$ , and  $\nu = \sum_j (b\hat{\mu}_j^\top \hat{\Sigma}_j^{-1} (b\hat{\mu}_j))/2$ , where  $b$  is a constant characterizing the potential differences between two population means, as well as the norm of the covariance matrix. If relevant data is not available, we can conduct sensitivity analyses on different combinations of  $(\nu, \rho_j)$ , and assess how the changes in  $(\nu, \rho_j)$  will influence the resulting decision rules and the worst-case allocation rates.

### 3. Theoretical results

In this section, we will establish some theoretical properties of the proposed MiLD method. Assume that  $(X_i, A_i, T_i), i = 1, \dots, n$ , are i.i.d observations from the trial. Additionally, we assume that

**Assumption 1** All covariates are bounded such that  $|X| \leq M$ .

**Assumption 2** Let  $\beta_1^*$  be the unique solution to (1).  $\beta_1^*$  lies in the interior of a compact set  $\mathbf{B}$ .

We assume the following condition on  $f^*(x) = E(T|X = x, A = 1) - E(T|X = x, A = -1)$ :

**Assumption 3** Margin condition: there exist  $K_1, \gamma > 0$  such that for all  $t > 0$

$$\tilde{P}(|f^*(X)| \leq t) \leq K_1 t^\gamma.$$

**Assumption 4**  $\hat{f}$  converges to  $f^m$ , which could be different from  $f^*$ . The convergence rate of estimated  $\hat{f}$  to  $f^m$  satisfies  $\|\hat{f} - f^m\|_2 = O_p(r_n)$ , where  $\|f\|_2 = E\{f(X)^2\}^{1/2}$  and the expectation is taken with respect to the distribution in the trial data.

**Assumption 5** The difference of the covariates is uniformly bounded

$$\max_{i, i' \leq n} |X_i - X_{i'}| \leq C_X,$$

for some constant  $C_X > 0$ .

**Assumption 6** Let  $\check{f}_i(x)$  be the estimate of  $f^*(x)$  using the Step 1 data except that the  $i^{\text{th}}$  observation  $(x_i, a_i, t_i)$  is replaced by an independent observation  $(\check{x}_i, \check{a}_i, \check{t}_i)$ . We assume that

$$\sup_X |\check{f}_i(X) - \hat{f}(X)| \leq \rho/n$$

for some constant  $\rho$ .

Assumption 3 is an analogue of the well-known *margin condition* (Tsybakov, 2004), which is commonly used to characterize the noise around the decision boundary in a binary classification problem. Here, the assumption describes the distribution of  $f^*(X)$  when  $X$  is near the boundary  $\{x : f^*(x) = 0\}$  in the target population, which usually contains more noise, and  $\gamma$  controls the size of these regions. Particularly, larger values of  $\gamma$  mean that the two treatment effects are less likely to be similar, and it is easier to distinguish patients who would have benefited from one treatment from those who would have benefited from the other. Usually,  $\gamma \in [0, m]$  for a smooth contrast function  $f^*(x)$ , unless  $f^*(x)$  does not cross 0 at any point, i.e., all the patients benefit from one treatment (Audibert et al., 2007). For example, if  $X \sim \text{Uniform}[-1, 1]$ ,  $P(A = 1|X = P(A = -1|X) = 1/2)$  and  $E(T|X, A) = XA$ , then  $\gamma = 1$ . The bounded covariate assumption, Assumption 5, is satisfied in most real applications. Assumption 6 indicates that small changes in the data will only lead to small changes in the estimates. For example, if we use Cox regression with nonlinear basis functions, or nonparametric kernel estimators, this condition will be satisfied (Devroye et al., 2013). The following results are proved in the supplementary materials (Zhao et al., 2019).

**Theorem 1** *Under the Assumptions 1-6, it holds that  $|\hat{\beta}_1 - \beta_1^*| = O_p(\|f^* - f^m\|_2^{\min(1, \frac{2\gamma}{\gamma+2})} + r_n^{\min(1, \frac{2\gamma}{\gamma+2})} + n^{-1/2})$ .*

The first term reflects the approximation error due to the initial estimator  $\hat{f}$ . If  $\hat{f}$  is a consistent estimator of  $f^*$ , where  $f^* = f^m$ , then the first term will disappear. The other terms bound the stochastic error, which arises from the variability inherent in a finite sample size, which captures the efficiency loss due to the first stage estimation. As an example, if we use random forest to estimate the required quantities,  $r_n = n^{-\theta}$  with  $\theta = \frac{0.75}{2\{S \log(2) + 0.75\}}$  under some mild assumptions, where  $S$  denotes the number of strong features used in the estimating process (Biau, 2012). The theorem indicates that the convergence rate of the estimated linear treatment rules depends on the the initial treatment rule, and the behavior of  $f^*(x)$  in the neighborhood of the boundary. For example, if a kernel estimator is used, the optimal rate of  $r_n$  would be  $n^{-2/(m+4)}$ , and the conclusion rate is  $n^{-4\gamma/(m+4)(\gamma+2)}$ .

#### 4. Simulation Studies

We conduct extensive simulations to evaluate the proposed methods. In all scenarios, the dimension of the covariate space is 10. Binary treatments  $A$  are generated from  $\{-1, 1\}$  with equal probability. Three different scenarios are presented, with outcomes generated as follows.

Scenario 1. We simulate the first half of patients from  $X_1, X_2 \sim N(1, 1)$ ,  $X_3, \dots, X_{10} \sim N(0, 1)$  and the second half of patients with  $X_1, \dots, X_{10}$  from  $N(0, 1)$ . The survival time is the minimum of  $\tau = 0.5$  and  $\tilde{T}$ , where  $\tilde{T}$  is generated with

$$\log \tilde{T} = \exp\{0.6 * X_1 - 0.8 * X_2 + A * c(X)\} + \log \epsilon.$$

Here,  $c(X) = 1$  for the first half of patients and  $c(X) = -1$  for the other half, and  $\epsilon$  is generated from an exponential distribution with mean 1. Censoring time  $C$  is generated from Uniform[0, 1]. The censoring percentage is around 24%. The optimal decision boundary is  $d^*(X) = \text{sign}(X_1 + X_2 - 1)$ .

Scenario 2.  $X_1, \dots, X_{10}$  are generated from  $N(0, 1)$ . The survival time is the minimum of  $\tau = 4$  and  $\tilde{T}$ , where  $\tilde{T}$  is generated with the hazard rate function

$$\lambda_{\tilde{T}}(t|X, A) = \exp[0.6X_1 + 0.8X_2 - 1 + \{2X_1 + 3(X_2 + 1)^2 - 2\}A].$$

Censoring time  $C$  is generated from Uniform[0, 5]. The censoring percentage is around 42%. The optimal decision boundary is  $d^*(X) = -\text{sign}\{2X_1 + 3(X_2 + 1)^2 - 2\}$ .

Scenario 3.  $X_1, \dots, X_{10}$  are generated from  $N(0, 1)$ . The survival time is the minimum of  $\tau = 4$  and  $\tilde{T}$ , which is generated with

$$\log(\tilde{T}) = X_1 + X_2 + 1 + A(2X_1^3 + 2X_2 + 0.5) + N(0, 1).$$

Censoring time  $C$  is generated from

$$\log(C) \sim X_1 + X_2 + X_3 + N(0, 1).$$

The censoring percentage is around 51%. The optimal decision boundary is  $d^*(X) = \text{sign}(2X_1^3 + 2X_2 + 0.5)$ .

We also consider settings where there is a mismatch between the trial participants and the target population, and we will denote these mismatch scenarios as Scenarios 1', 2', and 3' respectively. In Scenario 1', half of the patients will gain benefits from treatment 1 in both populations. We modify it in Scenario 1' such that the proportion of patients with  $d^*(x) = 1$  is 1/3 in the trial population, and 1/2 in the target population. In Scenario 2', we let the covariate distribution in the trial data with  $X_1 \sim N(-0.25, 1.5)$  and other covariates following  $N(0, 1.5)$ , which are different from the distribution in the target population with all covariates generated from  $N(0, 1)$ . Eligibility criteria are usually applied for the trial recruitment, and thus trials might selectively enroll patients from the target population. Instead of randomly choosing patients for participation in the trial, patients are selectively enrolled with certain probability, denoted by  $\pi(X)$ , into the trial data. In Scenario 3', the model for the enrollment is  $\text{logit}\{\pi(X)\} = -2X_1^3 - 1$ , where  $\text{logit}(t) = \log\{t/(1-t)\}$ . Hence, covariates predictive of participation in the trial could be predictive of treatment effects, where patients with  $d^*(x) = 1$  are less likely to participate in the trial. Subsequently, the covariate distributions in trial data and the target population are not the same.

We hope to discover a simple linear decision rule of high quality even if the true treatment effect might be nonlinear. MiLD will be evaluated for this purpose. In this paper, we utilize random survival forest to obtain the initial estimate of  $d^*(x)$ , but other methods can also be applied. Then we apply both

$\widehat{W}_1(x)$  and  $W_2(x) \equiv 1$  in MiLD, targeted to optimize the value and the allocation rate to the optimal rule, respectively. We denote them as MiLD-V and MiLD-P. The results might be affected with possible deviations between the target and the trial populations. When we do not have prior information on the target population, we can vary  $(\nu, \rho_j)$  to evaluate the resulting changes. As suggested in Section 2.5, we set  $\rho_j = b\|\widehat{\Sigma}_j\|_F$ ,  $\nu = \{\sum_j (b\widehat{\mu}_j)^\top (\widehat{\Sigma}_j + b\|\widehat{\Sigma}_j\|_F I_p) (b\widehat{\mu}_j)\}/2$ ,  $j = \pm 1$ . Here,  $\|\widehat{\Sigma}_j\|_F$  is the Frobenius norm of  $\widehat{\Sigma}_j$ , and  $b$  is a constant that reflects the deviations relative to the covariance matrix norms and the means, in the trial populations. We consider two cases with  $b = 0$  and  $b = 0.1$  for both MiLD-V and MiLD-P.

We compare the proposed methods to the following two approaches.

1. OWL (outcome weighted learning): we find the best linear decision rule by directly targeting maximizing the overall expected outcome (Zhao et al., 2012).
2. COX: we find the linear decision rule based on a Cox regression model, where interactions between  $X$  and  $A$  are included.

All methods are performed by deriving the best linear treatment rules using trial data. We then calculate the misclassification rate under the estimated rules via Monte Carlo methods. Specifically, a large testing dataset of 10,000 from the target population is generated. Training datasets representing the trial population are repeatedly simulated, and each time the rules yielded by various methods are evaluated on this testing set. Different training data sample sizes  $n = 250, 500$  and  $1000$  are considered. We report the average misallocation to non-optimal treatments over 500 replicates. We also report the expected overall survival of the estimated rules using the derived treatment rules by different methods in the supplementary materials.

The results in Figures 1–3 present evidence that the proposed methods perform well compared with the other methods. The Cox regression method relies on model fitting, and thus many subjects are recommended the wrong treatments by the resulting linear decision rule, labeled as misallocation rates in Figures 1–3. Even in Scenarios 1 and 1' where the optimal treatment rule is linear, the non-proportional hazards model negatively impacts the performance of Cox regression. The OWL method in general leads to a larger variability. On the other hand, both MiLD-V and MiLD-P perform well throughout, and the results are insensitive to the value of  $b$ . The limits of misallocation rates that MiLD-V and MiLD-P converge to are the lowest among all competitors. It can be seen that when the trial population is not representative as illustrated in Scenarios 1', 2' and 3', MiLD-V and MiLD-P are robust to the bias, and they yield favorable results. We note that the results in MiLD with  $b = 0.1$ , which allows a larger deviation between the trial and the target population, give a slightly better result most of the time in these cases. We also illustrate the efficiency loss due to the first stage estimation of  $f^*(x)$  numerically. Details can be found in supplementary material.

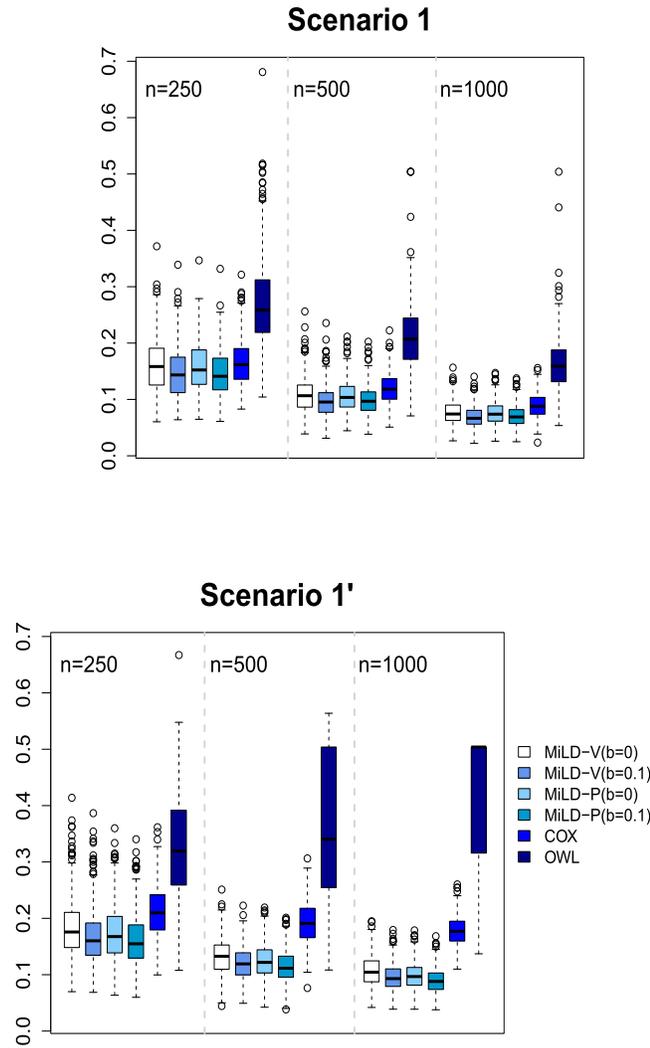


FIG 1. Misallocation rates in Scenarios 1 and 1' using trial data with sample sizes varying from 250 to 1000. The misallocation rate is evaluated by using 500 Monte Carlo repetitions.

### 5. Data Analysis

Prior studies suggest that elevated markers of bone turnover are prognostic for poor survival in castration-resistant prostate cancer, but their predictive value for the bone-targeted therapy has not been fully investigated. We illustrate the proposed methods using data from SWOG 0421 study, a North American Intergroup phase III trial (participants: SWOG, Eastern Cooperative Oncology Group, Cancer and Leukemia Group B/Alliance) for men with metastatic

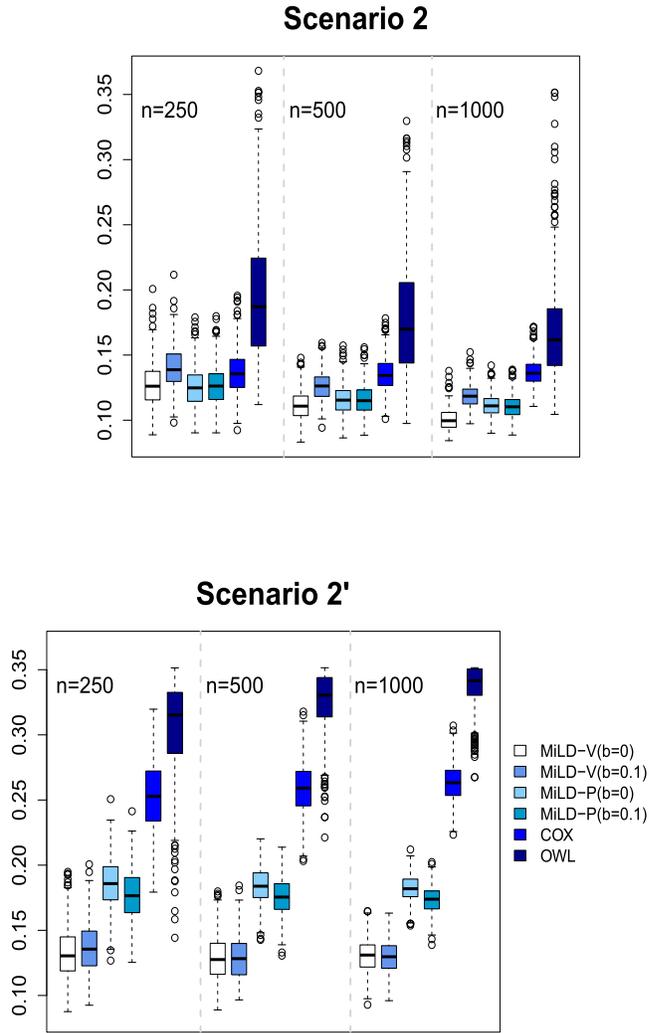


FIG 2. Misallocation rates in Scenarios 2 and 2' using trial data with sample sizes varying from 250 to 1000. The misallocation rate is evaluated by using 500 Monte Carlo repetitions.

castration-resistant prostate cancer. They were randomly assigned in a blinded fashion in a 1:1 ratio to docetaxel administered every 21 days at a dose of 75 mg/m<sup>2</sup> with or without the bone-targeted oral agent atrasentan taken daily for up to 12 cycles (Quinn et al., 2013; Lara et al., 2014). S0421 enrolled 1038 eligible patients; of these, 855 submitted serum for the bone biomarkers and 778 patients had usable specimens at baseline. After removing missing observations, the sample size for our analysis is 751, where 371 patients are in the docetaxel + atrasentan arm and 380 patients are in the docetaxel + placebo arm.

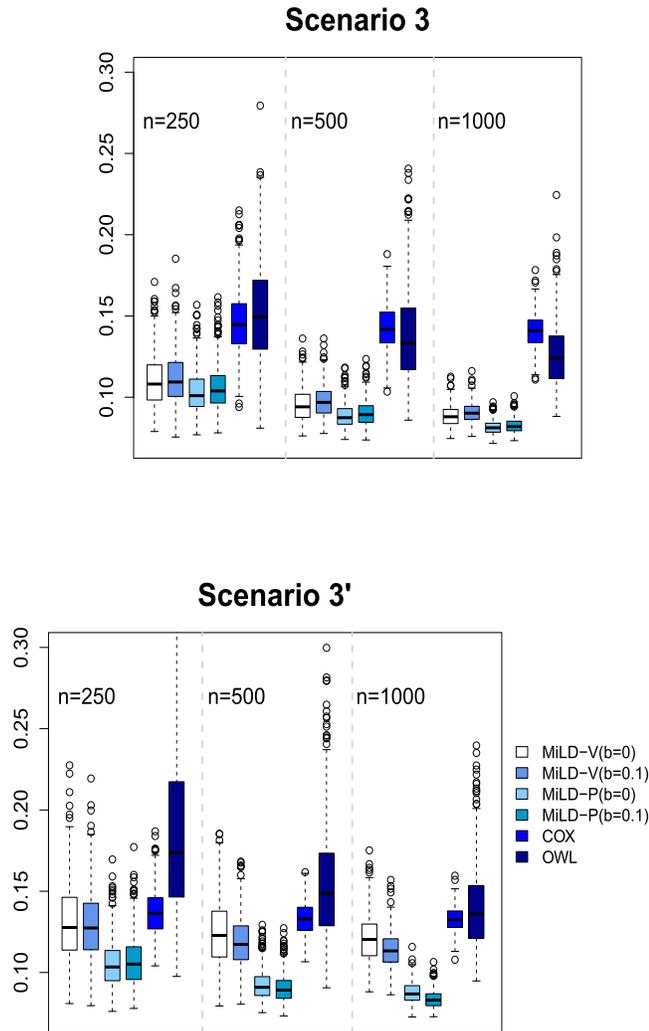


FIG 3. Misallocation rates in Scenarios 3 and 3' using trial data with sample sizes varying from 250 to 1000. The misallocation rate is evaluated by using 500 Monte Carlo repetitions.

Co-primary endpoints in this study were progression-free survival and overall survival. In our analysis, we use overall survival as our outcome, which is truncated at the end of the study, and could be censored due to loss of follow-up. We use 10 baseline covariates to inform the optimal treatment rule, including age (range 40–92), serum prostate-specific antigen (PSA, range 0.1–10414.1), indicator of bisphosphonate usage (61%), indicator of metastatic disease beyond the bones (55%), indicator of pain at baseline (60%), indicator of performance status (2-3 versus 0–1, 56% 2-3), and bone marker levels. Four bone markers are

measured, including Bone alkaline phosphatase (BAP, range 1.9–1761.0 u/L), C-terminal of type 1 collagen (CICP, range 1.4–273.6 ng/mL), N-telopeptides of type 1 collagen (NTx, range 1.4–480.0 nM) and pyridinoline (PYD, range 0.3–15.0 nmol/L). The distribution of bone marker concentrations and serum PSA were skewed with a wide range; therefore, we use log transformation for these variables. All covariates are then standardized for analysis. We intend to find the optimal linear treatment rule that is robust to a potential difference between the trial population and the future population.

We compare the proposed methods with Cox regression and OWL method. Practitioners often directly generalize the results from a clinical trial to the general patient population. However, It is not uncommon that the participants in clinical trials are in general more healthier than the patient population, due to the restrictions on patient eligibility. Hence, in our data analysis, we mimic this phenomenon by changing the distribution of healthier patients in the trial population and the target population. We categorize patients to healthier patients whose serum PSA is below the median level and sicker patients whose serum PSA is above the median level. We employ a cross-validated type analysis. At each run, we partition the whole data set into 5 pieces, where 2 parts of the data are used as training data to estimate the optimal rules, and the remaining part as the validation set for evaluating the estimated rules. Specifically, each training data set consists of 2/3 healthier patients and 1/3 sicker patients; on the other hand, each validation set contains 1/3 healthier patients and 2/3 sicker patients. Thus, there is a substantial discrepancy between the training set and the validation set, which represents the trial population and the target population, respectively. The cross-validated values are obtained by averaging the empirical value on all 3 validation subsets. To adjust for censoring when calculating the empirical values, we use inverse probability of censoring weighting techniques, where the empirical value for a treatment decision rule  $d$  is calculated by  $\mathbb{P}_n[\tilde{Y}I\{A = d(X)\}]/\mathbb{P}_n[\tilde{I}\{A = d(X)\}]$ , with  $\tilde{Y}$  equaling to  $\Delta Y/\hat{S}_C(Y|A, X)$ . We use a kernel estimator developed in Li et al. (1999) to obtain  $\hat{S}_C(t|A, X)$ , which does not require a model assumption. The procedure is repeated 200 times. The averages and standard errors of these values are reported in Table 1, where a larger value corresponds to a longer expected survival time. The results show better performances of both MiLD-V and MiLD-P procedures compared to other methods. MiLD-V performs the best perhaps because it targets to optimize the value directly.

We then apply the proposed methods to the whole data set. The coefficients in the treatment decision rule recommended by MiLD-V and MiLD-P are presented

TABLE 1  
Mean (s.e.) cross-validated values (days)

COX	OWL	MiLD-V		MiLD-P	
		$b = 0$	$b = 0.1$	$b = 0$	$b = 0.1$
749.2 (69.8)	711.1 (71.2)	764.6 (70.2)	765.0 (68.9)	753.5 (69.4)	753.7 (69.4)

“s.e.” denotes standard errors.

TABLE 2  
Coefficients for the estimated linear decision rules by MiLD-V and MiLD-P using the SWOG0421 data

	MiLD-V	MiLD-P
Intercept	0.021	0.037
Age	0.538	0.150
Baseline serum PSA	-0.510	-0.734
Bisphosphonate usage (YES = 1)	0.787	1.235
Metastatic disease beyond the bones (YES = 1)	-0.301	0.008
Pain (YES = 1)	-0.272	0.050
Performance Status ('2-3' = 1)	0.515	0.796
BAP	0.382	0.972
CICP	-0.023	0.302
NTx	0.137	0.657
PYD	-0.339	-0.326

in Table 2. The results yielded by MiLD-V and MiLD-P are close, where 80% of the patients receive the same treatment recommendation. In Figure 4(a), we compare the Kaplan–Meier curves for overall survival between two treatment arms. There does not appear to be separation. However, when comparing the group whose treatment assignments were in accordance with the treatments recommended by MiLD-V or MiLD-P with the other group, the Kaplan–Meier survival curves show a clear separation, as shown in Figures 4(b) and (c).

## 6. Discussion

In practice, it is preferable to use interpretable decision rules when communicating with clinical practitioners about treatment recommendation. A canonical example is linear decision rule, which is attractive because it can be easily understood, and thus can be used to guide future research. Our present work shows that it is possible to identify high-quality linear decision rule that leads to a greater overall benefit, even if the truth may be nonlinear. Furthermore, the proposed method is robust across future populations, taking into account the fact that the study sample may not be representative.

We consider survival time outcomes in this paper. Such endpoints are critical in many settings especially in oncology. However, MiLD can be readily applied for binary or continuous outcomes, provided that we can use existing nonparametric methods to obtain preliminary estimates of the decision boundaries. Popular methods for binary outcomes include support vector machine and boosting (Hastie et al., 2009), and for continuous outcomes, we can apply random forest (Breiman, 2001) and support vector regression (Vapnik et al., 1997). Our current proposal suggests an  $L_2$  penalty for handling high-dimensional covariates. However, this does not provide sparse solutions. In certain circumstances, several important variables characterize the optimal treatment rules, where the means and covariance matrices of those unimportant variables do not matter. It would be interesting to develop robust methods that conduct variable selection

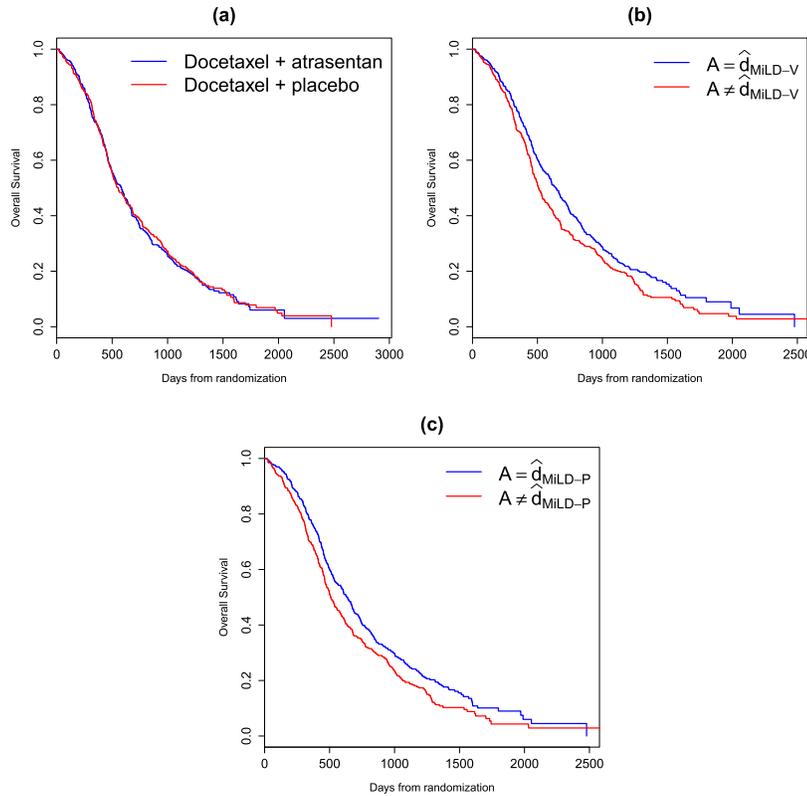


FIG 4. Kaplan–Meier survival curves of overall survival in castration-resistant prostate cancer patients: (a) by treatment received; (b) by accordance between treatment recommended by MiLD-V and treatment received; (c) by accordance between treatment recommended by MiLD-P and treatment received.

simultaneously, which would eliminate the unimportant variables and further improve the ease of interpretation. Another interesting extension of the current work is to consider settings involving more than two treatments. While it is straightforward to conduct a series of pairwise comparisons, further development is required to identify the best rule among all treatments. It will also be interesting to investigate extensions to Boolean combination of linear rules, i.e., rules of the form  $\{X_1\beta_{11} + \beta_{01} \geq 0\} \cap \{X_2\beta_{21} + \beta_{02} \geq 0\}$ .

## Appendix A: Appendix

**Solving minimax linear decisions.** It follows from the generalized Chebychev inequality (Marshall and Olkin, 1960) that (2) is equivalent to

$$\begin{aligned} \max_{\alpha, \beta_1, \beta_0} \alpha \quad s.t. \quad & \inf_{X^\dagger \sim (\tilde{\mu}_1^\dagger, \tilde{\Sigma}_1^\dagger) \in \mathcal{U}_1^\dagger} \beta_0 + \tilde{\mu}_1^{\dagger\top} \beta_1 \geq \kappa(\alpha) \sqrt{\beta_1^\top \tilde{\Sigma}_1^\dagger \beta_1} \\ & \inf_{X^\dagger \sim (\tilde{\mu}_{-1}^\dagger, \tilde{\Sigma}_{-1}^\dagger) \in \mathcal{U}_{-1}^\dagger} -\beta_0 - \tilde{\mu}_{-1}^{\dagger\top} \beta_1 \geq \kappa(\alpha) \sqrt{\beta_1^\top \tilde{\Sigma}_{-1}^\dagger \beta_1}, \end{aligned}$$

where  $\kappa(\alpha) = \sqrt{\alpha/(1-\alpha)}$ . As shown in Lanckriet et al. (2003),

$$\min_{\mu_j: (\tilde{\mu}_j^\dagger - \mu_j^\dagger)^\top \tilde{\Sigma}_j^{\dagger-1} (\tilde{\mu}_j^\dagger - \mu_j^\dagger) \leq \nu^2} \beta_1^\top \tilde{\mu}_j^\dagger = \beta_1^\top \mu_j^\dagger - \nu \sqrt{\beta_1^\top \tilde{\Sigma}_j^\dagger \beta_1},$$

and  $\max_{\tilde{\Sigma}^\dagger: \|\tilde{\Sigma}^\dagger - \Sigma^\dagger\|_F \leq \rho} \beta_1^\top \tilde{\Sigma}^\dagger \beta_1 = \beta_1^\top (\Sigma^\dagger + \rho I_p) \beta_1$ , where  $I_p$  is the identified matrix. Recall that  $\mathcal{U}_j^\dagger = \{(\tilde{\mu}_j^\dagger, \tilde{\Sigma}_j^\dagger) : (\tilde{\mu}_j^\dagger - \mu_j^\dagger)^\top \tilde{\Sigma}_j^\dagger (\tilde{\mu}_j^\dagger - \mu_j^\dagger) \leq \nu^2, \|\tilde{\Sigma}_j^\dagger - \Sigma_j^\dagger\|_F \leq \rho_j\}, j = \pm 1$ . We have  $\beta_0 + \beta_1^\top \tilde{\mu}_j^\dagger = \beta_0 + \beta_1^\top \mu_j^\dagger - \nu \sqrt{\beta_1^\top \tilde{\Sigma}_j^\dagger \beta_1}$ , and  $\beta_1^\top \tilde{\Sigma}_j^\dagger \beta_1 = \beta_1^\top (\Sigma_j^\dagger + \rho_j I_p) \beta_1$ . We thus further rewrite the optimization problem to

$$\begin{aligned} \max_{\alpha, \beta_1, \beta_0} \alpha \quad s.t. \quad & \beta_0 + \beta_1^\top \mu_1^\dagger \geq \kappa(\alpha) \sqrt{\beta_1^\top (\Sigma_1^\dagger + \rho_1 I_p) \beta_1} \\ & -\beta_0 - \beta_1^\top \mu_{-1}^\dagger \geq \kappa(\alpha) \sqrt{\beta_1^\top (\Sigma_{-1}^\dagger + \rho_{-1} I_p) \beta_1}. \end{aligned}$$

We will assume that  $\mu_1^\dagger \neq \mu_{-1}^\dagger$ ; otherwise, the above problem is not identifiable. We also assume that  $\Sigma_j^\dagger, j = \pm 1$  are both positive definite. Since  $\kappa(\alpha)$  is a monotone increasing function of  $\alpha$ , we can rewrite our problem as

$$\max_{\kappa} \kappa \quad s.t., \quad \beta_1^\top \mu_{-1}^\dagger + \kappa \sqrt{\beta_1^\top (\Sigma_{-1}^\dagger + \rho_{-1} I_p) \beta_1} \leq \beta_0 \leq \beta_1^\top \mu_1^\dagger - \kappa \sqrt{\beta_1^\top (\Sigma_1^\dagger + \rho_1 I_p) \beta_1},$$

and these inequalities will become equalities at the optimum. We can eliminate the equality constraint by letting  $\beta_1 = \beta_{10} + Fu$ , where  $u \in \mathbb{R}^{p-1}$ ,  $\beta_{10} = (\mu_1^\dagger - \mu_{-1}^\dagger) / \|\mu_1^\dagger - \mu_{-1}^\dagger\|_2^2$ , and  $F \in \mathbb{R}^{p \times (p-1)}$  is an orthogonal matrix whose columns span the subspace of vectors orthogonal to  $\mu_1^\dagger - \mu_{-1}^\dagger$ . Hence, the optimization problem can be written as in (2). The optimal

$$\beta_0^* = \beta_1^{*\top} \tilde{\mu}_1^\dagger - \kappa^* \sqrt{\beta_1^{*\top} (\tilde{\Sigma}_1^\dagger + \rho_1 I_p) \beta_1^*} = \beta_1^{*\top} \tilde{\mu}_{-1}^\dagger + \kappa^* \sqrt{\beta_1^{*\top} (\tilde{\Sigma}_{-1}^\dagger + \rho_{-1} I_p) \beta_1^*},$$

where  $\kappa^*$  is the optimal value of  $\kappa$  with

$$\kappa^* = \left\{ \sqrt{\beta_1^{*\top} (\tilde{\Sigma}_{-1}^\dagger + \rho_{-1} I_p) \beta_1^*} + \sqrt{\beta_1^{*\top} (\tilde{\Sigma}_1^\dagger + \rho_1 I_p) \beta_1^*} \right\}^{-1}.$$

The lower bound on the worst case allocation rate  $\alpha^* = (\kappa^* - \nu)^2 / \{1 + (\kappa^* - \nu)^2\}$ , and  $\nu$  should not exceed  $\kappa^*$ .

**Proof of Theorem 1.** We first show that  $|\hat{\mu}_j^\dagger - \mu_j^\dagger| = O_p(r_n^{\frac{2\gamma}{\gamma+2}} + n^{-1/2})$  and  $\|\hat{\Sigma}_j^\dagger - \Sigma_j^\dagger\| = O_p(r_n^{\frac{2\gamma}{\gamma+2}} + n^{-1/2}), j = \pm 1$ , where  $W(X) = \widehat{W}_1(X) = |\hat{f}(X)|$  and  $\|A\| = \sup_{x \in \mathcal{X}} \|Ax\|_2$ . For the convergence rate in means,

$$|\hat{\mu}_j^\dagger - \mu_j^\dagger| = |\mathbb{P}_n X \hat{f}(X) I\{\hat{d}(X) = j\} - \mathbb{P} X f^*(X) I\{d^*(X) = j\}|$$

$$\begin{aligned}
 &\leq |\mathbb{P}_n X \hat{f}(X) I\{\hat{d}(X) = j\} - \mathbb{P} X \hat{f}(X) I\{\hat{d}(X) = j\}| \\
 &+ |\mathbb{P} X \hat{f}(X) I\{\hat{d}(X) = j\} - \mathbb{P} X f^m(X) I\{\hat{d}(X) = j\}| \\
 &+ |\mathbb{P} X f^m(X) I\{\hat{d}(X) = j\} - \mathbb{P} X f^*(X) I\{\hat{d}(X) = j\}| \\
 &+ |\mathbb{P} X f^*(X) I\{\hat{d}(X) = j\} - \mathbb{P} X f^*(X) I\{d^*(X) = j\}| \\
 &= (I) + (II) + (III) + (IV).
 \end{aligned}$$

We will use McDiarmid’s inequality to bound (I).

$$\begin{aligned}
 &\left| \frac{1}{n} \sum_{k=1}^n x_k \hat{f}(x_k) I\{\hat{d}(x_k) = j\} - \frac{1}{n} \sum_{k=1, k \neq i}^n x_k \check{f}_i(x_k) I\{\check{d}(x_k) = j\} \right. \\
 &\quad \left. - \frac{1}{n} \check{x}_i \check{f}_i(\check{x}_i) I\{\check{d}(\check{x}_i) = j\} \right| \\
 &\leq \left| \frac{1}{n} \sum_{k=1}^n x_k \{\hat{f}(x_k) - \check{f}_i(x_k)\} I\{\hat{d}(x_k) = j\} \right| \\
 &\quad + \left| \frac{1}{n} \sum_{k=1}^n x_k \check{f}_i(x_k) [I\{\hat{d}(x_k) = j\} - I\{\check{d}(x_k) = j\}] \right| \\
 &\quad + \frac{1}{n} \left| (x_i - \check{x}_i) \hat{f}(x_i) I\{\hat{d}(x_i) = j\} + \check{x}_i \{\hat{f}(x_i) - \check{f}_i(\check{x}_i)\} I\{\hat{d}(x_i) = j\} \right. \\
 &\quad \left. + \check{x}_i \check{f}_i(\check{x}_i) [I\{\hat{d}(x_i) = j\} - I\{\check{d}(\check{x}_i) = j\}] \right| \leq M_\rho/n,
 \end{aligned}$$

where  $M_\rho$  depends on  $M, C_X$  and  $\rho$ . The last inequality follows from Assumptions 1, 5 and 6.

By McDiarmid’s inequality,

$$P(|\mathbb{P}_n X \hat{f}(X) I\{\hat{d}(X) = j\} - \mathbb{P} X \hat{f}(X) I\{\hat{d}(X) = j\}| \geq \epsilon) \leq e^{-2n\epsilon^2/M_\rho^2}.$$

Then with probability  $\geq 1 - \delta$ ,

$$|\mathbb{P}_n X \hat{f}(X) I\{\hat{d}(X) = j\} - \mathbb{P} X \hat{f}(X) I\{\hat{d}(X) = j\}| \leq c_M \sqrt{\frac{\log(1/\delta)}{n}}. \quad (7)$$

For (II), by Cauchy-Schwaz inequality and the boundedness of  $X$ ,

$$\begin{aligned}
 &|\mathbb{P} X \hat{f}(X) I\{\hat{d}(X) = j\} - \mathbb{P} X f^m(X) I\{\hat{d}(X) = j\}| \\
 &\leq \sqrt{\mathbb{P}\{\hat{f}(X) - f^m(X)\}^2} \sqrt{\mathbb{P}(X^2)} \\
 &= O_p(r_n),
 \end{aligned}$$

where the last step follows from Assumption 4. For (III),

$$|\mathbb{P} X f^m(X) I\{\hat{d}(X) = j\} - \mathbb{P} X f^*(X) I\{\hat{d}(X) = j\}|$$

$$\begin{aligned} &\leq \sqrt{\mathbb{P}\{f^m(X) - f^*(X)\}^2} \sqrt{\mathbb{P}(X^2)} \\ &\leq M \|f^* - f^m\|_2. \end{aligned}$$

For (IV),

$$\begin{aligned} &|\mathbb{P}X f^*(X)I\{\hat{d}(X) = j\} - \mathbb{P}X f^*(X)I\{d^*(X) = j\}| \\ &= |\mathbb{P}X f^*(X)I\{d^*(X) = j, \hat{d}(X) \neq d^*(X)\}| \\ &\leq |\mathbb{P}X f^*(X)I\{d^*(X) = j\}I\{0 < |f^*(X)| \leq |f^*(X) - \hat{f}(X)|\}| \\ &\leq |\mathbb{P}[X f^*(X)I\{d^*(X) = j\}]\mathbb{P}\{0 < |f^*(X)| \leq |f^*(X) - \hat{f}(X)|\}| \\ &\leq |\mathbb{P}[X f^*(X)I\{d^*(X) = j\}]\mathbb{P}\{I\{0 < |f^*(X)| \leq \epsilon\}I\{|f^*(X) - \hat{f}(X)| \leq \epsilon\} \\ &\quad + I\{|f^*(X) - \hat{f}(X)| > \epsilon\}|\}| \\ &\leq \mu_j^\dagger \left( \frac{\|f^* - \hat{f}\|_2^2}{\epsilon^2} + \epsilon^\gamma \right), \end{aligned}$$

where  $\|f\|_2 = E\{f(x)^2\}^{1/2}$ . Choosing the optimal  $\epsilon$  as  $\|f^* - \hat{f}\|_2^{\frac{2}{\gamma+2}}$ , we have that

$$\begin{aligned} &|\mathbb{P}X f^*(X)I\{\hat{d}(X) = j\} - \mathbb{P}X f^*(X)I\{d^*(X) = j\}| \\ &\leq \mu_j^\dagger \|f^* - \hat{f}\|_2^{\frac{2\gamma}{\gamma+2}} \\ &\leq M \left( r_n^{\frac{2\gamma}{\gamma+2}} + \|f^* - f^m\|_2^{\frac{2\gamma}{\gamma+2}} \right). \tag{8} \end{aligned}$$

Combining the above results, we obtain that with probability  $\geq 1 - \delta$ ,

$$|\hat{\mu}_j^\dagger - \mu_j^\dagger| \leq M \left\{ \|f^* - f^m\|_2^{\min(1, \frac{2\gamma}{\gamma+2})} + r_n^{\min(1, \frac{2\gamma}{\gamma+2})} \right\} + c_M \sqrt{\frac{\log(1/\delta)}{n}}.$$

Therefore,  $|\hat{\mu}_j^\dagger - \mu_j^\dagger|$  is  $O_p(\|f^* - f^m\|_2^{\min(1, \frac{2\gamma}{\gamma+2})} + r_n^{\min(1, \frac{2\gamma}{\gamma+2})} + n^{-1/2})$ .

We then consider the bound on  $\|\hat{\Sigma}_j^\dagger - \Sigma_j^\dagger\|, j = \pm 1$ . Let  $\hat{F}(X) = \sum_{i=1}^n \hat{f}(X_i)I\{\hat{d}(X_i) = j\}$ ,  $F^m(X) = \sum_{i=1}^n f^m(X_i)I\{\hat{d}(X_i) = j\}$  and  $F^*(X) = \sum_{i=1}^n f^*(X_i)I\{\hat{d}(X_i) = j\}$ . With probability  $\geq 1 - \delta$ ,

$$\hat{\Sigma}_j^\dagger = \sum_{i=1}^n \left[ \frac{W(X_i)I\{\hat{d}(X_i) = j\}}{\sum_{i=1}^n W(X_i)I\{\hat{d}(X_i) = j\}} \right]^2 (X_i - \hat{\mu}_j^\dagger)^\top (X_i - \hat{\mu}_j^\dagger).$$

$$\begin{aligned} &\|\hat{\Sigma}_j^\dagger - \Sigma_j^\dagger\| \\ &\leq \left\| (\mathbb{P}_n - \mathbb{P}) \left[ \left\{ \frac{\hat{f}(X)}{\hat{F}(X)} \right\}^2 (X - \hat{\mu}_j^\dagger)^\top (X - \hat{\mu}_j^\dagger) I\{\hat{d}(X) = j\} \right] \right\| \end{aligned}$$

$$\begin{aligned}
 & + \left\| \mathbb{P} \left[ \left\{ \frac{\hat{f}(X)}{\hat{F}(X)} \right\}^2 (X - \hat{\mu}_j^\dagger)^\top (X - \hat{\mu}_j^\dagger) I\{\hat{d}(X) = j\} \right] \right. \\
 & - \left. \mathbb{P} \left[ \left\{ \frac{f^*(X)}{F^*(X)} \right\}^2 (X - \mu_j^\dagger)^\top (X - \mu_j^\dagger) I\{d^*(X) = j\} \right] \right\| \\
 & \leq c_\delta \sqrt{\frac{\log(1/\delta)}{n}} + \left\| \mathbb{P} \left[ \left\{ \frac{\hat{f}(X)}{\hat{F}(X)} \right\}^2 (X - \hat{\mu}_j^\dagger)^\top (X - \hat{\mu}_j^\dagger) I\{\hat{d}(X) = j\} \right] \right. \\
 & - \left. \mathbb{P} \left[ \left\{ \frac{\hat{f}(X)}{\hat{F}(X)} \right\}^2 (X - \mu_j^\dagger)^\top (X - \mu_j^\dagger) I\{\hat{d}(X) = j\} \right] \right\| \\
 & + \left\| \mathbb{P} \left[ \left\{ \frac{\hat{f}(X)}{\hat{F}(X)} \right\}^2 (X - \mu_j^\dagger)^\top (X - \mu_j^\dagger) I\{\hat{d}(X) = j\} \right] \right. \\
 & - \left. \mathbb{P} \left[ \left\{ \frac{\hat{f}(X)}{\hat{F}(X)} \right\}^2 (X - \mu_j^\dagger)^\top (X - \mu_j^\dagger) I\{d^*(X) = j\} \right] \right\| \\
 & + \left\| \mathbb{P} \left( \left[ \left\{ \frac{\hat{f}(X)}{\hat{F}(X)} \right\}^2 - \left\{ \frac{f^*(X)}{F^*(X)} \right\}^2 \right] (X - \mu_j^\dagger)^\top (X - \mu_j^\dagger) I\{d^*(X) = j\} \right) \right\| \\
 & \leq c_\delta \sqrt{\frac{\log(1/\delta)}{n}} + 2 \left\| \mathbb{P} \left[ \left\{ \frac{\hat{f}(X)}{\hat{F}(X)} \right\}^2 (X - \mu_j^\dagger)^\top (\hat{\mu}_j^\dagger - \mu_j^\dagger) I\{\hat{d}(X) = j\} \right] \right\| \\
 & + \left\| \mathbb{P} \left[ \left\{ \frac{\hat{f}(X)}{\hat{F}(X)} \right\}^2 \{(\hat{\mu}_j^\dagger - \mu_j^\dagger)^\top (\hat{\mu}_j^\dagger - \mu_j^\dagger)\} I\{\hat{d}(X) = j\} \right] \right\| \\
 & + \left\| \mathbb{P} \left[ \left\{ \frac{\hat{f}(X)}{\hat{F}(X)} \right\}^2 (X - \mu_j^\dagger)^\top (X - \mu_j^\dagger) I\{d^*(X) = j, \hat{d}(X) \neq d^*(X)\} \right] \right\| \\
 & + \left\| \mathbb{P} \left( \left[ \left\{ \frac{\hat{f}(X)}{\hat{F}(X)} \right\}^2 - \left\{ \frac{f^m(X)}{F^m(X)} \right\}^2 \right] (X - \mu_j^\dagger)^\top (X - \mu_j^\dagger) I\{d^*(X) = j\} \right) \right\| \\
 & + \left\| \mathbb{P} \left( \left[ \left\{ \frac{f^m(X)}{F^m(X)} \right\}^2 - \left\{ \frac{f^*(X)}{F^*(X)} \right\}^2 \right] (X - \mu_j^\dagger)^\top (X - \mu_j^\dagger) I\{d^*(X) = j\} \right) \right\| \\
 & \leq O_p(\|f^* - f^m\|_2^{\min(1, \frac{2\gamma}{\gamma+2})} + r_n^{\min(1, \frac{2\gamma}{\gamma+2})} + n^{-1/2}).
 \end{aligned}$$

Now we prove Theorem 1’s result. Define

$$\begin{aligned}
 M(u) & = \sqrt{(\beta_{10} + Fu)^\top (\Sigma_1^\dagger + \rho_1 I_p) (\beta_{10} + Fu)} \\
 & + \sqrt{(\beta_{10} + Fu)^\top (\Sigma_{-1}^\dagger + \rho_{-1} I_p) (\beta_{10} + Fu)},
 \end{aligned}$$

and

$$M_n(u) = \sqrt{(\hat{\beta}_{10} + \hat{F}u)^\top (\hat{\Sigma}_1^\dagger + \rho_1 I_p) (\hat{\beta}_{10} + \hat{F}u)} + \sqrt{(\hat{\beta}_{10} + Fu)^\top (\hat{\Sigma}_{-1}^\dagger + \rho_{-1} I_p) (\hat{\beta}_{10} + \hat{F}u)},$$

where we use the estimated  $\hat{\Sigma}_{\pm 1}, \hat{F}$  and  $\hat{\beta}_{10}$  in  $M(u)$ . Since  $u$  belongs to a compact set,  $\hat{\beta}_{10} \rightarrow \beta_{10}, \hat{\Sigma}_j \rightarrow \Sigma_j$  and  $\hat{F} \rightarrow F$ ,  $\sup_u |M_n(u) - M(u)| \rightarrow 0$  in probability. Given that  $M(u)$  has a unique maximizer and  $u$  is in a compact set, we have  $\sup_{\|u-u^*\| \geq \epsilon} M(u) < M(u^*)$ . Hence the conditions of Theorem 5.7 of van der Vaart (1998) are satisfied, and it follows that  $\hat{u} \rightarrow u^*$  in probability, and subsequently  $\hat{\beta}_1 \rightarrow \beta_1^*$  in probability.

We now show the convergence rate of  $\hat{\beta}_1$  to  $\beta_1^*$ . Given that  $u \mapsto M(u)$  is twice differentiable at  $u^*$ , we have  $M(u) - M(u^*) \lesssim -c\|u - u^*\|^2$  for all  $u$  in the neighborhood of  $u^*$  and some  $c > 0$ , where  $u^*$  is the unique maximizer of  $M(u)$ . Since  $|\hat{\mu}_j - \mu_j| = O_p(O_p(\|f^* - f^m\|_2^{\min(1, \frac{2\gamma}{\gamma+2})} + r_n^{\min(1, \frac{2\gamma}{\gamma+2})} + n^{-1/2}))$  and  $\|\hat{\Sigma}_j - \Sigma_j\| = O_p(\|f^* - f^m\|_2^{\min(1, \frac{2\gamma}{\gamma+2})} + r_n^{\min(1, \frac{2\gamma}{\gamma+2})} + n^{-1/2})$ ,

$$\begin{aligned} & |(M_n - M)(u) - (M_n - M)(u^*)| \\ &= |\{M_n(u) - M_n(u^*)\} - \{M(u) - M(u^*)\}| \\ &\lesssim \left| \frac{\partial M_n(u)}{\partial u} - \frac{\partial M(u)}{\partial u} \right|_{u=u^*} \|u - u^*\| \\ &\approx (\|f^* - f^m\|_2^{\min(1, \frac{2\gamma}{\gamma+2})} + r_n^{\min(1, \frac{2\gamma}{\gamma+2})} + n^{-1/2}) \|u - u^*\|. \end{aligned}$$

Thus,

$$\begin{aligned} & E^* \sup_{\|u-u^*\| < \delta} |(M_n - M)(u) - (M_n - M)(u^*)| \\ &\lesssim (\|f^* - f^m\|_2^{\min(1, \frac{2\gamma}{\gamma+2})} + r_n^{\min(1, \frac{2\gamma}{\gamma+2})} + n^{-1/2}) \delta. \end{aligned}$$

When  $\delta \mapsto \phi(\delta) = \sqrt{n}(\|f^* - f^m\|_2^{\min(1, \frac{2\gamma}{\gamma+2})} + r_n^{\min(1, \frac{2\gamma}{\gamma+2})} + n^{-1/2})\delta$ ,  $\phi(\delta)/\delta^\eta$  is decreasing for any  $\eta \in (1, 2)$ . In addition,  $\hat{u} \rightarrow u^*$  in probability, and  $\hat{u}$  maximize  $M_n(u)$ . The conditions of Theorem 14.4 in Kosorok (2008) are satisfied with  $\phi(\delta) = \sqrt{n}(\|f^* - f^m\|_2^{\min(1, \frac{2\gamma}{\gamma+2})} + r_n^{\min(1, \frac{2\gamma}{\gamma+2})} + n^{-1/2})\delta$ . Hence,  $\|u - u^*\| = O_p(\|f^* - f^m\|_2^{\min(1, \frac{2\gamma}{\gamma+2})} + r_n^{\min(1, \frac{2\gamma}{\gamma+2})} + n^{-1/2})$ , and the desired result follows.

### Acknowledgments

The authors gratefully acknowledge support by R01DK108073, R01GM124104, U10CA180819, P30 CA015704, S10 OD020069 awarded by the National Institutes of Health, and Coltman Early Career Fellowship awarded by Hope Foundation.

## Supplementary Material

### Supplementary Materials for “Robustifying Trial-Derived Optimal Treatment Rules for A Target Population”

(doi: [10.1214/19-EJS1540SUPP](https://doi.org/10.1214/19-EJS1540SUPP); .pdf).

## References

- Audibert, J.-Y., Tsybakov, A. B., et al. “Fast learning rates for plug-in classifiers.” *The Annals of statistics*, 35(2):608–633 (2007). [MR2336861](#)
- Biau, G. “Analysis of a random forests model.” *Journal of Machine Learning Research*, 13(Apr):1063–1095 (2012). [MR2930634](#)
- Breiman, L. “Random forests.” *Machine learning*, 45(1):5–32 (2001). [MR3874153](#)
- Brinkley, J., Tsiatis, A., and Anstrom, K. J. “A generalized estimator of the attributable benefit of an optimal treatment regime.” *Biometrics*, 66(2):512–522 (2010). [MR2758831](#)
- Buchanan, A. L., Hudgens, M. G., Cole, S. R., Mollan, K., Sax, P. E., Daar, E., Adimora, A. A., Eron, J., and Mugavero, M. “Generalizing evidence from randomized trials using inverse probability of sampling weights.” *Pharmacy Practice Faculty Publications*, Paper 77 (2016).
- Cole, S. R. and Stuart, E. A. “Generalizing Evidence From Randomized Clinical Trials to Target Populations The ACTG 320 Trial.” *American journal of epidemiology*, 172(1):107–115 (2010).
- Devroye, L., Györfi, L., and Lugosi, G. *A probabilistic theory of pattern recognition*, volume 31. Springer Science & Business Media (2013). [MR1383093](#)
- Goldberg, Y. and Kosorok, M. R. “Q-learning with censored data.” *Annals of statistics*, 40(1):529 (2012). [MR3014316](#)
- Greenland, S. “Randomization, statistics, and causal inference.” *Epidemiology*, 1(6):421–429 (1990).
- Hamburg, M. A. and Collins, F. S. “The path to personalized medicine.” *New England Journal of Medicine*, 363(4):301–304 (2010).
- Hastie, T., Tibshirani, R., and Friedman, J. H. *The Elements of Statistical Learning*. New York: Springer-Verlag New York, Inc., second edition (2009). [MR2722294](#)
- Hayes, D. F., Thor, A. D., Dressler, L. G., Weaver, D., Edgerton, S., Cowan, D., Broadwater, G., Goldstein, L. J., Martino, S., Ingle, J. N., et al. “HER2 and response to paclitaxel in node-positive breast cancer.” *New England Journal of Medicine*, 357(15):1496–1506 (2007).
- Hernán, M. A. and Robins, J. M. “Estimating causal effects from epidemiological data.” *Journal of epidemiology and community health*, 60(7):578–586 (2006).
- Huang, X., Ning, J., and Wahed, A. S. “Optimization of individualized dynamic treatment regimes for recurrent diseases.” *Statistics in medicine*, 33(14):2363–2378 (2014). [MR3256672](#)
- Ishwaran, H., Kogalur, U. B., Blackstone, E. H., and Lauer, M. S. “Random survival forests.” *The annals of applied statistics*, 2:841–860 (2008). [MR2516796](#)

- Kang, C., Janes, H., and Huang, Y. “Combining biomarkers to optimize patient treatment recommendations.” *Biometrics*, 70(3):695–707 (2014). [MR3261788](#)
- Kosorok, M. R. *Introduction to empirical processes and semiparametric inference*. New York: Springer-Verlag (2008). [MR2724368](#)
- Lanckriet, G. R., Ghaoui, L. E., Bhattacharyya, C., and Jordan, M. I. “A robust minimax approach to classification.” *The Journal of Machine Learning Research*, 3:555–582 (2003). [MR1991086](#)
- Lara, P. N., Ely, B., Quinn, D. I., Mack, P. C., Tangen, C., Gertz, E., Twardowski, P. W., Goldkorn, A., Hussain, M., Vogelzang, N. J., et al. “Serum biomarkers of bone metabolism in castration-resistant prostate cancer patients with skeletal metastases: results from SWOG 0421.” *JNCI: Journal of the National Cancer Institute*, 106(4) (2014).
- Li, K.-C., Wang, J.-L., Chen, C.-H., et al. “Dimension reduction for censored regression data.” *The Annals of Statistics*, 27(1):1–23 (1999). [MR1701098](#)
- Marshall, A. W. and Olkin, I. “Multivariate chebyshev inequalities.” *The Annals of Mathematical Statistics*, 31(4):1001–1014 (1960). [MR0119234](#)
- Orellana, L., Rotnitzky, A., and Robins, J. M. “Dynamic regime marginal structural mean models for estimation of optimal dynamic treatment regimes, part I: main content.” *The International Journal of Biostatistics*, 6(2) (2010). [MR2602551](#)
- Qian, M. and Murphy, S. A. “Performance Guarantees for Individualized Treatment Rules.” *The Annals of Statistics*, 39:1180–1210 (2011). [MR2816351](#)
- Quinn, D. I., Tangen, C. M., Hussain, M., Lara, P. N., Goldkorn, A., Moynour, C. M., Garzotto, M. G., Mack, P. C., Carducci, M. A., Monk, J. P., et al. “Docetaxel and atrasentan versus docetaxel and placebo for men with advanced castration-resistant prostate cancer (SWOG S0421): a randomised phase 3 trial.” *The Lancet Oncology*, 14(9):893–900 (2013).
- Robins, J. M., Hernán, M. Á., and Brumback, B. “Marginal structural models and causal inference in epidemiology.” *Epidemiology*, 11(5):550–560 (2000). [MR1766776](#)
- Rubin, D. B. “Bayesian Inference for Causal Effects: The Role of Randomization.” *The Annals of Statistics*, 6:34–58 (1978). [MR0472152](#)
- Tsybakov, A. B. “Optimal aggregation of classifiers in statistical learning.” *Annals of Statistics*, 32:135–166 (2004). [MR2051002](#)
- van der Vaart, A. *Asymptotic Statistics*. New York: Cambridge University Press (1998). [MR1652247](#)
- Vapnik, V., Golowich, S. E., and Smola, A. “Support vector method for function approximation, regression estimation, and signal processing.” *Advances in neural information processing systems*, 281–287 (1997).
- Zhang, B., Tsiatis, A. A., Laber, E. B., and Davidian, M. “A robust method for estimating optimal treatment regimes.” *Biometrics*, 68(4):1010–1018 (2012). [MR3040007](#)
- Zhao, Y.-Q., Zeng, D., Laber, E. B., Song, R., Yuan, M., and Kosorok, M. R. “Doubly robust learning for estimating individualized treatment with censored data.” *Biometrika*, 102(1):151–168 (2015). [MR3335102](#)
- Zhao, Y. Q., Zeng, D., Rush, A. J., and Kosorok, M. R. “Estimating Indi-

vidualized Treatment Rules using Outcome Weighted Learning.” *Journal of American Statistical Association*, 107:1106–1118 (2012). [MR3010898](#)  
Zhao, Y. Q., Zeng, D., Tangen, C. M., and LeBlanc, M. L. Supplementary Materials for “Robustifying Trial-Derived Optimal Treatment Rules for A Target Population”, (2019) DOI: [10.1214/19-EJS1540SUPP](#)