The Annals of Applied Statistics
2016, Vol. 10, No. 1, 32–53
DOI: 10.1214/15-AOAS849
© Institute of Mathematical Statistics, 2016

SEQUENTIAL ADVANTAGE SELECTION FOR OPTIMAL TREATMENT REGIME[1]

BY AILIN FAN, WENBIN LU[1] AND RUI SONG[2]

North Carolina State University

Variable selection is gaining more attention because it plays an important role in deriving practical and reliable optimal treatment regimes for personalized medicine, especially when there are a large number of predictors. Most existing variable selection techniques focus on selecting variables that are important for prediction. With such methods, some variables that are poor in prediction but are critical for treatment decision making may be ignored. A qualitative interaction of a variable with treatment arises when the treatment effect changes direction as the value of the variable varies. Variables that have qualitative interactions with treatment are of clinical importance for treatment decision making. Gunter, Zhu and Murphy [*Stat. Methodol.* **8** (2011) 42–55] proposed the S-score method to characterize the magnitude of qualitative interaction of an individual variable with treatment. In this paper, we develop a sequential advantage selection method based on a modified S-score. Our method sequentially selects variables with a qualitative interaction and can be applied in multiple decision-point settings. To select the best candidate subset of variables for decision making, we also propose a BIC-type criterion that is based on the sequential advantage. The empirical performance of the proposed method is evaluated by simulation and an application to depression data from a clinical trial.

1. Introduction. Personalized medicine is emerging as a new strategy for treatment that takes individual heterogeneity in background characteristics, clinical measurements and genetic information into consideration. In this paradigm, treatment duration, dose and type are adjusted over time and are tailored according to an individual's information with the aim of optimizing the effectiveness of treatment. This approach is different from the traditional "one-size-fits-all" treatment, which ignores the long-term benefits and individual heterogeneities. Great interest lies in finding optimal treatment regimes based on data from clinical trials and observational studies [e.g., Moodie, Richardson and Stephens (2007), Murphy (2003), Robins (2004)].

A large number of approaches have been developed to estimate optimal treatment regimes, including marginal structural models [Murphy, van der Laan and

Received March 2015.

[1]Supported in part by Grant NIH R01 CA140632 and Grant NCI P01 CA142538.

[2]Supported in part by Grant NSF-DMS 1309465 and Grant NCI P01 CA142538.

Key words and phrases. Optimal treatment regime, qualitative interaction, sequential advantage, variable selection.

Robins (2001), Robins (1997)], Q-learning [Chakraborty, Murphy and Strecher (2010), Murphy (2005a, 2005b), Song et al. (2011), Watkins (1989), Watkins and Dayan (1992), Zhao et al. (2011)], A-learning [Murphy (2003), Robins (2004)] and value-based optimization methods [Zhang et al. (2012a, 2012b, 2013), Zhao et al. (2012)].

As the amount of information able to be collected on individuals continues to increase, more and more covariates are measured and are available in clinical studies. For example, a clinical trial may collect a large amount of information on a patient's demographics, medical history, intermediate outcomes and side effects. However, it may be expensive or time-consuming to collect all of this information in clinical practice, and redundancy in covariate information may impair the accuracy of optimal treatment decisions as well as its interpretation. Thus, a natural problem that arises in the estimation of optimal treatment regimes is how to identify the important covariates for treatment decision making.

Our work was motivated from the Sequenced Treatment Alternatives to Relieve Depression (STAR*D) study [Fava et al. (2003), Rush et al. (2004)]. The STAR*D study was a sequential, multiple-assignment, randomized trial [SMART, see Murphy (2005a), Qian, Nahum-Shani and Murphy (2013)] for patients with nonpsychotic major depressive disorder. This study aimed to determine which antidepressant medications, in what order and what combination, should be given to patients to yield the optimal treatment effect. A large number of covariates were collected at baseline, such as patient demographic characteristics and medical history. In addition, several intermediate medical measurements were taken to assist in treatment decision making at the second or higher treatment decision points. It is hard to select covariates useful for making decisions from such a large number of covariates based on experts' opinions only. Thus, variable selection is crucial for deriving the optimal treatment regimes in the STAR*D trial.

Although variable selection is an important area in modern statistical research, current variable selection techniques mainly focus on selecting variables for prediction. Such approaches may not be able to adequately predict the interactions of variables with treatment and thus may neglect variables that are vital for decision making. In medical decision making settings, variables that have qualitative interactions with treatments are clinically important [Peto (1982)]. These variables are called prescriptive variables, which help prescribe the optimal treatment regimes. These variables should be distinguished from predictive variables, which help to increase prediction accuracy.

Scarce research has been done to study variable selection techniques for decision making. Qualitative interaction tests [Gail and Simon (1985), Piantadosi and Gail (1993), Yan (2004)] have been used to test a small number of expert determined prespecified interactions. However, many of the tests were designed to test only qualitative interactions between categorical variables and treatments.

Moreover, when the number of covariates is large, these tests are too conservative when controlling the error rate for multiple testing. Penalized methods have also been studied to identify variables important for making treatment decisions. Among others, Qian and Murphy (2011) developed a two-step procedure, where they first estimate the conditional mean response using the penalized least squares regression with the L_1 penalty and then derive the estimated optimal treatment regimes from this estimated conditional mean. Lu, Zhang and Zeng (2013) proposed a penalized least squares regression in an A-learning framework, which does not require the correct specification of the baseline mean model and directly selects variables with nonzero interactions with treatment. However, both methods do not directly target prescriptive variables that are important for treatment decision making. Gunter, Zhu and Murphy (2011) proposed a variable-ranking measure that characterizes the qualitative interaction of an individual variable with treatment, namely, the S-score. Then, a hybrid algorithm that combines S-score ranking and weighted LASSO was used to select variables for treatment decision making.

In this paper, we propose a variable selection method to identify prescriptive variables for deriving optimal treatment regimes with single-stage and multi-stage treatment decisions. In particular, we develop a quantity named *sequential advantage*, which can be viewed as a sequential S-score. This quantity characterizes additional information provided by a new variable to treatment decision making, conditional on the effects of the covariates that are included from previous steps. We also propose a BIC-type criterion that is based on sequential advantage to choose the best candidate model for treatment decision making. As sequential advantage measures the potential for a qualitative interaction with treatment, our method targets prescriptive variables.

Compared to the S-score method, our method is more accurate in the sense that it tends to select more prescriptive variables but selects fewer variables overall. This behavior is due to the sequential advantage selection, which can incorporate the correlation among variables. Thus, our method can largely exclude spurious variables that are marginally important but jointly unimportant. The proposed method has satisfactory performance in each stage of dynamic treatment regimes. Because the proposed method starts from the null model, the implementation is feasible in high-dimensional settings provided that the true model is sufficiently sparse.

The remainder of the paper is organized as follows. In Section 2 we introduce the framework for deriving optimal dynamic treatment regimes and S-score ranking for selecting prescriptive variables. Section 3 provides the proposed sequential advantage selection method for variable selection in optimal treatment decision making. We demonstrate the method's performance in Section 4 by simulation studies in various scenarios and illustrate the method using data from the STAR*D clinical trial in Section 5.

2. Overview for dynamic treatment regime and S-score ranking.

2.1. *Dynamic treatment regime*. Suppose that treatment decisions are made at a finite number of time points, which are denoted by $t_1, \ldots, t_K$ with t_1 being the baseline. The data for a single individual are summarized as $(\mathbf{X}_1, A_1, \ldots, \mathbf{X}_K, A_K, Y)$, where $\mathbf{X}_1$ are the baseline covariates obtained prior to the first treatment decision, $\mathbf{X}_k$ are the covariates accrued between t_{k-1} and t_k, $k = 2, \ldots, K$, A_k is the treatment given at t_k, $k = 1, \ldots, K$, and Y is the outcome of interest with larger values indicating better response. For simplicity, assume that $A_k = 0$ or 1 for all k. Overbar notation is used to denote the history of time-dependent variables. That is, $\bar{\mathbf{X}}_k = (\mathbf{X}_1, \ldots, \mathbf{X}_k)$, $\bar{A}_k = (A_1, \ldots, A_k)$, $k = 1, \ldots, K$. The observed data for n subjects are summarized as

$$\{(\mathbf{X}_{1i}, A_{1i}, \ldots, \mathbf{X}_{Ki}, A_{Ki}, Y_i), i = 1, \ldots, n\},$$

which are independent and identically distributed (i.i.d.) across i.

A dynamic treatment regime is a set of rules that dictates how treatments are assigned to an individual over time based on past information. We denote a dynamic treatment regime as $d = (d_1, \ldots, d_K)$, where $d_k : \Gamma_k \to \mathcal{A}_k = \{0, 1\}$ is a map of the information available at time t_k to the possible treatment decisions that could be made at t_k. In the mapping, $\Gamma_k = \{(\bar{\mathbf{x}}_k, \bar{a}_{k-1}) \in \bar{\mathcal{X}}_k \times \bar{\mathcal{A}}_{k-1}\}$, which is the set of historical information including both covariates and treatments. To define the optimal dynamic treatment regime that maximizes the expected response, we need to introduce potential outcomes [Robins (1986), Rubin (1978)]. Specifically, for a fixed treatment regime $\bar{a}_K \in \bar{\mathcal{A}}_K$, the potential outcomes are given by

$$W = \{\{\mathbf{X}_1, \mathbf{X}_2^*(\bar{a}_1), \ldots, \mathbf{X}_K^*(\bar{a}_{K-1}), Y^*(\bar{a}_K)\}, \text{ for all } \bar{a}_K \in \bar{\mathcal{A}}_K\},$$

where $\mathbf{X}_k^*(\bar{a}_{k-1})$ denote the potential intermediate covariates that would accrue between t_{k-1} and t_k given the treatment history $\bar{a}_{k-1}$ $(k = 2, \ldots, K)$, and $Y^*(\bar{a}_K)$ denotes the potential outcome that would result if treated according to $\bar{a}_K$. The optimal dynamic treatment regime is then defined as $d^{\text{opt}} = \arg\max_{d \in \mathcal{D}} \mathbb{E}[Y^*(d)]$, where $\mathcal{D}$ is a class of possible treatment regimes.

To estimate the expected potential outcome following a dynamic treatment regime from the observed data, two assumptions are typically needed: the stable unit treatment value assumption [Rubin (1978)] and the sequential randomization assumption [Robins (1997)]. The first assumption is usually reasonable but cannot be verified generally. The second assumption is met in a sequentially randomized trial such as the STAR*D study. With these two assumptions, the expected potential outcome following dynamic treatment regime d can be expressed as

$$\mathbb{E}[Y^*(d)] = \mathbb{E}[\mathbb{E}[\cdots \mathbb{E}[Y|\bar{\mathbf{X}}_K, \bar{A}_{K-1}, A_K = d_K] \cdots |\mathbf{X}_1, A_1 = d_1]].$$

Therefore, the expected potential outcome following a dynamic treatment regime can be estimated from the observed data. Furthermore, an optimal dynamic treatment regime can be derived, for example, using Q-learning or A-learning.

2.2. *S-score ranking*. When deriving optimal treatment regimes, only variables that have qualitative interaction effects with the treatment play a role. Gunter, Zhu and Murphy (2011) pointed out two factors that affect the degree of a qualitative interaction: the magnitude of interaction between the variable and the treatment, and the proportion of patients for whom the optimal treatment changes given the knowledge of the variable. Based on these two factors, they proposed the S-score, which characterizes the degree of qualitative interaction of a variable. For single treatment decision A, the S-score for the jth covariate, X_j, is defined as

$$S_j = \sum_{i=1}^{n} \Big[\max_a \{\hat{\mathbb{E}}(Y_i | X_{ij} = x_{ij}, A_i = a)\} - \hat{\mathbb{E}}(Y_i | X_{ij} = x_{ij}, A_i = \hat{a}) \Big], \tag{1}$$

where $\hat{\mathbb{E}}(Y_i | X_{ij} = x_{ij}, A_i = a)$ is an estimator of $\mathbb{E}(Y_i | X_{ij} = x_{ij}, A_i = a)$, and $\hat{a} = \arg\max_a \hat{\mathbb{E}}(Y | A = a)$, that is, the treatment that leads to the largest treatment-specific mean response. The S-score is always nonnegative, and a higher valued S-score indicates a greater potential for the covariate to have a qualitative interaction with treatment.

To show that the S-score captures both the magnitude of interaction and the proportion of subjects whose optimal treatment changes, we illustrate with an example. Consider the model $\mathbb{E}(Y | X_j, A) = \beta_0 + \beta_1 X_j + \beta_2 A + \beta_3 X_j A$, and let $(\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2, \hat{\beta}_3)^T$ denote the estimates of $(\beta_0, \beta_1, \beta_2, \beta_3)^T$. The S-score for X_j is then given by

$$S_j = \sum_{i=1}^{n} (\hat{\beta}_2 + \hat{\beta}_3 x_{ij}) [\mathbf{1}(\hat{\beta}_2 + \hat{\beta}_3 x_{ij} \geq 0) - \hat{a}]. \tag{2}$$

In equation (2), $(\hat{\beta}_2 + \hat{\beta}_3 x_{ij})$ represents the magnitude of the treatment effect as a function of X_{ij}, and $\mathbf{1}(\hat{\beta}_2 + \hat{\beta}_3 x_{ij} \geq 0) - \hat{a}$ indicates whether the optimal treatment for patient i changes given the knowledge of X_{ij}. Therefore, both factors are reflected in the S-score.

Although the S-score has very appealing properties for characterizing qualitative interaction of an individual covariate, there are some limitations with the S-score ranking. First, when the number of covariates is large, the S-score is not very effective for selecting qualitative interactions; variables that have no qualitative interaction with treatment can have nonzero S-scores due to correlations among covariates. In the algorithm proposed by Gunter, Zhu and Murphy (2011), a weighted LASSO is used to select important interactions based on a linear model built on variables with nonzero S-scores, and the inverses of individual S-scores are used as the weights in the weighted LASSO selection. In addition, an adjusted gain in value criterion is used to select the best subset of variables along the solution path for those selected nonzero S-scores. This hybrid algorithm helps to pick variables among the pool of variables with nonzero S-scores. Second, because the S-score evaluates each variable individually, some variables that are jointly crucial for optimal treatment decision making may be neglected. Third, the S-score method

proposed by Gunter, Zhu and Murphy (2011) is only studied for a single-stage treatment decision. These limitations motivate us to develop a forward-selection procedure based on a modified S-score, named *sequential advantage*, for selecting variables having qualitative interactions with treatment for both single-stage and multi-stage treatment decisions.

3. Sequential advantage selection. In this section we introduce sequential advantage and describe sequential advantage selection algorithms for both single-stage and multi-stage treatment decisions.

3.1. *Sequential advantage.* We introduce sequential advantage in a single-stage treatment decision study. Let $\mathcal{M} = \{j^1, \ldots, j^k\}$ denote an arbitrary model with $X_{j^1}, \ldots, X_{j^k}$ as the selected covariates and $\mathcal{F} = \{1, \ldots, p\}$ denote the full model. In addition, let $\mathbf{X}_i$ denote the covariate for subject i and $\mathbf{X}_{i(\mathcal{M})} = \{X_{ij} : j \in \mathcal{M}\}$ denote the associated covariates corresponding to model $\mathcal{M}$. The sequential advantage of variable X_j, $j \in \mathcal{F} \setminus \mathcal{M}^{(k-1)}$, is defined as

$$
\begin{aligned}
S_j^{(k)} = \frac{1}{n}\sum_{i=1}^{n}\Big[&\max_a\big\{\hat{\mathbb{E}}(Y|\mathbf{X}_{\mathcal{M}_j^{(k)}} = \mathbf{x}_{i\mathcal{M}_j^{(k)}}, A = a)\big\} \\
&- \hat{\mathbb{E}}\big(Y|\mathbf{X}_{\mathcal{M}_j^{(k)}} = \mathbf{x}_{i\mathcal{M}_j^{(k)}}, A = a_{\mathrm{opt}}^{(k-1)}(\mathbf{x}_{i\mathcal{M}^{(k-1)}})\big)\Big],
\end{aligned} \tag{3}
$$

where $\mathcal{M}^{(k-1)} = \{j^1, \ldots, j^{k-1}\}$ is the model selected at the $(k-1)$th step, $\mathcal{M}_j^{(k)} = \mathcal{M}^{(k-1)} \cup \{j\}$, $\hat{\mathbb{E}}(Y|\mathbf{X}_{\mathcal{M}_j^{(k)}} = \mathbf{x}_{i\mathcal{M}_j^{(k)}}, A = a)$ is the estimated conditional mean response based on an assumed model with predictors $\mathbf{X}_{\mathcal{M}_j^{(k)}}$ and A, and $a_{\mathrm{opt}}^{(k-1)}(\mathbf{x}_{i\mathcal{M}^{(k-1)}})$ is the optimal treatment regime obtained based on the variables in $\mathcal{M}^{(k-1)}$. In practice, a linear model with main effects of $\mathbf{X}_{\mathcal{M}_j^{(k)}}$ and A as well as interaction effects between $\mathbf{X}_{\mathcal{M}_j^{(k)}}$ and A can be used to obtain $\hat{\mathbb{E}}(Y|\mathbf{X}_{\mathcal{M}_j^{(k)}} = \mathbf{x}_{i\mathcal{M}_j^{(k)}}, A = a)$. Similarly, $a_{\mathrm{opt}}^{(k-1)}(\mathbf{x}_{i\mathcal{M}^{(k-1)}})$ can be obtained based on the fitted model $\hat{\mathbb{E}}(Y|\mathbf{X}_{\mathcal{M}^{(k-1)}} = \mathbf{x}_{i\mathcal{M}_{(k-1)}}, A = a)$. The sequential advantage defined in (3) is similar to the S-score in spirit, but represents the additional benefit of including variable X_j to improve the optimal treatment regime estimated based on previously selected variables.

3.2. *Sequential advantage selection algorithm.* In this section we propose a variable selection method based on sequential advantage in a forward selection manner. We first describe the sequential advantage selection (SAS) algorithm for selecting variables that have a qualitative interaction with treatment in a single treatment decision study, and then extend SAS to accommodate multiple treatment decisions using Q-learning in the next section. The SAS algorithm for a single-stage treatment decision is given as follows:

(i) Initialization. Set $\mathcal{M}^{(0)} = \varnothing$. Compute $a_{\text{opt}}^{(0)} = \arg\max_a \hat{\mathbb{E}}(Y|A=a)$, and let $S^{(0)} = \hat{\mathbb{E}}(Y|A=a_{\text{opt}}^{(0)}) - \hat{\mathbb{E}}(Y)$.

(ii) *Sequential advantage selection*. In the kth step ($k \geq 1$), we have $\mathcal{M}^{(k-1)}$. For every $j \in \mathcal{F} \setminus \mathcal{M}^{(k-1)}$, we consider candidate covariates $\mathcal{M}_j^{(k)} = \mathcal{M}^{(k-1)} \cup \{j\}$ and compute the sequential advantage (3) corresponding to the jth covariate in the kth step. The kth variable to be selected is the one with the largest sequential advantage in this step: $j^k = \arg\max_{j \in \mathcal{F} \setminus \mathcal{M}^{(k-1)}} \{S_j^{(k)}\}$. Update $\mathcal{M}^{(k)} = \mathcal{M}^{(k-1)} \cup \{j^k\}$ and the estimated optimal treatment regime based on the first k selected variables $\mathbf{X}_{\mathcal{M}^{(k)}}$, that is, $a_{\text{opt}}^{(k)}(\mathbf{x}_{\mathcal{M}^{(k)}}) = \arg\max_a \hat{\mathbb{E}}(Y|\mathbf{X}_{\mathcal{M}^{(k)}} = \mathbf{x}_{\mathcal{M}^{(k)}}, A=a)$. Let $S^{(k)} = S_{j^k}^{(k)}$.

(iii) *Selection of best subset*. Iterate step (ii) to obtain a solution path for the first m selected variables: $\mathcal{M}^{(m)} = \{j^1, \ldots, j^m\}$, where m is a predefined integer that is usually chosen to be less than $n/2$. We use a BIC-type criterion to select the best subset of variables:

$$\text{BIC}(l) = -\log\left(\sum_{i=0}^{l} S^{(i)}\right) + l\log(n)/n.$$

Let $\hat{m} = \arg\min_{0 \leq l \leq m} \text{BIC}(l)$. Then, $\mathcal{M}^{(\hat{m})}$ is the set of selected important variables for the treatment decision, and $a_{\text{opt}}^{(\hat{m})}(\mathbf{x}_{\mathcal{M}^{(\hat{m})}})$ is the estimated optimal treatment regime obtained based on the selected variables $\mathbf{X}_{\mathcal{M}^{(\hat{m})}}$.

In the SAS algorithm, $S^{(k)}$ is the sequential advantage based on the kth selected variable, and the proposed BIC-type criterion balances between the accumulated sequential advantages for making the optimal treatment decision and the size of the model.

3.3. *Extension to multi-stage treatment decisions*. For a study with multiple treatment decisions that has the data structure as shown in Section 2.1, we use a modified Q-learning algorithm to estimate the optimal dynamic treatment regime via backward induction. We apply the SAS algorithm at each stage to select important variables for treatment decision making and use these variables to model Q-functions. The sequential advantage selection algorithm for multiple treatment decisions is given as follows:

(i) At the Kth stage, the response is Y and the covariates are $H_K = \{\mathbf{X}_1, A_1, \ldots, A_{K-1}, \mathbf{X}_K\}$. Following the SAS algorithm, $\hat{m}_K$ variables are selected, and the set of indexes of selected variables is denoted by $\widehat{\mathcal{M}}_K$. The Q-function at the Kth stage based on the selected variables is

$$Q_K(h_{K,\widehat{\mathcal{M}}_K}, a_K) = \mathbb{E}(Y|H_{K,\widehat{\mathcal{M}}_K} = h_{K,\widehat{\mathcal{M}}_K}, A_K = a_K).$$

In addition, the contrast function is $C_K(h_{K,\widehat{\mathcal{M}}_K}) = Q_K(h_{K,\widehat{\mathcal{M}}_K}, 1) - Q_K(h_{K,\widehat{\mathcal{M}}_K}, 0)$. Then, the corresponding optimal treatment regime and value function at the

Kth stage are

$$d_K^{\text{opt}}(h_{K,\widehat{\mathcal{M}}_K}) = I\{C_K(h_{K,\widehat{\mathcal{M}}_K}) \geq 0\},$$

$$V_K(h_{K,\widehat{\mathcal{M}}_K}) = Y + C_K(h_{K,\widehat{\mathcal{M}}_K})\{d_K^{\text{opt}}(h_{K,\widehat{\mathcal{M}}_K}) - a_K\}.$$

(ii) At the kth stage ($k = K-1, \ldots, 1$), use $V_{k+1}(h_{k+1,\widehat{\mathcal{M}}_{k+1}})$ from the previous stage as the response, and the covariates at this stage are $H_k = \{\mathbf{X}_1, A_1, \ldots, A_{k-1}, \mathbf{X}_k\}$. Following the SAS algorithm, $\hat{m}_k$ variables are selected. Similar to the Kth stage, we can define $\widehat{\mathcal{M}}_k$ and derive the Q-function $Q_k(h_{k,\widehat{\mathcal{M}}_k}, a_k)$ and contrast function $C_k(h_{k,\widehat{\mathcal{M}}_k})$ based on the selected variables. Then, the corresponding optimal treatment regime and value function at the kth stage are

$$d_k^{\text{opt}}(h_{k,\widehat{\mathcal{M}}_k}) = I\{C_k(h_{k,\widehat{\mathcal{M}}_k}) \geq 0\},$$

$$V_k(h_{k,\widehat{\mathcal{M}}_k}) = V_{k+1}(h_{k+1,\widehat{\mathcal{M}}_{k+1}}) + C_k(h_{k,\widehat{\mathcal{M}}_k})\{d_k^{\text{opt}}(h_{k,\widehat{\mathcal{M}}_k}) - a_k\}.$$

In the above algorithm, the value function is estimated based on the contrast function, which is different from the classical Q-learning algorithm where the value function is estimated based on the Q-function directly. Compared with the Q-function-based value estimation, the contrast-function based value estimation is more robust in the sense that the baseline model is not required to be specified correctly. As the method targets prescriptive variables that have qualitative interaction with treatment, the contrast-function-based value estimation is more suitable.

4. Simulation studies. In this section we conducted simulation studies to evaluate the performance of the proposed method in both single-stage and multi-stage treatment decisions studies.

4.1. *Single-stage treatment decision study.* The performance of the proposed SAS method is evaluated and compared with the S-score method and the method proposed by Lu, Zhang and Zeng (2013) under various settings. The S-score method was implemented as follows: we first identified all variables with nonzero S-scores and ranked the importance of variables based on their S-scores in decreasing order. Finally, we selected the variables with the largest k nonzero S-scores, where k was chosen as the number of important prescriptive variables selected by our SAS method for easy comparison. Note that the focus here is to compare the performance of sequential advantage and S-score in terms of variable ranking. Therefore, the S-score method considered here is different from the original S-score method proposed by Gunter, Zhu and Murphy (2011), which is a hybrid algorithm that combines S-score ranking and weighted LASSO selection.

For the method of Lu, Zhang and Zeng (2013), we considered LASSO selection based on a least square loss with constant baseline, that is,

$$\min_{\alpha,\boldsymbol{\beta}} \sum_{i=1}^{n} [Y_i - \alpha - \{A_i - \pi(\mathbf{X}_i)\}\boldsymbol{\beta}^T \tilde{\mathbf{X}}_i]^2 + \lambda \sum_{j=0}^{p} |\beta_j|,$$

where $\pi(\mathbf{X}_i) = P(A_i = 1|\mathbf{X}_i)$ is the propensity score, $\tilde{\mathbf{X}}_i = (1, \mathbf{X}_i^T)^T$, and $\boldsymbol{\beta} = (\beta_0, \beta_1, \ldots, \beta_p)^T$. In our simulations, $\pi(\mathbf{X}_i)$ is constant and is estimated by the sample proportion. This method was implemented using the R-package `glmnet`, and the tuning parameter λ was chosen by the built-in cross-validation. We refer to this method as LASSO.

We consider the following four models to generate simulation data:

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

Although all four models have linear interaction forms between covariates and treatment, they have different functional forms for the baseline effects. In our SAS method, the forward selection is based on the working model: $\mathbb{E}(Y) = \boldsymbol{\gamma}^T \tilde{\mathbf{X}} + A\boldsymbol{\beta}^T \tilde{\mathbf{X}}$, which is correctly specified under Models I and III but is misspecified under Models II and IV. Models I and II have three important prescriptive variables (X_1, X_9, X_{10}), while Models III and IV have eight important prescriptive variables $(X_1, X_9, X_{10}, X_{21}, X_{22}, X_{23}, X_{29}, X_{30})$. Covariates $\mathbf{X} = (X_1, \ldots, X_p)^T$ are generated from a multivariate normal distribution: each entry is normal with mean zero, variance one, and the correlation between covariates is $\mathrm{Corr}(X_j, X_k) = \rho^{|j-k|}$, for $j \neq k$, $j, k = 1, \ldots, p$. Here, ρ is chosen to be 0.2, 0.5 and 0.8, representing weak, moderate and strong correlations. We considered randomized trials, where A is generated from a Bernoulli distribution with the success probability of 0.5. The error term, ε, is normally distributed with mean zero and variance 0.25. We ran 500 simulations for each scenario with $n = 200$ and $p = 1000$.

Because the generative models are complex, it becomes rather difficult to evaluate the degree of qualitative interaction of each variable with treatment. As an illustration, we show in Figure 1 the marginal interaction plots of variables X_1, X_9 and X_{10} with treatment under two scenarios: $\rho = 0.2$ and $\rho = 0.8$. These marginal plots are for one simulated data under Model I, where X_1, X_9 and X_{10} are important prescriptive variables. Based on Figure 1, when the correlation is weak ($\rho = 0.2$), all variables show clear qualitative interaction with treatment; when the correlation is strong ($\rho = 0.8$), either variable X_9 or X_{10} (here is X_9) has nearly no qualitative interaction with treatment. This is possibly due to the fact that these two variables have strong positive correlation but opposite covariate effects. This result implies that the S-score method may fail to identify one of the variables because the method relies on the measures for the marginal qualitative interaction.

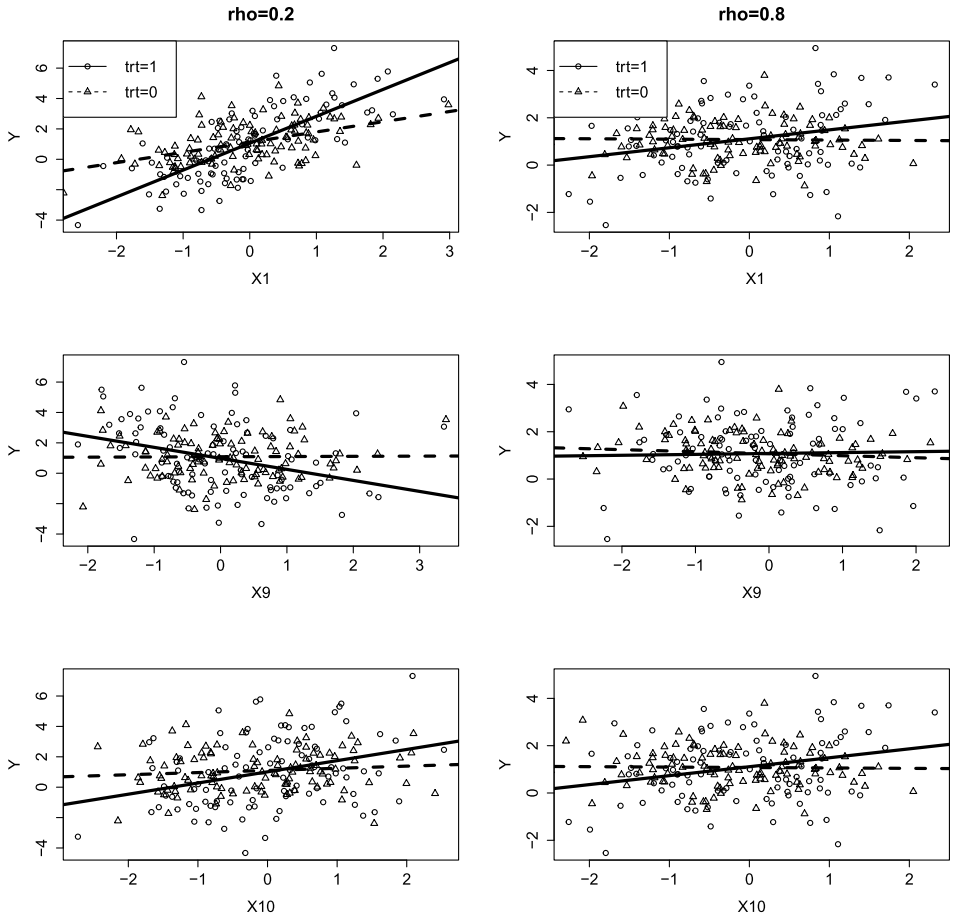

FIG. 1. *Plots of the marginal interaction of covariates* X_1, X_9 *and* X_{10} *with treatment* (*triangles are for treatment* 1, *and circles are for treatment* 0). *The fitted lines for treatment* 1 (*dashed*) *and treatment* 0 (*dotted*) *are from simple linear regression. The left panel is for* $\rho = 0.2$; *the right panel is for* $\rho = 0.8$.

Table 1 summarizes simulation results for variable selection and estimated optimal treatment regimes of the three methods. For variable selection, we report size and true positive (TP), which are the average numbers of selected variables and correctly identified prescriptive variables over 500 simulations, respectively. For assessing estimated optimal treatment regimes, we compute the mean value ratio between the value following the estimated optimal treatment regime, $Q(\hat{g}^{\text{opt}})$, and the value following the true optimal treatment regime, $Q(g^{\text{opt}})$, denoted by $\text{VR} = Q(\hat{g}^{\text{opt}})/Q(g^{\text{opt}})$. Here, the value of a given treatment regime is computed by averaging outcomes generated from the true model with the treatment dictated by the considered regime using Monte Carlo simulations with 10,000 replicates. In addition, we report the mean error rates of the estimated optimal treatment regimes

TABLE 1
Simulation results of sequential advantage selection (SAS), S-score and LASSO methods in the single-stage treatment decision study

	SAS				S-score				LASSO			
ρ	Size	TP	VR	ER	Size $S \neq 0$	TP	VR	ER	Size	TP	VR	ER
					Model I							
0.2	6.70	2.98	99.1	5.6	765.28	1.61	84.7	27.5	21.94	1.75	79.5	32.8
	(0.06)	(0.01)	(0.1)	(0.1)	(5.16)	(0.02)	(0.2)	(0.3)	(0.74)	(0.03)	(0.4)	(0.3)
0.5	7.56	2.91	98.5	7.4	757.18	1.31	86.7	26.6	15.04	1.37	84.2	29.4
	(0.08)	(0.01)	(0.1)	(0.2)	(5.24)	(0.02)	(0.1)	(0.2)	(0.59)	(0.03)	(0.3)	(0.3)
0.8	8.21	1.76	94.2	17.2	738.44	1.04	94.2	18.8	11.44	1.10	93.0	20.6
	(0.07)	(0.03)	(0.1)	(0.2)	(5.48)	(0.01)	(0.0)	(0.1)	(0.45)	(0.01)	(0.1)	(0.2)
					Model II							
0.2	11.14	2.24	89.8	28.3	765.27	1.73	87.7	31.5	15.84	1.48	86.3	34.3
	(0.10)	(0.03)	(0.2)	(0.3)	(5.23)	(0.02)	(0.2)	(0.3)	(0.68)	(0.03)	(0.2)	(0.3)
0.5	11.81	1.82	88.9	30.8	758.89	1.30	87.3	33.7	13.36	1.10	87.5	33.5
	(0.09)	(0.03)	(0.2)	(0.2)	(5.34)	(0.02)	(0.1)	(0.2)	(0.70)	(0.03)	(0.2)	(0.3)
0.8	10.84	1.36	90.4	29.5	749.84	1.09	92.2	26.7	11.65	0.98	92.1	26.0
	(0.09)	(0.02)	(0.1)	(0.2)	(5.38)	(0.01)	(0.1)	(0.2)	(0.48)	(0.02)	(0.2)	(0.3)
					Model III							
0.2	11.73	5.13	84.2	18.3	783.39	3.38	74.7	29.1	27.09	4.03	73.4	30.9
	(0.13)	(0.12)	(0.7)	(0.5)	(7.13)	(0.05)	(0.4)	(0.3)	(1.15)	(0.10)	(0.4)	(0.3)
0.5	10.41	4.67	87.0	18.6	776.15	2.88	78.3	28.1	25.05	3.24	77.6	29.3
	(0.11)	(0.10)	(0.5)	(0.4)	(7.38)	(0.04)	(0.3)	(0.2)	(1.17)	(0.08)	(0.3)	(0.3)
0.8	7.74	3.01	90.0	19.6	760.68	2.93	90.9	20.3	17.37	2.49	88.5	22.2
	(0.10)	(0.05)	(0.1)	(0.2)	(7.62)	(0.03)	(0.1)	(0.2)	(0.67)	(0.04)	(0.2)	(0.2)
					Model IV							
0.2	11.85	3.29	81.4	29.1	779.14	3.21	81.2	30.5	23.80	3.50	80.5	32.6
	(0.11)	(0.09)	(0.4)	(0.4)	(7.56)	(0.05)	(0.2)	(0.3)	(1.12)	(0.10)	(0.3)	(0.4)
0.5	11.68	2.80	82.9	29.6	769.07	2.67	82.3	31.1	18.83	2.51	82.4	32.2
	(0.11)	(0.07)	(0.3)	(0.3)	(7.65)	(0.04)	(0.2)	(0.2)	(1.01)	(0.07)	(0.3)	(0.3)
0.8	9.68	2.47	88.6	25.7	758.69	2.82	90.8	23.7	15.72	2.18	89.2	25.6
	(0.11)	(0.05)	(0.2)	(0.2)	(7.58)	(0.03)	(0.2)	(0.2)	(0.64)	(0.04)	(0.2)	(0.3)

Size: the average number of selected variables; Size $S \neq 0$: the average number of variables with nonzero S-scores; TP: the average number of correctly identified prescriptive variables (the true value is three under Models I and II, and is eight under Models III and IV); VR: the mean value ratio; ER: the mean error rate. Sample standard deviations are shown in parentheses.

for treatment decision making compared with the true optimal treatment regimes, denoted by ER. The numbers given in parentheses are the associated sample standard deviations.

The results in Table 1 show that the SAS method selects more true prescriptive variables in most cases but with fewer selected variables. For example, under Model I, SAS has size $= 6.7$ and TP $= 2.98$, whereas S-score has TP $= 1.61$ and

LASSO has TP = 1.75 with size = 21.94. In addition, it is observed that there are too many variables with nonzero S-scores and that the LASSO method tends to select more variables than the SAS method, especially when $\rho = 0.2$ and 0.5. Compared to the marginal S-score method, the SAS method includes more true positives in most cases, which indicates that the sequential advantage is a better characterization of prescriptive variables than the S-score. Under Model I with weak and moderate correlations, the SAS method can recover almost all of the important variables. However, for the other three models, all three methods missed a few important variables due to the weak signals of these variables and/or model misspecification.

Based on results about values and error rates on estimated optimal treatment regimes in Table 1, the SAS method provides good estimates of optimal treatment regimes with values close to the true optimal values and low error rates among all three methods. The error rates provided by the LASSO method are high in most cases; this is partly because the LASSO estimates tend to have large bias due to shrinkage. As correlation increases, the values of estimated treatment regimes are less affected by the quality of the variable selection because some variables may be good surrogates for the true variables when estimating the optimal treatment regime.

We also compare solution paths of the three methods in Figure 2. Here, we define a solution path as the trajectory of the number of identified important variables as the number of selected variables increases according to the selection order. For demonstration purposes, we only plot the solution paths for the first 30 selected variables. The SAS method has a natural order of selected variables. For the S-score method, we ranked the variables in descending order of the S-scores of these variables. The LASSO method has a solution path of $\boldsymbol{\beta}$, which can be used to determine the order of variables entering the model. The solution path plots allow us to evaluate the ability of each method to identify important variables given that the same number of variables is selected.

Figure 2 indicates that when the size is fixed, the SAS method includes the largest number of important variables in most cases. When the data are generated under Model I, the SAS method can include all of the important variables quickly under weak and moderate correlations. However, under strong correlations, they are likely to be missed by all three methods because some important variables are highly correlated. When the model is misspecified, the SAS method is slightly better in Model II, while all three methods do not differ significantly in Model IV.

Overall, the SAS method performs well on both the aspect of variable selection and the aspect of estimating optimal treatment regime. The SAS method can select most important variables at a moderate size of the selected variables. When the model is correctly specified and the correlations between covariates are not too high, the SAS method is able to identify all important variables. Moreover, the error rates of the optimal treatment regime based on the SAS method are low, and the estimated values are close to the true optimal value.

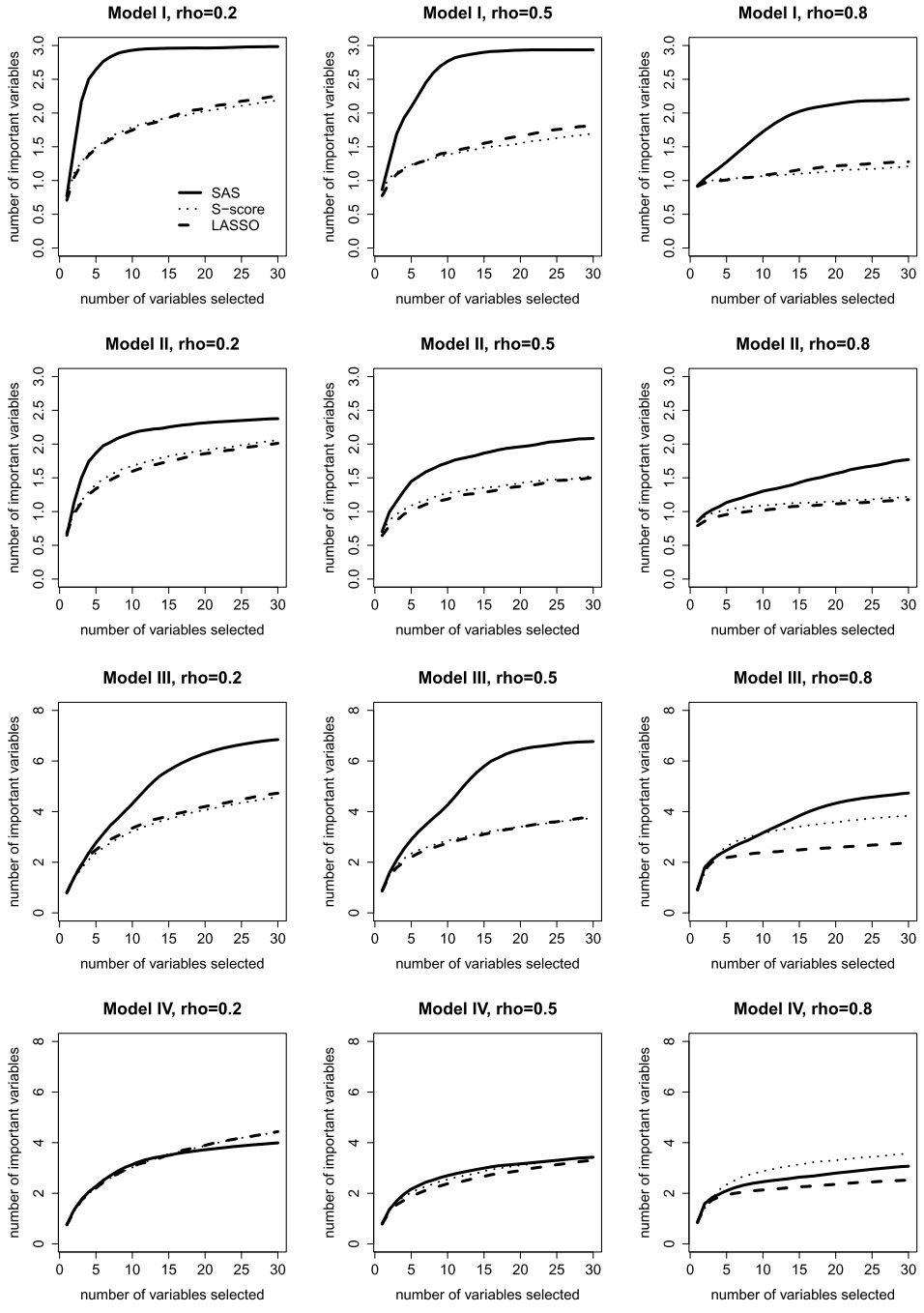

FIG. 2. *Solution paths of sequential advantage selection* (*SAS*, *solid line*), *S-score* (*doted line*) *and LASSO* (*dashed line*) *methods for the single-stage treatment decision simulation study.*

4.2. *Multi-stage treatment decisions study.* To illustrate the sequential advantage selection algorithm for the multi-stage treatment decisions (MTD) study, we applied the SAS algorithm to simulated data with two-stage treatment decisions based on the following generative model for the final response:

$$Y = A_1A_2 + A_2(a + \boldsymbol{\beta}_{12}^T\mathbf{X}_1 + \boldsymbol{\beta}_{21}^T\mathbf{X}_2) + A_1(a + \boldsymbol{\beta}_{11}^T\mathbf{X}_1) + \varepsilon, \tag{4}$$

where A_k, the treatment at stage k, follows a Bernoulli distribution with parameter 0.5 for $k = 1$ and 2. The covariates collected at baseline, $\mathbf{X}_1$, include $p_1 = 500$ variables and are denoted as $\mathbf{X}_1 = (X_{1,1}, X_{1,2}, \ldots, X_{1,p_1})^T$. We generate $\mathbf{X}_1$ from a multivariate normal distribution with mean zero, variance one and correlation $\operatorname{corr}(X_{1,j}, X_{1,l}) = 0.2^{|j-l|}$, $j \neq l$. The intermediate covariates collected at the second stage are denoted by $\mathbf{X}_2$. For demonstration purposes, we consider a one-dimensional intermediate covariate X_2 and assume that $X_2 = c_0 + c_1X_{1,1} + c_2A_1 + c_3A_1X_{1,1} + e$, where the normal random error e has mean zero and variance σ_2^2. The random error for response Y, ε, is normally distributed with mean zero and variance σ_1^2.

The parameter values for the above two-stage model are chosen as follows: $\boldsymbol{\beta}_{12} = (0, 0, 1, -1, \mathbf{0}_{p_1-4})^T$, $\beta_{21} = 1$, $\boldsymbol{\beta}_{11} = (\mathbf{0}_4, 1, -1, \mathbf{0}_{p_1-6})^T$, $a = 0$. For the standard deviations of two random errors, we choose $\sigma_1 = \sigma_2 = 0.5$. For the parameter $\mathbf{c} = (c_0, c_1, c_2, c_3)^T$ in the model for X_2, we consider three sets of values to evaluate the carry-on effects of baseline variables through intermediate covariates: $\mathbf{c} = (0, 1, 0, 0)^T$, $(0, 0, 1, 0)^T$ and $(0, 1, 1, 1)^T$.

Based on the generative model (4), it is clear that the optimal treatment regime at stage 2 is $g_2^{\text{opt}}(\mathbf{x}_1, a_1, x_2) = 1(a_1 + \boldsymbol{\beta}_{12}^T\mathbf{x}_1 + \beta_{21}x_2 \geq 0)$. Thus, four variables $(X_2, A_1, X_{1,3}, X_{1,4})$ determine the optimal treatment regime at stage 2. At stage 1, the Q-function is

$$\begin{aligned} Q_1(\mathbf{X}_1, A_1) &= E\{|A_1 + \boldsymbol{\beta}_{12}^T\mathbf{X}_1 + \beta_{21}X_2|_+|\mathbf{X}_1, A_1\} + A_1(\boldsymbol{\beta}_{11}^T\mathbf{X}_1) \\ &= \sigma_2\frac{1}{\sqrt{2\pi}}\exp\left\{-\frac{\mu_1^2}{2\sigma_2^2}\right\} + \mu_1[1 - \Phi(-\mu_1/\sigma_2)] + A_1(\boldsymbol{\beta}_{11}^T\mathbf{X}_1), \end{aligned}$$

where $\mu_1 = A_1 + \boldsymbol{\beta}_{12}^T\mathbf{X}_1 + \boldsymbol{\beta}_{21}[c_0 + c_1X_{1,1} + c_2A_1 + c_3A_1X_{1,1}]$. The optimal treatment regime at stage 1 is $g_1^{\text{opt}}(\mathbf{x}_1) = 1\{Q_1(\mathbf{x}_1, 1) > Q_1(\mathbf{x}_1, 0)\}$. There are five important variables $(X_{1,1}, X_{1,3}, X_{1,4}, X_{1,5}, X_{1,6})$ for determining the optimal treatment regime at stage 1 when $\mathbf{c} = (0, 1, 0, 0)^T$ and $(0, 1, 1, 1)^T$, and four important variables $(X_{1,3}, X_{1,4}, X_{1,5}, X_{1,6})$ when $\mathbf{c} = (0, 0, 1, 0)^T$. Table 2 summarizes the optimal treatment regimes and the important variables in this simulation study. Although the optimal treatment regime and the corresponding important variables at stage 2 are explicitly defined, the optimal treatment regime at stage 1 takes a complex nonlinear form. Therefore, the effects of the important variables are difficult to evaluate.

We applied the SAS algorithm to the simulated data for sample sizes $n = 100$, 200 and 400 over 100 replications. Simulation results are summarized in Tables 3

TABLE 2
Optimal treatment regimes and corresponding important variables in multi-stage simulation study

	Optimal treatment regime		**Important variables**
Stage 2	$1(a_1 + \boldsymbol{\beta}_{12}^T \mathbf{x}_1 + \beta_{21} x_2 \geq 0)$		$(X_2, A_1, X_{1,3}, X_{1,4})$
Stage 1	$1\{Q_1(\mathbf{x}_1, 1) > Q_1(\mathbf{x}_1, 0)\}$	$\mathbf{c} = (0, 1, 0, 0)^T$ and $(0, 1, 1, 1)^T$	$(X_{1,1}, X_{1,3}, X_{1,4}, X_{1,5}, X_{1,6})$
		$\mathbf{c} = (0, 0, 1, 0)^T$	$(X_{1,3}, X_{1,4}, X_{1,5}, X_{1,6})$

and 4. Table 3 presents results on variable selection and estimated optimal treatment regimes at both stages 1 and 2, where the same statistics as in Table 1 are reported (Size, TP, VR and ER). For the mean value ratio at stage 2 (VR2), we adopt random treatment regimes at stage 1 to calculate the outcome values because the optimal treatment regimes at stage 1 have not been estimated at this stage. Table 4 reports the proportions of each important variable being selected for all scenarios.

According to Table 3, the numbers of true positives at stage 2 increase and get close to the true number (four) when the sample size gets large; at stage 1, the numbers of true positives also increase, but 1–2 variables are missed. We will examine which variables are missed when analyzing results in Table 4. Based on the results for values and error rates, SAS provides good estimated optimal treatment regimes

TABLE 3
Simulation results of SAS–MTD method for two-stage treatment decisions study

	Stage 2				**Stage 1**			
n	**Size**	**TP**	**VR2**	**ER**	**Size**	**TP**	**VR1**	**ER**
				$\mathbf{c} = (0, 1, 1, 1)^T$				
100	5.22 (0.46)	2.08 (0.09)	85.3	17.8	6.29 (0.36)	0.70 (0.07)	73.1	26.9
200	4.08 (0.11)	3.25 (0.06)	94.3	8.8	5.43 (0.25)	2.34 (0.10)	93.4	14.5
400	4.02 (0.06)	3.75 (0.04)	97.5	4.4	3.61 (0.11)	3.14 (0.04)	98.2	8.5
				$\mathbf{c} = (0, 1, 0, 0)^T$				
100	6.70 (0.28)	1.80 (0.11)	67.7	24.9	8.88 (0.35)	0.45 (0.07)	49.1	39.8
200	6.38 (0.21)	3.48 (0.06)	89.7	13.5	11.80 (0.27)	1.78 (0.08)	78.2	26.6
400	5.82 (0.19)	3.88 (0.03)	96.1	7.8	13.01 (0.36)	2.41 (0.07)	92.7	16.6
				$\mathbf{c} = (0, 0, 1, 0)^T$				
100	5.93 (0.29)	1.94 (0.10)	80.2	21.6	4.96 (0.41)	0.63 (0.08)	71.4	25.0
200	5.13 (0.15)	3.15 (0.04)	94.8	11.5	5.75 (0.34)	1.89 (0.08)	91.0	15.6
400	4.70 (0.16)	3.53 (0.05)	97.8	7.0	4.12 (0.23)	2.91 (0.09)	97.4	8.5

Size: the average number of selected variables; TP: the number of correctly identified important variables; VR2: the mean value ratio, where random treatment regimes are adopted at stage 1; VR1: the mean value ratio between the estimated and true treatment regimes; ER: the mean error rate. VR1, VR2 and ER are presented in the percentage scale. Sample standard errors are shown in parenthesis.

TABLE 4
Proportion of each important variable being selected at stage 2 *and stage* 1

	Stage 2				Stage 1				
n	$X_{1,3}$ 1	$X_{1,4}$ −1	X_2 1	A_1 1	$X_{1,3}$ ∗	$X_{1,4}$ ∗	$X_{1,5}$ 1	$X_{1,6}$ −1	$X_{1,1}$ ∗/−
				$\mathbf{c} = (0, 1, 1, 1)^T$					
100	0.58	0.49	0.85	0.16	0.01	0.01	0.13	0.08	0.47
200	0.95	0.96	0.97	0.37	0.02	0.03	0.72	0.71	0.86
400	1.00	1.00	1.00	0.75	0.07	0.07	1.00	1.00	1.00
				$\mathbf{c} = (0, 1, 0, 0)^T$					
100	0.59	0.51	0.55	0.15	0.01	0.00	0.24	0.16	0.04
200	0.99	1.00	0.92	0.57	0.03	0.04	0.82	0.84	0.05
400	1.00	1.00	0.98	0.90	0.17	0.07	1.00	1.00	0.17
				$\mathbf{c} = (0, 0, 1, 0)^T$					
100	0.57	0.49	0.55	0.33	0.05	0.03	0.26	0.29	–
200	1.00	1.00	0.79	0.36	0.13	0.10	0.82	0.84	–
400	1.00	1.00	0.96	0.57	0.54	0.39	0.99	0.99	–

Important variables at stage 2 are $X_{1,3}, X_{1,4}, X_2$ and A_1, and the corresponding coefficients in the true optimal treatment regime at stage 2 are $(1, -1, 1, 1)$. Important variables at stage 1 are $X_{1,3}, X_{1,4}, X_{1,5}, X_{1,6}, X_{1,1}$ for the first two scenarios and $X_{1,3}, X_{1,4}, X_{1,5}, X_{1,6}$ for the third scenario. The coefficients for $X_{1,5}$ and $X_{1,6}$ are 1 and -1 in the true optimal treatment regime at stage 1; $X_{1,3}, X_{1,4}$ and/or $X_{1,1}$ appear in the true optimal treatment regime at stage 1 with a nonlinear form. "*" means this variable is important for the treatment decision and the coefficient is unknown, while "–" means this variable is not important for the treatment decision.

at both stages 1 and 2 when the sample size is large. We note that performances for $\mathbf{c} = (0, 1, 0, 0)^T$ at stage 1 are worse than the other scenarios. This indicates that the manner in which the intermediate variable depends on covariates from the last stage also affects the quality of the estimated optimal treatment regime. It is not apparent why the case with $\mathbf{c} = (0, 1, 0, 0)^T$ performs worse, and results in Table 4 partially explain this phenomenon.

Table 4 shows more detailed variable selection results. At stage 2, all but A_1 among the four important variables can almost always be selected when the sample size is large. A_1 can be selected more often for $\mathbf{c} = (0, 1, 0, 0)^T$ than for the other two scenarios; this may be because X_2 depends on A_1 in these cases, which partially eliminates the effects of A_1 on the final response, Y. At stage 1, only variables $X_{1,5}$ and $X_{1,6}$ can always be selected for all three scenarios when $n = 400$. These two variables appear in the optimal treatment regime at stage 1 in a linear form. On the contrary, variables $X_{1,3}$, $X_{1,4}$ and $X_{1,1}$ that present in a nonlinear form are not always selected. The probabilities of selecting $X_{1,3}$ and $X_{1,4}$ for all three scenarios are low. This may be because these two variables do not have substantial effects on the optimal treatment regime; the high values and low error rates

at stage 1 in Table 3 also verify this argument. The probabilities of including $X_{1,1}$ differ between the first two scenarios. A possible explanation is that $X_{1,1}$ interacts with A_1 in μ_1 for the first scenario, which makes its sequential advantage for being selected large. This may also explain why the scenario with $\mathbf{c} = (0, 1, 1, 1)^T$ performs better than the scenario with $\mathbf{c} = (0, 1, 0, 0)^T$ in Table 3.

Based on these results, the SAS algorithm performs well on both variable selection and optimal treatment regime estimation. The complex form of the optimal treatment regime at stage 1 makes it more difficult to identify important variables and brings a challenge for variable selection.

5. Application to STAR*D study. We apply the proposed method to data from the STAR*D study, which was conducted to determine the effectiveness of different treatments for patients with major depressive disorder (MDD) who had not been adequately benefiting from initial treatment with an antidepressant. There were 4041 participants (ages 18–75) with nonpsychotic MDD enrolled in this study. Initially, these participants were treated with citalopram (CIT) up to 14 weeks. Subsequently, 3 more levels of treatments were provided for participants without a satisfactory response to CIT. At Level 2, participants were eligible for seven treatment options, which may be conceptualized as two treatment strategies: medication or psychotherapy switch, and medication or psychotherapy augmentation. Available treatments for participants to switch were as follows: sertraline (SER), venlafaxine (VEN), bupropion (BUP) and cognitive therapy (CT); available treatments for patients to augment were the following: augmenting CIT with bupropion (CIT+BUP), buspirone (CIT+BUS) or cognitive therapy (CIT+CT). Participants without a satisfactory response to CT were provided additional medication treatments, which is called Level 2A. All participants who did not respond satisfactorily at Level 2 or 2A were eligible for Level 3, where possible treatments were medication switch to mirtazapine (MIRT) or nortriptyline (NTP), and medication augmentation with either lithium (Li) or thyroid hormone (THY). Participants without satisfactory response to Level 3 were re-randomized at Level 4 to either tranylcypromine (TCP) or a combination of mirtazapine and venlafaxine (MIRT+VEN). Participants who responded satisfactorily were followed up to 1 year. See Fava et al. (2003) and Rush et al. (2004) for a more detailed description of this STAR*D design.

For illustration, we focus on a subset of participants who were given treatment BUP or SER at Level 2, did not receive satisfactory responses, and were randomized to treatment MIRT or NTP at Level 3. There were 73 participants who meet this condition. Among these participants, 36 were treated with BUP and 37 were treated with SER at Level 2, and 33 were treated with NTP and 40 were treated with MIRT at Level 3. Our goal is to identify relevant prescriptive predictors and estimate optimal dynamic treatment regimes at Levels 2 and 3 that maximize the mean response at the end of Level 3. We consider 381 covariates as possible relevant predictors, which are listed in Table 5. These covariates include participant

TABLE 5
*List of covariates used in the analysis of STAR*D study*

Participant features	
1 Gender	2–6 Ethnicity
7 Economic study consent	8 Depressed mood
9 Diminished interest or pleasure	10 Weight loss while not dieting
11 Insomnia or hypersomnia	12 Psychomotor agitation or retardation
13 Fatigue or loss of energy	14 Feelings of worthlessness or guilt
15 Diminished ability to concentrate	16 Recurrent thoughts of death or suicide
17 Age	18 Number of relatives living with patient
19 Number of friends living with patient	20 Total number of persons in household
21 Years of schooling completed	22 Highest degree received
23 On medical or psychiatric leave	24 Medicare
25 Medicaid	26 Private insurance
27 Better able to make important decisions	28 Better able to enjoy things
29 Impact of your family and friends	30–35 Current marital status
36–41 Current employment status	42–44 Currently a student
45–46 Currently do volunteer work	
Illness features	
47–60 Cumulative Illness Rating Scale	61–78 Hamilton rating scale for depression
79–82 Medication history	83–221 Psychiatric diagnostic screening questionnaire
222 Baseline Axis I psychiatric condition	224 Family hx depression
223 Baseline Axis II psychiatric condition	226 Family hx alcohol abuse
225 Family hx bipolar disorder	228 Family hx suicide
227 Family hx drug abuse	
Care features	
229 Type of clinical site	
Intermediate medical conditions at level 1	
230 QIDS-C score change rate	231 AIDS-C percent improvement
232 QIDS-SR score change rate	233 FISER frequency score change rate
234 FISER intensity score change rate	235 GRSEB score change rate
236 CGII score change rate	237 Patient presently a suicide risk
238 Patient in remission	239 Study medical daily dose
240–290 Patient rated inventory of side effects	291–305 Quick Inventory of Depressive Symptomatology
Intermediate medical conditions at level 2	
306 QIDS-C score change rate	307 AIDS-C percent improvement
308 QIDS-SR score change rate	309 FISER frequency score change rate
310 FISER intensity score change rate	311 GRSEB score change rate
312 CGII score change rate	313 Patient presently a suicide risk
314 Patient in remission	315 Study medical daily dose
316–366 Patient rated inventory of side effects	367–381 Quick Inventory of Depressive Symptomatology

features such as age, gender, socioeconomic status and ethnicity; illness features such as medication history and family history of mood disorders; and care features such as clinician type. Intermediate medical conditions from Levels 1 and 2, such

as degree of symptom improvement and side effect burden, are also considered. For treatment regime at Level 3, all 381 covariates and the treatment at Level 2 are considered as possible predictors. For the treatment regime at Level 2, the intermediate medical conditions at Level 2 are no longer available, thus there are only 305 covariates considered for treatment decision making. We used negative 16-item Quick Inventory of Depressive Symptomatology-Clinician-Rated (QIDS-C_{16}) at the end of Level 3 as the final response, which is a measurement of symptomatic status. Because low QIDS-C_{16} stands for remission, the negative QIDS-C_{16} was used such that a larger value indicates better response.

We apply the SAS algorithm to this data set. The results are as follows. At Level 3, there are four covariates selected based on the BIC criterion: "ringing in ears" in patient rated inventory of side effects at Level 2 (EARNG-Level2), "hard to control worrying" in psychiatric diagnostic screening questionnaire at baseline (WYCRL), "feeling of worthlessness or guilt" in baseline protocol eligibility (DSMFW), and "fatigue or loss of energy" in baseline protocol eligibility (DSMLE). All four covariates are binary covariates with 1 indicating "Yes" and 0 indicating "No." The estimated optimal treatment regime is $I(-18.57 + 13.79 \times (\text{EARNG-Level2}) - 8.46 \times \text{WYCRL} + 6.36 \times \text{DSMFW} + 16.88 \times \text{DSMLE} \geq 0)$, where 1 stands for treatment NTP and 0 stands for treatment MIRT. This optimal treatment regime assigns 25 participants to NTP and the remaining 48 participants to MIRT. At Level 2, there are seven covariates selected based on the BIC criterion: "TE flashbacks of traumatic event" in the psychiatric diagnostic screening questionnaire at baseline (TEFSH), "EM worry saying something stupid" in the psychiatric diagnostic screening questionnaire at baseline (EMSTP), "QIDS psychomotor agitation" in the quick inventory of depressive symptomatology-clinician at Level 1 (CAGIT), "think drink too much" in the psychiatric diagnostic screening questionnaire at baseline (DKMCH), "QIDS outlook (self)" in the quick inventory of depressive symptomatology-clinician at Level 1 (CVWSF), "IM convinced others spying" in the psychiatric diagnostic screening questionnaire at baseline (IMSPY), and "sleep at least 1–2 hours less 2 weeks" in the psychiatric diagnostic screening questionnaire at baseline (LSL2W). Among these seven covariates, TEFSH, EMSTP, DKMCH, IMSPY and LSL2W are binary, with 1 indicating "Yes" and 0 indicating "No"; CAGIT and CVWSF are categorical covariates with 4 levels indicated by 0 to 3. The estimated optimal treatment regime is $I(-5.50 + 3.91 \times \text{TEFSH} + 11.17 \times \text{EMSTP} + 3.76 \times \text{CAGIT} - 4.65 \times \text{DKMCH} + 4.29 \times \text{CVWSF} + 6.57 \times \text{IMSPY} - 8.48\text{LSL2W} \geq 0)$, where 1 stands for treatment BUP and 0 stands for treatment SER. This optimal treatment regime assigns 39 participants to BUP and the remaining 34 participants to SER.

To further examine the estimated optimal dynamic treatment regime, we estimate the value of the estimated optimal dynamic treatment regime, that is, the mean outcome following the estimated optimal treatment regime, using the inverse

TABLE 6
Estimated values of different treatment regimes and confidence intervals for the differences of values

Treatment regime	Estimated value	Diff	95% CI on Diff
Optimal regime from SAS-MTD	−5.26		
BUP + NTP	−13.27	8.01	[2.83, 13.64]
BUP + MIRT	−11.71	6.45	[1.83, 11.02]
SER + NTP	−13.15	7.89	[1.94, 13.72]
SER + MIRT	−12.63	7.37	[2.88, 11.86]

probability weighted estimator proposed by Zhang et al. (2013), defined as

$$\text{IPW} = \frac{1}{n}\sum_{i=1}^{n}\frac{Y_i I(A_{i,1} = g_1(\mathbf{X}_i), A_{i,2} = g_2(\mathbf{X}_i))}{\pi(A_{i,1})\pi(A_{i,2})}.$$

Here Y_i is the outcome for ith individual, $A_{i,1}$ and $A_{i,2}$ are the treatments given to the ith individual at stage 1 and stage 2, respectively, $g_1(\mathbf{X}_i)$ and $g_2(\mathbf{X}_i)$ are the estimated treatment regimes at stages 1 and 2, and $\pi(A_{i,1})$ and $\pi(A_{i,2})$ are the probabilities of receiving treatment $A_{i,1}$ at stage 1 and treatment $A_{i,2}$ at stage 2, respectively. The estimated value for the estimated optimal dynamic treatment regime based on the SAS algorithm is compared to the estimated values when all subjects are treated with the nondynamic treatment regimes: BUP + NTP, BUP + MIRT, SER + NTP and SER + MIRT. The estimated values are shown in Table 6. We also report the 95% confidence intervals for the differences between values of the estimated optimal dynamic treatment regime and the four nondynamic treatment regimes based on 1000 bootstrap samples. The results show that the value of the estimated optimal dynamic treatment regime based on the SAS algorithm is significantly larger than those of the nondynamic treatment regimes.

6. Discussion. In this article we propose a forward-stepwise variable selection method based on sequential advantage for deriving optimal treatment regimes in both single-stage and multi-stage treatment decision studies. Our method generalizes S-score ranking and directly targets prescriptive variables that are important for decision making. We also propose a BIC-type criterion to select the number of important prescriptive variables needed for treatment decision making. The proposed method can be extended to other types of outcomes, such as categorical or censored survival data.

As suggested by a reviewer, a two-step procedure may be used for selecting important prescriptive variables. For example, in the first step, we fit a flexible regression model of Y given A and X using some tree-based methods, such as BART or GBM. Let $\hat{Q}(X)$ denote the estimated interaction effects of covariates X and treatment indicator A. In the second step, we can consider a classification problem with responses $\text{sign}\{\hat{Q}(X)\}$ and covariates X, and select important

covariates based on high-dimensional classification methods such as penalized logistic regression or support vector machine (SVM). Such a two-step procedure can potentially decrease the chance of missing important covariates as compared to a one-step approach such as the proposed method. Although a two-step procedure looks appealing, it also has limitations. First, when p is much larger than n, the estimation of the interaction effects using a tree-based method is usually quite challenging, especially when the effects are only small to moderate. Second, if the interaction effects are badly estimated in the first step, the resulting classification and selection in the second step can be erroneous.

In addition, inspired by the composite algorithm proposed by Gunter, Zhu and Murphy (2011), a two-step hybrid procedure can be built based on the SAS method. Specifically, in the first step, we use a penalized regression method to select important variables both in the main effects and in the interaction effects based on an assumed model. In the second step, we apply the SAS method based on selected variables from the first step. Such a hybrid procedure generally may have better selection performance than a single-step selection method.

Acknowledgments. The authors thank the National Institute of Mental Health for providing the STAR*D data. They thank the Editor, the Associate Editor and a referee for their comments that substantially improved the article. They also thank Dr. Shannon Holloway for her constructive input and proofreading to improve the manuscript.

REFERENCES

CHAKRABORTY, B., MURPHY, S. and STRECHER, V. (2010). Inference for nonregular parameters in optimal dynamic treatment regimes. *Stat. Methods Med. Res.* **19** 317–343. MR2757118

FAVA, M., RUSH, A. J., TRIVEDI, M. H., NIERENBERG, A. A., THASE, M. E., SACKEIM, H. A., QUITKIN, F. M., WISNIEWSKI, S., LAVORI, P. W., ROSENBAUM, J. F. et al. (2003). Background and rationale for the sequenced treatment alternatives to relieve depression (STAR*D) study. *Psychiatric Clinics of North America* **26** 457–494.

GAIL, M. and SIMON, R. (1985). Testing for qualitative interactions between treatment effects and patient subsets. *Biometrics* **41** 361–372.

GUNTER, L., ZHU, J. and MURPHY, S. A. (2011). Variable selection for qualitative interactions. *Stat. Methodol.* **8** 42–55. MR2741508

LU, W., ZHANG, H. H. and ZENG, D. (2013). Variable selection for optimal treatment decision. *Stat. Methods Med. Res.* **22** 493–504. MR3190671

MOODIE, E. E. M., RICHARDSON, T. S. and STEPHENS, D. A. (2007). Demystifying optimal dynamic treatment regimes. *Biometrics* **63** 447–455. MR2370803

MURPHY, S. A. (2003). Optimal dynamic treatment regimes. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **65** 331–366. MR1983752

MURPHY, S. A. (2005a). An experimental design for the development of adaptive treatment strategies. *Stat. Med.* **24** 1455–1481. MR2137651

MURPHY, S. A. (2005b). A generalization error for Q-learning. *J. Mach. Learn. Res.* **6** 1073–1097. MR2249849

MURPHY, S. A., VAN DER LAAN, M. J. and ROBINS, J. M. (2001). Marginal mean models for dynamic regimes. *J. Amer. Statist. Assoc.* **96** 1410–1423. MR1946586

PETO, R. (1982). Statistical aspects of cancel trials. In *Treatment of Cancer* (K. E. Halnan, ed.) 867–871. Chapman, London, UK.

PIANTADOSI, S. and GAIL, M. H. (1993). A comparison of the power of two tests for qualitative interactions. *Stat. Med.* **12** 1239–1248.

QIAN, M. and MURPHY, S. A. (2011). Performance guarantees for individualized treatment rules. *Ann. Statist.* **39** 1180–1210. MR2816351

QIAN, M., NAHUM-SHANI, I. and MURPHY, S. A. (2013). Dynamic treatment regimes. In *Modern Clinical Trial Analysis* 127–148. Springer, New York.

ROBINS, J. (1986). A new approach to causal inference in mortality studies with a sustained exposure period—Application to control of the healthy worker survivor effect. *Math. Modelling* **7** 1393–1512. MR0877758

ROBINS, J. M. (1997). Causal inference from complex longitudinal data. In *Latent Variable Modeling and Applications to Causality* (*Los Angeles*, *CA*, 1994). *Lecture Notes in Statist.* **120** 69–117. Springer, New York. MR1601279

ROBINS, J. M. (2004). Optimal structural nested models for optimal sequential decisions. In *Proceedings of the Second Seattle Symposium in Biostatistics*. *Lecture Notes in Statist.* **179** 189–326. Springer, New York. MR2129402

RUBIN, D. B. (1978). Bayesian inference for causal effects: The role of randomization. *Ann. Statist.* **6** 34–58. MR0472152

RUSH, A. J., FAVA, M., WISNIEWSKI, S. R., LAVORI, P. W., TRIVEDI, M. H., SACKEIM, H. A., THASE, M. E., NIERENBERG, A. A., QUITKIN, F. M., KASHNER, T. M. et al. (2004). Sequenced treatment alternatives to relieve depression (STAR*D): Rationale and design. *Controlled Clinical Trials* **25** 119–142.

SONG, R., WANG, W., ZENG, D. and KOSOROK, M. R. (2015). Penalized Q-learning for dynamic treatment regimens. *Statist. Sinica* **25** 901–920.

WATKINS, C. J. (1989). Learning from delayed rewards. Ph.D. thesis, Univ. Cambridge, England.

WATKINS, C. J. and DAYAN, P. (1992). Q-learning. *Mach. Learn.* **8** 279–292.

YAN, X. (2004). Test for qualitative interaction in equivalence trials when the number of centres is large. *Stat. Med.* **23** 711–722.

ZHANG, B., TSIATIS, A. A., LABER, E. B. and DAVIDIAN, M. (2012a). A robust method for estimating optimal treatment regimes. *Biometrics* **68** 1010–1018. MR3040007

ZHANG, B., TSIATIS, A. A., DAVIDIAN, M., ZHANG, M. and LABER, E. (2012b). Estimating optimal treatment regimes from a classification perspective. *Stat.* **1** 103–114.

ZHANG, B., TSIATIS, A. A., LABER, E. B. and DAVIDIAN, M. (2013). Robust estimation of optimal dynamic treatment regimes for sequential treatment decisions. *Biometrika* **100** 681–694. MR3094445

ZHAO, Y., ZENG, D., SOCINSKI, M. A. and KOSOROK, M. R. (2011). Reinforcement learning strategies for clinical trials in nonsmall cell lung cancer. *Biometrics* **67** 1422–1433. MR2872393

ZHAO, Y., ZENG, D., RUSH, A. J. and KOSOROK, M. R. (2012). Estimating individualized treatment rules using outcome weighted learning. *J. Amer. Statist. Assoc.* **107** 1106–1118. MR3010898

DEPARTMENT OF STATISTICS
NORTH CAROLINA STATE UNIVERSITY
2311 STINSON DR.
RALEIGH, NORTH CAROLINA 27695 USA
E-MAIL: afan@ncsu.edu
lu@stat.ncsu.edu
rsong@ncsu.edu