

University of Michigan School of Public Health

The University of Michigan Department of Biostatistics Working
Paper Series

Year 2016

Paper 120

Variable Selection for Estimating the Optimal Treatment Regimes in the Presence of a Large Number of Covariate

Baqun Zhang*

Min Zhang[†]

*School of Statistics, Renmin University, zhangbaqun@ruc.edu.cn

[†]University of Michigan - Ann Arbor, mzhangst@umich.edu

This working paper is hosted by The Berkeley Electronic Press (bepress) and may not be commercially reproduced without the permission of the copyright holder.

<http://biostats.bepress.com/umichbiostat/paper120>

Copyright ©2016 by the authors.

Variable Selection for Estimating the Optimal Treatment Regimes in the Presence of a Large Number of Covariate

Baqun Zhang and Min Zhang

Abstract

Most of existing methods for optimal treatment regimes, with few exceptions, focus on estimation and are not designed for variable selection with the objective of optimizing treatment decisions. In clinical trials and observational studies, often numerous baseline variables are collected and variable selection is essential for deriving reliable optimal treatment regimes. Although many variable selection methods exist, they mostly focus on selecting variables that are important for prediction (predictive variables) instead of variables that have a qualitative interaction with treatment (prescriptive variables) and hence are important for making treatment decisions. We propose a variable selection method within a general classification framework to select prescriptive variables and estimate the optimal treatment regime simultaneously. In this framework, an optimal treatment regime is equivalently defined as the one that minimizes a weighted misclassification error rate and the proposed method forward sequentially select prescriptive variables by minimizing this weighted misclassification error. A main advantage of this method is that it specifically targets selection of prescriptive variables and in the meantime is able to exploit predictive variables to improve performance. The method can be applied to both single- and multiple- decision point setting. The performance of the proposed method is evaluated by simulation studies and application to an clinical trial.

Variable selection for estimating the optimal treatment regimes in the presence of a large number of covariate

Baqun Zhang

School of Statistics, Renmin University, Beijing, R.R. China

and

Min Zhang

Department of Biostatistics, University of Michigan, Ann Arbor

August 22, 2016

Abstract

Most of existing methods for optimal treatment regimes, with few exceptions, focus on estimation and are not designed for variable selection with the objective of optimizing treatment decisions. In clinical trials and observational studies, often numerous baseline variables are collected and variable selection is essential for deriving reliable optimal treatment regimes. Although many variable selection methods exist, they mostly focus on selecting variables that are important for prediction (predictive variables) instead of variables that have a qualitative interaction with treatment (prescriptive variables) and hence are important for making treatment decisions. We propose a variable selection method within a general classification framework to select prescriptive variables and estimate the optimal treatment regime simultaneously. In this framework, an optimal treatment regime is equivalently defined as the one that

minimizes a weighted misclassification error rate and the proposed method forward sequentially select prescriptive variables by minimizing this weighted misclassification error. A main advantage of this method is that it specifically targets selection of prescriptive variables and in the meantime is able to exploit predictive variables to improve performance. The method can be applied to both single- and multiple-decision point setting. The performance of the proposed method is evaluated by simulation studies and application to an clinical trial.

1 Introduction

Personalized medicine that explicitly recognizes individual heterogeneity in response to treatments and focuses on making treatment decisions for a patient based on his/her own characteristics (eg., demographic, clinical, genetic information, etc.) has received much attention recently. The idea of personalized medicine can be formalized using the concept of treatment regimes, which are one or a sequence of decision rules that specify which treatment (among available options) a given subject receives based on a subject's characteristics at the time of the decision. In the last decade, there has been increasing interest and more vigorous research on developing methodologies for estimating the optimal treatment regimes (Murphy, 2003; Robins, 2004; Moodie, et al., 2007; Robins, et al., 2008; Brinkley, et al., 2009; Qian and Murphy, 2011; Chakraborty et al., 2010; Zhang et al., 2012ab, 2013; Zhao et al., 2012 and 2015; Geng et al., 2015; Barrett et al., 2014; Young et al., 2011; Tian, et al., 2014).

Most of existing methods, with few exceptions, focus on estimation and are not designed for selecting important variables from among a large number of covariates for optimizing treatment decisions. Clinical trials and observational studies (e.g., clinical registries), on

which estimation of the optimal treatment regime is based, often collect a large amount of potentially useful patient information. Although it is likely that many of those variables are useful for predicting outcomes, realistically perhaps only a small number of patient characteristics are useful in making treatment decisions since only those variables with a qualitative interaction with treatment are useful in making treatment decisions. The importance of variables that have qualitative interactions with treatments in medical decision-making setting has been noted previously (Peto, 1982) and are referred to as prescriptive variables. Therefore, variable selection from a high dimensional set of covariates targeted towards optimal decision making is an essential step in constructing a meaningful and practically useful treatment decision rule.

Variable selection has been an active research area in statistics; however, as pointed out by Gunter, Zhu and Murphy (2011), current variable selection work has been focused on prediction and their use in decision making has not been well developed and tested. As a matter of fact, variable selection approaches focused on prediction may neglect variables vital for decision making since the effect of interactions is often weaker than that of the main effect. Fairly recent literature started to see more research on variable selection methods for making treatment decisions (Gunter, Zhu and Murphy, 2011; Qian and Murphy, 2011; Lu, Zhang and Zeng, 2013; Fan, Lu and Song, 2015). Penalized least squares methods were proposed in the framework of Q-learning by Qian and Murphy (2011) and in the framework of A-learning by Lu, Zhang and Zeng (2013) to select important variables in the corresponding outcome regression, leading to estimated regimes with fewer variables. However, the variable selection in the two penalized methods are not directly targeted towards selecting prescriptive variables. Gunter, Zhu and Murphy (2011) proposed a variable selection ranking method for variable selection, where the ranking is based a measure that specifically characterizes the qualitative interaction of a variable with treatment. As

a result, the method of Gunter, Zhu and Murphy (2011) focuses specifically on selection of prescriptive variables. As noted by Fan, Lu and Song (2015), since the ranking method considers each variable separately and ignores correlations between covariates, they may identify too many covariates as potential prescriptive variables or miss some variables important for decision making. Building upon the work of Gunter, Zhu and Murphy (2011), Fan, Lu and Song (2015) proposed a sequential advantage selection method which takes into account variables already selected in previous steps and assesses the additional value of a new variable instead of considering variables individually. One advantage of this method is that it avoids selecting variables that are marginally important for decision making but are jointly unimportant. Although not directly focused on estimating optimal treatment regimes, one other recent relevant work is that of Tian, et al, (2014), which considers estimating interactions between treatment and a large number of covariates.

In this paper, we propose a new method to select important prescriptive variables for estimating the optimal treatment regimes in a classification framework. This method is motivated by targeting the objective function of the optimal treatment regime, ie., the expectation of potential outcomes under the optimal regime if it is followed by the entire population. We show that optimizing the objective function is equivalent to minimizing a weighted misclassification error rate for classifying patients to classes corresponding to their optimal treatment. The weighted misclassification error is defined in terms of contrast between treatments given covariates and hence directly targets on selection of variables with qualitative interaction with treatments. Our proposed algorithm is then based on forward sequentially minimizing the weighted misclassification error rate and, as Fan, Lu and Song (2015), in each step it takes into account previously selected variables. The performance and merit of the proposed method relative to the sequential advantage selection method of Fan, Lu and Song (2015) are evaluated by various simulation studies.

The remainder of the paper is organized as follows. In Section 2, we describe the framework and objective function for variable selection for optimal treatment regimes and propose a forward minimal misclassification error selection algorithm that selects variables important for decision making. We evaluate the performance of the proposed method by simulations studies in Section 3 and illustrate the use of the proposed method using data from the Nefazodone CBASP trial in Section 4, followed by a discussion in Section 5.

2 Method

2.1 Notation and Assumptions

We first focus on presenting the method in the simpler setting where only a single treatment decision point is involved; Section 2.4 considers the extension to multiple decision point setting. Consider a clinical trial or observational study involving n subjects, who receive either treatment $A = 0$ or $A = 1$. Let X be a p -dimensional vector of subject characteristics collected before the treatment. Let Y denote the observed outcome of interest and, without loss of generality, assume that larger values of Y are preferred. The observed data are then (X_i, A_i, Y_i) , $i = 1, \dots, n$, which are assumed to be independent and identically distributed (i.i.d.) across i . The goal is to use the data to find the optimal treatment decision rule which determines which treatment a patient should receive based on his/her baseline characteristics.

Formally, a treatment regime or rule, g , is a function which maps the values of X to the domain of A , eg, $\mathcal{A} = \{0, 1\}$. Let $Y^*(0)$ and $Y^*(1)$ denote the potential outcomes for a subject that would be observed had the subject received treatment 0 or 1, respectively. Then for each treatment regime g , there is a corresponding potential outcome, which is

defined as $Y^*(g) = Y^*(1)g(X) + Y^*(0)\{1 - g(X)\}$. The optimal treatment regime g^{opt} is the one that leads to the optimal expected potential outcomes if followed by the entire population; i.e., $g^{opt} = \arg \max_g E\{Y^*(g)\}$. We make the commonly assumed stable unit treatment value assumption (SUTVA), which states that the observed outcome is the same as the potential outcome under the treatment actually received; i.e., $Y = Y^*(1)A + Y^*(0)(1 - A)$. This assumption allows identification of the optimal treatment regime based on the observed data. We also assume the standard no unmeasured confounders assumption, i.e., $\{Y^*(0), Y^*(1)\} \perp\!\!\!\perp A | X$, where $\perp\!\!\!\perp$ denotes statistical independence. This assumption holds automatically for clinical trials and has to be evaluated for observational studies. See a review paper by Schulte, et al. (2014) for more background on potential outcomes and optimal treatment regimes.

2.2 Framework for Variable Selection for Optimal Treatment Regimes

By definition, one should target optimizing $E\{Y^*(g)\}$ for estimating the optimal treatment regime and indeed all existing methods estimate and optimize $E\{Y^*(g)\}$ directly or indirectly. Naturally, one should also aim to maximize $E\{Y^*(g)\}$ in selecting prescriptive variables for constructing the optimal treatment regime. However, how to achieve this is much more challenging and to the best of our knowledge none of the existing prescriptive variable selection methods directly target this objective function. We present a framework for estimating the optimal treatment regime and prescriptive variable selection by explicitly optimizing $E\{Y^*(g)\}$.

Intuitively, it is easy to see that we can view subjects as coming from two latent classes corresponding to $g^{opt}(X) = 0$ or 1 ; i.e., the class corresponding to $g^{opt}(X) = a$ includes all subjects whose expected potential outcome under treatment a is greater than that under treatment $1 - a$. Equivalently, only the contrast between treatments is relevant in

determining which class a subject belongs to as well as whether a variable is prescriptive or not. Zhang, et al. (2012b) formalized these ideas and proposed to estimate the optimal treatment regime using a classification framework. We refer to this general classification framework as C-learning with “C” stands for “classification”. Denoting $\mu(a, X) = E(Y|A = a, X)$, it can be shown that

$$E\{Y^*(g)\} = E_X\{C(X)g(X) + \mu(0, X)\}, \quad (1)$$

where $C(X) = \mu(1, X) - \mu(0, X)$, denoting the contrast between treatment 1 and 0. According to (1), it is clear that the optimal treatment regime corresponds to $I\{C(X) > 0\}$. It is also clear from (1) that $g^{opt}(X)$ maximizes $E_X\{C(X)g(X)\}$. In the classification perspective of Zhang, et al. (2012b), it separates the information in $C(X)$ to two components, $Z \equiv I\{C(X) > 0\}$ and $|C(X)|$, which are the sign and magnitude of $C(X)$ respectively. Zhang, et al. (2012b) show that $C(X)g(X) = Z|C(X)| - |C(X)|I\{Z \neq g(X)\}$ and as a result $g^{opt}(X)$ can be defined alternatively as

$$g^{opt}(X) = \arg \min E[|C(X)|I\{Z \neq g(X)\}]. \quad (2)$$

This alternative definition corresponds exactly to the intuition described above. That is, we can view each subject as belonging to one of two latent classes with the class label denoted by $Z = I\{C(X) > 0\}$. We can then view $E[|C(X)|I\{Z \neq g(X)\}]$ as a weighted misclassification error rate corresponding to treatment regime (or classifier) $g(X)$. That is, if $g(X) \neq Z$, then an error is made since the treatment decision according to $g(X)$ is not optimal and the loss corresponding to this error is $|C(X)|$, the difference in expected outcomes between $g(X)$ and $g^{opt}(X)$. In this sense, the optimal treatment regime minimizes a weighted misclassification error rate and the interpretation of the classification is consistent with the goal of optimal treatment decisions. In practice, the class label Z as

well as the weight $|C(X)|$ is unknown and Zhang, et al. (2012b) discussed various ways to estimate $C(X)$ and then Z . Denoting an estimate of $C(X)$ by $\widehat{C}(X)$, then $g^{opt}(X)$ can be estimated by $\arg \min \sum_{i=1}^n |\widehat{C}(X_i)| I\{\widehat{Z}_i \neq g(X_i)\}$, where $\widehat{Z}_i = I\{\widehat{C}(X_i) > 0\}$.

It may seem that the definition (2) and the resulting C-learning are unnecessarily complicated since by (1) one can directly estimate $g^{opt}(X)$ by $I\{\widehat{C}(X) > 0\}$, whereas the C-learning involves an additional step of optimization after obtaining $\widehat{C}(X)$. As a matter of fact, this alternative classification perspective has several advantages and is the key to variable selection for prescriptive variables. Among them, one advantage is that it leads to a direct optimization method, as opposed to outcome-regression based methods, that optimizes $E\{Y^*(g)\}$ or $E[|C(X)|I\{Z \neq g(X)\}]$ explicitly and as a result the performance of the estimated treatment regimes is not completely determined by outcome regression models used for estimating $C(X)$ and is more robust to model misspecification. See Zhang, et al., (2012a and b) for discussions on advantages of direct optimization methods and of the classification perspective in (2). The advantage of direct optimization method is also discussed in, for example, Kang et al.(2014) and discussion papers. Equally important, (2) provides a natural objective function for prescriptive variable selection for estimating the optimal treatment regime since it only depends on the contrast function, the part relevant for determining whether a variable is prescriptive or not.

2.3 Forward Minimal Misclassification Error Rate (ForMMER) Selection

In C-learning, the estimated weighted misclassification error rate corresponding to regime $g(X)$ is given by

$$\frac{1}{n} \sum_{i=1}^n [\widehat{W}_i I\{\widehat{Z}_i \neq g(X_i)\}], \quad (3)$$

where $\widehat{Z}_i = I\{\widehat{C}(X_i) > 0\}$, $\widehat{W}_i = |\widehat{C}(X_i)|$ and $\widehat{C}(X_i)$ is an estimate of $C(X_i)$. Various methods can be used for estimating $C(X)$; see Zhang, et al (2012b). For example, one can postulate parametric models for $\mu(a, X)$ and estimate $C(X)$ by $\widehat{C}_{reg}(X_i) \equiv \widehat{\mu}(1, X_i) - \widehat{\mu}(0, X_i)$, where $\widehat{\mu}(a, X)$ estimates $\mu(a, X)$ based on the fitted model. Alternatively, one can estimate $C(X_i)$ by the AIPWE estimator studied by Zhang, et al (2012b), ie,

$$\widehat{C}_{AIPWE}(X_i) = \frac{A_i}{\widehat{\pi}_i} Y_i - \frac{A_i - \widehat{\pi}_i}{\widehat{\pi}_i} \widehat{\mu}(1, X_i) - \left\{ \frac{1 - A_i}{1 - \widehat{\pi}_i} Y_i - \frac{\widehat{\pi}_i - A_i}{1 - \widehat{\pi}_i} \widehat{\mu}(0, X_i) \right\}, \quad (4)$$

where $\widehat{\pi}_i$ estimates $\pi_i = P(A = 1|X_i)$ and is simply the sample proportion corresponding to $A = 1$ for a randomized clinical trial. It can be shown that (3) coupled with $\widehat{C}_{AIPWE}(X)$ consistently estimates $E[|C(X)|I\{Z \neq g(X)\}]$ in (2) in a randomized study regardless of whether the postulated model for $\mu(a, X)$ is correct or not. This robustness is due to the double robustness property of the AIPWE and randomization.

This weighted misclassification error rate is for a given regime $g(X) \equiv g(X_1, \dots, X_p)$. Now we use this to define a measure that is helpful for quantifying the importance of a potential prescriptive variable given a set of already selected prescriptive variables. For that, we define the weighted misclassification error rate corresponding to a set of variables $\{X_{j^1}, \dots, X_{j^m}\}$ as

$$err(X_{j^1}, \dots, X_{j^m}) = \min_{\beta = \{\beta_0, \dots, \beta_m\}} n^{-1} \sum_{i=1}^n \widehat{W}_i I\{\widehat{Z}_i \neq I(\beta_0 + \beta_1 X_{j^1} + \dots + \beta_m X_{j^m} > 0)\},$$

which can be interpreted as the minimum weighted misclassification error rate among a subclass of regimes that are constructed by linear combinations of the set of variables $(X_{j^1}, \dots, X_{j^m})$. Then naturally we can quantify the importance of a potential prescriptive variable, say X_j , given a set of selected prescriptive variables $\{X_{j^1}, \dots, X_{j^m}\}$ by the difference in misclassification error, i.e., $err(X_{j^1}, \dots, X_{j^m}) - err(X_{j^1}, \dots, X_{j^m}, X_j)$.

Based on the idea described above, we propose the following forward minimal misclassification error rate (ForMMER) selection algorithm to sequentially select variables that

are important for treatment decision making. Our algorithm starts with an empty set corresponding to the case where there is no prescriptive variable and hence the optimal treatment is a fixed treatment for everyone.

Step 1 (Initial Step). Let

$$err^{(0)} \equiv err(\text{null set}) = \min\left\{n^{-1} \sum_{i=1}^n \widehat{W}_i I(\widehat{Z}_i \neq 1), n^{-1} \sum_{i=1}^n \widehat{W}_i I(\widehat{Z}_i \neq 0)\right\},$$

where the equality is due to that there are only two treatment regimes ($a = 0, 1$) when the set of covariates under consideration is null. Here $err^{(0)}$ is the weighted misclassification error rate by assigning the treatment with better average treatment effect to all patients regardless of their characteristics. We term this as the baseline weighted misclassification error rate and use it as a reference in the criterion for the initial selection that selects the first important prescriptive variable.

When the number of candidate variables is huge, it is preferable to initially screen variables for consideration in subsequent steps. For each X_j , let $err(X_j) = \min_{\beta=\{\beta_0, \beta_1\}} n^{-1} \sum_{i=1}^n \widehat{W}_i I\{\widehat{Z}_i \neq I(\beta_0 + \beta_1 X_j > 0)\}$ and calculate

$$err^{(0)} - err(X_j), j = 1, \dots, p.$$

As explained above, this difference characterizes the degree of reduction in weighted misclassification error rate under the optimal treatment regime within a subclass of regimes based on variable X_j , relative to the optimal treatment regime based on a null set of covariates. Therefore, the ranking (from the largest to the lowest) of $err^{(0)} - err(X_j), j = 1, \dots, p$, quantifies the relative importance of variables in treatment decision making. We propose to use the ranking of $err^{(0)} - err(X_j)$ to initially select the set of covariates considered in subsequent steps when the dimension of covariates is high. For example, one may consider the first 30 or 40 variables based on the ranking of $err^{(0)} - err(X_j)$ and include them into

a set \mathcal{F}^0 , which include variables considered in subsequent steps. In addition, one may add into \mathcal{F}^0 any variables that are thought to be potentially important in making treatment decisions based on clinical/scientific reasons or on evidence from empirical data. The purpose of this step is only to screen variables when the number of candidate variables is very large to make it computationally easier; for example, in our simulations we considered 1000 covariates. When the dimension is not super large, for example, in our real data analysis there are only 50 baseline covariates, then the screening step can be omitted, i.e., all candidate variables are included in \mathcal{F}^0 .

We comment that, even with the screen, the proposed variable selection method is still selecting prescriptive variables from the entire p -dimensional covariates instead of only selecting among variables in \mathcal{F}^0 . The reason is that in the first step in selecting X_{j^1} (as is clear from the formula below), it is selecting among all (X_1, \dots, X_p) using a principled approach based on importance of variables. Our simulations demonstrate that it works well in practice even with a large number of candidate variables.

Step 2 (Forward Selection). Let

$$X_{j^1} = \arg \min_{X_j \in (X_1, \dots, X_p)} \text{err}(X_j), \text{ and } \text{err}^{(1)} \equiv \min_{X_j \in (X_1, \dots, X_p)} \text{err}(X_j).$$

Then X_{j^1} is the first selected variable and $\mathcal{S}^{(1)} = \{X_{j^1}\}$, denoting the set of selected variables from step (1). We note that the selected X_{j^1} is always included in \mathcal{F}^0 as \mathcal{F}^0 includes variables ranked high based on $\text{err}^{(0)} - \text{err}(X_j)$ and hence X_{j^1} .

In the m -th step ($m > 1$), we have $\mathcal{S}^{(m-1)} = \{X_{j^1}, \dots, X_{j^{m-1}}\}$, which denotes the set of selected variables in steps prior to the m -th step. For every $X_j \in \mathcal{F}^0 \setminus \mathcal{S}^{(m-1)}$, we compute each $\text{err}(\mathcal{S}^{(m-1)}, X_j)$, which is the minimum weighted misclassification error rate for regimes constructed using variables in $\mathcal{S}^{(m-1)}$ and X_j . The m -th variable to be selected is the one

with the smallest weighted misclassification error rate in this step, i.e.

$$X_{j^m} = \arg \min_{X_j \in \mathcal{F}^0 \setminus \mathcal{S}^{(m-1)}} \text{err}(\mathcal{S}^{(m-1)}, X_j).$$

We update the set of selected variables, i.e., $\mathcal{S}^{(m)} = \mathcal{S}^{(m-1)} \cup \{X_{j^m}\}$. weighted misclassification error rate corresponding to the optimal treatment regime among regimes that are based on the m variables in $\mathcal{S}^{(m)}$ is also updated accordingly as follows,

$$\text{err}^{(m)} \equiv \min_{X_j \in \mathcal{F}^0 \setminus \mathcal{S}^{(m-1)}} \text{err}(\mathcal{S}^{(m-1)}, X_j).$$

Step 3 (Stopping Criterion). Continue forward selection until $\text{prop}^{(m)} \leq \alpha$, where α is a pre-specified cut-off point and

$$\text{prop}^{(m)} = \frac{\text{err}^{(m-1)} - \text{err}^{(m)}}{\text{err}^{(m-1)}}.$$

In our simulations as well as the data analysis, the optimization is implemented using a genetic algorithm discussed by Goldberg (1989), implemented in the `rgenoud` package in R (Mebane and Sekhon, 2011). As well as the sequential advantage selection (SAS) method of Fan, Lu and Song (2015), the proposed ForMMER algorithm sequentially select potential prescriptive variables by assessing the added advantage of a new variable relative to existing ones, in contrast to the S-score based method of Gunter, Zhu and Murphy (2011) that considers each variable individually. Therefore, ForMMER enjoys the same advantage as SAS, namely, it tends not to select those unnecessary variables that are only marginally important but not important given other variables. One key difference between ForMMER and SAS lies in the function used in quantifying the advantage. In SAS, the sequential advantage of a variable, say X_j , given a set of selected prescriptive variables $S^{m-1} = \{X_{j^1}, \dots, X_{j^{m-1}}\}$, is given by

$$\frac{1}{n} \sum_{i=1}^n \left\{ \max_a \widehat{E}(Y_i | S_i^{m-1}, X_{ij}, A_i = a) - \widehat{E}(Y_i | S_i^{m-1}, X_{ij}, A_i = a_{opt}(S_i^{m-1})) \right\}, \quad (5)$$

where $a_{opt}(S_i^{m-1})$ is the optimal decision based on variables in S_i^{m-1} . This sequential advantage extends the S-score of Gunter, Zhu and Murphy (2011) in that the second term of the sequential advantage additionally conditions on S^{k-1} , whereas the second term of S-score conditions on X_j only. Subject i contributes to the sequential advantage only if the optimal decision based on (S_i^{m-1}, X_{ij}) is different from the optimal decision based on S_i^{m-1} and hence the sequential advantage quantifies the importance of X_j in addition to S^{m-1} for decision making. At each step in the sequential advantage selection, it fits a model conditional on potential prescriptive variables selected in previous steps and a new variable. As a result of the SAS algorithm, variables with only main effect but no qualitative interactive effect tend to be not selected, which of course is an intended property; however, since at each step it builds conditional models conditional on only variables selected in previous steps, those only predictive but not prescriptive variables will not be able to be used in the outcome-regression models in subsequent steps, which is clearly not desirable as it misses the chance of exploiting those predictive variables to improve performance. In the proposed ForMMER method, the forward selection algorithm for selecting prescriptive variables are separated from estimation of the contrast function. In principle any model selection methods developed for prediction can be used to best model the outcome given covariates including those predictive but not prescriptive variables in the estimation of the contrast function. Therefore, ForMMER is able to exploit predictive variables for improving efficiency in the outcome-regression step or equivalently the contrast function estimation step. In the meantime, the variable selection step focuses on selecting prescriptive variables, aiming towards minimizing a weighted misclassification error rate or equivalently maximizing the expected potential outcome of a regime. This difference explains the superior performance of ForMMER relative to SAS, especially when the number of predictive variables is large, as illustrated by our simulation studies.

2.4 Extension to Multiple Decision Point Setting

The ForMMER algorithm extends naturally to multi-stage treatment decision problems where decisions are made at K decision points and at each stage there are two treatment options (0 or 1). Suppose data are obtained from sequentially randomized clinical trials or observational studies where the no unmeasured confounders assumption holds. We denote the treatment received at stage k as A_k and the observed treatment history up to decision k as $\bar{A}_k = (A_1, \dots, A_k)$. Let X_k be the covariate information observed between decision $k - 1$ and k and $\bar{X}_k = (X_1, \dots, X_k)$ be the observed covariate history up to k . The overall outcome of interest is still denoted by Y . A dynamic treatment regime is a set of sequential decision rules, $g = (g_1, \dots, g_K)$, where g_k is a function of \bar{x}_k and \bar{a}_{k-1} , denoted as $g_k(\bar{x}_k, \bar{a}_{k-1})$, that determines the treatment decision at stage k based on patient's covariate and treatment history available up to decision k . We denote $L_k = (\bar{X}_k, \bar{A}_k)$.

Similar to the single decision point setting, only the treatment contrast at each stage is relevant for treatment decision. Analogously we define a contrast function at each stage, ie., $C_k(L_k) = Q_k(L_k, a_k = 1) - Q_k(L_k, a_k = 0)$, where a_k is a treatment decision at stage k and $Q_k(L_k, a_k)$ is the so-called Q-functions with “Q” for “quality”. At the last stage K , $Q_K(L_K, a_K) = E(Y|L_K, A_K = a_K)$. The Q-functions at stage $k < K$ are defined recursively and can be interpreted as the conditional expected outcomes given that the optimal decisions are made in the future. Therefore, the contrast function $C_k(L_k)$ represents the contrast in the quality between treatment 1 or 0 at stage k assuming the optimal decisions are made in the future. We refer readers to Schulte, et al. (2014) for more details.

Intuitively, at each stage, subjects can be viewed as coming from two latent classes for whom the optimal decision at stage k is 1 or 0 (or equivalently $Z_k \equiv I\{C_k(L_k) > 0\}$), assuming the optimal decisions are made in the future. Zhang and Zhang (2015) show that

for multiple-stage decision problems, the optimal treatment regimes can be identified by backward sequentially minimizes $E[|C_k(L_k)|I\{Z_k \neq g_k(L_k)\}]$, which is the expected loss of misclassifying a patient at stage k by decision rule g_k . In practice, Z_k and $C_k(L_k)$ have to be estimated and, as in the single-decision point setting, various methods (eg, parametric regression method \widehat{C}_{reg} and the AIPWE \widehat{C}_{AIPWE}) can be used to estimate the contrast functions as well as Z_k . Then the empirical analog of $E[|C_k(L_k)|I\{Z_k \neq g_k(L_k)\}]$ can be used as a objective function in estimating the optimal treatment regime at each stage. We refer readers to Zhang and Zhang (2015) for details on the C-learning framework for multiple-decision point setting. The proposed ForMMER can naturally be embedded in the C-learning framework to select variables important for decision-making and estimate the optimal treatment regime at each stage. Specifically, one only needs to modify the objective function in (3) to $\frac{1}{n} \sum_{i=1}^n [\widehat{W}_{ki} I\{\widehat{Z}_{ki} \neq g_k(L_i)\}]$, where $\widehat{W}_{ki} = |C_k(L_{ki})|$, and modify $err^{(m)}$ accordingly. ForMMER can then be used to identify the linear decision rule that minimizes the weighted misclassification error rate at each stage.

3 Simulations

We conducted simulation studies to evaluate the performance of the proposed methods. All data generating scenarios except for one are adopted from Fan, Lu and Song (2015). We consider both single-decision point and multiple-decision point settings. We compare our methods with SAS developed by Fan, Lu and Song (2015) since in their simulations they demonstrated that SAS has superior performance than the S-score method of Gunter, Zhu and Murphy, 2011, and the method of Lu, Zhang and Zeng, 2013, with LASSO selection.

For the single decision point setting, data were generated according to five scenarios, where scenarios I-IV are directly adopted from Fan, Lu and Song (2015). Specifically, Co-

variables $X = (X_1, \dots, X_p)^T$, $p = 1000$, are generated from multivariate normal distribution with mean zero, variance 1 and correlation $\text{corr}(X_j, X_k) = \rho^{|j-k|}$, where $\rho = 0.2$ or 0.8 . Treatment A is generated from a Bernoulli distribution with probability 0.5 and the error term ϵ is normally distributed with mean 0 and variance 0.25. Denoting $\tilde{X} = (1, X^T)^T$, the outcomes are generated according to:

- Scenario I: $Y = 1 + \gamma_1^T X + A\beta^T \tilde{X} + \epsilon$ with $\gamma_1 = (1, -1, \mathbf{0}_{p-2})^T$, $\beta = (0.1, 1, \mathbf{0}_7, -0.9, 0.8, \mathbf{0}_{p-10})$;
- Scenario II: $Y = 1 + 0.5 \sin(\pi\gamma_1^T X) + 0.25(1 + \gamma_2^T X)^2 + A\beta^T \tilde{X} + \epsilon$ with γ_1 and β the same as in scenario I and $\gamma_2 = (1, 0_2, -1, \mathbf{0}_5, 1, \mathbf{0}_{p-10})^T$;
- scenario III: $Y = 1 + \gamma_1^T X + A\beta^T \tilde{X} + \epsilon$ with γ_1 the same as in scenario I, and $\beta = (0.1, 1, \mathbf{0}_7, -0.9, 0.8, \mathbf{0}_{10}, 1, 0.8, -1, \mathbf{0}_5, 1, -0.8, \mathbf{0}_{p-30})$;
- Scenario IV: $Y = 1 + 0.5 \sin(\pi\gamma_1^T X) + 0.25(1 + \gamma_2^T X)^2 + A\beta^T \tilde{X} + \epsilon$ with γ_1 , and γ_2 the same as in scenario II, and β the same as in scenario III.
- Scenario V: $Y = 1 + \gamma_1^T X + A\beta^T \tilde{X} + \epsilon$ with $\gamma_1 = (1, -0.8, 1, 0.9, 0.8, 1, 0.9, 0.8, \mathbf{0}_{p-8})^T$, $\beta = (0.1, \mathbf{0}_8, 1, 0.8, \mathbf{0}_{p-10})$.

Scenarios I and II have three prescriptive variables and scenarios III and IV have eight prescriptive variables. In scenarios I and III, the relationship between outcome and covariates are linear, whereas in scenarios II and IV, the relationship is nonlinear. In scenarios I-IV, the number of prescriptive variables is more than the number of predictive variables. However, in reality, it is perhaps more plausible or often believed that many covariates have an main effect but not a qualitative interaction effect with treatment. Considering this, we also generated data from scenario V, which is modified based on scenario I but more variables have an main effect and less variables have a qualitative interaction with

treatment. The sample size we considered in the single-decision point setting is $n = 200$ and 400.

As for the multiple-decision point setting, again we adopted the same data generating process as in Fan, Lu and Song (2015). Specifically, data are generated to mimic a two-stage decision problem, where the outcome Y is generated according to

$$Y = A_1 A_2 + A_2(a + \beta_{12}^T X_1 + \beta_{21}^T X_2) + A_1(a + \beta_{11}^T X_1) + \epsilon,$$

where treatment A_1 and A_2 follow Bernoulli (0.5), baseline covariats $X_1 = (X_{1,1}, \dots, X_{1,p_1})$ follow multivariate normal distribution with mean 0, variance 1 and $\text{corr}(X_{1,j}, X_{1,k}) = 0.2^{|j-k|}$, $j \neq k$, and ϵ follows normal distribution with mean 0 and variance 0.25. The intermediate covariate X_2 is generated according to $X_2 = c_0 + c_1 X_{1,1} + c_2 A_2 + C_3 A_1 X_{1,1} + e$ with e generated from normal with mean 0 and variance 0.25. The parameter values are chosen as: $\beta_{12} = (0, 0, 1, -1, \mathbf{0}_{p_1-4})$, $\beta_{11} = (\mathbf{0}_4, 1, -1, \mathbf{0}_{p_1-6})^T$, $a = 0$, and $c = (0, 1, 0, 0)^T$. Also we chose $p_1 = 500$.

We implemented the proposed methods using two ways. In ForMMER-reg, parametric regression models are used to model $\mu(a, X)$ and the contrast functions are estimated by \hat{C}_{reg} . Specifically, forward selection based on BIC is used to build models for $\mu(0, X)$ and $\mu(1, X)$ where the maximum number of steps is set to be 10. In ForMMER-AIPWE, the contrast functions are estimated using \hat{C}_{AIPWE} , where the same parametric models as in ForMMER-reg are used in the augmentation terms of the AIPWEs. We used the union of the variables selected in the model for $\mu(1, X)$ and the first 10 variables based on the ranking of $\text{err}^{(0)} - \text{err}(X_j)$ as the initially selected set to be considered in the ForMMER procedure. We set $\alpha = 0.05$ in the ForMMER algorithm.

Results are shown in Tables 1-3. Following Fan, Lu and Song (2015), we evaluate the performance of each methods using three metrics. TP (true positive) is the number

of correctly identified prescriptive variables. VR (value ratio) is the value ratio of the estimated regime relative to the true optimal regime, i.e., $VR = E\{Y^*(\hat{g}^{opt})\} / E\{Y^*(g^{opt})\}$, where the value of a regime $E\{Y^*(g)\}$ is calculated by the average of outcomes generated from the true model with treatment determined by the regime using 100,000 Monte Carlo Replicates. ER (error rate) is the rate of incorrect treatment decision of the estimated regime, ie, an incorrect decision is made if the treatment decision determined by a regime is different from the correct optimal decision. Reported results are averages across 500 Monte Carlo simulations and the standard deviation are reported in parenthesis. Although all three metrics are useful in evaluating the performance of a method, from the perspective of optimizing expected potential outcomes, VR is in our view the most relevant one as it takes into account whether or not an correct decision is made and the magnitude of the consequence of an incorrect decision, whereas ER only accounts for whether the optimal decision is made ignoring the magnitude of loss associated with an incorrect decision. In addition, the number (size) of prescriptive variables in the estimated regime is also reported, which is important for interpreting TP.

Results on scenarios I-IV for sample size $n=200$ are summarized in Table 1; results for $n=400$ are similar and are not reported. Table 1 shows that under scenarios II and IV, and scenario III($\rho = 0.2$), the proposed methods (both implementations) have better performances than SAS and ForMMER-reg has the best performance. Under scenario III ($\rho = 0.8$), the proposed methods and SAS have comparable performance with ForMMER-reg being slightly better. Under scenario I the proposed methods and SAS have comparable performance with SAS being slightly better.

Table 2 shows results under scenario V, which is similar to scenario I except for that there are more predictive variables and less prescriptive variables. Under this scenario, the proposed ForMMER (both implementations) has considerably better performances than

SAS. This difference in performance is expected to be even larger under other scenarios (e.g., scenarios similar to II and IV) since scenario V is modified based on scenario 1 where SAS has relatively the best performance compared with other scenarios in Table 1. This result is consistent with and supports our conjecture that SAS may not perform well when many covariates are predictive but not prescriptive as explained at the end of Section 2.2. This is because in SAS predictive but not prescriptive variables will not be selected in previous steps and as a result cannot be used in the conditional models in subsequent steps to improve performance, even though predictive variables (regardless of being prescriptive or not) are useful for improving the performance of the models and estimation of optimal regimes. Our methods do not suffer from this issue and is able to take advantages of predictive variables to improve efficiency while still targeting selection of only prescriptive variables in the forward selection algorithm. This is achieved by decoupling the step for estimating the contrast function and the step for variable selection in the optimization step. This difference also explains the better performance of ForMMER in Table 1.

Results on the two-stage setting are shown in Table 3. We report on the size of selected prescriptive variables, true positive (TP), and error rate (ER) of treatment decisions for each stage separately. The value ratio (VR) of the final estimated dynamic two-stage treatment regime is reported as overall VR. Table 3 shows that the proposed methods and SAS have comparable performance at the last stage (stage 2) with SAS selects a slightly larger number of variables and hence slightly larger number of true prescriptive variables. At the stage 1, SAS tends to select considerably larger number of variables with true positive only slightly different from the proposed methods. ForMMER (both implementations) have considerably better error rate than SAS at stage 1. Overall, for both sample size $n = 200$ and 400, the proposed methods have better value than SAS; for example, for $n=200$, the value ratio using SAS is 79.1% and 91.6% using the proposed ForMMER-reg. As the value

of a regime takes into account the magnitude of incorrect decisions and the overall and long-term effect of decisions at multiple stages, it is the most relevant metric in evaluating the overall performance of an estimated dynamic treatment regime.

4 Application to Nefazodone CBASP trial

We applied the proposed method to the Nefazodone CBASP trial, where 681 patients with nonpsychotic chronic major depressive disorder (MDD) were randomized to receive either Nefazodone, cognitive behavioral analysis system of psychotherapy (CBASP) or the combination of the two treatments (Keller, et al, 2000). Subjects were followed for 12 weeks with various assessments taken throughout the study. We consider the score on the 24-item Hamilton Rating Scale for Depression (HRSD) at 12 weeks after treatment as our outcome of interest and our analysis includes 577 subjects for whom the HRSD score at 12 weeks are available. We considered a total of 50 baseline variables in constructing optimal treatment regimes. Baseline variables considered in this analysis are listed in Table 5. Lower HRSD indicates low depression and better outcome. Previous analyses have found that the combination treatment leads to lower HRSD score than the other two single treatments, whereas there is no significance difference between the two single treatments. Based on this results, in our analysis, we firstly combined the two single treatment arms into one arm and the treatment decision to be made is either combination treatment or single treatment. Then, we limit our analysis to patients who were randomized to receive single treatment (Nefazodone or CBASP) and consider the treatment decision being either Nefazodone or CBASP. Also in our analysis, we consider $-HRSD$ as our outcome such that larger value means better outcome.

We analyzed the data using the proposed ForMMER-reg and SAS. For ForMMER, we

first build parametric linear regression model for each treatment group and the final models were chosen using forward selection based on BIC information number. Then the contrast for each subjects were used in the ForMMER algorithm to estimate the optimal treatment regime. Since the number of candidate variables in our data is only 50 and far less than the number of covariates (1000) in our simulations, we chose to use a less stringent cutoff value $\alpha = 0.02$. For combination ($A = 1$) versus single treatment ($A = 0$), the estimated optimal regimes by ForMMER and SAS are, respectively,

$$\begin{aligned}\widehat{g}_{ForMMER}^{opt} &= I(55 - X_6 > 0), \\ \widehat{g}_{SAS}^{opt} &= I(4.55 - 1.97X_{14} + 0.16X_{18} - 5.88X_{35} > 0).\end{aligned}$$

The estimated values of the two regimes using the inverse probability weighted method are almost the same: -9.8 (95% CI: -10.9,-8.7) and -9.8 (95% CI: -10.9, -8.6). They are very close to the value (-9.9) of a regime that assigns everyone to the combined treatment ($A = 1$) and much better than a regime that assigns everyone to single treatment (Table 4). These results are consistent with the previously published results, which indicate that the combined treatment is superior to single treatment.

For Nefazodone ($A = 1$) versus CBASP ($A = 0$), the estimated optimal regimes by ForMMER and SAS are, respectively,

$$\begin{aligned}\widehat{g}_{ForMMER}^{opt} &= I(-0.55 - 0.30X_1 - 0.17X_8 + 0.20X_{12} - 0.12X_{13} - 0.60X_{15} \\ &+ 0.33X_{16} + 0.71X_{40} - 0.88X_{50} > 0),\end{aligned}$$

$$\begin{aligned}\widehat{g}_{SAS}^{opt} &= I(-15.03 - 5.15X_1 - 2.71X_8 + 1.01X_{12} + 6.60X_{14} + 4.17X_{16} + \\ &0.09X_{22} - 2.49X_{28} + 6.16X_{31} - 14.69X_{33} - 7.69X_{36} - 12.53X_{37} - 4.56X_{38} \\ &+ 5.37X_{40} - 17.87X_{42} - 5.87X_{46} - 7.59X_{48} + 7.98X_{49} - 6.71X_{50} > 0).\end{aligned}$$

The estimated value of $\widehat{g}_{ForMMER}^{opt}$ and \widehat{g}_{SAS}^{opt} are -11.0 (95% CI: -12.6, -9.4) and -12.1 (95% CI: -13.8, -10.4) respectively, with $\widehat{g}_{ForMMER}^{opt}$ having larger estimated value while selecting less variables. The value of both regimes perform much better than the two regimes that assign everyone to Nefazodone or CBASP. More results are reported in Table 4.

5 Discussion

Within the classification framework (C-learning) for estimating the optimal treatment regimes, in this article we further developed a variable selection algorithm for selecting variables that have qualitative interactions with treatment and hence are important for making treatment decisions, namely, prescriptive variables. This variable selection algorithm directly targets prescriptive variables with the objective of optimizing treatment rules, in contrast to methods focusing on selecting predictive variables and prediction. Within the C-learning framework, the optimal treatment regime can be equivalently defined as the classifier that minimizes a weighted misclassification error, where the objective of the classifier is to, based on patient's characteristics, classify patients to the treatment option that leads to larger expected potential outcomes. A major advantage of this framework is that it naturally accommodates a strategy for variable selection targeting prescriptive variables, since only prescriptive variables are relevant in determining the contrast functions and the weighted misclassification error. In the proposed ForMMER algorithm, it forward sequentially selects important prescriptive variables and estimates the optimal treatment regimes simultaneously.

The ForMMER algorithm selects prescriptive variables sequentially and at each step it assesses the additional merit of a new variable given variables that have already been selected. As a result, similar to SAS, it tends not to select those variables that are only

marginally important for decision making but are not important jointly. Therefore, as SAS, it tends to select fewer variables overall but more true prescriptive variables than methods that consider each variables individually. Furthermore, the proposed ForMMER algorithm decouples the step for estimating the contrast functions from the step for optimization and prescriptive variable selection and, as a result, it is able to target directly on prescriptive variables while still taking advantage of predictive variables in the outcome-regression step to improve performance. This is one of the main differences between ForMMER and SAS and in SAS variables that are only predictive but not prescriptive tend not to be selected and hence will not be able to be exploited in subsequent steps to improve performance. This point is discussed in detail at the end of Section 2.3 and illustrated in simulations, especially in Table 2. To summarize, the flexibility of modeling the contrast functions using various ways, the sequential selection strategy, and the separation of the optimization step for variable selection and optimizing decision rules from the estimation of the contrast functions together contribute to the superior performance of the proposed ForMMER method. Overall, ForMMER selects less variables yet with better value than SAS and the same statement can be made for its performance relative to other methods evaluated in Fan, Lu and Song (2015) (i.e., the S-score method of Gunter, Zhu and Murphy, 2011, and the method of Lu, Zhang and Zeng, 2013, with LASSO selection) as our simulation settings are adopted from Fan, Lu and Song (2015). Finally, we note the measure (weighted misclassification error) used in the forward sequential variable selection in our method has a very intuitive interpretation and is directly related to the definition of an optimal treatment regime.



References

- BARRETT, J.K., HENDERSON, R. & ROSTHØJ, S. (2014). Doubly robust estimation of optimal dynamic treatment regimes. *Statistics in Biosciences* **6**, 244–260.
- BRINKLEY, J., TSIATIS, A.A. & ANSTROM, K.J. (2009). A generalized estimator of the attributable benefit of an optimal treatment regime. *Biometrics* **21**, 512–522.
- CHAKRABORTY, B., MURPHY, S.A. & STRECHER, V. (2010). Inference for non-regular parameters in optimal dynamic treatment regimes. *Statistical Methods in Medical Research* **19** 317–343.
- FAN, A., LU, W. & SONG, R. (2015). Sequential Advantage Selection for Optimal Treatment Regimens. *Annals of Applied Statistics*, in press.
- GENG, Y., LU, W. & ZHANG, H.H. (2015). On Optimal Treatment Regimes Selection for Mean Survival Time *Statistics in Medicine* **34**, 1169–1184.
- GOLDBERG, D.E. (1989). *Genetic Algorithms in Search, Optimization, and Machine Learning*. Reading, MA: Addison-Wesley.
- GUNTER, L., ZHU, J. & MURPHY, S. (2011). Variable selection for qualitative interactions. *Statistical methodology* **8**, 42-55.
- KELLER, M.B., MCCULLOUGH, J.P., KLEIN, D.N., ARNOW, B., DUNNER, D.L., GELLENBERG, A.J., MAREKOWITZ, J.C., NEMEROFF, C.B., RUSSELL, J.M., THASE, M.E., TRIVEDI, M.H. & ZAJECKA, J. (2000). A comparison of nefazodone, the cognitive behavioral-analysis system of psychotherapy, and their combination for treatment of chronic depression. *New England Journal of Medicine* **342**, 331–336.

- KANG, C., JANES, H. & HUANG, Y. (2014). Combining biomarkers to optimize patient treatment recommendations. *Biometrics* **70**, 695–720.
- LU, W., ZHANG, H.H. & ZENG, D. (2013). Variable selection for optimal treatment decision. *Statistical methods in medical research* **22**, 493–504.
- MEBANE, W.R. & SEKHON, J.S. (2011). Genetic optimization using derivatives: the rgenoud package for R. *Journal of Statistical Software* **42**, 1–26.
- MOODIE, E.E.M., RICHARDSON, T.S. & STEPHENS, D.A. (2007). Demystifying optimal dynamic treatment regimes. *Biometrics* **63**, 447–455.
- MURPHY, S.A. (2003). Optimal dynamic treatment regimes (with discussion). *Journal of Royal Statistical Society Series B* **58**, 331–366.
- PETO, R. (1982). Statistical aspects of cancer trials. In *Treatment of Cancer* (K. Halnan, ed.) 867–871. Chapman, London, UK.
- QIAN, M. & MURPHY, S.A. (2011). Performance Guarantees for Individualized Treatment Rules. *Annals of Statistics* **39**, 1180–1210.
- ROBINS, J.M. (2004). Optimal structured nested models for optimal sequential decisions. In *Proceedings of the Second Seattle Symposium on Biostatistics*, D. Y. Lin and P. J. Heagerty (eds), 189–326. New York: Springer.
- ROBINS, J., ORELLANA, L. & ROTNITZKY, A. (2008). Estimation and extrapolation of optimal treatment and testing strategies. *Statistics in Medicine* **27**, 4678–4721.
- SCHULTE, P.J., TSIATIS, A.A., LABER, E.B. & DAVIDIAN, M. (2014). Q- and A-learning Methods for Estimating Optimal Dynamic Treatment Regimes. *Statistical Science* **29** 640–661.

- TIAN, L., ALIZADH, A.A., GENTLES, A.J. & TIBSHIRANI, R. (2014). A simple method for estimating interactions between a treatment and a large number of covariates. *Journal of the American Statistical Association* **109**, 1517-1532.
- YOUNG, J.G., CAIN, L.E., ROBINS, J.M., O'REILLY, E.J., HERNÁN, M.A. (2011). Comparative effectiveness of dynamic treatment regimes: an application of the parametric g-formula. *Statistics in Biosciences* **3**, 119–143.
- ZHANG, B., TSIATIS, A.A., LABER, E.B. & DAVIDIAN, M. (2012a). A robust method for estimating optimal treatment regimes. *Biometrics* **68**, 1010–1018.
- ZHANG, B., TSIATIS, A.A., LABER, E.B. DAVIDIAN, M., ZHANG, M. & LABER, E.B.(2012b). Estimating optimal treatment regimes from a classification perspective *Stat* **1**, 103–114.
- ZHANG, B., TSIATIS, A.A., LABER, E.B., & DAVIDIAN, M.(2013). Robust estimation of optimal dynamic treatment regimes for sequential treatment decisions *Biometrika* **100**, 681–694.
- ZHANG, B., ZHANG, M.(2015). C-learning: a New Classification Framework to Estimate Optimal Dynamic Treatment Regimes. The University of Michigan Department of Biostatistics Working Paper Series, paper 116.
- ZHAO, Y., ZENG, D., RUSH, A.J. & KOSOROK, M.R. (2012). Estimating individualized treatment rules using outcome weighted learning. *Journal of the American Statistical Association* **107**, 1106–1118.
- ZHAO, Y., ZENG, D., LABER, E.B & KOSOROK, M.R. (2015). New Statistical Learning

Methods for Estimating Optimal Dynamic Treatment Regimes. *Journal of the American Statistical Association* **510**, 583–598.



Table 1: Simulation results for single-decision point setting based on 500 replications based on sample size $n=200$ (Scenarios adopted from Fan, Lu and Song, 2015). Size: number of selected prescriptive variables; TP: number of true positive (important) prescriptive variables; ER: error rate of the treatment decision; VR: ratio of the value of the estimated regime relative to that of the true optimal regime. Numbers in parenthesis are Monte Carlo standard deviation.

Method	ρ	Size	TP	ER	VR
Scenario I					
SAS	0.2	6.81 (1.65)	2.98 (0.25)	6.2 (4.8)	99.0 (3.4)
ForMMER-AIPWE		4.93 (1.05)	2.93 (0.26)	10.7 (5.1)	97.6 (2.6)
ForMMER-reg		5.78 (1.47)	2.93 (0.26)	10.3 (6.3)	97.6 (3.3)
SAS	0.8	8.03 (2.12)	1.79 (0.97)	19.0 (7.3)	94.3 (3.6)
ForMMER-AIPWE		4.20 (1.29)	1.31 (0.57)	21.6 (4.4)	93.3 (2.4)
ForMMER-reg		5.64 (1.59)	1.41 (0.64)	21.4 (4.8)	93.3 (2.6)
Scenario II					
SAS	0.2	11.90 (2.87)	2.09 (1.12)	32.7 (10.3)	88.7 (6.2)
ForMMER-AIPWE		4.82 (1.23)	2.49 (0.69)	22.6 (9.2)	94.0 (4.4)
ForMMER-reg		6.37 (1.58)	2.76 (0.48)	20.7 (7.5)	95.0 (3.4)
SAS	0.8	11.32 (2.44)	1.30 (0.73)	33.4 (5.7)	89.7 (3.1)
ForMMER-AIPWE		4.15 (1.27)	1.43 (0.75)	24.2 (6.2)	94.2 (2.7)
ForMMER-reg		5.78 (1.64)	1.66 (0.83)	23.7 (6.0)	94.4 (2.6)
Scenario III					
SAS	0.2	11.06 (2.92) ²⁸	5.08 (2.56)	23.0 (15.0)	84.2 (15.5)
ForMMER-AIPWE		7.50 (1.28)	6.79 (1.39)	14.2 (7.2)	94.3 (5.6)
ForMMER-reg		8.58 (1.13)	7.17 (1.14)	12.0 (7.5)	95.5 (5.4)
SAS	0.8	7.63 (2.02)	3.21 (1.35)	22.0 (4.7)	90.5 (3.6)
ForMMER-AIPWE		4.99 (1.34)	3.00 (1.15)	21.4 (3.8)	91.1 (3.0)
ForMMER-reg		6.56 (1.62)	3.49 (1.35)	20.4 (3.9)	91.8 (2.9)

Table 2: Simulation results for single-decision point setting based on 500 replications. Size: number of selected prescriptive variables; TP: number of true positive (important) prescriptive variables; ER: error rate of the treatment decision; VR (value ratio): ratio of the value of the estimated regime relative to that of the true optimal regime. Numbers in parenthesis are Monte Carlo standard deviations).

Method	ρ	Size	TP	ER	VR
Scenario V					
n=200					
SAS	0.2	14.49 (2.07)	0.72 (0.60)	43.9 (5.3)	71.8 (5.6)
ForMMER-AIPWE		3.99 (0.95)	1.97 (0.16)	11.8 (5.1)	97.3 (2.5)
ForMMER-reg		4.39 (1.04)	1.98 (0.15)	12.5 (6.3)	96.7 (2.9)
SAS	0.8	8.55 (2.03)	0.96 (0.35)	21.3 (6.3)	90.9 (5.1)
ForMMER-AIPWE		3.57 (1.00)	1.81 (0.39)	7.7 (2.9)	98.7 (0.9)
ForMMER-reg		3.84 (1.00)	1.99 (0.11)	4.4 (1.8)	99.6 (0.4)
n=400					
SAS	0.2	24.96 (3.18)	1.56 (0.54)	35.6 (7.0)	80.0 (6.3)
ForMMER-AIPWE		3.25 (0.78)	2.00 (0.04)	7.3 (2.9)	99.0 (0.8)
ForMMER-reg		4.15 (0.98)	2.00 (0.00)	8.2 (6.1)	98.3 (1.9)
SAS	0.8	9.59 (2.07)	1.47 (0.50)	13.5 (4.1)	96.3 (2.2)
ForMMER-AIPWE		3.14 (0.80)	1.92 (0.27)	5.7 (2.3)	99.3 (0.5)
ForMMER-reg		3.87 (1.05)	2.00 (0.00)	2.7 (1.0)	99.9 (0.1)

Table 3: Simulation results for two-stage setting based on 500 replications (scenario adopted from Fan, Lu and Song, 2015). Size: number of selected prescriptive variables; TP: number of true positive (important) prescriptive variables; ER: error rate of the treatment decision at the stage; overall VR (value ratio): ratio of the value of the estimated regime relative to that of the true optimal regime. Numbers in parenthesis are Monte Carlo standard deviations.

Method	stage 2			stage 1			Overall VR
	Size	TP	ER	Size	TP	ER	
n=200							
SAS	6.59	3.53	14.0	11.70	2.19	30.2	79.1
	(2.16)	(0.59)	(5.5)	(2.68)	(0.93)	(7.0)	(8.7)
ForMMER-AIPWE	4.59	3.14	14.0	4.07	2.19	13.4	87.3
	(1.05)	(0.61)	(4.1)	(1.23)	(0.82)	(4.4)	(6.4)
ForMMER-reg	5.12	3.34	11.8	4.75	2.61	11.3	91.6
	(1.21)	(0.54)	(3.6)	(1.35)	(0.82)	(3.8)	(4.2)
n=400							
SAS	5.77	3.93	7.6	13.00	3.95	16.2	93.3
	(1.80)	(0.26)	(3.6)	(3.75)	(1.13)	(5.3)	(3.9)
ForMMER-AIPWE	3.88	3.27	10.8	3.00	2.20	10.5	94.1
	(0.81)	(0.48)	(3.0)	(0.78)	(0.51)	(3.3)	(2.7)
ForMMER-reg	4.36	3.37	9.1	3.55	2.45	8.7	96.0
	(0.96)	(0.49)	(3.4)	(1.01)	(0.62)	(3.5)	(1.6)

Table 4: Data analysis results: estimated value of the estimated optimal treatment. The values are estimated using inverse probability weighted method. $g = 1$ is a regime that assigns everyone to treatment 1 and $g = 0$ is a regime that assign everyone to treatment 0. Numbers in parenthesis are 95% confidence intervals.

	Combination ($A = 1$) vs. Single ($A = 0$)	Nefazodone ($A = 1$) vs. CBASP ($A = 0$)
ForMMER	-9.8 (-10.9,-8.7)	-11.0 (-12.6,-9.4)
SAS	-9.8 (10.9,-8.6)	-12.1 (-13.8,-10.4)
$g = 1$	-9.9 (-11.0,-8.8)	-14.9 (-16.4,-14.9)
$g = 0$	-14.9 (-15.8,-13.9)	-14.8 (-16.2,-13.4)



Table 5: List of covariates used in the analysis of Nefazodone CBASP trial

1 Female	
2 White	
3-4 Marital Status (single, widowed/divorced/separated)	
5 Body mass index	
6 Age of MDD onset	7 Age at screening
8 Live alone	9 IDSSR Anxiety/Arousal Score
10 IDSSR General/Mood Cognition Score	11 IDSSR total score
12 IDSSR sleep score 1	13 IDSSR sleep score 2
14 HAMD Anxiety/Somatic Symptoms	15 HAMD Cognitive Disturbance
16 HAMD Retardation Score	17 HAMD Sleep Disturbance factor score
18 Total HAMD-17 score	19 Total HAMD-24 score
20 MOS36 Cognitive Functioning Factor Score	21 MOS36 General Health Factor Score
22 MOS36 Mental Health Factor Score	23 MOS36 Social Functioning
24 Total HAMA score	25 HAMA Psychic Anxiety Score
26 HAMA Somatic Anxiety Score	
27-28 MDD type (neither melancholic or atypical, melancholic)	
29-30 Main study diagnosis (no antecedent, continuous)	
31-32 MDD current severity (mild, moderate)	
33 Anxiety disorder NOS	
34-35 Alcohol (abuse, dependence)	
36-37 Anxiety(sub-threshold, threshold)	
38 Other psychological problems	
39 Body dysmorphic current	
40 Drug abuse	
41 Anorexia or bulimia nervosa	
42 Obsessive compulsive	32
43-44 Specific phobia (sub-threshold, threshold)	
45-46 Social phobia (sub-threshold, threshold)	
47-48 Post traumatic stress(sub-threshold, threshold)	
49-50 Panic (sub-threshold, threshold)	
